Neonatal EEG Classification using Atomic Decomposition

Sunil Belur Nagaraj
BE, ME

NATIONAL UNIVERSITY OF IRELAND, CORK

School of Engineering

Department of Electrical and Electronic Engineering
Irish Centre for Fetal and Neonatal Translational Research

Thesis submitted for the degree of Doctor of Philosophy

21 August 2015

Supervisors: Dr. Gordon Lightbody
Prof. Liam Marnane
Prof. Geraldine Boylan
Dr. Nathan Stevenson
Head of Department/School: Prof. Nabeel A. Riza

Research supported by Science Foundation Ireland

I, Sunil Belur Nagaraj, certify that this thesis is my own work and I have not obtained a degree in this university or elsewhere on the basis of the work submitted in this thesis.

Sunil Belur Nagaraj

To my ***parents***, my dear wife, ***Sowmya*** and our daughter, ***Shraddha***.....

Acknowledgements

This thesis was possible due to the guidance and knowledge provided by my team members in Neonatal Brain Research Group, help from my friends in Ireland and continuous support provided by my family members.

I would like to record my gratitude to my academic supervisor Dr. Gordon Lightbody for his constant supervision, advice, and guidance from the very early stage of this research as well as giving me extraordinary experiences throughout the work. Above all, and the most needed, he provided me unflinching encouragement and support in various ways.

This thesis would not have been possible without the help of Dr. Nathan Stevenson who offered suggestions for successful completion of this thesis. His scientific intuitiveness has created several ideas and passions in science, which exceptionally inspired and enriched my growth as a student, a researcher, and a scientist. He has played a major role in helping me understand the concept of time-frequency analysis techniques. In addition, the proposed methods in this project were inspired by his publications.

I have also benefited by advice and guidance from Dr. Liam Marnane who always kindly granted me his time to answer questions about signal processing techniques. I would also like to thank Drs. Andrey Temko and John O'Toole for their valuable suggestions related to machine learning and time-frequency analysis techniques, respectively.

Next, I would like to thank Prof. Geraldine Boylan who gave me an opportunity to work with this wonderful interdisciplinary team in Neonatal Brain Research Group at University College Cork and has given valuable suggestions into the medical applications of this research.

Many thanks to my friends in Cork Harlequins Cricket Club who are responsible for several memorable events during my stay in Ireland and always helped me to relieve the stress during my research.

Last, but not least, I would like to thank my parents and my wife, Sowmya for their patience, understanding and continuous encouragement which allowed me to spend more time with the thesis.

Abstract

The electroencephalogram (EEG) is an important noninvasive tool used in the neonatal intensive care unit (NICU) for the neurologic evaluation of the sick newborn infant. It provides an excellent assessment of at-risk newborns and formulates a prognosis for long-term neurologic outcome. In general, the examination of EEG data requires specific expertise and is performed by visual interpretation of the neonatal EEG by an experienced neurophysiologist or pediatric neurologist. However this expertise is not readily available on a 24hr basis in most NICUs. For this reason, methods for the automated analysis of neonatal EEG data in the NICU can provide valuable information to the clinician facilitating medical intervention.

Recent advances in signal processing and machine learning techniques have led to the development of automated analysis of EEG data. These techniques have reduced the EEG analysis and interpretation time from hours to minutes, thereby reducing the burden on neurophysiologists. However, typical EEG signals are contaminated by extra-cerebral sources known as artefacts and the separation of actual brain electrical activity from extraneous artefacts is critical to the interpretation of the neonatal EEG which significantly distort the EEG signal. Typically these artefacts can be grouped into four broad categories: environment (power-line noise, radio transmitters etc), instrumentation-patient interface (electrode contacts, head positioning etc), instrumentation (movement of electrode leads), and physiologic potentials of noncerebral origin (electrocardiogram, electrooculogram, etc). These artefacts makes the interpretation of the EEG difficult, thereby reducing the performance of the automated EEG analysis.

The aim of this thesis is to develop a system for automatic classification of neonatal EEG which can be mainly divided into two parts: (1) classification of neonatal EEG seizure from nonseizure, and (2) classifying neonatal background EEG into several grades based on the severity of the injury.

Atomic decomposition techniques use redundant time-frequency dictionaries for sparse signal representations or approximations. Recently, this technique had gained importance in the field of EEG (adult and neonatal) classification. In this thesis the performance of three decomposition techniques, matching pursuit, orthogonal matching pursuit and basis pursuit was investigated for neonatal EEG classification. It was shown that orthogonal matching pursuit provides best classification at a given decomposition level. Due to this, the orthogonal matching pursuit was used as a preferred decomposition technique throughout this thesis.

The first novel contribution of this thesis is the development of a novel time-frequency dictionary coherent with the neonatal EEG seizure states. This dictionary was able to track the time-varying nature of the EEG signal. It was shown that by using atomic decomposition (AD) and the proposed novel dictionary, the neonatal EEG transition from nonseizure to seizure states could be detected efficiently. A new measure known as *relative structural complexity* was developed to detect this transition.

The second novel contribution of this thesis is the development of a neonatal seizure detection algorithm (NSDA) using several time-frequency features from the proposed novel dictionary. It was shown that the time-frequency features obtained from the atoms in the novel dictionary improved the seizure detection accuracy when compared to that obtained from the raw EEG signal. An AD-based NSDA using support vector machines was proposed which provided better seizure detection accuracy at clinically relevant false detections per hour.

The third novel contribution of this thesis is a system for automatic grading of EEG of term neonates with hypoxic-ischemic encephalopathy (HIE). With the assistance of a supervised multiclass SVM classifier and several time frequency features, several methods to automatically grade EEG were explored. In particular, the Hilbert transform and atomic decomposition methods were compared and the results were analyzed in detail. Additionally, a novel AD artefact detector is proposed to improve the performance of the automatic grading system (AGS). The proposed AGS was shown to obtain state-of-the-art performance of 87% by using only 8 time-frequency features.

In summary, the novel techniques proposed in this thesis contribute to the application of advanced signal processing techniques for automatic assessment of neonatal EEG recordings.

Contents

List of Figures

List of Tables

Acronyms

AD	Atomic decomposition
ADS	Artefact detection system
AGS	Automatic grading system
AED	Anti epileptic drug
AM	Amplitude modulation
ASI	Activation synchrony index
AUC	Area under the curve
BP	Basis pursuit
CNS	Central nervous system
CA	Conceptional age
ECoG	Electrocorticogram
EEG	Electroencephalogram
FD	False detection
FT	Fourier transform
FFT	Fast Fourier transform
GDR	Good detection rate
HIE	Hypoxic-ischemic encephalopathy
IBI	Inter-burst interval
IF	Instantaneous frequency
IQR	Inter-quantile range
LFM	Linear frequency modulation
LOO	Leave one out
MP	Matching pursuit
MMP	Multi-channel matching pursuit
MBD	Modified bilinear distribution
NICU	Neonatal intensive care unit
NSDA	Neonatal seizure detection algorithm
OMP	Orthogonal matching pursuit
PTE	Percentage time error
RSC	Relative structural complexity
ROC	Receiver-operator characteristic curve
RBF	Radial basis function
TFA	Time-frequency analysis
TFD	Time-frequency distribution
QTFD	Quadratic time-frequency distribution
SVM	Support vector machine
SDA	Seizure detection algorithm
SDR	Seizure detection rate
SPWVD	Smooth-pseudo Weigner-ville distribution
SVD	Singular value decomposition
SER	Signal-to-error ratio
TF	Time-frequency
TFR	Time-frequency representation
WT	Wavelet transform
WVD	Weigner-ville distribution

Chapter 1

Introduction

> We can't solve problems by using the same kind of thinking we used when we created them. ––Albert Einstein

1.1 Introduction

The necessity to monitor the functionality of the brain is an important issue in neonatal intensive care units (NICUs). It is essential to track the progress of the brain injury caused by insufficient supply of oxygen to the brain which can lead to hypoxic ischaemic encephalopathy (HIE). Based on the diagnosis, the clinicians can make decisions about the treatment given to the baby with such complications.

Compared to modern technologies such as computed tomography (CT) scan or magnetic resonance imaging (MRI), which only provide the image of the brain structure at a given instant, it is essential to have a system that can continuously monitor the functioning of the brain. Electroencephalography (EEG) is a non-invasive way to monitor brain activities and is the gold standard technology in NICUs (Boylan et al., 2013, Niedermeyer and da Silva, 2005). Several researchers have shown that the functional state of the brain can be deduced by studying the EEG signals (Marret et al., 1997, Watanabe et al., 1999). Due to this, the EEG is more suitable for long term monitoring of patients and for viewing variation in neurological state such as seizures and sleep state analysis.

EEG signals are typically of the order of microvolts (μV) and are usually measured using metal electrodes placed on the patient's scalp. The electrical activities of

the neuronal cells in the brain generate a potential difference between the scalp electrodes which are recorded in the EEG (Fisch, 1999). Several recent studies have shown that the information obtained from the EEG signal can be used to detect and predict abnormalities in the brain such as seizures (Mathieson et al., 2015), HIE (Ahmed et al., 2014, Löfgren, 2005, Stevenson et al., 2013) etc.

In NICUs, most clinical neurophysiologists evaluate the EEG recordings through visual inspection which provides intermittent evaluations, and cannot be used for a continuous bedside monitoring. Moreover visual interpretation of the EEG is time-consuming and not all NICUs have 24hr access to experienced neurophyiologists. Amplitude-integrated EEG (aEEG) is commonly used to monitor the brain activity of a patient in NICUs which provides a single or double channel EEG on a smaller time scale. However, it has been shown that this method is not robust since the presence of artefacts in the compressed time scale can be misinterpreted as activity in the brain (Rosén, 2006) and can be less sensitive to low amplitude seizure events (Toet et al., 2002).

Due to these limitations, it is necessary to develop automated methods for the analysis and classification of activities in neonatal EEG. The method should be robust enough for use in a NICU environment which can provide information about the real-time conditions of the brain. The clinician should be able to interpret the information obtained from the method for making a decision about a treatment or to adjust the dosage of a medication. This long term neural activity monitoring will help to track the development of the status of the patient over time which enables the clinicians to adjust their treatments actively.

1.2 Aim and scope of this thesis

EEG is the recording that represents the electrical activity of the brain using sensors placed on the scalp. The electrical activity recorded by EEG can be classified as normal or abnormal. The EEG has high time resolution (in the order of ms) and is noninvasive which means there is no pain caused to the patient. Various activity levels are indicated by the components present in EEG. There are several applications of EEG which are primarily used for the diagnosis of sleep disorders and also to study the effect of drugs on the functionality of the brain.

Several neonatal EEG analysis algorithms reported in the literature either use

a small dataset of artefact free neonatal EEG or simulated or selective datasets and report their performance with no apparent validation of results. Moreover, the lack of datasets that can best represent the real-time situation in the NICU also hinders the development of automated algorithms. In order to effectively use automated neonatal EEG analysis systems in real-time, the algorithm must be *robust* enough to deal with the variety of EEG signals that may arise in noisy, medical environments. Here, robustness refers to the ability of the algorithm to detect a signal of interest with minimum false detections per hour.

The main objective of this thesis is to classify neonatal EEG using *Atomic Decomposition* (AD) which is a *Time-frequency* signal analysis technique. Since the newborn EEG is nonstationary and nonlinear, nonstationary signal analysis is necessary to capture the time-varying statistical properties of the signal. It has been shown that AD is well suited for the analysis of EEG signals. The atoms selected during the decomposition provide meaningful information about the underlying signal. The main advantage of using AD is that an application based dictionary can be designed. In addition, AD techniques provide high resolution time-frequency representation. It can also track time-varying properties of EEG (Durka et al., 2005, Jouny et al., 2003, Rankine et al., 2007).

Automated analysis of neonatal EEG is an important problem in the NICUs. In NICUs, neonatal seizures and HIE due to perinatal asphyxia occur at a rate of 1-3 per 1000 births (Glass et al., 2009). In order to have a bed-side neonatal EEG analysis system, the system must be able to transfer the information real-time to the neurophysiologists in a simpler way. Early detection of the brain abnormalities in neonates is essential to prevent death or long term neurological disabilities. In addition, the automated system can also be used to diagnose and monitor the effect of anti-epileptic drugs given to the patients. A robust, reliable and automated neonatal EEG analysis is thus essential to aid as a tool to neurophysiologits for diagnosing neonatal brain functionality.

This thesis is focussed on developing advanced nonstationary signal processing features for automated analysis of neonatal EEG. Classification of the neonatal EEG is mainly divided into two parts:

1. **Seizure detection** - Classification of seizure from nonseizure/background EEG, and
2. **Hypoxic-ischaemic encephalopathy** (HIE) grading - Classification of HIE using EEG based on severity of the brain injury.

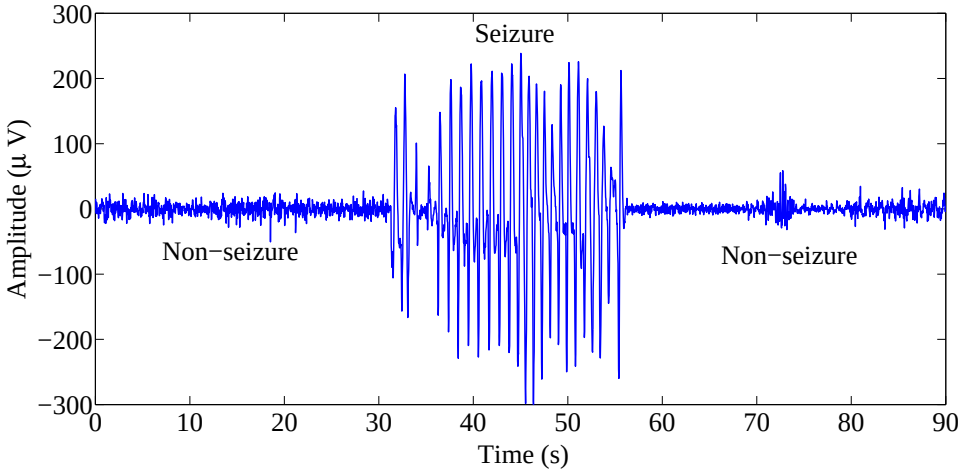


Figure 1.1: Example of a newborn EEG with an obvious seizure event.

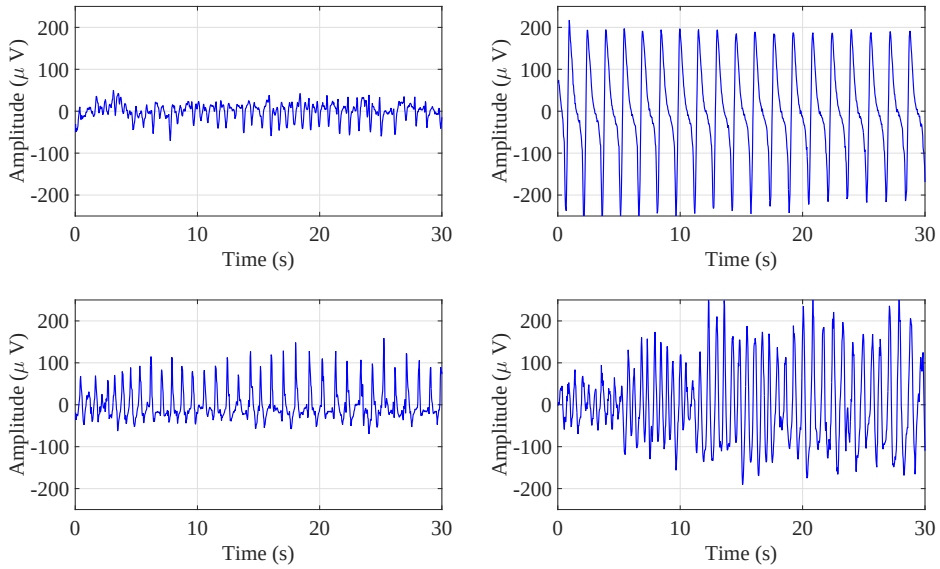


Figure 1.2: Example of different morphologies of newborn EEG seizure.

In *seizure detection*, the primary aim is to design a novel time-frequency dictionary and develop a neonatal seizure detection algorithm using atomic decomposition which can significantly improve the performance of previously reported neonatal seizure detectors. In this methodology, the activities of fixed length epochs (or segments) of EEG are quantified using a set of features obtained using atomic decomposition. The set of features obtained are then classified using machine learning techniques. An example of an EEG segment with a seizure event is shown in figure 1.1. Examples of several seizure morphologies (artefact free) across different patients is shown in figure 1.2 clearly indicating the rhythmic characteristics of neonatal EEG seizure activity with time varying amplitude

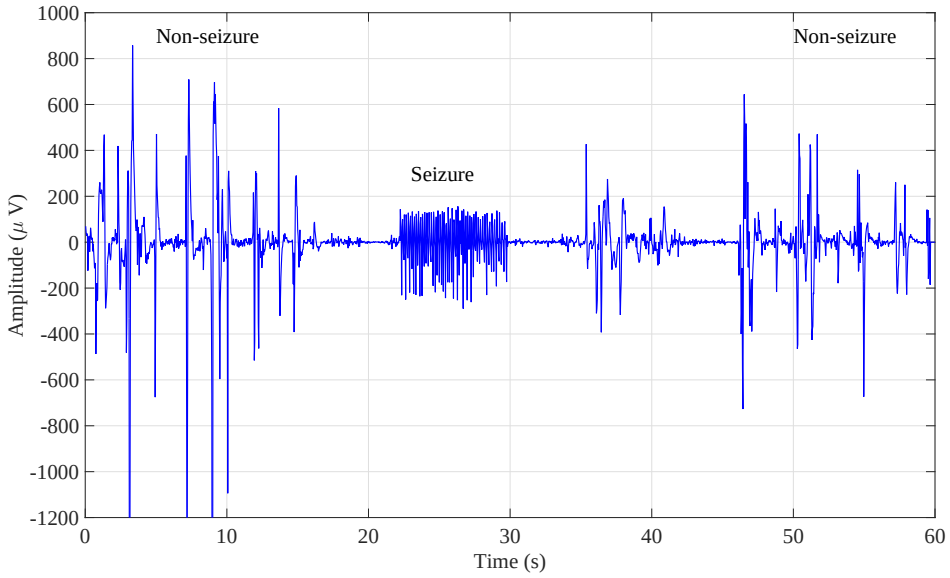


Figure 1.3: Example of a newborn EEG with a seizure event and several artefacts in nonseizure.

and frequencies. In this case (figure 1.2), the seizure events can be easily distinguished from non-seizure due to the absence of artefacts. However figure 1.3 shows a sample EEG with several artefacts which makes it difficult to identify the seizure event.

HIE grading, involves the grading of the EEG into four categories based on the severity of the brain injury or HIE (Murray et al., 2009). In clinical practice, both seizure detection and grading HIE requires the presence of highly qualified neurophysiologists and this expertise is not widely available in all NICUs. Since 1992, automated neonatal seizure detection has been developed, however none of them are robust enough for the clinical use (Faul et al., 2005). Automatic grading of HIE using EEG is relatively a new area and is being developed since 2011 (Korotchikova et al., 2011). Examples of the different grades of clean EEG based on the severity of HIE are shown in figure 1.4. These are typically corrupted by artefacts (shown in figure 1.5) and represent a difficult classification problem to automate.

1.3 Literature review

This section reviews some of the algorithms and their accuracy for the detection of neonatal seizures and EEG based HIE grading.

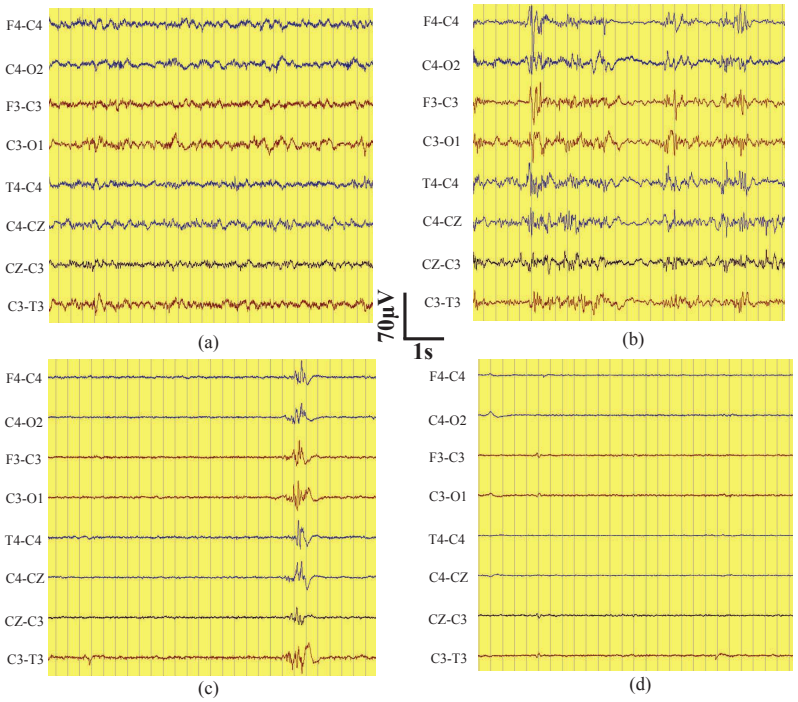


Figure 1.4: Sample 8-channel 30s ideal EEG epochs showing different grades of HIE (a) Grade 1: Normal/Mild abnormalities. (b) Grade 2: Moderate abnormalities. (c) Grade 3: Major abnormalities. (d) Grade 4: Inactive.

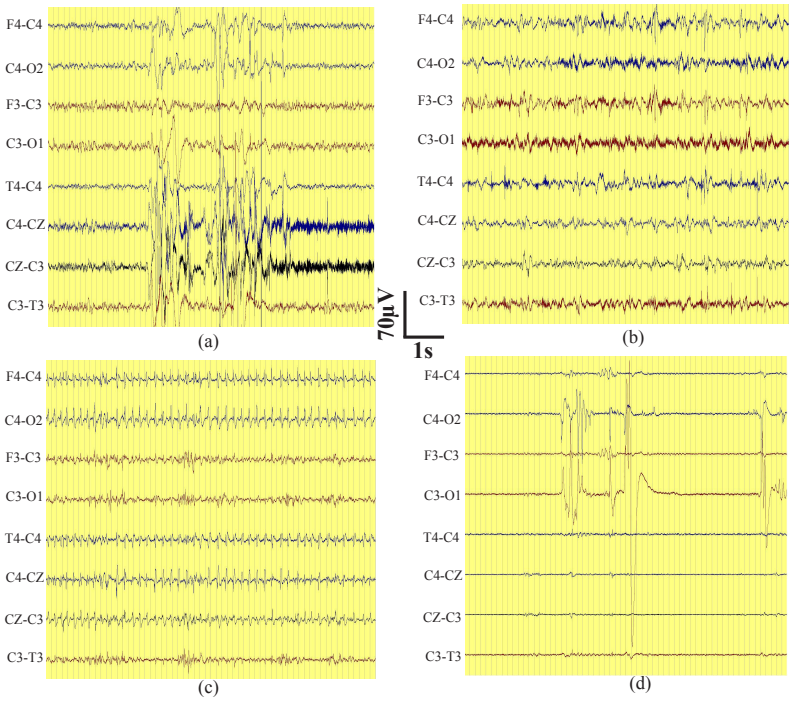


Figure 1.5: Sample 8-channel 60s noisy EEG epochs showing different grades of HIE (a) Grade 1: Normal/Mild abnormalities. (b) Grade 2: Moderate abnormalities. (c) Grade 3: Major abnormalities. (d) Grade 4: Inactive.

1.3.1 Neonatal seizure detection

Several neonatal seizure detection methods have been proposed using many signal processing techniques which can be divided into two main categories - *stationary* and *nonstationary*. Since neonatal EEG is nonstationary, the performance of the methods using stationary techniques can lead to suboptimal performance (Boashash and Mesbah, 2003). Only some of the techniques relevant to this thesis are reviewed in this section.

1.3.1.1 Initial methods using stationary techniques

Liu et al. first proposed a method for detecting seizures in neonates (Liu et al., 1992). In this method, the periodic or rhythmic discharges of the electrographic seizures in the newborn which are distinctively different from the normal background cerebral activity were utilised to identify the electrographic seizure activity. Autocorrelation analysis was used for detecting seizures using a novel scoring system known as scored autocorrelation moment (SAM) analysis on a dataset of approximately 1h of EEG (from 9 neonates in total). SAM analysis showed a sensitivity of 84% and a specificity of 98%.

The method proposed by Gotman et al. primarily used frequency domain characteristics from the frequency spectrum of the EEG (Gotman et al., 1997) to detect neonatal seizures. The method was developed using EEGs obtained from 55 newborns, recorded at 3 hospitals (a total of 281 h of recordings containing 679 seizures). An accuracy of 78% was reported at a false detection rate of 1.7/h. Celka et al. proposed a seizure detector based on a model of the background EEG (Celka and Colditz, 2002a). Singular values corresponding to the complexity of the signal were used to determine the presence of seizure in the EEG. An accuracy of 93% was reported at a false detection rate of 4/h. However, based on a study by Faul et al., the performance of these methods was found not suitable in NICU (Faul et al., 2005) and the authors proposed the necessity for more advanced features and classification algorithms.

A major problem with these methods is the assumption of stationarity. Since the newborn EEG signal is nonstationary, the poor performance of these methods is due to the overlap of frequency characteristics between classes which may result in false detections even during the absence of seizures (Boashash and Mesbah, 2001, Rankine et al., 2005). Several model based approaches have also been proposed in literature which is primarily based on the model of the newborn

EEG (Celka and Colditz, 2002b, Da Silva et al., 1974, Roessgen et al., 1998). The drawback of these methods is that the preprocessing stage uses a time-invariant model of the neonatal EEG for whitening the signal. Since neonatal EEG is a dynamic nonstationary signal, some of the sections of background EEG may not be whitened properly and can lead to false detections.

1.3.1.2 Methods using nonstationary techniques

The wavelet transform (WT) is a popular nonstationary signal processing technique that separates a transient signal into different frequency scales. Several seizure detection algorithms have been proposed using the features obtained from the WT (Nagasubramanian et al., 1997, Zarjam et al., 2003). These methods were trained and tested on a small database and since EEG patterns vary significantly, their performance is yet to be assessed on large recordings of EEG.

A number of seizure detection systems have been proposed using time-frequency representation (TFR) analysis of neonatal EEG. Due to the time-varying frequency content of the EEG signal, this representation has an advantage of tracking the variation of frequency over time. The quadratic time-frequency distribution (QTFD) is a popular TFR method that has been used in development of seizure detection algorithms[1].

A seizure detection method based on detecting repetitive spike activity in a high frequency region of QTFD was proposed in (Hassanpour et al., 2004). It was shown that the singular values and singular vectors obtained from the singular value decomposition (SVD) of the TFR using QTFD provide useful information about seizures and can be used in seizure detection in Hassanpour et al. (Hassanpour et al., 2003). However, these methods require further validation as only a small dataset was used to test the algorithm (50 minutes of EEG from 5 patients).

Several seizure detection methods for adult EEG using atomic decomposition (AD) have been proposed (Bergey and Franaszczuk, 2001, Jouny et al., 2003). The main advantages of AD are that it can provide an artefact free, high resolution TFR (Blinowska and Durka, 2001) and also track the nonlinearity of the signal (Bergey and Franaszczuk, 2001). The parametric form of the signal can also be obtained using AD (Durka, 2004).

Rankine et al. demonstrated that AD using matching pursuit (MP) can be used to

[1] more details about TFR and QTFD are given in chapter 3

detect changes in the signal states of the EEG (Rankine et al., 2007). Recently, AD was used for the detection of neonatal EEG seizure (Khlif et al., 2008). The performance of these methods was based on optimum thresholds and is not practical as it requires apriori knowledge of the data. Moreover, their performance was validated on a small artefact free data set and their performance is yet to be reassessed in the presence of artefacts. The summary of the performance of some of the recent advanced classifier based neonatal seizure detection methods is given in Table 1.1 and are used in thesis for the purpose of comparison.

Table 1.1: Performance of some of the neonatal seizure detection methods. (* Authors use database obtained from Cork University Maternity Hospital which is also used in this thesis).

Authors	No. of features	Dataset (hours)	Seizure detection rate(%)	False detections per hour	No. of neonates
Navakatikyan et al. (Navakatikyan et al., 2006)	1	24.4	87	2	55
Greene et al. (Greene et al., 2007)	12	154	81.4	3.15	17
Deburchgraeve et al. (Deburchgraeve et al., 2008)	1	217	75	0.66	26
Mitra et al. (Mitra et al., 2009)	8	121	80	0.74	76
Temko et al.* (Temko et al., 2011)	55	268	89	1.0	17
Stevenson et al.* (Stevenson et al., 2012)	3	826	79	1.0	18
Mathieson et al. (Mathieson et al., 2015)	55	4060	75	0.36	70

1.3.2 Hypoxic-ischaemic encephalopathy grading

Hypoxic-ischaemic encephalopathy (HIE) is caused due to the lack of oxygen to the neonatal brain during birth and is the most common cause of long-term neonatal neurological dysfunction (Vannucci, 2000) or sometimes even neonatal deaths with an incidence of 2-3 per 1000 births (Graham et al., 2008). It was shown that the visual interpretation of the background EEG is a useful tool in monitoring the recovery of brain activity after HI injury (Murray et al., 2009, Toet et al., 1999). Some of the EEG characteristics such as amplitude, frequency content, sharp waves, symmetry and synchrony measures are visualised to classify the background EEG as normal or abnormal. The abnormal EEG is further graded into four types based on the degree of abnormality or severity of the brain injury.

A first attempt to classify HIE-EEG based on the quantitative analysis of background EEG was investigated by Korotchikova et al. (Korotchikova et al., 2011). In this approach, 9 quantitative EEG (qEEG) measures that replicate the neurophysiologist's approach of visually classifying neonatal EEG were used to grade HIE. The authors reported a classification accuracy of 91% on a dataset consisting of recordings of 54 full term neonates with HIE (approximately one hour long EEG selected from each recording). The data was recorded in the NICU of Cork University Maternity Hospital, Cork, Ireland[2]. However, the results obtained were estimated on artefact free training data and their performance dropped to 72% when validated on a dataset with artefacts.

Stevenson et al. extended this study by using features obtained from the background EEG model (Stevenson et al., 2013). An automated grading system (AGS) using 15 features from the model of background EEG was proposed in this study. The AGS was trained on artefact free data and tested on unseen data with artefacts. The AGS provided an accuracy of 83.3% and it was reported that the majority of misclassifications were due to the presence of artefacts and several abnormal patterns in the EEG such as sharp waves, asymmetry and asynchrony.

Ahmed et al. proposed an AGS based on a cross disciplinary method of using support vector machine and supervectors (Ahmed et al., 2014). This system used a total of 55 features which provided a generic EEG description computed from the time, frequency and information theory domains. The AGS was trained and tested on data with artefacts and an accuracy of 85% was reported. The dataset recorded in the NICU of Cork University Maternity Hospital, Cork, Ireland has been used in all these methods (Ahmed et al., 2014, Korotchikova et al., 2011, Stevenson et al., 2013).

1.4 Thesis contributions

The following *novel contributions* are presented in this thesis:

1. A novel time-frequency (TF) dictionary consisting of atoms coherent with the newborn EEG seizure was developed. Using this TF dictionary and a novel signal complexity measure referred as *relative structural complexity*, an *atomic decomposition* based neonatal seizure detection algorithm that could classify seizure from nonseizure states was developed. (*Chapter 4*)

[2]Same dataset is used in this thesis

2. A robust neonatal seizure detection algorithm using nonstationary TF analysis techniques. Several TF features from the novel dictionary were derived and improved the performance of NSDA using support vector machine. (*Chapter 5*)

3. A novel *atomic decomposition* based automated HIE grading system was developed. This also includes a novel *atomic decomposition* based artefact detector. (*Chapter 6*)

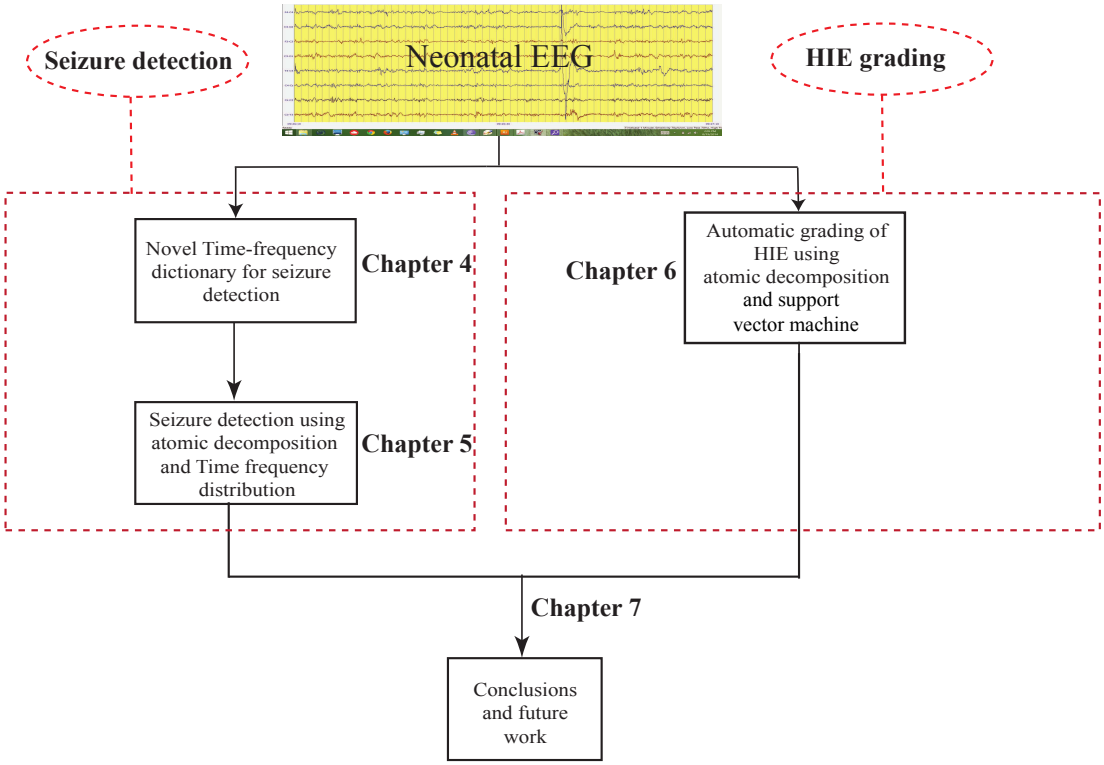


Figure 1.6: Block diagram of the major contributions of the thesis.

1.5 Outline of thesis

The thesis is organized into the following chapters:

Chapter 1 describes the aim and scope of this thesis and also reviews some earlier techniques for neonatal seizure detection and HIE grading. Some of their limitations are also described. The objectives and major contributions of this thesis are also described in this chapter.

Chapter 2 provides an introduction to the characteristics of neonatal EEG signal. The characteristics and morphology of the EEG signal in terms of both

clinical and signal processing perspectives are summarized.

Chapter 3 introduces several time-frequency methods for analyzing neonatal EEG signals. This includes Fourier and wavelet transforms, atomic decomposition (AD) and quadratic time-frequency distribution techniques (QTFD). The analysis of the EEG signals using these methods is demonstrated. Some advantages of atomic decomposition methods over traditional stationary analysis methods is also addressed.

Chapter 4 proposes a novel TF dictionary (referred to as the pseudo-periodic Duffing dictionary) that is coherent with neonatal EEG seizure structures. This TF dictionary uses a Duffing oscillator based neonatal EEG seizure model to generate the atoms. The performance of several atomic decomposition methods to discriminate seizure from nonseizure are compared. It was concluded that the orthogonal matching pursuit provides best discrimination and is used in this thesis.

Using a signal complexity measure referred as *relative structural complexity*, an atomic decomposition neonatal seizure detection algorithm (NSDA) is developed. The decomposition techniques use complexity measure from the TF dictionary to detect seizure in EEG signal. The dataset employed throughout this thesis for seizure detection is detailed and the performance metrics and validation techniques are explained. The patterns in the EEG contributing to false alarms and the properties of small duration seizures which are not detected by the NSDA were also analyzed. The performance of the proposed NSDA is compared with the performance of several state-of-the-art TF dictionaries and also neonatal seizure detectors.

Chapter 5 proposes a support vector machine (SVM) based NSDA using time-frequency features obtained from the pseudo-periodic Duffing dictionary. The robustness of TF features for neonatal seizure detection was studied in this chapter. The proposed NSDA uses several QTFD based TF features for seizure detection using SVM. It is shown that the TF features obtained from pre-processed EEG signal provide poor performance at clinically relevant false detection thresholds in the larger dataset used in this thesis. Alternatively, an AD based NSDA using TF features obtained from the novel pseudo-periodic Duffing dictionary was proposed which provided improved performance when compared to the TF features obtained directly from the EEG signal.

The performance of the proposed NSDA was also compared with several state-of-

the-art NSDA's. It was concluded that the TF features obtained from the QTFD of the pseudo-periodic Duffing dictionary were robust and could be incorporated in the NSDA using larger feature-set used proposed in literature.

Chapter 6 introduces a novel AD based automated HIE-EEG grading system (AGS) using SVM. The proposed AGS uses several amplitude and frequency features obtained from the QTFD of the EEG signal. Several state-of-the art methods to extract these TF features were also studied and their performance is reported in this chapter.

It was observed that the presence of artefacts significantly reduced the performance of the AD based AGS. A novel AD based patient independent artefact detection system is proposed to improve the performance of the proposed AD based AGS. Only the artefacts that were $\geq$ 16s in duration were considered in this system since their contribution was higher for the misclassifications by the AGS. A multiclass SVM was used in this study for classification of HIE-EEG. Several limitations of AD based methods to automatically grade HIE-EEG are also discussed in this chapter.

Chapter 7 summarises the main findings of this thesis and also provide a brief overview of individual chapters. Several limitations of the proposed methods in the thesis are discussed and some future research directions are proposed.

1.6 Publications arising from the thesis

Journal:

1. **Nagaraj, S.B.**, Stevenson, N.J., Marnane, W.P., Boylan, G.B., Lightbody, G., "Neonatal seizure detection using atomic decomposition with a novel dictionary", IEEE Transactions on Biomedical Engineering, vol. 61, no. 11, pp. 2724–2732, Nov. 2014.

2. **Nagaraj, S.B.**, Stevenson, N.J., Marnane, W.P., Boylan, G.B., Lightbody, G., "Neonatal seizure detection using atomic decomposition and time-frequency distribution features from a novel dictionary" to be submitted to Medical Engineering & Physics.

3. **Nagaraj, S.B.**, Stevenson, N.J., Marnane, W.P., Boylan, G.B., Lightbody, G., "Automated grading of Neonatal EEG with Hypoxic ischaemic

encephalopathy using Atomic decomposition" to be submitted to Medical Engineering & Physics.

Conference:

1. **Nagaraj, S.B.**, Stevenson, N.J., Marnane, W.P., Boylan, G.B., Lightbody, G., "Robustness of Time Frequency Distribution based Features for Automated Neonatal EEG Seizure Detection," Engineering in Medicine and Biology Society (EMBC), 2014 Annual International Conference of the IEEE, pp. 2829–2832, Aug. 2014.

2. **Nagaraj, S.B.**, Stevenson, N.J., Marnane, W.P., Boylan, G.B., Lightbody, G., "A novel dictionary for neonatal EEG seizure detection using atomic decomposition," Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE , pp. 1073–1076, Aug. 2012.

References

Ahmed, R., Temko, A., Marnane, W., Boylan, G., Lightbody, G., 2014. Grading brain injury in neonatal EEG using svm and supervector kernel. In: Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. IEEE, pp. 5894–5898.

Bergey, G. K., Franaszczuk, P. J., 2001. Epileptic seizures are characterized by changing signal complexity. Clinical Neurophysiology 112 (2), 241–249.

Blinowska, K. J., Durka, P. J., 2001. Unbiased high resolution method of EEG analysis in time-frequency space. Acta Neurobiologiae Experimentalis 61 (3), 157–174.

Boashash, B., Mesbah, M., 2001. A time-frequency approach for newborn seizure detection. Engineering in Medicine and Biology Magazine, IEEE 20 (5), 54–64.

Boashash, B., Mesbah, M., 2003. Time-frequency methodology for newborn electroencephalographic seizure detection. Applications in Time-Frequency Signal Processing 9, 339–369.

Boylan, G. B., Stevenson, N. J., Vanhatalo, S., 2013. Monitoring neonatal seizures. In: Seminars in Fetal and Neonatal Medicine. Vol. 18. Elsevier, pp. 202–208.

Celka, P., Colditz, P., 2002a. A computer-aided detection of EEG seizures in infants: a singular-spectrum approach and performance comparison. Biomedical Engineering, IEEE Transactions on 49 (5), 455–462.

Celka, P., Colditz, P., 2002b. Nonlinear nonstationary wiener model of infant EEG seizures. Biomedical Engineering, IEEE Transactions on 49 (6), 556–564.

Da Silva, F. L., Hoeks, A., Smits, H., Zetterberg, L., 1974. Model of brain rhythmic activity. Kybernetik 15 (1), 27–37.

Deburchgraeve, W., Cherian, P., De Vos, M., Swarte, R., Blok, J., Visser, G. H., Govaert, P., Van Huffel, S., 2008. Automated neonatal seizure detection mimicking a human observer reading EEG. Clinical Neurophysiology 119 (11), 2447–2454.

Durka, P. J., 2004. Adaptive time-frequency parametrization of epileptic spikes. Physical Review E 69 (5), 051914.

Durka, P. J., Matysiak, A., Montes, E. M., Sosa, P. V., Blinowska, K. J., 2005. Multichannel matching pursuit and EEG inverse solutions. Journal of Neuroscience Methods 148 (1), 49–59.

Faul, S., Boylan, G., Connolly, S., Marnane, L., Lightbody, G., 2005. An evaluation of automated neonatal seizure detection methods. Clinical Neurophysiology 116 (7), 1533–1541.

Fisch, B. J., 1999. Fisch and spehlmann's EEG primer. Elsevier, Amsterdam, ed 3, 301–303.

Glass, H. C., Pham, T. N., Danielsen, B., Towner, D., Glidden, D., Wu, Y. W., 2009. Antenatal and intrapartum risk factors for seizures in term newborns: a population-based study, california 1998-2002. The Journal of Pediatrics 154 (1), 24–28.

Gotman, J., Flanagan, D., Zhang, J., Rosenblatt, B., 1997. Automatic seizure detection in the newborn: methods and initial evaluation. Electroencephalography and Clinical Neurophysiology 103 (3), 356–362.

Graham, E. M., Ruis, K. A., Hartman, A. L., Northington, F. J., Fox, H. E., 2008. A systematic review of the role of intrapartum hypoxia-ischemia in the causation of neonatal encephalopathy. American journal of obstetrics and gynecology 199 (6), 587–595.

Greene, B. R., Boylan, G. B., Reilly, R. B., de Chazal, P., Connolly, S., 2007. Combination of EEG and ECG for improved automatic neonatal seizure detection. Clinical Neurophysiology 118 (6), 1348–1359.

Hassanpour, H., Mesbah, M., Boashash, B., 2003. Enhanced time-frequency features for neonatal EEG seizure detection. In: Circuits and Systems, 2003.

ISCAS'03. Proceedings of the 2003 International Symposium on. Vol. 5. IEEE, pp. V–29.

Hassanpour, H., Mesbah, M., Boashash, B., 2004. Time–frequency based newborn EEG seizure detection using low and high frequency signatures. Physiological Measurement 25 (4), 935.

Jouny, C. C., Franaszczuk, P. J., Bergey, G. K., 2003. Characterization of epileptic seizure dynamics using Gabor atom density. Clinical Neurophysiology 114 (3), 426–437.

Khlif, M., Mesbah, M., Boashash, B., Colditz, P., 2008. Detection of neonatal EEG seizure using multichannel matching pursuit. In: Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE. IEEE, pp. 907–910.

Korotchikova, I., Stevenson, N., Walsh, B., Murray, D., Boylan, G., 2011. Quantitative EEG analysis in neonatal hypoxic ischaemic encephalopathy. Clinical Neurophysiology 122 (8), 1671–1678.

Liu, A., Hahn, J., Heldt, G., Coen, R., 1992. Detection of neonatal seizures through computerized EEG analysis. Electroencephalography and Clinical Neurophysiology 82 (1), 30–37.

Löfgren, N., 2005. The EEG of the Newborn Brain-Detection of Hypoxia and Prediction of Outcome. Chalmers University of Technology.

Marret, S., Parain, D., Ménard, J.-F., Blanc, T., Devaux, A.-M., Ensel, P., Fessard, C., Samson-Dollfus, D., 1997. Prognostic value of neonatal electroencephalography in premature newborns less than 33 weeks of gestational age. Electroencephalography and Clinical Neurophysiology 102 (3), 178–185.

Mathieson, S. R., Stevenson, N. J., Low, E., Marnane, W. P., Rennie, J. M., Temko, A., Lightbody, G., Boylan, G. B., 2015. Validation of an automated seizure detection algorithm for term neonates. Clinical Neurophysiology.

Mitra, J., Glover, J. R., Ktonas, P. Y., Kumar, A. T., Mukherjee, A., Karayiannis, N. B., Frost Jr, J. D., Hrachovy, R. A., Mizrahi, E. M., 2009. A multi-stage system for the automated detection of epileptic seizures in neonatal EEG. Journal of Clinical Neurophysiology: official publication of the American Electroencephalographic Society 26 (4), 218.

Murray, D. M., Boylan, G. B., Ryan, C. A., Connolly, S., 2009. Early EEG find-

ings in hypoxic-ischemic encephalopathy predict outcomes at 2 years. Pediatrics 124 (3), e459–e467.

Nagasubramanian, S., Onaral, B., Clancy, R., 1997. On-line neonatal seizure detection based on multi-scale analysis of EEG using wavelets as a tool. In: Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE. Vol. 3. IEEE, pp. 1289–1292.

Navakatikyan, M. A., Colditz, P. B., Burke, C. J., Inder, T. E., Richmond, J., Williams, C. E., 2006. Seizure detection algorithm for neonates based on wave-sequence analysis. Clinical Neurophysiology 117 (6), 1190–1203.

Niedermeyer, E., da Silva, F. L., 2005. Electroencephalography: basic principles, clinical applications, and related fields. Lippincott Williams & Wilkins.

Rankine, L., Hassanpour, H., Mesbah, M., Boashash, B., 2005. EEG simulation using fractal dimension analysis. In: Thirteen Iranian Conference on Electrical Engineering.

Rankine, L., Mesbah, M., Boashash, B., 2007. A matching pursuit-based signal complexity measure for the analysis of newborn EEG. Medical & Biological Engineering & Computing 45 (3), 251–260.

Roessgen, M., Zoubir, A. M., Boashash, B., 1998. Seizure detection of newborn EEG using a model-based approach. Biomedical Engineering, IEEE Transactions on 45 (6), 673–685.

Rosén, I., 2006. The physiological basis for continuous electroencephalogram monitoring in the neonate. Clinics in Perinatology 33 (3), 593–611.

Stevenson, N., Korotchikova, I., Temko, A., Lightbody, G., Marnane, W., Boylan, G., 2013. An automated system for grading EEG abnormality in term neonates with hypoxic-ischaemic encephalopathy. Annals of Biomedical Engineering 41 (4), 775–785.

Stevenson, N. J., O'Toole, J. M., Rankine, L. J., Boylan, G. B., Boashash, B., 2012. A nonparametric feature for neonatal EEG seizure detection based on a representation of pseudo-periodicity. Medical Engineering & physics 34 (4), 437–446.

Temko, A., Thomas, E., Marnane, W., Lightbody, G., Boylan, G., 2011. EEG-based neonatal seizure detection with support vector machines. Clinical Neurophysiology 122 (3), 464–473.

Toet, M., Hellström-Westas, L., Groenendaal, F., Eken, P., De Vries, L., 1999. Amplitude integrated EEG 3 and 6 hours after birth in full term neonates with hypoxic–ischaemic encephalopathy. Archives of Disease in Childhood-Fetal and Neonatal Edition 81 (1), F19–F23.

Toet, M. C., van der Meij, W., de Vries, L. S., Uiterwaal, C. S., van Huffelen, K. C., 2002. Comparison between simultaneously recorded amplitude integrated electroencephalogram (cerebral function monitor) and standard electroencephalogram in neonates. Pediatrics 109 (5), 772–779.

Vannucci, R. C., 2000. Hypoxic-ischemic encephalopathy. American Journal of Perinatology 17 (03), 113–120.

Watanabe, K., Hayakawa, F., Okumura, A., 1999. Neonatal EEG: a powerful tool in the assessment of brain damage in preterm infants. Brain and Development 21 (6), 361–372.

Zarjam, P., Mesbah, M., Boashash, B., 2003. Detection of newborn EEG seizure using optimal features based on discrete wavelet transform. In: Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on. Vol. 2. IEEE, pp. II–265.

Chapter 2

Overview of Neonatal EEG

> Whoever undertakes to set himself up as a judge of Truth and Knowledge is shipwrecked by the laughter of the gods. ––Albert Einstein

2.1 Introduction

The electroencephalogram (EEG) is an important tool used in neonatal intensive care units (NICUs) to evaluate the function of the neonatal brain. In 1938, Loomis et al. and Smith et al. published the first EEG recordings of term babies (Loomis et al., 1936, Smith, 1938) and cerebral activities of the preterm babies was first published by Hughes et al. (Hughes et al., 1951). Based on these early works, the field of neonatal EEG has been developed extensively for diagnosis of the infant brain.

EEG is an excellent, noninvasive method to assess at-risk newborns and is used to formulate a prognosis for long-term neurological outcome. EEGs are primarily used for (see chapter 4 of (Aminoff, 2012)):

1. assessment of the central nervous system (CNS),
2. diagnosis and treatment of neonatal seizures,
3. estimation of the conceptional age (CA), and
4. determination of prognosis and long-term neurological outcome.

In this chapter, the basic characteristics of neonatal EEG signal are reviewed.

The international standard for recording neonatal EEG is described in section 2.2. The descriptions of neonatal EEG signal characteristics including seizure, normal and abnormal background states are given in section 2.3. A description of different types of background EEG and their grading depending on the severity of the brain injury is also presented in this section.

The following terms (defined in (Itai Berger, 2008)) will be frequently used in this thesis for the description of neonatal EEG:

- *Gestational age* (GA): Time from the first day of the last menstrual period to delivery (in weeks).
- *Conceptional age* (CA): Time from birth in additional to gestational age (in weeks).
- *Continuous tracing*: Continuous (or uninterrupted) EEG activities in all channels (or leads).
- *Discontinuous tracing*: Discontinuous or interrupted EEG activities by periods of inactivity of variable length.
- *Active sleep*: EEG activities characterized by irregular breathing pattern and by facial movements including ocular movements and gentle body movements, which is also called Rapid Eye Movement (REM) sleep.
- *Quiet sleep*: Characterized by regular breathing pattern and some body movements, which is also referred to as Non REM (NREM) sleep.
- *Interburst intervals* (IBIs): Inactive periods in the EEG.
- *Asynchrony*: Refers to the EEG activity between hemispheres during discontinuous tracing.
- *Asymmetry*: Voltage difference between two hemispheres.
- *Trace alternant*: EEG activity alternating between low and high voltages.

2.2 Electroencephalogram

EEG can be described as a record of the electric signal generated by the cooperative action of brain cells at different locations throughout the brain (Niedermeyer and da Silva, 2005). The main advantage of using the EEG is that it provides continuous monitoring of brain activity rather than the snapshot provided by modern

technologies such as computerized tomography (CT) scan and magnetic resonance imaging (MRI). This helps clinicians or neurophysiologists to continuously monitor changes in brain activity, such as the reaction to treatment/medication. This makes EEG an ideal choice for the long term continuous monitoring of neonatal brain activities. The EEG can be measured by placing electrodes on the scalp (scalp EEG) or directly on the cortex, known as the electrocorticogram (ECoG). The International 10-20 System which is used for the placement of recording electrodes (Jasper, 1958) is shown in figure 2.2 where a letter and a number is assigned to individual electrodes. In this system, the placement of electrodes is determined by measurements from four standard positions on the head - the nasion, inion, and right and left preauricular points. In this 10-20 system, the letters refer to the **C**entral, **F**rontal, **O**ccipital, **P**arietal, and **T**emporal lobes of the brain; even numbers correspond to right hemisphere and odd numbers correspond to left hemisphere of the brain. The electrode placed on the mid-line, which is a reference electrode is denoted as **Z**. There is no fixed optimal number of electrodes, but clinical experience has shown that nine electrodes are sufficient to characterize the neonatal EEG (Mizrahi et al., 2004).

Typically, the amplitude of the EEG is in the order of microvolts (μV) and the amplified potential differences between electrodes are then recorded as EEG. In general, predetermined patterns, or montages are used to connect the electrodes with amplifiers to record EEG in sequence. The potential difference can be measured either between pairs of electrodes known as a *bipolar montage* or between individual electrodes and a common reference point known as *monopolar montage* (Bozek-Juzmicki et al., 1994). In this thesis, the EEG recordings were recorded in the NICU of the Cork University Maternity Hospital, Cork, Ireland using eight EEG channels in the bipolar montage: F4-C4, C4-O2, F3-C3, C3-O1, T4-C4, C4-Cz, Cz-C3, and C3-T3. The bipolar montage is preferred clinically as it tends to result in the most interpretable EEG (see chapter 2 of (Rennie et al., 2008)). Figure 2.1 shows a newborn in NICU with EEG electrodes placed on the scalp for recording and figure 2.3 shows a sample EEG recording using the bipolar montage used in this thesis.

2.3 Characteristics of the neonatal EEG

The EEG of a neonate shows well defined patterns related to different activities in the brain such as awake, active and quiet sleep etc. Depending upon the

Figure 2.1: A neonate in the NICU (Cork University maternity Hospital, Cork, Ireland). EEG electrodes are placed on the scalp for recording.

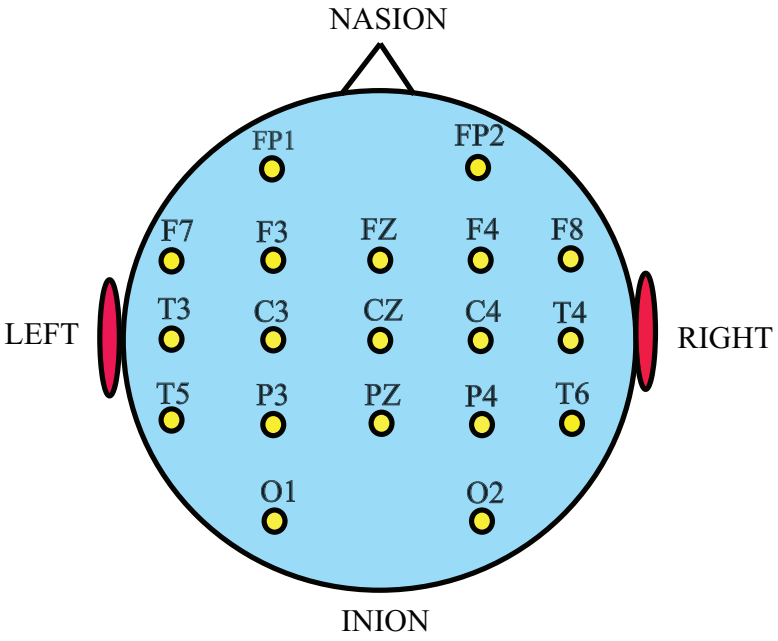


Figure 2.2: Illustration of a birds-eye-view of the International 10-20 electrode placement system used for neonates.

activities in the EEG, it can be categorized as normal or abnormal, or pathological (indicative of disease). In this thesis the dataset obtained from full term neonates (GA ranging between 39-42 weeks) was used, so some of the characteristics of full term neonatal EEG are presented in this section.

Figure 2.3: Example of a 1-minute neonatal EEG recording using the bipolar montage from the NicOne EEG machine in the neonatal intensive care unit at Cork University Maternity Hospital. The EEG is sampled at a frequency of 256 Hz.

