UNIVERSITY OF SOUTHAMPTON

SIGNAL PROCESSING AND CONTROL GROUP

Institute of Sound and Vibration Research

Advances in Epileptic Seizure Onset Prediction in the EEG with ICA and Phase Synchronization

by

Disha Gupta

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

April 2009

Acknowledgements

This research has been a challenge from the very start, and I would like to thank all the people who contributed and helped in its completion. I especially thank my supervisor Dr. Christopher James for the opportunity he offered me to perform this research and his support, academic guidance and encouragement throughout the study. He has helped refine many a skills and I am grateful to have been working with him. I thank him for the motivation and opportunities he gave me for presenting this work at international conferences.

I thank Prof. William Gray for supporting the clinical aspect of the research, especially the data collection. I acknowledge everyone at the Clinical Neurophysiology Dept. of Southampton General Hospital, for their understanding and hospitality during the data collection. I especially thank Jon Flannery for his time, unending support and enthusiasm for the same. I thank Dr. Richard van der Star for the discussions and support he offered. I am thankful to the R&D at SUHT for facilitating the ethical approval process to use human physiological data.

I thank Prof. Paul White for the reviews, feedback and discussions at the various stages of the research process. I also thank Dr. David Simpson for the feedback, discussions and the biomedical courses.

I further thank the Institute of Sound and Vibration Research (ISVR)(University of Southampton) and the Life Sciences Interface for funding this research. The staff at the ISVR has been fabulous, always ready to assist with administrative and technical computing matters, especially Maureen and Joyce.

I wish to thank all my friends and colleagues at the University, June, Suogang, Joao, Aida, Norma, Charmaine, Jens, Su Yeon, Emery, Cristobal, Paulo, Benny and Neven for their support and company during the ups and downs of the PhD. I am grateful to Dr. Anna Barney the head of Foundation Year of School of Engineering for giving me the opportunity to work as a teaching assistant for three years. The experience has left profound marks, opening my heart to the joy of teaching and helping me in many ways to be a better teacher and a student. It has been great to know and work with Sandra, Dr. Martin Counihan, Ollie, Francois, Joao, Parisa and to have found beautiful friends in Dr. John Mills and Dr. Patricia Mills.

Finally, my deepest gratitude goes to my family for their help and moral support. I thank my parents Dr. Pawan Gupta and Dr. Promila Gupta, my sister Shaloo and brother Abhishek for their confidence in my abilities and constant encouragement. I thank them for instigating in me the strength to look beyond the limits of possibilities and explore the realms of my dreams.

To my parents, who stood by my travels
across the rough paths I chose, urging me to look up and ahead.
Your hard work and sacrifices have been
and will be my inspiration forever.

I am a wanderer, belong to many, and to none.
Have a heart that speaks, and a mind that sings,
to myself a silent song.

Dreams stage a play, pot-pouri of all,
that went in the head, I never could recall.

Eyes wide shut, mind wanders astray,
to make a fable, book of rules per say.

I'd let the heart speak, and the mind run,
I'd dare and strive in the silent song.
I am a wanderer, belong to many and to none.

- Disha

Declaration of Authorship

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

STATEMENT 1

This dissertation is being submitted in partial fulfilment of the requirements for the degree of a PhD in the Institute of Sound and Vibration Research at University of Southampton.

STATEMENT 2

This dissertation is the result of my own independent work/investigation/analysis, except where otherwise stated. Other sources are acknowledged by citations, giving explicit references (appended).

STATEMENT 3

I hereby give consent for my dissertation, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organizations.

_____

Disha Gupta

_____

Date

Abstract

Seizure onset prediction in epilepsy is a challenge which is under investigation using many and varied signal processing techniques, across the world. This research thesis contributes to the advancement of digital signal analysis of neurophysiological signals of epileptic patients. It has been studied especially in the context of epileptic seizure onset prediction, with a motivation to help epileptic patients by advancing the knowledge on the possibilities of seizure prediction and inching towards a clinically viable seizure predictor.

In this work, a synchrony based multi-stage system is analyzed that brings to bear the advantages of many techniques in each substage. The 1st stage of the system unmixes and de-noises continuous long-term (2-4 days) multichannel scalp Electroencephalograms using spatially constrained Independent Component Analysis. The 2d stage estimates the long term significant phase synchrony dynamics of narrowband (2-8 Hz and 8-14 Hz) seizure components. The synchrony dynamics are assessed with a novel statistic, the PLV-d, analyzing the joint synchrony in two frequency bands of interest.

The 3rd stage creates multidimensional features of these synchrony dynamics for two classes ('seizure free' and 'seizure predictive') which are then projected onto a 2-dimensional map using a supervised Neuroscale, a topographic projection scheme based on a Radial Basis Neural Network. The 4th stage evaluates the probability of occurrence of predictive events using Gaussian Mixture Models used in supervised and semi-supervised forms.

Preliminary analysis is performed on shorter data segments and the final system is based on nine patient's long term (2-4 days each) continuous data. The training and testing for feature extraction analysis is performed on five patient datasets. The features extracted and the parameters ascertained with this analysis are then applied on the remaining four long-term datasets as a test of performance. The analysis is tested against random predictors as well. We show the possibility of seizure onset prediction (performing better than a random predictor) within a prediction window of 35-65 minutes with a sensitivity of 65-100% and specificity of 60-100% across the epileptic patients.

Abbreviations

AED: Anti-Epileptic Drugs

BSS: Blind Source Separation

BOLD: Blood Oxygenation Level

cICA: Constrained ICA

C(r): Correlation Integral

Contra: Contralateral (away from focus)

DWT: Discrete Wavelet Transform

EEG: Electroencephalogram

ECG: Electrocardiogram

EMG: Electromyogram

fMRI: functional Magnetic Resonance Imaging

FIR: Finite Impulse Response

FPR: False Positive Rate

FP: False Positives

FN: False Negatives

FFT: Fats Fourier Transform

GMM: Gaussian Mixture Model

HT: Hilbert Transform

ICA: Independent Component Analysis

ICs: Independent Components

Ictal: EEG of an epileptic patient having a seizure

Inter-ictal: Seizure free EEG of epileptic patient

ILAE: International League Against Epilepsy

Ipsi: Ipsilateral (near focus)

LDC: Least Dependent Component

MEG: Magnetoencephalogram

min: Minutes

MRI: Magnetic Resonance Imaging

MTLE: Mesial Temporal Lobe Epilepsy

Neuroscale: Radial Basis function Neural Network

PET: Positron Emission Tomography

PCA: Principal Component Analysis

PLS:Phase Locking Statistic

PLV: Phase Locking Value

PLV-d: Phase Locking Value-difference

REC: Research Ethics Committee

ROC: Receiver Operating Characteristics

SPECT: Single Positron Emission Tomography

SVD: Singular Value Decomposition

TDSEP: Temporal Decorrelation algorithm

TN: True Negatives

TP: True Positives

VNS: Vagus Nerve Stimulator

Symbols

A: Mixing Matrix

d: distance

f: Frequency

Hz: Hertz

m: mili 10^{-3}

s: seconds

S: Source signal

S: Sensitivity

Sp: Specificity

t: Time

V: Volts

W: Unmixing Matrix

x: Time domain signal

α: Control parameter for extent of supervision in learning

ϕ: Phase

μ: micro 10^{-6}

ω: Angular frequency

2-D: Two dimensional

C3,Cz,C4: Central (left/medial/right)

F3,F7/Fz/F4,F9: Frontal (left/medial/right)

Fp1/Fp2: Fronto-polar (left/right)

P3,Pz,P4: Parietal (left/medial/right)

T3,T5/T4,T6: Temporal (left/right)

O1/O2: Occipital (left/right)

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Epilepsy is a common neurological disorder affecting 1% of the world population. Its manifestation is seen as random and (often) recurrent seizures (a sudden attack usually lasting a few minutes associated (sometimes) with convulsions/loss of consciousness/loss of control/aggression). For centuries, the lack of knowledge about the cause and cure of epilepsy has led to its association with curses and other superstitious beliefs adding to the physical and emotional problems related to it. Today, with the ongoing research in various fields such as microbiology, genetics, neurology, neurosurgery and brain-imaging, there is sufficient understanding of epilepsy as a neurological disorder and help is available in many forms such as anti-epileptic drugs (AEDs), surgery, and vagus nerve stimulators. But the real cause and cure of this disorder is still unexplained and is under research. The present treatment methods aim at controlling, preventing or reducing the number of seizures but do not cure epilepsy. Also because of the vast variation seen in the way the disorder manifests in different people, and sometimes even across seizures in a single person, the treatment becomes restricted. Nevertheless, two-thirds of the patients can be treated by AEDs. Another 8% may benefit from surgery. But 25% of people with an epilepsy disorder continue to have seizures and no treatment may be suitable for them. The main problem with seizures is their apparent unpredictability, the random occurrence of the seizure(s) can restrict involvement of the patient in many activities, leading to anxiety, depression or feelings of loss of control. It may affect their social and professional life and thus have an indirect adverse impact on the quality of life of the sufferer. It may also be life threatening in the severity of the disorder or due to circumstances that occur at the time of seizure. Until research yields more information about the underlying dynamics/causes of this disorder in order to treat it, current research aims at advancing the knowledge of seizure onset prediction techniques.

This study aims to look into seizure prediction with a novel approach, based on a combination of phase synchronization and a state-of-the-art signal processing technique: Independent Component Analysis. The idea is to track synchrony dynamics across

time in the underlying independent signals extracted from the mixed brain activity signals (the EEG). Such a system would be useful as a warning system to assist the wearer in getting appropriate help or move away from a dangerous situation. It could also be used in diagnosis or further research where data collection requires tracer injection in a patient before a seizure onset, in order to capture brain activity at the time of a seizure. Alternatively such a prediction could be a step forward for an advanced treatment mechanism for epilepsy that may help to halt seizures by inhibiting synchronized neuronal activity [14, 15]. It could also be the precursor to a deep brain stimulation device [16, 17], a rapid localized cooling mechanism [18, 19] or a system that could infuse drugs before a seizure onset [20, 21]. These ideas seem plausible and are mentioned by researchers across different aspects of science and technology. Preliminary efforts translating seizure prediction into implantable devices was carried out by B.Litt *et al* [22], although it is in need of much detailed research in many areas of engineering and medicine for such a complete system to be realized.

Nevertheless, such a predictor would be required to have a practically useful prediction time, allowing for a pharmacological or electrotherapeutic system to intervene. The appropriateness of the prediction time would be quite dependent on the intervention system. For example, an intervening pharmacological system that automatically infuses a dose of anticonvulsant drugs before seizure onset, would depend on the minimum time the medication (if it exists) would require to act and curb a seizure. If the side effects of the drug are milder, false alarms would be more tolerable and a wider prediction window might suffice. If the drug side effects are strong, a shorter and more specific prediction horizon might be sought. In the case of a electrotherapeutic system such as electrical deep brain stimulation it might be critical to assess the prediction time and specificity depending on the strength of the stimulation and the corresponding effects induced.

The aim of this research is advancement of knowledge for seizure prediction using long term continuous brain activity. To date seizure prediction methods have been applied on short term brain activity and are in need of a specificity analysis along with the reported sensitivity for a clinically viable prediction system. It is understood that such a system would ideally be expected to perform in real time, have a low rate of false positives, high sensitivity and a practically feasible prediction horizon. The difficulty in generalizing the manifestation of epilepsy across different patients or even in one patient, motivates a patient specific model and at most a generalized model that will essentially need to be adapted to each patient.

The work in this thesis is based on continuous long-term multichannel brain activity: the electroencephalogram (EEG). The analysis is based on a novel approach for prediction analysis: tracking synchrony dynamics across time on underlying independent signals from the brain as opposed to raw EEG signals. Previously various methods have

been used for prediction, such as non-linear dynamics, complexity, correlation dimension, similarity, accumulated energy, etc. but usually on univariate or bivariate signals and on short term EEG signals. We have applied a state-of-the-art Blind Source Separation (BSS) technique called Independent Component Analysis (ICA) on EEG signals as a pre-processing step. The ICA technique helps to extract underlying independent signals mixed in the EEG, as well as cleans it of artifacts and noise. Furthermore, constrained ICA has been used to refine this search by placing spatial / temporal limits to the signal(s) being located. ICA has been combined in a novel way with phase synchronization analysis to track the synchrony dynamics in signals extracted from a spatially constrained unmixing algorithm. Synchrony analysis of unmixed signals has not been performed previously for prediction analysis in epileptic EEG.

Phase synchronization has been used as the basis of analysis because previous research in medical and engineering fields has presented a promising hypothesis, involving transitions in strength of synchrony among the neuronal networks in an epileptic brain as it prepares for a seizure. It is hypothesized that in epilepsy the normal synchronization among neurons is affected as the brain gets ready for a seizure. The neurons near the seizure focus tend to lose synchronization with other cortical areas, isolating the focus and providing an idle population of neurons, facilitating the development of a hyperexcitable state. It is such spatial and power transitions in synchronization that are being tracked using independent source signals from the multichannel EEG. Theoretically the use of phase synchronization in independent signals appears contradictory, as independent signals should ideally be free from dependence of any form (synchronization being a form of dependence). But in practice when real life complex multichannel biomedical signals such as EEG are used, remnant dependencies have been observed to be present [23, 24, 25]. One of the reasons being: linear ICA is not able to remove nonlinear dependencies, which are essentially captured by phase synchronization (a measure of nonlinear correlations). Nevertheless, ICA helps to reach closer to the underlying sources in the brain as opposed to using mixed raw EEG signals.

The data used here is short term and long term continuous multichannel scalp EEG of patients with epilepsy. The patients are of various ages, and have various types of epilepsies. The short term recordings have been used for preliminary research work. The long term recordings consist of 3-4 days of continuous multichannel EEG for 10 patients with epilepsy (being monitored for presurgical evaluation). The phase synchrony has been measured for pairs of contralateral and ipsilateral spatially constrained IC's. These have shown interesting features such as a possible underlying circadian-like rhythm, certain patterns that appear to indicate seizure onset on visual subjective evaluation of various patient results. In order to make the process objective, feature selection was carried out on 5 data sets using artificial intelligence based classification algorithm: a Radial Basis Neural Network (Neuroscale) and Gaussian Mixture Probability Modeling. Neuroscale was used for visualization and topographic dimensional

reduction followed by Gaussian Mixture Models (GMM) applied to the dimension reduced data set, to find the probability of prediction across time. Features that show a prediction possibility of about 35-65 minutes were found with 65-100% sensitivity and 65-80 % specificity. The specificity and sensitivity results have been tabulated along with ROC curves in Chapter 8.

As parts of the research process, there were few diversions (some resulting in publications), such as acquiring long term human epileptic EEG data with ethical approval, selection of seizure sources objectively using temporal and topographical information, de-noising EEG objectively, selection of constraints for spatially constrained ICA, using narrow band vs. broad band signals for phase synchronization and using statistical means to check the significance of phase synchrony. These have been summarized in following chapters.

1.1 Thesis Structure

Chapter 1 introduces the problem and gives an overview of the approach used to analyze it. Figure 1.1 shows the major steps of analysis in this research work and each is covered in the corresponding numbered Chapter. Chapter 1 is followed by a detailed description of the system being studied: Epilepsy and the brain. Chapter 2 also describes the multivariate brain signals, the 'normal' and the 'epileptic' EEG, which are being used for the present analysis. Chapter 3 gives an overview of the various analysis techniques with their advantages and disadvantages, that have been used previously in literature, for epileptic seizure onset prediction. Chapter 4 describes the pre-processing for EEG using the BSS technique of ICA. The ICA technique, algorithms, assumptions and limitations have also been summarized. Chapter 4 also shows the preliminary results for un-mixing and de-noising with ICA for epileptiform EEG. This is followed by the assessment of synchrony dynamics covered in Chapter 5. Synchrony dynamics have been assessed using phase synchronization, which has been explained along with the analysis of the parameters involved. It also covers the test of phase synchrony significance based on surrogate analysis. The procedure followed in this research using ICA and phase synchrony for seizure detection and tracking synchrony dynamics has also been explained in Chapter 5. The preliminary analysis performed for seizure prediction using short term multichannel EEG data sets has been described in Chapter 6. It includes the use of phase synchrony and ICA for de-noising multichannel EEG and its use for objective selection of seizure sources. The long term synchrony dynamics have been individually analyzed for each patient data set in its usefulness in the context of seizure detection/prediciton. The feature extraction procedure used for the analysis of the synchrony dynamics obtained in Chapter 7 is described in Chapter 8. This chapter also summarizes the results of the research on seizure prediction. Chapter 9 concludes the thesis with discussion of this research work, and poses the questions that would need further research.