2.3.1 Normal background EEG

The patterns of the neonatal EEG become stable with maturation i.e., immature activities disappear and mature activities become prominent. In full-term neonates, the background EEG activities consist of stable patterns without significant change in amplitude (voltage) and frequency (Lombroso, 1985, Niedermeyer and da Silva, 2005). Based on the frequency range, the EEG can be grouped into different bands as shown in Table 2.1 (Gevins and Rémond, 1987, Lombroso, 1985). In neonates, most cerebral activities are found in the range 1-20 Hz and it is important the information in these extremely low frequencies should not be filtered away (Lofhede et al., 2006).

Table 2.1: Properties of some common EEG rhythms (Thomas et al., 2011)

Name	Frequency range	properties
δ (Delta)	0-4	predominant in Infants.
θ (Theta)	4-7	predominant in children.
α (Alpha)	7-12	Visible during awake, resting with closed eyes.
β (beta)	above 12Hz	High-frequency activities during active state.

2.3.2 Abnormal background EEG

Several abnormal patterns are observed in the neonatal EEG. However, only certain patterns that are most commonly seen and within the scope of this thesis

are presented here. Additional discussions on the abnormal patterns of EEG in full-term newborns can be found in (Aminoff, 2012).

A *spike* is defined as "a transient, clearly distinguished from the background activity, with a pointed peak at conventional paper speed and a duration between 20-70msec" (Niedermeyer and da Silva, 2005). The main component is negative with a variable amplitude. The spikes can be distinguished from the background activity based on the wave morphology and amplitude. In several cases, the voltage of the spike is significantly greater than the background activity. However, if the voltages are equal, then the spike can be identified by its faster character (or shorter duration).

A *sharp wave* is defined as "a transient, clearly distinguished from background activity, with pointed peak at conventional paper speeds and duration of 70-200msec, i.e., more than approximately 1/14 to 1/5 sec" (Niedermeyer and da Silva, 2005). The rising phase of the sharp wave and spikes have the same order of magnitude but the descending phase is prolonged in sharp waves. An example of sharp wave is shown in figure 2.4.

Figure 2.4: Example of a sharp wave in 1-minute neonatal EEG.

Electrocerebral Inactivity (Isoelectric) pattern is characterized by extremely low voltage activities. This implies that there is no visible cerebral electrical activity, even at high sensitivities. This EEG pattern is seen in various clinical settings mostly due to severe asphyxia, circulatory collapse, and massive intracerebral hemorrhage. The majority of neonates with inactive EEGs either die in the neonatal period or survive with severe neurological deficits (Aminoff, 2012).

Burst-Suppression (Paroxysmal) pattern is characterized by an isoelectric background activity (amplitude $< 10\mu$V) interrupted by nonperiodic bursts of abnor-

mal activity (amplitude $> 10\mu$V): bursts or short runs of diffuse or focal alpha or theta activity that is occasionally rhythmic lasting approximately 1-10 seconds. In the most severe form, this pattern is invariant and minimally altered by stimuli, and persists throughout waking and sleeping states (Aminoff, 2012, Niedermeyer and da Silva, 2005). Figure 2.5 shows an example of such a pattern.

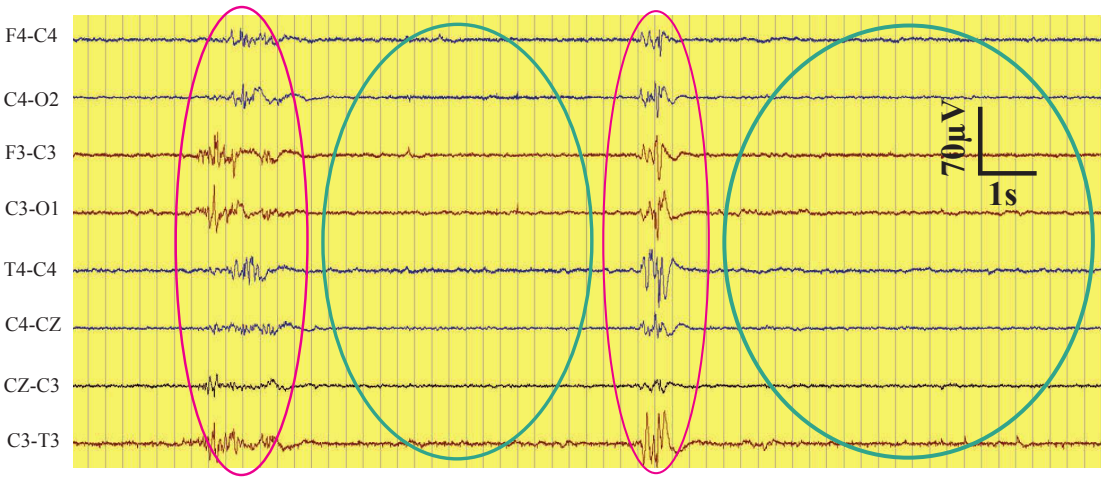


Figure 2.5: Example of burst suppression activity (green-suppression, pink-burst) in 1 minute neonatal EEG. Note that the EEG is inactive for most of the time with some bursts at frequent intervals.

Ictal EEG or seizures

Neonatal seizures result from conditions such as hypoxic ischaemic encephalopathy (HIE), stroke, meningitis and occur at the rate of 1-3/1000 births (Volpe, 2008). Seizures are caused by an imbalance between the excitatory and inhibitory influences in the neonatal brain and may or may not give rise to clinical symptoms (Niedermeyer and da Silva, 2005). Seizures can be focal (appear in one channel), multi-focal (appear in two or more channels) and generalised (appear in all channels). While the risk of developing epilepsy is significantly increased in these neonates, the majority will not experience seizures in their adult life (Garcias Da Silva et al., 2004). Neonatal seizures occur within several days of birth and are associated with increased damage to the brain (Björkman et al., 2010, Lynch et al., 2012, Wusthoff et al., 2011). In general, seizures can be classified into two groups: *Partial* (focal, local) seizures which affects a limited region of the brain and *Generalized* seizures (convulsive or nonconvulsive) affecting the whole brain (Niedermeyer and da Silva, 2005).

The patterns in ictal EEG are highly variable consisting of a variety of frequencies

with complex and varied morphologies (Aminoff, 2012). The seizure patterns are more highly structured and characterized in the newborn EEG, by periods of rhythmic or repeated sharp waves (Gotman et al., 1997, Pellock et al., 2007). An example clearly indicating the rhythmic characteristics of neonatal EEG seizure activity is shown in figure 2.6 (channels corresponding to F3, C3, F4, C4).

There is no "well-defined" minimum duration for ictal patterns in the neonatal EEG. Most of the clinicians specify a limit of 10 seconds to consider ictal pattern as a seizure event, however some specify a limit of 20 seconds (Clancy and Legido, 1987, Niedermeyer and da Silva, 2005).

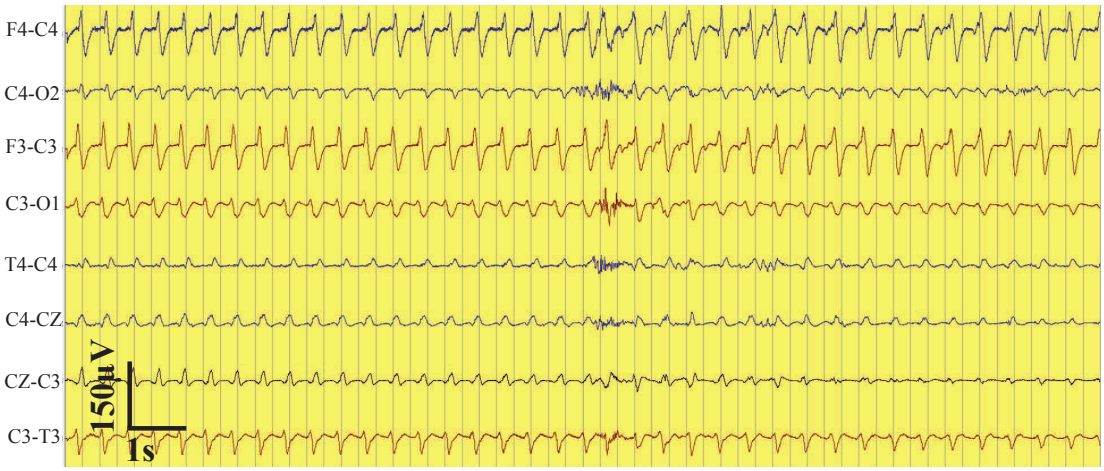


Figure 2.6: Example of a seizure activity in a 1-minute neonatal EEG. Note the high amplitude, repetitive, rythmic, spiking characteristics of seizure in channels corresponding to F3, C3, F4, C4.

Hypoxic-ischaemic encephalopathy

Hypoxic-ischaemic encephalopathy (HIE) is the brain injury caused by the lack of supply of oxygen to the brain. HIE due to fetal or neonatal asphyxia is among the leading cause of death or severe brain damage/impairment among infants with an incidence of 2-3/1000 births (Graham et al., 2008, Inder and Volpe, 2000). Within the first 24 hours of life, the infant may develop seizures with abnormal EEG results. Some of the clinical symptoms includes coma, irregular breathing, absence of neonatal reflexes, disturbances of ocular motion, irregular heart rate and blood pressure (Fenichel, 2009).

The severity of the HIE insult determines the outcome of HIE. Mild HIE can result in a normal outcome, moderate encephalopathy can increase the risk of neurological disability to 20-40% and severe encephalopathy can lead to severe

neurological disability, or in some cases, death. The grading of HIE can be done either by clinical assessment or by using the electroencephalogram (EEG) or amplitude integrated EEG (aEEG). EEG based HIE grading is mainly done through visual analysis of the background activity. Several features such as the continuity of the EEG signal, interhemispheric symmetry and synchrony, amplitude, frequency content, and presence or absence of sleep-wake cycling (SWC) have been used to grade HIE (Murray et al., 2009). HIE are graded into four main types which are summarized in Table 2.2 and sample epochs corresponding to each grade are shown in figure 2.7.

Table 2.2: Properties of HIE

EEG/HIE grade	Abnormality level	Characteristics
1	normal/mild	Continuous background pattern with mild asymmetries and voltage depression (30-50 μV). Presence of poorly defined SWC.
2	moderate	Discontinuous activity with IBI $\leq$10 s. Presence of clear asymmetry or asynchrony and disrupted SWC.
3	major	Discontinuous activity with IBI 10-60s,severe disruption of background patterns (<30 μ V), absence of SWC.
4	severe	Background activity of $\leq$ 10 μV, or severe discontinuity with IBI $\geq$ 60 s.

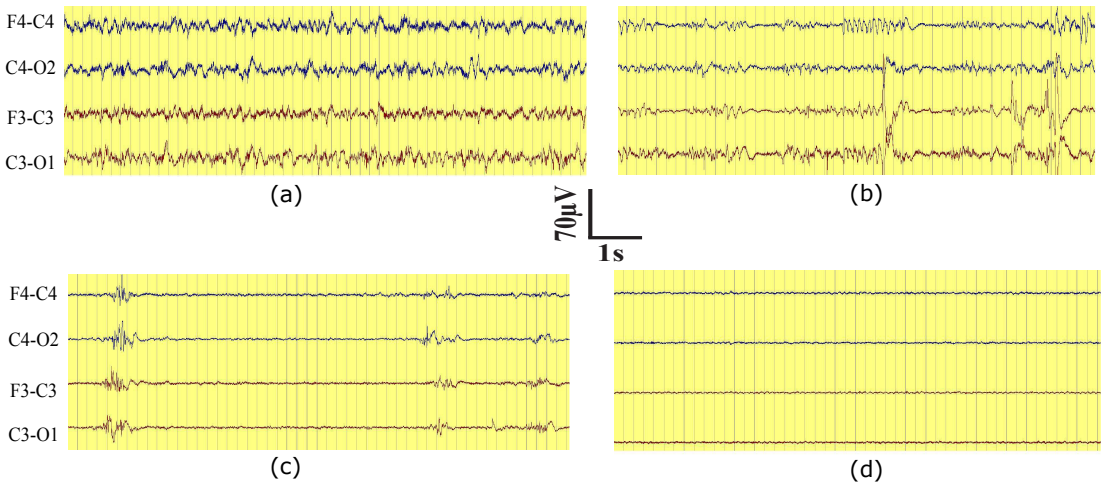


Figure 2.7: Example of a 60s EEG showing different grades of HIE (a) Grade 1: Normal/Mild abnormalities. (b) Grade 2: Moderate abnormalities. (c) Grade 3: Major abnormalities. (d) Grade 4: Inactive.

2.4 Summary

A brief overview of the neonatal EEG signal is provided in this chapter. The EEG recordings can be obtained by placing electrodes on the scalp. Neonatal EEG is useful for detection of seizures, for continuous long-term assessment of brain maturation, for prognosis after brain injury such as hypoxic ischaemic encephalopathy, brain deformations etc. The most common application of performing neonatal EEG is to assess the presence of neonatal seizures. It is one of the most common neurological signs and is considered a clinal challenge which is difficult to detect. It is important to detect seizures on time, especially in the immature brain as they may be the only sign of brain disorder. Early detection can have great impact on the prediction of outcome and the choice of therapy.

The majority of the seizure activities in the neonatal EEG are concentrated in the 0.1-8 Hz frequency band. Several abnormal patterns can be clearly visible in the EEG which can be used to obtain information about the state of the neonatal brain. Neonatal EEG can also provide information about the developmental outcome of the child. Neonates with normal/mild abnormal EEG can have good outcome, whereas those with moderate/severe have poor outcome (or even result in death). These results significantly contribute to the prognosis estimation regardless of when they are analyzed.

References

Aminoff, M. J., 2012. Aminoff's Electrodiagnosis in Clinical Neurology: Expert Consult-Online. Elsevier Health Sciences.

Björkman, S., Miller, S., Rose, S., Burke, C., Colditz, P., 2010. Seizures are associated with brain injury severity in a neonatal model of hypoxia–ischemia. Neuroscience 166 (1), 157–167.

Bozek-Juzmicki, M., Colella, D., Jacyna, G., Oct 1994. Feature-based epileptic seizure detection and prediction from ecog recordings. In: Time-Frequency and Time-Scale Analysis, 1994., Proceedings of the IEEE-SP International Symposium on. pp. 564–567.

Clancy, R. R., Legido, A., 1987. The exact ictal and interictal duration of electroencephalographic neonatal seizures. Epilepsia 28 (5), 537–541.

Fenichel, G. M., 2009. Clinical pediatric neurology: a signs and symptoms approach. Elsevier Health Sciences.

Garcias Da Silva, L. F., Nunes, M. L., Da Costa, J. C., 2004. Risk factors for developing epilepsy after neonatal seizures. Pediatric neurology 30 (4), 271–277.

Gevins, A., Rémond, A., 1987. Methods of Analysis of Brain Electrical and Magnetic Signals. Handbook of electroencephalography and Clinical Neurophysiology: Revised series. Elsevier.

Gotman, J., Flanagan, D., Zhang, J., Rosenblatt, B., 1997. Automatic seizure detection in the newborn: methods and initial evaluation. Electroencephalography and Clinical Neurophysiology 103 (3), 356–362.

Graham, E. M., Ruis, K. A., Hartman, A. L., Northington, F. J., Fox, H. E., 2008. A systematic review of the role of intrapartum hypoxia-ischemia in the

causation of neonatal encephalopathy. American journal of obstetrics and gynecology 199 (6), 587–595.

Hughes, J. G., Davis, B. C., Brennan, M. L., 1951. Electroencephalography of the newborn infant vi. studies on premature infants. Pediatrics 7 (5), 707–712.

Inder, T. E., Volpe, J. J., 2000. Mechanisms of perinatal brain injury. In: Seminars in neonatology. Vol. 5. Elsevier, pp. 3–16.

Itai Berger, M. S., 2008. Hot Topics in Neonatal Neurology. Nova science publishers.

Jasper, H., 1958. Report of the committee on methods of clinical examination in electroencephalography1957. Electroencephalography and Clinical Neurophysiology 10 (2), 370–375.

Lofhede, J., Lofgren, N., Thordstein, M., Flisberg, A., Kjellmer, I., Lindecrantz, K., 2006. Detection of bursts in the EEG of post asphyctic newborns. In: Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE. IEEE, pp. 2179–2182.

Lombroso, C. T., 1985. Neonatal polygraphy in full-term and premature infants: a review of normal and abnormal findings. Journal of Clinical Neurophysiology 2 (2), 105–156.

Loomis, A. L., Harvey, E. N., Hobart, G., 1936. Electrical potentials of the human brain. Journal of experimental Psychology 19 (3), 249.

Lynch, N. E., Stevenson, N. J., Livingstone, V., Murphy, B. P., Rennie, J. M., Boylan, G. B., 2012. The temporal evolution of electrographic seizure burden in neonatal hypoxic ischemic encephalopathy. Epilepsia 53 (3), 549–557.

Mizrahi, E. M., Hrachovy, R. A., Kellaway, P., Stockard-Pope, J., 2004. Atlas of neonatal electroencephalography. Vol. 31. Lippincott Williams & Wilkins Philadelphia, PA:.

Murray, D. M., Boylan, G. B., Ryan, C. A., Connolly, S., 2009. Early eeg findings in hypoxic-ischemic encephalopathy predict outcomes at 2 years. Pediatrics 124 (3), e459–e467.

Niedermeyer, E., da Silva, F. L., 2005. Electroencephalography: basic principles, clinical applications, and related fields. Lippincott Williams & Wilkins.

Pellock, J. M., Bourgeois, B. F., Dodson, E., Nordli, D. R., Sankar, R., Bourgeois,

B., Dodson, E., Nordli, D., 2007. Pediatric epilepsy: diagnosis and therapy. Demos Medical Publishing.

Rennie, J. M., Hagmann, C. F., Robertson, N. J., 2008. Neonatal cerebral investigation. Cambridge University Press.

Smith, J. R., 1938. The electroencephalogram during normal infancy and childhood: I. rhythmic activities present in the neonate and their subsequent development. The Pedagogical Seminary and Journal of Genetic Psychology 53 (2), 431–453.

Thomas, E. M., Temko, A., Lightbody, G., Marnane, W. P., Boylan, G. B., 2011. Advances in automated neonatal seizure detection. In: New Advances in Intelligent Signal Processing. Springer, pp. 93–113.

Volpe, J. J., 2008. Neurology of the Newborn. Elsevier Health Sciences.

Wusthoff, C. J., Dlugos, D. J., Gutierrez-Colina, A., Wang, A., Cook, N., Donnelly, M., Clancy, R., Abend, N. S., 2011. Electrographic seizures during therapeutic hypothermia for neonatal hypoxic-ischemic encephalopathy. Journal of child neurology 26, 724–728.

Chapter 3

Overview of Time-Frequency Signal Analysis

> The important thing is not to stop questioning. Curiosity has its own reason for existing. ––Albert Einstein

3.1 Introduction

Time-frequency analysis (TFA) or time-frequency representation (TFR) is a technique that represents the signal in both time and frequency domains (2D), in contrast with the traditional time domain or frequency domain (1D) analysis. For several years spectral estimation has been well developed for stationary signals - *signals whose statistical properties do not change over time*. For non-stationary signals that we encounter in everyday applications, the TFR proves to be essential to track the variation of frequency content of the signal over the evolution of time. The energy contents of a non-stationary signal can be visualized using the TFR. This also helps in understanding the properties of the signal in order to determine the best approach for analysis.

In this chapter, several time-frequency representations of signals are described that are relevant to the thesis. In section 3.2, the limitations of stationary techniques (Fourier Transform) to analyze non-stationary signals are demonstrated. Section 3.3 reviews several linear TFR's which is followed by description of advanced TFR (Wigner-ville distribution and its improvements) in section 3.4. Finally, in section 3.5, various methods of atomic decomposition are described and

their application for the analysis of the non-stationary signal is shown.

3.2 Necessity for non-stationary signal analysis techniques

For several years, the Fourier transform (FT) has been the simplest and most intuitive way to represent the signal in the frequency domain. The FT of a signal decomposes the original signal into its harmonic components, giving the spectral content of the signal. For a signal $x(t)$, its FT is defined as:

$$X(f) = \mathcal{F}\{x(t)\} = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}dt. \tag{3.1}$$

However, when analyzing signals with time varying frequencies, the amplitude of the FT, also known as magnitude spectrum ($|X(f)|$) does not give complete information regarding the behavior of the system (Bracewell, 2000). To illustrate this, a sample Gaussian test signal ($x(t)$) given by:

$$x(t) = e^{-\pi(t-\alpha_1)^2}\sin(\beta_1 t) + e^{-\pi(t-\alpha_2)^2}\sin(\beta_2 t) \tag{3.2}$$

which is centered at (α_1, β_1) and (α_2, β_2), respectively is used. This is a special case of the summation of two Gabor atoms with a scaling factor = 1(which is described later in section 3.6).

Figure 3.1 shows the test signal and its magnitude spectrum using the FT which gives information about the frequency content of the signal. Since the information from the FT at a given frequency f is computed by averaging the contributions for the entire duration, it is not possible to obtain the instantaneous frequency content. This limitation of FT led to the development of time-frequency representations (TFR) of the signal.

3.3 Linear time-frequency representations

The energy of a signal can be computed in the time or in the frequency domain as:

$$E_x = \int_{-\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} |X(f)|^2 df \tag{3.3}$$

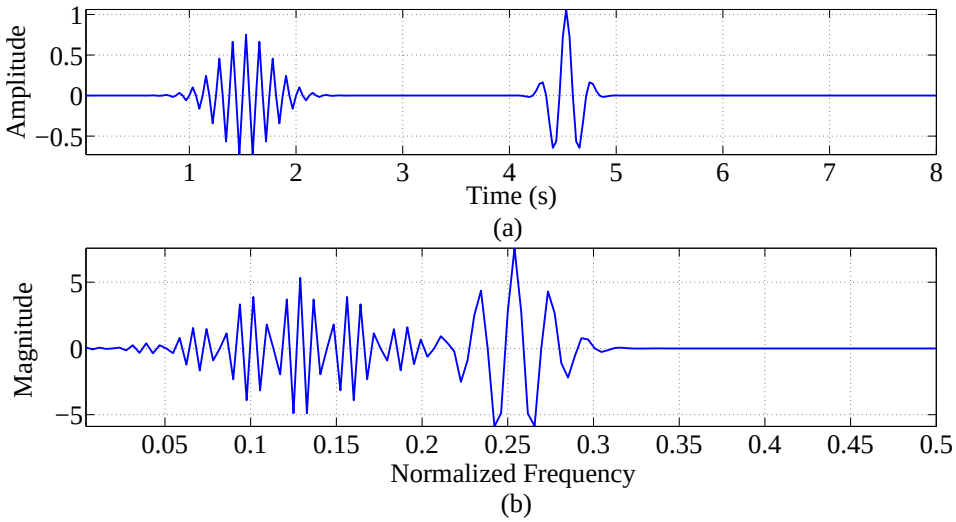


Figure 3.1: (a) The time domain and (b) its corresponding frequency domain representation of the test signal using the FT. No clear distinction is obtained between two components in the signal.

where $|x(t)|^2$ and $|X(f)|^2$ correspond to time and frequency domain energy density, respectively. This idea of representing the energy density of the signal can be extended to joint time-frequency (TF) space. The signal energy should be completely represented by the TF energy density $\rho_x(t, f)$ and fulfill the following conditions at a given TF point:

1. The total energy is conserved:

$$E_x = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho_x(t, f) dt df, \tag{3.4}$$

2. The marginal distributions are preserved:

$$\int_{-\infty}^{\infty} \rho_x(t, f) df = |x(t)|^2 \tag{3.5}$$

$$\int_{-\infty}^{\infty} \rho_x(t, f) dt = |X(f)|^2. \tag{3.6}$$

Linear TFR's are based on the superposition principle which states that if $x(t)$ can be represented as a linear combination of $x_1(t)$ and $x_2(t)$ then, the TFR of $x(t)$ is given by the linear combination of TFR of $x_1(t)$ and $x_2(t)$, i.e.,

$$x(t) = c_1(x_1(t)) + c_2(x_2(t)) \implies \rho_x(t, f) = c_1(\rho_{x_1}(t, f)) + c_2(\rho_{x_2}(t, f)) \tag{3.7}$$

where $\rho_x(t,f)$ is the TFR of $x(t)$ (Hlawatsch and Boudreaux-bartels, 1992). In general, a linear TFR is obtained by correlating the signal with a collection of waveforms concentrated in time and frequency known as *time-frequency atoms*(Mallat, 2008). The general formulation of a linear TFR in terms of inner product[1] is given by

$$\rho_x(t,f) = \int_{-\infty}^{\infty} x(t)\phi^*_{t,f}(t)dt = \langle x, \phi^*_{t,f}\rangle \tag{3.8}$$

where $\phi_{t,f}$ is the time-frequency atom concentrated around t and f.

Due to the limitation of the FT in representing non-stationary signals, the short-time Fourier transform (STFT) was developed. The STFT of a signal is given by:

$$\rho_x^{STFT}(t,f) = \int_{-\infty}^{\infty} x(\tau)w(\tau-t)e^{-j2\pi f\tau}d\tau \tag{3.9}$$

which projects the signal on the basis function $\{w(\tau-t)e^{-j2\pi f\tau}\}$. In other words, the TF atom in (3.8) is given by $\phi_{t,f} = w(\tau-t)e^{j2\pi f\tau}$. In STFT, the input signal $x(t)$ is multiplied by a window function $w(t)$, and then the Fourier transform is applied. This method is repeated to obtain the FT with windowed signals that emphasize the original signal at different instances which results in the STFT (Oppenheim et al., 1999). By squaring the STFT, the signal's energy spectrum in the TF domain is obtained known as the *spectrogram*. In STFT, the signal is segmented into small blocks that are assumed to be short enough to be considered as stationary. Though the spectrogram provides a clear picture of change in frequency content with respect to time, it has a limitation with respect to time and frequency resolution. The resolution in frequency direction is decreased with the increased resolution in time domain (using narrow window), and vice-versa by using a wider window (Boashash, 1992).

Another popular linear TFR known as the wavelet transform (WT) is a technique to analyze time varying signals (Mallat, 2008). The WT of signal $x(t)$ is defined as

$$\rho_x^{WT}(t,a) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} x(\tau)\phi^*\left(\frac{\tau-t}{a}\right)d\tau \tag{3.10}$$

where $\phi(t)$ is called the analysing or mother wavelet, and a is the scale which can be written in terms of frequency f as $a = \frac{f_0}{f}$ where f_0 is the center frequency of the FT of $\phi(t)$. The WT can be considered as the projection of the signal on the basis function $\left\{\sqrt{\frac{f}{f_0}}\phi\left(\frac{f}{f_0}(\tau-t)\right)\right\}$.

[1]The inner product of the two functions $f_1(t)$ and $f_2(t)$ is given by $\langle f_1, f_2\rangle = \int_{-\infty}^{\infty} f_1(t)f_2^*(t)dt$.

Both the STFT and the WT are bounded by the limitation of fixed resolution which is indirectly related to the *Heisenberg uncertainty principle* which states that time and frequency resolution cannot be improved simultaneously, i.e.,

$$\Delta t \Delta f \geq \frac{1}{4\pi} \tag{3.11}$$

where Δt and Δf corresponds to time duration and bandwidth of the TF atoms, respectively (Boashash, 1992). Due to this limitation, we cannot precisely know the frequency component of a signal at a give time instance also known as Instantaneous frequency (IF).[2]

3.4 Quadratic Time-Frequency Representations

To overcome the above mentioned limitations, several TFR techniques have been developed to represent the signal in the time-frequency domain. The Wigner-Ville distribution (WVD) was initially proposed for this reason (Boashash, 1992, Cohen, 1995). The WVD is defined for the analytic associate of $x(t)$ which is given by

$$\rho_x^{WVD}(t,f) = \int_{-\infty}^{\infty} z(t+\frac{\tau}{2}) z^*(t-\frac{\tau}{2}) e^{-2\pi f \tau} d\tau \tag{3.12}$$

where $z(t) = x(t) + j\mathcal{H}\{x(t)\}$ is a complex signal, $\mathcal{H}\{x(t)\}$ is the Hilbert transform of $x(t)$, and $z^*(t)$ is the complex conjugate of $z(t)$.

Several desirable properties such as time and frequency shift invariance, time and frequency marginal satisfaction, time and frequency support (Boashash, 2003) of WVD makes it as an ideal choice to explore TFRs. However, its main drawback is the existence of oscillatory cross-terms (or artefacts) which makes the interpretation of WVD difficult. Due to the quadratic representation of the WVD, the cross-term artefact appears if the signal has multiple TF components. For example, let us consider a signal $x(t)$ composed of two signals $x_1(t)$ and $x_2(t)$ such that $x(t) = x_1(t) + x_2(t)$. The WVD of this signal can be expressed as:

$$\rho_x^{WVD}(t,f) = \rho_{x_1}^{WVD}(t,f) + \rho_{x_2}^{WVD}(t,f) + 2\Re\{\rho_{x_1,x_2}^{WVD}(t,f)\}. \tag{3.13}$$

The last term in (3.13) is called a cross-term. Figure 3.2 shows the WVD of the test signal (given in equation 3.2). We can clearly see the presence of cross-terms

[2]Instantaneous frequency of a signal is defined as the derivative of the phase i.e., $IF(t) = \frac{1}{2\pi}\frac{d\theta(t)}{dt}$ where $\theta(t)$ is the phase function of the signal $x(t)$.

(outer artefacts) at the centre which masks the IF's of the individual components, thereby decreasing the performance of TFR.

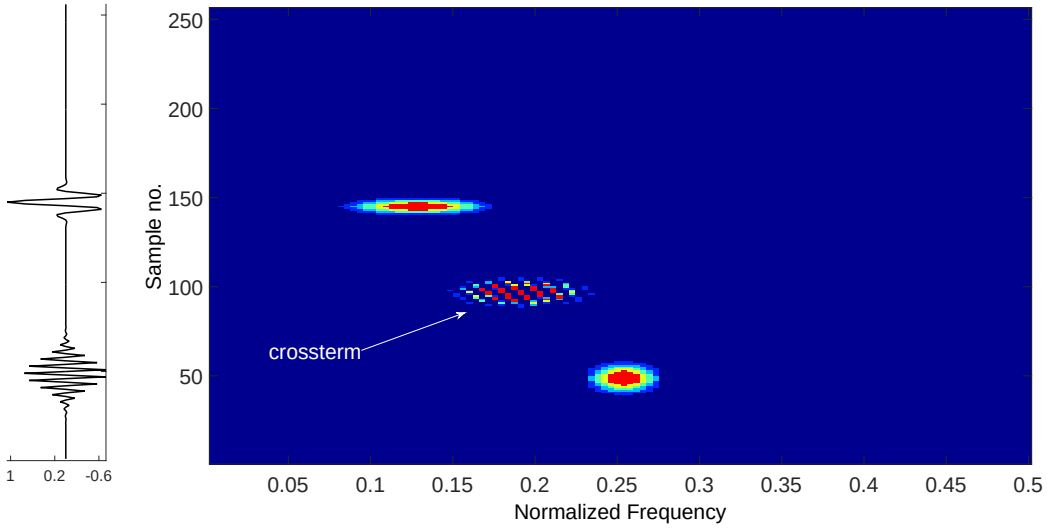


Figure 3.2: Time frequency representation of the test signal obtained using WVD. Note the presence of cross-terms between the two signal components.

The quadratic time frequency distribution (QTFD) is an improved version of the traditional WVD in which the WVD is smoothed with a time-lag kernel, $G(t,\tau)$. By using an appropriate $G(t,\tau)$, the cross-term artefacts in WVD can be eliminated. The general form of QTFD is given by (see page 67 of (Boashash, 2003))

$$\begin{aligned}\rho_z^{QTFD}(t,f) &= \underset{\tau\to f}{\mathcal{F}} G(t,\tau) \underset{t}{*} z\left(t+\frac{\tau}{2}\right) z^*\left(t-\frac{\tau}{2}\right) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(u,\tau) z\left(t-u+\frac{\tau}{2}\right) z^*\left(t-u-\frac{\tau}{2}\right) du e^{-2\pi f\tau} d\tau,\end{aligned} \tag{3.14}$$

Several time-lag kernels have been designed to improve the resolution of QTFD by tuning several parameters in $G(t,\tau)$. Table 3.1 gives some examples of QTFDs and their corresponding time-lag kernels. It can be seen from table 3.1 that the tunable parameters (σ for Choi-Williams (CW) and β for Modified B (MB) and B distributions) can be adjusted for various signals to enhance the resolution of the TFR obtained by these distributions (Boashash, 2003).

Figure 3.3 shows the TFR of the test signal obtained using the B-distribution. We can clearly see that the cross-term artefacts are removed thereby enhancing the resolution of the TFR.

Table 3.1: QTFD and their time-lag kernels

QTFD	time-lag kernel, $G(t,\tau)$
Wigner-Ville	$\delta(t)$
Modified B	$\frac{\cosh^{-2\beta} t}{\int_{-\infty}^{\infty} \cosh^{-2\beta} \xi d\xi}$
B	$\lvert\lambda\rvert^{\beta} \cosh^{-2\beta} t$
Choi-Williams	$\frac{\sqrt{\pi\delta}}{\lvert\lambda\rvert} e^{-\pi^2\delta t^2}/\lambda^2$
Spectrogram	$w\left(t+\frac{\tau}{2}\right) w\left(t-\frac{\tau}{2}\right)$

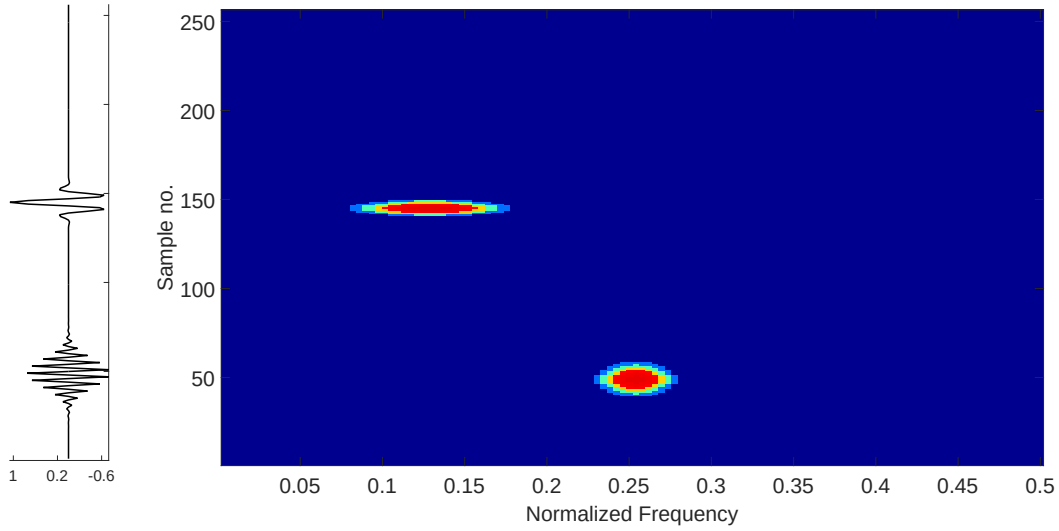


Figure 3.3: Time frequency representation of the test signal obtained using the B-distribution. Note that the artefacts are removed by tuning the parameter β.

3.5 Atomic Decomposition

Atomic decomposition (AD) uses a time-frequency redundant dictionary to optimally represent a given signal. A dictionary is a collection of elementary signals or *atoms* such that $\mathbf{D} = \{\boldsymbol{\phi}_\lambda\}_{\lambda\in\Lambda}$, where Λ is the set of parameters used to build ϕ, and can be used for signal representation. The atoms in the dictionary are obtained through the transformations of the fundamental atom $\boldsymbol{\phi}_\lambda(t)$. The transformations can be obtained both in time and frequency to obtain a complete or over-complete time-frequency dictionary. The dictionary is said to be complete if it spans the Hilbert space, H (Mallat and Zhang, 1993). However, if the dictionary contains more atoms than are needed to span H, it is called a *redundant* dictionary, which is also referred to as *overcomplete* (Chen et al., 2001). Therefore, a method that can best represent the signal structure using atoms from the dictionary is desirable.

AD techniques provide a signal representation (or approximation) as a linear superposition of the selected atoms from the dictionary as

$$\mathbf{x} = \sum_{\lambda \in \Lambda} \boldsymbol{\gamma}_\lambda \boldsymbol{\phi}_\lambda \tag{3.15}$$

where $\boldsymbol{\gamma}_\lambda$ is the coefficient associated with $\boldsymbol{\phi}_\lambda$. The approximate signal obtained after K iterations [3].

$$\hat{\mathbf{x}} = \sum_{k=1}^{K-1} \boldsymbol{\gamma}_{\lambda_k} \boldsymbol{\phi}_{\lambda_k} \tag{3.16}$$

There are several AD techniques available which includes: matching pursuit (Mallat and Zhang, 1993), orthogonal matching pursuit (Pati et al., 1993) and basis pursuit (Chen et al., 2001). These methods are briefly described in the following subsections.

3.5.1 Matching Pursuit

Matching pursuit (MP) is a greedy algorithm used for sparse signal representation (Mallat and Zhang, 1993). It finds linear approximations of signals by iteratively projecting them over a redundant, non-orthogonal set of atoms in the dictionary. Since MP is a greedy algorithm, the approximation obtained is not optimal, however, it is useful in the case of higher order signal approximation. MP has found its applications in the area of signal compression, in particular audio, video and image compression. In addition, MP has also found its application in feature extraction for classification problems (Cotton and Ellis, 2009, Hsu and Huang, 2001, Rankine et al., 2007).

Given a dictionary of M atoms, $\mathbf{D} = [\boldsymbol{\phi}_1 \ \boldsymbol{\phi}_2 \ \ldots \ \boldsymbol{\phi}_M] \in \mathbb{R}^{N \times M}$, where each column represents a single atom of length N and $M \geq N$, MP approximates the signal $\mathbf{x}$ using an iterative greedy process as a linear combinations of K atoms which can be summarized as follows:

1. Initialize the residual $\mathbf{r}_0 = \mathbf{x}$, the index set $\Psi_0 = \emptyset$, the dictionary $\mathbf{D}_0 = \emptyset$ and the iteration count $k = 1$.
2. Find $\check{i} = \max |\mathbf{r}_{k-1}^T \boldsymbol{\phi}_i|$, $\boldsymbol{\phi}_i \in \mathbf{D}$.
3. Update the index set $\Psi_k = \Psi_{k-1} \cup \check{i}$ and the active dictionary, $\mathbf{D}_k = [\mathbf{D}_{k-1} : \boldsymbol{\phi}_{\check{i}}] \in \mathbb{R}^{N \times k}$.

[3] It should be noted that number of iterations and number of atoms will be used interchangeably which refer to number of atoms used in AD.

4. Estimate $\boldsymbol{\gamma}_k$ by finding maximum inner product $|\langle \mathbf{D}_k, \boldsymbol{\gamma} \rangle|$;

5. Update the residual, $\mathbf{r}_k = \mathbf{r}_{k-1} - \mathbf{D}_k \boldsymbol{\gamma}_k$.

6. Update the iteration count, $k = k + 1$. Repeat from step 2 until the given stopping criteria is reached ($k = K$).

After K iterations, the MP provides the sparse approximation of the input signal $\mathbf{x}$ as

$$\hat{\mathbf{x}} = \mathbf{D}_K \boldsymbol{\gamma}_K. \tag{3.17}$$

The MP iterative procedure is shown in Figure 3.4.

Major improvements to the MP algorithm have been proposed in the literature which deal with several aspects of the MP algorithm such as computation speed, optimality of the solution, accuracy of approximation and sparsity of the representation. Since MP is a greedy algorithm, with a finite K, the approximation $\hat{\mathbf{x}}$ will be suboptimal. Let $\mathbf{S_K}$ be the span of all chosen atoms from dictionary such that $\mathbf{S_K} = span\{\boldsymbol{\phi}_1 \ \boldsymbol{\phi}_2 \ \ldots \ \boldsymbol{\phi}_K\}$. The approximation $\hat{\mathbf{x}}$ is said to be optimal only if the residue after K iterations lies in the span of orthogonal complement of $\mathbf{S_K}$, i.e., $\mathbf{r}_K \in \mathbf{S_K}^{\perp}$. However, MP only guarantees that the residue is perpendicular to the final dictionary atom and therefore, the approximation given by the MP is usually suboptimal. Due to this limitation, the MP algorithm requires more iterations to obtain a residual $\mathbf{r}_K$ less than the given threshold (Mallat and Zhang, 1992, Pati et al., 1993). This limitation is removed by orthogonal matching pursuit (OMP).

3.5.2 Orthogonal matching pursuit

OMP is an iterative, greedy algorithm that calculates the locally optimum signal approximation at each iteration. At every iteration, an optimal approximation is obtained from the linear combination of selected atoms from the dictionary with residue orthogonal to all of the chosen dictionary atoms (Tropp and Gilbert, 2007). OMP uses the mean square error, rather than the inner product to optimize the selection of $\boldsymbol{\gamma}$. $\boldsymbol{\gamma}_k$ in step 4 of the MP algorithm is estimated by solving a least squares problem using the Moore-Penrose pseudoinverse method (Penrose, 1955) as

$$\boldsymbol{\gamma}_k = \arg\min_{\boldsymbol{\gamma} \in \mathbb{R}_k} ||\mathbf{x} - \mathbf{D}_k \boldsymbol{\gamma}||_2. \tag{3.18}$$

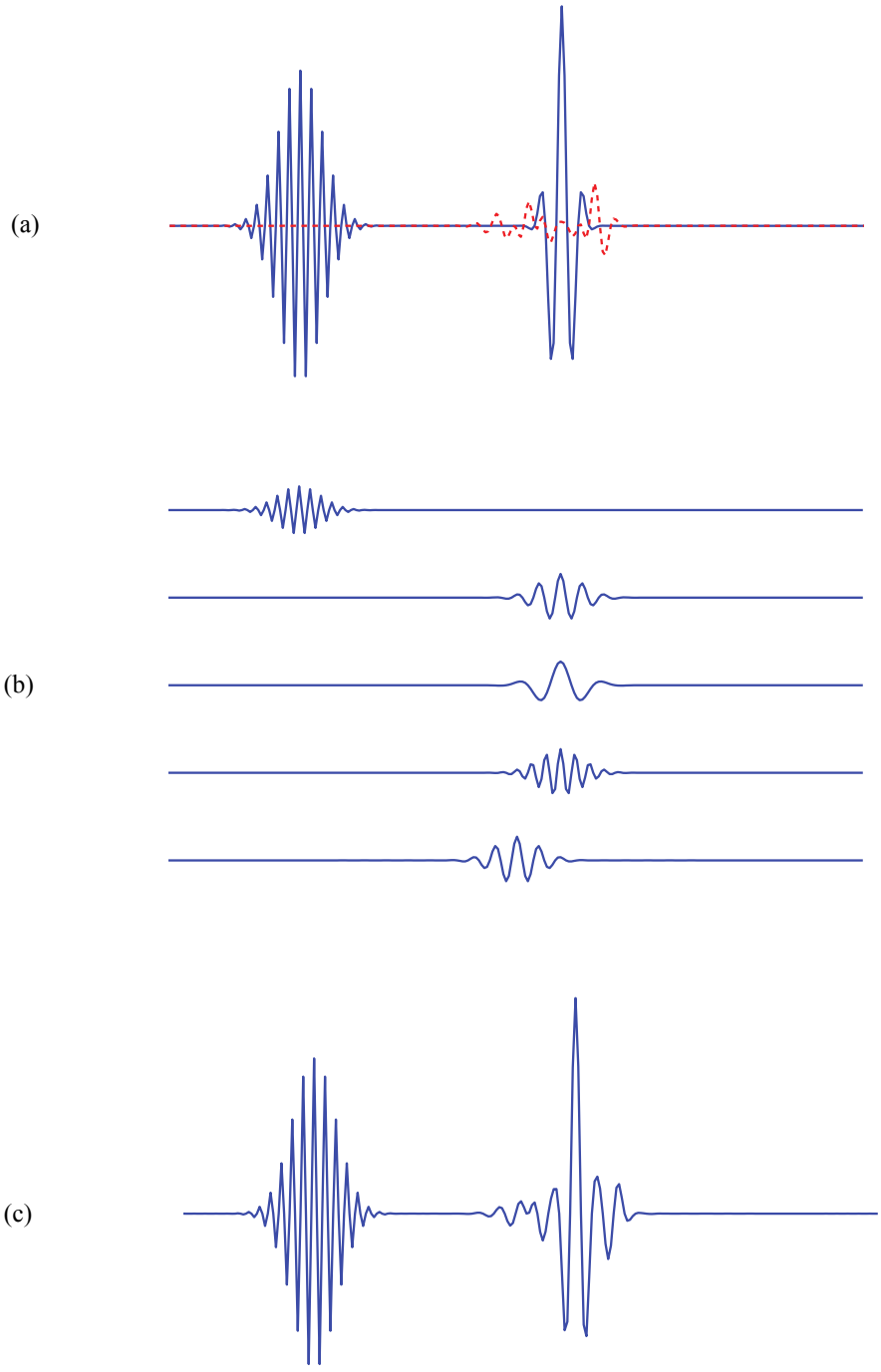


Figure 3.4: Illustration of MP decomposition. (a) The test signal to be decomposed (blue line) and its residual obtained after five iterations (red dotted line), (b) the first five atoms selected during MP, (c) the weighted sum of five atoms.

This ensures that the residue $\mathbf{r}_K \in \mathbf{S_K}^{\perp}$ and an optimal approximation of $\mathbf{x}$ is obtained for the given set of K dictionary atoms. OMP also differs from MP in

that all possible values of $\boldsymbol{\gamma}$ are estimated at each iteration, rather than a single value. This means that the same atom can never be selected twice and as a result, OMP converges faster.

3.5.3 Basis pursuit

Both MP and OMP use a greedy approach to select the atoms from the dictionary to give a speedy approximation but not necessarily a guaranteed globally optimal solution. Basis Pursuit (BP) is an AD technique proposed as an alternative to MP to provide the sparsest representation of a signal over the given dictionary through convex optimization (Chen et al., 2001). BP finds the best representation of a signal by minimizing the L_1-norm [4] of the components of $\boldsymbol{\gamma}$, i.e. the coefficients in the representation. BP can be considered as an optimization method to solve the following problem:

$$\min ||\boldsymbol{\gamma}||_1 \text{ subject to } \mathbf{x} = \sum_{\boldsymbol{\lambda} \in \boldsymbol{\Lambda}} \boldsymbol{\gamma}_\lambda \boldsymbol{\phi}_\lambda. \tag{3.19}$$

During the process of finding $\boldsymbol{\gamma}$, since the nonzero coefficients correspond to columns of the dictionary, the indices of the nonzero components of $\boldsymbol{\gamma}$ can be used to identify the columns of $\mathbf{D}$ that are needed to accurately represent the given signal. The BP problem can be written as a linear programming problem (LP) of the form:

$$\text{Minimize } \mathbf{c}^T \boldsymbol{\gamma} \text{ subject to } \mathbf{D}\boldsymbol{\gamma} = \mathbf{x}; \boldsymbol{\gamma} > 0. \tag{3.20}$$

Any LP technique can be used to solve this optimization problem like the simplex or interior-point method (Boyd and Vandenberghe, 2004). More details about BP can be found in (Chen et al., 2001).

3.6 Dictionaries for atomic decomposition

The performance of the AD algorithm mainly depends on the choice of the dictionary. A dictionary is a collection of parameterized signals $\mathbf{D} = \{\boldsymbol{\phi}_\lambda\}_{\lambda \in \Lambda}$, also called *atoms* which contain several characteristics of the signal. Depending on the application, dictionaries can be either *complete* with n atoms or *overcomplete*

[4] L_1 norm of a function $f(t)$ is defined as $||f||_1 = \int_{-\infty}^{\infty} |f(t)|dt$

with more than n atoms. In special cases, dictionaries with less than n atoms are used which are known as *undercomplete* dictionaries. Several dictionaries have been proposed in the literature which can be grouped into three categories:

3.6.1 Frequency dictionaries

A dictionary consisting of orthogonal bases such as the Fourier basis is an example of a frequency dictionary. The Fourier dictionary is a collection of sinusiodal waveforms ϕ_λ which is indexed by $\lambda = (\omega, v)$. In this case, $\omega \in [0, 2\pi]$ is the variable angular frequency and $v \in [0, 1]$ is the phase type (sine or cosine) given by (Chen et al., 2001):

$$\phi_{(\omega,0)} = \cos(\omega t), \quad \text{and} \quad \phi_{(\omega,1)} = \sin(\omega t). \tag{3.21}$$

In a standard Fourier dictionary, λ runs through the set of all cosines with Fourier frequencies $\omega_k = \frac{2\pi k}{n} f_s,\ k = 0, 1, \ldots, n/2$ and all sines with Fourier frequencies $\omega_k,\ k = 1, 2, \ldots, n/2 - 1$. There will be n atoms available in this dictionary and the atoms in this dictionary will be mutually orthogonal (Chen et al., 2001).

3.6.2 Time-scale dictionaries

A time-scale dictionary is a collection of translated and dilated atoms from a *mother wavelet*. A wavelet is a wave like function with finite energy that allows the analysis of transient, or a time-varying event/phenomena. Wavelets are orthogonal basis vectors formed by shifting and dilating a mother wavelet, $\phi(t)$

$$\phi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \phi\left(\frac{t-b}{a}\right) \quad \forall a, b \in \mathbb{R} \tag{3.22}$$

where a is the scale/dilation parameter and b is the location/time-shifting parameter (Chen et al., 2001). For a given signal of length N, each atom in the wavelet dictionary is indexed by $\lambda = (a, b)$ with $0 \leq a \leq \log_2(N), 0 \leq p \leq 2^{-a}N$. Several orthogonal wavelet functions are available which includes the families of Haar, Daubechies, coiflets, and symlets (Chen et al., 2001).

Traditionally, dictionaries consisting of orthogonal bases such as Fourier basis and orthogonal wavelet bases have a minimum number of atoms to span the Hilbert space. As a result, their performance is limited due to the amount of

information provided. These limitations were outlined in (Mallat and Zhang, 1993) where it was demonstrated that the Fourier and Wavelet bases are poor in representing signal structures that are well localized in time and have narrow high frequency support, respectively (Chen et al., 2001).Therefore, it is necessary to design a dictionary with TF atoms which can represent all types of signal structures with high resolution. Figure 3.5 shows a sample atom; in a Fourier and wavelet dictionary.

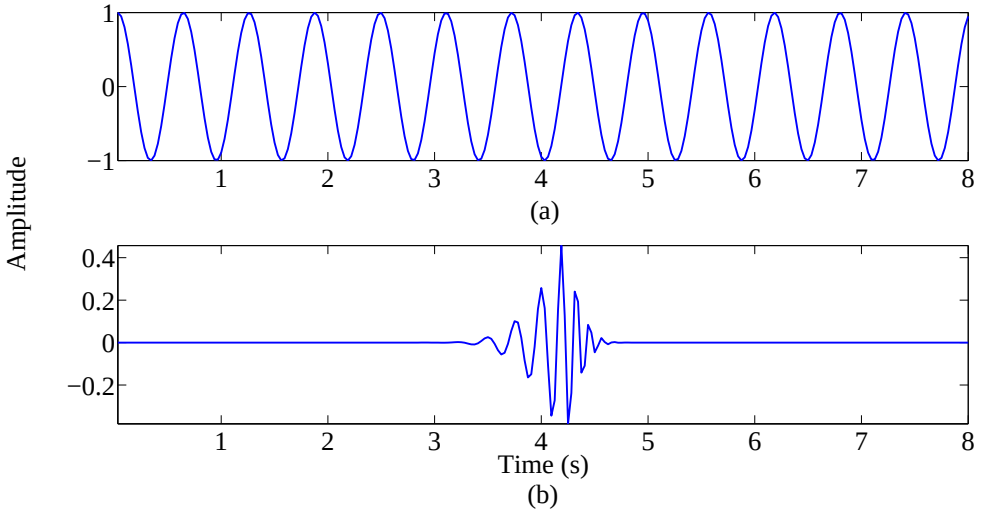


Figure 3.5: Example of an atom from (a) Fourier and (b) wavelet dictionary.

3.6.3 Time-Frequency dictionaries

For many applications of AD, parametric dictionaries are preferred to reduce the computation speed and storage space required. Parametric dictionaries refer to a group of atoms with a well-defined parametric form and the entire group can be generated by varying the parameters of the atoms. Using a group of signals as a dictionary has an advantage of reducing storage requirements since only the parametric form of the dictionary and the corresponding parameters for each atom are sufficient. In addition, the computational speed of the AD algorithms can be increased by exploiting the inherent structure of the dictionary.

Time-Frequency (TF) atoms are functions that are well localized in both time and frequency domains. The Wavelet transform is an example of a signal decomposition method that uses TF atoms which can have different properties based on the selection of TF atoms. Due to this reason, the selection of the family of TF atoms for decomposition mainly depends upon the type of data being analyzed.

A general family of time-frequency atoms can be generated by scaled, translated and modulated version of a single window function $\phi(t) \in H$. Each atom in $\mathbf{D}$ is normalized such that its L_2 norm[5] is unity i.e., $||\phi_\lambda(t)||_2 = 1; \forall\lambda$.

A Gabor dictionary is an example of a TF dictionary that is commonly used in signal approximation (Chen et al., 2001). The Gabor dictionary consists of translated (α), scaled (m) and modulated (β) versions of a Gaussian window such that[6]:

$$\phi(t;\lambda) = \frac{1}{\sqrt{m}} e^{-\pi\left(\frac{t-\alpha}{m}\right)^2} \sin(\beta t) \tag{3.23}$$

where $\lambda = [\alpha, m, \beta]^T \in \mathbf{\Lambda}$ and $\mathbf{\Lambda} = \mathbb{R}^+ \times \mathbb{R}^2$ are the time-frequency parameters. Figure 3.6 shows an example of atoms in the Gabor dictionary.

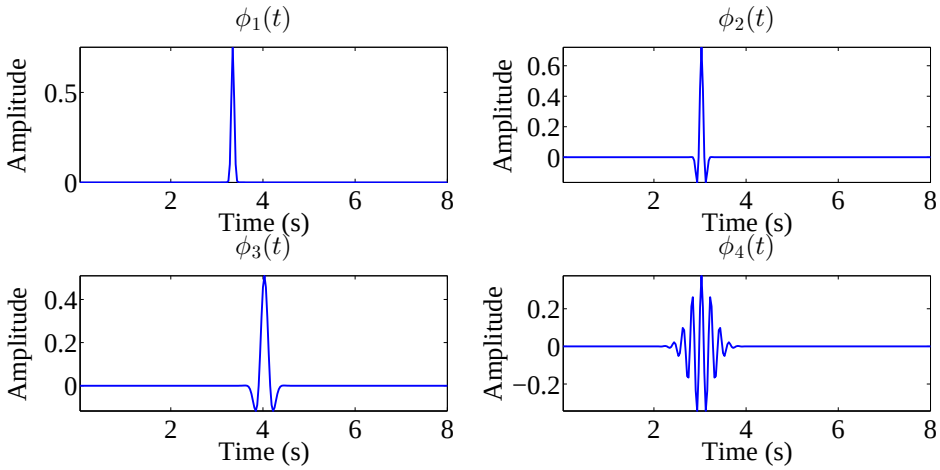


Figure 3.6: Example atoms in the Gabor dictionary.

3.7 Time-Frequency representation using atomic decomposition

The AD of signal can be used to produce the time-frequency distribution. In this approach, the best properties of QTFD and ADs can be combined to provide a joint TFR. The WVD of the AD can be written as:

$$\rho_x^{AD}(t,f) = \sum_{k=0}^{K} |\langle \mathbf{r}_k, \phi_{\lambda_k} \rangle|^2 \rho_{\phi_{\lambda_k}}(t,f) \tag{3.24}$$

[5] L_2 norm of a function is defined as $||f||_2 = \int_{-\infty}^{\infty} f(t) f^*(t) dt$

[6] more details about Gabor dictionary is given in chapter 4

where $\mathbf{r}_k$ is the residual obtained using $\boldsymbol{\phi}_{\lambda_k}$ atoms during AD, and $\rho_{\phi_{\lambda_k}}(t, f)$ is the WVD of individual atoms. This TFR has no cross-terms and therefore offers the highest resolution. Figure 3.7 shows the TFR obtained using AD techniques.

In addition to the TFR, the dictionaries in AD can be designed to be coherent with desired structures in the given signal. For example the Gabor dictionary with MP has been used in detecting structural changes in EEG (Jouny et al., 2003). It was shown that the features obtained from AD can provide necessary information about the change in structures of the EEG signal which could be useful in detecting components with varying time instances of the rising and decaying parts.

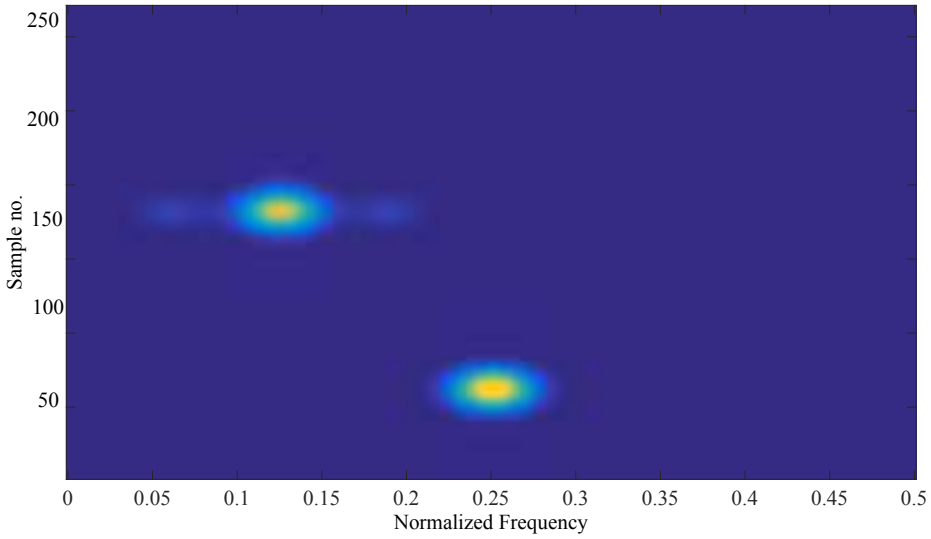


Figure 3.7: WVD of the first five weighted atoms selected from the Gabor dictionary during the decomposition of test signal in equation (3.2). The selected atoms are shown in figure (3.4b).

3.8 Summary

In this chapter, time-frequency analysis and its applications in nonstationary signal analysis are discussed. It was shown that TFA is an ideal choice for analyzing nonstationary signals. The positive properties of traditional STFT and WVD are the speed of computation and the time and frequency shift invariance, which makes the interpretation of the resulting TFR maps easy. However, the major limitations are: (i) limited or bounded time and frequency resolution in

the TF space resulting in smearing of the TFR, (ii) the presence of cross-terms or artefacts between the neighbouring time-frequency structures.

AD is mainly useful for randomly appearing signal components in the time series. The TFR derived from the AD has explicitly no cross-term, which leads to clean and easy-to-interpret time-frequency maps of energy density. However, this is obtained at the cost of higher computational complexity which can be simplified using a well designed dictionary. The sparsity of AD can be obtained from the very redundant set of atoms, which represents the signal structures as a summation of limited number of atoms. The parameterization of the signal structures in terms of the amplitude, frequency, time of occurrence, time, and frequency span can be obtained from the TF dictionary used in AD.

References

Boashash, B., 1992. Time-frequency signal analysis–methods and applications. Longman Cheshire.

Boashash, B., 2003. Time Frequency Signal Analysis and Processing: Method and Applications. Elsevier, Oxford.

Boyd, S., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press, New York, NY, USA.

Bracewell, R. N., 2000. The Fourier Transform and its Applications. McGraw Hill Higher Education.

Chen, S., Donoho, D., Saunders, M., 2001. Atomic decomposition by basis pursuit. Society for Industrial and Applied Mathematics: Review 43 (1), 129–159.

Cohen, L., 1995. Time-frequency Analysis. Electrical Engineering signal processing. Prentice Hall PTR.

Cotton, C. V., Ellis, D. P., 2009. Finding similar acoustic events using matching pursuit and locality-sensitive hashing. In: 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, October 18-21, 2009, Mohonk Mountain House, New Paltz, NY, USA. IEEE, pp. 125–128.

Hlawatsch, F., Boudreaux-bartels, G., April 1992. Linear and quadratic time-frequency signal representations. Signal Processing Magazine, IEEE 9 (2), 21–67.

Hsu, S.-H., Huang, C.-L., 2001. Road sign detection and recognition using matching pursuit method. Image and Vision Computing 19 (3), 119–129.

Jouny, C. C., Franaszczuk, P. J., Bergey, G. K., 2003. Characterization of epileptic

seizure dynamics using Gabor atom density. Clinical Neurophysiology 114 (3), 426 – 437.

Mallat, S., 2008. A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way, 3rd Edition. Academic Press.

Mallat, S., Zhang, Z., 1992. Adaptive time-frequency decomposition with matching pursuits. Time-Frequency and Time-Scale Analysis, 1992., Proceedings of the IEEE-SP International Symposium, 7–10.

Mallat, S. G., Zhang, Z., 1993. Matching pursuits with time-frequency dictionaries. IEEE Trans. on Signal Proc. 41 (12), 3397–3415.

Oppenheim, A. V., Schafer, R. W., Buck, J. R., 1999. Discrete-time Signal Processing (2Nd Ed.). Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Pati, Y., Rezaiifar, R., Krishnaprasad, P. S., 1993. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. Twenty-Seventh Asilomar Conference on Signals, Systems and Computers 43, 40–44.

Penrose, R., 1955. A generalized inverse for matrices. Mathematical Proceedings of the Cambridge Philosophical Society 51, 406–413.

Rankine, L., Mesbah, M., Boashash, B., 2007. A matching pursuit-based signal complexity measure for the analysis of newborn EEG. Medical & Biological Engineering & Computing 45 (3), 251–260.

Tropp, J., Gilbert, A., 2007. Signal recovery from random measurements via orthogonal matching pursuit. Information Theory, IEEE Transactions on 53 (12), 4655–4666.

Chapter 4

Neonatal Seizure Detection using Atomic Decomposition with a Novel Dictionary

> You have to learn the rules of the game. And then you have to play better than anyone else. ––Albert Einstein

4.1 Introduction

Neonatal seizures can be identified and diagnosed using the EEG which requires continuous access to trained expertise for the interpretation in neonatal intensive care units (NICU). Currently, the detection of neonatal seizures NICUs requires a specific expertise to visually interpret the neonatal EEG by an experienced neurophysiologist or paediatric neurologist. This expertise is not available 24/7 in most NICUs. Due to this it is essential to develop methods for the automated detection of seizures in the NICU which can provide valuable information and can aid as a tool to the clinician facilitating medical intervention (Celka and Colditz, 2002b).

The automated detection of seizures attempts to classify the EEG into two states: non-seizure and seizure. The non-seizure EEG consists of background EEG, which can be modelled as an amplitude modulated, coloured random noise process contaminated by several sources of artefact of electrical and biological origin (Rankine et al., 2007, Roessgen et al., 1998, Stevenson et al., 2013). The seizure patterns in the newborn EEG, however, are more highly structured and characterized by periods of rhythmic spiking or repeated sharp waves (Gotman et al., 1997, Pellock

et al., 2007).

A number of groups have published methods to automatically detect neonatal seizures using linear, nonlinear and nonstationary signal analysis (Aarabi et al., 2007, Celka and Colditz, 2002a, Deburchgraeve et al., 2008, Gotman et al., 1997, Liu et al., 1992, Navakatikyan et al., 2006). A method for detecting seizures based on atomic decomposition (AD) was outlined in Rankine et al. (Rankine et al., 2007) and was inspired by the analysis of Durka et al. (Durka et al., 2001). AD is a non-stationary signal analysis technique, in which a non-stationary signal is optimally decomposed into atoms drawn from an over-complete (redundant) dictionary. The main advantage of AD is that the dictionary of atoms can be designed to suit the application, resulting potentially in a sparse signal representation; this sparsity is useful in signal classification. Several techniques have been proposed for the optimal selection of atoms in AD; these include the matching pursuit (MP), basis pursuit (BP) and orthogonal matching pursuit (OMP) algorithms (Chen et al., 2001, Mallat and Zhang, 1993, Tropp and Gilbert, 2007). Rankine et al. proposed a novel measure called relative structural complexity (RSC) which is based on analysing the convergence of an AD implemented with a MP estimate of atom weights. In this context, convergence refers to the rate at which AD via MP approximates the signal (Rankine et al., 2007). The RSC, however, is highly dependent on the coherence of the dictionary with the signal classes under analysis. For high seizure discrimination the atoms in the dictionary should have a high coherence with seizure (low RSC) and low coherence with non-seizure (high RSC). Rankine et al. used a dictionary, based on the modelling of neonatal EEG seizure, that consisted of linear frequency modulated (LFM) atoms and Gabor atoms to complete the dictionary (Rankine et al., 2007).

Recent developments in modelling the neonatal EEG may provide a better choice of dictionary (Stevenson et al., 2010). These advances suggest that using a dictionary based on short duration wideband atoms rather than narrowband atoms may result in an RSC that is better able to differentiate between seizure and non-seizure.

In this chapter, a novel dictionary for automated seizure detection is proposed. This begins with an experiment in which the sparsity of signal approximations using basis pursuit (BP), matching pursuit (MP) and orthogonal matching pursuit (OMP) techniques are compared. This is done to determine the best decomposition method and dictionary type (time-scale, frequency, time-frequency) to provide optimal discrimination between seizure and non-seizure. A novel dic-

tionary based on the Duffing oscillator model (Stevenson et al., 2010) is then developed, in which the atoms are highly coherent with neonatal seizure. The RSC feature is then modified for computational reliability as RSC is dependent on the type of dictionary used in AD. The use of the RSC with the proposed dictionary was compared to the RSC with several other dictionaries for the detection of neonatal seizures. The effect of dictionary size and decomposition level was also investigated. The performance of the RSC for neonatal seizure detection was then estimated on a large database of neonatal EEG using leave-one-out (LOO) cross validation.

4.2 Neonatal EEG seizure dataset

The dataset consisted of EEG recordings from 18 neonates who had seizures. The EEG recordings were recorded in the NICU of Cork University Maternity Hospital, Cork, Ireland. The patients were full-term neonates ranging in gestational age from 39 to 42 weeks. The EEG was recorded using the Viasys NicOne EEG system, with a sampling frequency of 256 Hz. The data were annotated using eight EEG channels in bipolar montage: F4-C4, C4-O2, F3-C3, C3-O1, T4-C4, C4-Cz, Cz-C3, and C3-T3. All seizures were annotated independently by two experienced neonatal electroencephalographers with the assistance of simultaneously recorded video. Any disagreement in annotations was resolved by consensus. The median recording length was 50.6h (IQR: 29.7–59.8h) with a combined recording duration of 826h. There were 1389 recorded seizures including both electrographic-only and electro-clinical seizures of focal, multi-focal and generalized types with a mean duration of 194s (median 249s, IQR 96–356s). The ratio of seizure to non-seizure duration had a median of 5.9%, IQR (4.2–12.1%). Additional details of this database can be found in table 4.1. All EEGs were recorded with informed parental consent and under ethical approval of the CUMH and University College Cork. All data were anonymized at the time of recording. The EEG was down-sampled from 256 to 32 Hz with an anti-aliasing filter set at 12.8 Hz, as the significant energy in the newborn EEG ($> 95\%$) does not exceed alpha band (8-12 Hz) (Scher et al., 1994, Victor et al., 2005). The data were also filtered with a single pole highpass filter with a cutoff of 0.5Hz and then segmented into 8 s epochs with a 4 s overlap.

Table 4.1: EEG dataset of newborns with seizures used in this thesis.

Patient	Record length(h)	Seizure event	Seizure duration		
			Mean	Min	Max
1	29.7	17	1′30″	17″	3′54″
2	24.7	3	6′10″	55″	11′09″
3	29.9	209	1′50″	11″	10′43″
4	47.5	84	1′38″	32″	9′58″
5	47.2	62	6′37″	20″	34′10″
6	19.2	46	1′8″	15″	4′17″
7	60.8	99	1′32″	14″	10′20″
8	49.5	17	5′56″	29″	19′14″
9	67.7	201	4′59″	13″	37′06″
10	59.8	41	4′51″	13″	34′46″
11	21.8	43	2′27″	17″	7′36″
12	54.4	150	1′36″	15″	10′08″
13	51.7	60	3′26″	19″	16′56″
14	22.8	21	8′13″	22″	39′03″
15	59.7	121	1′31″	10″	7′08″
16	76.4	190	5′03″	26″	34′37″
17	30.7	21	5′31″	27″	23′16″
18	63.0	4	9′34″	7′19″	13′22″
Total	816.7	1389			

4.3 Comparison of performance of AD techniques

The Gabor dictionary with MP has been used in detecting structural changes in EEG (Jouny et al., 2003). It was shown that the features obtained from AD can provide necessary information about the change in EEG signal structures. In this section, two experiments were performed which demonstrated the effect of different AD techniques and dictionary types on the performance of neonatal seizure detection. 100 seizure and non-seizure neonatal EEG epochs from the neonatal EEG database (each epoch is 8 s in duration downsampled from 256 Hz to 32 Hz) were used [1]. Figure 4.1 shows sample epochs used in these experiments.

[1]More details about the preprocessing and filtering is given in section 4.4.

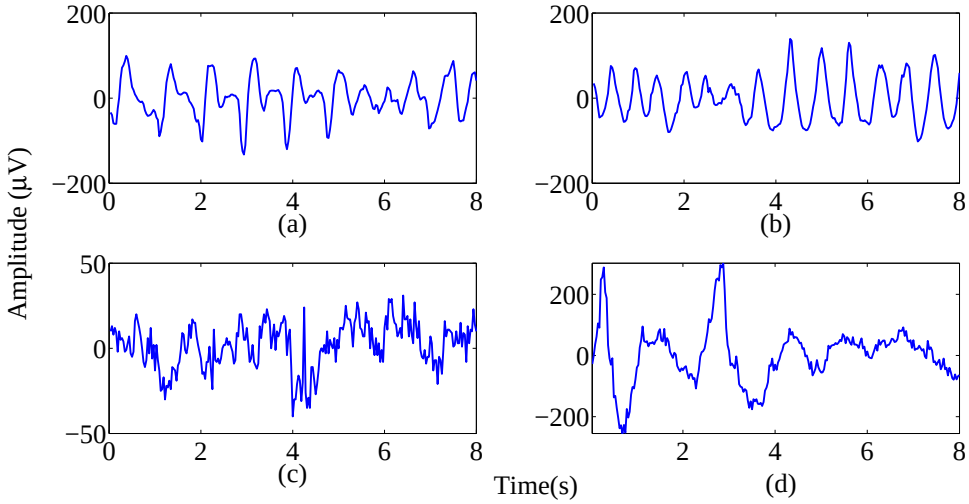


Figure 4.1: Sample epochs used in this experiment to demonstrate the performance of AD techniques to discriminate seizure (a,b) from non-seizure epochs (c,d).

4.3.1 Experiment 1: Effect of dictionary types on seizure detection

In this experiment, through the analysis of real newborn EEG data, it is shown that the selection of dictionary can influence the seizure detection accuracy. Three different dictionaries were used in this experiment[2]:

1. A Gabor dictionary consisting of translated, scaled and modulated versions of a Gaussian window (Mallat and Zhang, 1993).
2. A Fourier dictionary consisting of sinusoidal waveforms, and
3. A wavelet packet dictionary built from a Daubechies 10 quadrature mirror filter, consisting approximately of $Nlog_2N$ waveforms which is a family of orthonormal wavelet bases (Mallat and Zhang, 1993).

Two times overcomplete dictionaries were used to run the MP algorithm. To evaluate the performance, a performance metric known as *percentage time error* ($PTE\%$) was used which is given by:

$$PTE\% = 100\sqrt{\frac{\sum_{n=1}^{K}(x_o(n) - x_r(n))^2}{\sum_{n=1}^{K}(x_o^2(n))}} \tag{4.1}$$

[2]The properties of these dictionaries was discussed in chapter 3

where x_o and x_r are the original and reconstructed signal, respectively. For a given number of atoms K, the reconstruction error ($PTE\%$) would naturally be expected to be lower for seizure when compared to nonseizure signals if the atoms in the dictionary are coherent with the seizure signal. The difference in PTE defined as :

$$dPTE\% = |PTE_S - PTE_{NS}|\% \tag{4.2}$$

gives the level of separability between seizure and nonseizure signals for a given dictionary. Here PTE_S and PTE_{NS} are the reconstruction errors for seizure and nonseizure signals, respectively. Here K was restricted to 25 atoms.

Figure 4.2 shows the performance of different dictionaries using the MP algorithm. It can be seen that the Gabor dictionary requires less atoms to separate seizure from non-seizure class when compared to the Fourier and wavelet dictionaries. This experiment suggests that for detecting neonatal EEG signal transitions, TF dictionaries are more efficient when compared to frequency dictionaries (Fourier) or time-scale dictionaries (wavelet).

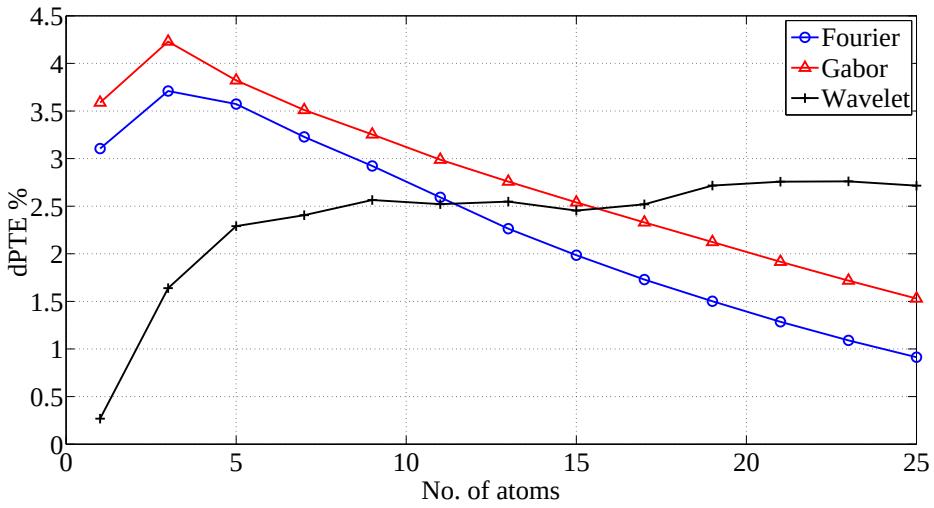
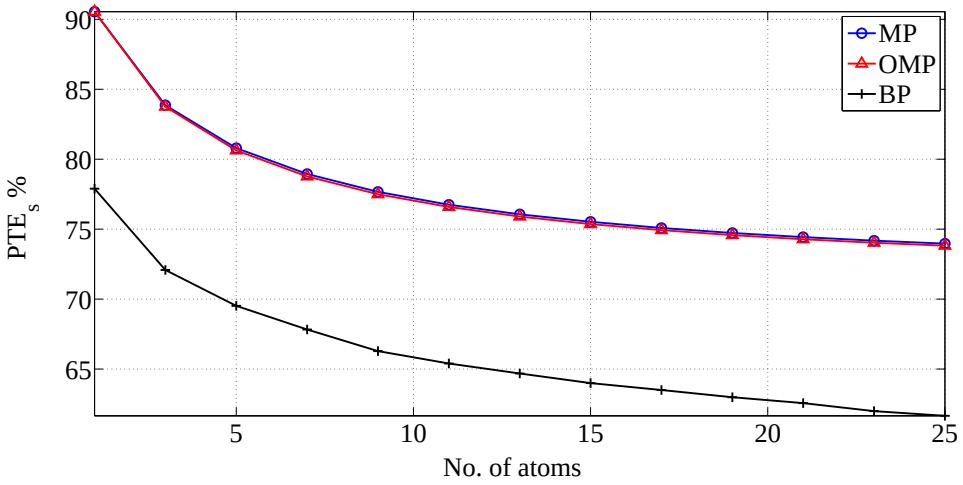


Figure 4.2: Illustration of the effect of dictionary types on discrimination level using the MP algorithm.

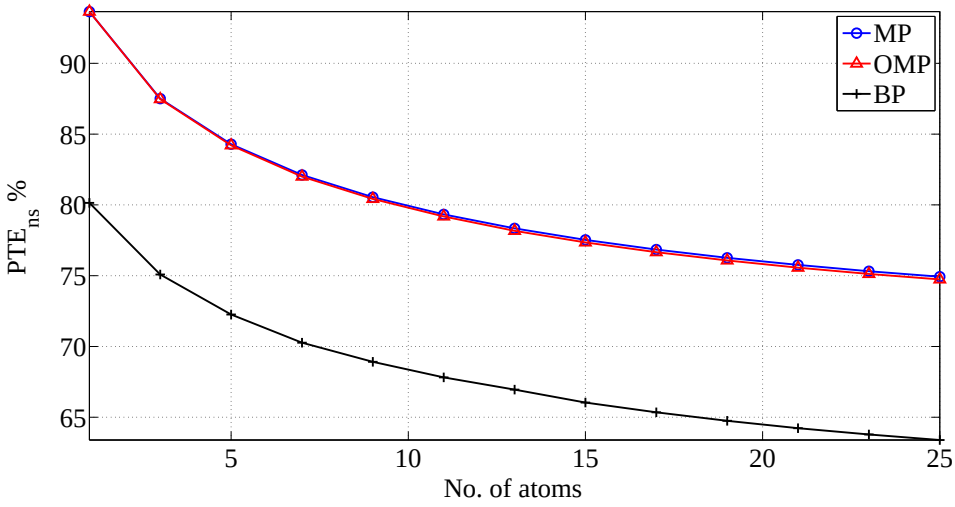
4.3.2 Experiment 2: Performance of AD techniques for seizure detection

In this experiment, the performance of MP, OMP and BP techniques were compared for discriminating seizure from non-seizure. Gabor dictionary was used for this experiment. Figure 4.3 shows the performance of the AD techniques in

representing seizure and non-seizure using $K = 25$ atoms. It can be seen that as the number of atoms are increased, the $PTE\%$ for both seizure and nonseizure decreases. However, from figure 4.4, it can seen that the OMP outperforms BP and MP in discriminating seizure from non-seizure epochs using 4-5 atoms. This suggests that features obtained from the AD using OMP is efficient in discriminating neonatal EEG seizure from non-seizure, or in other words, in neonatal seizure detection. For this reason, OMP was used as the preferred AD technique in this thesis.



(a)



(b)

Figure 4.3: Illustration of the reconstruction error obtained using 100 eight second epochs each of (a) seizure, and (b) non-seizure EEG using different AD techniques for the increasing number of decomposition atoms.

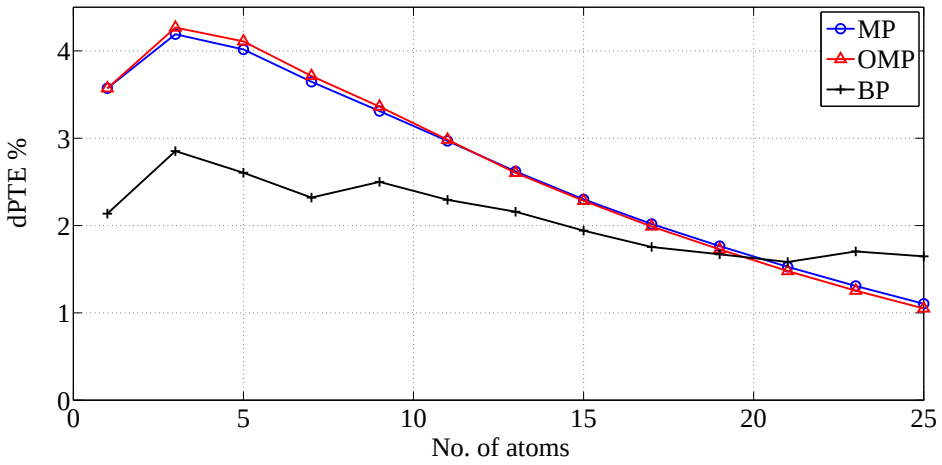


Figure 4.4: Illustration of the variation in the discrimination level with the increase in number of atoms. OMP always outperformed MP when tested on several epochs.

4.4 Seizure detection using atomic decomposition

AD is a technique that uses a large dictionary of signal atoms to efficiently represent a given signal. Given a real dictionary of M atoms $\mathbf{D} = [\boldsymbol{\phi}_1\ \boldsymbol{\phi}_2\ \ldots\ \boldsymbol{\phi}_M] \in \mathbb{R}^{N\times M}$, where each column represents a single atom of length N and $M \geq N$, a signal $\mathbf{x} \in \mathbb{R}^N$ can be represented as

$$\mathbf{x} = \mathbf{D}\boldsymbol{\gamma}, \tag{4.3}$$

where $\boldsymbol{\gamma}$ is a $M \times 1$ vector containing the analysis coefficients (Goodwin and Vetterli, 1999). The aim of AD is to estimate $\boldsymbol{\gamma}$ to yield a suitable approximation of the signal. Ideally, $\boldsymbol{\gamma}$ should contain a minimum number of non-zero coefficients for a given approximation accuracy (a sparse solution). The MP was originally proposed by Mallat and Zhang in 1993 (Mallat and Zhang, 1993). It is a greedy algorithm that selects atoms using a maximum inner product criteria. The coefficient value is the maximum inner product. MP is iterated until a certain level of signal approximation is achieved. BP uses linear programming to minimize the L_1-norm, but unlike MP, the BP decomposition technique does not select atoms iteratively. The use of the L_1-norm attempts to minimize the number of atoms used to approximate a signal at a given level. It was shown that the MP provides a sparser signal approximation when compared to the BP (Rankine et al., 2007). The OMP uses the same criteria as the MP for the atom selection but estimates

the entire coefficient vector using a pseudo-matrix inversion at each iteration in order to minimise the mean square error. As a result, the OMP will not select an atom more than once and provides a better signal approximation as described in section 4.3.

In this work, the OMP was utilised to decompose the neonatal EEG with a *complex dictionary* (Tropp and Gilbert, 2007) as the OMP has a faster convergence rate. The use of a complex dictionary improves the convergence rate of the OMP algorithm as additional phase information can improve the alignment of atoms in time (Goodwin and Vetterli, 1999). Here convergence refers to the rate at which the OMP approximates the signal. The OMP algorithm using a complex dictionary can be summarized as follows (Goodwin and Vetterli, 1999, Tropp and Gilbert, 2007):

1. Initialize the residual $\mathbf{r}_0 = \mathbf{x}$, the index set $\Psi_0 = \emptyset$, the dictionary $\mathbf{D}_0 = \emptyset$ and the iteration count $k = 1$.
2. Find $\check{i} = \arg \max_{i \notin \Psi_{k-1}} |\mathbf{r}_{k-1}^T \boldsymbol{\phi}_i|$, $\boldsymbol{\phi}_i \in \mathbf{D}$.
3. Update the index set $\Psi_k = \Psi_{k-1} \cup \check{i}$ and the active dictionary, $\mathbf{D}_k = [\mathbf{D}_{k-1} : \boldsymbol{\phi}_{\check{i}}] \in \mathbb{C}^{N \times k}$.
4. Estimate $\boldsymbol{\gamma}_k$ by solving a least squares problem using Moore-Penrose pseudoinverse method (Penrose, 1955), $\boldsymbol{\gamma}_k = \arg \min_{\gamma \in \mathbb{C}_k} ||\mathbf{x} - 2\Re\{\mathbf{D}_k \boldsymbol{\gamma}\}||_2$. .
5. Update the residual, $\mathbf{r}_k = \mathbf{r}_{k-1} - 2\Re\{\mathbf{D}_k \boldsymbol{\gamma}\}$.
6. Update the iteration count, $k = k + 1$. Repeat from step 2 until the given stopping criteria is reached ($k = K$).

After K iterations, the OMP provides the sparse approximation of the input signal $\mathbf{x}$ as,

$$\hat{\mathbf{x}} = 2\Re\{\mathbf{D}_K \boldsymbol{\gamma}_K\}. \tag{4.4}$$

Given a full rank dictionary and large enough K, then $\mathbf{x} = \hat{\mathbf{x}}$ can be achieved.

AD via MP has been used as the basis of a method to detect neonatal seizures in the EEG (Rankine et al., 2007). This method exploited differences in the rate of convergence of the MP algorithm for seizure detection (the dictionary was chosen so that AD converged faster for seizures compared to non-seizure). The rate of convergence was measured using the number of atoms required to reach a set signal to error ratio (SER). The SER at a given level of decomposition, K, is

given by

$$\text{SER}^K = 20\log_{10}\left(\frac{||\mathbf{x}||_2}{||\mathbf{r}_K||_2}\right) \tag{4.5}$$

where $\mathbf{x}$ is the input EEG signal and $\mathbf{r}_K$ is the residual obtained after K iterations of AD. In this present study, the SER after K OMP iterations was used to define the RSC which is an improvement on the measure used in (Rankine et al., 2007). A set decomposition level, rather than approximation level, was used as it ensures that the OMP algorithm will be computationally tractable.

In order to enhance the ability of the RSC to separate seizure from non-seizure an application specific dictionary was introduced by Rankine et al. (Rankine et al., 2007). This dictionary was based on a model of EEG seizure and consisted of a series of linear frequency modulated (LFM) sinusoids supplemented by Gabor atoms in order to generate a full rank dictionary. This current study now investigates the use of alternate dictionaries with the aim of improving seizure detection. The results of this investigation were used to develop a seizure detection algorithm (SDA) based on AD (see figure 4.5). The input signal is passed through an anti aliasing filter set at 12.8Hz, downsampled to 32 Hz and segmented into 8 s epochs with 4 s overlap. Sharp transients in the RSC output are then suppressed using a median filter of 32s in duration. The maximum RSC across all channels was then computed. The RSC decision threshold was varied and seizure was detected if it exceeded a set threshold. An adaptive collar was then applied to extend the duration of the preliminary detection. An automated annotation of seizure is then obtained. The main structural difference to the SDA proposed by Temko et al. (apart from the feature used) is in the post-processing stage (Temko et al., 2011a). A mean operation is substituted with a median operation and the fixed collar is replaced with an adaptive collar. The adaptive collar extends the detection results in proportion to the initial detection duration and is constrained from 30 to 80s such that $ac = 30$s when $T_d < 30$s, $ac = T_d$ when $30 \leq T_d \leq 80$s, and $ac = 80$s when $T_d > 80$s, where T_d is the duration of the detected seizure (in seconds). The assumption of this collar is that the decay in seizure amplitude is in proportion to its duration.

4.5 Dictionaries for seizure detection

In this section, the design of different dictionaries used in this thesis are described. Consider the general mother atom $g(t;\lambda)$ for some particular parameter vector

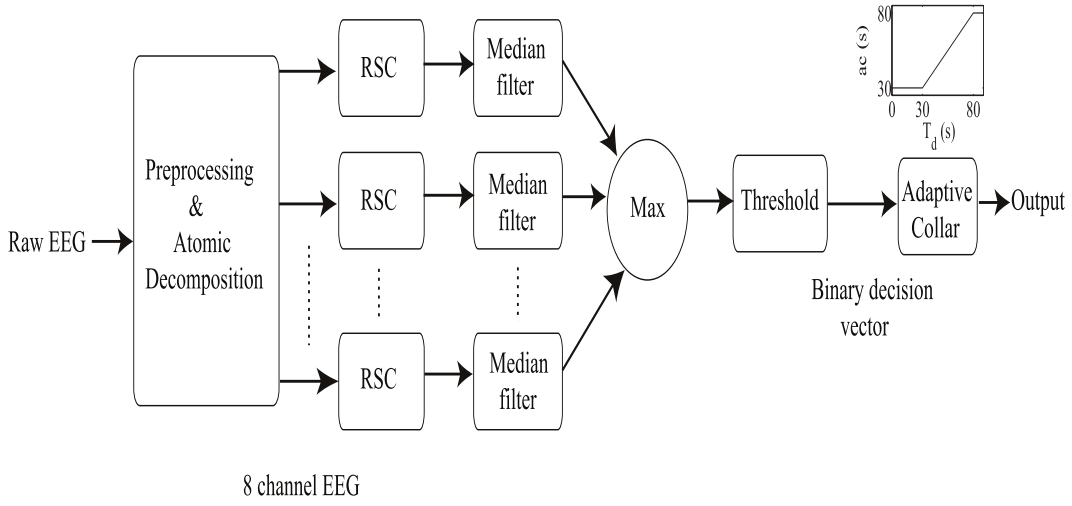


Figure 4.5: Neonatal seizure detection system based on atomic decomposition.

$\lambda \in \mathbf{\Lambda}$. The analytic associate of each atom $\bar{g}(t;\lambda) = \mathcal{H}(g(t;\lambda)$ can then be generated using the Hilbert transform.

The complex dictionary $\mathbf{D} \in \mathbb{C}^{N\times M}$ is then constructed from the sampled version of the complex mother atom. The complex dictionary is $\mathbf{D} = [\boldsymbol{\phi}_1\ \boldsymbol{\phi}_2\ \ldots \boldsymbol{\phi}_M]$, where the i^{th} complex atom $\boldsymbol{\phi}_i$ is the time sampled version of the complex mother atom with parameter vector choice $\lambda_i \in \mathbf{\Lambda} = \{\lambda_1, \lambda_2, \ldots, \lambda_M\}$. Hence, for a given signal duration of T and sample time T_s seconds, the n^{th} element of $\boldsymbol{\phi}_i$ can be expressed as

$$\boldsymbol{\phi}_i(n) = g(nT_s;\lambda_i) + j\bar{g}(nT_s;\lambda_i) \tag{4.6}$$

where $0 \leq nT_s < T$.

4.5.1 Gabor dictionary: $\mathbf{D}_G \in \mathbb{C}^{N\times M_G}$

As a benchmark, a Gabor dictionary consisting of translated (α), scaled (m) and modulated (β) versions of a Gaussian window is first investigated to test the proposed SDA,

$$g_G(t;\lambda_G) = \frac{1}{\sqrt{m}} e^{-\pi\left(\frac{t-\alpha}{m}\right)^2} \sin(\beta t). \tag{4.7}$$

Here $\lambda_G = [\alpha, m, \beta]^T \in \mathbf{\Lambda}_G$ are the time-frequency parameters, $T = 8s$ is the time duration of the EEG epoch, $T_s = 1/f_s$, $f_s = 32$ Hz and $N = 256$.

The result of AD on an epoch of seizure and non-seizure EEG is shown in Fig. 4.6(i). Fig. 4.6(ii) shows the changes in RSC with respect to the number of

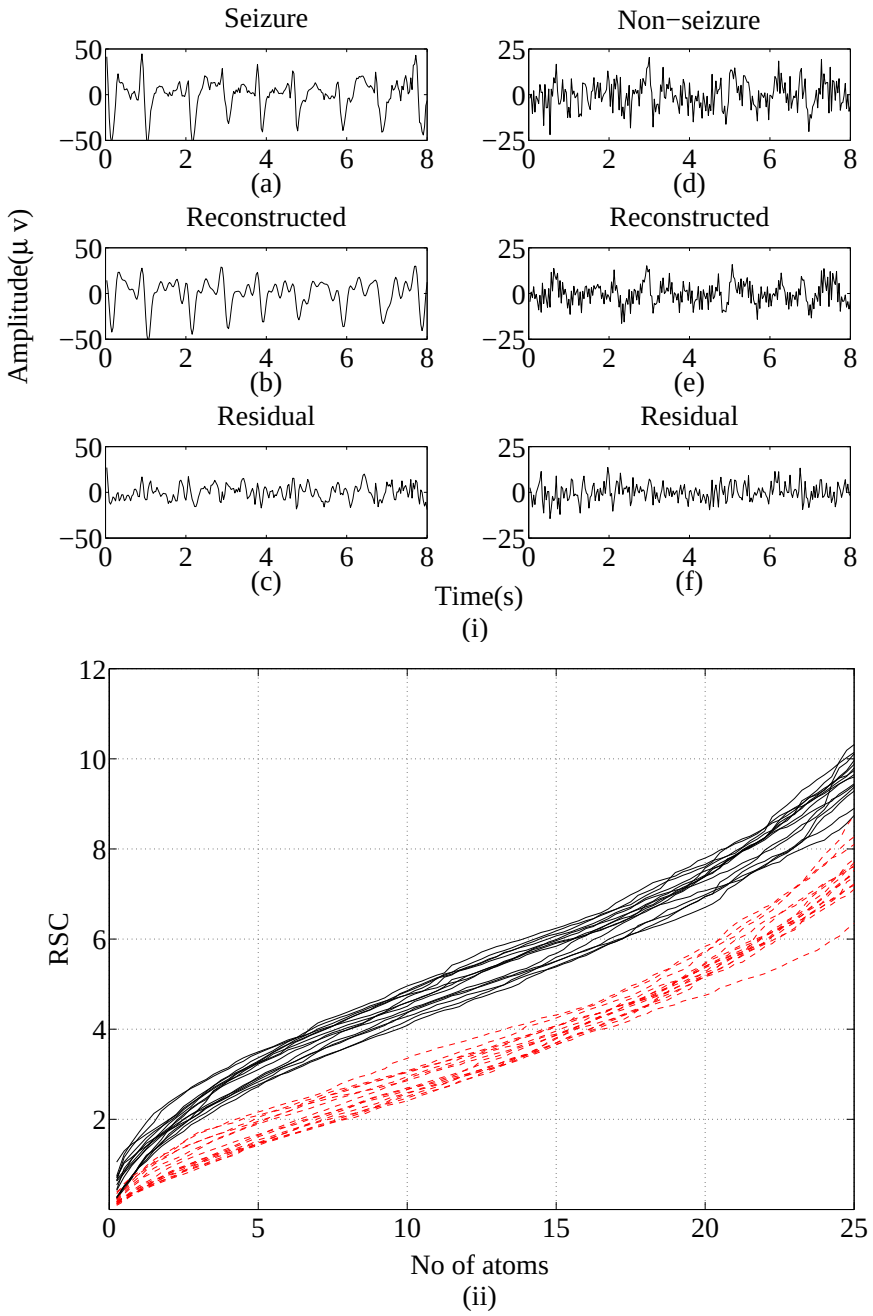


Figure 4.6: Illustration of atomic decomposition of an 8 s EEG epoch using a variable number of Gabor atoms. (i) a,b,c are the actual ($\mathbf{x}$), reconstructed ($\hat{\mathbf{x}}$), residual seizure epoch ($\mathbf{r}$), and d,e,f are the actual ($\mathbf{x}$), reconstructed ($\hat{\mathbf{x}}$), residual non-seizure epoch ($\mathbf{r}$) respectively using 15 atoms, and (ii) RSC for fifteen 8 s seizure and non-seizure epochs for 1–25 atoms (−seizure, −− non-seizure).

atoms used for decomposition. As the Gabor dictionary is more coherent with seizure, due to its inherent structure, AD provides a more accurate representation of seizure, at a set decomposition level, than non-seizure.

4.5.2 Linear frequency modulated (LFM) dictionary: $\mathbf{D}_L \in \mathbb{C}^{N \times M_L}$

The dictionary proposed by Rankine et al., consisted of LFM atoms as it had been shown that the variation in the frequency content of seizure could be broadly categorized with a piecewise linear function (Rankine et al., 2007). The set of LFM atoms used were of the form

$$g_L(t; \lambda_L) = A(\lambda_L) \cos\left(\frac{\pi}{f_s}\left[\frac{at^2}{f_s} + 2bt\right]\right). \tag{4.8}$$

Here the parameter vector $\lambda_L = [a\ b]^T \in \mathbf{\Lambda}_L$, f_s is the sampling frequency (Hz), a (Hz/s) is the frequency rate and b (Hz) is the initial frequency. For each choice of $\lambda_L \in \mathbf{\Lambda}_L$, the normalizing constant $A(\lambda_L)$ is chosen such that $||g_L(t; \lambda_L)||_2 = 1$.

4.5.3 Duffing dictionary: $\mathbf{D}_D \in \mathbb{C}^{N \times M_D}$

Recent advances in modelling the neonatal EEG seizure signal offer the potential to improve the dictionary used for seizure detection (Stevenson et al., 2010). An alternate model of neonatal EEG seizure was based on a Duffing oscillator. Seizure was modelled as

$$\ddot{x}(t) = u(t) - c\dot{x}(t) - k_2 x^3(t), \tag{4.9}$$

where $x(t)$ represents the recorded EEG voltage, $[c, k_2]$ are the nonlinear spring constant and damping coefficient, respectively, and $u(t)$ is an impulse train with time-varying period (Stevenson et al., 2010). The parameters $[c, k_2]$ control the time and frequency characteristics of the Duffing oscillator, respectively. The impulse response of the Duffing oscillator has time-varying frequency content and was used to approximate the morphology of the underlying seizure waveform which is not sinusoidal as shown in figure 4.7. In addition, a range of seizure morphologies (examples shown in figure 1.2) can be modelled by selecting appropriate values of $[c, k_2]$ as shown in figure 4.8.

The set of Duffing atoms used were of the form

$$g_D(t; \lambda_D) = h\left((t - \tau); c, k_2\right) \tag{4.10}$$

where $\lambda_D = [c, k_2, \tau]^T \in \mathbf{\Lambda}_D$, $\tau \in \{T_s, 2T_s, \ldots, NT_s\}$ is the time shift, and $h(t)$ is

the impulse response of the Duffing oscillator.

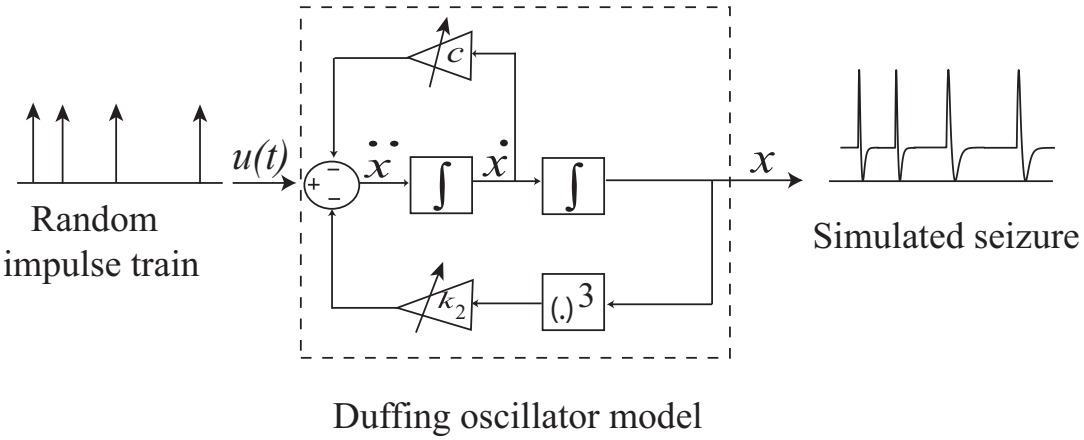


Figure 4.7: The model of the Duffing oscillator to generate simulated seizure signals. x represents the simulated seizure, $[c, k_2]$ controls the time and frequency characteristics of the Duffing oscillator and u is the unit impulse voltage with time varying period.

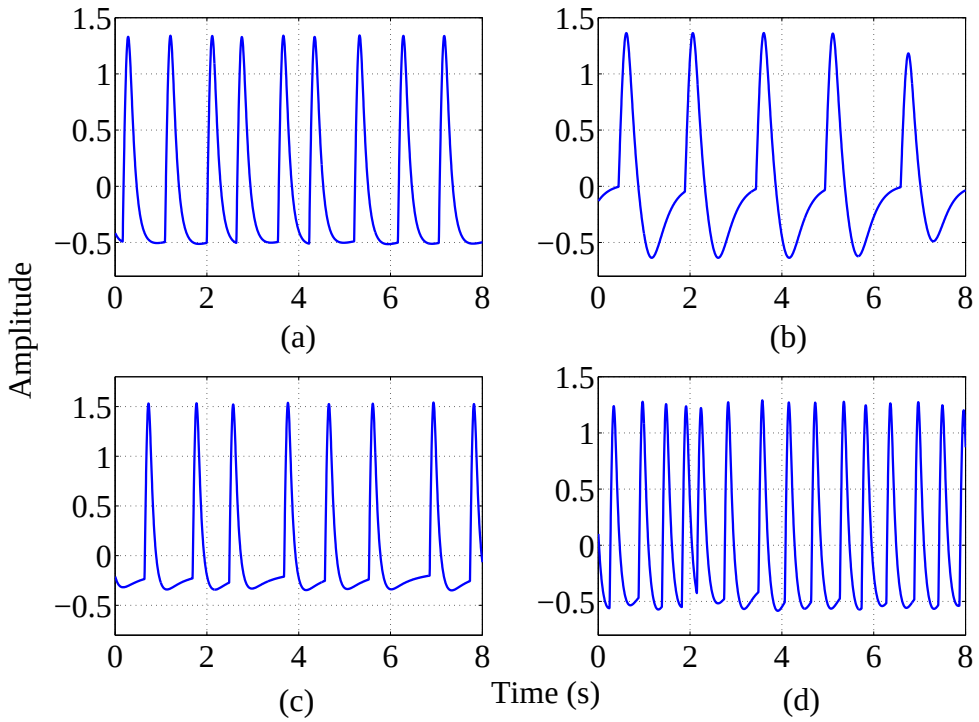


Figure 4.8: Examples of several types of newborn EEG seizure morphology simulated using Duffing oscillator model of figure 4.7.

4.5.4 Pseudo-periodic Duffing dictionary: $\mathbf{D}_{\tilde{D}} \in \mathbb{C}^{N \times M_{\tilde{D}}}$

This dictionary was based on a pseudo-periodic sequence of Duffing outputs which captures both the morphological and repetitive characteristics of seizure in a single atom. As the random nature of the periodicity used in Stevenson et al., (Stevenson et al., 2010) precludes a systematic implementation of a dictionary, the linear instantaneous frequency laws outlined in Rankine et al. (Rankine et al., 2007) were used instead. This results in an atom defined as,

$$g_{\tilde{D}}(t;\lambda_{\tilde{D}}) = \sum_{p=0}^{P(a,b)-1} h\left((t-\tau_p(a,b));c,k_2\right) \tag{4.11}$$

where $\lambda_{\tilde{D}} = [c, k_2, a, b]^T$,

$$\begin{aligned} \tau_p(a,b) &= f_s\left(\frac{-b+\sqrt{b^2+a(2p+\frac{1}{2})}}{a}\right), \\ p &= [0,1,\cdots P(a,b)-1], \text{ and} \\ P(a,b) &= \left\lfloor \frac{2aT^2+4\frac{b}{a}T-1}{4} \right\rfloor, \end{aligned} \tag{4.12}$$

where ($\lfloor . \rfloor$) is the floor operator.

The various atoms described above are illustrated in Fig. 4.9(a-d). As an example, a complex Duffing oscillator atom, obtained using the Hilbert transform is shown in Fig. 4.9e. In addition to the above mentioned dictionaries, a standard Fourier dictionary ($\mathbf{D}_F$) is also used for the purpose of comparison.

4.6 Training and testing

A leave-one-out (LOO) cross-validation method was used in this work to assess the performance of the proposed algorithm for patient independent seizure detection. This approach is known to provide an almost unbiased estimation of the generalization error (Vapnik and Kotz, 2006) and is in keeping with the development of a patient independent seizure detection algorithm (Temko et al., 2011a). For each iteration within the LOO framework adopted here, the continuous unpruned dataset from one neonate was first reserved as an unseen testing set. A training set consisting of 85min of seizure data and 850min of non-seizure data was then selected from the EEG data of the remaining 17 neonates (5min of

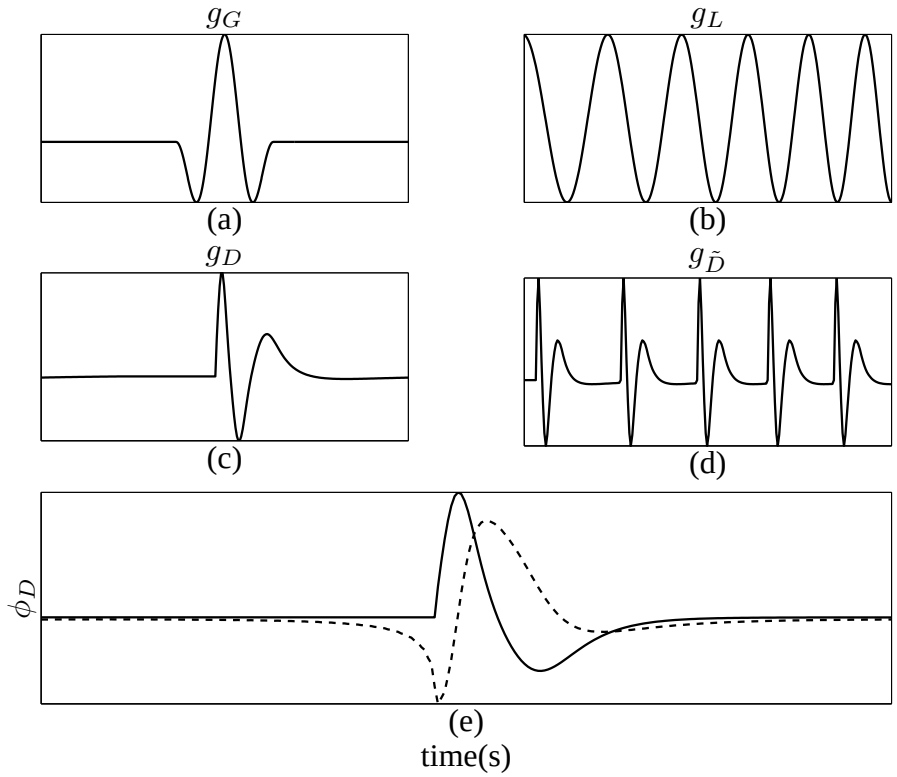


Figure 4.9: Example atoms in the dictionaries corresponding to (a) Gabor, (b) LFM, (c) Duffing, and (d) pseudo-periodic Duffing dictionaries, respectively. (e) A complex Duffing oscillator atom obtained after Hilbert transform ($-$ real, $--$ imaginary atom).

seizure and 50min of non-seizure per patient). The parameters of the seizure detection scheme (dictionary type, size and decomposition level) were then selected to provide the optimal performance over the training set. This trained algorithm was then tested on the continuous unpruned multi-channel EEG data from the remaining (left out) neonate and the performance reported. This process was repeated (total number of iterations here is 18) until the data from each neonate had been used once for testing.

The area under the receiver operator characteristic (ROC) curve (AUC) was used as an epoch based performance metric for the training phase and for each unseen test neonate. The ROC curve displays the relationship between the sensitivity of the algorithm and its specificity as a single classification threshold is varied. The *sensitivity* and *specificity* are defined as $TP/(TP + FN)$ and $TN/(TN + FP)$ respectively, where TP is the number of seizure epochs correctly detected as seizure, FP is the number of non-seizure epochs incorrectly detected as seizure, TN is the number of non-seizure epochs correctly detected as non-seizure, and FN is the number of seizure epochs incorrectly detected as non-seizure. The

individual ROC curves of all the test neonates were combined (mean or median) to provide an estimate of the performance curve of the algorithm over unseen data. The AUC provides a good estimate of the ability of the algorithm to discriminate between a seizure and a non-seizure epoch from a new, unseen neonate (Temko et al., 2011b).

For each dictionary (except $\mathbf{D}_F$), an overcomplete dictionary was first constructed from d subdictionaries as

$$\mathbf{D}_D = [\mathbf{D}_1 \ \mathbf{D}_2 \cdots \mathbf{D}_d] \in \mathbb{C}^{N \times dN}. \tag{4.13}$$

In the training phase, the P "best" subdictionaries are then selected from this overcomplete dictionary, to produce the optimal AUC. This then yields an optimal trained dictionary of size $M = PN$.

For the Gabor overcomplete dictionary, 7 subdictionaries were constructed with atom parameters chosen from the dyadic sequence of integers: $m = 2^q, 0 \leq q \leq L$, $N = 2^L$, $\alpha \in \{2, 4, 8, 16, 32, 64, 128\}$ and $\beta \in \{128, 64, 32, 16, 8, 4, 2\}$.

For the LFM overcomplete dictionary, a single $(N \times N/2)$ LFM subdictionary was constructed for $N/2$ possible choices of a and b ($a \in [-0.06, 0.06]Hz/s$ and $b \in [0.25, 8.5]Hz$). In order to span the TF plane, additional Gabor atoms were used to fill the high frequency segment of the TF plane (8-16) Hz (Mallat and Zhang, 1993, Rubinstein et al., 2010). This dictionary was augmented with 6 $(N \times N)$ Gabor subdictionaries to form an overcomplete dictionary.

In the Duffing and pseudo-periodic Duffing dictionary, it was important that the overcomplete dictionary captured the wide variety of seizure impulses seen in practice. For the Duffing overcomplete dictionary, 37 $(N \times N)$ subdictionaries were generated by sampling the distribution of $[c, k_2]$ as shown in figure 4.10. This distribution was used to model several morphologies of neonatal seizure in (Stevenson et al., 2010). However, for the pseudo-periodic Duffing overcomplete dictionary each subdictionary corresponded to N choices of a and b obtained from the LFM dictionary.

The optimized method for estimating the RSC was then tested on the full recording of the remaining neonate and used to generate an automated annotation of seizure (see Fig. 4.5). The level of agreement between the human and automated annotation of seizure was then assessed using time ($sensitivity$ and $specificity$) and event (seizure detection rate (SDR) and false detections per hr (FD/h)) based metrics (Temko et al., 2012). After testing, a subset of false and true seizure de-

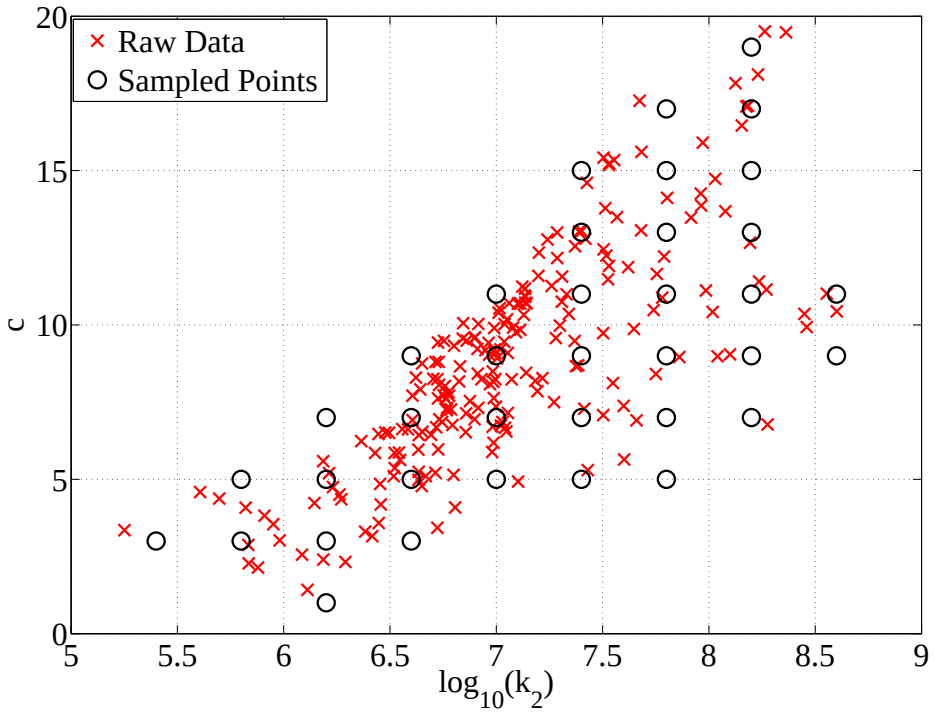


Figure 4.10: The joint distribution of the parameters of the Duffing oscillator that best represents neonatal EEG seizure, where **o** denotes the parameters that were used to generate mother atoms to build the Duffing dictionary.

tections were analyzed (Greene et al., 2008).