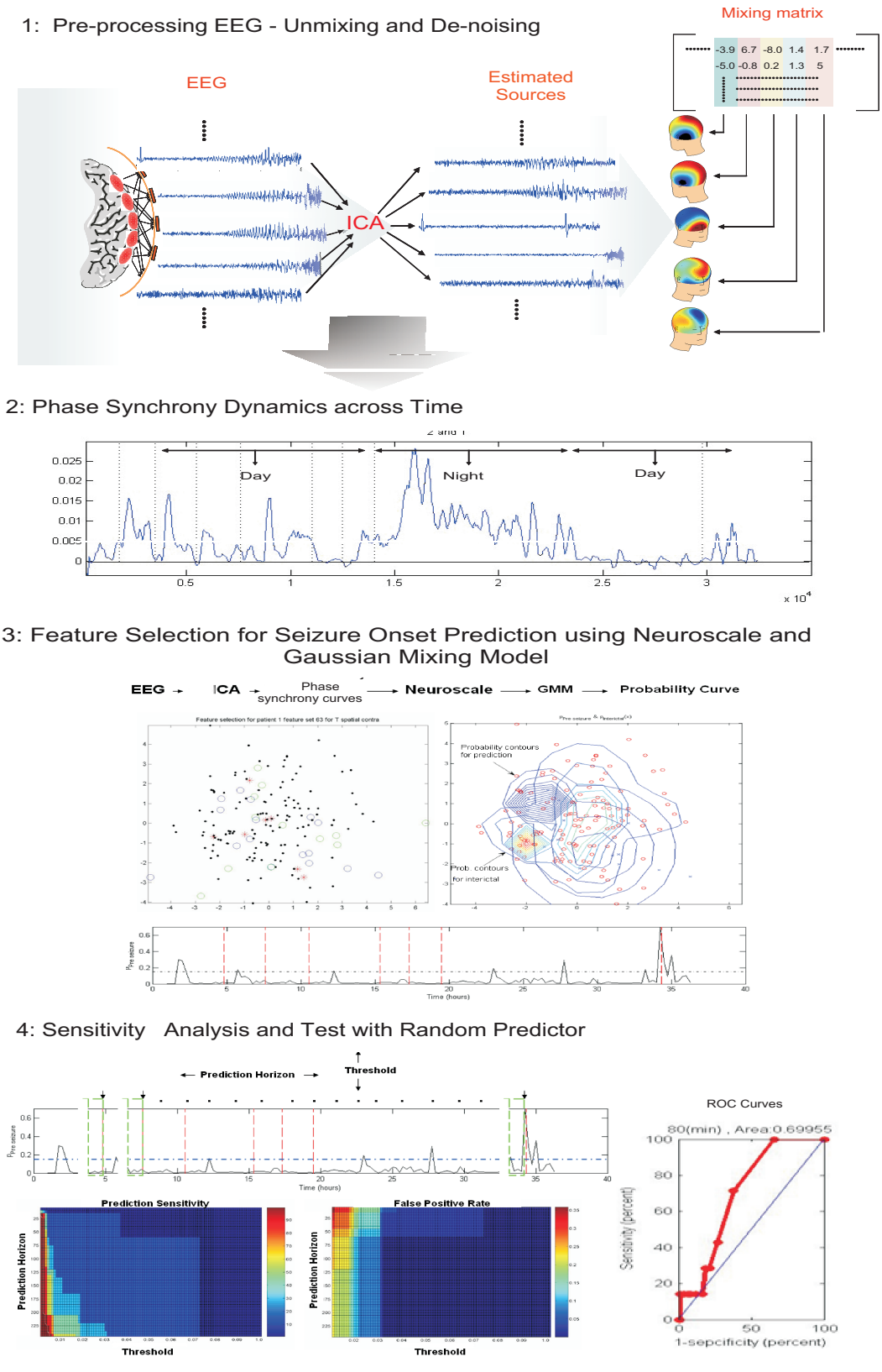


Figure 1.1: This figure summarizes the main steps in this research on seizure onset prediction using ICA and phase synchronization. (1) Preprocessing of continuous multichannel EEG data: Unmixing and De-noising (2) Tracking synchrony dynamics across time in the brain with phase synchronization and ICA (3) Feature selection from synchrony time series based on Neuroscale and Gaussian Mixing Probability Model (4) Sensitivity analysis and ROC curves for various prediction horizons. (This is finally compared with a random predictor).

Chapter 2

Background: The Electroencephalogram and Epilepsy

2.1 Epilepsy: An introduction

"Epilepsy : The most common serious brain disorder worldwide with no age, racial, social class, national nor geographic boundaries." - World Health Organization(WHO)[26].

For many centuries epilepsy has been regarded as an inherited illness with an unknown cause that does not have a cure, although this is not an entirely true statement. As mentioned by [27] epilepsy is not a disease. It is a chronic recurrence of sudden unpredictable abnormal reactions of the brain in the form of *epileptic seizures*. It may involve the entire brain or parts of it, and the extent of involvement largely determines the type of seizure. The seizure may affect how a person feels or acts for a short time. Some seizures can be hardly noticed while some can be totally disabling. As mentioned by the World Health Organization,"There are over 50 million sufferers in the world today and an estimated 2.4 million new cases occur each year globally." The International League Against Epilepsy (ILAE)[28] states: 456,000 or one in every 131 people in the UK has epilepsy; only 52 % of people with epilepsy in the UK are seizure-free; there are only approximately 29 specialist epilepsy clinics in the country - 14,000 patients per clinic. Compared to the 245 specialist diabetes clinics, each catering for only 800 people with insulin dependent diabetes; only about 8 out of 150 neurosurgeons in the UK specialize in epilepsy.

The cause of this seemingly 'strange' phenomenon remained unexplained for many decades and epilepsy got associated with curses, possessions, punishment of past sins and the supernatural. In around 400 BC the Greek physician Hippocrates wrote the first book on epilepsy, titled "On the Sacred Disease". He recognized that epilepsy was a brain disorder and spoke out against the ideas that seizures were a curse from the Gods and that people with epilepsy held the power of prophecy. But the strong belief

of curses and demons was not given up easily. Many myths and misunderstandings prevailed even in early 19th century when people with an epileptic disorder were cared for in asylums and seizures were considered contagious. Major work on epilepsy began in the mid-1800s, by three English neurologists: Russell Reynolds, John Hughlings Jackson and Sir William Richard Gowers. In a study, Jackson defines a seizure as an occasional, an excessive, and a disorderly discharge of nerve tissue on muscles. In 1909 the International League Against Epilepsy (ILAE) was founded with chapters in 60 countries and was followed by the International Bureau for Epilepsy (IBE) which had about 55 national chapters. The first drug for helping people with an epileptic disorder was created in 1912. The research in this field gained momentum after 1929, when a German psychiatrist called Hans Berger discovered a non-invasive way to record and graphically depict the electric currents generated in the brain. This method was called electroencephalography (EEG). In 1997 the ILAE and IBE joined with the WHO to establish the Global Campaign Against Epilepsy and aimed at improving awareness, treatment, care and services for people with epilepsy disorder. Along with these efforts the continuous research and development in the fields of human brain and brain imaging, in the past century, has helped to provide a lot of knowledge about epilepsy, although a lot remains to be learned in areas of the underlying cause of epilepsy, seizures, better treatment methods, medication.

Today, it is known that a seizure can be caused by a number of temporary or permanent brain or cell function disorders and be facilitated by the presence of a genetic predisposition. Thus 'epilepsy' is better known as an epileptic seizure disorder. Brain damage acquired during the earliest years of life is a major cause of epileptic seizure disorder. Epilepsy can also start at any later age due to trauma, brain tumor, space occupying lesions (abscesses, haematomas), chronic alcoholism, infection in central nervous system, etc. The onset, spread and clinical manifestation of a seizure can vary vastly from person to person and even among seizures experienced by an individual. This can make its diagnosis and treatment complicated.

2.1.1 Symptoms of an Epileptic Seizure

Some of the symptoms of a seizure can be very dynamic. They vary from person to person and sometime even from seizure to seizure. Some people experience a distinctive feeling when a seizure is about to happen, this warning sign is called an aura. It is usually in the form of a certain taste, smell, bodily sensation, etc. and may help the patient by giving time to prevent any self harm. The seizure may involve semiconscious or confused actions. The common signs observed before, during and after an epileptic seizure are as follows.

Warning signs (aura) before an epileptic seizure: Sensory: *deja vu*, smell, sound, taste, visual loss or blurring, racing thoughts, stomach sensations, tingling sensation;

emotional: fear/panic, pleasant feeling; physical: dizziness, headache, light-headedness, nausea, numbness.

Signs during a seizure: Sensory: black out, confusion, deafness/sounds, feeling of electric shock , loss of consciousness, spacing out, out of body experience, visual loss or blurring; emotional: fear/ panic; physical: chewing movements, convulsion, difficulty talking, drooling, eyes rolling up, falling down, foot stomping, hand waving, inability to move, incontinence, lip smacking, making sounds, shaking, staring, stiffening, swallowing, sweating, teeth clenching/grinding, tongue biting, tremors, twitching movements, breathing difficulty, heart racing.

Signs after an epileptic seizure: Sensory: Memory loss, writing difficulty; emotional: Confusion, depression and sadness, fear, frustration, shame/embarrassment; physical: Bruising, difficulty talking, injuries, sleeping, exhaustion, headache, nausea, pain, thirst, weakness.

2.1.2 Types of Epileptic Seizures

Classifying epileptic seizures has been a difficult task because of the amount of variability involved in the disorder. An international Commission of classification and terminology gave the following classification of epileptic seizures in 1981 which has been followed internationally; it was re-iterated in 1994. This classification considers two main broad types of seizures, as follows [27]:

1. Partial (focal) seizures: This type of seizure is characterized by abnormal electrical activity in a small part of the brain, but it may spread to other parts as well.
   (a) Simple partial (consciousness not impaired)
   (b) Complex partial seizures (with impairment of consciousness)
      i. Simple partial onset followed by impairment of consciousness
      ii. With impairment of consciousness at onset
   (c) Partial seizure evolving to secondarily generalized seizures
      i. Simple partial seizure evolving to generalized seizures
      ii. Complex partial seizure evolving to generalized seizures
      iii. Simple partial seizure evolving to complex partial seizure evolving to generalized seizures
2. Generalized seizures (convulsive or non convulsive): This type seizure involves abnormal electrical activity in both sides of the brain. The seizure may start in one area of the brain and then spread across the brain.
3. Unclassified epileptic seizures: These are subdivided into many more classes based on features of the seizure.

2.1.3 Treatment Available for Epilepsy Disorders

According to the information from the WHO, up to 5% of the world's population may have a seizure at some time in their lives. But fortunately, now 70% of the people newly diagnosed with epilepsy seizure disorder can be successfully 'treated'. The treatment can be of many forms depending on the type of disorder. Epilepsy treatment does not focus on curing the cause of epilepsy but more often is in the form of suppression of the seizure. A brief overview of the treatment methods is as follows:

- **Antiepileptic drugs (AEDs)**[29] The cause of epilepsy seizures is still unclear, thus medications that have been developed mainly aim to prevent or reduce the frequency of seizures. AEDs cannot cure epilepsy but they help in control over seizures as long as the medication is taken regularly. The first main drug to be used for epilepsy was in 1912 called Phenobarbital and is still in use today. But its usefulness is limited to generalised tonic-clonic seizures and has a sedative effect on some people. It was followed by Phenytoin in 1938. Between 1960 and 1974 only one new drug Diazepam (Valium) was approved to be used in epilepsy. In 1974, Carbamazepine (Tegretol) was approved, and in 1978, Valproic acid (Depakene). AEDs like any drug has side effects too, usually dose related. AED side effects include: sedation, irritability, nausea, rash, and clumsiness. Some drugs produce changes in emotions, memory or behavior, or affect learning. Another side effect is allergic reaction to medicine which may even lead to fatal liver damage. None of the drugs have been proven safe to be used during pregnancy. Since 1978, a series of drugs have been developed like Felbatol, Lamictal, Topamax, Gabitril, Keppra, etc. aiming for lesser side effects.
  - Success rate: It has been found that AEDs help in control in more than half the epilepsy patients and reduce the number of seizures for another 20-30 percent people. But medication is usually continued for many years (sometimes life long) and can have numerous side effects. Chronic daily intake of medication sometimes develops tolerance to anticonvulsant effect. About 20 percent of people do not respond to the medication and their condition is called 'intractable' or 'refractory to treatment'.
- **Ketogenic diet** [29] The Ketogenic diet is one of the oldest forms of treatment for epilepsy. Though it is yet unexplained as to how this particular diet helps, it has been found to be effective in some epileptic disorders, since the 1920s. The Ketogenic diet is a diet with high-fat, low-protein and negligible amount of carbohydrate and was noted to have anticonvulsant effects, similar to starvation. This diet is maintained for about 2 years. The diet is indicated for use primarily in young children with a certain type of epilepsy disorder. But this treatment was rarely used after AEDs were developed.

- Success rate: 30 to 50 percent of children respond to this diet. Those who respond show dramatic improvement, with 50 percent reduction in seizure frequency within two to three weeks.

- **Vagus nerve stimulation (VNS)** [30, 31] VNS is a treatment system which has been found to be effective for some seizure types. The underlying mechanism of its action is yet unclear. It is designed to prevent seizures by sending regular (mild) pulses of electrical energy to the brain via the vagus nerve (the vagus nerve is a part of the nervous system that controls the involuntary functions of the body (like heart rate)). The pulses are provided by a pacemaker like generator that is surgically implanted in the chest wall and the vagus nerve is reached by an electrode that is inserted through an incision in the neck. VNS was approved by U.S. Food and Drug Administration (FDA) in 1997 for use in combination with AEDs in partial epilepsy and was commercialized by Cyberonics, Inc. [31] on obtaining the FDA approval.

  - Success rate: Its success is comparable to AEDs with a seizure reduction of about 50 percent in 40 percent of patients.

- **Surgery** [32] With the continued development of diagnostic techniques and surgical equipment and number of surgeons becoming involved with epilepsy surgery, surgery for epilepsy is increasingly gaining importance. Epilepsy surgery involves removal of the seizure causing part of the brain (if the seizure focus is localized and if its removal does not interfere with normal tasks of the brain) or preventing the spread of the seizure. Thus it can be done in following ways:

  - Selective amygdalo hippocampectomy: The removal of structures in the temporal lobe which are commonly the site of seizure activity.
  - Temporal lobectomy: The removal of a larger part of the temporal lobe.
  - Sub-pial resection: Making fine cuts in the motor areas of the brain that do not affect the motor function but prevent the spread of seizures.
  - Hemispherectomy: The removal of damaged side of the brain. It is used to treat very severe epilepsy in children with damage to one whole side of the brain.
  - Corpus callosotomy: Cutting the fibres connecting the two halves of the brain. This is also used to treat children with very severe epilepsy.
  - Removal of a lesion such as a tumor or a cyst.

  Surgery is a much better option compared to taking AED medication for many years. But it is possible only in specific types of epilepsy disorders and there is no guarantee that after the operation the seizures would stop. It also involves the risks common to a surgery. Before surgery the patient has to undergo a thorough pre-surgery evaluation. This may involve monitoring over days or weeks

with EEG, video, depth electrodes, MRI (magnetic resonance imaging) scans, PET (positron emission tomography) scans, SPECT (Single Positron Emission Tomography), etc. The risks involved are apart from the risks of any surgery, brain surgery may inflict damage to other brain functions, sometimes leading to paralysis.

  - Success rate: Very select hospitals are involved with such surgery and the total cost could be around $40,000-$50,000 in the U.S.A. Despite the difficult nature of brain surgery most patients make a rapid recovery and within 8-15 weeks can return to work.

- **Other alternative therapies:** These are unconventional, non medical therapies that focus on integration of the body as a whole. Some examples are: Herbal medicines, Reiki, Acupuncture, Yoga, Epilepsy dogs, Aromapathy, Homeopathy, Relaxation exercises, Biofeedback.

  - Success rate: Research in these areas is limited and there is little scientific evidence to support their effectiveness. Some people have felt that these have improved their quality of life and hence such therapies cannot be completely ignored. It can only be said that a scientific statement to their affect is still in need of more research and analysis.

2.2 The Brain: Introduction

The human brain, weighing about three pounds, is the most complex part of the human body. It is understood as the center of control of most of the human body's biological, physical and emotional functions and reactions. The brain is made up of many cells (about 10^{13}), only 10% of these cells (about 10^{12}) are involved in processing and communicating information. These integral units are called neurons (the nerve cells) and work in a complicated but coordinated chemical and electrical system to transmit and process information. Inside the neuron the information travels in the form of an electrical signal and information is communicated across neurons through release of chemicals at specialized junctions called synapses.

The brain is enclosed within the skull and is surrounded by a protective dense layer of fibres called dura matter. The brain weighs a significant amount but has no skeletal support. To overcome this it is immersed in a fluid called the cerebrospinal fluid. This fluid allows the brain to float and offsets the pressure and effects of shocks which would otherwise be destructive. It mainly consists of three main parts: cerebrum, cerebellum and the brainstem.

2.2.1 Basic Anatomy

Cerebrum The cerebrum is the largest part of the brain. It is divided into two hemispheres called the left and right cerebral hemispheres. These hemispheres appear symmetrical and are connected by a commissure (bundle of neuronal axons connecting left and right side of the brain) called corpus collosum. Each cerebral hemisphere has a covering of highly folded structure called the *cerebral cortex*. The cerebral cortex is the critical structure which is a 3 mm thick layer of nerve cells and their connecting pathways and is involved with important functions as perception, movement, language and reasoning. The folded structure of the cortex helps in packing more cortical surface in a smaller space and also brings some regions closer thus reducing neuronal conduction times between different areas. The smoothed out human cortex is about $2200cm^2$ to $2400cm^2$. The cortex contains high density of the neuron cell bodies, their dendrites and some axons. This gives it a greyish appearance and is referred to as the grey matter, as compared to the underlying regions containing mainly the myelin coated axons of the neurons which are called the white matter. This cerebral cortex is divided into four different sections called 'lobes' namely: frontal, parietal, occipital and temporal. These lobes have been shown in Figure 2.1 . The lobes are associated with different tasks of the brain as mentioned in Table 2.1 . Each sensory field of the cerebral cortex like as auditory or visual, contains a topographic map (spatial layout) of its receptor surface.

Lobe	**Task**	**Sensory field**
Frontal	Movement, reasoning, planning, emotions, problem solving	Motor
Parietal	Movement, orientation, recognition , perception of stimuli	Somatosensory
Occipital	Visual processing	Visual
Temporal	Memory, speech, recognition of auditory stimuli	Auditory

Table 2.1: The different tasks associated with different lobes of the human brain

Cerebellum The cerebellum also known as the little brain is divided into hemispheres like the cerebrum and is covered by a similar cortex. It is mainly associated with coordination of movement, posture and balance. By itself the cerebellum does not directly control movements but it integrates information about the body and motor commands and allows smooth coordinated movements.

Brain stem The brain stem consists of the midbrain, pons and medulla. These are the continuation of the spinal cord in the brain and contain the fibres tracts connecting the brain and the spinal cord. The midbrain is involved in the control of movement. The pons has an important role in feeding behavior and facial expression. The medulla is concerned with several autonomic functions as respiration, cardiovascular, consciousness and gastrointestinal.