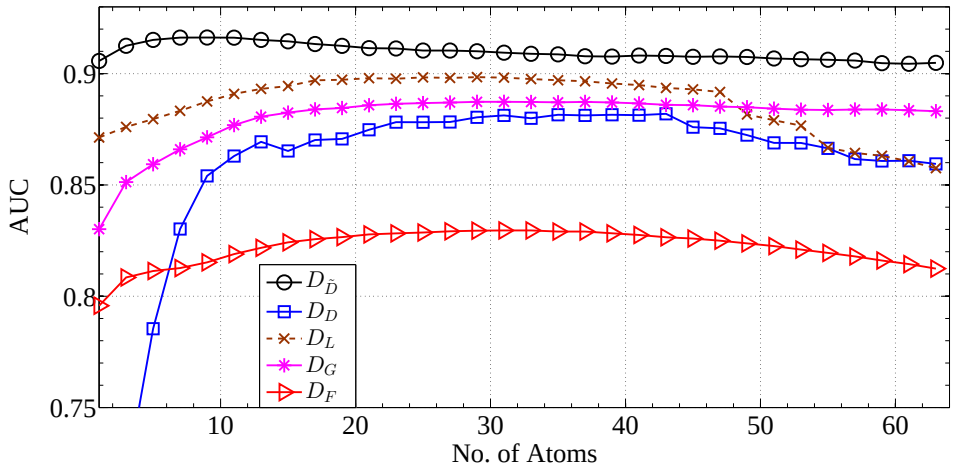
4.7 Results and discussion

The effect of dictionary type, dictionary size and decomposition level is shown in Fig. 4.11. The maximum likelihood estimate across the training iterations would result in a RSC calculated using a pseudo-periodic Duffing dictionary ($\mathbf{D}_{\tilde{D}}$) of size $2N$ and a decomposition level of 5 atoms. A median AUC of 0.92 (mean = 0.91) was obtained across training folds.

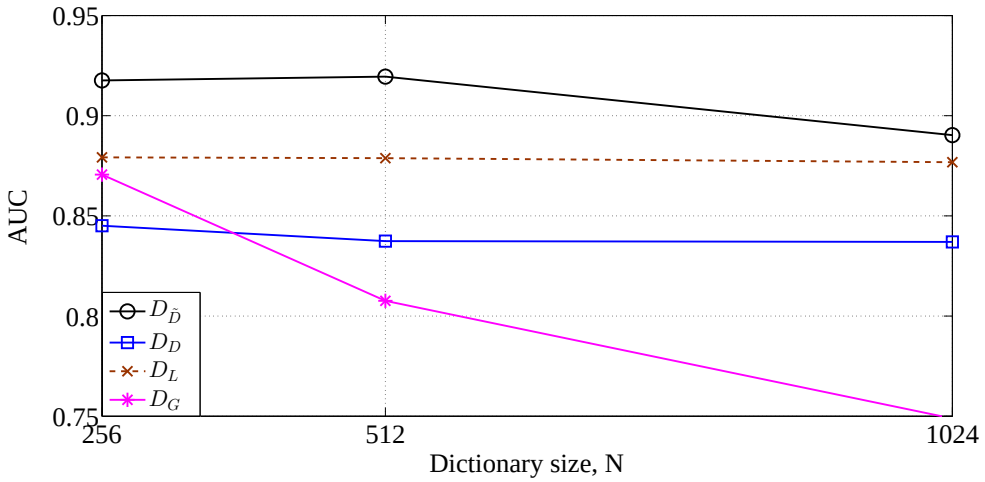
The performance of the SDA for each patient (test fold) is given in Table 4.3. The results shown in Table 4.4 demonstrate how the performance of the algorithm, as measured by the mean or median AUC of the ROC across all testing folds, depends on the dictionary type and the use of the adaptive collar. The use of the pseudo-periodic Duffing dictionary as a basis for AD and the additional of post-processing stages improved the performance of the SDA. Time and event based metrics at 0.1, 0.5 and 1.0 FD/h are shown in Table 4.5. The SDA has a

Table 4.2: A summary of the optimal parameters selected during training. Results presented as median (IQR) or † most commonly selected (selections)

Parameters	Optimal value
Atoms	5 (4.0-5.2)
Dictionary type†	$\mathbf{D}_{\tilde{D}}(18)$
Dictionary size†	2N (18)
AUC	0.915 (0.91-0.92)
RSC (seizure), dB	5.3 (4.7-6.5)
RSC (non-seizure), dB	1.6 (1.3-2.2)



(a)



(b)

Figure 4.11: Training results for several different dictionaries on the training sets. a) AUC v number of atoms b) AUC v dictionary size. AUC is the median AUC across 18 training folds of the LOO. Note, the Fourier dictionary cannot be increased above 256 atoms.

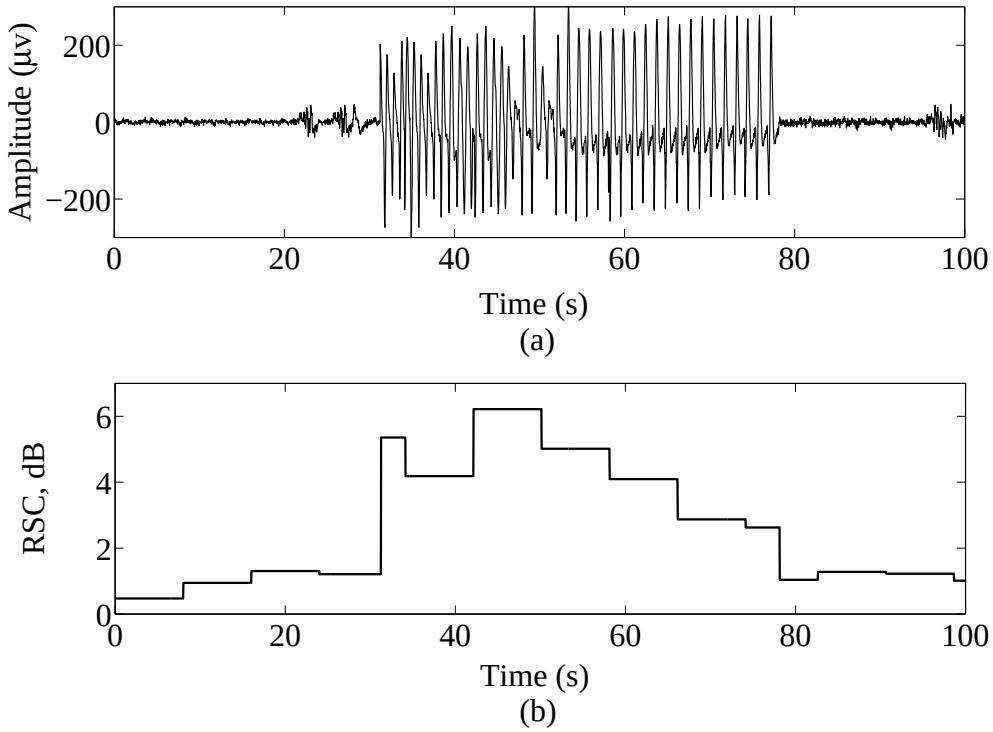


Figure 4.12: Example of the behavior of the proposed RSC for patient 3. (a) raw 100 seconds single channel (F4-C4) EEG with a seizure event emerging from low voltage activity, and (b) its corresponding RSC. Note that RSC increases in the presence of seizure as the atoms in the dictionary are coherent with seizure which provides improved reconstruction of seizure compared to non-seizure.

median seizure detection rate of 82.3% at 1FD/h and 56.5% at 0.1FD/h (a more clinically relevant threshold (Lawrence and Inder, 2010)). An example illustrating the behaviour of the RSC corresponding to seizure and non-seizure is shown in Fig. 4.12. It can be seen that the RSC increases in the presence of seizure since the atoms selected during AD are coherent with seizure, thereby decreasing the reconstruction error. A comparison of the SDA performance and other methods reported in the literature is shown in Fig. 4.13. While the comparison of SDA performance is fraught with difficulty (Temko et al., 2011a), the proposed SDA has comparable performance with existing methods. Figure 4.14 shows the affect of the postprocessing stage on the SDA performance.

The signal characteristics of the false and detected seizures for a defined false alarm rate of 1FD/h were investigated with other quantitative measures of EEG (Greene et al., 2008). The results are shown in Table 4.6. The spectral entropy and the AR model fit show significant difference between correctly detected seizures and false detections across the cohort of 18 neonates.

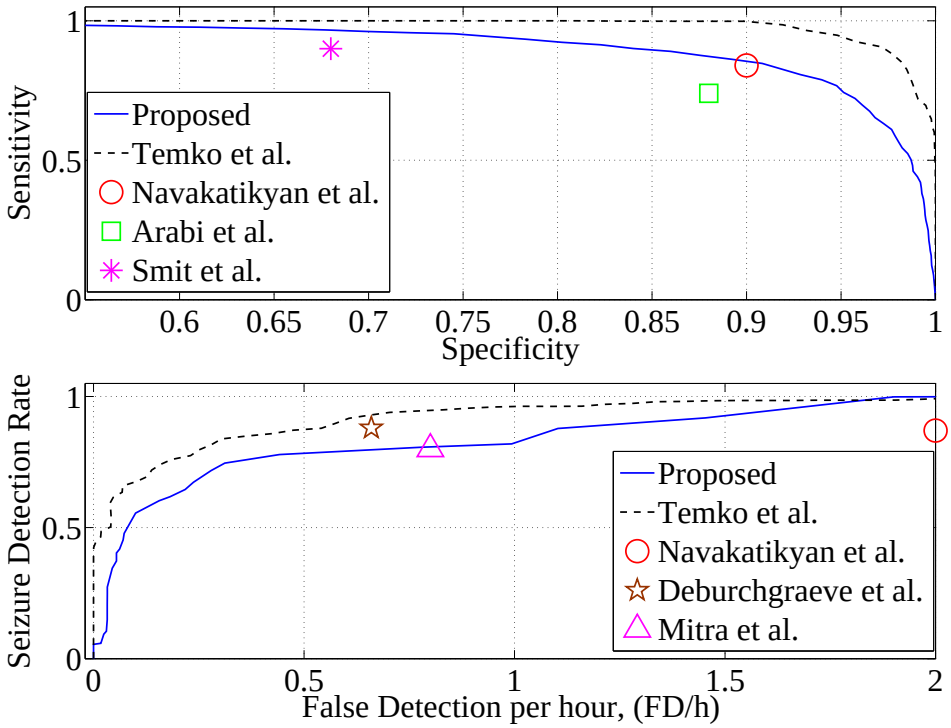


Figure 4.13: The performance curves of the proposed SDA. These are the ROC and the seizure detection rate/false detections per hour curves, determined as the median over the 18 unseen neonates in the LOO cross-validation. These curves present an almost unbiased estimate of the performance of the proposed SDA on unseen data. As a comparison, the performance curves of Temko et al. (Temko et al., 2011a) are included. The reported results of other key research groups are also reported here for comparison (Aarabi et al., 2007, Deburchgraeve et al., 2008, Mitra et al., 2009, Navakatikyan et al., 2006, Smit et al., 2004).

Fig. 4.15 shows the performance of the proposed SDA with respect to the seizure duration. Seizures less than 30s in duration were more difficult to detect and the largest amount of false detections were less than 30s in duration. EEG recordings from 3 neonates resulted in significantly reduced SDA performance (figure 4.17). Fig. 4.17 shows the distribution of mean RSC for high (15/18) and low (3/18) performing neonates.

It should be noted that there is a large difference in the training and testing results (mean AUC = 0.91 vs 0.79). This is most likely due to the fact that the training data consisted of single channel EEG epochs whereas testing is performed on multichannel EEG recordings which require a post-processing stage (a maximum across EEG channels) to generate a decision. The affect of this maximum operation is not examined during training. The use of further post-processing (a median filter and collaring operation), however, alleviated this difference (mean

Table 4.3: Performance of the proposed SDA for each unseen test patient. Here, ID=Patient ID, Sens=sensitivity (%), Spec=specificity (%), SN=Number of seizures, MSD=Mean seizure duration (seconds) and $SDR_{0.5}$=Seizure detection rate at 0.5 FD/h.

ID	AUC	Sens	Spec	SN	MSD	$SDR_{0.5}$
1	0.74	64.3	66.5	17	90	52.9
2	0.70	65.2	67.8	3	370	66.6
3	0.94	86.4	87.5	209	110	72.1
4	0.90	84.2	85.6	84	98	77.6
5	0.63	58.6	56.8	62	397	55.3
6	0.85	74.3	76.4	46	68	65.4
7	0.91	83.1	81.7	99	92	85.5
8	0.89	81.3	78.5	17	356	76.5
9	0.92	82.5	87.6	201	299	75.3
10	0.91	83.4	84.5	41	291	68.1
11	0.99	95.6	94.4	43	147	86.2
12	0.96	90.3	89.5	150	96	84.5
13	0.95	91.2	90.4	60	206	98.5
14	0.95	91.6	88.7	21	493	85.6
15	0.92	82.3	81.5	121	91	62.3
16	0.92	84.5	81.3	190	303	81.2
17	0.91	83.4	87.6	21	331	57.5
18	0.96	91.2	90.5	4	574	99.4

Table 4.4: The effect of dictionary type and post-processing on the performance of the proposed SDA when applied to each test fold (N=18). Mean, median and IQR of the AUC are presented.

Dictionary type	without post-processing		with adaptive collaring	
	Mean	Median (IQR)	Mean	Median (IQR)
$\mathbf{D}_{\tilde{D}}$	0.79	0.81 (0.72–0.84)	0.88	0.91 (0.87–0.95)
$\mathbf{D}_{L}$	0.72	0.73 (0.68–0.81)	0.76	0.78 (0.70–0.87)
$\mathbf{D}_{D}$	0.68	0.69 (0.58–0.77)	0.75	0.77 (0.64–0.87)
$\mathbf{D}_{F}$	0.63	0.64 (0.58–0.68)	0.75	0.79 (0.66–0.85)
$\mathbf{D}_{G}$	0.66	0.67 (0.62–0.77)	0.70	0.71 (0.62–0.84)

AUC = 0.91 (training) vs 0.88 (testing)).

There are many SDAs proposed in the literature, although a fair comparison be-

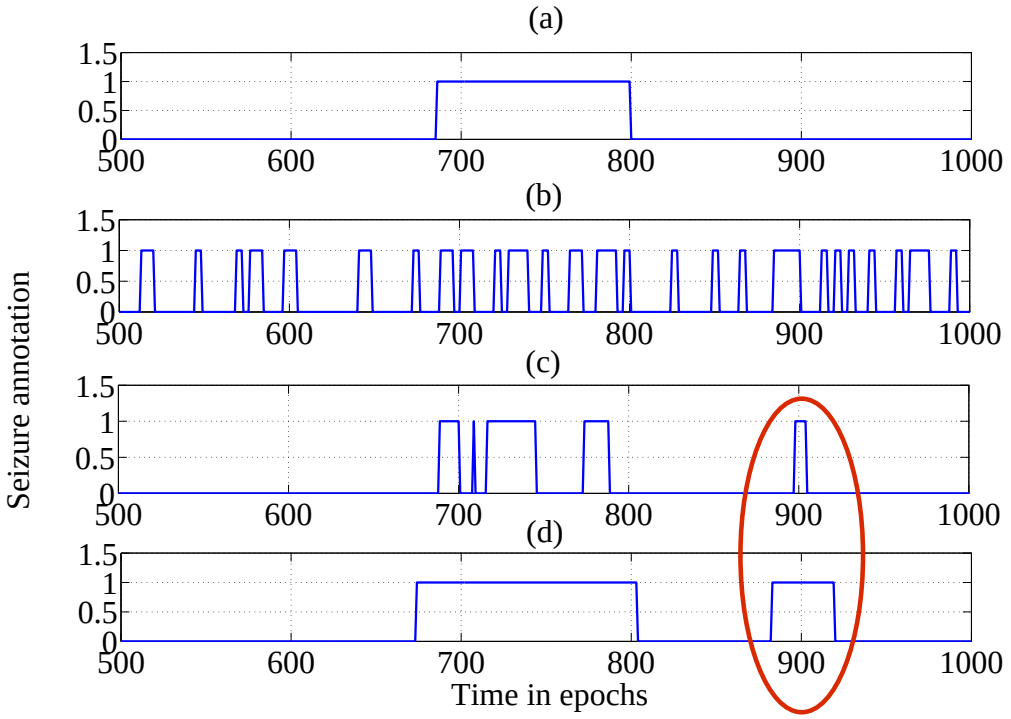


Figure 4.14: Effects of the post-processing scheme. (a) Ground truth or human annotation where 1 indicates seizure, (b) The binary decisions obtained from the SDA, (c) The smoothed binary decisions after a 32s median filter is applied, and (d) The final binary decisions after the adaptive collar operation, which increases the duration of all positive decisions. An example of a false detection event is also shown (highlighted by the red ellipse).

Table 4.5: Testing results (median) using the proposed SDA at several FD/h.

FD/h	SDR (IQR) %	Sensitivity (IQR) %	Specificity (IQR) %
0.1	56.5 (50.5–65.2)	53.5 (48.4–65.6)	97.5 (93.6–98.8)
0.5	75.9 (65.4–85.5)	63.9 (58.6–75.8)	93.4 (86.3–95.4)
1.0	82.3 (78.8–93.2)	82.5 (76.2–90.3)	86.5 (75.9–91.7)

tween methods is difficult (Temko et al., 2011a), as there are (a) differences within datasets used and (b) methodological differences particularly when validating a SDA. The differences in datasets mean that only some datasets reflect the reality of neonatal monitoring for seizures and so permit the robustness of a SDA to be tested. There are also large differences in the prevalence index (typically, EEG consists of 90% more non-seizure than seizure) of the datasets which means that traditional methods of assessing agreement such as Cohen's Kappa statistic are highly variable and, therefore, not sufficient. This results in a large array of different metrics for the analysis of SDA performance (Temko et al., 2011a). Methodological differences include cross-pollination between training and testing data resulting in optimistic assessments of SDA performance.

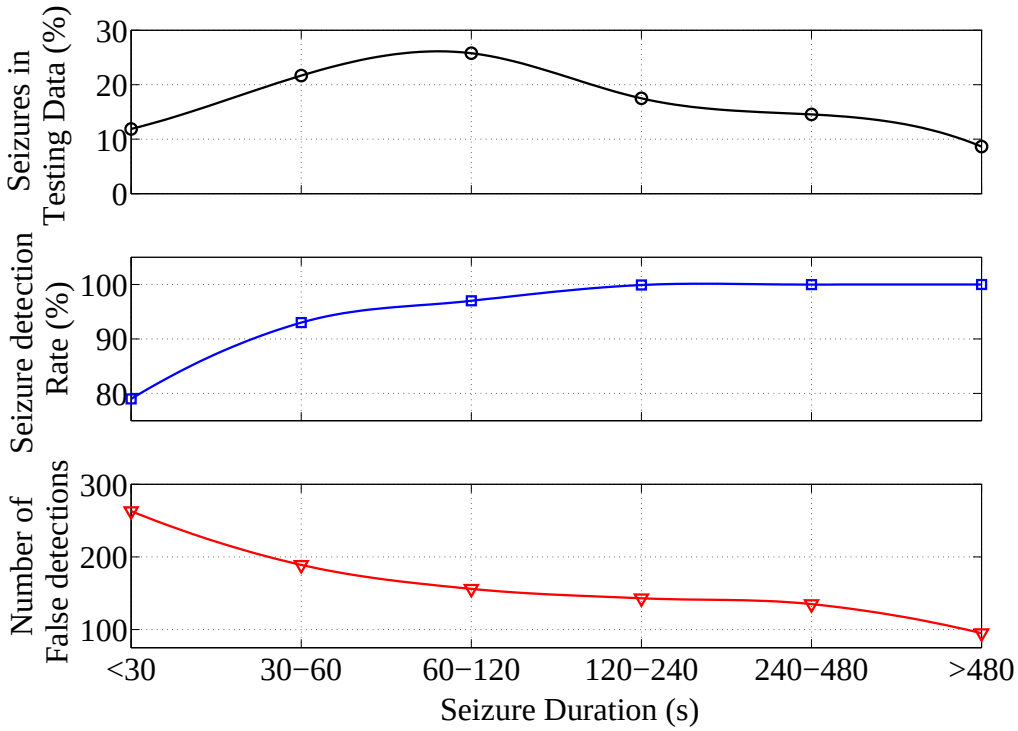


Figure 4.15: Illustration of the seizure detection rate (%) and corresponding number of false detections of seizures of different duration by the proposed SDA at 1 FD/h using testing data. Most false detections occur when the duration of seizure is less than 30s. As the duration of seizure increases, the seizure detection rate increases and the number of false detections decreases.

The proposed seizure detection algorithm is conceptually similar to Navakatikyan et al. (Navakatikyan et al., 2006) and Deburchraeve et al. (Deburchgraeve et al., 2008), but rather than attempting to estimate periodicity by looking at the correlation between the outputs of an adaptive segmentation procedure, the correlation is performed between the EEG signal and a dictionary of possible nonstationary atoms. The performance of the proposed SDA (see Fig. 4.13) is comparable to these methods (mean SDR = 0.82 vs 0.89, 0.88) (Deburchgraeve et al., 2008, Navakatikyan et al., 2006) at 1 FD/h. It should be noted that Navakatikyan et al. and Deburchraeve et al. trained and tested on the same data. The performance of the proposed SDA was similar to that of the multi-stage neonatal SDA proposed by Mitra et al. (GDR = 80% at 0.74 FD/h) in which 121h of EEG recordings with artefacts were used to test the SDA (Mitra et al., 2009). However in this study, the results were obtained on the larger dataset, which comprises 826 h of unedited neonatal EEG. Proper comparison is difficult since only performance points rather than performance curves are provided from the literature. The performance of the proposed SDA exceeded the single feature

method of Stevenson et al. (Stevenson et al., 2012), but was less than the SDA proposed by Temko et al. (Temko et al., 2011a). This comparison offers a better reflection of true results as the datasets used in these methods are the same. The superior performance of the method of Temko et al. (a method based on 55 features) is not uniform across the cohort but is significant on a subset of 3 neonates on which the proposed SDA performs poorly (mean AUC = 0.69 vs 0.92). In these 3 neonates, the RSC does not provide sufficient discrimination between seizure and non-seizure states, primarily due to a lack of response to seizure epochs (see Figs. 4.16 and 4.17). The use of additional features as used by Temko et al. (Temko et al., 2011a) may better represent seizure in this subset of poorly performing neonates and reduce the number of false alarms. Table 4.6 gives the analysis of the performance of additional features to discriminate detected seizures from false alarms. From this analysis, AR fit, bandwidth and spectral entropy measures provide the best additional discriminative power for reducing the false detections.

Table 4.6: Analysis of some additional features on true (*TD*) and false (*FD*) seizure detections by the SDA at 1FD/h. The median was used to summarise the feature value across each neonate. A Mann Whitney U Test was used to test the difference between TDs and FDs across the cohort of neonates (N=18). For details on the features see (Greene et al., 2008).

Feature	Detected seizures	False alarms	p-value
	median (IQR)	median (IQR)	
RMS amplitude(μV)	15.3 (7.78–24.25)	16.3 (13.72–20.98)	p =0.889
Maxima and minima	50 (26–56)	38 (23–46)	p =0.289
Number of extrema	4.74 (3.44–7.87)	5.79 (4.25–7.16)	p =0.962
AR Fit	0.64 (0.34–1.11)	1.19 (1.10–1.39)	p =0.002
Bandwidth	2.95 (2.15–4.23)	3.67 (3.20–4.19)	p =0.091
Spectral Entropy	0.42 (0.34–0.55)	0.63 (0.52–0.68)	p =0.001

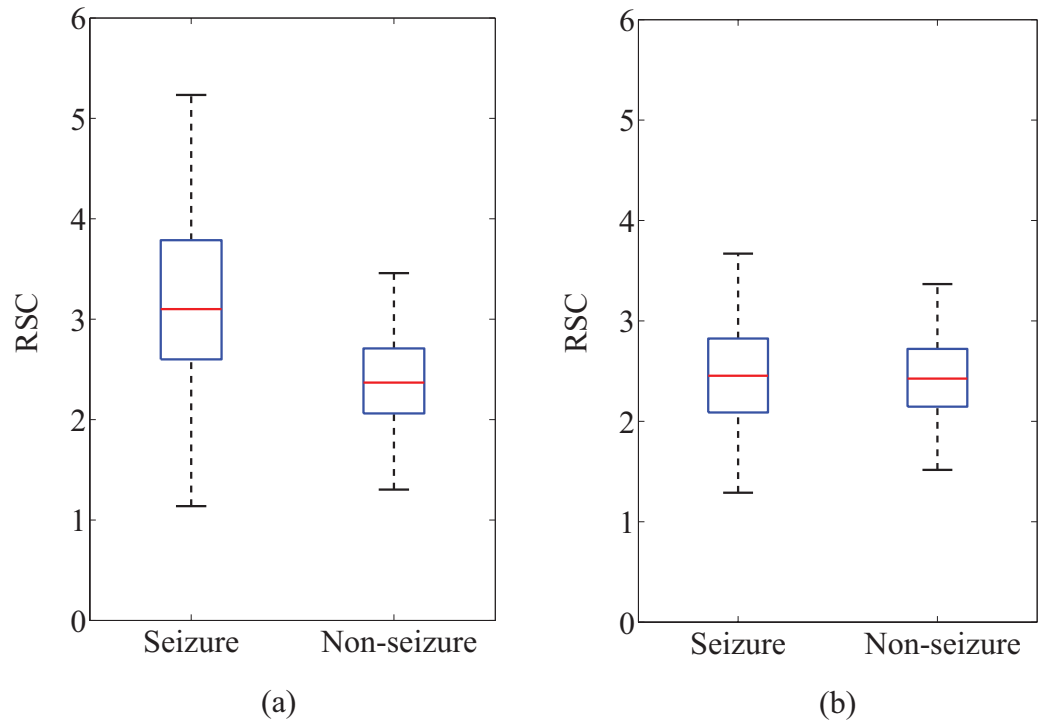


Figure 4.16: Boxplot showing the distribution of RSC (mean values) obtained for all seizure and non-seizure epochs for (a) 15/18 high performing, and (b) 3/18 low performing neonates using testing data.

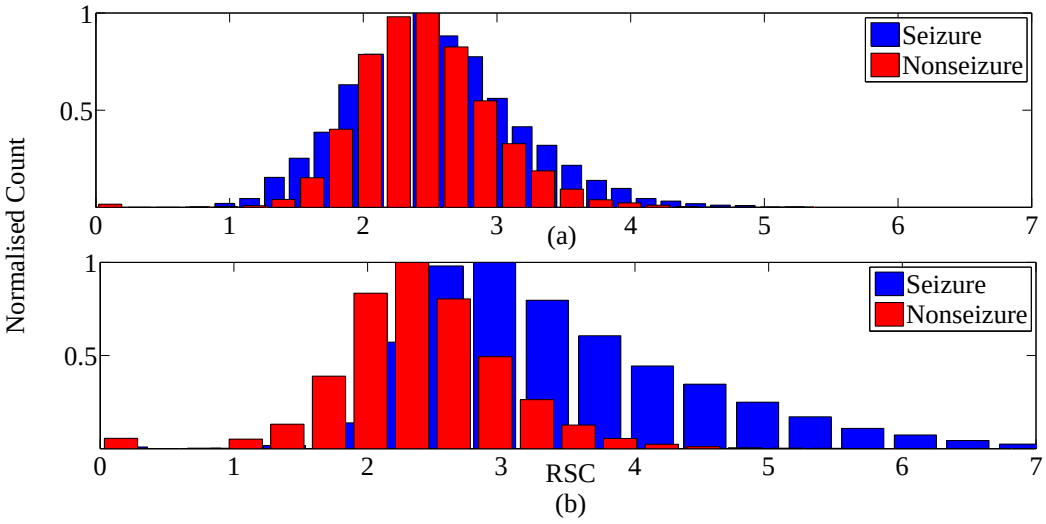


Figure 4.17: Histogram of RSC obtained for all seizure and non-seizure epochs for each testing data showing the distribution of the performance of the proposed SDA in (a) 3/18 low performing neonates, and (b) 15/18 high performing neonates.

4.8 Summary

In this chapter, a method for the detection of neonatal seizures from EEG is presented. The method was based on a single feature (RSC) of the EEG signal

which was defined as the signal to error ratio obtained using a set number of atoms for decomposition. It was observed that the number of atoms used to decompose the signal and the size of the dictionary can significantly improve the detection performance of the RSC feature. The major factor in RSC performance was, however, the type of dictionary used for AD. It was shown that a novel pseudo-periodic dictionary based on Duffing atoms results in an RSC with superior seizure detection performance. Atoms in the pseudo-periodic Duffing dictionary have a high correlation with seizure waveforms and a low correlation with non-seizure waveforms. As a result, the RSC will increase in the presence of seizure. As seizure tends towards a higher energy than non-seizure, further increases in the RSC are observed, due to the improved signal to error ratio.

A short-coming of the proposed SDA is its inability to detect short-duration seizures (the seizure detection rate drops below 80% for seizures less than 30s in duration). A significant proportion of false detections (12%) are also less than 30s in duration. A limit on seizure duration will improve the performance and may also permit the lengthening of the epoch length from 8 seconds resulting in more data to be processed for a decision. The simplest solution to the short seizure duration problem would be to redefine the length of a seizure to 30s (the minimum duration of 10s that is currently used is relatively arbitrary) as the majority of seizures in the database were longer than 93s. A change in minimum seizure duration would improve the median AUC to 0.94 (IQR 0.89–0.97) and the mean AUC to 0.92.

References

Aarabi, A., Grebe, R., Wallois, F., 2007. A multistage knowledge-based system for eeg seizure detection in newborn infants. Clinical Neurophysiology 118 (12), 2781–2797.

Celka, P., Colditz, P., 2002a. A computer-aided detection of eeg seizures in infants: a singular-spectrum approach and performance comparison. Biomedical Engineering, IEEE Transactions on 49 (5), 455–462.

Celka, P., Colditz, P., 2002b. Nonlinear nonstationary wiener model of infant EEG seizures. Biomedical Engineering, IEEE Transactions on 49 (6), 556–564.

Chen, S., Donoho, D., Saunders, M., 2001. Atomic decomposition by basis pursuit. Society for Industrial and Applied Mathematics: Review 43 (1), 129–159.

Deburchgraeve, W., Cherian, P., De Vos, M., Swarte, R., Blok, J., Visser, G. H., Govaert, P., Van Huffel, S., 2008. Automated neonatal seizure detection mimicking a human observer reading eeg. Clinical Neurophysiology 119 (11), 2447–2454.

Durka, P. J., Ircha, D., Blinowska, K. J., 2001. Stochastic time-frequency dictionaries for matching pursuit. Signal Processing, IEEE Transactions on 49 (3), 507–510.

Goodwin, M. M., Vetterli, M., 1999. Matching pursuit and atomic signal models based on recursive filter banks. Signal Processing, IEEE Transactions on 47 (7), 1890–1902.

Gotman, J., Flanagan, D., Zhang, J., Rosenblatt, B., 1997. Automatic seizure detection in the newborn: methods and initial evaluation. Electroencephalography and Clinical Neurophysiology 103 (3), 356–362.

Greene, B., Faul, S., Marnane, W., Lightbody, G., Korotchikova, I., Boylan, G., 2008. A comparison of quantitative eeg features for neonatal seizure detection. Clinical Neurophysiology 119 (6), 1248–1261.

Jouny, C. C., Franaszczuk, P. J., Bergey, G. K., 2003. Characterization of epileptic seizure dynamics using Gabor atom density. Clinical Neurophysiology 114 (3), 426 – 437.

Lawrence, R., Inder, T., 2010. Neonatal status epilepticus. In: Seminars in pediatric neurology. Vol. 17. Elsevier, pp. 163–168.

Liu, A., Hahn, J., Heldt, G., Coen, R., 1992. Detection of neonatal seizures through computerized eeg analysis. Electroencephalography and Clinical Neurophysiology 82 (1), 30–37.

Mallat, S. G., Zhang, Z., 1993. Matching pursuits with time-frequency dictionaries. IEEE Trans. on Signal Proc. 41 (12), 3397–3415.

Mitra, J., Glover, J. R., Ktonas, P. Y., Kumar, A. T., Mukherjee, A., Karayiannis, N. B., Frost Jr, J. D., Hrachovy, R. A., Mizrahi, E. M., 2009. A multi-stage system for the automated detection of epileptic seizures in neonatal eeg. Journal of Clinical Neurophysiology: official publication of the American Electroencephalographic Society 26 (4), 218.

Navakatikyan, M. A., Colditz, P. B., Burke, C. J., Inder, T. E., Richmond, J., Williams, C. E., 2006. Seizure detection algorithm for neonates based on wave-sequence analysis. Clinical Neurophysiology 117 (6), 1190–1203.

Pellock, J. M., Bourgeois, B. F., Dodson, E., Nordli, D. R., Sankar, R., Bourgeois, B., Dodson, E., Nordli, D., 2007. Pediatric epilepsy: diagnosis and therapy. 3rded. New York:Demos Medical Publishing.

Penrose, R., 1955. A generalized inverse for matrices. In: Mathematical proceedings of the Cambridge philosophical society. Vol. 51. Cambridge Univ Press, pp. 406–413.

Rankine, L., Mesbah, M., Boashash, B., 2007. A matching pursuit-based signal complexity measure for the analysis of newborn eeg. Medical & Biological Engineering & Computing 45 (3), 251–260.

Roessgen, M., Zoubir, A. M., Boashash, B., 1998. Seizure detection of newborn eeg using a model-based approach. Biomedical Engineering, IEEE Transactions on 45 (6), 673–685.

Rubinstein, R., Bruckstein, A. M., Elad, M., 2010. Dictionaries for sparse representation modeling. Proceedings of the IEEE 98 (6), 1045–1057.

Scher, M. S., Sun, M., Steppe, D. A., Guthrie, R. D., Sclabassi, R. J., 1994. Comparisons of eeg spectral and correlation measures between healthy term and preterm infants. Pediatric neurology 10 (2), 104–108.

Smit, L., Vermeulen, R., Fetter, W., Strijers, R., Stam, C., 2004. Neonatal seizure monitoring using non-linear eeg analysis. Neuropediatrics 35 (06), 329–335.

Stevenson, N., Korotchikova, I., Temko, A., Lightbody, G., Marnane, W., Boylan, G., 2013. An automated system for grading eeg abnormality in term neonates with hypoxic-ischaemic encephalopathy. Annals of Biomedical Engineering 41 (4), 775–785.

Stevenson, N. J., Mesbah, M., Boylan, G. B., Colditz, P. B., Boashash, B., 2010. A nonlinear model of newborn eeg with nonstationary inputs. Annals of Biomedical Engineering 38 (9), 3010–3021.

Stevenson, N. J., O'Toole, J. M., Rankine, L. J., Boylan, G. B., Boashash, B., 2012. A nonparametric feature for neonatal eeg seizure detection based on a representation of pseudo-periodicity. Medical Engineering and physics 34 (4), 437–446.

Temko, A., Stevenson, N., Marnane, W., Boylan, G., Lightbody, G., 2012. Inclusion of temporal priors for automated neonatal eeg classification. Journal of Neural Engineering 9 (4), 046002.

Temko, A., Thomas, E., Marnane, W., Lightbody, G., Boylan, G., 2011a. Eeg-based neonatal seizure detection with support vector machines. Clinical Neurophysiology 122 (3), 464–473.

Temko, A., Thomas, E., Marnane, W., Lightbody, G., Boylan, G., 2011b. Performance assessment for eeg-based neonatal seizure detectors. Clinical Neurophysiology 122 (3), 474–482.

Tropp, J., Gilbert, A., Dec 2007. Signal recovery from random measurements via orthogonal matching pursuit. Information Theory, IEEE Transactions on 53 (12), 4655–4666.

Vapnik, V., Kotz, S., 2006. Estimation of dependences based on empirical data. Springer Series in Statistics, Springer-Verlag, New York, 2nd edition.

Victor, S., Appleton, R. E., Beirne, M., Marson, A. G., Weindling, A. M., 2005.

Spectral analysis of electroencephalography in premature newborn infants: normal ranges. Pediatric research 57 (3), 336–341.

Chapter 5

Robustness of Time-Frequency Distribution Features for Neonatal Seizure Detection

> As far as the laws of mathematics refer to reality, they are not certain, as far as they are certain, they do not refer to reality. ––Albert Einstein

5.1 Introduction

In the real world, the majority of signals such as the electroencephalogram (EEG) are nonstationary in nature since their properties change with time. This may be due to external/internal events which influence the dynamics, characteristic time scale, transient processes and change in system parameters (Paluš, 1996). The changes in EEG waveforms are directly influenced by several behavioral and mental states (Gribkov and Gribkova, 2000, Kohlmorgen et al., 2000) and are clearly visualized especially during epileptic seizures (Jefferys et al., 1990). The duration of this change in EEG properties is not fixed and can vary between 10s - 1min or even longer (Niedermeyer and da Silva, 2005). In this chapter, the performance of the neonatal seizure detection algorithm (NSDA) proposed in chapter 4 was improved using additional time-frequency distribution (TFD) features obtained from the pseudo-periodic Duffing dictionary. The main advantage of using TFD is that it can provide key instantaneous frequency values which characterize the nonstationary nature of the EEG signal.

5.2 EEG seizure detection using time frequency (TF) analysis techniques

Several seizure detection algorithms have been proposed based on the time, frequency, nonlinear and nonstationary characteristics of EEG signals. The basic architecture of these methods is illustrated in figure 5.1 which involves the analysis of single channel EEG for seizure detection. Most of these methods either use short duration EEG dataset (varying between 30 mins to 2 hours) or use carefully selected artefact free segments and do not achieve a level of performance similar to the inter-observer agreement seen between human experts (Wilson et al., 2003).

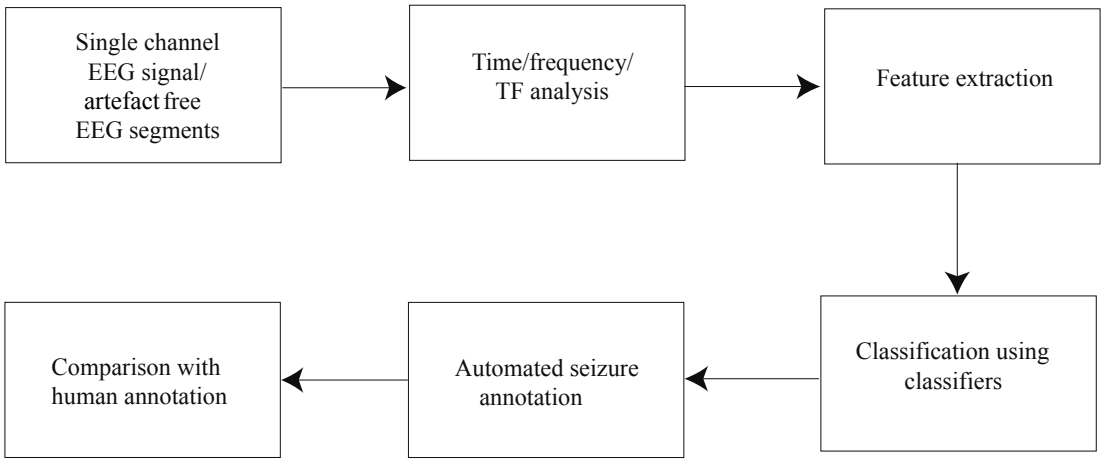


Figure 5.1: Illustration of basic methodology for automated seizure detection using Time-Frequency analysis.

Some of these methods use basic time domain features (Gotman, 1990), spectral features (Widman et al., 2000) and autoregressive modelling approach (Subasi et al., 2006). Other methods use some advanced features which include a combination of time and frequency features (Srinivasan et al., 2005), discrete wavelet transform (DWT) of EEG signals (Subasi, 2007), energy measures from the TFD of the EEG signals (Tzallas et al., 2007) and time frequency (TF) matched filtering methods (Boashash and Mesbah, 2001). A combination of time, frequency and time-scale domain features are used for neonatal seizure detection in (Greene et al., 2008, Temko et al., 2011).

The non-stationarity nature of the neonatal EEG has led to the application of non stationary signal processing and segmented analysis for the seizure detection problem. It was shown that TF and instantaneous frequency (IF) based methods

were more suitable for automated seizure detection and classification (Boashash and Mesbah, 2001). A seizure detection method using nonstationary features obtained from TFD of the EEG signals was proposed in (Boashash et al., 2011). TFDs inherently deal with nonstationarity by projecting a signal onto a joint TF domain. An interesting approach was outlined by Boashash et al. in which features obtained from the TFD were used to detect neonatal seizures (Boashash et al., 2012b). This method was tested on a small dataset of artefact free neonatal EEG (50 segments of seizure and non-seizure with individual segments of length 12.8s) with no apparent validation of results.

A significant problem with the development of a neonatal seizure detection algorithm (NSDA) is the lack of commonly available datasets to allow for the comparison of proposed methods. This problem, in conjunction with a lack of data on inter-observer variability between human experts, hinders the progress of NSDA development. In this chapter, the NSDA proposed by Boashash et al. (Boashash et al., 2012b) (TFD feature-set classified by a support vector machine (SVM)) was applied to a large dataset of neonatal EEG that better represents the application of a NSDA in the NICU (a NSDA as a surrogate neurophysiologist for long duration assessment of the neonatal EEG). The usefulness of such a process is twofold in that it provides information on the robustness of this method when applied to a real world setting and permits the comparison between this method and other methods developed on the same datasets (Stevenson et al., 2012a, Temko et al., 2011, Thomas et al., 2010).

5.3 Support vector machine

The support vector machine (SVM) was introduced by Vapnik et al. in 1995 (Vapnik and Cortes, 1995). The main idea behind classification is to identify certain patterns in a data set and classify accordingly into two classes. The classification process requires the SVM to be trained on a specific data set which is divided into *training* and *testing* sets. The patterns in the training set correspond to a particular class which is labeled with a certain *target class*. The SVM model parameters are tuned to provide the best classification accuracy over the training set and then the model is tested on the unseen testing set (see figure 5.2). Since this thesis is mainly focussed on developing advanced features for neonatal EEG classification, it is out of scope of this thesis to explain the extensive theory of SVMs. Only a basic understanding of classification using SVMs is therefore

provided. For a more detailed description, readers are referred to (Cristianini and Shawe-Taylor, 2000, Vapnik and Cortes, 1995).

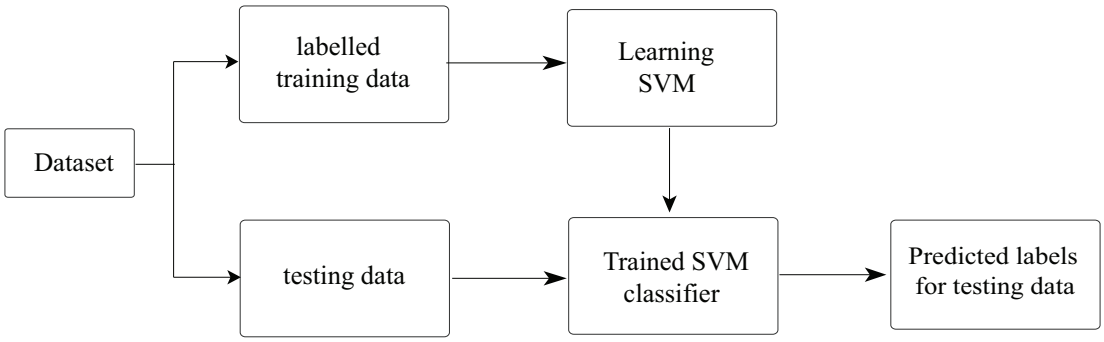


Figure 5.2: Basic SVM learning process involved in classification. With the help of a particular training set of data with labels, the learning algorithm generates a decision rule which can then be used to predict the labels of new unlabelled/testing data.

Consider a classification problem with a training set $\mathbf{X}$ consisting of M pattern-label instances with patterns belonging to either positive (w_1) or negative class (w_{-1}) are defined as $y \in \{1, -1\}$. A vector $\mathbf{x}_j \in \mathbb{R}^n$, where $1 \leq j \leq M$ represents a vector of n features which are used to characterise the data. The training set $\mathbf{X}$ can now be denoted as:

$$\mathbf{X} = \{(\mathbf{x}_j, y_j)\}_{j=1}^{M}, \quad \mathbf{x} \in \mathbb{R}^n, \; y_j \in \{1, -1\} \tag{5.1}$$

The goal is to find a *decision surface*, also known as *hyperplane* that can correctly separate the patterns in the training set into their corresponding classes which can be written in the linear form as

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b \tag{5.2}$$

where $\mathbf{w} \in \mathbb{R}^n$ and b is the bias. Depending on the values of $\mathbf{w}$ and b, there are many hyperplanes that can separate the training set (see figure 5.3), however it is important to find the best hyperplane that achieves maximum separation.

Let $\mathbf{H}_0$, $\mathbf{H}_1$, $\mathbf{H}_{-1}$ be the hyperplanes corresponding to $f(\mathbf{x}) = 0, f(\mathbf{x}) = 1$ and $f(\mathbf{x}) = -1$ respectively. The optimal hyperplane which is also denoted as the *maximum-margin hyperplane* is obtained by *maximizing* the distance between hyperplanes, also known as *maximum margin* (see figure 5.4) given by

$$Maximum\ margin = \frac{2}{||\mathbf{w}||}. \tag{5.3}$$

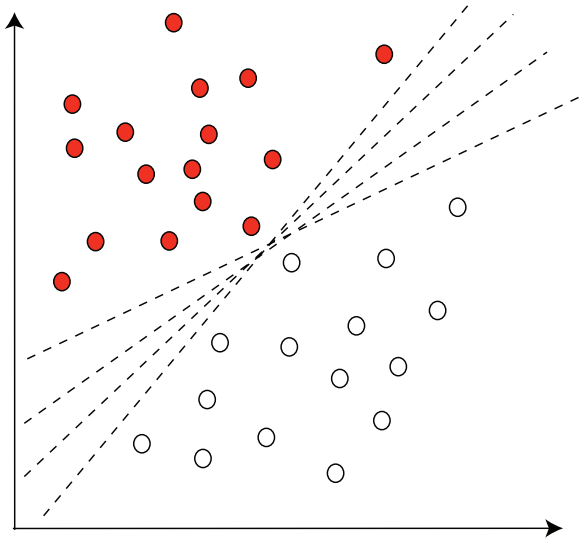


Figure 5.3: Illustration of several hyperplanes that can classify the data. The real problem is to find the optimal hyperplane that best separates the data.

The following constraints must be satisfied in order to ensure that data points are on/outside *maximum margin*:

$$\mathbf{w}^T\mathbf{x}_j + b \geq +1 \;\; if, \;\; y_j = +1, \;\; \text{and} \tag{5.4a}$$

$$\mathbf{w}^T\mathbf{x}_j + b \leq -1 \;\; if, \;\; y_j = -1 \;\; \forall j \tag{5.4b}$$

which can be written in simple form as:

$$y_j(\mathbf{w}^T\mathbf{x}_j + b) \geq +1 \;\; \forall \, j. \tag{5.5}$$

Maximizing the margin is equivalent to minimizing $\mathbf{w}$ and can be written in the form of optimization problem as:

$$\text{minimise } \frac{1}{2}||\mathbf{w}||^2, \;\; \text{subject to } \; y_j(\mathbf{w}^T\mathbf{x}_j + b) \geq +1 \; \forall \, j. \tag{5.6}$$

This formulation using constraints ensures that the maximum margin classifier can classify the data correctly, given that the data is linearly separable. However, in practice, data is often not linearly separable (which is generally the case with biomedical signals) and the optimization problem would diverge and grow arbitrarily. A set of slack variables ϵ are then introduced in equation 5.4(a,b) to

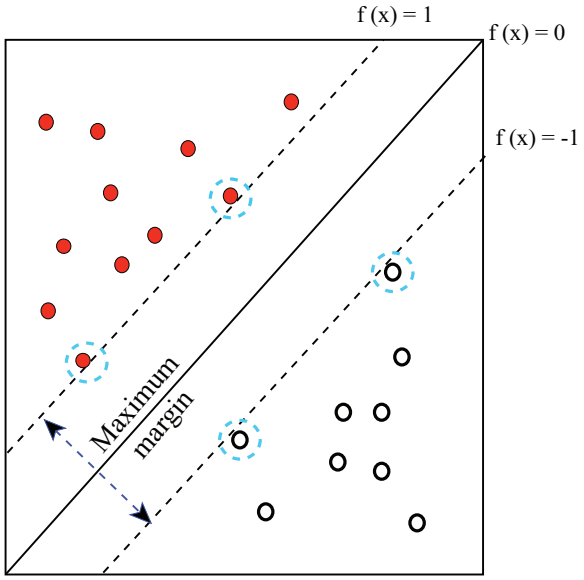


Figure 5.4: Example of a maximum-margin hyperplane and the margins for an SVM trained on two classes. The support vectors are indicated with a blue dotted circle.

ensure convergence such that

$$\mathbf{w}^T\mathbf{x}_j + b \geq +1 - \epsilon_j \;\; if, \;\; y_j = +1, \;\; \text{and} \tag{5.7a}$$

$$\mathbf{w}^T\mathbf{x}_j + b \leq -1 + \epsilon_j \;\; if, \;\; y_j = -1 \tag{5.7b}$$

where $\epsilon_j \geq 0$ allows a feature vector $\mathbf{x}_j$ given data to lie inside the margin and possible be misclassified if $\epsilon_j \geq 1$. Since data can be misclassified if the value of its slack variable is greater than 1, a regularisation constant C is introduced into the objective function in equation 5.6 which can be written as:

$$\text{minimise } \frac{1}{2}||\mathbf{w}||^2 + C\sum_{j}^{M}\epsilon_j, \;\; \text{subject to } \; y_j(\mathbf{w}^T\mathbf{x}_j + b) \geq +1 \; \forall \; j. \tag{5.8}$$

The constant $C > 0$ sets the relative importance of maximizing the margin and minimizing the amount of slack and was introduced by Cortes and Vapnik (Vapnik and Cortes, 1995). The convex optimization problem in equation 5.8 is solved by the introduction of Lagrange multipliers, $\alpha_j \geq 0, \; \forall \; j$ (Boyd and Vandenberghe, 2009). The training samples $\mathbf{x}_j$ for which $\alpha_j > 0$ are on or within the margin are called *support vectors* (see figure 5.4). Given a test vector $\tilde{\mathbf{x}} \in \mathbb{R}^n$, the classification process now simplifies to assigning one of the two classes defined in

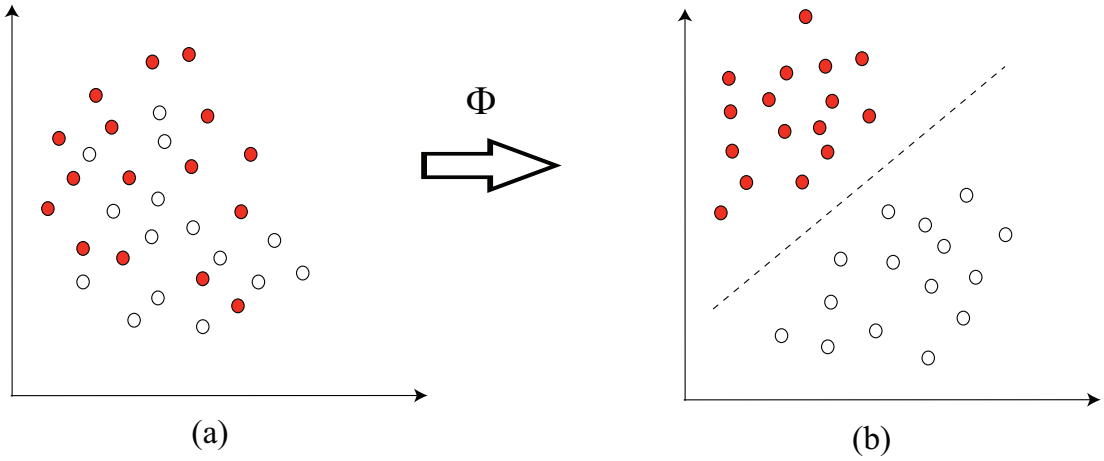


Figure 5.5: Simple illustration of mapping linearly inseperable data (a) to a seperable data (b) using mapping.

y using a SVM such that,

$$f(\mathbf{x}) = \sum_{j \in N_{sv}} \alpha_j \mathbf{y}_j \tilde{\mathbf{x}}_j^T \mathbf{x}_j + b. \tag{5.9}$$

Here N_{sv} is the set of indices for the support vectors extracted from the training data. A hyperplane can be efficiently used to separate the data, if the data is linear. However, in most of the cases, the data is far from linear and is inseparable. In such cases, the non-separable data is mapped into a high-dimensional linearly separable feature space using kernels (illustrated in figure 5.5) such that:

$$f(\mathbf{x}) = \sum_{j \in N_{sv}} \alpha_j \mathbf{y}_j \mathcal{K}\left(\mathbf{x}_j^T, \tilde{\mathbf{x}}_j\right) + b. \tag{5.10}$$

where $\mathcal{K}\left(\mathbf{x}_i^T, \tilde{\mathbf{x}}_i\right)$ is the kernel function.

There are several kernel functions available, however the Gaussian radial basis function (rbf) is used in this thesis and is defined as,

$$\mathcal{K}(\mathbf{x}_j, \tilde{\mathbf{x}}_j) = e^{-\frac{(\mathbf{x}_j^T - \tilde{\mathbf{x}}_j)^T (\mathbf{x}_j^T - \tilde{\mathbf{x}}_j)}{2\sigma^2}} \tag{5.11}$$

where σ is the parameter that controls the width of Gaussian kernel. It has been demonstrated that Gaussian kernel SVM classifiers are effective in neonatal seizure detection (Temko et al., 2011, Thomas et al., 2010). The kernel $\mathcal{K}(\mathbf{x}_j, \tilde{\mathbf{x}}_j)$ can be interpreted as the non-linear similarity measure between two data points which can be used to replace the mapping function using a property known as the *kernel trick* (Scholkopf, 2001). With the inclusion of a kernel function, the

SVM classifier can now be written as (Maji et al., 2008):

$$y(\tilde{\mathbf{x}}_j) = sign(f(\tilde{\mathbf{x}}_j)) = sign\left(\sum_{j \in N_{sv}} \alpha_j y_j \mathcal{K}(\mathbf{x}_j, \tilde{\mathbf{x}}_j) + b\right). \tag{5.12}$$

It should be noted that the decision function in equation 5.12 will provide hard decisions i.e the output of the SVM varies from $[-\infty$ to $\infty]$. The output of the SVM can be converted to a probabilistic measure of seizures, bounded within [0,1] via Platt scaling (Platt, 1999). The output of SVM can be converted to a probabilistic measure of data belonging to the positive class w_1, using a sigmoid function:

$$P(w_1|\mathbf{x}) = \frac{1}{1 + exp(Af(\mathbf{x}) + B)}, \tag{5.13}$$

where $P(w_1|\mathbf{x})$ is the probability of the presence of a seizure in an epoch, $f = f(\mathbf{x})$ is the output of the SVM classifier (distance to the separating hyperplane), A and B are the parameters of the sigmoid function which are estimated over the training dataset (Platt, 1999).

5.4 TF features from atomic decomposition

Nonstationary signals with time-varying frequency content can be best represented by a TFD which represents the energy distribution of the signal over a two-dimensional (2D) TF space. The TFD provides the variation of frequency of the signal components with time, also defined as *instantaneous frequency* (IF) (Boashash, 1992). By using a peak detector, the IF can be obtained from the TFD by selecting the frequency with the maximum value in the TFD as a function of time. Therefore, it is appropriate to use TFD for detection and classification of EEG abnormalities based on their nonstationarity nature (Boashash, 1991).

TFD based on the Wigner-Ville distribution (WVD) can be used to analyze and classify non-stationary signals. However, due to the bilinear nature of the signal kernel, the WVD is contaminated by cross-terms (described in section 3.4 of chapter 3). By using an appropriate smoothing window (a 2D low pass filter), these cross-terms can be minimized. It was demonstrated that the quadratic time-frequency distributions (QTFD) were efficient for analysis of non-stationary signals which can be written for a given input signal x as (Boashash, 1991)

$$\rho_x(t,f) = \rho_x^{WVD}(t,f) \underset{t,f}{**} G(t,f) \tag{5.14}$$

where $\rho_x(t,f)$ is the QTFD, $\rho_x^{WVD}(t,f)$ is the Wigner Ville distribution (WVD), $G(t,f)$ is the TF kernel or a 2D smoothing window. Depending on the selection of kernels or smoothing window, several QTFD can be obtained which can be adapted to a particular class of signals (Boashash, 1991) some of which are described in table 3.1 of chapter 3.

The QTFD of a given real discrete time signal $x[n]$, $n = 0, 1, \ldots, N-1$, is defined in terms of its analytic associate $z[n]$ (Boashash et al., 2011) as:

$$\rho[n,p] = 2\underset{n\to p}{\mathrm{DFT}}\{G[n,m] \underset{n}{*} z[n+m]z^*[n-m]\} \tag{5.15}$$

where $G[n,m]$ is the kernel of the TFD and $\rho[n,p]$ is an $N \times N$ matrix. The modified B distribution (MBD) and smoothed WVD (SWVD) was used in this work as it has been shown that these TFDs provide superior performance for neonatal and adult seizure detection (Boashash et al., 2012b), respectively. The discrete time kernel functions for MBD and SWVD are listed in Table 5.1.

Table 5.1: QTFD and their time-lag kernels. Here β is positive and real, $w[n]$ is the Hanning window function.

QTFD	time-lag kernel, $G[n,m]$
MBD	$\frac{\cosh^{-2\beta}}{\sum_n \cosh^{-2\beta} n}$
SWVD	$\delta[n]w[n]$

In the previous chapter, a method for the detection of neonatal seizures from EEG was presented. The method was based on a single feature denoted as *relative structural complexity* (RSC) of the EEG signal which was defined as the signal to error ratio obtained using a set number of atoms for decomposition. It was observed that the number of atoms used to decompose the signal and the size of the dictionary can significantly improve the detection performance of the RSC feature. The major factor in RSC performance was, however, the type of dictionary used for AD. It was shown that a novel pseudo-periodic Duffing dictionary ($\mathbf{D}_{\tilde{D}}$) based on Duffing atoms results in an RSC with superior seizure detection performance. Atoms in the pseudo-periodic Duffing dictionary have a high correlation with seizure waveforms and a low correlation with non-seizure waveforms. As a result, the RSC will increase in the presence of seizure. As seizure tends towards a higher energy than non-seizure, further increases in the RSC are observed, due to the improved signal to error ratio. The ($\mathbf{D}_{\tilde{D}}$) of size $2N$ (two times over complete dictionaty and N is the length of the signal) and a decomposition level of 5 atoms was found optimal for seizure detection. This

suggested that the atoms selected in the ($\mathbf{D}_{\tilde{D}}$) dictionary during AD can provide valuable information about the presence of seizures.

In this study, the TFD of the atoms from the ($\mathbf{D}_{\tilde{D}}$) dictionary selected during AD was obtained as

$$\rho_{AD}[n,p] = \sum_{k=1}^{K} TFD\left\{\boldsymbol{\gamma}_k \boldsymbol{\phi}_k\right\} \tag{5.16}$$

where K is the number of iterations/atoms used in AD, $\boldsymbol{\gamma}_k$ is the weight associated with the corresponding atom $\boldsymbol{\phi}_k$.

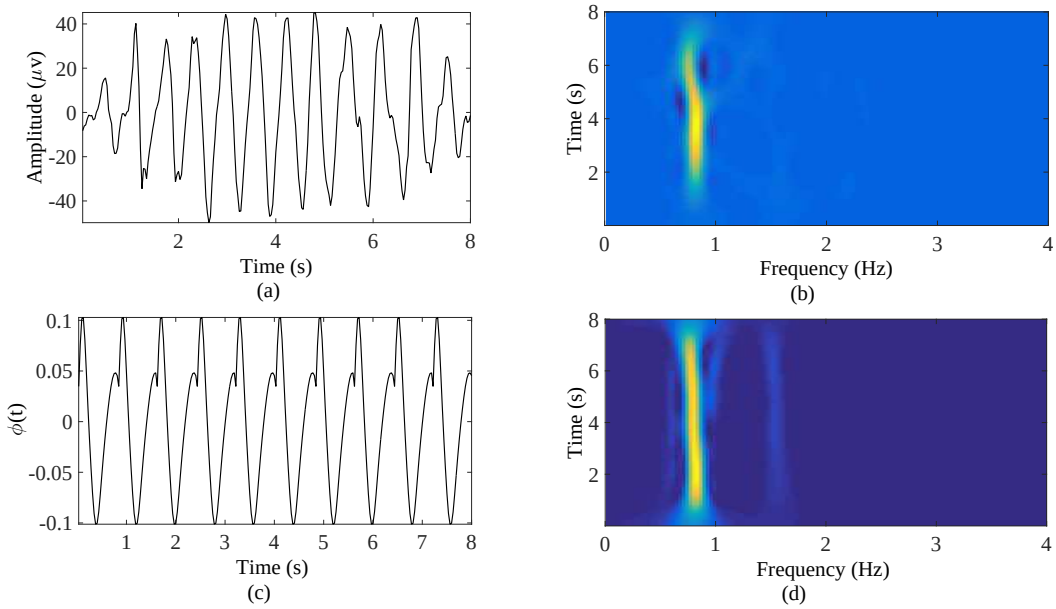


Figure 5.6: Example of (a) 8s seizure epoch (b) and its corresponding TFD using MBD ($\beta = 0.01$), (c) a pseudo periodic Duffing oscillator atom and (d) its corresponding TFD using MBD ($\beta = 0.01$).

The proposed feature extraction using AD is primarily based on the assumption that the most significant information of the EEG signal can be provided by the synthesizing atoms with the maximum energy during AD. Since OMP selects atoms in a sequence by removing the largest residual energy and also does not select the same atom twice, it tends to provide most useful atoms in just fewer iterations. Since atoms in the ($\mathbf{D}_{\tilde{D}}$) dictionary are designed to be highly coherent with seizure, by extracting the information from the atoms selected during the OMP the presence of seizure can be detected efficiently. Moreover, the TFD obtained from AD provides a noise-free representation and increases the time-frequency resolution (demonstrated in chapter 3 section 3.7). Figure 5.6 shows a sample 8s seizure epoch, pseudo periodic Duffing oscillator atom and their corresponding TFDs.

8 features were derived from the TFD of the atoms selected from the dictionary during AD. These features were initially used by Boashash et al. (derived from the TFD of the EEG signal) for multichannel-based newborn seizure detection (Boashash et al., 2012b). However in this study, the features extracted from the TFD of atoms selected during the AD from the dictionary were used; this should reduce the effect of artefacts on the performance of NSDA. Here the atoms are represented as $\phi[n]$, sampling frequency as f_s and TFD as $\rho_{AD}[n,p]$. The following 8 features were used in this study.

1. *Instantaneous frequency (IF) features*: The instantaneous frequency (IF) of a nonstationary signal provides information about its change in frequency content with time (Boashash, 1992). For a given signal, the peaks of the constant-time cross-sections of the TFD provides the IF of the signal (Boashash et al., 2012b). The IF can be obtained from the TFD of the selected atoms as

$$f_i[n] = \frac{f_s}{2N} \arg\left\{\max_p \rho_{AD}[n,p]\right\}. \tag{5.17}$$

The mean

$$F_1 = \frac{1}{N}\sum_{n=1}^{N} f_i[n] \tag{5.18}$$

and deviation

$$F_2 = \Delta f_i = \max_n(f_i[n]) - \min_n(f_i[n]) \tag{5.19}$$

are selected as features 1 and 2 respectively.

2. *Singular value decomposition (SVD) based features*: The SVD of the TFD matrix, ρ_{AD} can be obtained as

$$\rho_{AD} = \mathbf{USV}^H. \tag{5.20}$$

The diagonal entries of $\mathbf{S}$ represent the singular values of ρ_{AD} and it was shown that these singular values provide useful information about EEG abnormalities (Hassanpour et al., 2004, Temko et al., 2011). The maximum (F_3) and the variance (F_4) of the diagonal entries of the singular values ($\mathbf{S}$) are taken as features 3 and 4, respectively.

3. *Time-frequency complexity measure* (TFCM): The TCFM was initially proposed in (Greene et al., 2008) and was extended to the TF plane by Boashash et al. in (Boashash et al., 2012a). This feature provides information about

both SVD and Shannon entropy measures defined as

$$F_5 = -\sum_{j=1}^{N} \bar{S}_j log \bar{S}_j \tag{5.21}$$

where $\bar{S}_j = \frac{S_j}{\sum_1^N S_j}$, for $j = 1, 2, \ldots, N$ are the normalized singular values.

4. *Energy concentration measure* (ECOME): The concentration of the important element at each grid in the TF domain can be obtained using ECOME (Sejdić et al., 2009) which is used as feature 6, given by

$$F_6 = \left(\sum\sum |\rho_{AD}[n,p]|^{\frac{1}{2}}\right)^2 \tag{5.22}$$

5. *Energies in sub-bands*: Additionally, two features which capture the sub-band energies corresponding to $\delta = 0 - 4Hz$ and $\theta = 4 - 8Hz$ respectively (Boashash et al., 2011) defined as:

$$F_7 = \sum_{n=1}^{N}\sum_{p=1}^{N_\delta} \rho_{AD}[n,p] \tag{5.23}$$

$$F_8 = \sum_{n=1}^{N}\sum_{p=N_\delta}^{2N_\delta} \rho_{AD}[n,p] \tag{5.24}$$

where $N_\delta = \lfloor 8N/f_s \rfloor$ and $\lfloor . \rfloor$ is the floor operator are used as features 7 and 8 respectively. These features provides information about the energy of the EEG signal $x[n]$ corresponding to their sub-bands.

5.5 Automated neonatal seizure detection system architecture

Figure 5.7 shows the architecture of the automatic NSDA. The initial step involves the preprocessing of signals from the EEG channels and segmenting into epochs. Features are extracted from each segmented EEG epoch. The feature vectors are then fed to the SVM classifier and the probability of seizure is obtained for each EEG epoch. The outputs of the SVM are then converted to probability-like values. The maximum of the probabilities across all channels is then computed to represent the final support of a seizure. The sharp transients in the SVM output

are then suppressed using a median filter. An adaptive collar that is related to the duration of the detected seizure is applied to the binary output to extend the detection. An automated seizure annotation is then obtained.

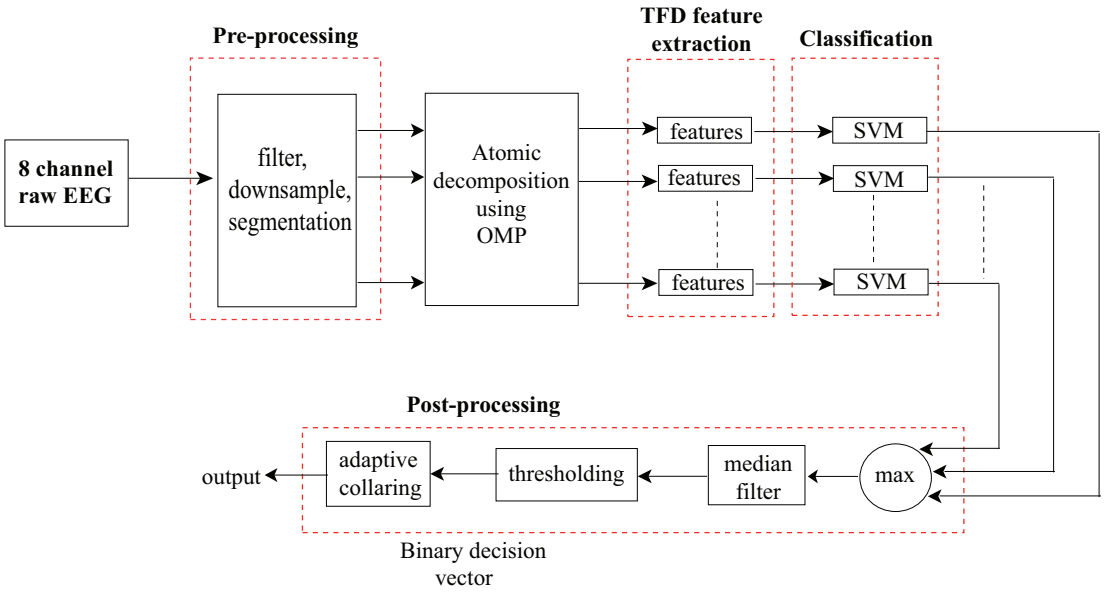


Figure 5.7: Atomic decomposition based neonatal seizure detection system.

5.5.1 Preprocessing

The 8 channel EEG signal is passed through an anti-aliasing filter set at 12.8Hz, downsampled from 256 to 32 Hz as the significant energy in the newborn EEG ($> 95\%$) does not exceed the alpha band (8-12 Hz) (Scher et al., 1994). The data is also filtered with a single pole highpass filter with a cutoff frequency of 0.5Hz. The EEG is then segmented into 8 s epochs with 50% (4 s) overlap between epochs. The minimum seizure duration recommended by the International Federation of Clinical Neurophysiology (IFCN) (Boylan et al., 2010) is 5 s for normal EEG background and 10 s in case of abnormal background EEG. Since most babies suffering from HIE or seizure have abnormal background EEG, a window length of 8 s was chosen in this work.

5.5.2 Feature extraction

Each 8s epoch was decomposed using OMP using a complex pseudo-periodic duffing dictionary ($\mathbf{D}_{\tilde{D}}$). The number of atoms was restricted to 5 as it was shown in chapter 4 that the dictionary of size $2N$ (here $N = 256$ is the length of

the signal) and decomposition level of 5 atoms provided optimal performance in seizure detection. The TFD of the individual atoms selected during this process was then obtained which was later combined to form a joint TFD of atoms. Eight features were then extracted from this joint TFD. This was repeated for all 8 channels.

The parameter of the MBD was chosen as $\beta = 0.01$ and a Hanning window of length $\lfloor N/4 \rfloor$ was chosen for SPWVD (here $N = 256$) as it was shown to provide good performance for the TFD of neonatal and adult seizure detection, respectively (Boashash et al., 2012b). It should be noted that the window function was not optimized in this work. The TF features from the pre-processed EEG signal were also obtained for the purpose of comparison.

5.5.3 SVM classification

SVM classification using a Gaussian kernel was used to classify the extracted feature vectors. More details about the working of the SVM classifier is described in section 5.3. The classification stage was divided into two parts - training and testing. The performance of the proposed NSDA was estimated using a leave-one-out (LOO) cross validation as it provided almost an unbiased estimation of the true generalization error. In this validation method 17 patients data were used in training and the left-out patients data was used for testing. This process was repeated until data from each patient was used for testing (18 different combinations of test/train sets). The mean and median value across all 18 test folds were then obtained.

To select suitable model parameters for the SVM, a nested cross-validation model selection on the training data was performed. Probability-like values were then obtained from the SVM. A subset of 5 minutes of seizure and 50 minutes of non-seizure were selected from 17 neonates at each step (a total of 85 minutes of seizure and 850 minutes of non-seizure data at each training iteration). The features extracted during training were then fed to an SVM classifier with a Gaussian rbf kernel and then tested on the full recording of the remaining neonate to generate an automated annotation of seizure.

All features were normalised using a Box-Cox transformation during training to have similar mean and standard deviation of the distribution of each feature (Box and Cox, 1964). The obtained SVM classifier was then used to test the testing data for each channel and the final decision was obtained after post-processing

described below.

5.5.4 Post-processing

The outputs of the SVM from each channel were converted to pseudo-probabilistic values using Platt's method (Temko et al., 2011). The maximum of the probabilities across all channels was computed to represent the final support of a seizure. Sharp transients in the support were suppressed using a median filter of 8s in duration. An initial decision was obtained after comparing the postprocessed value to the threshold of 0.5 (equal priority for seizure and nonseizure classes). The adaptive collar technique was then applied in which every seizure decision was extended proportional to the duration of the detected seizure on either side (illustrated in figure 5.8). An automated annotation of the seizure was then obtained.

5.5.5 Performance assessment metrics

Similar to the assessment used in chapter 4, the area under the receiver operator characteristic (ROC) curve (AUC) was used as an epoch based performance metric for the training stage as well as for each unseen test subject. In addition, the event based good detection rate (GDR) for varying number of false detections per hour (FD h^{-1}) is also reported.

5.6 Results and discussion

Table 5.2 shows time and event based metrics at selected thresholds for the TF feature-sets generated with both SPWVD and MBD. It can be seen that the TF features set based on MBD provided slightly superior performance than SPWVD. This was also noted in Boashash et al. (SPWVD=0.93 and MBD=0.96) (Boashash et al., 2012b). Moreover, the performance of the NSDA obtained using features from AD was better when compared to that obtained from the pre-processed EEG signal at clinically relevant threshold (< 1 FD/h).

The performance of individual features for the detection of EEG seizures is shown in figure 5.9. Features F_1 and F_8 provide the highest discrimination between

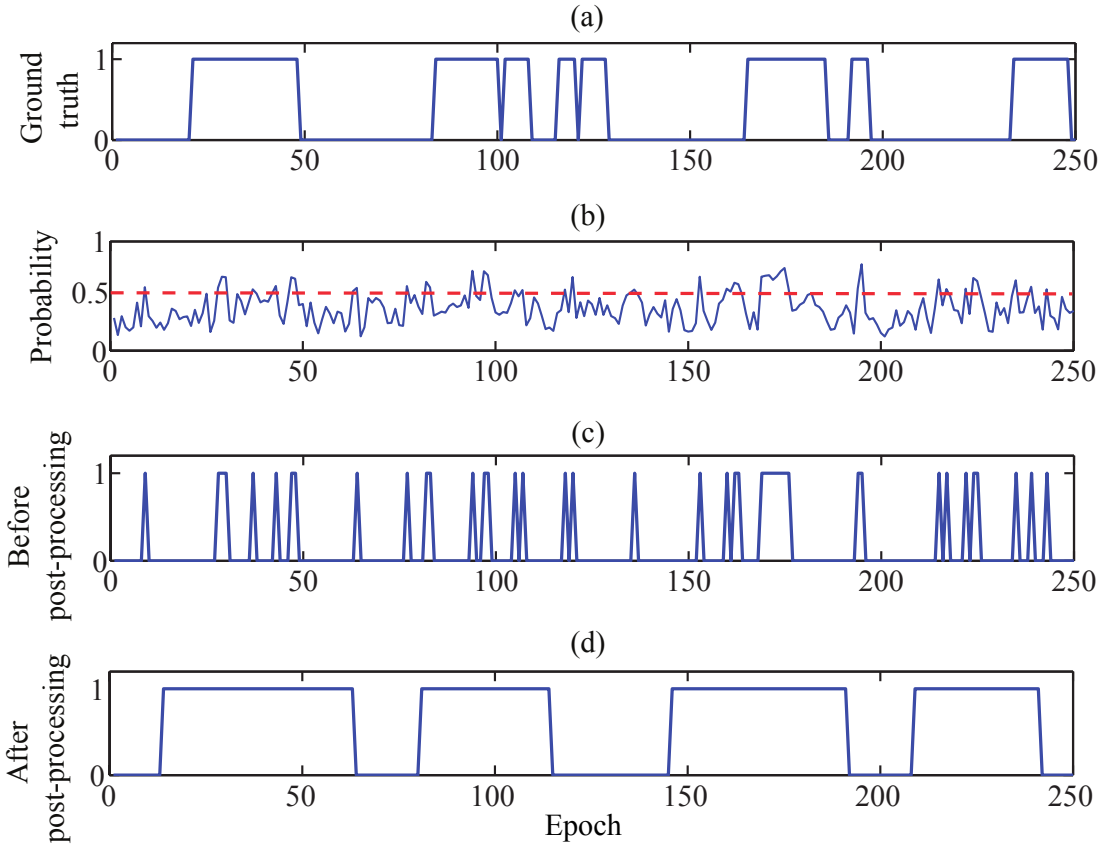


Figure 5.8: Illustration of the post-processing scheme. (a) The ground truth, where 1 indicates seizure. (b) The output of SVM converted to a probability via a sigmoid function. (c) The binary decisions resulting from applying a threshold of 0.5 to the probability of seizure . (d) The final binary decision after median filtering and adaptive collar operation, which increases the duration of the detected seizure proportional to its duration on either side.

seizure and non-seizure. The results obtained using the proposed NSDA for individual recordings is shown in Fig. 5.10. The median AUC of the NSDA obtained using AD was 0.91 and 0.93 (SPWVD and MDB, respectively). These values are significantly lower than outlined in Boashash et al. (0.96 vs 0.93) (Boashash et al., 2012b). This was expected as the large unedited dataset used in this study includes a large proportion of confounding EEG, such as artefact, that may not be available in smaller selected datasets. The performance of the proposed NSDA compared to other NSDA's are shown in figures 5.11 and 5.12.

From these results, it can be concluded that proper detection and measurement of the components of the nonstationary EEG signal can provide improved features for seizure detection in neonatal EEG. The proposed NSDA using AD offers significant improvement over other stationary and nonstationary methods for the detection of nonstationary variations observed in neonatal EEG. However, it

Table 5.2: Testing results (mean, median, IQR)% using the proposed NSDA. The performance using MBD outperforms SPWVD.

	FD/h	SDR	Sensitivity	Specificity
AD based TF features				
MBD	0.01	42.1, 47.3 (36.5–51.2)	40.3, 41.2 (38.1–45.7)	97.5, 99.6 (99.1–100)
	0.1	53.5, 55.3 (51.6–58.4)	41.3, 43.6 (40.8–51.3)	98.1, 97.2 (95.4–99.4)
	1.0	83.5, 87.2 (84.6–93.3)	75.2, 78.9 (72.6–84.3)	83.8, 85.4 (81.5–93.2)
SPWVD	0.01	40.5, 45.6 (35.4–50.6)	38.1, 40.4 (37.6–44.4)	97.1, 98.7 (98.5–100)
	0.1	52.1, 54.2 (50.1–57.2)	40.2, 42.3 (38.9–49.8)	97.5, 96.3 (93.4–97.6)
	1.0	81.4, 85.7 (83.5–92.1)	73.5, 76.7 (71.5–82.6)	81.2, 83.8 (80.1–91.3)
EEG signal based TF features				
MBD	0.01	6.8, 7.2 (6.1–8.8)	4.2, 5.1 (3.5–5.8)	99.5, 99.8 (99.2–100)
	0.1	26.5, 28.3 (22.6–30.4)	13.8, 16.6 (10.8–20.3)	98.1, 99.3 (97.2–99.8)
	1.0	90.8, 92.5 (86.6–94.3)	81.7, 83.6 (74.6–89.3)	91.4, 94.8 (86.5–97.2)
SPWVD	0.01	6.1, 6.4 (5.8–8.6)	3.6, 4.0 (3.2–5.4)	99.6, 99.8 (99.5–100)
	0.1	23.2, 25.1 (21.4–28.7)	12.6, 14.2 (11.5–18.7)	97.3, 98.5 (96.8–99.6)
	1.0	89.5, 91.4 (85.3–93.2)	78.3, 80.2 (73.5–88.2)	90.6, 93.4 (85.8–96.7)

should be noted that there are several issues in comparing the results with other NSDA's. The difference in databases can have a high impact on the performance of an NSDA as short duration EEG recordings have low artefact burden and have approximately equal seizure to nonseizure duration ratios (Stevenson et al., 2012b).

The performance of the proposed NSDA based on the TF feature-set obtained from AD is less than the methods of Temko et al. and Thomas et al. which are reported on the same datasets (AUCs of 0.96, 0.93) (Temko et al., 2011, Thomas et al., 2010). A more useful comparison at a clinically relevant threshold of 0.1 FD/h shows approximately a 20% reduction in seizure detection rate (62% vs 42.5%) as can be seen from figures 5.12 and table 5.2. It must be noted that the TF feature-set is significantly smaller than the feature-set used in these methods

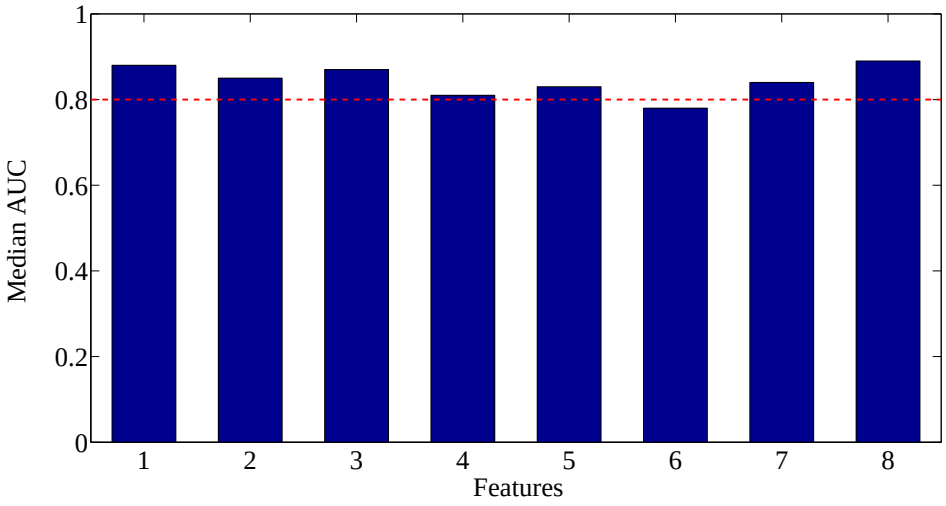


Figure 5.9: Performance of individual features obtained using OMP and MBD.

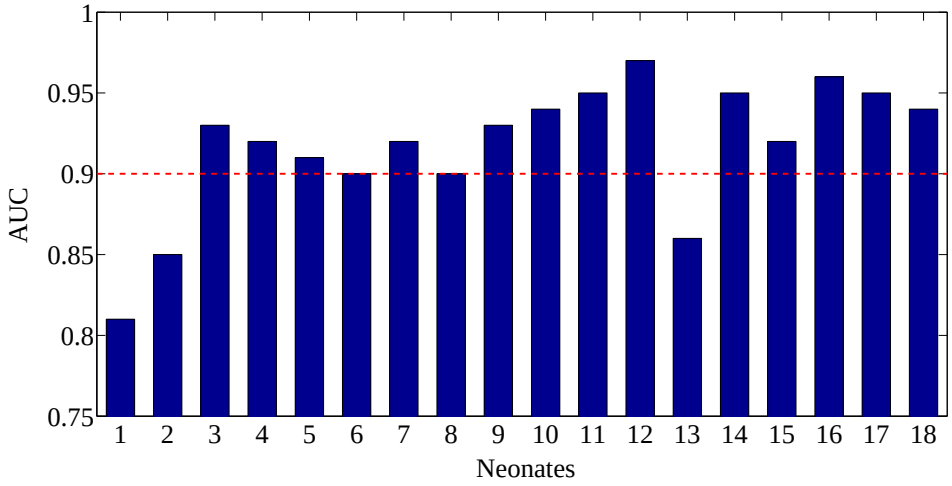


Figure 5.10: Performance of the proposed AD based NSDA for each neonate using MBD. It be clearly seen that the AUCs for all neonates are above 0.90 (marked as red dotted line) except for neonates 1,2 and 13 which also underperformed in the method proposed in chapter 4.

(8 vs 55 features) and may not be able to represent the variety of seizure and non-seizure EEG seen in the larger dataset.

Nevertheless, it was surprising that the TF feature-set which was developed on such a small dataset of neonatal EEG (Boashash et al., 2012b) has relatively good performance in the large dataset used in this thesis. This suggests the TF feature-set using AD is robust and its incorporation in the larger feature-set used in Temko et al. (Temko et al., 2011) and Thomas et al. (Thomas et al., 2010) may further improve NSDA performance.

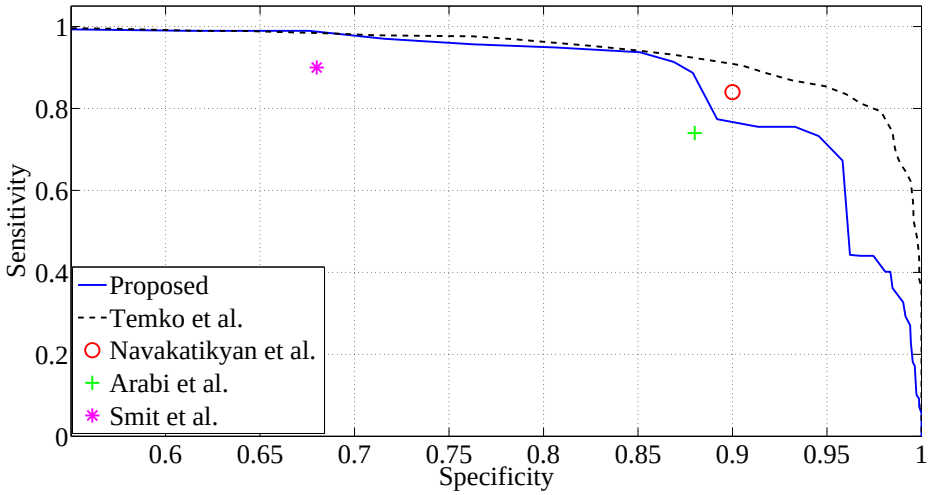


Figure 5.11: Comparison of the performance of the proposed AD based NSDA (median values) with several methods (epoch based metrics) currently disclosed in the literature using MBD (Aarabi et al., 2007, Navakatikyan et al., 2006, Smit et al., 2004, Temko et al., 2011).

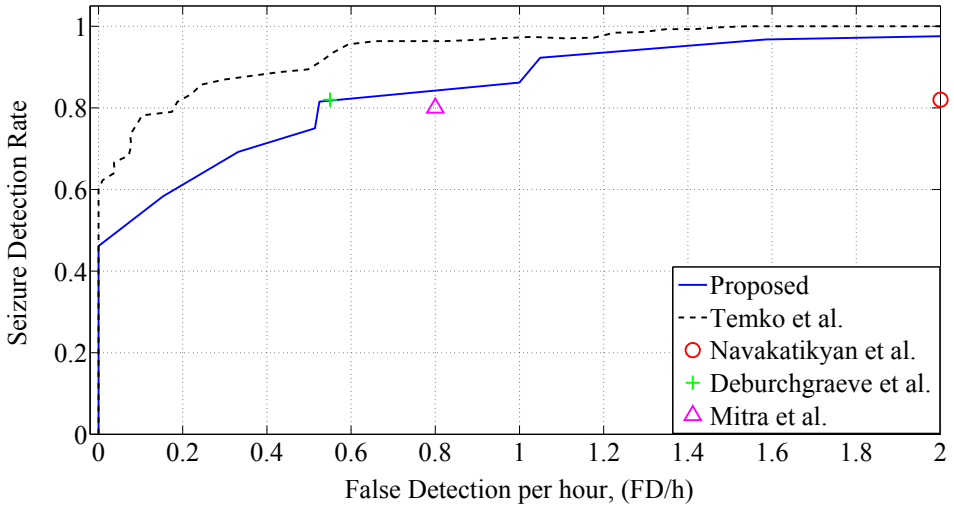


Figure 5.12: Comparison of the performance of the proposed AD based NSDA (median values) with several methods (event based metrics) currently disclosed in the literature using MBD (Deburchgraeve et al., 2008, Navakatikyan et al., 2006, Smit et al., 2004, Temko et al., 2011).

5.7 Summary

In this chapter, the robustness of TF features for neonatal seizure detection was studied. The proposed seizure detection system is based on the ability to track the nonstationarity of the EEG signal using QTFD analysis.

It has been previously shown that a SVM based seizure detection system us-

ing QTFD based TF features can provide good seizure detection performance (Boashash et al., 2012b). In this chapter, it was seen that the TF features obtained from pre-processed EEG signal provides poor performance at clinically relevant false detection thresholds in the larger dataset. Alternatively, an AD based NSDA using TF features obtained from the novel pseudo-periodic Duffing dictionary was proposed which provided improved performance when compared to the TF features obtained directly from the EEG signal.

The performance of the proposed NSDA was also compared with several NSDA's proposed in literature. It was concluded that the TF features obtained from the QTFD of the pseudo-periodic Duffing dictionary is robust and can be incorporated in the larger feature-set used proposed by Temko et al. (Temko et al., 2011) and Thomas et al. (Thomas et al., 2010) to improve the NSDA performance.

References

Aarabi, A., Grebe, R., Wallois, F., 2007. A multistage knowledge-based system for EEG seizure detection in newborn infants. Clinical Neurophysiology 118 (12), 2781–2797.

Boashash, B., 1991. Time-frequency signal analysis. Prentice Hall.

Boashash, B., 1992. Estimating and interpreting the instantaneous frequency of a signal. i. fundamentals. Proceedings of the IEEE 80 (4), 520–538.

Boashash, B., Boubchir, L., Azemi, G., 2011. Time-frequency signal and image processing of non-stationary signals with application to the classification of newborn EEG abnormalities. In: Signal Processing and Information Technology (ISSPIT), 2011 IEEE International Symposium on. IEEE, pp. 120–129.

Boashash, B., Boubchir, L., Azemi, G., 2012a. Improving the classification of newborn EEG time-frequency representations using a combined time-frequency signal and image approach. In: Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on. IEEE, pp. 280–285.

Boashash, B., Boubchir, L., Azemi, G., 2012b. A methodology for time-frequency image processing applied to the classification of non-stationary multichannel signals using instantaneous frequency descriptors with application to newborn EEG signals. EURASIP Journal on Advances in Signal Processing 2012 (1), 1–21.

Boashash, B., Mesbah, M., 2001. A time-frequency approach for newborn seizure detection. Engineering in Medicine and Biology Magazine, IEEE 20 (5), 54–64.

Box, G. E., Cox, D. R., 1964. An analysis of transformations. Journal of the Royal Statistical Society. Series B (Methodological), 211–252.

Boyd, S., Vandenberghe, L., 2009. Convex optimization. Cambridge University Press.

Boylan, G., Burgoyne, L., Moore, C., O'Flaherty, B., Rennie, J., 2010. An international survey of EEG use in the neonatal intensive care unit. Acta Paediatrica 99 (8), 1150–1155.

Cristianini, N., Shawe-Taylor, J., 2000. An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press.

Deburchgraeve, W., Cherian, P., De Vos, M., Swarte, R., Blok, J., Visser, G. H., Govaert, P., Van Huffel, S., 2008. Automated neonatal seizure detection mimicking a human observer reading EEG. Clinical Neurophysiology 119 (11), 2447–2454.

Gotman, J., 1990. Automatic seizure detection: improvements and evaluation. Electroencephalography and Clinical Neurophysiology 76 (4), 317–324.

Greene, B., Faul, S., Marnane, W., Lightbody, G., Korotchikova, I., Boylan, G., 2008. A comparison of quantitative EEG features for neonatal seizure detection. Clinical Neurophysiology 119 (6), 1248–1261.

Gribkov, D., Gribkova, V., 2000. Learning dynamics from nonstationary time series: analysis of electroencephalograms. Physical Review E 61 (6), 6538.

Hassanpour, H., Mesbah, M., Boashash, B., 2004. Time-frequency feature extraction of newborn EEG seizure using svd-based techniques. EURASIP Journal on Applied Signal Processing 2004, 2544–2554.

Jefferys, J., et al., 1990. Basic mechanisms of focal epilepsies. Exp Physiol 75 (2), 127–162.

Kohlmorgen, J., Müller, K.-R., Rittweger, J., Pawelzik, K., 2000. Identification of nonstationary dynamics in physiological recordings. Biological Cybernetics 83 (1), 73–84.

Maji, S., Berg, A. C., Malik, J., 2008. Classification using intersection kernel support vector machines is efficient. In: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, pp. 1–8.

Navakatikyan, M. A., Colditz, P. B., Burke, C. J., Inder, T. E., Richmond, J., Williams, C. E., 2006. Seizure detection algorithm for neonates based on wave-sequence analysis. Clinical Neurophysiology 117 (6), 1190–1203.

Niedermeyer, E., da Silva, F. L., 2005. Electroencephalography: basic principles, clinical applications, and related fields. Lippincott Williams & Wilkins.

Paluš, M., 1996. Nonlinearity in normal human EEG: cycles, temporal asymmetry, nonstationarity and randomness, not chaos. Biological Cybernetics 75 (5), 389–396.

Platt, J. C., 1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Advances in large margin classifiers.

Scher, M. S., Sun, M., Steppe, D. A., Guthrie, R. D., Sclabassi, R. J., 1994. Comparisons of EEG spectral and correlation measures between healthy term and preterm infants. Pediatric neurology 10 (2), 104–108.

Scholkopf, B., 2001. The kernel trick for distances. Advances in Neural information processing systems, 301–307.

Sejdić, E., Djurović, I., Jiang, J., 2009. Time–frequency feature representation using energy concentration: An overview of recent advances. Digital Signal Processing 19 (1), 153–183.

Smit, L., Vermeulen, R., Fetter, W., Strijers, R., Stam, C., 2004. Neonatal seizure monitoring using non-linear EEG analysis. Neuropediatrics 35 (06), 329–335.

Srinivasan, V., Eswaran, C., Sriraam, N, 2005. Artificial Neural network based epileptic detection using time-domain and frequency-domain features. Journal of Medical Systems 29 (6), 647–660.

Stevenson, N. J., O'Toole, J. M., Rankine, L. J., Boylan, G. B., Boashash, B., 2012a. A nonparametric feature for neonatal EEG seizure detection based on a representation of pseudo-periodicity. Medical Engineering & physics 34 (4), 437–446.

Stevenson, N. J., O'Toole, J. M., Rankine, L. J., Boylan, G. B., Boashash, B., 2012b. A nonparametric feature for neonatal EEG seizure detection based on a representation of pseudo-periodicity. Medical Engineering and physics 34 (4), 437–446.

Subasi, A., 2007. EEG signal classification using wavelet feature extraction and a mixture of expert model. Expert Systems with Applications 32 (4), 1084–1093.

Subasi, A., Erçelebi, E., Alkan, A., Koklukaya, E., 2006. Comparison of subspace-based methods with ar parametric methods in epileptic seizure detection. Computers in Biology and Medicine 36 (2), 195–208.

Temko, A., Thomas, E., Marnane, W., Lightbody, G., Boylan, G., 2011. EEG-based neonatal seizure detection with support vector machines. Clinical Neurophysiology 122 (3), 464–473.

Thomas, E., Temko, A., Lightbody, G., Marnane, W., Boylan, G., 2010. Gaussian mixture models for classification of neonatal seizures using EEG. Physiological measurement 31 (7), 1047.

Tzallas, A., Tsipouras, M., Fotiadis, D., 2007. Automatic seizure detection based on time-frequency analysis and artificial Neural networks. Computational Intelligence and Neuroscience 2007.

Vapnik, V., Cortes, C., 1995. Support-vector networks. Machine learning 20 (3), 273–297.

Widman, G., Schreiber, T., Rehberg, B., Hoeft, A., Elger, C., 2000. Quantification of depth of anesthesia by nonlinear time series analysis of brain electrical activity. Physical Review E 62 (4), 4898.

Wilson, S. B., Scheuer, M. L., Plummer, C., Young, B., Pacia, S., 2003. Seizure detection: correlation of human experts. Clinical Neurophysiology 114 (11), 2156–2164.

Chapter 6

Automatic Grading of EEG Abnormality with Hypoxic Ischaemic Encephalopathy using Atomic Decomposition and the Support Vector Machine

> Music does not influence research work, but both are nourished by the same sort of longing, and they complement each other in the release they offer. ––Albert Einstein

6.1 Introduction

Hypoxic-ischaemic encephalopathy (HIE) is a brain injury caused by the lack of supply of oxygen and impairment to the blood supply in the neonatal brain. HIE due to fetal or neonatal asphyxia is a leading cause of death or severe impairment among infants, with an incidence of 2-3/1000 births (Graham et al., 2008, Inder and Volpe, 2000). The babies with such complications are monitored closely in the neonatal intensive care unit (NICU) for further assessment. It is difficult to determine the state of the baby's brain at an early stage (Scher, 2001). Within the first 24 hours of life, the infant with such complications can develop symptoms of apnea and seizures with abnormal EEG. Some of the clinical symptoms include stupor or coma, irregular breathing, absence of neonatal reflexes, disturbances of ocular motion, irregular heart rate and blood pressure (Fenichel,

2009).

Within the NICU the heart rate, blood pressure, oxygen saturation of the babies are monitored regularly (Nicklin et al., 2004), but these parameters only provide information on their general condition. It is important to obtain the long term functioning of the brain which can be done through the electroencephalogram (EEG) (Boylan et al., 2013). Through EEG, continuous monitoring of the brain can be carried out over a long time with high time resolution.

The severity of the HIE insult determines the outcomes of HIE. Mild HIE can result in a normal outcome: moderate encephalopathy can increase the risk of neurological disability to 20-40% and severe encephalopathy can lead to severe neurological disability, or in some cases, death (Gray et al., 1993). It was shown that the effects of the developing encephalopathy could be reduced by Therapeutic Hypothermia (which is also known as *cooling of brain*) for the newborn (Azzopardi et al., 2009) within 6 hours of birth. Since the treatment is less effective after 6 hours of birth (Shankaran, 2012), it is important for the clinical neurophysiologists to detect the severity of the HIE brain injury as early as possible.

The grading of HIE can be done either by clinical assessment or by using the EEG or amplitude integrated EEG (aEEG). EEG based HIE grading is mainly done through visual analysis of the background activity. Several features such as the continuity of the EEG signal, interhemispheric symmetry and synchrony, amplitude, frequency content, and presence or absence of sleep-wake cycling (SWC) are used to grade HIE (Murray et al., 2009). HIE is graded into four main types which are summarized in Table 6.1 and various patterns corresponding to each grade of HIE-EEG across different patients are shown in figures 6.1 and 6.2. The grading of HIE is usually done over an hour-long recording of EEG. It can be seen from figures 6.1 and 6.2 that the HIE-EEG exhibits various patterns across different HIE grades. Sometimes, these patterns may be similar across HIE grades, however the measurement of the variability of these patterns over the entire duration of the EEG is important for assigning grades.

Automatic grading of HIE using EEG (HIE-EEG) is a relatively new area. A first attempt to classify HIE-EEG based on the quantitative analysis of background EEG was investigated by (Korotchikova et al., 2011). In this approach, 9 quantitative EEG (qEEG) measures that replicates the neurophysiologist's approach of visually classifying neonatal EEG were used to grade HIE and are given in Table 6.2. The authors reported a classification accuracy of 91% on a dataset

Table 6.1: Properties of HIE

EEG/HIE grade	Abnormality level	Characteristics
1	normal/mild	Continuous background pattern with mild asymmetries and voltage depression (30-50 μV). Presence of poorly defined SWC.
2	moderate	Discontinuous activity with IBI $\leq$10 s. Presence of clear asymmetry or asynchrony and disrupted SWC.
3	major	Discontinuous activity with IBI 10-60s,severe disruption of background patterns (<30 μ V), absence of SWC.
4	severe	Background activity of $\leq$ 10 μV, or severe discontinuity with IBI $\geq$ 60 s.

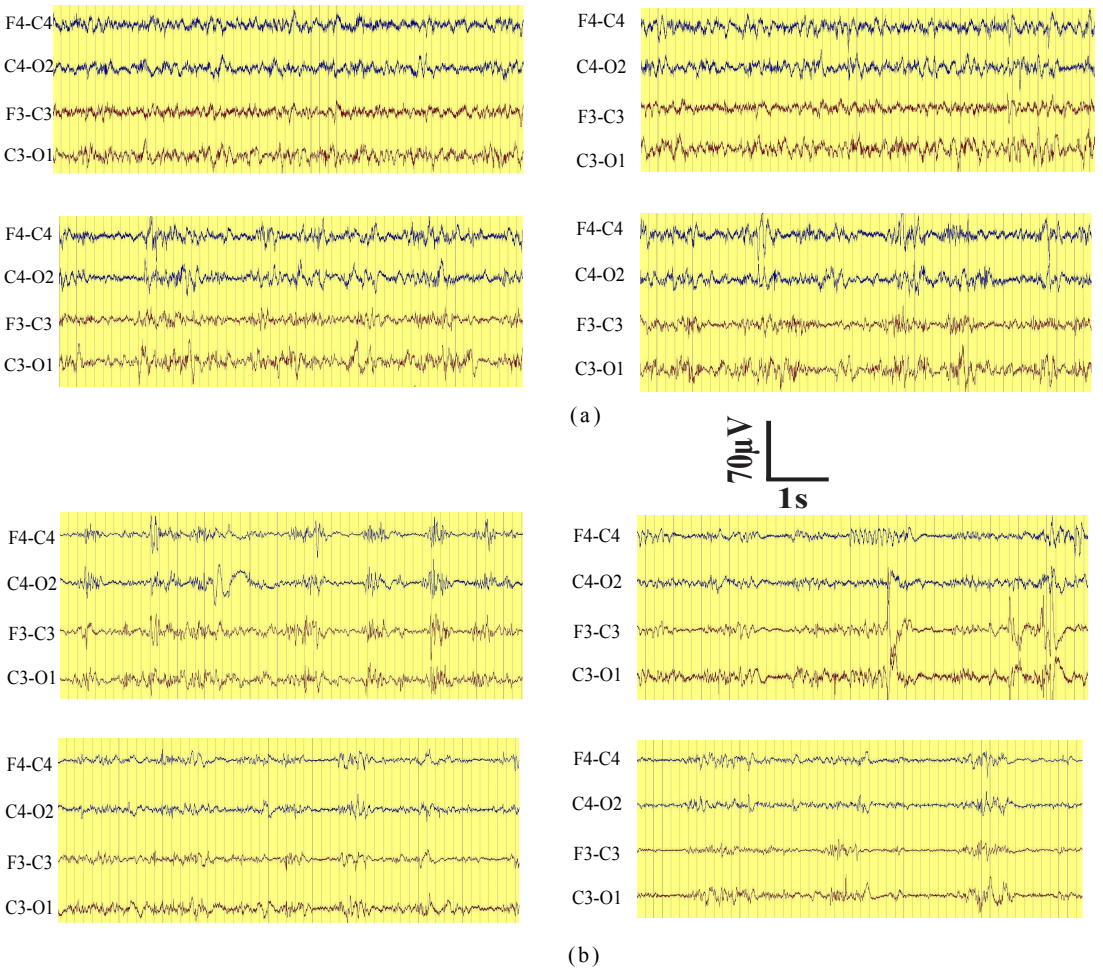


Figure 6.1: Clear examples of 60s EEG from different patients showing grades of HIE (a) Grade 1: Normal/Mild abnormalities. (b) Grade 2: Moderate abnormalities.

consisting of approximately one hour long EEG selected from the recordings of 54 full term neonates with HIE. However, the results obtained were estimated on artefact free data and their performance dropped to 72% when validated on artefactual data. Stevenson et al. extended this study by using features ob-

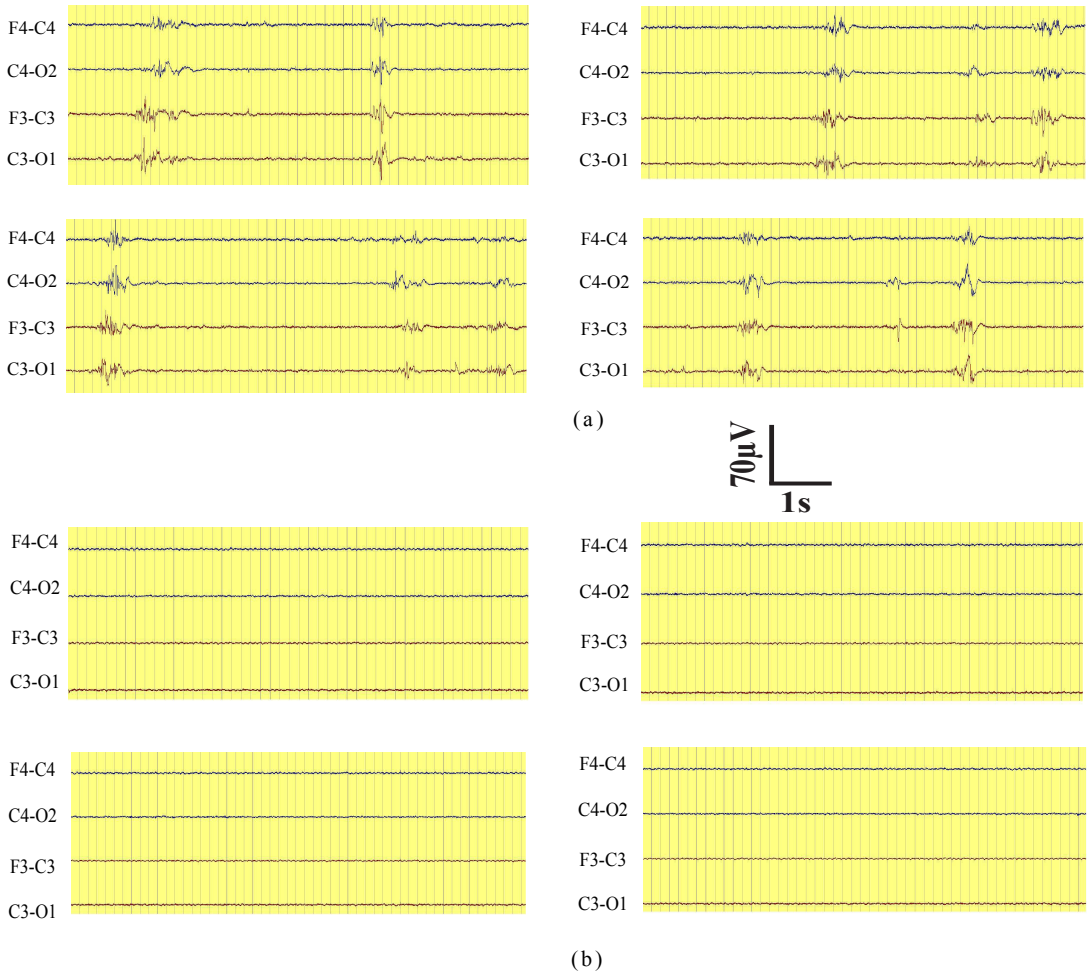


Figure 6.2: Clear examples of 60s EEG from different patients showing grades of HIE (a) Grade 3: Major abnormalities. (b) Grade 4: Inactive.

Table 6.2: Quantitative measures replicating visual characteristics used to grade HIE in (Korotchikova et al., 2011) .

Visual characteristics	Quantitative measure
Continuity of EEG signal	Skewness, kurtosis, discontinuity
Amplitude	Amplitude integrated EEG
Frequency content	Fractal dimension, relative delta power, spectral edge frequency
Symmetry	Revised brain symmetry index
Synchrony	Linear correlation coefficient

tained from the background EEG model (Stevenson et al., 2013). An automated grading system (AGS) using 15 features from the model of background EEG was proposed in this study. The AGS was trained on artefact free data and tested on unseen data with artefacts. The AGS provided an accuracy of 83.3% and it was reported that the majority of misclassifications were due to the presence of arte-

facts and several abnormal patterns in the EEG such as sharp waves, asymmetry and asynchrony. Ahmed et al. proposed an AGS based on a cross disciplinary method of using support vector machine (SVM) and supervectors (Ahmed et al., 2014). This system used a total of 55 features which provided a generic EEG description computed from the frequency, time and information theory domains. The AGS was trained and tested on data with artefacts and an accuracy of 85% was reported. Recently, an holistic approach for grading background EEG was proposed by Matic et al. (Matic et al., 2014) based on adaptive EEG segmentation and its mapping on segment's feature space. This AGS was trained and tested on 272 EEG epochs (each epoch was 1 hour in duration) selected from 34 neonates which provided an accuracy of 89% in grading HIE.

The main aim of developing an automated grading system (AGS) is to mimic the visual interpretation of the neurophysiologist for grading HIE-EEG. It has been shown that neonatal EEG background is a band-limited nonstationary process (Boashash and Mesbah, 2001, Notley and Elliott, 2003). Prior work on neonatal seizure detection used short epochs of 8s in duration based on the assumption that the neonatal EEG is nearly stationary over such a short duration (Nagaraj et al., 2014, Temko et al., 2011). However, for grading the background EEG a much longer duration epoch is necessary to capture the variation of several time varying components such as burst suppression which plays an important role in discriminating different HIE grades. It was shown that an epoch length of 64 s provided optimal performance in grading HIE-EEG (Korotchikova et al., 2011, Stevenson et al., 2013) on the same database used in this thesis.

In this chapter, the performance of several methods to automatically classify HIE-EEG using the support vector machine (SVM) are compared. This includes the Hilbert transform methods described in (Stevenson et al., 2013), AD based methods (Tropp and Gilbert, 2007) and the multichannel matching pursuit (MMP) method (Durka et al., 2005). After several post-processing steps, the automatic grading of HIE is obtained. The misclassified recordings were further examined for the presence of abnormal patterns such as artefacts, runs of sharp waves, asymmetry, and asynchrony.

6.2 HIE dataset

The dataset used in this study consisted of EEG recordings from 54 full term neonates with HIE with each recording approximately one hour long. The EEG

data were free of seizures and major movement artefacts (amplitude higher than 250 μV lasting for more than 3s). The data were recorded in the NICU of Cork University Maternity Hospital (CUMH), Cork, Ireland using the NicoletOne EEG system (Carefusion Neurocare, Wisconsin, USA) with a sampling frequency of 256 Hz. The EEG was recorded within 12h of birth and it was continued for 24-72h to monitor the evolution of the developing HIE (Mizrahi et al., 2004, Niedermeyer and da Silva, 2005). The neonates were not treated with therapeutic hypothermia. The data were annotated using eight EEG channels in bipolar montage: F4-C4, C4-O2, F3-C3, C3-O1, T4-C4, C4-Cz, Cz-C3, and C3-T3. The segments of the EEG recording selected had constant HIE grade within an hour. The one hour EEG recordings were graded independently by two neonatal EEG experts (**IK** and **GBB**)[1] using the system proposed in (Murray et al., 2009) which is summarized in Table 6.1. The neonatal EEG was assigned one of the 4 grades based on the degree of abnormality: *grade 1*-moderate/mild abnormalities, *grade 2* - moderate abnormalities, *grade 3* - major abnormalities and *grade 4* - inactive EEG. Overall the database used in this study consisted of 62h of 8-channel EEG recordings with a median duration of 65 min per recording (IQR: 62-67 minutes). There were 22 neonates with grade 1 EEG, 14 neonates with grade 2 EEG, 12 neonates with grade 3 EEG and 6 neonates with grade 4 EEG. The same dataset had been previously used in (Stevenson et al., 2013) and a direct comparison of the results can thus be obtained.

6.3 Neonatal background EEG models

Several newborn EEG models have been proposed in literature based on the assumption that the neonatal EEG background is a random process. The neonatal background EEG signal consists of both stochastic and chaotic characteristics with amplitude and frequency contents varying over time (Niedermeyer and da Silva, 2005, Varsavsky et al., 2010). This change with time (nonstationarities) can last for a few seconds or sometimes can extend to hours. The model proposed by Roessgen et al. uses a colored noise process to simulate the neonatal background EEG (Roessgen et al., 1998). However, this model does not take into account the nonstationary characteristics of the neonatal EEG background. As an extension to the Roessgen model, Celka and Colditz proposed a Weiner filter based model to incorporate the nonlinear nature of the neonatal EEG back-

[1] **IK** - Dr. Irene Korotchikova and **GBB** - Prof. Geraldine B. Boylan are in Department of Pediatrics and Child health, CUMH.

ground (Celka and Colditz, 2002). This model was stationary (time-invariant) and therefore failed to incorporate the nonstationary behaviour of the neonatal EEG background (Boashash and Mesbah, 2002, Liu et al., 1992).