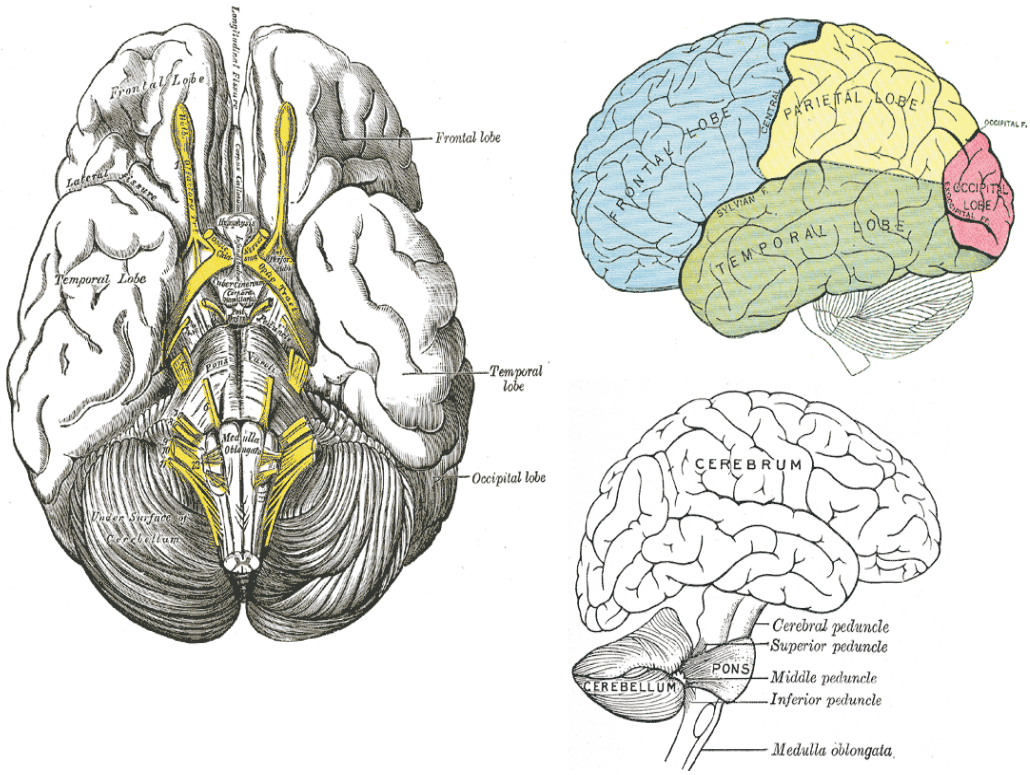


Figure 2.1: The left image shows the inferior view of the principal lobes of the brain, the images on the right show a lateral view of the principal lobes, structures and fissures of the cerebrum [1]

2.2.2 The Neuron and Signal Transmission

The elementary unit that makes up the brain is a neuron (Figure 2.2). The neuron also has structures called dendrites and axons as shown in Figure 2.2. Dendrites are tree like fibre extensions of the cell. Axon is also a fibre extending from the cell body like a stem of the neuron and its end is called a synapse.

Electrical Transmission Neurons create electrical signals using an electrochemical gradient based on ions. In the neuron there are 4 main ions which are significant: Sodium (Na^+), Potassium (K^+), Calcium (Ca^{2+}), and Chloride (Cl^-). A neuron is encapsulated in a semi-permeable (allowing only selective substances to cross) membrane. At rest the concentration of Na^+ and K^+ in a neuron is such that there is a potential of -55 mV to -70 mV across the neuron membrane. Certain stimuli can reduce the charge across the membrane to a certain threshold level (depolarization) which triggers influx of Na^+ ions through that portion of the membrane. This also triggers outflow of K^+ ions which is slower as compared to Na^+. With the flow of ions the potential keeps increasing and reaches its peak value and by that time the outflow of K^+ and inflow of Na^+ pulls the potential back to resting potential, stopping the flow of ions. On reaching resting potential the neuron membrane goes through a resetting

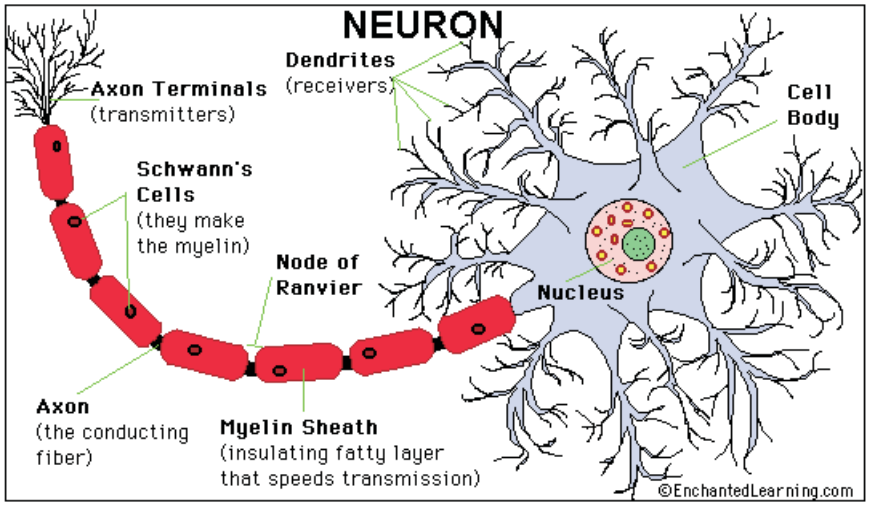


Figure 2.2: The basic neuron and its nucleus [2]

phase of about 0.001-0.002 seconds, called refractory period, preparing for a new action potential. The sudden large increase and a subsequent decrease in the membrane voltage is called an action potential. This spreads to adjacent membrane of the neuron leading to a spike traveling along the neuron axon (uni-directionally due to refractory phase) at a speed of about 100 m/s and the neuron is said to be firing. This process has been diagrammatically shown in Figure 2.3.

Synaptic transmission For neurons to communicate with other neurons and muscles they must transmit the electrical signal. This action is called synaptic transmission. It occurs at the synapse of the neurons by transmission of chemicals called neurotransmitters. The elements in this process are a presynaptic ending (from which neurotransmitter is released), a synaptic cleft (across which the transmitter is diffused) and a postsynaptic element (the receptor of neuron to which the transmitter binds). Usually the presynaptic ending is an axon terminal and the postsynaptic ending is the dendrite, though any part of the neuron can be presynaptic to any part of another neuron. Neurotransmitters are small molecules like amino acids which are packaged for release in the presynaptic endings in synaptic vesicles. Their release is triggered by increase in Ca^{2+} concentration in the presynaptic. The Ca^{2+} concentration is maintained by action potentials. Thus, when the action potential spreads in a neuron, it causes Ca^{2+} influx into the presynaptic ending, which leads to release of neurotransmitter from the presynaptic ending as shown in Figure 2.4. The amount of transmitter released in response to an action potential is dependent on the history of activity. Several processes (short-lived or long -lived) lead to certain controlled amount of transmitter release, which forms the basis of learning and memory. This neurotransmitter is diffused into the synaptic cleft and is taken up by the postsynaptic element of another neuron. The neurotransmitter then generates a response in the postsynaptic element. The response can be depolarizing (excitatory postsynaptic potentials (EPSPs)) or hyperpolarizing

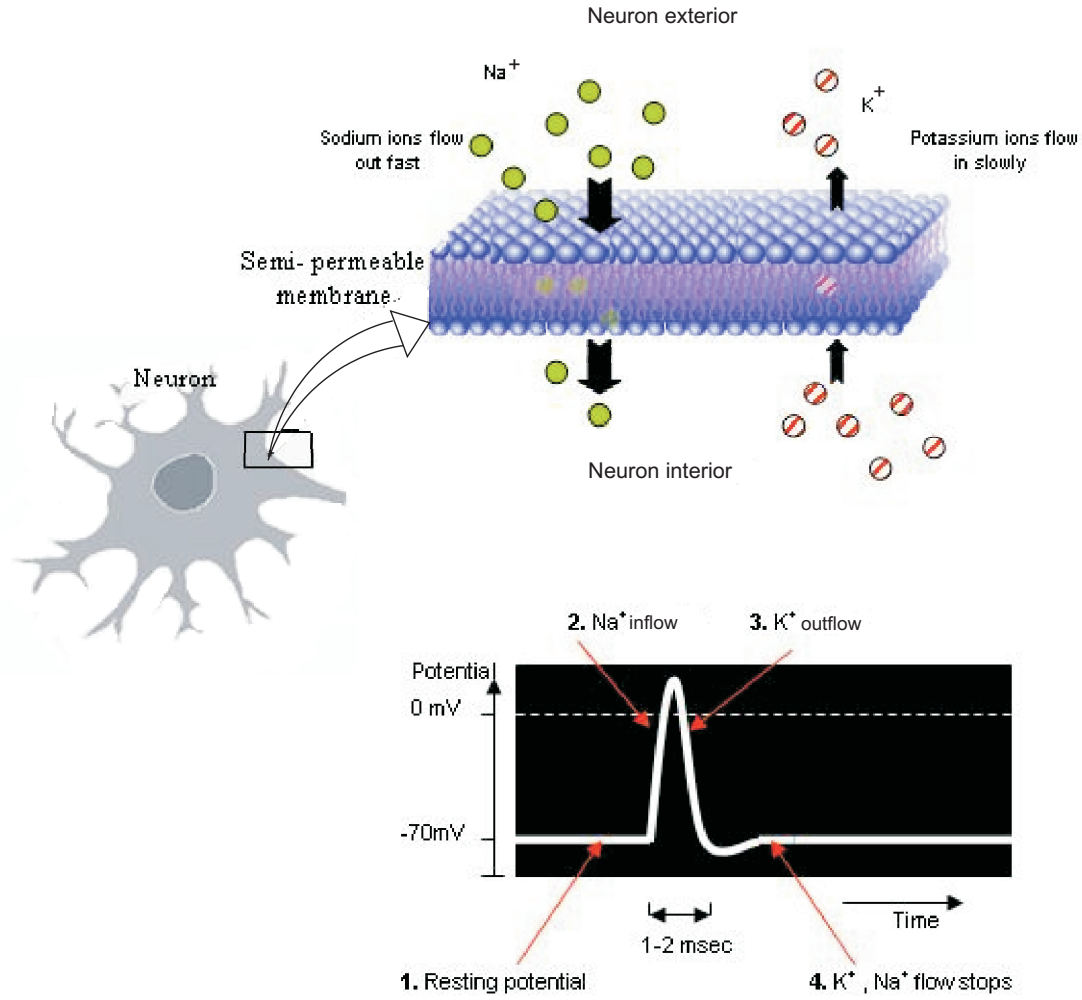


Figure 2.3: Generation of an action potential in a neuron in the brain

(inhibitory postsynaptic potentials (IPSPs)).

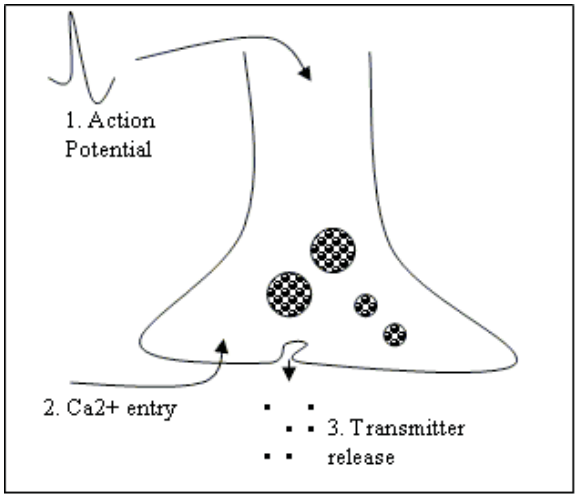


Figure 2.4: Transmitter release at the synapse when action potential spreads in the neuron causing calcium influx.

A single neuron may have thousands of other neurons synapsing to it. Some of these release polarizing neurotransmitters while some release hyperpolarizing neurotransmitters. The neuron integrates these inputs (temporally and spatially) and if the sum exceeds the threshold (about -20mV) an action potential is generated which travels down its axon. Thus an electrical signal gets communicated to a neuron. The synaptic transmission has been shown in Figure 2.5.

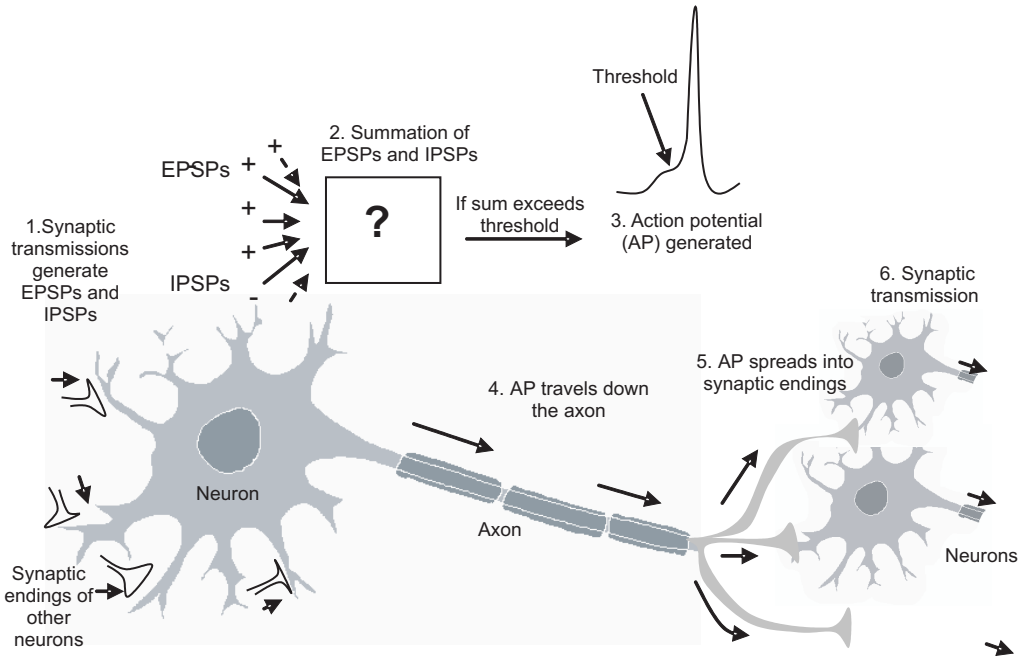


Figure 2.5: Synaptic transmission of an electrical signal

2.2.3 The Forward and Inverse Problem

The behavior of electric and magnetic fields created by neuronal currents are predicted to appear on the scalp obeying normal physical laws (also called the *forward problem*). These electric potentials or their corresponding magnetic fields are then measured on (or near to) the scalp to analyze the underlying brain activity.

The surface potential maps thus obtained are also used to solve for the *inverse problem*, that relates to finding the intracerebral sources (also referred to as *source localization* or *source estimation*) with a known potential distribution on the scalp. This is not a trivial problem and does not have a unique solution, which means different combinations of intracerebral sources can lead to the same potential distributions measured on the scalp.

2.3 Measuring Brain Activity

There are various techniques to measure the brain activity. They have been briefly discussed in Table 2.2 with respect to their measuring technique, advantages and disadvantages. Their temporal and spatial resolution is compared in Table 2.3. The selection of a technique to use for research depends on various criteria:

- The activity of interest (application)
- Temporal resolution required
- Spatial resolution required
- Feasibility of acquiring data
- Costs involved, etc.

For the present study high temporal resolution would be highly appropriate and good spatial resolution might also be helpful. The system should be practically feasible and economically viable. EEG and MEG have superior temporal resolution as compared to PET or fMRI. But their spatial resolution is not as good fMRI. This is basically due to the comparatively lesser number of spatial measurements channels (10's or 100's in EEG and MEG compared to 10's of thousands in PET or fMRI, even though the inverse problem can still not be solved exactly). [33]. The temporal resolution of MEG and EEG is very suitable for analysis of epilepsy. MEG has a considerably better spatial coverage as it has closely spaced sensors and more sensors (more than 250 as compared to 19-25 in standard EEG recordings). However MEG apparatus is available only in a few places. More so, for this study on epilepsy, the patient's seizure data is

Technique	Advantages	Disadvantages
EEG: measures electric field generated by large populations of neurons	• Less expensive compared to MEG, fMRI • Does not require much special apparatus • Can be used for 24 hour recordings for a single patient • Is non-invasive / invasive	Difficult to localize signal source as signals get smeared when passing through skull and tissue.
MEG: measures magnetic fields generated by large populations of neurons	• 3D source localization better than EEG as magnetic fields pass unaffected through tissue but this benefit gets cancelled due to larger distance between MEG sensors and the brain. • Is non-invasive	• Expensive set up • Not available easily • Using can be costly
fMRI: measures haemodynamic reflection of neural activity i.e. measure the local changes in blood flow which are caused by neural activity. It is sensitive to the blood oxygenation level (BOLD)	• More sensitive than PET • Is non-invasive	• Expensive apparatus • Poor temporal resolution (although it is improving)
PET: measure regional cerebral blood flow (CBF) which reflects neural activity.	• Is non-invasive	• Requires injection of radioactive tracers, which might be difficult to inject when person starts to have a seizure. • Has poor spatial and temporal resolution

Table 2.2: Table of advantages and disadvantages of various brain activity measurement techniques

required and this can be collected only if the patient has a seizure when MEG is being recorded. MEG requires the use of a specific solid enclosure around the head which can be very dangerous for a patient having a seizure. Thus, EEG has been selected for current analysis for the study and will be discussed in detail in the next section.

Technique	Spatial resolution	Temporal resolution
EEG	Low (about 20-30mm)	High (about msec)
MEG	Low (several mm)	High (about msec)
fMRI	High (about 1 mm)	Poor (about 2 sec)
PET	medium (few mm -1 cm)	Poor (1 min)

Table 2.3: Table of spatial and temporal resolutions for various brain activity measurement techniques

2.4 Electroencephalography

The EEG is a measure of the electrical potentials that are produced by the synaptic and axonal neural currents in the brain. The method to measure these currents was first described by Hans Berger in 1924. EEG provides a multichannel signal that provides both spatial information along with temporal cerebral information.

2.4.1 EEG Recording

The potentials are measured between electrodes placed in different known positions either on the scalp or inside the brain. They may be measured relative to a standard reference potential or between pairs of electrodes. The standard reference potential is assigned the value of zero. This potential is selected so that it is infinitely far away from the source of electrical activity - an 'inactive' electrode location. Ear, mastoid and chin locations are common referential potential locations for scalp electrodes. For recording a ground electrode is also used, and can be the forehead or ear [34].

2.4.1.1 Placement of electrodes

The electrodes can be placed on the scalp or inside the brain and are known as follows:

1. **Scalp surface-electroencephalography:** Scalp surface electrodes are small cup shaped electrodes of silver-silver chloride of about 0.5-1 cm in diameter as shown in Figure 2.6. These are held in place by an adhesive paste like collodion. A small amount of an electrolyte gel is inserted into the electrode cups using blunt needles. To allow exchange and interpretation of data across laboratories and to have uniformity in nomenclature, in 1958, the International Federation in electroencephalography and Clinical Neurophysiology adopted a standardization for electrode placement called 10-20 electrode placement system [4, 3, 5, 35] (see Figure 2.7 (a,b)). Even today it is the most commonly used system. This system standardized the physical locations and designations of the electrodes to be placed on the scalp. It divides the head into proportional distances (labeled as 10-20

Figure 2.6: Silver/Chloride scalp electrode

percentages) from prominent skull landmarks (nasion, inion, preauricular points) such as to provide adequate coverage of all regions of the brain. Electrodes are designated according to adjacent brain areas: F (frontal), C (central), P (posterior), and O (occipital). These letters are accompanied by odd numbers (for electrodes at the left side of the head) and even numbers (for electrodes at the right side of the head). (Left and right convention is from the point of view of the subject). Such a system can be replicated consistently and across laboratories. Also electrodes are available pre-positioned according to such a system and fitted in a cap. Though 10-20 system mainly refers to 21 electrodes, additional electrodes can be used, as given by American EEG Society in 1991 [3, 6](see Figure 2.7(c)).