An alternate neonatal EEG background model was proposed by Rankine et al. in which the EEG background was modelled as a random signal with time varying spectrum using the inverse power law (Rankine et al., 2007). Since large number of parameters were randomly selected, this model failed to account for the dependence between parameters; the background EEG generated by this model were not representative of real newborn EEG background. An improved version of the newborn EEG background model was proposed by Stevenson et al. using a minimal set of model parameters (Stevenson et al., 2010). A nonlinear dynamic system driven by stationary white Gaussian noise was used in this model to simulate neonatal EEG background. The nonlinear dynamic system used different parameters to best represent neonatal EEG background. The nonlinear and nonstationary characteristics of the neonatal EEG background were incorporated by the models proposed by (Rankine et al., 2007) and (Stevenson et al., 2010).

In all the above mentioned models, the neonatal background EEG has been modelled as a colored noise process. The general form of the model can be written as (Stevenson et al., 2013)

$$EEG(t) = X(t) = T_{filter}\{Y(t)\}, \tag{6.1}$$

where $T_{filter}\{.\}$ is the linear or nonlinear transformation and $Y(t)$ is white Gaussian noise. The frequency response of the simulating filter T_{filter} follows a power law response of $1/f^{\alpha}$ (Celka and Colditz, 2002, Roessgen et al., 1998, Stevenson et al., 2007). However, this model fails to take into account the presence of amplitude modulation (AM) seen in trace alterant (associated with quiet sleep) and burst-suppression (associated with EEG grades 2 and 3) EEG patterns of the term neonates (Walsh et al., 2011). With the inclusion of AM in equation 6.1, the EEG model can be written as (Stevenson et al., 2013)

$$EEG(t) = a_m(t)X(t), \tag{6.2}$$

where $a_m(t)$ is the component representing AM. Due to the dual form of equation 6.2, several combinations of $a_m(t)$ and $X(t)$ can be used to generate the background EEG signal, $EEG(t)$. However, by constraining the frequency content of the $a_m(t)$ to be much less than $X(t)$, the frequency domain representation of

$X(t)$ can provide information on the underlying signal transformation $T_{filter}\{.\}$ via Instantaneous Frequency (IF) estimates. Therefore, the interpretation of the $a_m(t)$ and $T_{filter}\{.\}$ can be considered similar to the interpretation of the EEG made by a neurophysiologist based on the amplitude and frequency contents respectively (Stevenson et al., 2013). Moreover, the AM and IF components may also represent the influence of the cortico-cortical and thalamo-cortical activities on EEG (Steriade et al., 1993, Vanhatalo and Kaila, 2006).

Inspired by the model of the neonatal EEG background in (Roessgen et al., 1998, Stevenson et al., 2007, 2010), a model of the background EEG was proposed by Stevenson et al. in the joint energy time-frequency domain in terms of AM and IF as (Stevenson et al., 2013):

$$|EEG(t,f)|^2 = \left|\frac{a_m(t)}{f^{\alpha(t)}}\right|. \tag{6.3}$$

6.4 Automatic grading system using time-frequency (TF) features

An overview of the proposed multiclass SVM based background EEG automatic grading system (AGS) is shown in figure 6.3. This is a multichannel system which consists several pre-processing and post-processing steps for extracting time-frequency features and classification. In the pre-processing step, N channels of the EEG were filtered with a highpass filter with a cutoff frequency f_c (which will be determined in section 6.9)and a transition width of 0.5 Hz. Since the EEG activity of interest in neonates is negligible over 32 Hz, the EEG was down-sampled to 64 Hz from 256 Hz. The EEG was then segmented into 64 s epochs with a 32 s overlap (50% overlap). The duration of the epoch was selected based on the definition of EEG grade 4 which states an interburst interval (IBI) $>$ 60s (see Table 6.1). In the feature extraction step, a basic statistical summary (mean, standard deviation, skewness, kurtosis) of the amplitude modulation (AM) and Instantaneous frequency (IF) obtained from the EEG epoch were estimated and used as key features to characterize the EEG summarized in table 6.1. 8 features were estimated from the AM and IF measures of the EEG epoch which are summarized in table 6.3 (Stevenson et al., 2013).

Each EEG epoch was therefore represented by 8 measures estimated from the AM and IF. These features provide basic statistical analysis of the variations in

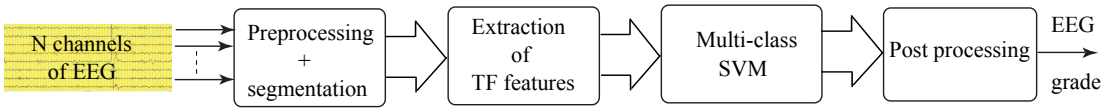


Figure 6.3: Illustration of the proposed AGS.

Table 6.3: The features obtained from AM and IF of each epoch.

Feature no.	EEG feature
1	mean(AM)
2	standard deviation (AM)
3	skewness(AM)
4	kurtosis(AM)
5	mean(IF)
6	standard deviation(IF)
7	skewness(IF)
8	kurtosis(IF)

the amplitude and frequency characteristics of the EEG epoch. This process was repeated for all 8 channels. Figures 6.4 and 6.5 provides the distribution of TF features across different HIE grades for a sample one hour long EEG recording from each grade using the Hilbert transform method (discussed in section 6.5). It can be seen that there is sufficient discriminatory information provided by these features that can be used to automatically classify different HIE grades. Moreover, these features mimic the amplitude and frequency characteristics of the visual interpretation of the EEG used to grade HIE (described in Table 6.2) and can be useful in automatic classification.

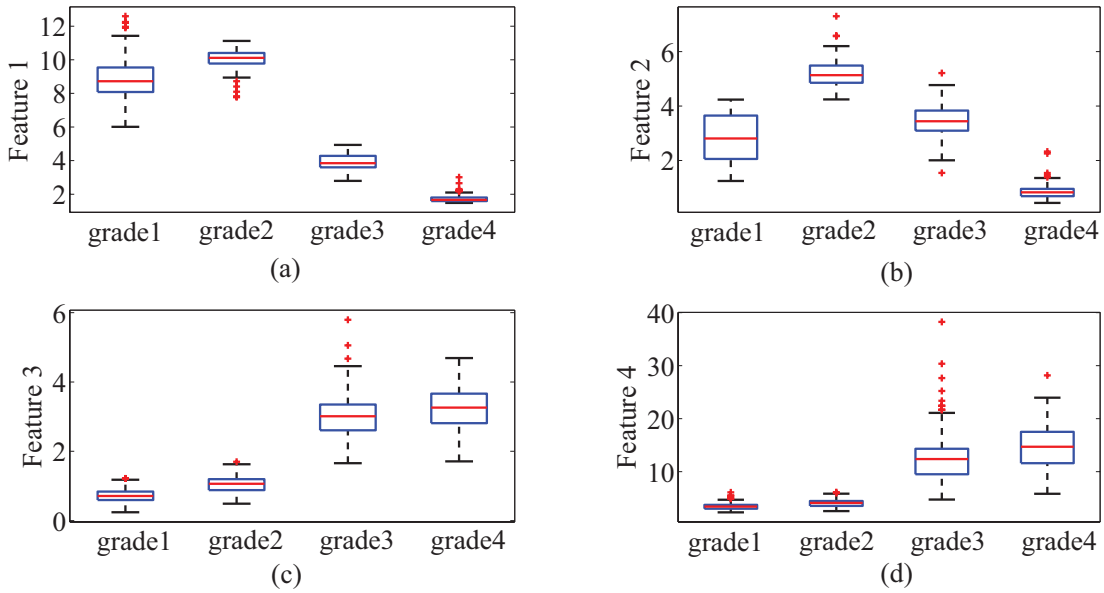


Figure 6.4: Boxplots showing the distribution of AM features across different HIE grades for a sample one hour EEG recording, from each grade.

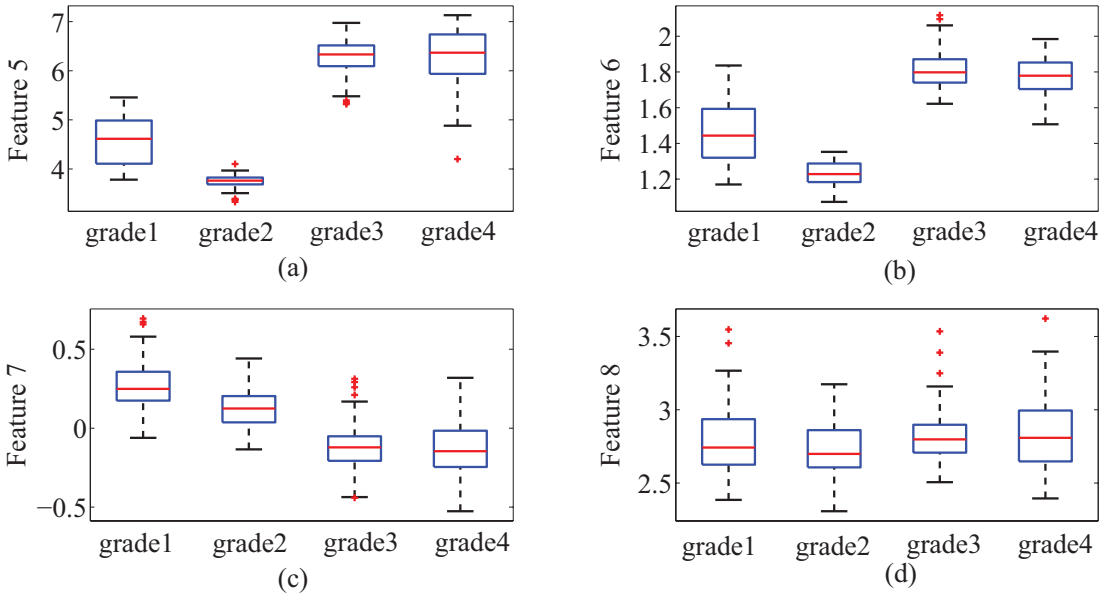


Figure 6.5: Boxplots showing the distribution of IF features across different HIE grades for a sample one hour EEG recording, from each grade.

6.4.1 Multichannel information fusion

In this study, since there are 8 EEG channels, selection of optimal channel/channels for classification is important, this can be considered as a feature selection problem. However, unlike feature selection where only certain features are selected, it is important to consider all features obtained from the individual channel to preserve the information. Selection of channels is important because:

1. with the increase in number of channels, the computational complexity of the classification model increases, and
2. irrelevant features or signals may result in additional noise which can reduce the accuracy of the classification model.

There are various approaches to deal with the multichannel nature of the data for grading HIE. Most EEG analysis algorithms do not utilize the information present in multichannel EEG and are limited to using single channel EEG. To overcome this limitation, there are several *fusion* approaches to include the EEG information present across all the channels (Wald, 1999) which are briefly described below.

1. *Multichannel feature fusion*: Multichannel feature fusion is most commonly used for the fusion of several features which is illustrated in figure 6.6. In this approach, the fusion combines features obtained from the individual

channels into a single global set of features. By this way, a large feature vector is obtained which can be used for classifying multiple channel EEG. The fusion can be done by taking the max or mean or median across the features obtained from different channels. The major advantage of this approach is that it requires only one classifier for the classification of N channel EEG.

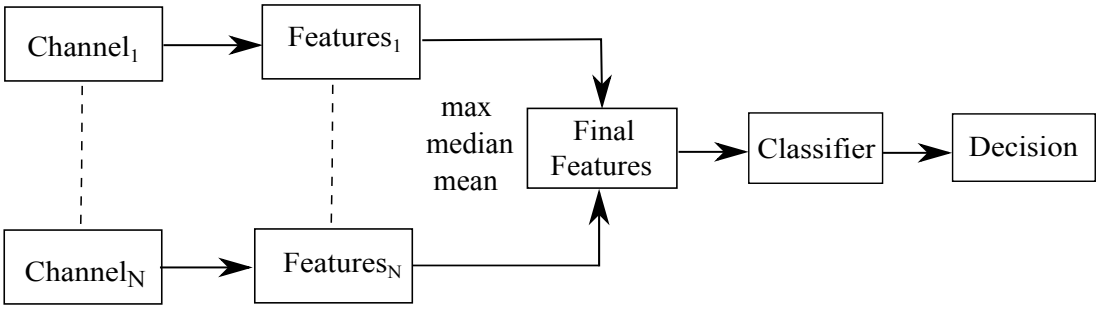


Figure 6.6: Overview of the multichannel features fusion method. The features from individual channels are combined to obtain a single feature set which is then used for classification.

2. *Multichannel classifier fusion*: In this approach, the outputs of each channel classifier are fused prior to making a decision as illustrated in figure 6.7. For each channel separate multi-class classifiers are used and the output of classifiers are fused (mean, max, median, majority voting etc) to make the final decision. The major disadvantage of this approach is that it requires N multi-class classifiers for N channel EEG data and therefore increases the computational complexity.

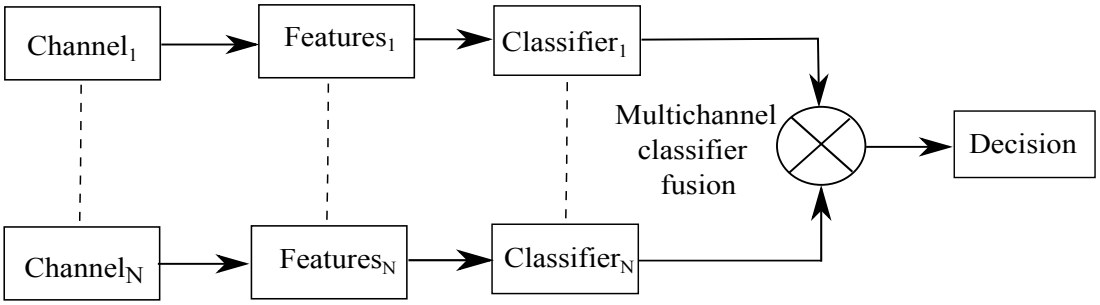


Figure 6.7: Overview of the multichannel classifier fusion method. The outputs from individual classifiers are combined to obtain a single classifier which is then used for decision making.

3. *Multichannel decision fusion*: As illustrated in figure 6.8, the decision making is done in two steps in this approach. The classifier output from each channel is used to make the initial decision. The individual channel based

decisions are then combined to make a final decision by postprocessing, majority voting etc. This approach also uses N multiclass classifiers.

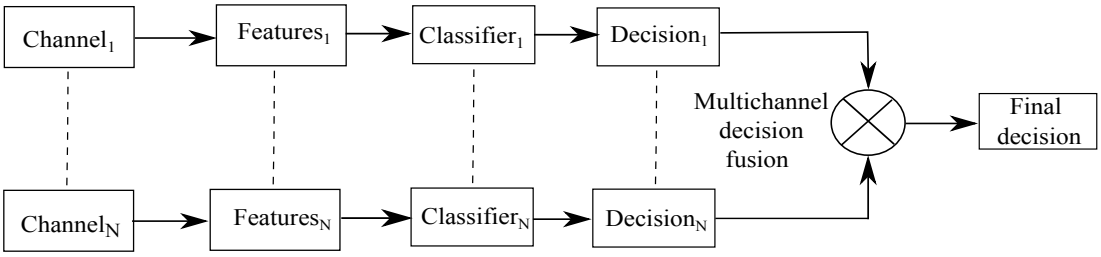


Figure 6.8: Overview of the multichannel decision fusion method. The decisions from individual classifiers are combined to obtain the final decision.

6.4.2 Classification using a multiclass support vector machine

The support vector machine (SVM) is basically a binary classifier. Ideally, the SVM maps the inseparable data in the input feature space to a high dimensional feature space to separate the data into two classes using hyperplanes. Several methods of extending the use of the SVM for multiclass problems have been suggested in the literature, which can be grouped into two categories:

Category 1: In this category, the whole dataset with all the classes is used simultaneously and solves the multiclass problem directly (Crammer and Singer, 2002, Vapnik and Vapnik, 1998). The main drawback of these methods is that due to the large number of variables that need to be optimized, they present numerical difficulties and are difficult to implement.

Category 2: In this category, the classification is decomposed into a binary classification problem (Vapnik and Vapnik, 1998) and is commonly used for multiclass classification problems. The methods used in this approach are: *one-against-all* and *one-against-one*. In both these methods, a binary SVM classifier is constructed separating the datapoints of one class against the other. After testing, each SVM classifier provides a decision value or class for the test datapoint and the label is assigned to the datapoint from the classifier with the highest positive decision value. *One-against-all* was used in the present study due to its feasibility and reduced computational complexity (Hsu and Lin, 2002).

In the *one-against-all* method, m SVM models corresponding to m classes are constructed. From the samples used for training, the i^{th} SVM is trained with the

examples in the i^{th} class assigned positive labels and all other examples with negative labels (Hsu and Lin, 2002). Given M training data, $(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \ldots (\mathbf{x}_M, \mathbf{y}_M)$ where $\mathbf{x}_i \in R^n,\ i = 1, \ldots, M$ and $\mathbf{y}_i \in \{1, \ldots, m\}$ is the class of $\mathbf{x}_i$, the problem of classification for the i^{th} SVM can be written in terms of optimization problem as:

$$\begin{aligned} \text{minimise } & \frac{1}{2}||\mathbf{w}^i||^2 + C\sum_{j=1}^{M}\epsilon_j^i \qquad suchthat \\ & (\mathbf{w^i})^T\boldsymbol{\phi}(\mathbf{x}_j) + b^i \geq +1 - \epsilon_j^i, \ \text{ if } y_j = i \\ & (\mathbf{w^i})^T\boldsymbol{\phi}(\mathbf{x}_j) + b^i \leq -1 + \epsilon_j^i, \ \text{ if } y_j \neq i \\ & \epsilon_j^i \geq 0, j = 1, \ldots M. \end{aligned} \tag{6.4}$$

Using the maximum operation, a final decision for a feature vector $\mathbf{x}_j$ could be determined as,

$$\hat{y}(\mathbf{x}_j) = \underset{i \in 1,2,\ldots,m}{\arg\max} \left((\mathbf{w}^i)^T\boldsymbol{\phi}(\mathbf{x}_j) + b^i \right). \tag{6.5}$$

Similar to the method described in chapter 5, the output of the SVM can be converted to a probabilistic measure of a HIE grade, bounded within [0,1] via Platt scaling (Platt, 1999) using a sigmoid function as:

$$P(w_1|\mathbf{x}) = \frac{1}{1 + exp(Af(\mathbf{x}) + B)}, \tag{6.6}$$

where $P(w_1|\mathbf{x})$ is the probability of the HIE grade assigned by the classifier for an epoch, $f = f(\mathbf{x})$ is the output of the SVM classifier, A and B are the sigmoid function parameters which are estimated over the training dataset (Platt, 1999).

6.4.3 Postprocessing

Majority voting is used here to obtain the final grade of a given sequence vector obtained from the multiclass SVM. In the first step, the output of the i^{th} SVM model is stored in a vector S_i. The majority voting of the decisions from these vectors (S_i) is performed to obtain the best grade from each SVM classifier in a vector S_{CL}. Later, in the second step, the output of the best performing SVM classifier is assigned as the final grade to the HIE-EEG. This process of majority voting for a 4-class classification problem is illustrated in figure 6.9.

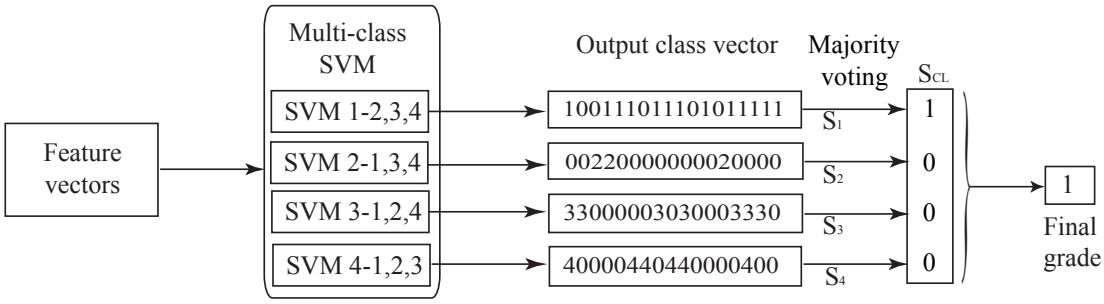


Figure 6.9: Illustration of the majority voting system for assigning an overall HIE grade.

6.5 AGS based on the Hilbert transform

The Hilbert transform based TF feature extraction method was proposed in (Stevenson et al., 2013). In this method, it was assumed that the quadratic time-frequency distribution (QTFD) of the EEG signal can be used to obtain the time varying amplitude and frequency measures which can be used to grade the severity of brain injury in the EEG of neonates with HIE. The general form of QTFD is given by

$$\begin{aligned}\rho_z(t,f) &= \underset{\tau\to f}{\mathcal{F}} G(t,\tau) \underset{t}{*} z\left(t+\frac{\tau}{2}\right) z^*\left(t-\frac{\tau}{2}\right) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(u,\tau) z\left(t-u+\frac{\tau}{2}\right) z^*\left(t-u-\frac{\tau}{2}\right) du e^{-2\pi f\tau} d\tau,\end{aligned} \tag{6.7}$$

where $z(t) = EEG(t) + j\mathcal{H}\{EEG(t)\}$ is a complex signal, $\mathcal{H}\{EEG(t)\}$ is the Hilbert transform of $EEG(t)$, $z^*(t)$ is the complex conjugate of $z(t)$, f is frequency, t is time, τ is the time lag and $G(t,\tau)$ is the time-lag kernel.

From $\rho_z(t,f)$, the AM and IF can be approximated as:

$$a_m(t) = \sqrt{\int_{-\infty}^{\infty} \rho_z(t,f) df} \tag{6.8a}$$

$$f_i(t) = \frac{\int_{-\infty}^{\infty} f\rho_z(t,f) df}{\int_{-\infty}^{\infty} \rho_z(t,f) df}. \tag{6.8b}$$

With an assumption of a slowly time-varying transformation, $T_{filter} = 1/f^{\alpha(t)}$ and a bandlimited process constrained by lower (f_1) and upper (f_2) frequencies, the relationship between IF and $\alpha(t)$ can be estimated as (Stevenson et al., 2013):

$$f_i(t) = \frac{\left(f_2^{2-2\alpha(t)} - f_1^{2-2\alpha(t)}\right)(1-2\alpha(t))}{\left(f_2^{1-2\alpha(t)} - f_1^{1-2\alpha(t)}\right)(2-2\alpha(t))}, \quad \alpha(t) > 1 \tag{6.9}$$

where f_2 is the Nyquist rate ($f_2 = 32$Hz in this case) and f_1 is optimally selected between 0.5 and 5 Hz for the purpose of eliminating low frequency artefacts from the EEG recording.

The QTFD of a given real discrete time signal $x[n]$, $n = 0, 1, \ldots, N-1$, is defined in terms of its analytic associate $z[n]$ as (Boashash, 2003):

$$\rho[n,p] = 2G[n,m] \underset{n,m}{**} \sum_{m=1}^{N} z[n+m]z^*[n-m]e^{\frac{-j2\pi pm}{N}} \tag{6.10}$$

where $G[n,m]$ is the kernel of the TFD and $\rho[n,p]$ is an $N \times N$ matrix. From $\rho[n,p]$, the AM and IF can be approximated as:

$$a_m[n] = \sqrt{\sum_{p=1}^{N} \rho[n,p]} \tag{6.11a}$$

$$f_i[n] = \frac{\sum_{p=1}^{N} f[p]\rho[n,p]}{\sum_{p=1}^{N} \rho[n,p]}. \tag{6.11b}$$

A smoothed $\rho[n,p]$ using a two-dimensional Hamming window was used to estimate the AM and IF of the EEG signal. The bandwidth-time product must be greater than one (which will be estimated in next section) for a smoothed $\rho[n,p]$ to provide a non-negative representation of the EEG signal (Janssen and Claasen, 1985). This provides non-negative AM and IF of the EEG signals. The AM and IF for a sample 64 s epoch of different HIE grade is shown in figure 6.10. We can clearly see the variation in AM and IF components across different grades. The measurement of these variations could provide important information about the particular HIE-EEG epoch.

The AGS using the Hilbert transform is shown in figure 6.11. In the preprocessing step, the 8 channel EEG was initially filtered with a highpass filter with a cutoff frequency f_c and down-sampled to 64 Hz from 256 Hz. The EEG was then segmented into 64 s epochs with a 32 s overlap (50% overlap). Since HIE is assumed to be a global injury in this study, the *median* feature vector across 8-channels (multichannel feature fusion) was obtained which combines the information across the EEG channels to form a single feature set (Stevenson et al., 2013). When compared to the *mean*, the *median* operator is a better option for skewed distributions as it is much more robust and not influenced by outliers. The features were normalized using Box-Cox transformation (Box and Cox, 1964). The normalized features were then passed to a multiclass support vector machine (SVM) classifier to obtain the HIE grade. The performance of the AGS was also tested

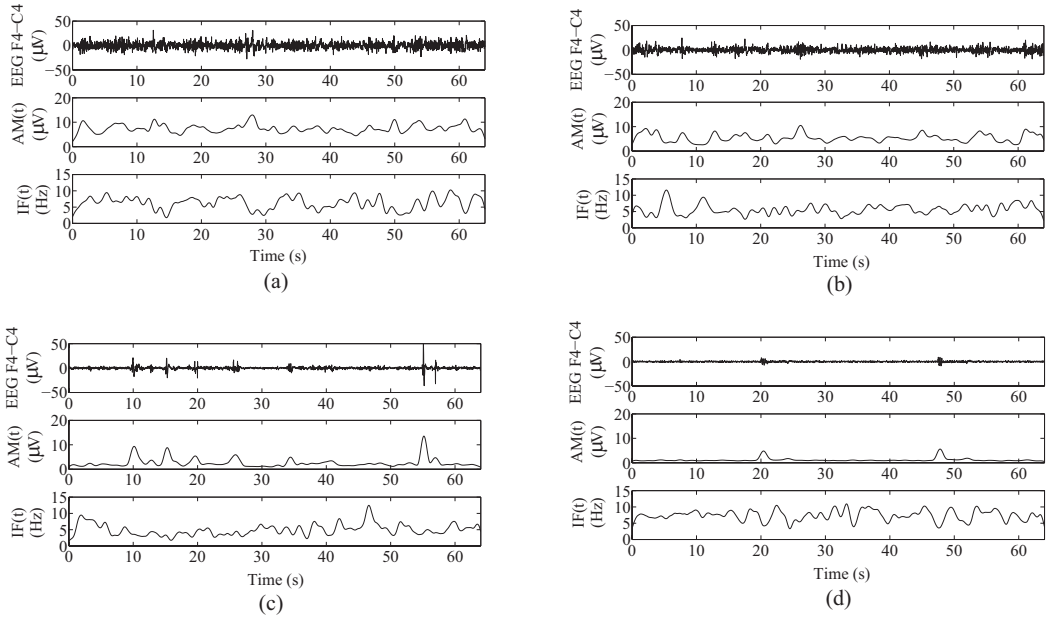


Figure 6.10: Example AM and IF of 64 s ideal EEG epochs for channel F4-C4 of HIE (a) grade 1, (b) grade 2, (c) grade 3 and (d) grade 4 using Hilbert transform method.

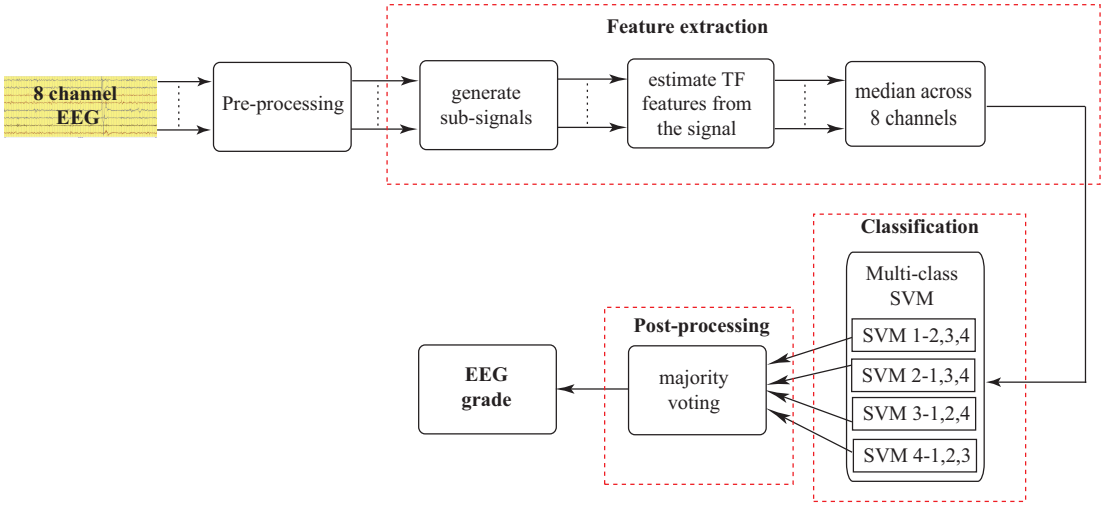


Figure 6.11: Illustration of the proposed AGS using the Hilbert transform and feature fusion method.

for other fusion techniques described in section 6.4.1.

6.6 AGS based on orthogonal matching pursuit

Given a complex dictionary of M atoms, $\mathbf{D} \in \mathbb{C}^{N \times M}$, any EEG signal $\mathbf{X} \in \mathbb{R}^{N \times 1}$ can be represented as

$$\mathbf{X} = 2\Re\{\mathbf{D}\boldsymbol{\gamma}\}, \tag{6.12}$$

where $\mathbf{D}$ is the overcomplete dictionary ($N \leq M$), N is the length (samples) of the signal, M is the number of atoms in the dictionary and $\boldsymbol{\gamma} \in \mathbb{R}^{M \times 1}$ is the set of sparse coefficients selected by the decomposition algorithm. The sparsity of the EEG signal representation can be increased using an appropriate dictionary $\mathbf{D}$ which is highly coherent with the given class of signals and thus minimizing the residual error.

The performance of the AD algorithm depends mainly on the choice of the decomposition dictionary. Several dictionaries have been proposed in the literature that can be used with AD including wavelets (Vera-Candeas et al., 2004), wavelet packets (Yang et al., 2007), Chirplets (Ghofrani et al., 2003), Fourier dictionary (Saito, 1999), Gabor dictionaries (Mallat and Zhang, 1993, Rubinstein et al., 2010) and so on. It is necessary to select a suitable dictionary that can encapsulate the nonstationary characteristics of the HIE-EEG signals. For this purpose, three different dictionaries were used in this section which include:

1. A Gabor dictionary (time-frequency dictionary) consisting of translated, scaled and modulated versions of a Gaussian window (Mallat and Zhang, 1993).
2. A Fourier dictionary (frequency dictionary) consisting of sinusoidal atoms.
3. A Wavelet packet dictionary (time-scaled dictionary) built from a Daubechies 4 quadrature mirror filter, consisting approximately $Nlog2N$ waveforms which is a family of orthonormal wavelet basis (Daubechies, 1988).

The OMP algorithm was applied to 50 one minute relatively artefact free HIE grade 1 EEG epochs using different dictionaries. The HIE-EEG epochs were down-sampled from 256Hz to 64 Hz since EEG activity is negligible in neonates in frequencies higher than 32Hz. The signal-to-error ratio (SER) which is defined as

$$\mathrm{SER}^K = 20\log_{10}\left(\frac{||\mathbf{x}||_2}{||\mathbf{r}_K||_2}\right), \tag{6.13}$$

where $\mathbf{x}$ is the input EEG signal and $\mathbf{r}_K$ is the residual obtained after K iterations of OMP was used to assess the performance of different dictionaries.

Up to 100 atoms were used to decompose grade 1 HIE-EEG epochs using OMP algorithm for different dictionaries. Figure 6.12 shows the mean SER obtained across 50 one minute HIE-EEG epochs using different dictionaries. Since the Fourier dictionary consists of non-local signals or atoms, it requires a large number of atoms for representation. Similarly, as the Daubechies 4 dictionary is a time-scale dictionary and does not span the TF plane completely, it requires more atoms to represent the HIE epoch when compared to the Gabor dictionary. Similar performance was obtained across all HIE-EEG grades (grades 2, 3 and 4) in which the Gabor dictionary outperformed other dictionaries in representation. Using Gabor atoms resulted in the lowest reconstruction error (or higher SER) when compared to the Daubechies 4 or the Fourier dictionary using the same number of atoms. This suggests that the Gabor dictionary is more suited for TF localized signals as it uses finite duration band limited signal or atoms for representation. This would provide flexibility to capture the time and frequency localization of HIE-EEG signals.

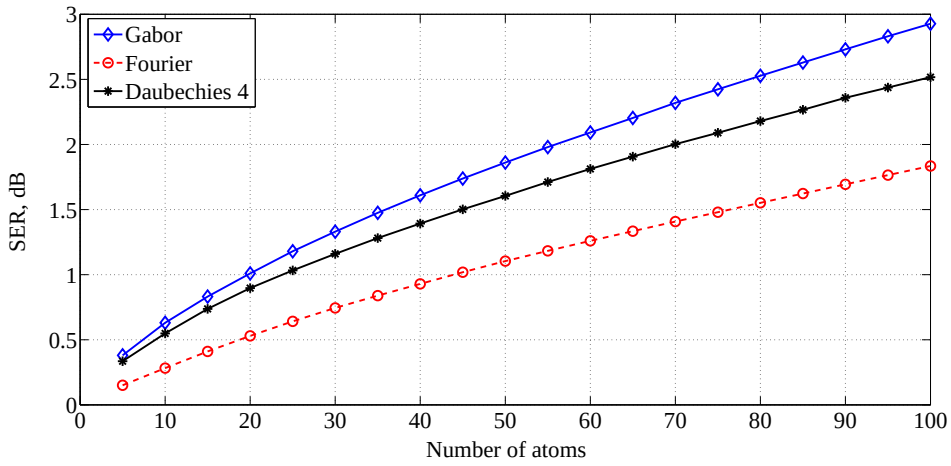


Figure 6.12: Comparison of the mean SER across 50 one minute HIE-EEG grade 1 epochs for increasing number of atoms using Gabor, Fourier and Daubechies 4 wavelet dictionaries. We can clearly see that the Gabor dictionary outperforms other dictionaries in representing HIE-EEG epoch.

Based on the experiment in this section, an overcomplete complex Gabor dictionary was used to extract TF features for classification using AD. The Gabor dictionary offers high time-frequency resolution of the HIE-EEG signal consisting of translated (α), scaled (m) and modulated (β) versions of a Gaussian window,

$$g_G(t;\lambda_G) = \frac{1}{\sqrt{m}} e^{-\pi\left(\frac{t-\alpha}{m}\right)^2} \sin(\beta t). \tag{6.14}$$

Here $\lambda_G = [\alpha, m, \beta]^T \in \mathbf{\Lambda}_G$ are the time-frequency parameters, $T = 64s$ is

the time duration of the EEG epoch, $T_s = 1/f_s$, $f_s = 64$ Hz and $N = 4096$. 7 subdictionaries were constructed to form the overcomplete Gabor dictionary, with atom parameters chosen from the dyadic sequence of integers: $m = 2^q, 0 \leq q \leq L$, $N = 2^L$, $\alpha \in \{2, 4, 8, 16, 32, 64, 128\}$ and $\beta \in \{128, 64, 32, 16, 8, 4, 2\}$. Thus, the dictionary consisted of 28672 Gabor atoms each of length 4096. The analytic associate of each atom was then generated using the Hilbert transform.

The AGS using OMP is shown in figure 6.13. The 8 channel EEG was initially

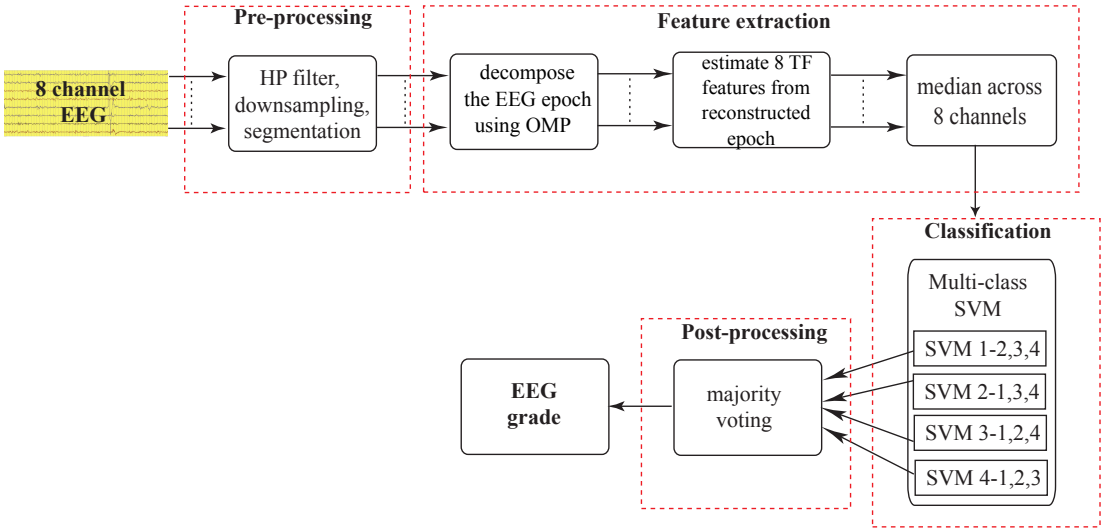


Figure 6.13: The proposed AGS using OMP and feature fusion method.

filtered with a highpass filter with a cutoff frequency f_c and down-sampled to 64 Hz from 256 Hz which was later segmented into 64 s epochs with a 32 s overlap (50% overlap). Each EEG epoch was then decomposed using OMP for a given number of atoms K. Later the TFD of the reconstructed EEG epoch was obtained as

$$\rho_{AD1}[n, p] = \text{TFD}\left\{\sum_{k=1}^{K} \gamma_k \boldsymbol{\phi}_k\right\} \tag{6.15}$$

where K is the number of iterations/atoms used in AD and γ_k is the weight associated with the corresponding atom $\boldsymbol{\phi}_k$ in the dictionary $\mathbf{D}_k$. After decomposition, the AM and IF measures were estimated from ρ_{AD_1} for each channel. This process can also be considered as a feature selection process in which only the significant set of features from a large feature space are considered for classification. The vital information from the signal for discrimination could be obtained from a few high energy basis vectors which is similar to the process where OMP selects these vectors by discarding the largest residual energy. This means that the initial few atoms selected by OMP will contain the most significant information which can be considered as the most significant features. By this way, OMP reconstructs

the essential parts of the signal using the most significant atoms which can then be used to obtain essential information for discrimination.

An example of the AM and IF for a 64 s reconstructed EEG epoch of different HIE grade obtained after OMP using 50 atoms is shown in figure 6.14. When compared to figure 6.10 we can see that the AM and IF components for certain components of the EEG epoch is non-zero which means that only certain components of the signal will be used by the AGS for grading HIE-EEG. For the sake of simplicity, this method is referred as OMP_{method_1}.

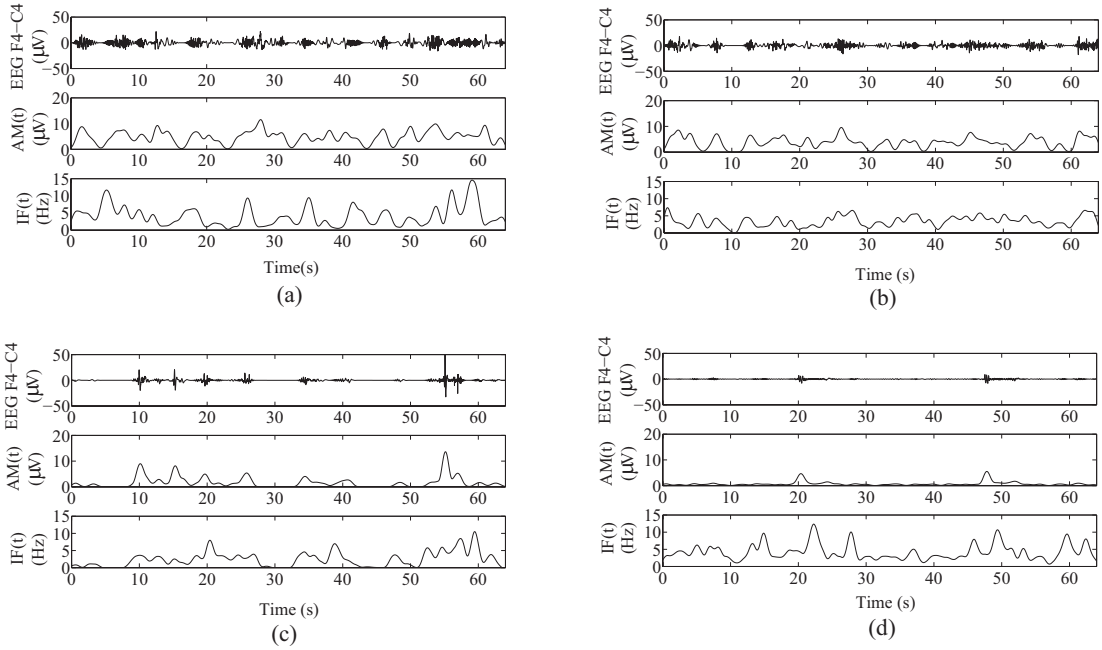


Figure 6.14: Example AM and IF of the reconstructed 64 s epochs using 50 atoms of HIE (a) grade 1, (b) grade 2, (c) grade 3 and (d) grade 4 using OMP_{method_1}.

In another approach, similar to chapter 5, the TFD of the atoms from the Gabor dictionary selected during OMP was obtained as

$$\rho_{AD_2}[n,p] = \sum_{k=1}^{K} \mathrm{TFD}\left\{\boldsymbol{\gamma}_k \boldsymbol{\phi}_k\right\}, \tag{6.16}$$

where K is the number of iterations/atoms used in AD, $\boldsymbol{\gamma}_k$ is the weight associated with the corresponding atom $\boldsymbol{\phi}_k$. The TFD of each atom selected during AD is first obtained which is then combined to form a joint TFD of all the atoms. In this method, the TF features are obtained from the linear combination of the TFD of the atoms as compared to the above method in which the TF features are obtained from the TFD of the reconstructed EEG epoch. This method is referred as OMP_{method_2}.

This method in which the features are extracted directly from the atoms selected during AD is used assuming that the most significant information of the EEG signal can be obtained by synthesizing atoms with maximum energy. Since atoms in the Gabor dictionary completely span the TF plane, by extracting the information from the atoms selected during OMP, the characteristics of the EEG signal could be extracted efficiently. In addition, the TFD obtained from OMP provides a noise-free representation and increased the time-frequency resolution (demonstrated in chapter 3) when compared to the TFD obtained directly from the signal. An example of the AM and IF features obtained using 50 atoms is shown in figure 6.15.

The TFD of the 64 s grade 1 HIE EEG epoch is shown in figure 6.16a, along with the TFD determined using OMP_{method_1} based on 50 atoms in figure 6.16b, and the TFD determined using OMP_{method_2} in figure 6.16c. We can clearly note the difference in TFD's obtained from OMP_{method_1} and OMP_{method_2} from figure 6.16.

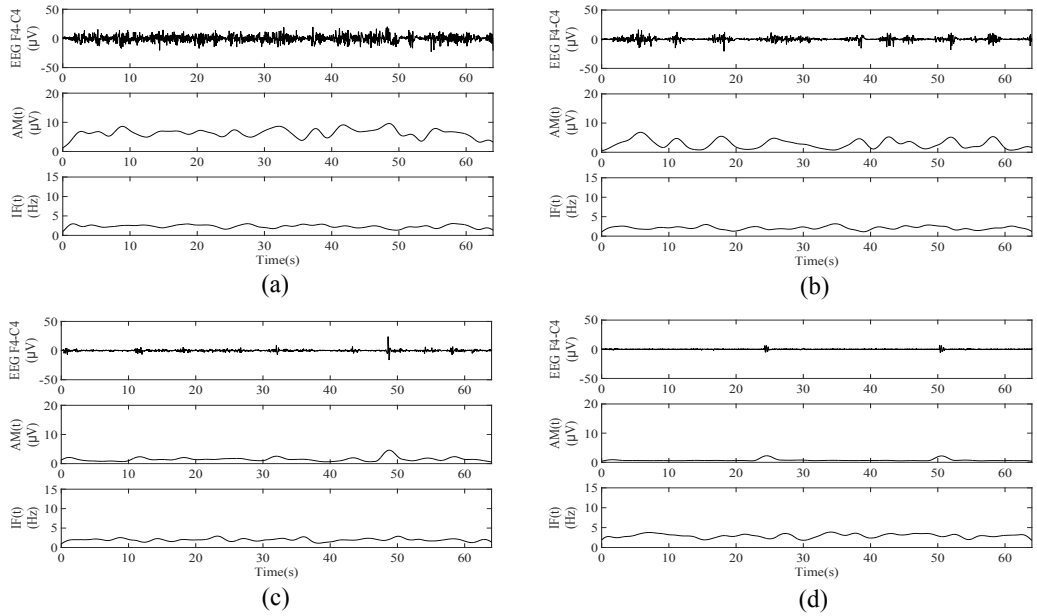


Figure 6.15: Example AM and IF of 64 s ideal EEG epochs for channel F4-C4 of HIE (a) grade 1, (b) grade 2, (c) grade 3 and (d) grade 4 obtained using OMP_{method_2}.

In both these methods, the *median* feature vector across 8-channels (multichannel feature fusion) was used to obtain the performance of AGS for HIE grading as it was found to be optimal (this will be demonstrated in section 6.9).

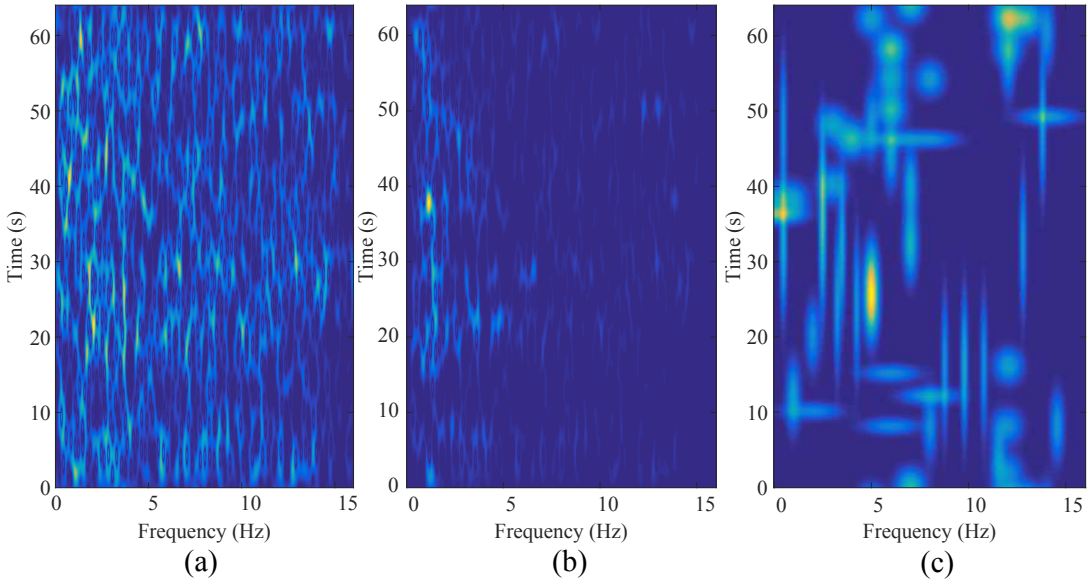


Figure 6.16: Example QTFD's of (a) 64 s ideal EEG epoch from HIE grade 1, (b) reconstructed EEG epoch using OMP and 50 atoms, and (c) 50 atoms selected from the Gabor dictionary during the decomposition of EEG epoch using OMP.

6.7 AGS using multichannel matching pursuit (MMP)

Matching pursuit (MP) is an iterative algorithm that is used to obtain a suboptimal solution to represent a given signal signal using a redundant signal family of atoms known as dictionary (Mallat and Zhang, 1993). It finds linear approximations of signals by iteratively projecting them over a redundant, non-orthogonal set of atoms in the dictionary. The atomic decomposition algorithms: matching pursuit (MP), orthogonal matching pursuit (OMP) and basis pursuit (BP) described in chapter 3 are *single-channel* sparse approximation algorithms in which they try to sparsely model a single channel signal. In certain cases which requires the extraction of certain coherent information from multiple channel observations, simultaneous approximation of a group of signals is necessary by using an identical set of expansion atoms with distinct expansion coefficients.

MP algorithm extension for the simultaneous decomposition of multichannel EEG signal was proposed by (Durka et al., 2005). The main objective of this approach is to identify a set of atoms from the decomposition dictionary which best represents the multichannel input signals. This means that the same atoms are used to model the EEG signal across multiple channels but with distinct expansion coefficients.

The outline of the MP and MMP algorithms are summarized in table 6.4. In short, the MMP algorithm identifies an atom in a dictionary of Gabor atoms that is highly correlated simultaneously across all the channels. The weighted selected Gabor atom is then used to reconstruct the signals across all the channels with the corresponding weighted coefficients. The subsequent steps to approximate the residues follow the standard MP principle. By this way, the MMP algorithm decomposes multichannel data into sum of parameterized MMP atoms.

Table 6.4: Comparison of the algorithm outlines of MP and MMP pursuit. Here $l = 1, 2, \ldots$ denotes the number of channels.

MP	MMP
Initialize the residual $\mathbf{r}_0 = \mathbf{x}$, the index set $\Psi_0 = \emptyset$, the dictionary $\mathbf{D}_0 = \emptyset$ and the iteration count $k = 1$.	Initialize the residual $\mathbf{r}_0^l = \mathbf{x}^l$, the index set $\Psi_0 = \emptyset$, the dictionary $\mathbf{D}_0 = \emptyset$ and the iteration count $k = 1$.
Find $\check{i} = \max_i \lvert \mathbf{r}_{k-1}^T \boldsymbol{\phi}_i \rvert$, $\boldsymbol{\phi}_i \in \mathbf{D}$.	Find $\check{i} = \max_i \lvert \sum_l \mathbf{r}_{k-1}^{l^T} \boldsymbol{\phi}_i \rvert$, $\boldsymbol{\phi}_i \in \mathbf{D}$.
Update the index set $\Psi_k = \Psi_{k-1} \cup \check{i}$ and the active dictionary, $\mathbf{D}_k = [\mathbf{D}_{k-1} : \boldsymbol{\phi}_{\check{i}}] \in \mathbb{R}^{N \times k}$.	Update the index set $\Psi_k = \Psi_{k-1} \cup \check{i}$ and the active dictionary, $\mathbf{D}_k = [\mathbf{D}_{k-1} : \boldsymbol{\phi}_{\check{i}}] \in \mathbb{R}^{N \times k}$.
Estimate $\boldsymbol{\gamma}_k$ by finding maximum inner product $\lvert \langle \mathbf{D}_k, \boldsymbol{\gamma} \rangle \rvert$	Estimate $\boldsymbol{\gamma}_k^l$ by finding maximum inner product $\lvert \langle \mathbf{D}_k, \boldsymbol{\gamma}^l \rangle \rvert$;$\forall\, l$.
Update the residual, $\mathbf{r}_k = \mathbf{r}_{k-1} - \mathbf{D}_k \boldsymbol{\gamma}$.	Update the residual, $\mathbf{r}_k^l = \mathbf{r}_{k-1}^l - \mathbf{D}_k \boldsymbol{\gamma}^l$ for each l.
Update the iteration count, $k = k + 1$.	Update the iteration count, $k = k + 1$.
After K iterations, $\hat{\mathbf{x}} = \mathbf{D}_K \boldsymbol{\gamma}_K$.	After K iterations, $\hat{\mathbf{x}}^l = \mathbf{D}_K \boldsymbol{\gamma}_K^l$; $\forall\, l$.

From Table 6.4, we can see that the MMP preserves the MP algorithm structure. Figure 6.17 demonstrates the working of the MMP algorithm when compared to the OMP algorithm. A sample 8s epoch was used from three channels of grade 1 HIE-EEG for the purpose of demonstration. We can clearly see that the OMP algorithm has better reconstruction accuracy (high SER) when compared to the MMP algorithm (low SER).

MMP implements the single MP for each individual signal across different channels; however, the pursuit search to identify the correlated atoms is not performed separately for individual channels. Instead, the search is performed once to identify the best correlated atoms based on the summation of individual channel metrics across multiple channels. In this manner, the computational cost is reduced significantly since the pursuit is performed only once across all the channels.

Similar to OMP_{method_1}, the TF features were obtained from the reconstructed signal using MMP for each channel. The *median* feature vector across 8-channels

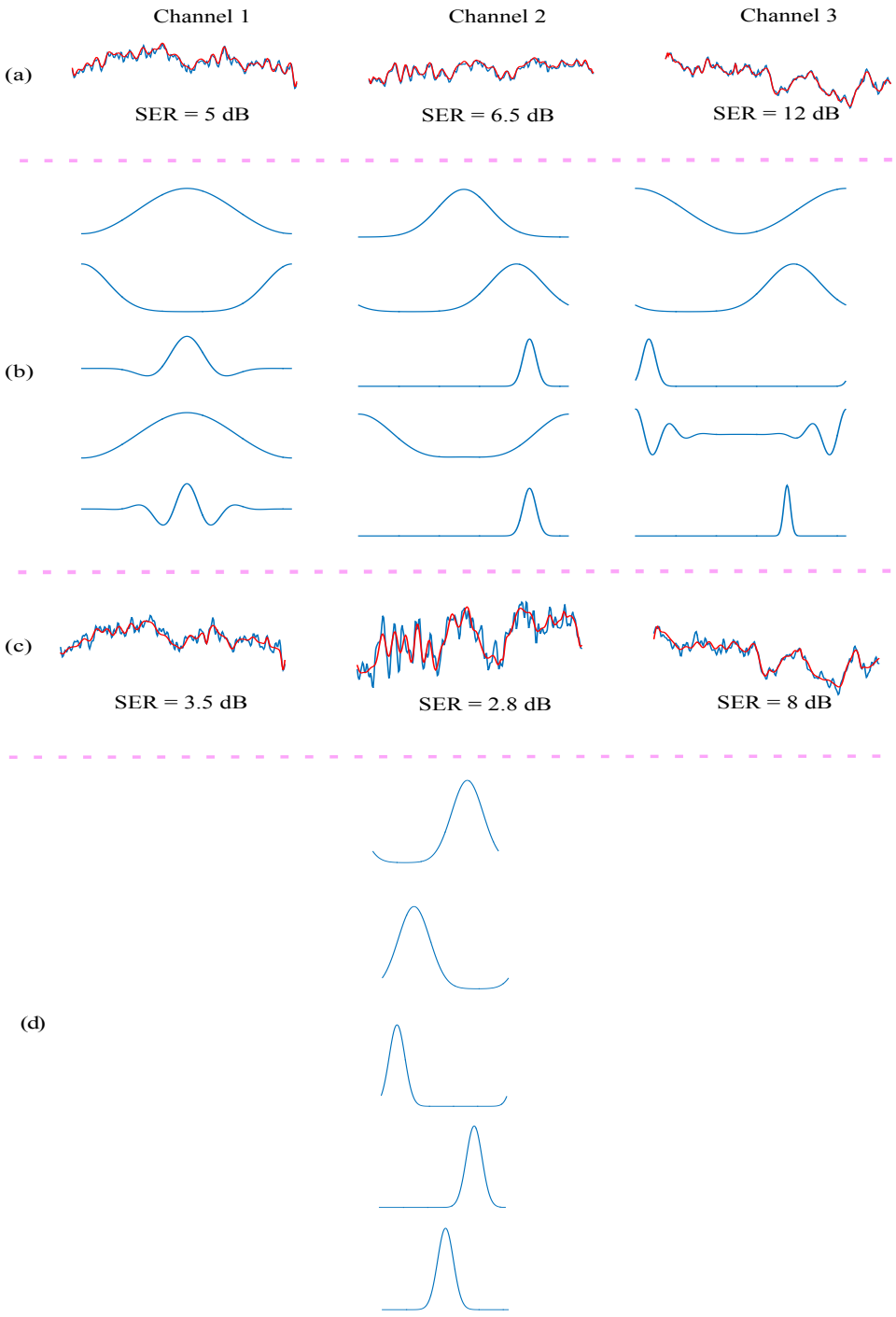


Figure 6.17: Illustration of the decomposition techniques using OMP and MMP using 30 atoms per channel. (a) Sample EEG signal (blue) and its reconstructed version (red) using OMP, (b) the first five Gabor atoms selected during OMP for all three channels, (c) the sample EEG signal (blue) and its reconstructed version (red) using MMP, and (d) the first five Gabor atoms selected during MMP. Note that the SER obtained using OMP is higher when compared to MMP.

(multichannel feature fusion) was used to obtain the performance of the AGS for HIE grading.

6.8 Performance assessment of the AGS

A LOO cross validation was used to assess the performance of the proposed AGS since it provides nearly an unbiased performance estimate of the proposed AGS (Vapnik, 1982). Approximately 530 minutes of data from 53 neonates (10 min of randomly selected data from each neonate) was used to train the SVM model in the LOO routine. Several parameters given in Table 6.5 were greedily searched within the training process using a nested LOO cross-validation on the training set of 53 subjects:

Table 6.5: The pre-processing and feature extraction parameters and their search range.

Parameter	Search range
Pre-processing HPF cutoff frequency, (f_c)	[0.5, 5]Hz
Duration of the Hamming window	[1,4]s
Bandwidth of the Hamming window	[1,4]Hz
Number of atoms, K (for $OMP_{method1}, OMP_{method2}$ and MMP)	[10,200]

The parameters that provided highest classification accuracy over the training set were selected. This trained AGS with the optimal parameters was then tested on the EEG data from the remaining (left out) neonate and the accuracy was obtained. This process was repeated until the recording from each neonate had been used once for testing (total number of iterations here is 54). The overall mean accuracy across 54 iterations is then reported.

6.9 Results of the AGS system

6.9.1 Hilbert transform method

The results of the proposed AGS with multi-class SVM is shown in Table 6.6. Figure 6.18 shows the effect of the HP cutoff frequency (f_c) on the AGS decision. The highest accuracy was obtained for either $fc = 2.5$ or 4 Hz (using a Hamming window of duration = 2s and bandwidth =2Hz). The overall accuracy of the AGS

using SVM and $f_c = 2.5$ Hz was 87% (47/54). Using the same set of TF features extracted from the HIE-EEG signal and a simple multi-class linear discriminant classifier, an accuracy of 77.8% was reported in (Stevenson et al., 2013). This suggests that by using an advanced classifier, the performance of the AGS can be improved. The classification accuracy of the AGS in this study was slightly better when compared to the accuracy obtained in (Ahmed et al., 2014) where an accuracy of 83.3% was reported using the same dataset. However, these results were obtained using 55 generic features when compared to 8 features obtained from the TF analysis in this study.

From Table 6.6, it can be seen that the proposed AGS can classify grade 1 and grade 3 efficiently (>90%) but the performance reduces in the case of grade 2. Only one grade 4 recording was misclassified as grade 3. The performance of each TF feature for the HIE-EEG grading is shown in figure 6.19. The mean and standard deviation of AM and IF (features 1, 2, 5 and 6) provided highest discriminatory information when compared to other features.

Table 6.6: Confusion matrix of the AGS output using Hilbert transform method (with feature fusion shown in figure 6.11) and the actual HIE grade assigned by the EEGer.

Actual HIE grade	AGS output					
	1	2	3	4	Total	Misclassified
1	22	0	0	0	22	0
2	4	9	1	0	14	5
3	0	1	11	0	12	1
4	0	0	1	5	6	1
Total	26	10	13	5		
Accuracy(%)	100	64	78.5	83.3		

The performance of the AGS using several fusing methods described in section 6.4 was compared. In this study, the multichannel feature fusion was achieved by taking the median of the EEG feature vectors across multiple channels. For the multichannel classifier fusion (shown in figure 6.20a), the majority voting was performed across the channels to assign a grade for each epoch. Later, the final HIE grade was assigned to the EEG recording after the majority voting of the classifier fusion output. Similarly, in the case of multichannel decision fusion (shown in figure 6.20b), two stage majority voting was performed. In the first stage, the initial HIE grades were obtained by the majority voting of the individual classifier output. This gives different/similar grades for each channel.

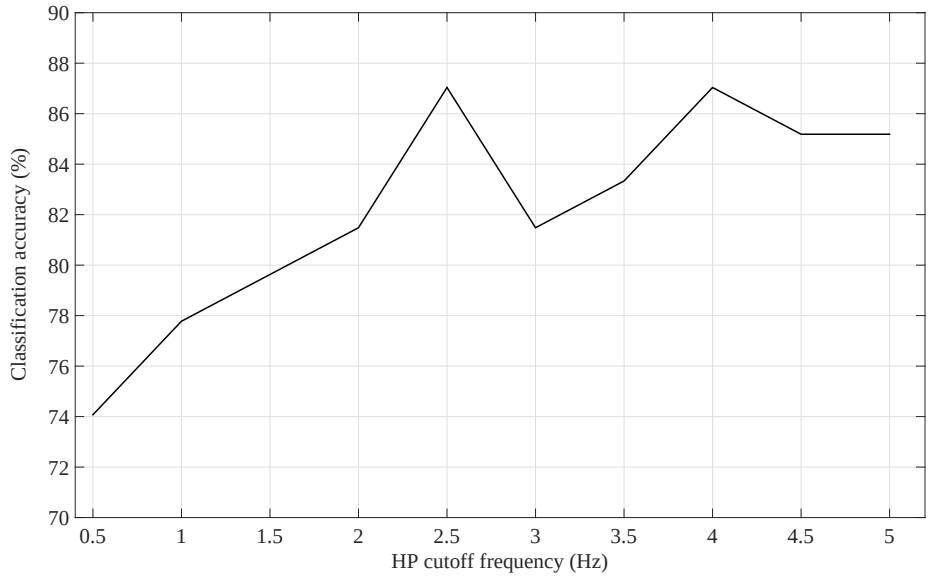


Figure 6.18: Classification accuracy of the AGS using Hilbert transform method (with feature fusion) for variable HP cutoff frequency.

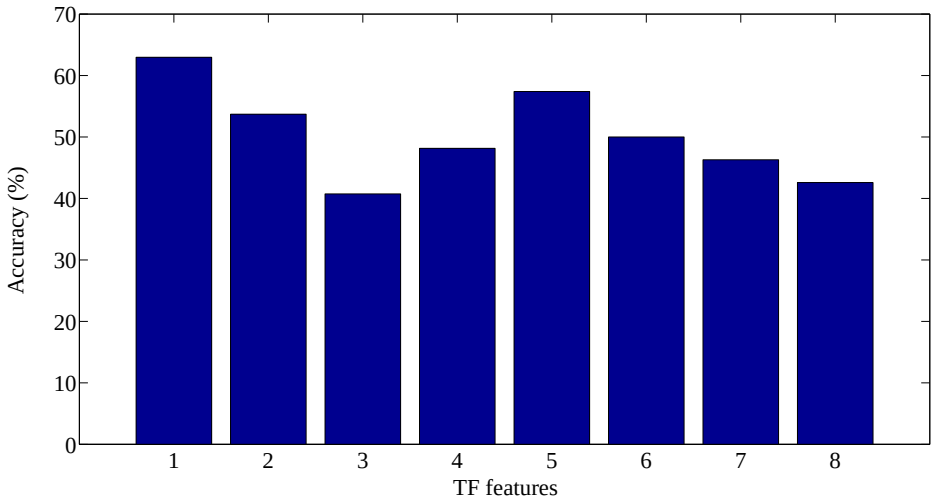


Figure 6.19: Performance of the individual TF features for discriminating HIE-EEG grades using Hilbert transform based AGS (with feature fusion).

Later, in the second stage, the majority voting of the initial HIE grades was obtained to get the final HIE grade. The results are summarized in Table 6.7.

From Table 6.7 we can see that the AGS based on multichannel feature fusion outperforms other fusion methods. In the multichannel classifier and decision fusion approaches, the outputs of each multichannel classifiers were combined to obtain the final HIE grade. Both these approaches do not take into account the simultaneous recording nature of the EEG channels and assume that the

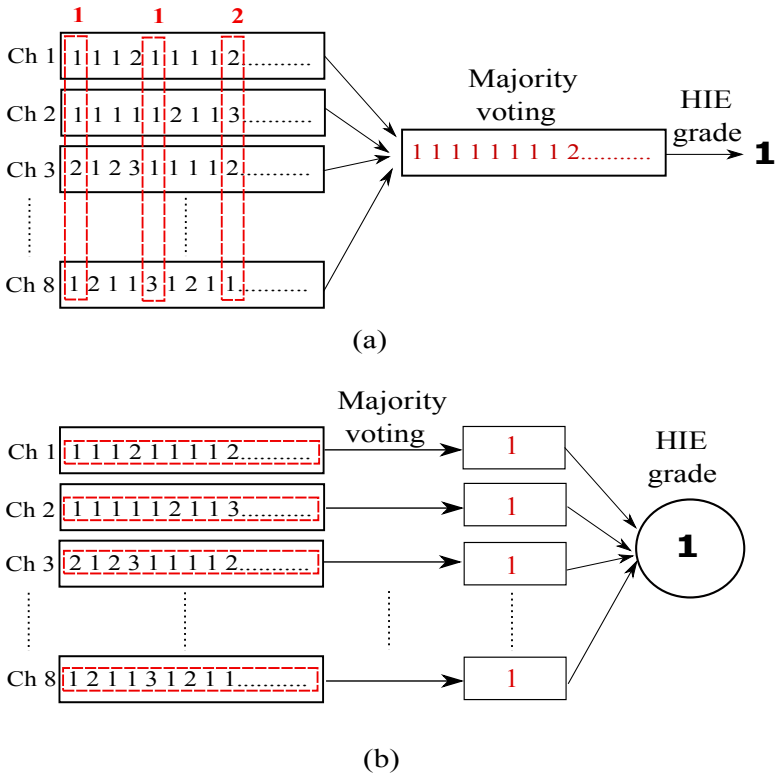


Figure 6.20: Illustration of the (a) multichannel classifier fusion, and (b) multichannel decision fusion techniques for assigning an overall HIE grade.

Table 6.7: Comparison of the Hilbert transform based AGS classification accuracy using different fusion techniques.

Fusion type	Accuracy (%)
Multichannel feature fusion	87%
Multichannel classifier fusion	74%
Multichannel decision fusion	72%

EEG channels are statistically independent from each other. As a result, the classification accuracy of the AGS reduced. However, the multichannel feature fusion approach was primarily developed based on the assumption of global injury nature of the HIE and the existence of the inter-dependence between some of the recorded EEG channels. By taking the median across the channels, the simultaneous recording nature of the EEG channels was taken into account which resulted in increased classification accuracy.

In total, the EEG recordings of 47 neonates were correctly classified by the AGS and 7 recordings were misclassified. In 4 misclassified recordings, the majority of the segments in the recording were grade 1 but were annotated as grade 2 by the neurophysiologists due to the presence of asymmetry, asynchrony, runs of sharp waves. Since the proposed AGS uses majority voting over one hour during

the post-processing stage for decision making, these abnormal patterns were not detected by the AGS and the recordings were misclassified. The presence of these abnormal patterns resulted in the downgrading of the EEG which are usually present in grade 2 and grade 3 EEG recordings. The remaining misclassifications were caused due to the presence of long periods of artefacts incorrectly graded by the AGS. In these recordings, the artefact free periods were correctly classified.

Figure 6.21 shows the distribution of certain and uncertain decisions made by the AGS. The post-processing stage was modified which required at least two thirds majority votes for a certain decision. Anything below this level was considered as uncertain decision. In total, 36 neonates were correctly classified with certainty, 11 neonates were correctly classified with uncertainty, 7 neonates were misclassified with uncertainty. It was interesting to note that there were no certain misclassifications. The distribution of decisions made by the proposed AGS for individual grades is shown in figure 6.22.

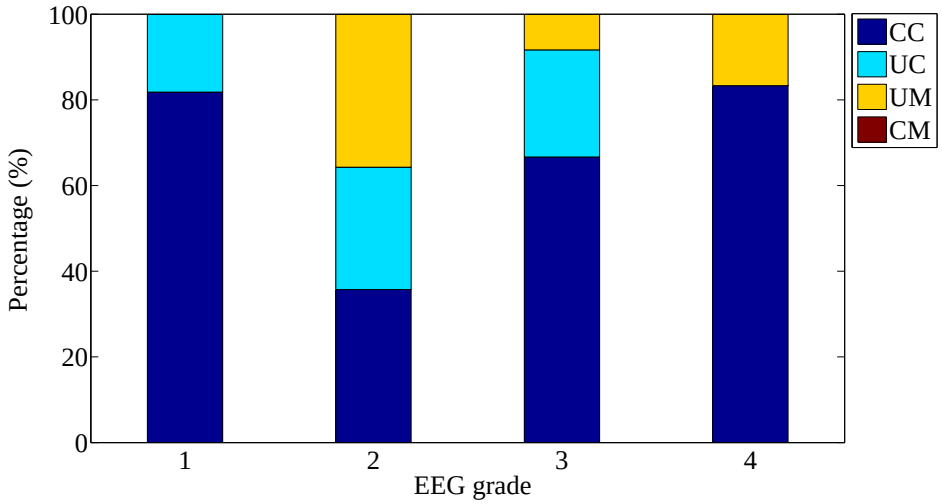


Figure 6.21: The distribution of the certain and uncertain AGS decisions using Hilbert transform method. CC = certainty in correct classification, UC = uncertainity in correct classification, CM = certainty in misclassification and UM = uncertainty in misclassification.

All grade 1 recordings were correctly classified. It can be seen that grade 2 was most commonly misclassified as grade 1 due to the similar morphology of the EEG signal present in both the grades (see figure 6.10). Even though some of the sequences were graded as grade 3 and grade 4, the final grading was not affected. Similarly, grade 2 sequences had a major influence on grade 3 decisions. As expected, grade 4 had no major influence on other grades since there are no major activities in grade 4 EEG. Most of the misclassifications occurred between

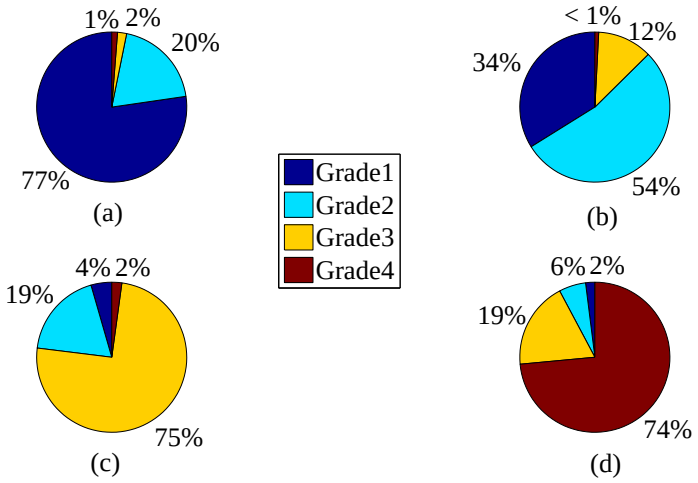


Figure 6.22: The mean distribution of the epochs classified by the proposed AGS using Hilbert transform method for (a) grade 1, (b) grade 2, (c) grade 3, and (d) grade 4.

grades 1 and 2.

6.9.2 Orthogonal matching pursuit methods

The classification accuracy obtained using AD methods is given in table 6.8. The best performance was obtained using TF features from the TFD of the reconstructed epoch when compared to the TF features obtained directly from the TFD of atoms. The overall accuracy of the OMP based AGS system was 83.3% (45/54) which was less than that using the Hilbert transform method (87%). However, from table 6.8, it can be concluded that by using OMP, one can reconstruct the important components of a given signal for the purpose of discrimination.

Table 6.8: Classification accuracy obtained using AD methods

AD method	No. of atoms	Accuracy
OMP_{method_1}	150	83.3%
OMP_{method_2}	30	74%

Some of the key findings using OMP methods are as follows:

1. The OMP selected different atoms for different HIE grades which suggests that different HIE grades exhibit different characteristics and each decomposition set contains the essential information within each EEG signal. However, between HIE grade 1 and grade 2 signals, the decomposition was

similar and the same atoms were selected in most of the epochs. Due to this, the performance of AGS was poor and the features obtained directly from the atoms of the Gabor dictionary (OMP_{method_2}) were not discriminatory enough to distinguish between HIE grades. This is due to the fact that the atoms in the Gabor dictionary are not trained for any particular HIE grades which means that the atoms in the dictionary are not coherent with any particular HIE-EEG structure. As a result, the type of features proposed in chapters 4 and 5 cannot be used for the classification of HIE grades.

2. The main aim of using OMP is to extract several discriminatory features for HIE classification without any intention to recover or compress the actual EEG signal. OMP provides sparse representation of the signal and reduces the residual energy of the EEG signal with few iterations or atoms. The main assumption using OMP for feature extraction is that the most important characteristics of the EEG signal are provided by the first few atoms with the highest energy selected by the OMP which provides a simple representation of the underlying EEG signal structure. OMP reconstructs the given EEG signal using the most useful atoms in just a few iterations. By extracting several features from the linear combination of atoms (or reconstructed signal) in OMP_{method_1}, it was possible to improve the performance of the AGS when compared to OMP_{method_2}.

The results of the AGS using OMP_{method_1} is shown in Table 6.9. The overall accuracy of the proposed AGS system was 83.3% (45/54). Table 6.9 also provides the accuracy obtained for individual grades. It can be seen that the proposed AGS can classify grade 1 and grade 4 efficiently (>90%) but the performance reduces in the case of grade 2 and grade 3 with grade 3 being the lowest among others. The performance of each TF feature for the HIE-EEG grading is shown in figure 6.23. The mean and standard deviation of AM (features 1 and 2) provided highest discriminatory information when compared to other features.

Figure 6.24 shows the distribution of certain and uncertain decisions made by the AGS. A modification to the post-processing stage was performed which required at least two thirds majority votes for a certain decision. Anything below this level was considered as an uncertain decision. In total, 35 neonates were correctly classified with certainty, 10 neonates were correctly classified with uncertainty, 7 neonates were misclassified with uncertainty. It was interesting to note that only 2 neonates were certainly misclassified. The distribution of the decisions made

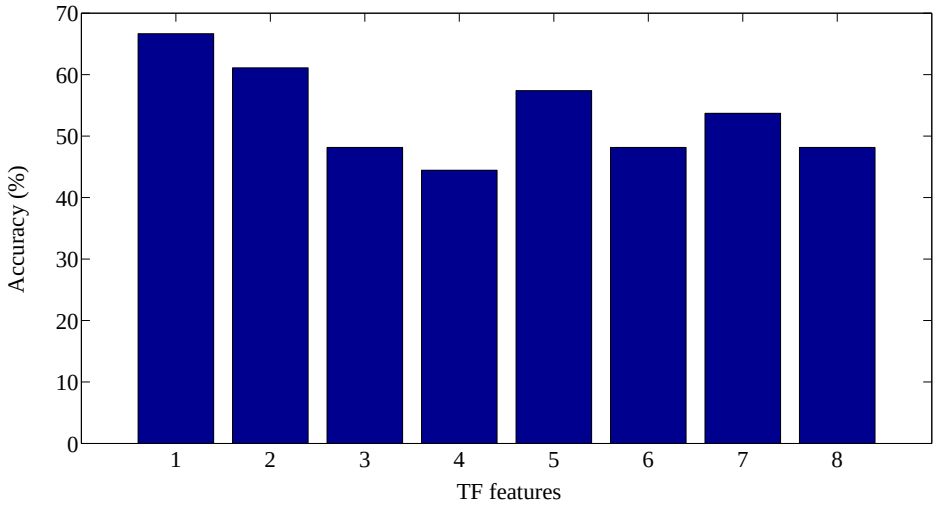


Figure 6.23: Performance of individual TF features for discriminating HIE-EEG grades using OMP_{method_1}.

Table 6.9: Confusion matrix of the AGS output and the actual HIE grade assigned by the EEGer using OMP_{method_1}.

Actual HIE grade	AGS output					
	1	2	3	4	Total	Misclassified
1	21	1	0	0	22	1
2	3	10	1	0	14	4
3	1	3	8	0	12	4
4	0	0	0	6	6	0
Total	25	14	9	6		
Accuracy(%)	95	71	66.6	100		

by the proposed AGS for individual grades is shown in figure 6.25. It can be seen that grade 2 is most commonly misclassified as grade 1. Similarly, grade 2 sequences had influence on grade 3 decisions. As expected, grade 4 had no major influence on other grades since there are no major activities in grade 4 EEG. Most of the misclassifications occurred between grades 2 and 3.

In order to demonstrate the working of the proposed AGS, a one hour long grade 1 EEG recording was graded using the proposed AGS and is shown in figure 6.26. Figures 6.26(a-d) shows the probability outputs of the individual grades obtained from all four SVM models. We can clearly see that the probability output of SVM model for grade 1 is greater than 0.5 for most of the epochs. Figure 6.26(e) shows the initial grading by taking the maximum of the overall probability outputs. The final grade is assigned to the EEG by the majority voting which is shown in figure 6.26(f).

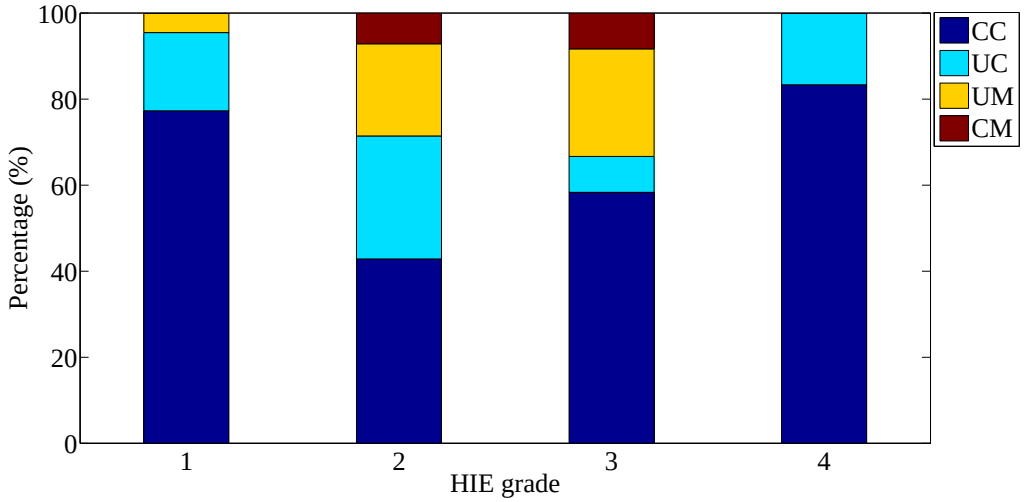


Figure 6.24: The distribution of the certain and uncertain AGS decisions using OMP_{method_1}. CC = certainty in correct classification, UC = uncertainity in correct classification, CM = certainty in misclassification and UM = uncertainty in misclassification.

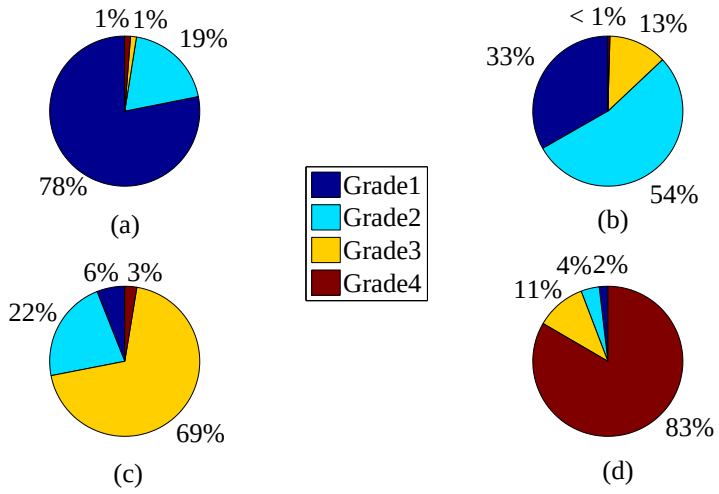


Figure 6.25: The mean distribution of the epochs classified by the proposed AGS using OMP_{method_1} for (a) grade 1, (b) grade 2, (c) grade 3, and (d) grade 4.

Nine recordings were misclassified in total and these recordings were further analysed. It was observed that the presence of abnormalities such as asymmetry, asynchrony and runs of sharp waves resulted in the misclassification of five recordings by the AGS. An example of the presence of a sharp wave in grade 2 EEG is shown in figure 6.27. Examples for the presence of asymmetry and asynchrony are shown in figures 6.28 and 6.29 respectively. Due to the presence of these abnormalities in the EEG recording, the AGS decision was downgraded. This means that if the original EEG grade assigned by the EEGer was grade 2, the AGS decision

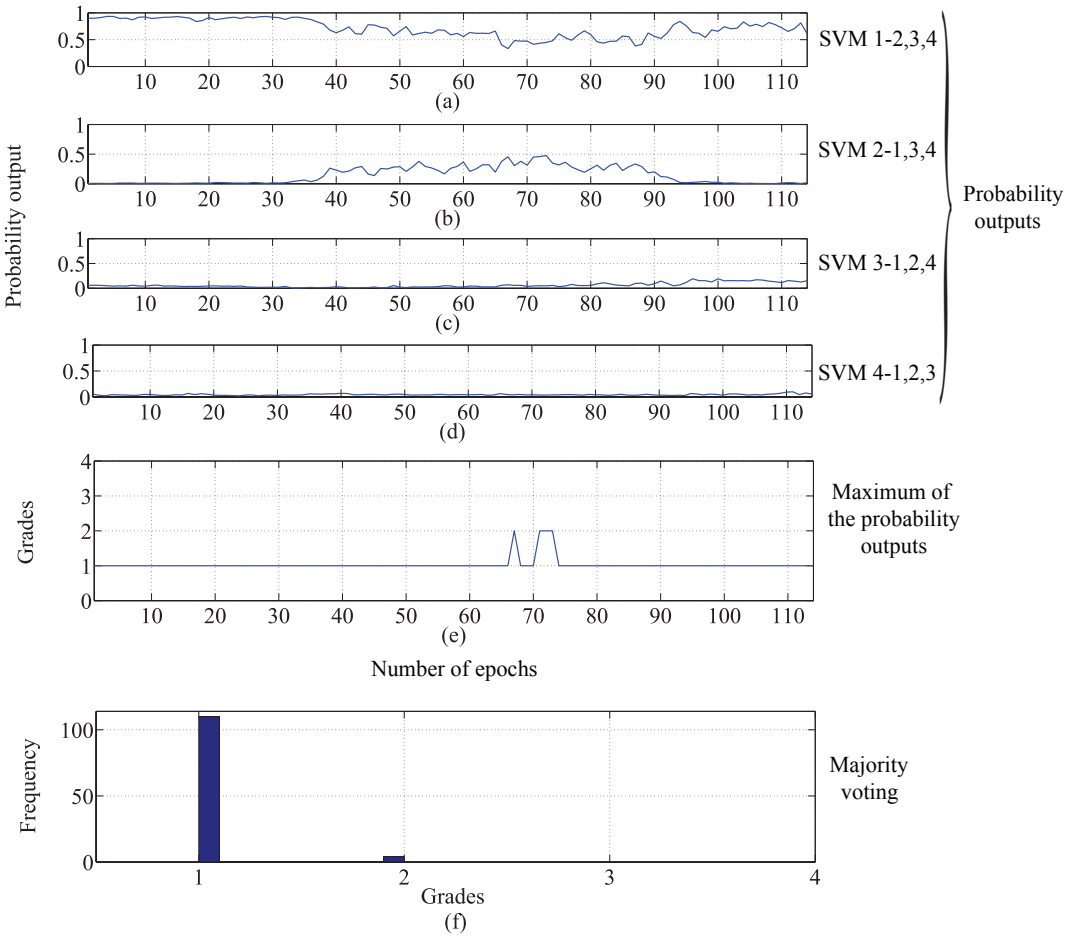


Figure 6.26: Illustration of the probabilistic method of assigning HIE grade by the AGS for a complete 1 hour grade 1 EEG recording.

was grade 1. It is difficult to determine the presence of these abnormalities as they appear intermittently. The incorporation of the methods to detect these abnormalities has a potential to improve the performance of the proposed AGS.

The remaining misclassifications were caused by the presence of long periods of EEG artefacts, however the artefact-free periods were correctly classified by the AGS. Overall, the influence of artefact resulted in an incorrect decision of 33% of the misclassified EEG recordings in which the majority was due to the presence of several artefacts $\geq$ 16s in duration (29%). In three misclassified EEG recordings, more than 50% of the recording were contaminated with artefacts which influenced the decision during the majority voting. An example of a 16s duration artefact is shown in figure 6.30. An example of the AGS output for a misclassified recording is shown in figure 6.31.

It can be seen from figure 6.31 that the presence of artefacts has a great influence on the AGS decision using $OMP_{method1}$. Even though the presence of these arte-

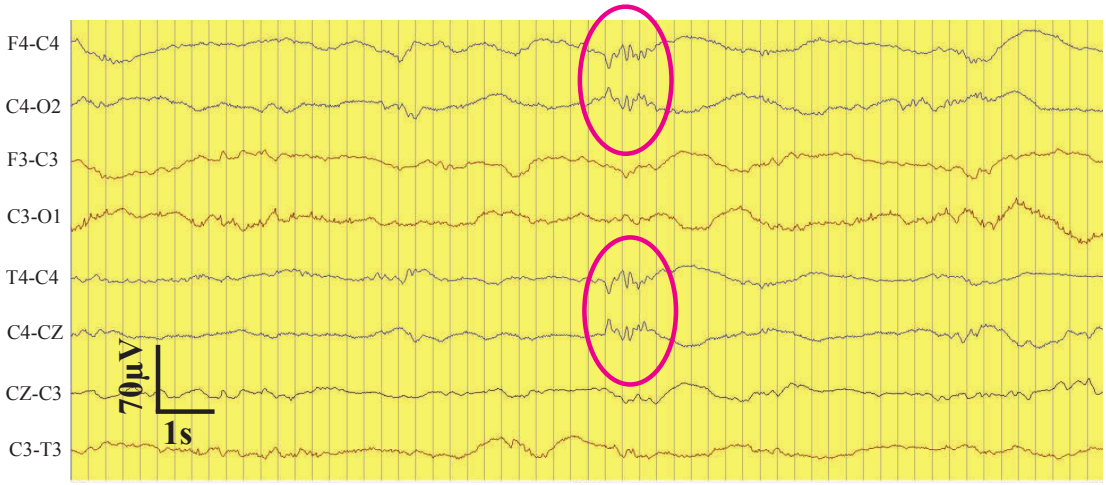


Figure 6.27: Example of sharp wave in a 64 s HIE grade 2.

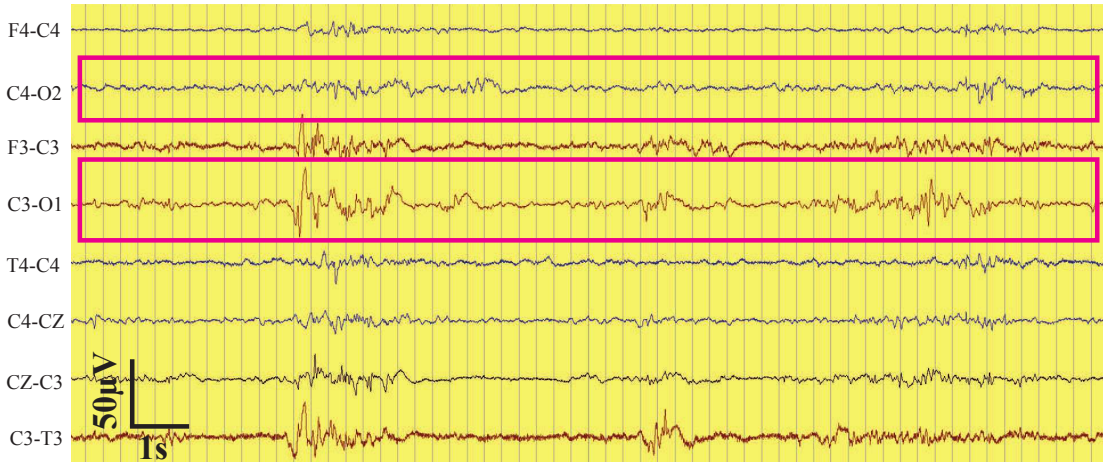


Figure 6.28: Example of an asymmetry event in a 64 s HIE grade 2 EEG.

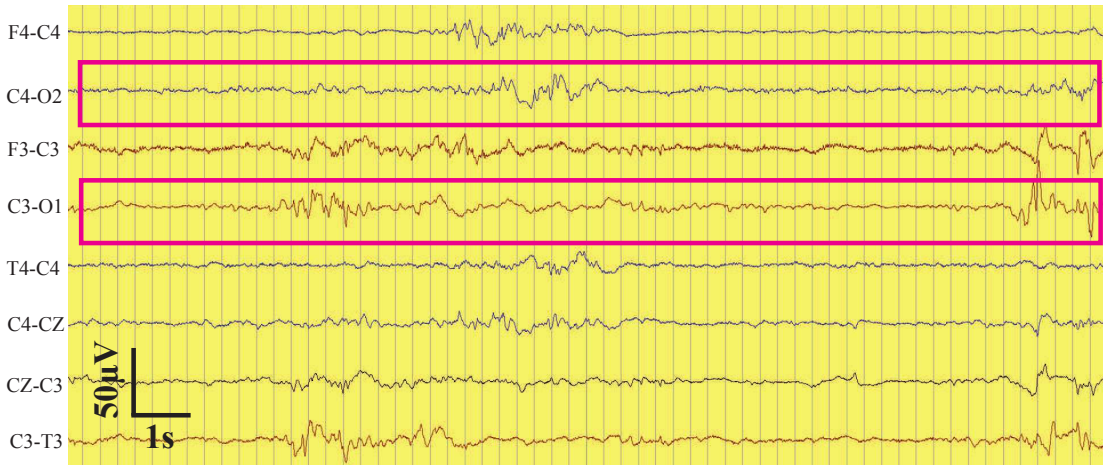


Figure 6.29: Example of an asynchrony event in a 64 s HIE grade 2 EEG.

facts had no influence on the decision of the AGS on most of the EEG recordings due to post-processing and majority voting in the Hilbert transform method, the

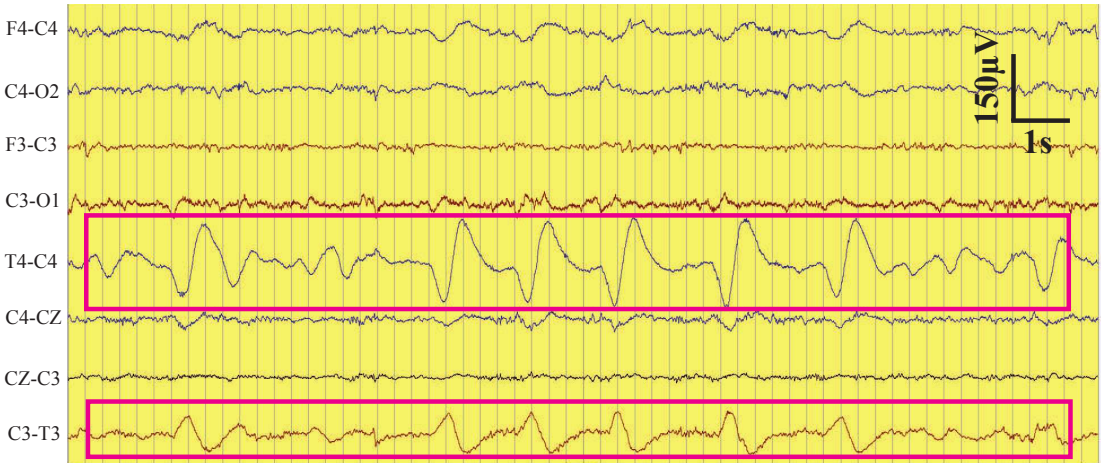


Figure 6.30: Example of a long duration artefact in channels T4-C4, C3-T3.

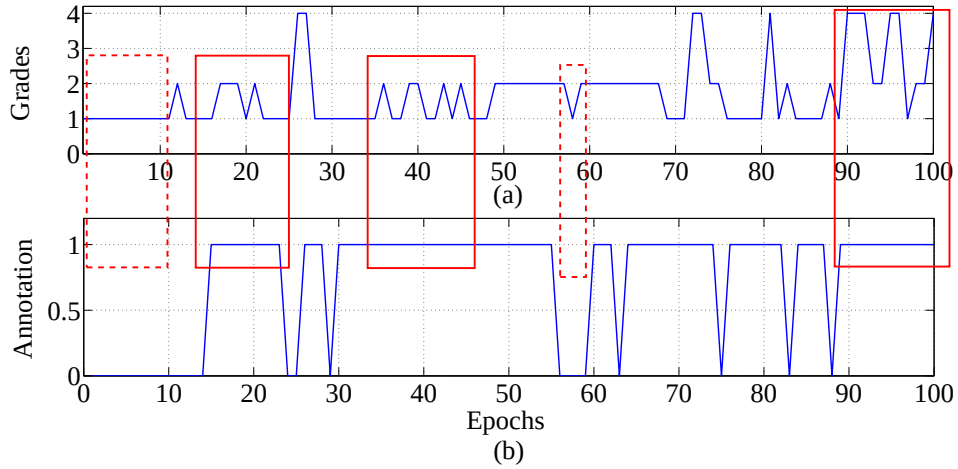


Figure 6.31: Illustration of the effect of artefacts on the OMP_{method_1} based AGS decisions. (a) output of the AGS, (b) Artefact annotation where '1' indicates the presence of artefact. The red solid box highlights the presence of artefacts and red-dotted box highlights sample artefact free epochs.

incorporation of an artefact detector to detect these long duration EEG artefacts could improve the performance of the OMP based AGS by removing the artefacts from the misclassified data.

6.9.3 MMP method

Figure 6.32 shows the accuracy obtained using MMP for increasing number of atoms. The confusion matrix of the AGS output using MMP method and the actual HIE grade is given in Table 6.10. The highest accuracy obtained using MMP was 76% using approximately 100 atoms which was less that the accuracy obtained using $OMP_{method1}$ (83.3%). The possible reasons for decreased accuracy

using MMP can be due to the presence of active sources from several channels at the same time. This means that for a given duration, source waveforms are assumed to have similar frequency and phase. Increase in number of sources may result in lower stability of inverse solutions (Achim, 1995) and could lead to an unrealistic approximation of the EEG signal (Nunez and Srinivasan, 2006). This was also observed in (Lelic et al., 2011) where it was demonstrated that the performance of the MMP algorithm was sensitive to the number of EEG channels used for simultaneous decomposition.

Table 6.10: Confusion matrix of the AGS output using MMP method (with feature fusion) and the actual HIE grade assigned by the EEGer.

Actual HIE grade	AGS output					
	1	2	3	4	Total	Misclassified
1	19	3	0	0	22	3
2	3	8	3	0	14	6
3	1	2	9	0	12	3
4	0	0	0	6	6	0
Total	23	13	12	6		
Accuracy(%)	86	57	75	100		

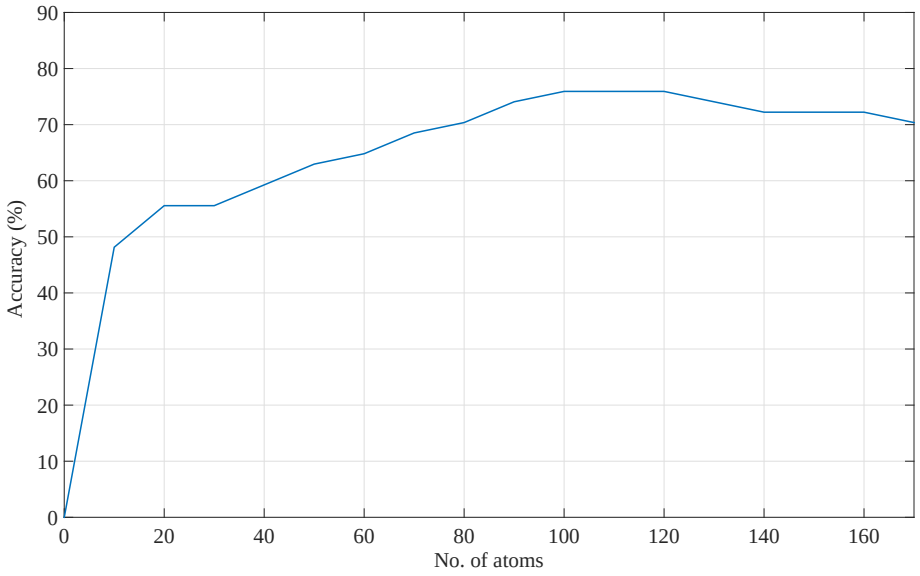


Figure 6.32: MMP accuracy versus variable number of atoms.

6.10 Artefact detection system using orthogonal matching pursuit

From the analysis of the results using OMP in section 6.9.2, it was observed that the performance of the AGS using OMP was sensitive to artefacts. In this section, a system for artefact detection in neonatal EEG using AD is proposed that could improve the performance of the AGS for HIE grading. Several types of artefacts are present in the neonatal EEG which includes body movement, respiration, eye blinking, etc. The normal background EEG is usually free from burst or seizure patterns. Artefacts that are about 10-100 times larger than the normal background EEG can be easily distinguished visually from the background EEG. In some cases, high energy artefacts mimic burst patterns which makes it difficult to distinguish from bursts. In addition, these artefacts are nonstationary which includes variations in the amplitude and shape (or morphology) over time and across channels. So a method for automatic EEG pattern detection should first properly detect and discard the artefacts from further consideration. AD using OMP is an ideal choice for this purpose as it decomposes signals into parameterized atoms. Based on the parameters of the atoms designed in the dictionary, the artefacts from clean EEG signal and bursts can be separated.

The structure of the proposed artefact detection system (ADS) is shown in figure 6.33 which is similar to the neonatal seizure detection system illustrated in chapter 5. The input signal is initially passed through a bandpass filter [1 70]Hz with a transition width of 0.5Hz. The signal was then downsampled from 256 Hz to 64 Hz and then segmented into 8s epochs with 4s overlap. Each epoch was then decomposed using OMP using a set number of atoms K. Several features were then extracted from the reconstructed EEG epoch after AD. The maximum of the SVM across all channels was obtained and sharp transients in the SVM output were then suppressed using a median filter of 4s in duration. The SVM decision threshold was varied and the artefact was detected if it exceeded a set threshold. A collar of 2s was applied that extends the initial decision 2s forward and backward in time. An automated annotation of artefacts is then obtained. Different types of dictionary can be used for OMP. In this study, a Gabor dictionary was used (see section 6.14) since Gabor atoms offer optimal time-frequency localization. With a highly redundant Gabor dictionary, it is possible to reconstruct an approximation of artefact using relatively a small set of parameterized atoms.

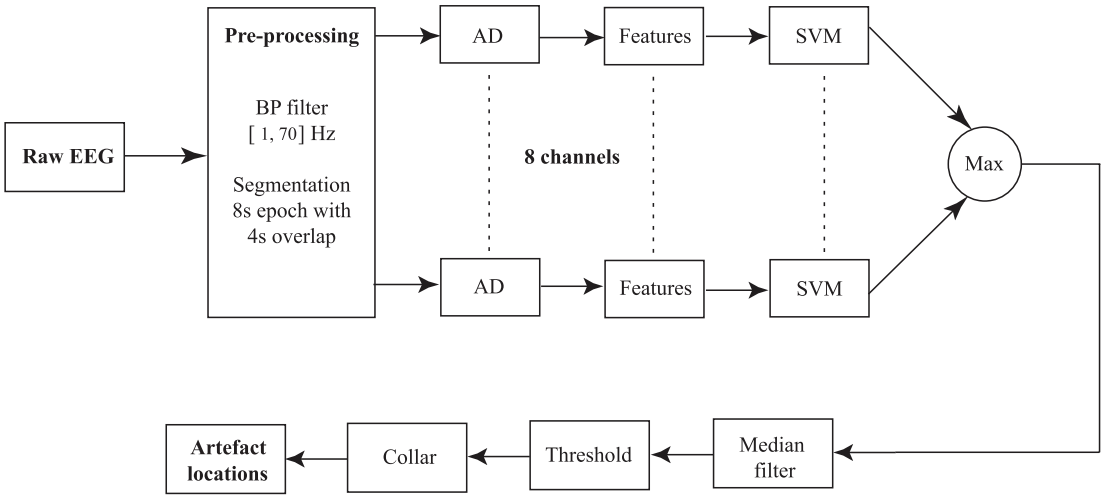


Figure 6.33: Overview of the proposed atomic decomposition based artefact detection system.

6.10.1 Performance assessment of the ADS

The dataset described in section 6.2 was used to test the performance of the proposed ADS. Three recordings were not used since they did not have any artefacts $\geq$ 16s. The parameter α which is the scale of the Gabor atom was varied in the range $[2, 4, 8, 16, 32, 64, 128]$. The decomposition level for OMP was varied from 1-10. A LOO cross-validation method was used in this work to assess the performance of the proposed ADS for patient independent artefact detection. The parameters of the dictionaries (scale of the Gabor atom and decomposition level) were then selected to provide the optimal performance over the training set. The training set consisted of features from randomly selected artefacts and clean EEG from 50 neonates (2 minutes of artefacts and 10 minutes of clean EEG from each neonate). This trained algorithm was then tested on the full recording of the left out neonate. This process was repeated (total number of iterations here is 51) until the data from each neonate had been used once for testing. The AUC was used as an epoch based performance metric for the training phase and for each unseen test neonate. The mean and median AUC were then obtained.

6.10.2 Feature extraction for artefact detection

Several amplitude and frequency related features were calculated from the reconstructed EEG epoch which are given below:

1. *Relative structural complexity (RSC)*: As mentioned in chapter 4, the RSC

at a given level of decomposition, K, is given by

$$F_1 = RSC = \text{SER}^K = 20\log_{10}\left(\frac{||\mathbf{x}||_2}{||\mathbf{r}_K||_2}\right), \tag{6.17}$$

where $\mathbf{x}$ is the input EEG epoch, SER is the signal-to-error ratio and $\mathbf{r} = \mathbf{x} - \tilde{\mathbf{x}}$ is the residual obtained after K iterations of AD. An example illustrating the behaviour of the RSC corresponding to artefacts and clean EEG is shown in figure 6.34 using the Gabor dictionary with $\alpha = 8$ and a decomposition level of 6 atoms (this provided optimal AUC which will be discussed in section 6.10.3). It can be seen that the RSC increases in the presence of artefacts since the atoms selected during AD are coherent with artefacts, thereby decreasing the reconstruction error.

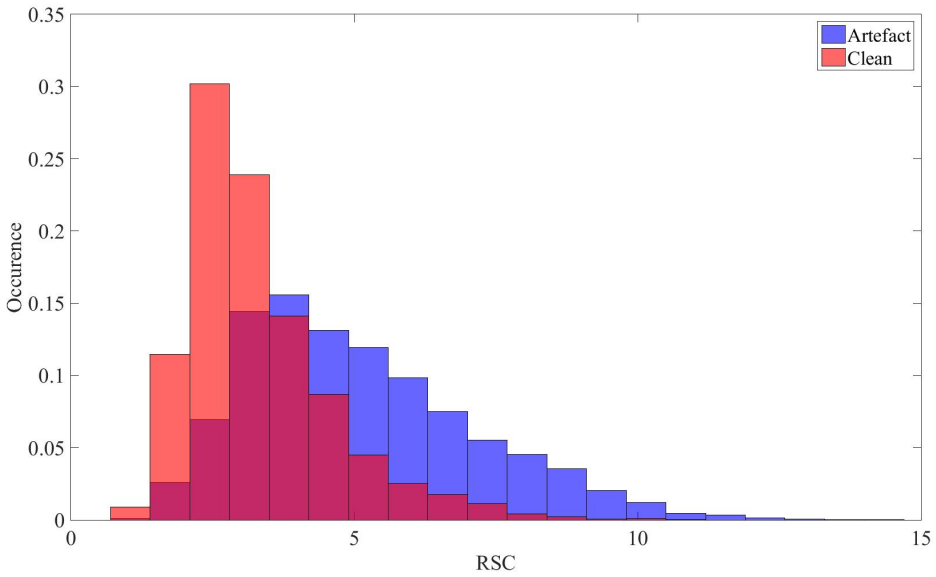


Figure 6.34: Class specific histogram for RSC.

2. *Root mean square (RMS) amplitude*: The RMS of the reconstructed epoch $\tilde{\mathbf{x}}$ can be obtained as

$$F_2 = RMS(\tilde{\mathbf{x}}) = \sqrt{\frac{1}{N}\tilde{\mathbf{x}}^T\tilde{\mathbf{x}}}. \tag{6.18}$$

The distributions of the classes over the sample epochs are plotted in figure 6.35. It can be confirmed from this plot that there is an increase in RMS amplitude for many artefacts. However, since there is overlap between the two classes, it indicates the presence of low amplitude artefacts as well.

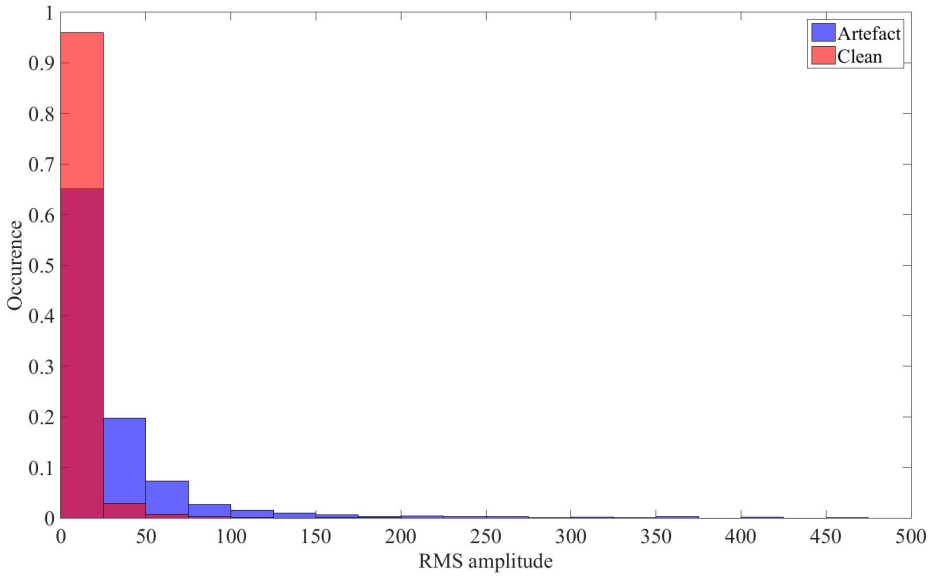


Figure 6.35: Class specific histogram for RMS amplitude.

3. *Nonlinear energy*: Nonlinear energy (NLE) was initially used for seizure prediction in epileptic patients by (D'Alessandro et al., 2003). The nonlinear energy is a function of both the amplitude of the signal, as well as the variation in the amplitude of the epoch. This feature for reconstructed epoch $\tilde{\mathbf{x}}$ is calculated as:

$$F_3 = NLE(\tilde{\mathbf{x}}) = \frac{1}{N-2}\sum_{j=2}^{N-1}(\tilde{\mathbf{x}}(j))^2\tilde{\mathbf{x}}(j-1)\tilde{\mathbf{x}}(j+1) \tag{6.19}$$

The distribution of classes for nonlinear energy is shown in figure 6.36. Similarly to RMS amplitude, artefacts show higher energy values for a number of epochs when compared to clean epochs.

4. Dominant/peak frequency: Several explicit (using the fast fourier transform) and implicit (periodicity of the EEG signal) properties are utilised by the neurophysiologists to examine EEG. Therefore it is important to include frequency domain features from the EEG to quantify changes in the spectrum of the EEG during the presence of artefact. By using a 256 point fast Fourier transform (FFT), the power spectral density (PSD) of the EEG epoch was obtained. The absolute value of the n_s complex coefficients from the FFT were extracted as the frequency coefficients $\mathbf{a} = [a(0), a(1), ..., a(i), ..., a_j(\frac{n_s}{2})]$, where $a(i)$ is the amplitude of a sinusoid of frequency $i\frac{f_s}{n_s}$, which represents the spectrum of a reconstructed EEG epoch

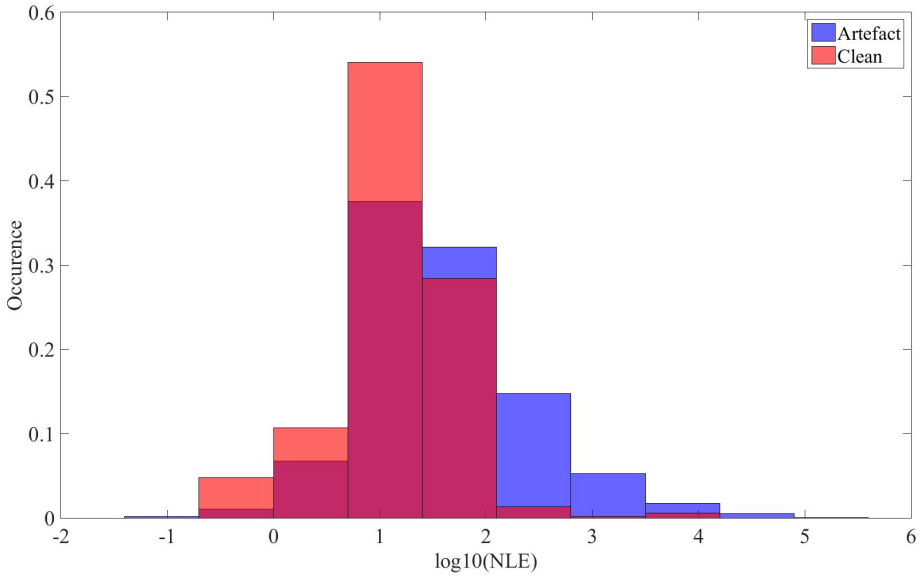


Figure 6.36: Class specific histogram for nonlinear energy.

$\tilde{\mathbf{x}}$, several frequency features can be obtained.

Gotman et al. initially proposed the use of dominant-peak frequency for detecting seizure in neonates (Gotman et al., 1997) which can be obtained as the frequency corresponding to the largest amplitude in the power spectral density (PSD) as:

$$F_4 = f_{peak}(\tilde{\mathbf{x}}) = i_{peak}\frac{n_s}{f_s} \tag{6.20}$$

where f_s = sampling frequency, and $i_{peak} = \arg\max_i \; a(i)$ is the index corresponding to the largest amplitude in the PSD. However, from figure 6.37 is can be seen that there is high overlap between the peak frequencies of both classes. Thus, despite several studies employing this feature, very little discriminating power is provided by the dominant frequency of the PSD.

6.10.3 Results of the ADS

The effect of the decomposition level is shown in table 6.11. A Gabor dictionary with $\alpha = 8$ and a decomposition level of 6 atoms provided an optimal performance of a mean AUC of 0.8853 (median = 0.9279 , IQR: [0.8355 - 0.9619]). Figure 6.38 shows an example 16s artefact and its reconstruction using 6 Gabor atoms. Figure 6.39 shows the median receiver-operator-characteristic (ROC) curve over the 51 unseen neonates in the LOO cross-validation.