(a) Advantages: This recording technique is non-invasive and does not involve very expensive equipment and does not involve risks.

(b) Disadvantages: The recording site is the size of the ball of paste used to attach the electrode. For longer recording times (45 min or longer) the paste may start drying which could lead to variable contact. Also sweat or movement may lead to falling or slipping of the electrodes. This causes artifacts in the EEG.

2. **Inside the cranium: intracranial electroencephalography**

   (a) **Cortical surface (inside dura matter): subdural electroencephalography / electrocorticography** [36]: These electrodes are in the form of an array with about 3.2 mm diameter stainless disks embedded in 1.5 mm thick medical grade silicone. These arrays are designed to be placed on the cortical surface of the brain beneath the dura by a craniotomy as shown in the surgical crossection in Figure 2.8. Each electrode is connected to a stainless steel wire insulated with Teflon. The wires are collected together and exit from the skull in a flexible tube. These wires are then connected to the EEG machine.

      i. Disadvantages: It is an invasive technique and involves all the risks

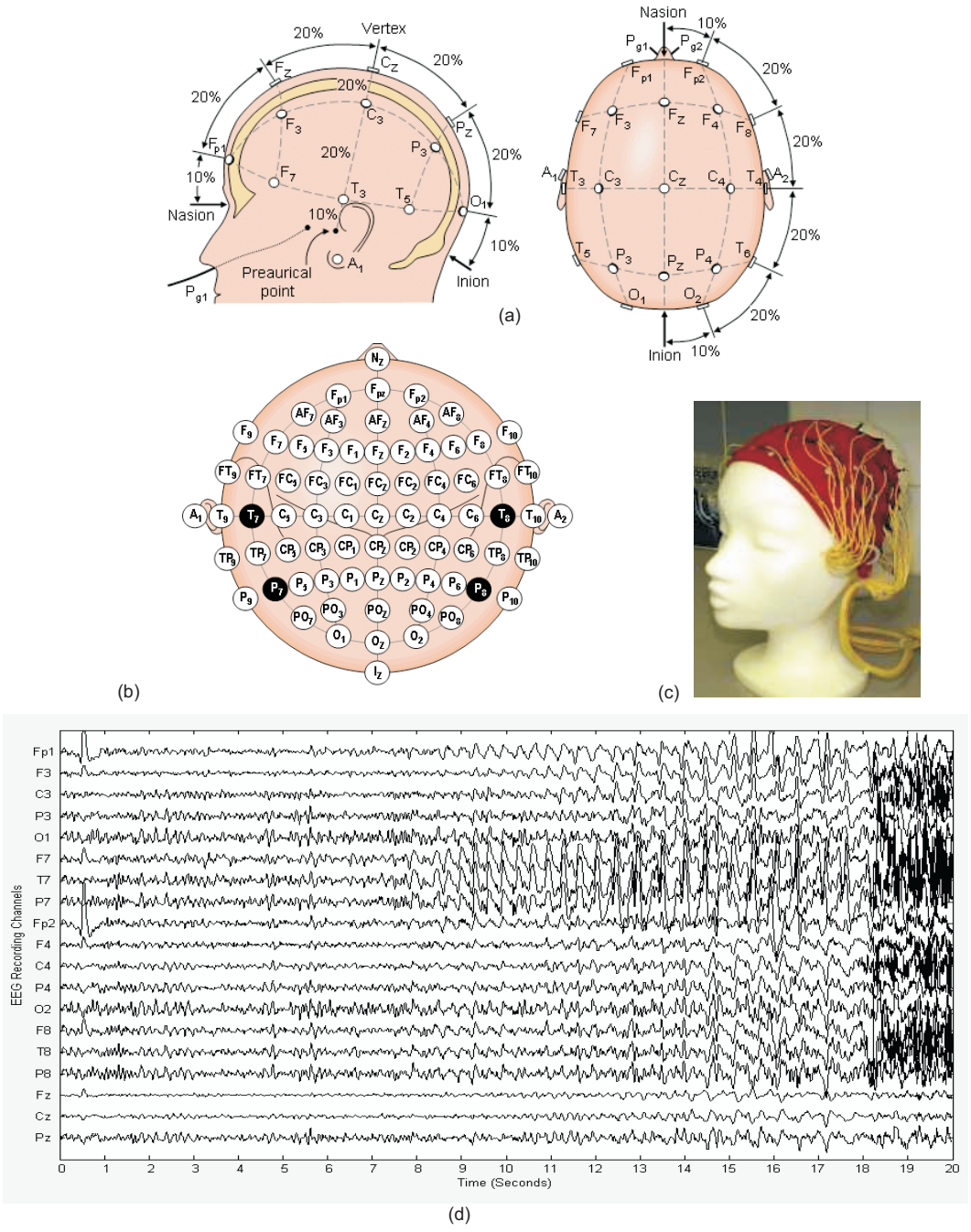


Figure 2.7: (a) The 10-20 International EEG system for scalp electrode placement [3] (reproduced from [4, 5]) (b) In addition to the 10-20 system, additional intermediate 10% positions are also sometimes used (standardized by the American Electroencephalographic Society) [6, 3] (c) Electrodes placed in the form of a cap (d) A multichannel scalp EEG recording (mV/s)

present in normal surgery. There is also some risk of localized infection in scalp or leakage of cerebral spinal fluid around the exiting cable. Both of which are treatable.

ii. Advantages: Recordings collected in this way are remarkably free from

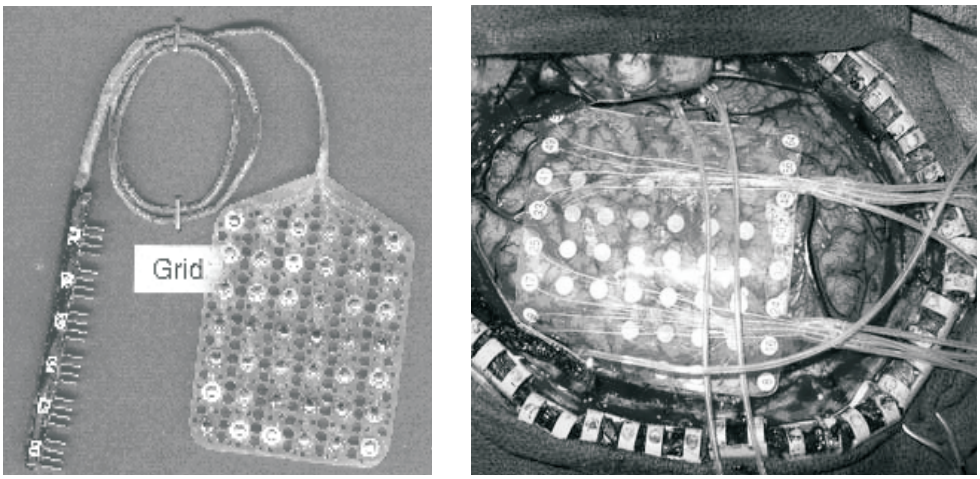


Figure 2.8: Intracranial grid and surgical crossection of a grid placed on the cortical surface of brain [7]

artifact and can be obtained for over 3 weeks in a patient without deterioration of the signal.

(b) **Cortical surface (outside dura matter): epidural electroencephalography**: The electrode usually consists of a 1x4 cm strip of an elastomer in which 3mm disks (centres 1 cm apart) are embedded. These strips are placed epidurally (outside dura).

i. Disadvantages: It is an invasive technique and involves the usual risks of surgery. They give limited information as compared to subdural electrodes.

ii. Advantages: It is free from artifact and is not as invasive as subdural electrodes (does not require rupturing the dura).

(c) **Inside the brain: depth electrode electroencephalography**: These are thin wires placed deep into the brain stereotactically.

i. Disadvantages: It is an invasive technique and involves risks of infection, and haemorrhage (rare). It provides limited information compared to subdural and epidural electrodes.

ii. Advantages: It is remarkably free from movement and EMG artifacts. It can be used for electrical stimulation of deeper structures in the brain.

Selecting the type of electrodes to be used is quite subjective. It is based on the analysis required, the area to be assessed, the condition of the patient, possibility of application of an electrode type, length of recording required, etc. Invasive techniques (depth, subdural and epidural) have an advantage of being free from many artifacts and may also allows one to reach deeper brain areas which may be inaccessible to the scalp electrodes. The disadvantage of such techniques is that they reflect activation of a small number of neurons which are in their immediate vicinity only, whereas scalp electrodes

reflect the activity of a larger area with fewer channels/signals. Alternatively, this can be useful to monitor specific areas with least interference from other sites. However, in the seizure analysis, as the seizure source is not known *a priori*, it is more useful to use scalp electrodes to scan wider brain regions quickly.

2.4.1.2 Amplification and filtering

The original potential at the cortical surface is about $1.50\mu V$ but it gets attenuated by the cerebrospinal fluid, dura matter, bone, scalp, etc. to $10 - 50\mu V$ by the time it reaches the scalp [27] (pp.151). This EEG recorded is an analog signal and it needs to be amplified to make it compatible with displays, analog to digital converters or recorders. For this bio-potential amplifiers are used, that are designed to provide amplification to the physiological signal without distorting them, reject interference signals as much as possible and protect patient and equipment from damage due to voltage or current surges. The amplification of the signal (amplifier gain) is the ratio of the output to the input signal. Usually the signal is amplified around 500 times. Also, to decrease the impact of electrically noisy environment the amplifier has a high common-mode-rejection ratio (CMMR), (at least 100 dB) [34] and high input impedance (at least 100 MΩ) [34]. EEG machines also have a number of high-pass and low-pass filters which may be used if required.

2.4.1.3 Calibration

Calibration of biological as well as machine signal is important and is performed before starting a recording. It involves provoking a known signal and measuring the output of the EEG machine to check if the machine records the signal accurately. The individual electrode impedance is also checked before recording. This is usually done by injecting a very small AC signal to the electrodes and measuring the resulting voltage.

2.4.1.4 Montage

EEG can be accessed in various forms also called *montages* eg. referential (potential with reference to a common reference electrode placed at an electrically quieter area of scalp), bipolar (potential with reference to the next electrode in predefined chains of bipolar electrode montages), etc. These may be interchangeable by digital post-formatting but not always, for example referential montage can be converted to bipolar but bipolar cannot be converted to referential.

The advantages of using referential montage is that they give an undistorted view of voltage changes whilst bipolar montages alter the wave shape and amplitude affecting

both recording electrodes connected to a channel [37]. The disadvantage of a referential montage is in finding an electrically inactive area for the reference electrode. Any change at the reference electrode would show up in all electrode recordings as they are all connected to the reference [37].

2.4.1.5 Temporal and Spatial View

Representation of potential change with time shows a temporal view of the potentials. Along with these, topographic maps are also used occasionally to view the distributions of scalp potential at an instant of time during an EEG recording. Such 'brain maps' are used to visualize the spatial distribution of an EEG waveform. They require value of the potential at all points on the mapping surface (ideal/real model of scalp surface). As EEG provides potentials only at discrete locations on the head surface (electrode positions), these potentials have to be interpolated to intermediate points between these positions, which is performed by using various interpolation algorithms such as the cubic splines, etc.

2.4.2 Characteristics of the EEG signal

A. Rhythms: EEG has a wide spectrum of frequencies. The ultra low (<1 Hz) and ultra high frequencies (>80 Hz) do not have a significant role [27] in many diagnosis and are usually filtered during analysis (though these are gaining attention recently). Thus the frequency range relevant for clinical purposes is usually 0.3 Hz - 80 Hz. These are further classified into many bands of interest. The major frequency bands are as follows:

1. **Delta ($< 3.5Hz$):** Delta activity is prominent in normal children during first year of life and may appear in posterior region throughout maturation. It is also occurs during deep sleep. In an awake adult, rhythmic delta activity is abnormal [38].

2. **Theta (4-7.5 Hz):** This rhythm is a low amplitude irregular rhythm and a usual feature present in the EEG of a normal awake adult, with greatest amplitude in the posterior temporal region. They are more prominent in states of drowsiness and sleep. Its development during a recording accompanied by a decrease in alpha activity may show a suspected onset of sleep [38] .

3. **Alpha (8-13 Hz):** This rhythm occurs during wakefulness over the posterior region of the head, generally with a higher voltage over the occipital areas. Amplitude is variable mostly about 50 μV in adults. It is best seen when the patient's

eyes are closed, in a physically relaxed position with relative mental inactivity. Attention, eye opening and mental effort blocks or attenuates it [27, 38].

4. **Beta ($> 13Hz$):** This rhythm is seen in the frontal and central regions of the brain during wakefulness and usually do not exceed 40 Hz. It shows considerable increase when minor tranquilizers and sedatives are taken.

5. **Gamma ($26 - 100Hz$):** Gamma activity is very high frequency activity and is usually associated with perception and consciousness. It is being linked to the synchronized effect of neuronal populations from different parts of the brain for higher mental tasks as perception [39], learning and recognition memory.

6. **Wave Complex :** Other transient waveforms are usually a response to extrinsic (evoked potential) or intrinsic changes. They may have morphology consistent with two or more rhythms of the brain but are still distinguishable from the background EEG. This is called a wave complex.

B. Artifacts: The EEG recording is usually contaminated by artifacts which may be due to power line interference, patient movements or other electrical sources. These artifacts can lead to misinterpretation of the EEG analysis and sometimes obscure the cerebral information completely. But these artifacts are not easy to filter out from the EEG by standard filters, as they do not always conform to known/recognisable fixed patterns and are spread over a broad frequency band affecting various measurement channels differently. Some commonly occurring artifacts are as follows:

Physiological artifacts:

- **Ocular:** Ocular artifacts are produced due to eye movement and blinking of eyes. Eye movements lead to artifacts prominent at the the frontal electrodes FP1 and FP2. Vertical eye movements occur as a sharp downward deflection (due to eye closure) and sharp upward deflection (due to eye opening) in a bipolar EEG recording, as shown in the Figure 2.9(a).

- **Myogenic:** Muscle potentials can also be picked up on the scalp and are termed as electromyogram (EMG). They could arise due to a movement of facial or scalp muscles like twitching or shivering, smiling, chewing or grinding of teeth. Most often these movements are seen at the frontal or temporal electrodes. On EEG these are seen as bursts of high frequency activity imposed on slower waves in synchrony with the movement as shown in Figure 2.9 (b) .This artifact can obscure the underlying cerebral activity and with high pass filtering may get transformed to a waveform indistinguishable from genuine beta activity.

- **Cardiac:** Cardiac activity or (Electrocardiogram (ECG) may also be picked up by signals recorded on scalp. As the electric potential of the heart is equipotential

on the scalp, ECG may not be picked up in bipolar recordings. But with a common reference (especially when reference is placed off the scalp), the QRS complex of the ECG can be seen to appear periodically on the EEG.

- **Pulse artifact:** Pulse artifact is usually restricted to the electrode neighboring a pulsating blood vessel. It is seen to be periodic and in step with the ECG tracing.
- **Other biological artifacts:** Perspiration, change in skin potential, movement of body, tremor can also lead to unusual artifacts.

Physical Artifacts:

- **Electrode movement:** sudden movement or detachment of an electrode causes a an abrupt rise and fall in EEG recording like a steep right angle rise and exponential decay as shown in Figure 2.9(c).
- **Power line interference:** The interference due to presence of electric and magnetic fields alternating at the frequency of the mains supply of 50 Hz (Europe) and 60 Hz (North America) is quite common in EEG. It shows up as a 50 Hz persistent wave as shown in Figure 2.9(c). Without proper screening of mains sources, even though the induced potential of the body and mains conductor is very small, but in comparison to the EEG it may be quite large and interfering. Usually the EEG systems have notch filters to remove such interference.

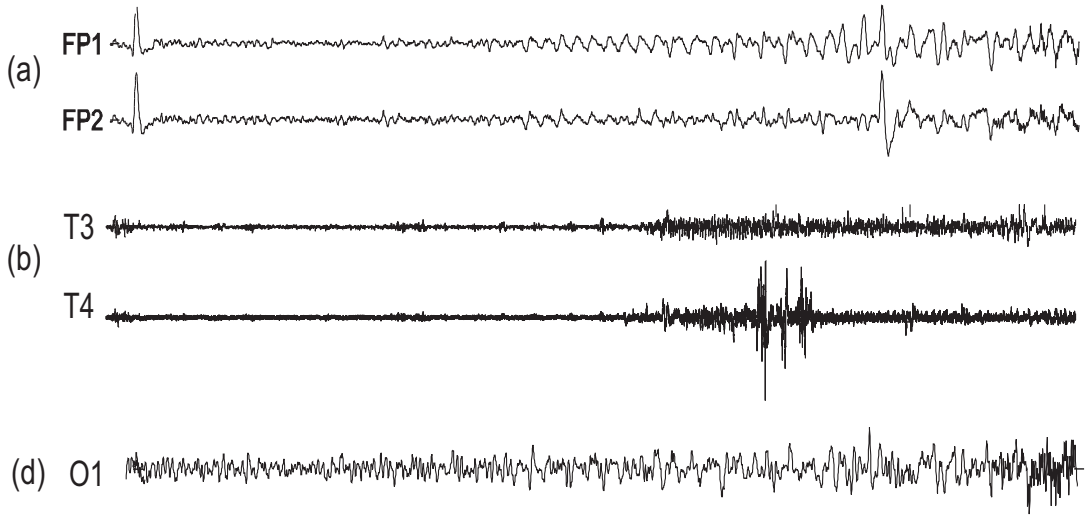


Figure 2.9: (a) Eye blink artifact (b) Muscle artifact (Electromyogram) (c) Power line interference

2.4.3 The 'Normal'/ Epileptic EEG

A. Normal EEG: 'Normal' EEG is not trivial to define. There are certain patterns which are known to be seen with normal brain functioning but their absence does not necessarily show abnormality as the EEG patterns vary vastly with age, state of consciousness, effect of medications, and generally from person to person as well. There are certain known patterns associated with abnormal functioning of the brain and usually their absence is more often used to define the normal EEG. These may occur in few channels or all of them and may appear regularly or transiently. Due to this, sometimes continuous EEG recordings are taken for days or weeks. Some abnormal patterns that are usually looked for (in context with the state, age, settings, etc.) in EEG are:

- Amplitude
- Presence of abnormal patterns
- Frequency
- Symmetry
- Response of alpha rhythm
- Atypical response to visual stimulation
- localized / generalized sudden onset of rhythmic EEG

For example: Slowing (increased low frequency activity) can reflect pathologies as edema, raised intracranial pressure, post ictal state, etc. Global reduction in amplitude may also reflect destruction of underlying neurons [38].