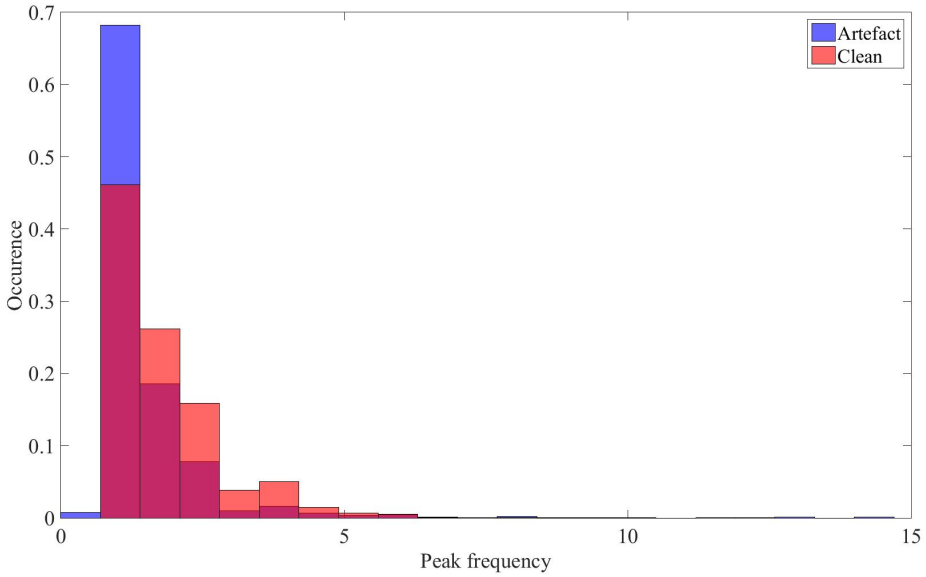


Figure 6.37: Class specific histogram for peak/dominant frequency.

Table 6.11: The mean, median(IQR) AUC obtained for varying number of atoms using OMP and a Gabor dictionary.

No. of atoms (K)	Mean AUC	Median AUC (IQR)
1	0.8475	0.8753 (0.7967 - 0.9487)
2	0.8515	0.8870 (0.8061 - 0.9426)
3	0.8600	0.8959 (0.8056 - 0.9460)
4	0.8767	0.9182 (0.8403 - 0.9566)
5	0.8712	0.9242 (0.8235 - 0.9564)
6	0.8853	0.9279 (0.8355 - 0.9619)
7	0.8795	0.9229 (0.8462 - 0.9511)
8	0.8825	0.9311 (0.8605 - 0.9630)
9	0.8820	0.9382 (0.8533 - 0.9581)
10	0.8708	0.9281 (0.8387 - 0.9665)

The ADS incorporates features that were integrated from the existing artefact or seizure detection methods (Nagaraj et al., 2014, Temko et al., 2011). The detection of major artefacts is important as they contribute to a total of 29% of the misclassifications by the AGS. The detection of minor artefacts (duration $\leq$ 16s) is more difficult due to limited data availability for the feature extraction. Moreover, it was observed that the minor artefacts only contributed to 4% of the total misclassifications by the AGS. The performance of the proposed ADS suggests that there is clearly a room for the development of real-time artefact detection of all durations. The novelty of the proposed ADS is that it is a patient independent artefact detection system and also takes into account the multichannel nature of

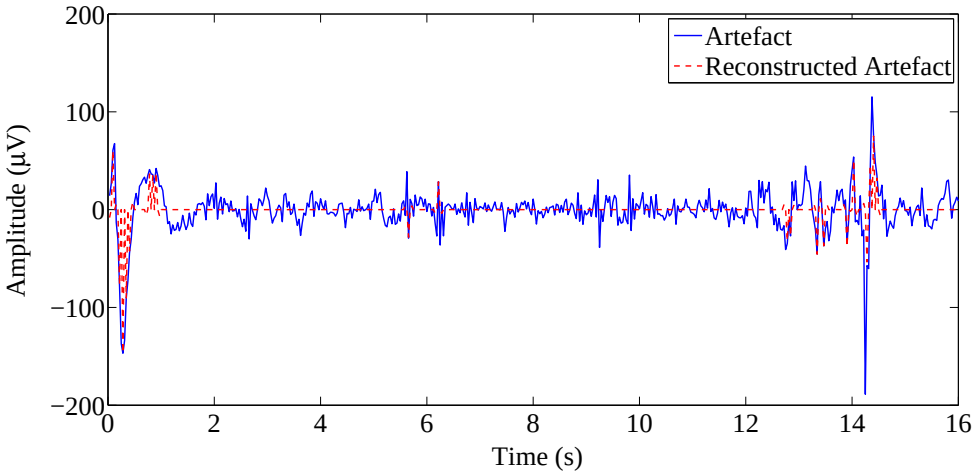


Figure 6.38: Example of a 16s artefact (blue) and its reconstruction (dotted red) using 6 Gabor atoms. Since the selected atoms during OMP were coherent with artefacts, we can see that Gabor atoms reconstruct some/most of the high amplitude sections of the artefact.

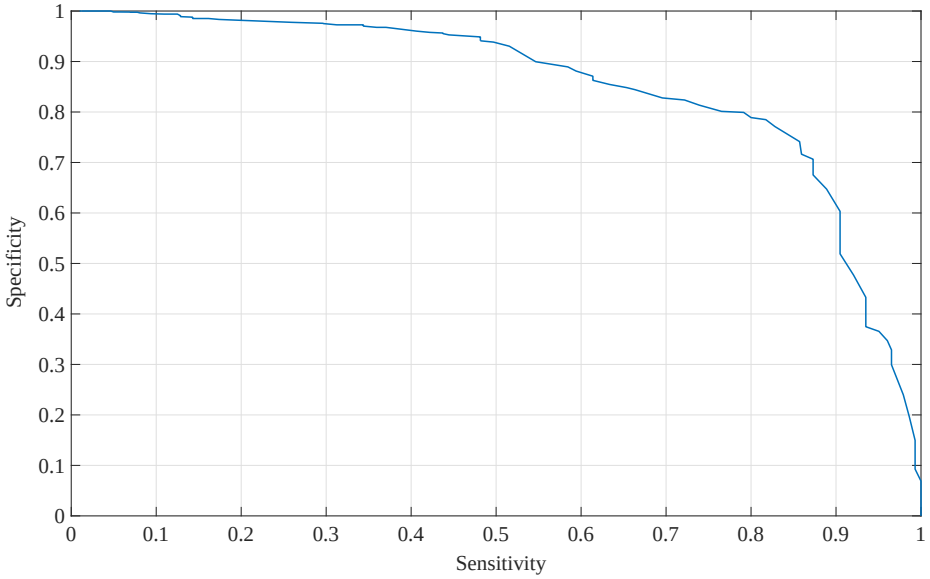


Figure 6.39: The performance curve (ROC) of the proposed ADS determined as the median over the 51 unseen neonates in the LOO cross-validation.

the artefacts in EEG. This is important as certain types of artefacts are present on multiple channels simultaneously. The proposed ADS in this chapter is thus suitable for artefact annotation of neonatal EEG recorded in a noisy environment. The proposed ADS could be incorporated into an automated neonatal EEG event detection (seizure detection, sleep states detection, HIE classification) either during the pre-processing (as an artefact rejector) or during post-processing stage (as a fusion of classifiers).

6.11 Combining the ADS with AGS using OMP for HIE-EEG grading

The architecture of the AGS using $OMP_{method1}$ which includes the ADS in the post-processing stage (AGS-ADS) is shown in figure 6.40. In order to generate an artefact annotation for the proposed AGS, a threshold for the ADS SVM output was selected corresponding to a specificity of 75% (from figure 6.39). The SVM output less than the threshold represents a clean EEG and greater than the threshold corresponds to an artefact. The features corresponding to artefacts were removed before passing it to the multi-class SVM for grading HIE. In this way, the effects of high energy, high amplitude artefacts were removed which mainly contribute to the false decisions made by AGS. The threshold was selected with an intention of preserving all clean EEG epochs and rejecting the majority of the high amplitude, high energy artefacts.

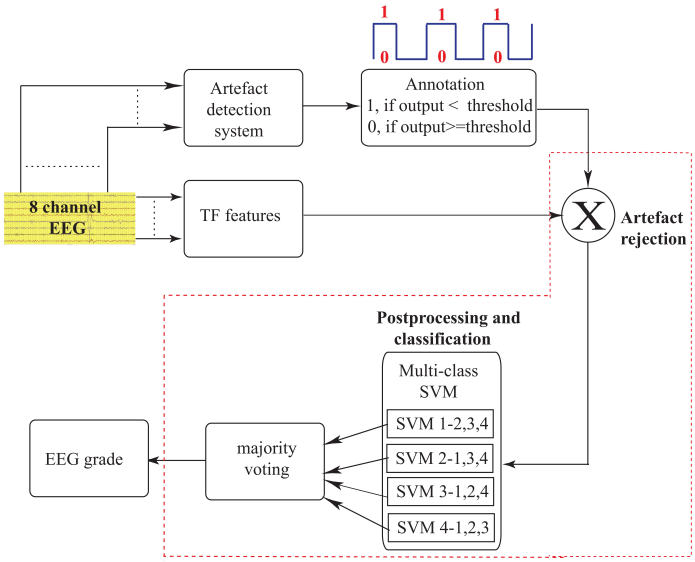


Figure 6.40: Illustration of the proposed AGS using $OMP_{method1}$ in combination with ADS. The inclusion of ADS removes the influence of artefacts on HIE grading.

Table 6.12 shows the performance of the proposed AGS-ADS system. We can clearly see that the overall performance is increased by 4% with the inclusion of the ADS. Two EEG grade 3 recordings that were misclassified as grade 2 in table 6.6 were correctly classified in this system since the artefacts were removed by the ADS. This shows that the fusion of the ADS into the AGS improves performance of the AGS to grade HIE-EEG. Table 6.13 compares the performance of the proposed AGS with several HIE grading algorithms in literature.

Table 6.12: Confusion matrix of the AGS-ADS output and the actual HIE grade assigned by the EEGer.

Actual HIE grade	AGS-ADS output					
	1	2	3	4	Total	Misclassified
1	21	1	0	0	22	1
2	3	10	1	0	14	4
3	2	0	10	0	12	2
4	0	0	0	6	6	0
Total	26	11	11	6		
Accuracy(%)	95	71	83	100		

Table 6.13: Performance of the proposed AGS with several methods in literature. (*)Same dataset was used in all these methods.

Method	No. of neonates	Duration (h)	Accuracy (%)
Proposed*(Hilbert, AGS+ADS)	54	62	87
Stevenson et al.* (Stevenson et al., 2013)	54	62	83.3
Korotchikova et al.* (Korotchikova et al., 2011)	54	62	72
Ahmed et al.* (Ahmed et al., 2014)	54	62	85
Matic et al. (Matic et al., 2014)	34	272	89

6.12 Discussion

The classification accuracy of the AD based AGS with multi-class SVM is shown in Table 6.14. The highest accuracy of the $OMP_{method1}$ based AGS obtained was 83.3% (43/54) using OMP alone and it increased to 87% with the inclusion of the ADS which was comparable to the accuracy obtained using the Hilbert transform method (87%). This suggests that TF features obtained directly from the TFD of the atoms selected during AD are not discriminative since the atoms in the dictionary are not coherent with particular HIE grades. Moreover, the computation time (calculated using a dual-core i5 Intel processor and 8gb ram) using this approach was very high as shown in figure 6.41. As we can see, it takes around 2 minutes to extract the features from a single channel of EEG, of 1 hour duration. Since this has to be repeated for all 8 channels, it takes around 16 minutes to obtain all 8 TF features from one EEG recording. This is due to the computation of TF features from every atom selected during AD. This means that if 10 atoms were used, then the TFD was obtained 10 times to compute the TF features from the final TFD. However, it only takes 30s to compute the TF features using the Hilbert based AGS for one hour EEG recording.

Table 6.14: Comparison of the performance of different AD techniques for automatic HIE-EEG grading.

AD type	Accuracy (%)
OMP(TF features from reconstructed signal)	83.3%
Multichannel matching pursuit (MMP)	77.8%
OMP (TF features from atoms)	74%
OMP(TF features from reconstructed signal)+ADS	87%

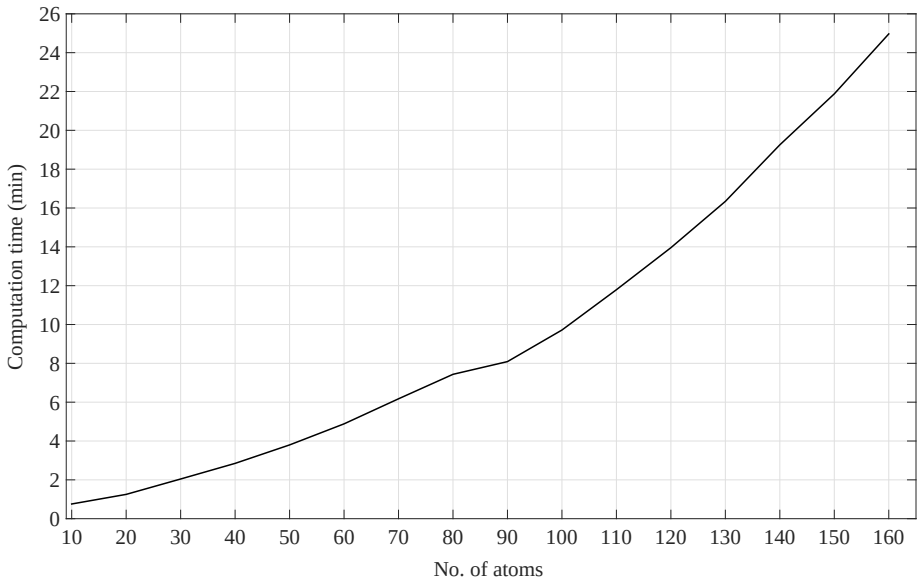


Figure 6.41: The computation time (in minutes) versus the number of atoms to extract TF features from a single channel 1 hour EEG recording (grade 1) using OMP.

Although AD provided promising results for seizure detection (chapters 4 and 5), its application to automatically grade HIE-EEG is limited. Table 6.14 gives the accuracy obtained using different versions of matching pursuit algorithm to automatically grade HIE-EEG. Due to the following reasons, AD based AGS may not be suitable for automatic grading of HIE-EEG:

1. The major limitation of using AD based techniques for automatic HIE-EEG grading is that each epoch can require up to 100 atoms for accurate representation and due to this it is a tedious process to study individual patient's data per channel for an hour long EEG recording. Moreover, since there is no inherent order of atoms (for example: the first atom selected across different patients of similar grade may not necessarily be same) selected during AD, it is difficult to obtain the general overview of a patient's EEG signal characteristics.

2. The choice of the decomposition dictionary is an important factor for AD. Since the Gabor dictionary provides optimal time-frequency resolution, it was used for grading HIE-EEG. However, it is a tedious process to identify the optimal number of atoms required to efficiently classify HIE grades. In addition, since the atoms in the dictionary are not coherent with any particular HIE-EEG grade, the information extracted from the atoms was not sufficiently discriminative.

3. The computation complexity of the AD based AGS was very high when compared to the Hilbert transform based AGS. Since the HIE-EEG is graded based on a decision taken over an hour long EEG, it is necessary to decompose the entire one hour EEG recording using AD. Moreover, the accuracy obtained using AD based methods were lower.

4. The performance tradeoff in AD is mainly regulated by the residual error. Smaller residual error results in larger amount of atoms in the dictionary and increased resolution of the AD but at a price of increased computation complexity. Since the residual error is directly dependent on the number of atoms used for decomposition, more effort is required to find the optimal threshold for accurate classification.

5. Another major issue in using AD is the size of the dictionary. Increasing the size of the dictionary increases the time-frequency resolution but will also significantly increase the computational cost. Moreover, it is a challenging task to find the exact resolution required for signal representation using AD.

6.13 Summary

In this chapter, a novel SVM based automated HIE-EEG grading system is proposed. The proposed AGS is based on the ability of the features obtained from QTFD analysis to track the nonstationarity of the EEG signal. Several methods to extract these TF features were studied and their performance was reported.

It has been previously shown that an HIE-EEG AGS using QTFD based TF features obtained from the Hilbert transform method can provide good HIE-EEG grading (Stevenson et al., 2013). In this chapter, it was seen that the performance of this AGS improved significantly using an SVM with an overall accuracy of 87%. It was also shown that the TF features obtained from the QTFD of the linear

combination of the Gabor atoms (or reconstructed signal) selected during AD provided highest accuracy when compared to other AD methods. The AD based AGS provided an accuracy of 83.3% in grading HIE-EEG.

A novel AD based patient independent artefact detection system was proposed to improve the performance of the proposed AD based AGS. Only the artefacts that were $\geq$ 16s in duration were considered in this system since their contribution was higher for the misclassifications by the AGS. The proposed AD based ADS resulted in a mean AUC of 0.89 for artefact detection. With the additional help of an AD based artefact detector, this performance was increased to 87% when the ADS was included as an artefact rejector in the $OMP_{method1}$ based AGS. A multiclass SVM was used in this study for classification. From the analysis of the misclassified recordings it was concluded that the presence of abnormal patterns such as asymmetry, asynchrony, runs of sharp waves should influence the decision of the AGS. Several limitations of AD based methods to automatically grade HIE-EEG were also discussed in this chapter.

The presented SVM based approach showed promising results in HIE-EEG grading, however there are certain opportunities to improve the performance of the proposed AGS. Detection of the short duration physiologically variable neonatal EEG artefacts is a challenging problem, in particular for the analysis and grading of HIE-EEG. In addition, the proposed AGS does not detect certain short duration abnormalities present in EEG such as asymmetry, asynchrony and runs of sharp waves. Inclusion of parallel detectors to detect such abnormalities has the potential to improve the performance of the proposed AGS significantly.

References

Achim, A., 1995. Cerebral source localization paradigms-spatiotemporal source modeling. Brain and cognition 27 (3), 256–287.

Ahmed, R., Temko, A., Marnane, W., Boylan, G., Lightbody, G., 2014. Grading brain injury in neonatal EEG using svm and supervector kernel. In: Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. IEEE, pp. 5894–5898.

Azzopardi, D. V., Strohm, B., Edwards, A. D., Dyet, L., Halliday, H. L., Juszczak, E., Kapellou, O., Levene, M., Marlow, N., Porter, E., et al., 2009. Moderate hypothermia to treat perinatal asphyxial encephalopathy. New England Journal of Medicine 361 (14), 1349–1358.

Boashash, B., 2003. Time Frequency Signal Analysis and Processing: Method and Applications. Elsevier, Oxford.

Boashash, B., Mesbah, M., 2001. A time-frequency approach for newborn seizure detection. Engineering in Medicine and Biology Magazine, IEEE 20 (5), 54–64.

Boashash, B., Mesbah, M., 2002. Time-frequency methodology for newborn electroencephalographic seizure detection. Applications in Time–Frequency Signal Processing, A. Papreou-Suppappola, Ed. Boca Raton.

Box, G. E., Cox, D. R., 1964. An analysis of transformations. Journal of the Royal Statistical Society. Series B (Methodological), 211–252.

Boylan, G. B., Stevenson, N. J., Vanhatalo, S., 2013. Monitoring neonatal seizures. In: Seminars in Fetal and Neonatal Medicine. Vol. 18. Elsevier, pp. 202–208.

Celka, P., Colditz, P., 2002. Nonlinear nonstationary wiener model of infant EEG seizures. Biomedical Engineering, IEEE Transactions on 49 (6), 556–564.

Crammer, K., Singer, Y., 2002. On the learnability and design of output codes for multiclass problems. Machine Learning 47 (2-3), 201–233.

D'Alessandro, M., Esteller, R., Vachtsevanos, G., Hinson, A., Echauz, J., Litt, B., 2003. Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: a report of four patients. Biomedical Engineering, IEEE Transactions on 50 (5), 603–615.

Daubechies, I., 1988. Orthonormal bases of compactly supported wavelets. Communications on pure and applied mathematics 41 (7), 909–996.

Durka, P. J., Matysiak, A., Montes, E. M., Sosa, P. V., Blinowska, K. J., 2005. Multichannel matching pursuit and EEG inverse solutions. Journal of Neuroscience Methods 148 (1), 49–59.

Fenichel, G. M., 2009. Clinical pediatric neurology: a signs and symptoms approach. Elsevier Health Sciences.

Ghofrani, S., McLernon, D., Ayatollahi, A., 2003. Comparing gaussian and chirplet dictionaries for time-frequency analysis using matching pursuit decomposition. In: Signal Processing and Information Technology, 2003. ISSPIT 2003. Proceedings of the 3rd IEEE International Symposium on. IEEE, pp. 713–716.

Gotman, J., Flanagan, D., Zhang, J., Rosenblatt, B., 1997. Automatic seizure detection in the newborn: methods and initial evaluation. Electroencephalography and Clinical Neurophysiology 103 (3), 356–362.

Graham, E. M., Ruis, K. A., Hartman, A. L., Northington, F. J., Fox, H. E., 2008. A systematic review of the role of intrapartum hypoxia-ischemia in the causation of neonatal encephalopathy. American journal of obstetrics and gynecology 199 (6), 587–595.

Gray, P. H., Tudehope, D. I., Masel, J. P., Burns, Y. R., Mohay, H. A., O'Callaghan, M. J., Williams, G. M., 1993. Perinatal hypoxic-ischaemic brain injury: Prdiction of outcome. Developmental Medicine & Child Neurology 35 (11), 965–973.

Hsu, C.-W., Lin, C.-J., 2002. A comparison of methods for multiclass support vector machines. Neural Networks, IEEE Transactions on 13 (2), 415–425.

Inder, T. E., Volpe, J. J., 2000. Mechanisms of perinatal brain injury. Seminars in neonatology 5 (1), 3–16.

Janssen, A., Claasen, T., 1985. On positivity of time-frequency distributions. Acoustics, Speech and Signal Processing, IEEE Transactions on 33 (4), 1029–1032.

Korotchikova, I., Stevenson, N., Walsh, B., Murray, D., Boylan, G., 2011. Quantitative EEG analysis in neonatal hypoxic ischaemic encephalopathy. Clinical Neurophysiology 122 (8), 1671–1678.

Lelic, D., Gratkowski, M., Hennings, K., Drewes, A. M., 2011. Multichannel matching pursuit validation and clustering—a simulation and empirical study. Journal of Neuroscience Methods 196 (1), 190–200.

Liu, A., Hahn, J., Heldt, G., Coen, R., 1992. Detection of neonatal seizures through computerized EEG analysis. Electroencephalography and Clinical Neurophysiology 82 (1), 30–37.

Mallat, S. G., Zhang, Z., 1993. Matching pursuits with time-frequency dictionaries. Signal Processing, IEEE Transactions on 41 (12), 3397–3415.

Matic, V., Cherian, P. J., Koolen, N., Naulaers, G., Swarte, R. M., Govaert, P., Van Huffel, S., De Vos, M., 2014. Holistic approach for automated background EEG assessment in asphyxiated full-term infants. Journal of Neural Engineering 11 (6), 066007.

Mizrahi, E. M., Hrachovy, R. A., Kellaway, P., Stockard-Pope, J., 2004. Atlas of neonatal electroencephalography. Vol. 31. Lippincott Williams & Wilkins Philadelphia.

Murray, D. M., Boylan, G. B., Ryan, C. A., Connolly, S., 2009. Early EEG findings in hypoxic-ischemic encephalopathy predict outcomes at 2 years. Pediatrics 124 (3), e459–e467.

Nagaraj, S. B., Stevenson, N. J., Marnane, W. P., Boylan, G. B., Lightbody, G., 2014. Neonatal seizure detection using atomic decomposition with a novel dictionary. IEEE transactions on bio-medical Engineering 61 (11), 2724–2732.

Nicklin, S., Wickramasinghe, Y. A., Andrew Spencer, S., 2004. Neonatal intensive care monitoring. Current Paediatrics 14 (1), 1–7.

Niedermeyer, E., da Silva, F. L., 2005. Electroencephalography: basic principles, clinical applications, and related fields. Lippincott Williams & Wilkins.

Notley, S. V., Elliott, S. J., 2003. Efficient estimation of a time-varying dimension parameter and its application to EEG analysis. Biomedical Engineering, IEEE Transactions on 50 (5), 594–602.

Nunez, P. L., Srinivasan, R., 2006. Electric fields of the brain: the neurophysics of EEG. Oxford university press.

Platt, J. C., 1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Advances in large margin classifiers. Citeseer.

Rankine, L., Stevenson, N., Mesbah, M., Boashash, B., 2007. A nonstationary model of newborn EEG. Biomedical Engineering, IEEE Transactions on 54 (1), 19–28.

Roessgen, M., Zoubir, A. M., Boashash, B., 1998. Seizure detection of newborn EEG using a model-based approach. Biomedical Engineering, IEEE Transactions on 45 (6), 673–685.

Rubinstein, R., Bruckstein, A. M., Elad, M., 2010. Dictionaries for sparse representation modeling. Proceedings of the IEEE 98 (6), 1045–1057.

Saito, N., 1999. Local fourier dictionary: a natural tool for data analysis. In: SPIE's International Symposium on Optical Science, Engineering, and Instrumentation. International Society for Optics and Photonics, pp. 610–624.

Scher, M., 2001. Perinatal asphyxia: timing and mechanisms of injury in neonatal encephalopathy. Current neurology and neuroscience reports 1 (2), 175–184.

Shankaran, S., 2012. Therapeutic hypothermia for neonatal encephalopathy. Current treatment options in neurology 14 (6), 608–619.

Steriade, M., McCormick, D. A., Sejnowski, T. J., 1993. Thalamocortical oscillations in the sleeping and aroused brain. Science 262 (5134), 679–685.

Stevenson, N., Korotchikova, I., Temko, A., Lightbody, G., Marnane, W., Boylan, G., 2013. An automated system for grading EEG abnormality in term neonates with hypoxic-ischaemic encephalopathy. Annals of Biomedical Engineering 41 (4), 775–785.

Stevenson, N., Mesbah, M., Boashash, B., 2007. Modelling newborn EEG background using a time-varying fractional brownian process. In: Proceedings of the 15th European Signal Processing Conference. Vol. 1. pp. 1246–1250.

Stevenson, N. J., Mesbah, M., Boylan, G. B., Colditz, P. B., Boashash, B., 2010. A nonlinear model of newborn EEG with nonstationary inputs. Annals of Biomedical Engineering 38 (9), 3010–3021.

Temko, A., Thomas, E., Marnane, W., Lightbody, G., Boylan, G., 2011. EEG-based neonatal seizure detection with support vector machines. Clinical Neurophysiology 122 (3), 464–473.

Tropp, J., Gilbert, A., Dec 2007. Signal recovery from random measurements via orthogonal matching pursuit. Information Theory, IEEE Transactions on 53 (12), 4655–4666.

Vanhatalo, S., Kaila, K., 2006. Development of neonatal EEG activity: from phenomenology to physiology. In: Seminars in Fetal and Neonatal Medicine. Vol. 11. Elsevier, pp. 471–478.

Vapnik, V., 1982. Estimation of Dependences Based on Empirical Data: Springer Series in Statistics (Springer Series in Statistics). Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Vapnik, V. N., Vapnik, V., 1998. Statistical learning theory. Vol. 2. Wiley New York.

Varsavsky, A., Mareels, I., Cook, M., 2010. Epileptic seizures and the EEG: measurement, models, detection and prediction. CRC Press.

Vera-Candeas, P., Ruiz-Reyes, N., Rosa-Zurera, M., Martinez-Munoz, D., Lopez-Ferreras, F., 2004. Transient modeling by matching pursuits with a wavelet dictionary for parametric audio coding. Signal Processing Letters, IEEE 11 (3), 349–352.

Wald, L., 1999. Some terms of reference in data fusion. IEEE Transactions on geoscience and remote sensing 37 (3), 1190–1193.

Walsh, B., Murray, D., Boylan, G., 2011. The use of conventional EEG for the assessment of hypoxic ischaemic encephalopathy in the newborn: a review. Clinical Neurophysiology 122 (7), 1284–1294.

Yang, G., Zhang, Q., Que, P.-W., 2007. Matching-pursuit-based adaptive wavelet-packet atomic decomposition applied in ultrasonic inspection. Russian Journal of Nondestructive Testing 43 (1), 62–68.

Chapter 7

Conclusions and future work

> Life is like riding a bicycle. To keep your balance, you must keep moving. ––Albert Einstein

This thesis focussed on a detailed study of the application of atomic decomposition techniques to classify neonatal EEG. Several time-frequency signal processing techniques were studied in this thesis which led to the development of:

1. A novel automated neonatal EEG seizure detection system.
2. An automated HIE grading system.

Several time-frequency based features were developed for this purpose. Based on the results obtained in this thesis, this chapter concludes and summarises the contents of each chapter. The contributions of this thesis are highlighted in the summary. Several future directions of this research are proposed which were identified during this study.

7.1 Summary of the thesis

A large number of topics have been discussed and analyzed in this thesis, related to neonatal EEG classification. This section briefly summarises the contents of individual chapter in terms of the analysis and experiments performed in this thesis.

Chapter 1 provided the aim and scope of this thesis. Several limitations in the current automated neonatal EEG analysis were described based on the literature

review of the current research in this field. It was observed that several neonatal EEG analysis algorithms reported in literature either use a small dataset of artefact free neonatal EEG or simulated or selective datasets and report their performance with no apparent validation of results. Moreover, a lack of datasets that can best represent the real-time situation in the NICU also hinders the development of automated algorithms. In order to effectively implement the automated neonatal EEG analysis systems in real-time, the algorithm must be robust enough to deal with variety of EEG signals that may arise in noisy, medical environments. Due to this, an automated detector was found to be useful to continuously monitor sick newborn patients in the NICU that could help clinicians with the interpretation of neonatal EEG. The primary objective to significantly improve the performance of previously reported neonatal seizure detectors and HIE grading systems was proposed in this chapter.

Chapter 2 provided a brief overview about the properties of neonatal EEG signals in terms of signal processing. From several case studies, it was shown that electroencephalogram (EEG) monitoring was required to identify neonatal seizures as they are often clinically silent. In addition, several characteristics of the background EEG with HIE were also reviewed. Several abnormal patterns in the background EEG were described including the grading of HIE-EEG based on several characteristics used by clinicians. It was deemed necessary to design an automated system to detect neonatal seizures and also to grade EEG based on HIE severity. This is required for continuous monitoring of sick newborn patients as early detection can have great impact on predicting of neurological outcome and choice of therapy.

Chapter 3 provided a detailed description of time-frequency analysis and its applications in nonstationary signal analysis. Several limitations of traditional stationary analysis for nonstationary signals were demonstrated and it was shown that time-frequency analysis was an ideal choice for analyzing nonstationary signals. Several atomic decomposition (AD) techniques were investigated in this chapter and it was shown that AD is mainly useful for randomly appearing signal components in the time series. It was demonstrated that the time-frequency representation (TFR) based on the AD offered a step-wise adaptive compromise between the time and frequency resolution which provides a time and frequency invariant decomposition. Also since the TFR derived from the AD has explicitly no cross-term, it can be used to obtain clean and easy-to-interpret time-frequency maps of the signal. Several properties of different types of dictionary were also discussed in this chapter.

Chapter 4 investigated several AD techniques for the analysis of neonatal EEG. AD is mainly used to provide sparse signal representations using redundant dictionaries. However, in several applications only a certain level of signal approximation is sufficient for the purpose of discrimination. It was demonstrated that the orthogonal matching pursuit (OMP) provides better discrimination of seizures from non-seizures when compared to basis pursuit (BP) and matching pursuit (MP). For this reason, OMP was used as a preferred AD technique throughout this thesis.

It was demonstrated that as the number of atoms in OMP decomposition increased, the coherence between signal structures and the atoms in the decomposition dictionary decreased. Based on this, it was shown that the number of atoms used in OMP decomposition can be used to discriminate between neonatal EEG seizure and nonseizure. A novel measure known as relative structural complexity (RSC) which is defined as the number of atoms requires to reach a set signal to error ratio (SER) was used as a measure to distinguish between seizure and nonseizure states. It was observed that the RSC was mainly dependent on the signal structures and the atoms in the dictionary.

A novel time-frequency dictionary denoted as the *pseudo-periodic Duffing dictionary* was designed based on a recent neonatal seizure model using the Duffing oscillator. The atoms in this dictionary were highly coherent with the neonatal EEG seizure structures. This dictionary was based on the TF characterization of the neonatal EEG seizure. Atoms in this dictionary had a high correlation with seizure waveforms and a low correlation with nonseizure waveforms. As a result, the RSC will increase in the presence of seizure. A novel RSC based neonatal seizure detection algorithm (NSDA) was then developed using the proposed novel dictionary and OMP. The OMP based NSDA achieved an overall mean AUC of 0.88 when tested on a large dataset using several post-processing techniques.

Chapter 5 improved the performance of the neonatal seizure detection algorithm (NSDA) proposed in chapter 4 using additional time-frequency distribution (TFD) features obtained from the pseudo-periodic Duffing dictionary. The main advantage of using the TFD was that key instantaneous frequency values can be obtained which can characterize the nonstationary nature of the EEG signal. A brief description about the support vector machine (SVM) classifier was presented. An AD based seizure detection algorithm using nonstationary features obtained from the TFD of the atoms from the pseudo-periodic Duffing dictionary was proposed in this chapter using SVM. This approach provides a

noise-free representation and increased the time-frequency resolution of the underlying signal. 8 features were derived from the TFD of the atoms selected from the dictionary during AD. It was seen that the proposed method had relatively good performance in the large dataset using minimal number of features when compared to the current *state-of-the-art* methods.

Chapter 6 proposed an automated grading system (AGS) for grading EEG based on the severity of HIE injury. The proposed AGS uses a nonlinear EEG background model to describe the characteristics of EEG. The proposed AGS used 8 TF features which provided simple statistics (mean, median, skewness, kurtosis) of the amplitude modulation and instantaneous frequency characteristics of the EEG signal over the window of 64s. The performance of the AGS proposed in (Stevenson et al., 2013) was improved using support vector machine (SVM). By using a multiclass SVM, the HIE-EEG classification accuracy was improved from 78% proposed in (Stevenson et al., 2013) to 87% using a leave-one-out (LOO) cross validation.

The performance of the AGS using several AD methods were also studied in this chapter. Several limitations of the application of AD for automatic grading of HIE-EEG was described. With the help of an additional artefact detection system, the AD based AGS resulted in an accuracy of 87% in classifying HIE EEG. It was concluded that the AD based AGS for automatic HIE-EEG may not be feasible due to increased computational complexity and poor performance.

Misclassified recordings were further analysed and it was observed that the presence of abnormalities such as asymmetry, asynchrony and runs of sharp waves resulted in the misclassification. Due to the presence of these abnormalities in the EEG recording, the AGS decision was downgraded. This means that if the EEG was graded as grade 2 by the annotators due to the presence of these abnormal patterns, the output of the AGS was grade 1. In addition, the influence of artefacts resulted in an incorrect decision of 33% of the misclassified EEG recordings in which the majority (29%) was due to the presence of artefacts $\geq$ 16s in duration.

7.2 Future research work

The features and methods described in this thesis provides a starting point for further investigation into neonatal EEG analysis and abnormality detection. In

this section some suggestions for further investigation are provided that take the advantages of atomic decomposition and time-frequency analysis to further improve the performance of the proposed systems in this thesis.

7.2.1 Neonatal seizure detection

In this thesis, a novel time-frequency dictionary using Duffing oscillator model which was coherent with newborn EEG seizures was proposed to detect neonatal EEG seizures. Several TF features were extracted from this dictionary and an automated neonatal EEG seizure detection system was developed. This is a model based dictionary and it is worth investigating the design of a data based dictionary where the seizure epochs from the database can be directly used to design a TF dictionary. With the help of available *dictionary learning* algorithms, a TF dictionary coherent with seizure epochs can also be designed to detect neonatal EEG seizures. It is interesting to test the performance of the proposed TF dictionary using the seizure detection algorithm used by Temko et al (Temko et al., 2011).

It was shown in (Boashash et al., 2012, Nagaraj et al., 2014) that the addition of image processing features can improve the performance of the seizure detection system. However, it is computationally expensive and more research is required to analyze the potential of image processing features for seizure detection. Inclusion of spatial and temporal information (Bye and Flanagan, 1995) of the seizures can also be useful in improving the performance of proposed seizure detection algorithm in this thesis .

Since most of the false detections were due to the presence of artefacts such as respiration and movement traces, an artefact detection system proposed by Stevenson et al. can be incorporated to reduce the false alarms (Stevenson et al., 2014). Furthermore, it will be interesting to develop a separate seizure detection system for short duration seizures ($\leq$ 30s) and then cascade it with the NSDA proposed in this thesis to efficiently detect short duration seizures as well.

7.2.2 Background EEG classification

The EEG of the fullterm infants have clear sleep patterns which can describe the continuity of their EEG signals. In EEG, the sleep-wake cycle normally appears as alternating discontinuous (sleep) and continuous (awake) periods. Therefore,

it is worthwhile designing an automated sleep-stage detector with which the state of the infant brain can be obtained from the continuity of the EEG (Stevenson et al., 2013). This can be implemented in parallel with the automated grading system proposed in this thesis to obtain the sleep-state disorders in neonates.

Only Gabor dictionary was used in this thesis for background classification. Since there is no well defined pattern in background EEG, it is difficult to design a model based dictionary for HIE classification. However, with the help of several dictionary learning algorithms the atoms in the dictionary can be trained to be coherent with specific HIE grades (Mairal et al., 2009, Yaghoobi et al., 2009, Yang et al., 2011, Zhang and Li, 2010). Several features can then be obtained from these trained dictionaries to grade HIE-EEG.

7.3 Asymmetry and asynchrony measurements

It was reported in (Ahmed et al., 2014, Stevenson et al., 2013) that the presence of background EEG abnormalities such as artefacts, asymmetry, asynchrony, spike and sharp waves decreased the performance of the automated grading systems (AGS) which was also observed in chapter 6 of this thesis. The incorporation of methods to detect asymmetry, asynchrony and runs of sharp waves has the potential to improve the AGS. The presence of these patterns in the EEG recording results in a downgrade of the estimated AGS grade (i.e if the actual grade of the EEG is grade 2 then the output of the AGS grades the EEG as grade 1). It is difficult to detect these abnormalities as they appear intermittently. In this section, a few symmetry and synchrony measures are analyzed and their effectiveness for the data is also demonstrated.

Traditionally, neonatal EEG is classified based on visual assessment of several clinical features such as symmetry, synchrony, amplitude, continuity. It was observed that the presence of these features was not detected by the AGS for automatic grading of HIE. In this section, some standard features are explored that provide some information about presence of symmetry and synchrony in an EEG signal on the misclassified recordings. Their ability to detect asynchronous/asymmetric events in the EEG signal is tested The main aim is to identify certain features that could be used as an additional feature to detect the abnormal patterns in HIE-EEG AGS.

7.3.1 Asymmetry measurement

Symmetry provides information about the similarity of the two hemispheres. Symmetry is mainly concerned with the general pattern observed in both hemispheres and how they are similar to each other. The EEG recording is said to be *asymmetric* if a particular pattern is present in one hemisphere and absent in the other. Symmetry is generally expected from EEG traces of all infants regardless of age or state. Asymmetry usually indicates brain injury or a delay in development in one of the hemispheres.

Symmetry is commonly present on most of the physiological recordings and its measurement is important in evaluation of several clinical conditions. Measurement of left-right symmetry is an important characteristic of EEG which provides information about different rhythms in left and right hemispheric EEG (Niedermeyer and da Silva, 2005). Several features such as the change in the spectral edge frequency, changes in relative power have been proposed in the literature to measure symmetry between the channels (Hanowell et al., 1992, Laman et al., 2005). Recently, a measure known as *brain symmetry index* (BSI) was proposed which quantifies the interhemispheric spectral symmetry of the EEG (Van Putten, 2006, Van Putten and Tavy, 2004). This measure was introduced to assist the visual interpretation of EEG. Later, the revised BSI was proposed which has an improved sensitivity in detecting interhemispheric asymmetry and diffuse changes (Van Putten, 2007). The revised BSI is used in this section to measure the asymmetry between the channels of the misclassified recordings. An example of asymmetry present in the misclassified recording is shown in figure 7.1.

Figure 7.1: Example of an asymmetry event in a 64s HIE grade 2 EEG.

The standard BSI proposed in (Van Putten, 2006) is defined as:

$$\text{BSI} = \frac{1}{N}\sum_{i=1}^{P}\left\|\frac{1}{M}\sum_{j=1}^{M}\frac{R_{i,j}-L_{i,j}}{R_{i,j}-L_{i,j}}\right\| \tag{7.1}$$

where N is the number of channel pairs, M is the number of Fourier coefficients, $R_{i,j}, L_{i,j}$ is the Fourier coefficient corresponding to right and left hemispheres, respectively. The revised BSI is based on the square value of the Fourier coefficients instead of the absolute value. This approach becomes more sensitive to the changes in hemispheric asymmetry. The revised BSI is defined as (Van Putten, 2007):

$$\text{rBSI} = \frac{1}{P}\sum_{i=1}^{P}\left|\frac{R_i^*-L_i^*}{R_i^*+L_i^*}\right| \tag{7.2}$$

where

$$R_i^* = \frac{1}{M}\sum_{j=1}^{M}a_{ij}^2 \tag{7.3}$$

for the right hemisphere. A similar expression is defined for left hemisphere. Here, a_{ij}^2 correspond to the magnitude of the Fourier coefficient at the i^{th} frequency for the j^{th} channel. If the signals are symmetric, then $rBSI$ is close to zero. On the other hand, if they are asymmetric, then $rBSI$ is closer to 1. In the current study, the rBSI measurements were obtained across the bilateral electrode pairs: F3-C3/F4-C4, CZ-C3/C4-CZ, C3-O1/C4-O2, T4-C4/C3-T3.

Figure 7.2 shows the distribution of the mean $rBSI$ across symmetric and asymmetric epochs in the EEG signal (for channel F3-C3/F4-C4 and using 3 misclassified recordings). We can clearly see that the $rBSI$ measurement was not able to detect interhemispheric asymmetry and diffuse changes. It was demonstrated that the BSI measure was sensitive to ischemia and captures the normalized difference in the spectral density of the two cerebral hemispheres when compared to other measures (Van Putten et al., 2004) in adult EEG. It was also shown that presence of artefacts also influenced the values of BSI. Currently, the BSI was implemented on the real EEG data and free from artefacts, however it still failed to detect asymmetry.

7.3.2 Asynchrony measurement

Synchrony of the EEG provides information about the activities with respect to time. In synchronous EEG trace activities present in one hemisphere are

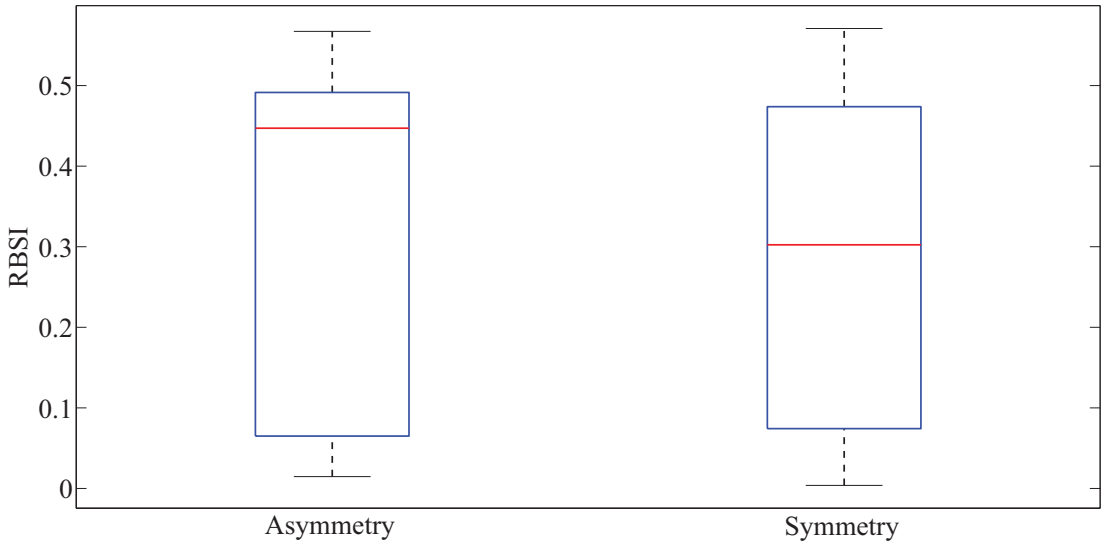


Figure 7.2: Box plot showing the distribution of $rBSI$ measures for channel F3-C3/F4-C4. There is no clear separation between the box plots.

also present in the other hemisphere of the brain at the same time. By finding the correlation between the two channels/hemispheres, the synchrony between the two channels can be obtained. An example of asynchrony present in the misclassified recording is shown in figure 7.3.

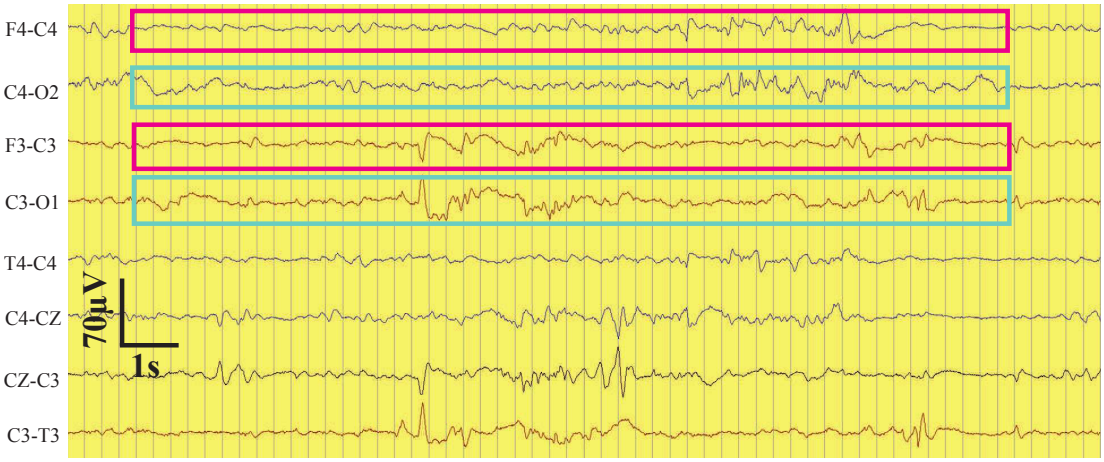


Figure 7.3: Example of an asynchrony event in a 64s HIE grade 2 EEG.

Several signal processing methods are available to detect the fluctuations of inter hemispheric synchrony (IHS) in EEG traces (Dauwels et al., 2010, Lachaux et al., 1999, Pereda et al., 2005, Uhlhaas and Singer, 2006). Recently, a quantitative measure to detect IHS in neonatal EEG known as activation synchrony index (ASI) was proposed in (Räsänen et al., 2013). The ASI measures the temporal delay between two signal energies and it was shown to correlate with the visually annotated IHS events. In this section, the ability of ASI to detect IHS on

the misclassified EEG recordings in the database is tested. The details of the algorithm to estimate ASI can be found in (Räsänen et al., 2013). The basic steps involved in estimating ASI are shown in figure 7.4. The ASI measure for synchronous signals is higher when compared to asynchronous signals. Figure

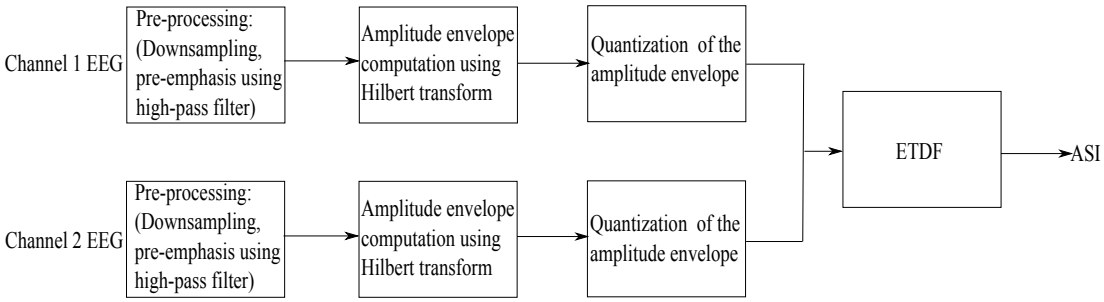


Figure 7.4: Schematic representation of the ASI feature extraction algorithm.

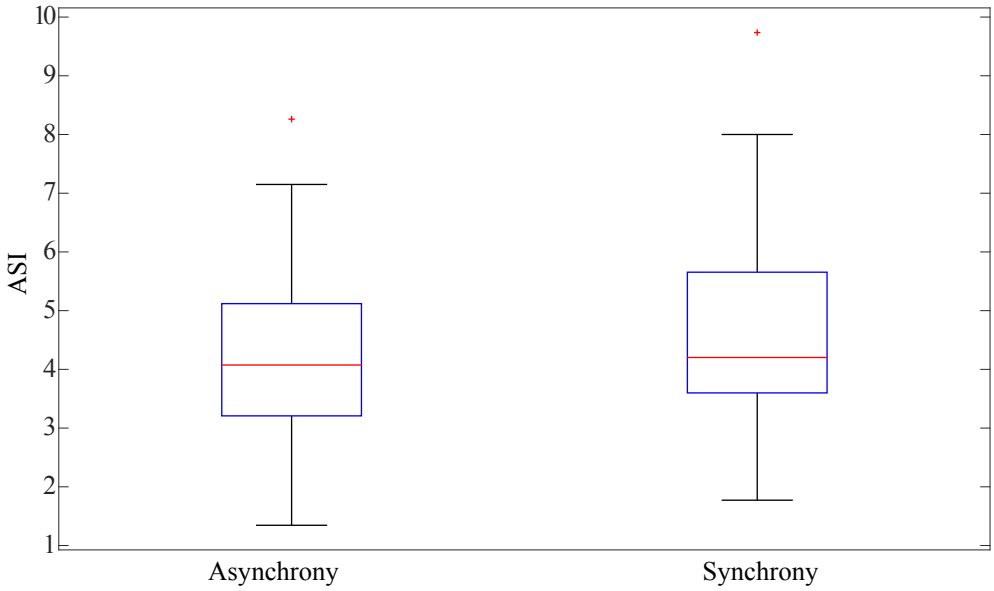


Figure 7.5: Box plot showing the distribution of ASI measures for channel F3-C3/F4-C4. We can observe that box plots are less separated.

7.5 shows the box plot of the distribution of ASI values obtained using three misclassified EEG recordings. Even though the median ASI was slightly lower for asynchronous signals when compared to that of synchronous EEG signals, there is no clear difference obtained that best separates two measures which makes it difficult to classify the recordings. It was also observed that the ASI measures obtained were very sensitive to the values of λ, the spectrum analysis window length and also the high-pass cutoff frequency. This suggests that using ASI for measuring interhemispheric synchrony from the neonatal EEG is not robust and

is relatively sensitive to the parameters used to measure ASI.

7.4 Conclusion

This thesis focussed on developing several time-frequency features using atomic decomposition technique for the anbalysis of neonatal EEG. Several newborn automated EEG techniques were proposed in this thesis. The thesis developed automated systems for seizure detection and HIE grading. The system uses EEG for brain function monitoring of neonates which can be used for early detection and prognosis of the brain abnormalities. The proposed systems were tested on unedited EEG recordings which makes it robust for NICU applications. The SVM based neonatal seizure detection system provides an accuracy of 91% and the features developed in this system can be used to improve the performance of state-of-the-art NSDA (Temko et al., 2011).

The SVM based AGS proposed in Chapter 6 results in an accuracy of 87% for grading HIE-EEG and there are several opportunities to improve the performance of the proposed AGS. Detection of the abnormal patterns (asymmetry, asynchrony, and sharp waves) is not trivial and is a challenging problem. It was observed that the standard measures to detect these abnormalities were not discriminatory between the grades and cannot be used as a feature in AGS. Measures to detect such abnormalities have a potential to improve the performance of the proposed AGS significantly.

References

Ahmed, R., Temko, A., Marnane, W., Boylan, G., Lightbody, G., 2014. Grading brain injury in neonatal eeg using svm and supervector kernel. In: Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. IEEE, pp. 5894–5898.

Boashash, B., Boubchir, L., Azemi, G., 2012. A methodology for time-frequency image processing applied to the classification of non-stationary multichannel signals using instantaneous frequency descriptors with application to newborn eeg signals. EURASIP Journal on Advances in Signal Processing 2012 (1), 1–21.

Bye, A., Flanagan, D., 1995. Spatial and temporal characteristics of neonatal seizures. Epilepsia 36 (10), 1009–1016.

Dauwels, J., Vialatte, F., Musha, T., Cichocki, A., 2010. A comparative study of synchrony measures for the early diagnosis of alzheimer's disease based on eeg. NeuroImage 49 (1), 668–693.

Hanowell, L. H., Soriano, S., Bennett, H. L., 1992. Eeg power changes are more sensitive than spectral edge frequency variation for detection of cerebral ischemia during carotid artery surgery: a prospective assessment of processed eeg monitoring. Journal of cardiothoracic and vascular anesthesia 6 (3), 292–294.

Lachaux, J.-P., Rodriguez, E., Martinerie, J., Varela, F. J., et al., 1999. Measuring phase synchrony in brain signals. Human brain mapping 8 (4), 194–208.

Laman, D. M., Wieneke, G. H., van Duijn, H., Veldhuizen, R. J., van Huffelen, A. C., 2005. Qeeg changes during carotid clamping in carotid endarterectomy: spectral edge frequency parameters and relative band power parameters. Journal of Clinical Neurophysiology 22 (4), 244–252.

Mairal, J., Ponce, J., Sapiro, G., Zisserman, A., Bach, F. R., 2009. Supervised dictionary learning. In: Advances in Neural information processing systems. pp. 1033–1040.

Nagaraj, S., Stevenson, N., Marnane, W., Boylan, G., Lightbody, G., Aug 2014. Robustness of time frequency distribution based features for automated neonatal eeg seizure detection. In: Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE. pp. 2829–2832.

Niedermeyer, E., da Silva, F. L., 2005. Electroencephalography: basic principles, clinical applications, and related fields. Lippincott Williams & Wilkins.

Pereda, E., Quiroga, R. Q., Bhattacharya, J., 2005. Nonlinear multivariate analysis of neurophysiological signals. Progress in neurobiology 77 (1), 1–37.

Räsänen, O., Metsäranta, M., Vanhatalo, S., 2013. Development of a novel robust measure for interhemispheric synchrony in the neonatal eeg: Activation synchrony index (asi). Neuroimage 69, 256–266.

Stevenson, N., Korotchikova, I., Temko, A., Lightbody, G., Marnane, W., Boylan, G., 2013. An automated system for grading eeg abnormality in term neonates with hypoxic-ischaemic encephalopathy. Annals of Biomedical Engineering 41 (4), 775–785.

Stevenson, N., O'Toole, J., Korotchikova, I., Boylan, G., Aug 2014. Artefact detection in neonatal eeg. In: Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE. pp. 926–929.

Temko, A., Thomas, E., Marnane, W., Lightbody, G., Boylan, G., 2011. Eeg-based neonatal seizure detection with support vector machines. Clinical Neurophysiology 122 (3), 464–473.

Uhlhaas, P. J., Singer, W., 2006. Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology. Neuron 52 (1), 155–168.

Van Putten, M. J., 2006. Extended bsi for continuous eeg monitoring in carotid endarterectomy. Clinical Neurophysiology 117 (12), 2661–2666.

Van Putten, M. J., 2007. The revised brain symmetry index. Clinical Neurophysiology 118 (11), 2362–2367.

Van Putten, M. J., Peters, J. M., Mulder, S. M., de Haas, J. A., Bruijninckx,

C. M., Tavy, D. L., 2004. A brain symmetry index (bsi) for online eeg monitoring in carotid endarterectomy. Clinical Neurophysiology 115 (5), 1189–1194.

Van Putten, M. J., Tavy, D. L., 2004. Continuous quantitative eeg monitoring in hemispheric stroke patients using the brain symmetry index. Stroke 35 (11), 2489–2492.

Yaghoobi, M., Blumensath, T., Davies, M. E., 2009. Dictionary learning for sparse approximations with the majorization method. Signal Processing, IEEE Transactions on 57 (6), 2178–2191.

Yang, M., Zhang, D., Feng, X., 2011. Fisher discrimination dictionary learning for sparse representation. In: Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, pp. 543–550.

Zhang, Q., Li, B., 2010. Discriminative k-svd for dictionary learning in face recognition. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, pp. 2691–2698.