B. Epileptic EEG: A typical EEG of patient with epilepsy can be broadly divided into *ictal* and *interictal* segments. EEG at the time of seizure is called *ictal* EEG and the period with no seizure is termed as *interictal*. These are further described below:
Ictal EEG: In patient with epilepsy, cerebral electrical activity changes dramatically with a seizure onset. Ictal EEG usually depicts a paroxysm i.e. a phenomenon with abrupt onset, rapid attainment of a maxima and sudden termination, lasting at least several seconds and distinguishable from the background EEG. An example of such activity is shown in Figure 2.11.

The following patterns are typically observed in ictal EEG:

- A characteristic onset of slow rhythmic waves (see Figure 2.10 and 2.12 (b)). These may be confined to focal channels (focal epilepsy) or may be spread over many channels (generalized epilepsy) or sometimes it may start at few channels and later spread to more channels.

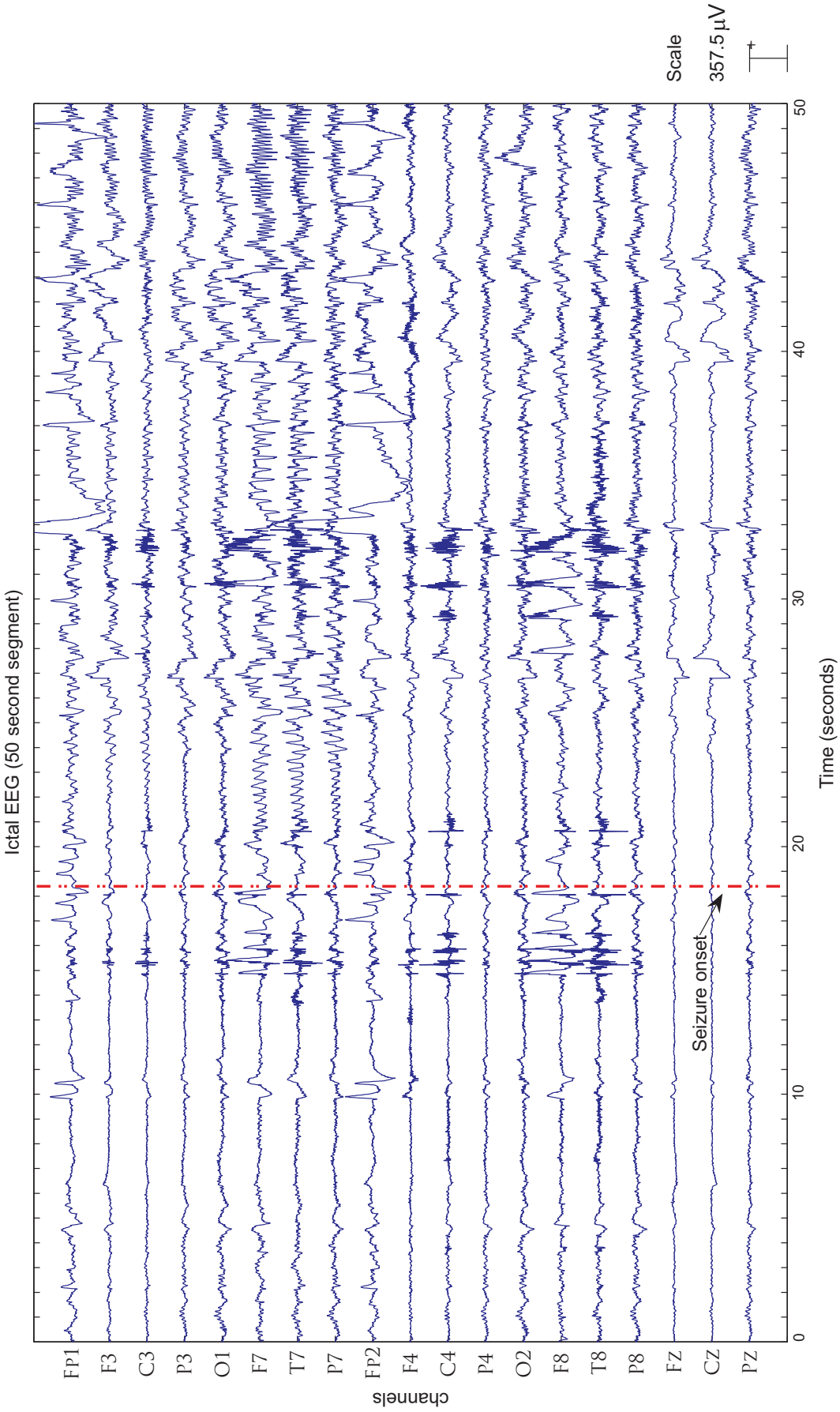


Figure 2.10: The figure shows a typical seizure onset in an epileptic scalp EEG. The seizure onsets is marked by the vertical line. The onset is markedly sudden, rhythmic and concentrated on the EEG channels F7, T7, P7 suggesting a localized left temporal seizure onset.

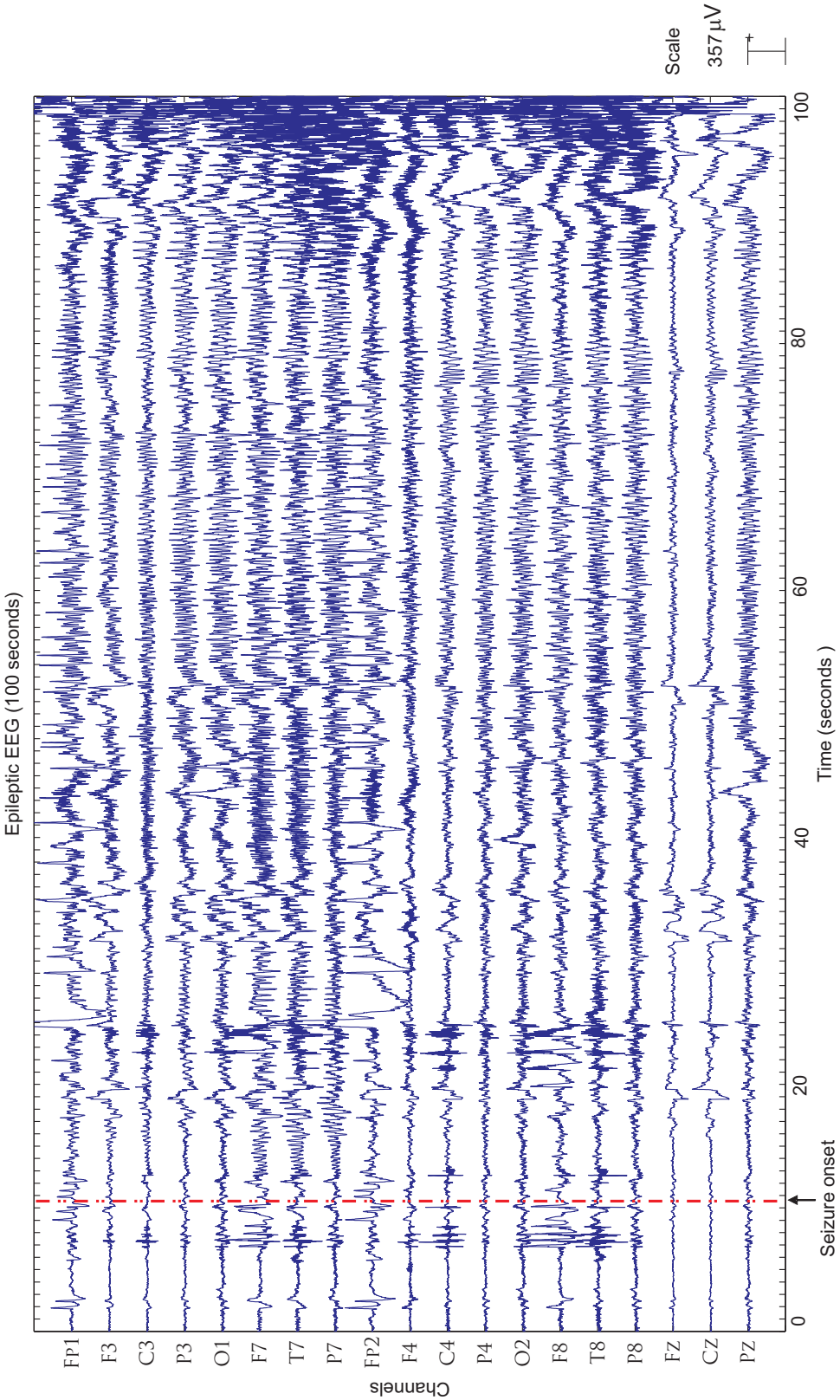


Figure 2.11: The figure shows a typical 100 seconds ictal scalp EEG of an epileptic patient having a seizure. Seizure onset is seen 10 seconds into the recording. A rapid attainment of a maxima is seen in the left temporal side quickly becoming generalized (spreading to all areas of the brain)

- Characteristic sharp waveforms with *spike and waves* are also observed in ictal EEG. *Spikes* are described as waves with sharp outline that stand out from the background EEG and last < 70 ms [38](see Figure 2.12 (c)).
- Many times *spike and wave complexes* are also associated with ictal EEG, where a spike is seen to be followed by a delta wave as shown in Figure 2.12 (a). The spike and slow wave complexes are further described as 3 Hz and slow wave, 6 Hz and slow wave and multiple spike complex.

Interictal EEG: '*interictal*' EEG, may also show signs of abnormality at rest or with provocation (photic stimulation or hyperventilation) that helps in diagnosis of epilepsy. These are called '*epileptiform*' features. Interictal EEG also shows spikes and spike and wave activity even when the patient is not experiencing any clinical seizure. They may be found to be lateralized (over one hemisphere) or generalized (over both the hemispheres).

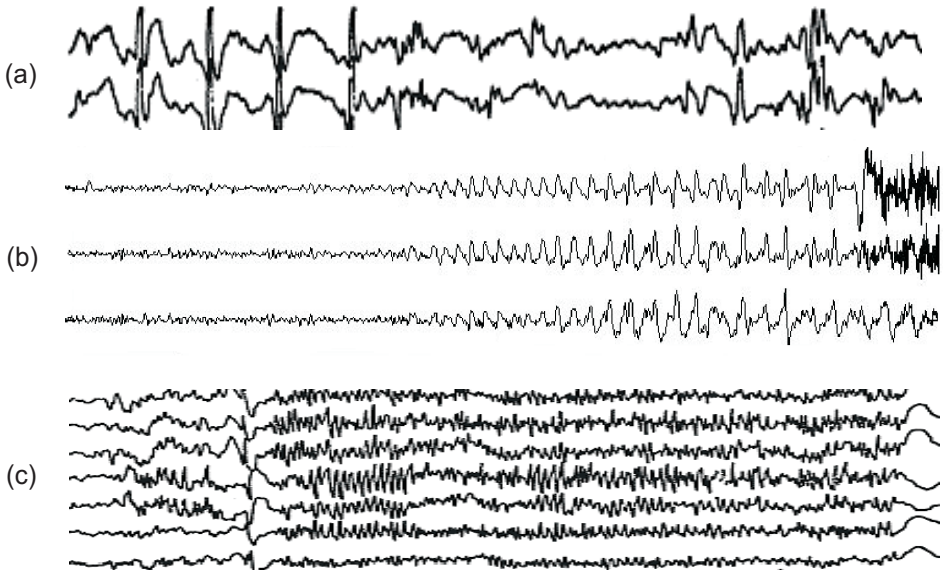


Figure 2.12: (a) Spike and Wave epileptiform activity, (b) Localized rhythmic seizure activity, (c) Paroxysmal activity (sharp transients)

2.5 Seizure Prediction with EEG

The changes in an epileptic brain observed through changes such as auras have inspired much research on ways to detect and predict a seizure. The changes observed in brain signals have made many people research into the underlying dynamics of the brain that leads to such interictal and ictal recordings, with an aim to shed some light on the mechanisms of epilepsy, its cure, detection or warning. But like with any research

area, this field is set with a number of obstacles and a lot of questions which need to be answered and are yet under research in both the medical and the engineering fields. The main steps for research in seizure prediction with brain signals can be listed as follows:

1. **Data collection:** The EEG signals are collected invasively or non-invasively. A good data set would have multichannel continuous EEG recordings over a few days with information about when the seizure starts and ends. Annotations by a clinician and or the monitoring videos help to identify the artifacts. Patient reports by the neurologist help to identify the type of epilepsy and concordance/disconcordance with any abnormalities seen in the patients' imaging results (PET, MRI, etc.) and other details about the variation in the seizures. Information about any drug reduction/sleep deprivation, etc. may prove helpful. Collecting data is a crucial aspect of research in seizure prediction, as there is yet no standardized, internationally available data bank of anonymised patient EEG data for such studies. Also as the requirement is of usually long/continuous multichannel recordings of EEG which requires attaching electrodes on the scalp or embedding them surgically, it has not been ethically, economically viable to conduct such studies only for research purposes. Thus the data is usually obtained from the hospital where epileptic patients are observed for a few days or a week for a possible surgical treatment for epilepsy. In such cases the patients undergo brain surgery to get the electrode grids or strips embedded in the brain for a few days, which are later removed and the EEG recordings are continuously collected and stored in the meantime.

   This data is very beneficial for a researcher but it still needs to be assessed with due consideration given to many parameters that may affect the analysis, such as:

   - Drug reduction: For pre-surgical evaluation the aim is to observe as many habitual seizures as possible in a limited time. Thus if the patient is on drugs to suppress seizures, then those drugs are slowly reduced to allow seizures to take place. This adds a dynamic parameter which may not follow any defined rules and sometimes the patient on drug reduction may not have his habitual seizures but a different type of seizure.
   - Type of epilepsy: The patients that are evaluated for surgery may be from a constrained profile of epilepsy (focal temporal or generalized) and hence it is difficult to generalize the research to many different types of epilepsies.
   - Spatial location of invasive electrodes: Compared to scalp EEG electrode placement systems, there are no standards to describe the placement of invasive EEG electrodes in the brain.

2. **Feature extraction:** A method of feature extraction and/or analysis algorithm is used to find a probable precursor for predicting a seizure. It can be based on univariate, bivariate signals or multivariate signals.

3. **Algorithm Testing:** The prototype is then tested across different seizures of the same patient, at different times of the day and then with other patients with the same seizure type or different seizure type, to test the robustness of the hypothesis. Its effect is then compared with its effectiveness on ictal and interictal data. The number of false alarms and sensitivity are the main factors of judging the probability of the effectiveness of the prototype for prediction. The prediction characteristics are then assessed for a check against random prediction as well.

It becomes difficult if not impossible to compare the results of such studies because of the differences in EEG types used, epilepsy types assessed, channels of EEG analyzed, other parameters such as time of day, drugs, state of consciousness being ignored or described, etc. and only a general overview is obtained. Thus, even though there has been research ongoing in seizure prediction since the 1970's, it may be said to be in its infancy with a long road ahead, before a clinically viable method of predicting a seizure is found and applied for real time warning or seizure prevention.

2.6 Summary

The very nature of epilepsy, especially the apparent randomness of the seizure onset can be very disabling for the person suffering from this disorder. Recurrent, unpredictable seizures can prevent involvement in many activities which may reduce the quality of life and sometimes even prove fatal. This can also lead to depression and/or anxiety. Treatment available in form of drugs can help to prevent or reduce seizures but they usually have side effects that can affect learning and memory. Also about 30% of epilepsy patients do not respond to the currently available drugs. Surgery is a possibility but for a certain limited type of epilepsy disorders only, and with the risks and costs involved, lack of neurosurgeons expert in epilepsy surgery, it is not within reach of many people. Thus there is still a need for some alternative form of help, whilst the cause of the disorder remains unexplained and until a complete cure becomes possible and available.

The first step in the research on epilepsy is analyzing the brain activity records of people having epileptic seizure disorders. In this chapter we have seen various methods for acquiring brain activity records and have assessed them in terms of their suitability and accessibility for the present research. EEG has been found to be the most suitable and it has been discussed in detail with regards to the method of recording (apparatus, electrode placement, amplification, calibration, assessing) and characteristics of the

EEG signals (amplitude, rhythms and frequency content, artifacts). The epileptic EEG and the basic terminology has been described as compared to *normal* . Thereafter a brief introduction to the research on epileptic seizure prediction using EEG has been given with an overview of the challenges in this field. The next chapter will discuss the methods that have been previously evaluated for seizure prediction and the results reported in literature.

Chapter 3

Epileptic Seizure Prediction

3.1 Are Seizures Predictable?

For decades epileptic seizures were understood to occur suddenly, without any transition phase from normal to ictal. But medical research is now strengthening the view that the seizures have a pre-seizure phase with a cascade of changes which might be detected. For example: Rajna et al.,[40] have found pre-seizure auras occurring in 50% of 562 interviewed epileptic patients. Federico *et al.,* [41] have shown a change in the blood oxygen level (BOLD) signal several minutes before seizure with fMRI studies of focal epilepsy patients. Baumgartner *et al.*, [42] and Weinand *et al.*, [43] have found a significant increase in cerebral blood flow in the epileptic temporal lobe, several minutes before a seizure. Novak *et al.,* [44] and Delamont *et al.,*[45] have shown changes in autonomic regulation of heart rhythm several minutes pre-seizure. Many studies have shown the existence of pre-ictal state based on intracranial and scalp EEG. The previous work in this field with EEG has been summarized in this chapter for an overview of the methods used, their results and the challenges they present.

3.2 An Overview of Seizure Prediction Techniques

The earliest work on seizure prediction dates back to the 1970's by Viglione *et al.* [46]. They had used feature extraction and pattern recognition to distinguish 10 min pre-ictal from non pre-ictal data segments. They were able to reasonably show some difference in pre-ictal and non-pre-ictal segments but had a high rate of false positives and the statistical performance was lacking. Research in this area declined for the next 20 years but with the advent of various aspects of technology such as the improvement in brain imaging techniques, EEG recording systems, computational power, computing memory, implantable sensors and signal processing methods it again gained momentum

in the 1990's. It was then that Iasemedis *et al.*[47] introduced the concept of using nonlinear dynamics for EEG analysis.

3.2.1 Non Linear Dynamics

The debate about whether the irregular EEG signal is random or nonlinear deterministic is ongoing. However, it is now known from chaos theory, that random input is not the only possible cause for an irregular output, but even nonlinear chaotic systems can produce very irregular data with purely deterministic equations of motion, in an autonomous way. Iasemedis and Sackellares [47, 48] showed that methods from nonlinear dynamical theory could hence be tapped in order to detect the abnormal activity in EEG recordings of epileptic patients and possibly to predict seizures.

In nonlinear dynamics, the time series (which is a sequence of scalar measurements of a quantity at fixed intervals of time depicts the current state of the system) is first converted into a set of state vectors. This is done by *delay reconstruction* in m dimensions, for n number of points, as follows:

$$\mathbf{NL}_n = (NL_{n-(m-1)\tau}, NL_{n-(m-2)\tau}, ..., NL_{n-\tau}, NL_n) \tag{3.1}$$

The time difference τ, in number of samples is called the *lag* or *delay time*. This forms a trajectory in pseudo-phase space. For a truly periodic systems these trajectories do not diverge and a system with a divergence is said to be in a chaotic state. The averaged exponent of this increase is characteristic of the systems underlying strength of chaos and is called the *Lyapunov Exponent.* A system has as many Lyapunov exponents as the number of states, and the negative ones show the rate of convergence. It is the measure of the largest Lyapunov exponent that is usually used to measure the systems transition to and from a chaotic state.

In the context of epileptic seizures it was hypothesized that seizures represented a transition from a less ordered or chaotic (normal) state to a more ordered (abnormal) state and then back to a normal chaotic state. Measures such as correlation dimension and largest Lyapunov exponent were used to characterize this transition from a high to low dimensional dynamical state.

Iasemedis and Sackellares [47] used intracranial EEG of one patient with 8 seizures occurring over a 5 day period. They used single channel EEG segments of 20 minutes, with a 10 min pre-ictal epoch. They were able to demonstrate pre-seizure changes up to 7.5 min before a seizure.

Studies in 2003 and 2004 have raised doubts about the suitability of using this index [49, 50], suggesting that the results of Iasemedis and Sackellares [47] were strongly

dependent upon the chosen electrode sites and the parameters used for nonlinear analysis.

Also, Aschenbrenner-Scheibe, [51] and Notley [35] show that with continuous intracranial EEG data and intracranial rat EEG data respectively, such a measure fails to produce robust clinically applicable ability to predict. The main problems identified with nonlinear dynamical analysis for EEG seizure prediction are as follows:

1. **The length of data used**: The Lyapunov exponent calculations need longer data sets and for short data sets where the structure of the system changes in intensity within msec (in brain) only an approximation or a local Lyapunov exponent can be defined but its interpretation can be very difficult [52](Ch. 5).
2. **Selection of the embedding dimension:** It is difficult to know the exact value of the embedding dimension and its evolvement over data segments (even as long as 20 minutes). Moreover, the exponent can vary with the embedding dimension used.

Despite the failure observed in the preliminary studies, the methods are being researched further, as the many parameters involved, such as the embedding dimension, type of EEG, dynamics of the patient data, etc. can influence the results and may be in need of more informed settings.

In a further study by Iasemidis *et al.*,[53], another measure known as *Short Term Lyapunov exponent (STL_{max})* was introduced to overcome the challenge of local Lyapunov exponents. They were able to show a warning in 91.3 % of 23 seizures (in two patients) at 89 min prior to seizure with a prediction horizon of 3 hours, their false warning rate was about one every 8.27 hours. The shortcoming was the lack of significance testing as the data used was limited to two patients.

3.2.2 Complexity and Correlation Dimension

In 1994-95 another nonlinear measure for EEG called *complexity* and the *correlation dimension* was introduced for seizure prediction. For epileptic seizures it was hypothesized that the neuronal networks of the brain were highly complex and in the epileptic process they underwent a complexity loss or a state of reduced dimensionality. This led to the interest in extracting features from EEG recordings that might help to identify the complexity and estimate the change in complexity with seizure onset. In work of Lehnertz, Elger *et al.* [54], the complexity was measured by *correlation dimension analysis*.

Correlation dimension is based on the correlation integral $C(r)$ which was introduced by Grassberger and Procaccia [55]. The correlation integral $C(r)$ for a time series

$v(j)$ in a reconstructed space is the probability that two randomly chosen embedding vectors $v(j_1)$ and $v(j_2)$ lie within a given distance r of each other and is estimated by :

$$C(r) = \frac{1}{N^2} \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} \theta(r - ||v(j_1) - v(j_2)||), \tag{3.2}$$

where $||.||$ is the norm distance and θ is the Heaveside unit function. If the system is chaotic for sufficiently large m, the embedding dimension, the correlation integral takes a scaling form independent of m :

$$C(r) \sim r^{D_2}, \tag{3.3}$$

where the exponent gives the correlation dimension. Thus D_2 can also be obtained from the slope of $lnC(r)$ vs lnr

$$D_2 = lim_{r \to 0} \frac{dlogC(r)}{dlogr}. \tag{3.4}$$

Lehnertz *et al.* [54], found a reduction in correlation dimension immediately before a seizure. A dimension drop of sufficient amplitude and duration was regarded as a feature to define seizure preceding state. They did a study on a group of 20 patients' intracranial EEG and showed a drop in 'neuronal complexity' in pre-seizure periods. In 1998 they provided the first statistical evidence for seizure prediction (16 patients, intracranial EEG) with a pre-seizure time of upto 25 min [56]. In a further study Lehnertz *et al.* [57] showed 67% temporal lobe and 29% neocortical seizures were preceded by such a drop of complexity. However, the shortcomings were the use of short (30-50 min) interictal data segments, and the dependency of this method to the variability of dimension in the interictal time windows. Also, their method shows its use only for electrodes confined to the seizure focus, and for data only directly before a seizure.

Their work was reiterated by Aschenbrenner-Scheibe *et al.* [51] and Harrison *et al.* [58] in 2003 and 2005 respectively. Aschenbrenner-Scheibe *et al.* tried to investigate the sensitivity and specificity of this method based on intracranial long term recordings of 21 patients. Their study showed that only one out of 88 seizures showed a significant pre-ictal correlation dimension drop in long term recordings due to increased variability of correlation dimension being accessed in longer inter ictal (24 hour) periods. They further tested the dimension drop within shorter time windows of up to 50 min prior to seizure and found that for a false prediction rate less than atleast 0.1/hour, the sensitivity was about 33% - 38%. They also tested this measure against a random predictor and found the results significant compared to random prediction, which was

encouraging as it showed there was indeed some specific change occurring before a seizure that may allow prediction. But they found that the false positives were as high as 70 false predictions per month while only one out of 3 seizures were correctly predicted in the month. This would be discouraging for the patient. Thus this measure was not found clinically useful. They concluded that the mean length and amplitude of the dimension drop in the pre-ictal data did not show any significant change as compared to the inter-ictal data.

In a separate study by Mormann *et al.*[59], complexity was again found to generate a high number of false positives and a low sensitivity. It was also dependent on subjective selection of EEG channels from *a posteriori* knowledge.

3.2.3 Spatio-temporal Correlation Dimension

Correlation dimension (sometimes referred to as *correlation density*) was further studied in 1998-2001 for seizure prediction, by Martinerie *et al.* [60], which included spatial information from the multichannel EEG records. They adapted the phase space reconstruction by using spatial as well as temporal information. A four time delay iteration and four intracranial electrode recording data set (close to the epileptogenic focus) was used resulting in a spatio-temporal dimension of 16. This was implemented as follows: If $x_k(j)$ is the electrical neuronal activity at time j at electrode site $k = 1, 2, 3, 4$ and four delayed amplitudes are recorded at each site, then at each time j, the state of the neuronal activity can be represented in a 16 dimensional space as:

$$v(j) = \{x_k(j), x_k(j - \tau_k), x_k(j - 2\tau_k), x_k(j - 3\tau_k)\}, \tag{3.5}$$

where k = 1 to 4, j=1 to N.

The data they used included 19 seizures from 11 patients having temporal lobe epilepsy. They showed that 17 out of 19 seizures could be predicted within a time window of 2.6 minutes. But they had used 40 min data epochs with only a 20 min pre-ictal (interictal) period. Thus their method was in need to be tested for correlation dimension variability across longer interictal recordings as well.

3.2.4 Dynamical Similarity Index

In 1999 this work on spatio-temporal dimension was used to describe a new method called *dynamical similarity* [61]. Essentially, it aimed to measure the dynamical similarity between different parts of the EEG, in long term recordings (data used: 23 seizures from 13 patients. The length of data segments used was 40 minutes in total with 20 minutes of data being prior to the seizure). This technique was basically trying to

overcome the challenges of the non-linear dynamical measures, such as the variability of the dimension of the nonstationary EEG. The nonstationarity of EEG had forced the use of shorter time windows, but these were not sufficient for a reliable estimation of the nonlinear quantities. To overcome this, the new nonlinear method called *similarity index* was proposed. The assumption (supported by their previous work) was that pre-ictal changes involving neuronal recruitment evolved gradually, and did not vary significantly over shorter time scales of seconds. Their method compared the dynamics of different EEG windows instead of a moving window strategy. In brief, the dynamical similarity method can be described as follows:

1. Measure dynamics of a reference window: The EEG time series is divided into small segments and the aim is to quantify the extent to which the underlying dynamics differs between pairs of distant segments. This is carried out by calculating the nonlinear characteristics of a *reference* EEG window (few minutes in length) free of seizure. Nonlinear characteristics are calculated using time delay embedding or the phases in a Poincare section.

2. Comparison with reference dynamics: Then the dynamics of the reference window are compared with the dynamics of distant EEG segments, up to 25 seconds in length.

3. Calculating dynamic similarity: The dynamic similarity is then estimated using (a) a cross-correlation integral defined as:

$$C(S_{ref}, S_t) = \frac{1}{N_{ref} N_t} \sum_{i=1,N_{ref}} \sum_{j=1,N_t} \theta(||Y_i(S_{ref}) - X_j(S_t)|| - r) \tag{3.6}$$

   where N_t are the number of elements in each set, S_{ref} is the reference segment and S_t is the data segment at time t.

   and (b) a cross-correlation ratio defined as :

$$\gamma(S_{ref}, S_t) = C(S_{ref}, S_t) / \sqrt{C(S_{ref}, S_{ref}) C(S_t, S_t)} \tag{3.7}$$

   This gives a probability of finding points within a neighborhood r of the reference state space.

4. Thresholding and prediction: When the difference between the dynamics of the reference and the scanned window crosses a predefined threshold, a pre-seizure state is said to be identified.

Using this method with a probability distance r as 30%, dynamical similarity was shown to identify pre-seizure time of 6 min with 13 patients (23 seizures). The data used was intracranial (9 patients, 17 seizures, 94% sensitivity of prediction) [62] and scalp EEG

[63] (23 patients, 26 seizures in total, 96% sensitivity of prediction with prediction time being 7 minutes, using data segments of only 60 minutes prior to seizure).

Nevertheless, this method is highly sensitive to the reference window selected and the threshold may require adaptation to patient, epilepsy type, seizure in same patient or change of vigilance. Also, their use of the similarity index was challenged by De Clercq *et al.* [64] and Winterhalder *et al.* [65] in 2003.

It should be noted that these nonlinear dynamic statistics depend on the state space on which they are defined, it is possible that as the distribution of the points on this state space changes, they affect the statistics. Thus to overcome this problem another method was introduced by McSharry et al.[66] based on quantifying the changes within the state space, called *multi-dimensional probability evolution (MDPE)*. They aimed to track how often different regions of a state space are visited while the EEG is in the inter-ictal state. This could help to detect changes in the underlying dynamics of the EEG signal. However, they mention that it was not found useful in comparison to variance based techniques for seizure detection.

3.2.5 Accumulated Energy

The concept of accumulated energy was explored by Litt *et al.*(in 2001 [67]) in context of seizure prediction. Accumulated energy was calculated on broadband data by a running sum of the square of the EEG amplitudes for a given epoch. The average energy was given by:

$$E_k = \frac{1}{N}\sum_{i=1}^{N} x_i^2(k), \tag{3.8}$$

calculated for a time window of 1.25 seconds, and then the accumulated energy is given by:

$$AE_m = \frac{1}{10}\sum_{k=10m-9}^{10m} E_k + AE_{m-1}, \tag{3.9}$$

where $m = 1, 2, ...$ and $AE_0 = 0$.

They had used 3-14 day intracranial EEG from five patients with Mesial Temporal Lobe Epilepsy (MTLE). 80 one hour epochs were analyzed, with 30 leading to a seizure. They had used the electrodes that showed the earliest seizure onset and compared it to the energy of baseline EEG electrodes and contralateral EEG electrodes. They found that accumulated energy increased dramatically about 50 min prior to a seizure. A mean time of 18.5 min pre-seizure warning was concluded, with a linear threshold.

Nevertheless, their work could not be reproduced by studies performed later in 2005 on intracranial EEG by Harrison *et al.*[68]. Another study on intracranial EEG of 21

patients with 88 seizures by Maiwald *et al.*[69] showed a lower sensitivity of 18-31%.

3.2.6 Linear Techniques

Apart from nonlinear techniques, various linear techniques were also explored, such as power spectral density, parametric modeling, wavelet decomposition.

- **Variance of signal amplitude**: The use of variance (second statistical moment) of EEG amplitude has been studied on EEG data to observe linear changes in the time series. Variance is similar to the accumulated energy concept as accumulated energy measures the sum of squares of the amplitudes. The variance is usually calculated on consecutive overlapping / non-overlapping windows. If S_i is the EEG signal at time i, then the variance of this EEG is given by:

$$\sigma^2 = \langle [S_i - \mu]^2 \rangle, \tag{3.10}$$

  where the mean is given by $\mu = \langle S_i \rangle$ and $\langle . \rangle$ is the average of the signal over a time interval.

  McSharry *et al.*[66] showed that variance displayed clear changes at seizure onset but muscle and movement artifact may interfere as they increase in variance simultaneously. Thus variance has helped in detection of seizure (results being similar to non linear analysis)[70] but it has not been found useful for prediction of seizures. No definitive amplitude changes have been identified in pre-ictal data so far.

- **Power spectral density**: Approaches using the power spectrum of the EEG were also tried as it is postulated that there might be certain power changes in a certain frequency at, or before, seizure onset which might be helpful [66, 35]. But not much work has been demonstrated for seizure prediction. The total integral of the power spectral density of the EEG signal over all non zero frequencies was also used to track a 'quiet' period before a seizure. It was not really found as a useful measure for seizure prediction [35] as it was not robust and would cause a high number of false alarms.

- **Parametric modeling (Autoregressive modeling)**: AR modeling has been used for a better frequency resolution (for a given time resolution) compared to a spectrogram. The aim was to show a change in the model order as the seizure onsets, but no consistent change was observed in the model order over time. Thus it was found that the AR model does not essentially help in providing any additional knowledge for prediction (when used for depth recordings [35].

- **Wavelet decomposition**: Wavelet decomposition has been found to be useful for analyzing changes in time-frequency spectrum. The discrete wavelet transform (DWT) decomposes a signal into time-frequency sub-bands using orthogonal basis functions called wavelets. It involves repeated filtering, downsampling and partitioning of the signal into detail (high pass) and approximation (low pass) components. At each step a transformation is carried out on the approximation part, whose length is halved successively. The transform coefficients measure the correlation between the signal and the wavelet at a particular time and resolution (scale). Thus wavelet acts like a filter that can adjust the window for better time / frequency resolution. It has been applied to EEG for calculating the wavelet entropy and relative wavelet energy of the signal over a fixed time, in order to study the multi-frequency dynamics associated with EEG. An important feature of this method is that it does not require the assumption of stationarity of the signal [71] and allows simultaneous analysis over a range of frequencies.

3.2.7 Statistical Testing and Prospective Studies

In the 1990's research studies a missing element was the significance statistics and comparison with controlled inter-ictal data, due to which these studies lacked specificity. The comparison of features for ictal and inter-ictal recordings was seen in late 90's with studies such as [72]. Mormann *et al.*, showed changes in phase synchrony in intracranial recordings occurring only prior to seizure and Navarro *et al.*(2002) [73] showed a more frequent drop in similarity index before a seizure compared to inter-ictal recordings.

Gradually, studies with statistical testing are also being carried out. Elger and Lehnertz (1998) [56] used correlation dimension as a measure of complexity for 16 patients and found a sensitivity of 94% with a 11.5 min pre-seizure period. Martinerie *et al.* (1998) [60] found an 89% sensitivity to predict at 3 min using correlation density. Le Van Quyen *et al.* (1999, 2000) [61, 62] used the similarity index for intracranial recordings and found 83% and 94% sensitivity of prediction in 13 and 9 patients respectively with a pre-seizure period of 4-5 min. They also found a 96% sensitivity in scalp EEG recordings for 23 patients [74]. Iasemidis *et al.*(2001) [75] found a 49 min pre-seizure with 91% sensitivity for 5 patients, using the concept of largest Lyapunov exponent. Mormann *et al.*(2003) [76, 77] used detection of changes in phase synchronization and found 86% sensitivity for seizure prediction with a preseizure period ranging from 4-200 minutes. However, specificity was still lacking and with the studies and their results being challenged the analysis was in need of bigger datasets. Also, usually the work with nonlinear dynamics did not account for the change in vigilance or state i.e. change of state from sleep or wake or effect of medication and as mentioned by Saab *et al.* [78] the transitions from different states of wakefulness can also lead to similar observations of change in complexity.

In 2003 and 2005 first few prospective studies were also carried out by Iasemidis *et al.*[79] and D'Alessandro *et al.*[80] but they were not found to be fit for clinical application as they showed a high rate of false alarms (26 false alarms per day). Their specificity of prediction was also not tested against randomness.

3.2.8 Synchronization

In the mean time, in this previous work and in the medical literature, a strong view that was developing was that epileptic seizures were associated with synchrony of critical neuronal masses. This link of synchronization with the genesis of epilepsy is not new and can be found in medical literature from the 1970's as well [81, 27]. Advanced signal processing techniques, knowledge of non-linear dynamics, better brain activity measurement apparatus and digital analysis has spurred the interest of measuring and analyzing the synchrony changes in brain in the engineering and medical community. It is being hypothesized that during the inter-ictal state, the neuron population around the epileptic focus are involved in large scale dynamics of normal brain functioning. During pre-ictal state, these neurons lose the synchronization with themselves and other cortical areas. This possibly isolates the epileptic focus from the normal involvement of the brain functioning and facilitates the development of a hyperexcitable focus. Also it provides an 'idle' population of neurons that can be easily recruited by the focus into the epileptic process. Seizure activity may get initiated when a critical mass of neurons gets recruited. Also, during pre-ictal state, the inhibitory control of the epileptogenic zone can also progressively break down. This is very well summarized by Le Van Quyen [8] in the Figure 3.1.

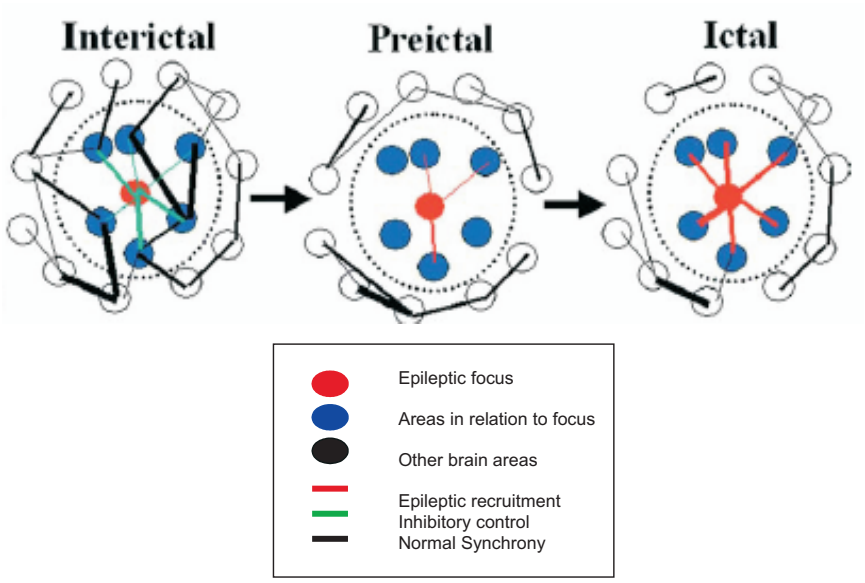


Figure 3.1: Synchronization and de-synchronization of areas near focus, as epileptic seizure onsets [8]

These theories give a strong indication of a possibility of finding a precursor for prediction. This would be based on measuring and quantifying the dynamics of synchronization of signals in the brain. This has spurred the research on analysis of synchrony in brain activity.

Initially, the basic and simpler measures of linear synchronization like coherence and cross-correlation were being employed. They were found to be useful for obtaining information that was not perceivable on visual inspection [82] (using ECoG of a rat), but was not found significant or robust enough for clinical use [83]. Rosenblum *et al.*[84], Pikovsky *et al.* [11] and Tass *et al*'s.[85] work on measuring phase synchronization of chaotic/noisy signals was very helpful in strengthening the interest in phase synchronization for EEG signals. Ongoing work on neuronal phase synchrony [86] also helped to support the view of underlying synchronies in the brain. The first study that showed a method for quantification of phase synchrony between brain signals was performed by Lachaux *et al.* [39] in 1999. They also described a way to check for the significance of the phase synchrony called phase locking statistics (PLS). Recently, in 2001, Bhattacharya [87] showed that the long range synchrony was reduced in patients with pathological condition (seizures, mania) as compared to healthy people. In 2000 more work was done on the various indices used to calculate the phase synchrony of epileptic EEG. Mormann *et al.* [72], used the mean phase coherence and showed the decrease in synchronization pre-seizure and a consequent return to the normal high level at seizure onset. Much work was being done to analyze the hypothesis of a pre-ictal decrease in synchronization in epileptic EEG, that may help to predict seizure onset [77, 76, 88]. Le Van Quyen *et al.* [88] used intracranial data of 5 patients and used a classification algorithm in order to identify pre-ictal from inter ictal segments by compiling a reference library of characteristic interictal synchronization patterns. They found 70% of seizures could be predicted in this way specifically within the 4-15 Hz frequency band data. However, this analysis may be quite sensitive to the reference library and may differ from patient to patient. Although, the data used was also not extensive, yet it showed the existence of changes in neuronal synchrony prior to a seizure, that might be helpful for prediction. Another study by Mormann *et al.* [76] described an automated detection of pre-seizure state based on the concept of decrease in synchronization prior to seizures. They used data of 10 patients and were able to detect pre-seizure states in 12 out of 14 seizures by identifying the hemisphere of seizure activity.

A study using a combination of phase synchronization and cross correlation in the form of a linear discriminator was carried out on 9 seizures with intracranial EEG [89]. The linear multivariate discriminator was defined as

$$\mathbf{X} = \begin{pmatrix} \bar{C}(0) \\ \bar{S} \end{pmatrix} \tag{3.11}$$

where $\bar{C}(0)$ is the average cross correlation and $\bar{S}$ is the phase synchronization. The findings in this study showed this discriminator to be unreliable in 7 out of 9 seizures. It should be noted that they had used broadband EEG signals for phase synchronization analysis as well. Broadband signals used for phase synchronization analysis can show spurious results as the estimation of phase and frequency of broadband signal may not be accurate. They also mention that the dynamics introduced by medication reduction and subsequent increase in seizure frequency may affect the seizure prediction analysis.

Synchrony dynamics were also considered spatially in the context of seizure analysis. Studies involving local versus distant phase synchronization was performed by Dominguez *et al.* [90]. He showed that seizures were characterized with enhanced local synchrony as compared to distant synchrony.

As the methods of calculating synchrony were being researched various comparative studies were carried out as well, for example Le Van Quyen *et al.* [91] compared the use of the Hilbert Transform and the wavelet transform for analyzing phase synchronization. They analyzed two selected intracranial EEG recordings (ictal) and also scalp EEG (non-epileptic) and showed that the Hilbert Transform (with filtering) and wavelet transform give similar results (reliability of results was tested with surrogate data).

It should be noted that the EEG signal is a broadband signal having no central dominant frequency. For this data, the Hilbert Transform may not be suitable as it is based on narrow band signals; yet many times this has been ignored and the Hilbert Transform has been applied on unfiltered EEG. Netoff *et al.* [92], tried to compare the phase synchrony analysis of broadband and narrow band signals but at a neuronal level. They used very little data and were not able to conclude as to which method is better for synchronization, though they found that the broadband analysis worked better in most of their cases. This result must be treated with caution as using broadband data with Hilbert Transform is basically incorrect in theory, and the results highly questionable.

As we see above, there were quite a few independent studies for synchrony dynamics that were carried out recently. But conclusions about synchrony as a seizure predictor have not been drawn yet. One of the main reasons being the difficulty in comparing the studies as they are performed on different data sets, different type of EEG (intracranial/scalp), etc. Efforts to compare the performance of synchrony with other methods was carried out by Quiroga *et al.* [82]. They analyzed nonlinear interdependencies, phase synchronization, mutual information, cross correlation and coherence for rat ECoG signals. They showed hemispheric interdependencies that were not apparent by visual analysis of EEG and found most of the linear and nonlinear measure performed quite similarly except for mutual information.

3.2.9 Multichannel Signal Analysis

To date more emphasis has been placed on univariate measures, such as largest Lyapunov exponent, correlation dimension, accumulated energy, neuronal complexity or bivariate studies, such as phase synchronization, cross-correlation or 2 channel Lyapunov exponent. However in these studies the results are usually sensitive to the EEG channel selection. Some researchers have found that using channels close to the focus [56, 61, 62, 60, 93] gave the requisite information for prediction of seizures while others found contralateral electrodes also providing such information [88, 77, 59, 80]. Recently another study by James *et al.,* [94] showed that the information about epileptic seizures can also be obtained from pairs of contralateral electrodes (away from seizure focus). Hence the channel selection for analysis remains a subjective element in such studies.

EEG is a multivariate signal and apart from the temporal information it displays, it also carries spatial information and may involve a complex spatio-temporal pattern leading to the seizure, which can be useful for prediction, but is rarely utilized. Few studies have used this embedded information for seizure detection and prediction, such as the following:

- James *et al.* [95, 37] used a self organizing feature map (SOFM) to detect seizures using multichannel EEG with feature vectors based on autoregressive parameters;
- D'Alessandro *et al.,*[80] used genetic algorithm based feature extraction and a probabilistic neural network for classification for seizure prediction. They used intracranial data from five patients with 10 min of pre-ictal epochs. They used half the seizure as the training set and other half as a validation set. They showed a sensitivity of 100% for two patients, complete failure in one patient and 1.1 false positives per hour for the fifth patient. They found the best electrode to be analyzed was the one contralateral to the seizure focus.

  This technique is sensitive to the features selected, which include the features of curve length, energy and nonlinear energy.
- Firpi *et al.* [96] used genetic programming (based on a technique called as genetic algorithm) for seizure prediction with features extracted from state-space trajectories of EEG.
- Kobayashi *et al.* [97] used singular value decomposition (SVD) and showed the ability to estimate seizure focal brain activity when its effect on scalp EEG is unclear to visual examination.

Nonetheless, the concept of EEG signals being mixed signals has gained attention. It is being accepted that EEG is a mixture of overlapped signals from the underlying cortex, that should be separated prior to any analysis [9]. Studies such as that of

Meinecke *et al.* and Hong *et al.* [98, 99] confirmed that direct feature analysis on raw EEG (mixtures) could lead to incorrect/spurious results. This was also confirmed for seizure prediction and detection recently by Gupta *et al.* [100]. More on this will be expanded later in the thesis.

3.2.10 Blind Source Separation

In the mean time another branch of signal processing called Blind Source Separation (BSS) became popular and made its way to biomedical signal analysis [9, 101]. BSS is a technique that allows blind separation of linearly, squarely mixed signals from given mixtures of signals. It can be performed in the time domain or the frequency domain and may use higher order statistics. Details on these techniques are covered in following chapter. Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are the main BSS techniques that have been explored for seizure analysis. James *et al.* [94], Gupta *et al.* [102, 100] and Sweeny-Reed [103] have showed the use of ICA and phase synchronization for EEG in seizure analysis. James *et al.* [104] and Nicolaou and Nasuto [105] compared the temporal decorrelation separation technique with statistical ICA and found temporal decorrelation to be more suited for EEG analysis. Min Jing *et al.* [106] carried out a study to predict dynamical changes of the brain using BSS and short-term largest Lyapunov exponent and were able to show a possibility to predict seizures, but they did not present specificity and sensitivity studies.

Algorithms such as constrained ICA: spatially constrained by Hesse *et al.* [107], temporally constrained by James *et al.* [108] and spectrally constrained by James *et al.* [109], that evolved from the basic ICA technique were found to be quite helpful in seizure analysis. They were also used along with nonlinear dynamics and an improvement was seen in the results of seizure detection with a possibility of seizure prediction [94, 110]. This technique has been analyzed further in this thesis.

3.3 Assessment of Seizure Prediction Time

An important question in seizure prediction studies is deciding an appropriate prediction time. Longer prediction times may allow diversified and much less intensive control measures (like drugs which may need time to be absorbed and reach the cortex) and early warnings, but it may need to be assessed for the clinical effects of false alarms and the psychological stress on patient who is warned about an impending seizure long before it happens. Shorter prediction times may require intensive fast acting intervening measures (like electrical stimulation that may need only few seconds) which may not be tolerable and may have unwanted side effects with false predictions. A good

prediction time would be a compromise between the prediction horizon, sensitivity to prediction, number of false alarms, accuracy in spatio-temporal localization and/or tolerance to warning times. Maiwald *et al.* [69] have made an effort to provide an assessment criteria called *seizure prediction characteristics* for assessing the prediction times. For a clinically viable system, it may also need to perform on-line in real time. To check the prediction against random alerts, various significance tests have been suggested in previous studies, such as: Monte Carlo simulation and surrogate data [76, 59, 89, 111, 112] or comparison with a random predictor [65, 113].

It should be noted that it is not trivial to compare the studies in this field because of the number of variabilities between studies such as the following:

- Lack of standardization of the data set: there is yet no standard database for epilepsy studies due to ethical issues and laws of acquiring and sharing patient data; and also because it is still an open field of research and there are many combinations of parameters that are slowly being assessed.
- Type of epilepsy: there is a big variation in the type of epilepsies being studied in the medical and biomedical research. Some concentrate on focal (temporal lobe, neocortical, frontal, parietal) epilepsy while others may use generalized epilepsy data, while others do not mention the type of epilepsy.
- Type of EEG: the type of EEG used also varies, being intracranial (focal grids, depth electrodes) or scalp.
- Selecting channels for assessment: The commonly used univariate or bivariate analysis concentrates on one or two channels of EEG. As EEG is a multichannel signal, selection of few channels usually involves subjective selection. Usually channels near the focus (ipsilateral) are used, hypothesizing the seizure focus to be maximally dynamic. Sometimes channels away from the focus (contralateral) are also used for various reasons. But this makes comparisons between studies difficult if not impossible.
- Effect of medication/vigilance states: usually the continuous epileptic EEG data is obtained from hospitals where patients are being monitored for epileptic presurgical evaluation for possible surgery. Patients are usually implanted with grids for a week and are monitored in this period for possible seizure focus localization, by observation of their seizures and changes in intracranial EEG. In order to reach a possible diagnosis of the seizure focus, maximum seizures are expected to be recorded in the limited time. To facilitate seizures, patients may/ may not be taken off their anti-epileptic medication. This can influence the analysis as the effect of such changes are not yet known properly and are either not accounted for or not mentioned making comparisons complicated.

- Subjectiveness in EEG analysis by epileptologists: Another difficulty in epileptic data analysis has been the lack of gold standards for comparisons. Even the clinical experts in epileptology may differ in their opinions about the electrical seizure onset / focus of seizure.

- Lack of statistics: Epilepsy data is not easy to acquire, which has also been a major cause of lack of statistical reports on the analysis in this area.

- Short term EEG vs long term continuous EEG: Many previous studies use short EEG segments with a seizure. However for seizure prediction, features that are highlighted in the ictal segments need to be checked for their occurrence in interictal segments as well. A complete analysis for prediction should be based on long term continuous EEG data to assess the false positive rate as well as the sensitivity of the predictor.

3.4 Summary

This chapter reviewed the literature on seizure prediction, mainly over the last decade. It can be seen that a lot of analysis has been performed using the nonlinear time series approach using the robustness of the fundamental nonlinearities in EEG signals, [114]. Such approaches still being actively pursued as there are still many dynamic parameters involved in all studies (medication effects, size and type of data set used, length of signals used, type of epilepsy, contralateral or ipsilateral EEG channel used for analysis).

The following Tables 3.1, 3.2 summarize the previous research work with regards to the type of data used, type of epilepsy evaluated (if known), amount of data analyzed, prediction time estimated and sensitivity of prediction of seizures. Most of this work was done on MTLE (focal), and few studies have been performed on Neocortical Epilepsy (Le Van Quyen *et al.* [74], Navarro *et al.*[73], Chavez *et al.*[115]). Not much work has been done on generalized seizures, one major reason is the lack of data, as the patients that come for pre-surgical evaluation as candidates for possible surgery are usually patients with focal epilepsy.

The main objective in these studies has been a search for a precursor for seizure prediction. This may have a form of an identifiable, significant pattern, feature or a dynamical change occurring before seizure onset. The main challenges that have been identified with seizure prediction are as follows:

1. **Noisy recordings:** Prior to the seizure onset, the information about it is usually immersed in the background EEG and noise as muscle contraction becomes strongly entrained in such a situation. Thus the signal-to-noise ratio is usually

Year	Author	Method	Pat.	Seiz.	S (%)	FP /hr	Pr (min)	Ref.
1998	Martinerie *et al.*	Correlation density	11	19	89	-	4	[60]
1998	Lehnertz and Elger	Correlation Dimension	16	16	94	0	12	[56]
1999	Le Van Quyen *et al.*	Similarity index	13	23	83	-	6	[61]
2000	Le Van Quyen *et al.*	Similarity index	9	17	94	-	4	[62]
2000	Mormann *et al.*	Phase synchronization	17	3	100	0	-	[72]
2001	Iasemidis *et al.*	Dynamical entrainment	5	58	91	-	49	[75]
2001	Le Van Quyen *et al.*	Phase synchronization	8	-	77	-	few min	[116]
2001	Lehnertz *et al.*	Correlation dimension	59	95	48	0	19	[57]
2001	Jerger *et al.*	Different measures	4	12	100	-	2	[117]
2001	Litt *et al.*	Accumulated Energy	5	30	90	0.12	19	[67]
2002	Navarro *et al.*	Similarity index	11	41	83	0.31	8	[73]
2003	Mormann *et al.*	Synchronization/correlation	10	14	86	0	86	[76]
2003	Mormann *et al.*	Phase synchronization	18	32	81	0	4-221	[77]
2003	Chavez *et al.*	Phase synchronization	2	6	-	-	30	[115]
2003	D'Alessandro *et al.*	Feature selection	4	46	63	0.28	3	[118]
2003	Iasemidis *et al.*	Dynamical entrainment	5	55	83	0.17	72	[79]
2003	Winterhalder *et al.*	Similarity index	21	88	42	0.15	30	[65]
2003	Aschenbrenner *et al.*	correlation dimension	21	88	34	0.10	50	[51]
2004	Maiwald *et al.*	Accumulated energy	21	88	30	0.15	-	[69]
2005	D'Alessandro *et al.*	Feature selection	2	19	100	1.10	2	[80]
2005	Esteller *et al.*	Accumulated Energy	4	42	71	0.11	85	[119]
2005	Harrison*et al.*	Accumulated Energy	5	51	-	-	-	[68]
2005	Iasemidis *et al.*	Dynamical entrainment	2	11	82	0.15	78	[53]
2005	Jouny *et al.*	Complexity/synchrony	2	25	-	-	-	[83]
2005	Le Van Quyen *et al.*	Phase synchronization	5	52	69	-	187	[88]
2005	Mormann *et al.*	30 measure	5	51	-	-	-	[59]
2005	Navarro *et al.*	Similarity index	13	129	64	-	13	[120]
2005	Harrison *et al.*	Correlation dimension	20	960	0	0	15-90	[58]
2006	Schelter *et al.*	Phase synchronization	4	20	70	0.15	-	[113]

Table 3.1: Previous work on seizure prediction with intracranial EEG. Pat.: Number of Patients; Seiz.: Number of Seizures; S: Sensitivity (%); FP/hr: Number of False Positives per hour; Pr: Prediction time (minutes).

Year	Author	Method Used	Pat.	Seiz.	S (%)	FP /hr	Pr. (min)	Ref.
2000	Hively *et al.*	Dissimilarity measures	-	20	100	-	52	[121]
2001	Le Van Quyen *et al.*	Similarity index	23	26	96	-	7	[63]
2002	Schindler *et al.*	Simulated neuronal cells	7	15	100	-	83	[122]
2003	De Clercq *et al.*	Similarity index	12	-	-	-	-	[123]
2003	Hively *et al.*	Dissimilarity measure	41	46	88	0.02	35	[124]

Table 3.2: Previous work on seizure prediction with scalp EEG. Pat.: Number of Patients; Seiz.: Number of Seizures; S: Sensitivity (%); FP/hr: Number of False Positives per hour; Pr: Prediction time (minutes).

very low and thus analytical techniques based on power/amplitude of the signal might not be useful.

2. **The complexity of a 'normal' brain network:** The pre-seizure data might be

sensitive to other parameters like rhythmicity, phase synchronization and phase de-synchronization, due to the dynamical changes that are known to occur in the brain [8].

3. **Lack of known pre-seizure features/patterns:** To date there are no known fixed / consistently occurring pattern of events / features recognized in pre-seizure data.

4. **Lack of a 'gold standard':** There is no 'gold standard' for comparison, for normal EEG data quality.

Apart from the challenges and shortcomings observed in the previous research on seizure prediction, these studies do suggest the existence of abnormal dynamics before and at seizure onset. The various measures such as correlation dimension, complexity loss, Lyapunov Exponent, reduction in chaos, phase synchrony, etc. all point to essentially to some form of hyper-synchronization of brain areas at seizure onset. The synchronization hypothesis suggests that a seizure like activity may be explained as being a result of activation of critical masses of synchronized neurons [125, 126, 127, 8, 128]. It is most likely preceded by recruitment of normal neuronal sites as the normal synchrony between neurons in itself is not strong enough to activate such dynamism such as seen in a seizure. Such progressive coupling of the focus with normal brain areas may occur over days, hours, minutes or seconds, depending on the type of epilepsy, patients, etc. It is this idea of synchrony that has been explored in the current research using measures of phase synchronization as discussed in following chapters.

Chapter 4

Preprocessing of EEG with Independent Component Analysis

4.1 The Need to Preprocess EEG

Scalp and Intracranial EEG signals have been extensively used for seizure onset prediction in their raw form. However the underlying principle of the EEG signals is about capturing synchronized activity of neuronal networks from a distance. With the complex anatomy of the brain and the distribution of the neurons, some neuronal/anatomical pathways may cause signals from far apart neuronal spaces to reach the same electrode (see Figure 4.1), also referred to as the *effect of volume conduction*. This makes the EEG signals a set of mixed signals. These properties effect the estimation of linear and nonlinear quantities leading to misinterpretations and inconsistent measures. It is not trivial to revert such mixing, due to lack of knowledge about the mixing process or the underlying true signals.

Along with such complex mixing and background noise, the signals at the electrodes also get mixed with physical and biological artifacts such as ocular, muscular and cardiac signals; power line noise, electrode movement and electrode pop. These artifacts are spread over different frequency bands across time and electrode channels in a complex way and removing them through usual filtering techniques is not possible, as that removes other relevant information as well. When such a signal travels from the neuronal source to the electrode through soft bone and tissue, it gets attenuated. An already weak signal (for example, from a seizure source prior to a seizure) can easily get overshadowed in the background EEG. The assessment of EEG signals is usually based on extracting linear or non-linear features using raw EEG. Thus, it is important to pre-process the EEG to improve the signal to noise ratio, remove artifacts and appropriately unmix it, before analysis. This motivates the first pre-processing step for the EEG signals namely: unmixing and de-noising.

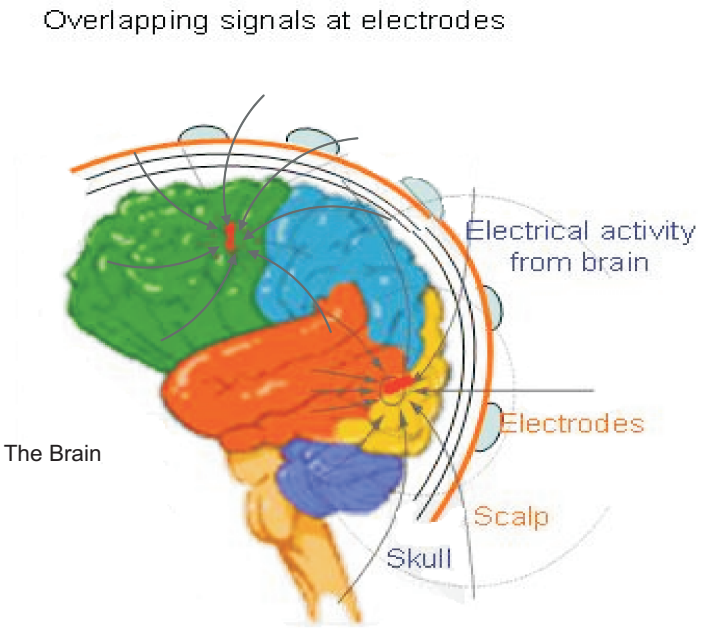


Figure 4.1: The figure shows the mixing that occurs in the EEG signals measured at the electrodes on the scalp (shown only for demonstration on quite a magnified scale). Signals from two sources embedded in the cortex emanate signals that are collected at the scalp by the electrodes placed at various fixed positions (according to the international EEG electrode placement system). One electrode receives signals from not only the sources directly below it but also neighboring sources, making the EEG a mixed signal.

4.2 Preprocessing 1: Unmixing and De-noising EEG

To remove the effect of mixing or volume conduction, previous studies have suggested the following:

1. The use of signals which have lesser dispersion of signals, for example MEG instead of EEG (as magnetic field does not get dispersed by the tissue in the same way as EEG)

2. Improve spatial precision of EEG by techniques such as inverse deblurring and Scalp Current Density (SCD) profiles [129]. SCD profiling uses the Laplacian of the scalp potentials which essentially gives the divergence (rate of flow through a unit area) of the current density.

However, there techniques have problems such as:

1. Signals from techniques such as MEG are difficult to obtain, expensive and not suitable for measuring activity at the time of an epileptic seizure,

2. MEG and SCD may not be able to record activity from deep lying sources that are accessible by EEG.

3. Inverse de-blurring requires additional information such as a model of the real head obtained from the patient's MRI reconstructions, which is not a trivial task.

Keeping in mind the limitations of the above methods we have opted to apply a different approach for unmixing and de-noising the multivariate EEG. We have used a state-of-the-art technique called Independent Component Analysis (ICA). ICA is a Blind Source Separation technique that aims to extract true or source signals when only a set of mixtures of these signals is known. It is thus useful in extracting artifacts and physiologically relevant signals from a given set of multichannel EEG [130]. Such a decomposition is based on separation criteria that uses higher order statistics [131, 132], or that exploits the structure in time series through time-delayed cross correlations [9, 133] or spatial time frequency distributions (STFD) [134]. Both statistical and temporal based ICA produce neurophysiologically meaningful results when applied to EEG. The statistical approach using FastICA [131] is most widespread but for the analysis of ongoing EEG, statistical ICA faces problems. One such problem is the need to have longer data segments for analysis, as the statistical methods rely on sample-estimates. This may not be possible with ongoing EEG as it is essentially non-stationary. It has been argued that the time structure based approach may be more appropriate for EEG decomposition, shown by James *et al.* [104]. They have compared the time structure based and statistically based ICA on short epileptiform EEG segments, and have shown that temporal ICA is able to extract neurophysiologically informative components with a better temporal resolution than statistical ICA. Temporal ICA is also able to extract signals from shorter data segments, where statistical ICA fails, hence overcoming the issue of non-stationarity.

The ICA technique, its assumptions, limitations and application in the neurophysiologic domain are described in the following sections.

4.3 Independent Component Analysis

The ICA algorithm is essentially a technique that allows multiple measurements from a mixture to be broken down into their fundamental constituents. It aims to recover statistically independent sources, when given a set of their mixtures. ICA extracts the sources from the mixtures with no information about the mixing process and mixing weights, hence is referred to as a 'blind' separation. In the process, it additionally recovers the mixing weights. This algorithm is based on higher order cumulants or time structure of the time series. It can be implemented in different ways subject to the cumulant being maximized or minimized or the way of processing (batch mode or

online mode). It is also possible to constrain the algorithm to search for a particular source(s) only, referred to as *constrained ICA (cICA)*. ICA and cICA have been helpful in EEG analysis for unmixing and de-noising as they are multivariate techniques and exploit the spatio-temporal and time-frequency information inherent in the EEG. They have certain assumptions and limitations which are described in the following sections.

4.3.1 Assumptions

The application of ICA assumes the following:

Statistical independence of the original sources The basic assumption of ICA is that the sources to be recovered are independent. The property of independence is stronger than uncorrelatedness and refers to a lack of joint distribution of the functions, having no higher order correlations [9, 135].

Non Gaussian sources The sources to be recovered are assumed to have non Gaussian distributions [135]. The higher order cumulants are non-zero for non Gaussians and this is essential for ICA model. If more than one Gaussian source is present in the mixtures, they will not be separated by ICA. Recently, another set of ICA algorithms are being used as well that bypass this assumption and are based on time-structure of the signal instead [133].

Linear instantaneous mixing The mixing is assumed to be instantaneous (without delay) with linear superposition of the source signals.

Noiseless mixing ICA assumes noise free mixing though noise free mixtures are ideal signals and practically it is very difficult to obtain noise free biomedical signals or to de-noise them prior to use. Thus applying ICA to such signals leads to independent sources or least dependent sources which may remain contaminated by noise [9].

Square mixing The assumption of square mixing implies that the number of sources is assumed to be equal to the number of mixed signals. This is an assumption with due consideration of the fact that it may not be possible to know the actual number of underlying sources in biomedical signals. Usually biomedical systems have fewer sources than mixtures and some independent components may get split into more than one component, distorting the recovered sources. This happens because the ICA aims to find (fit) as many sources as the number of mixtures (an assumption of ICA being square mixing). This is also referred to as 'overestimation' or 'overfitting'. Overfitting can affect the interpretation of the sources obtained. To combat overfitting, dimension reduction techniques are applied, to reduce the number of sources to be estimated.

Stationary Mixing Another assumption with ICA is that the mixing matrix remains stationary and is not changing with time. Essentially, it refers to the stationarity of the sources in space that appears as fixed set of weights in mixing matrix over the time range used for ICA.

4.3.2 The Separation Algorithm

Fundamentally, the problem that ICA aims to solve is for a set of p data points measured at time instant t, $\boldsymbol{v}(t) = [v_1, v_2, v_3..., v_p]^T$ assumed to be a linear combination (represented by a mixing matrix $\mathbf{A}$) of q unknown statistically independent sources $\boldsymbol{s}(t) = [s_1, s_2, s_3, ..., s_q]^T$ (where $q \leq p$) such that

$$\boldsymbol{v}(t) = \mathbf{A}\boldsymbol{s}(t), \tag{4.1}$$

ICA aims to find the de-mixing matrix $\mathbf{W}$ such that

$$\boldsymbol{s}(t) = \mathbf{W}\boldsymbol{v}(t). \tag{4.2}$$

The de-mixing matrix helps to find the estimates of sources $\boldsymbol{s}(t)$. The mixing/demixing process can be described as a change of coordinates. The data vector remains the same, while it undergoes passive transformation as it is expressed in a different coordinate system. The mixing matrix gives the spatial information of the sources. As shown in Figure 4.2, the mixing matrix column describes the distribution of a source across all the mixed signals. While the row of the mixing matrix shows the distribution weights of all extracted sources in one mixed signal. To simplify the estimation of the independent sources, *whitening* or *spehering* is done initially. This involves diagonalization of the covariance matrix of $\boldsymbol{v}(t)$ and making the data unit variance. ICA then uses higher order statistical information (kurtosis, etc.) or time structure to estimate the independent sources.

It is noteworthy here, that Principal Component Analysis (PCA) is also a BSS technique that aims to generate signals that are orthogonal to each other such that the signals are uncorrelated with each other. While ICA aims to find independent signals, where independence includes uncorrelatedness; uncorrelatedness does not necessarily imply independence. **ICA output: Sources and Topographic Maps**

ICA sources can be viewed temporally as time series or spatially as topographic maps. An example of the temporal and corresponding spatial form of the ICs from a multichannel EEG is shown in Figure 4.3. Along with the temporal form, the topographic map is useful for visualizing the distribution of the potential due to the source. If the EEG is a scalp EEG then the map is plotted on the scalp using a head model. As the

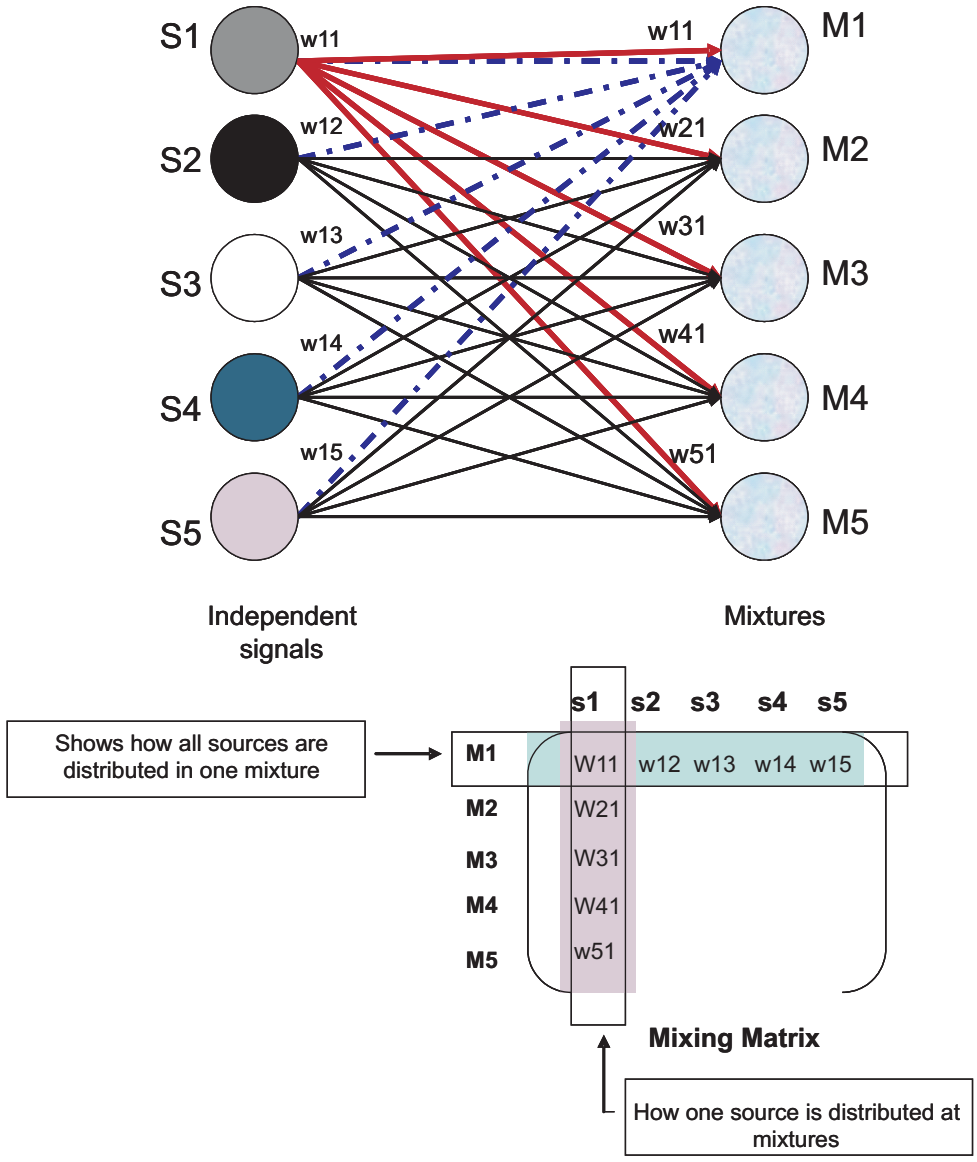


Figure 4.2: An example of source mixing is shown here. The different source signals are mixed in different proportions to form mixed signals. The weights $w_{1,1}, w_{2,1}, w_{3,1}, w_{4,1}, w_{5,1}$ are the information that tells how one source s_1 contributes to all mixed signals $(m_1....m_5)$. The weights $w_{1,1}, w_{1,2}, w_{1,3}, w_{1,4}, w_{1,5}$ are the information that shows how all source signals $s_1...s_5$ contribute to one mixed signal m_1. This information is found by the ICA algorithm in the form of a mixing matrix $\mathbf{A}$. The columns of the matrix $\mathbf{A}$ have the estimate of weights $w_{1,1}, w_{2,1}, w_{3,1}, w_{4,1}, w_{5,1}$ and hence the spatial information, while the rows contain the $w_{1,1}, w_{1,2}, w_{1,3}, w_{1,4}, w_{1,5}$ about how all sources contribute to one mixed signal.

mixing weights of a source are assumed to remain constant over time, hence the spatial map shows the distribution of the source for the time window evaluated.

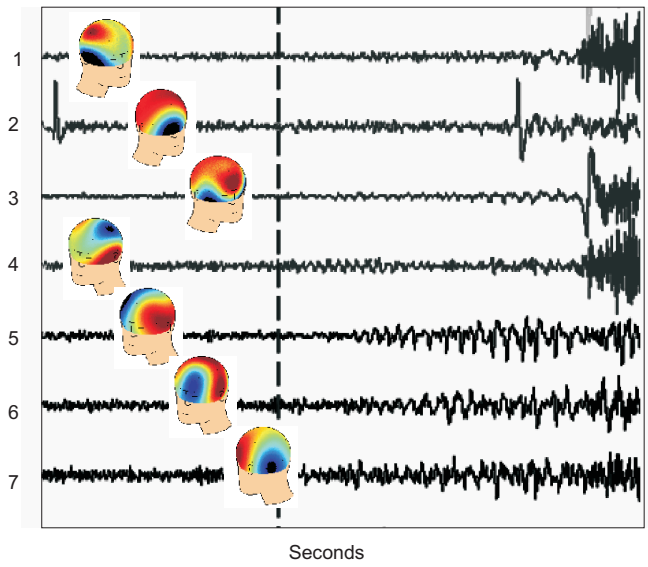


Figure 4.3: The figure shows the source signals (1-7) which have been found by ICA (IC's) and the head models show the corresponding topographic maps of each source.

In the neurophysiologic domain, the mixed signals correspond to the EEG and the source signals to the underlying signals from the brain. The measured EEG is assumed to be formed as a linear summation (unknown linear mixing criteria) of a finite number of signals from underlying unknown independent sources in the brain. Thus if the set of source signals is represented by $\mathbf{S}$ and the linear mixing criteria as the mixing matrix $\mathbf{A}$, then the EEG signals obtained are essentially a set of mixtures $\mathbf{M}$. These mixtures (EEG) may be acquired from the electrodes placed on the scalp, or inserted into the brain (cortical, depth, intracranial). These $\mathbf{M}$ are then used as inputs to the ICA algorithms (with/without constraints), and the aim is to estimate the sources $\widehat{\mathbf{S}}$ along with the mixing matrix $\mathbf{A}$ and the un-mixing matrix $\mathbf{W}$.

The estimate of sources can then be viewed as topographic maps using the mixing matrix $\mathbf{A}$ or temporal using the estimated source signals $\mathbf{S}$.

The use of ICA on electromagnetic brain signals, in particular the neurophysiologic domain, is valid as the assumptions required for ICA are generally met in this domain. The correlates of the required basic assumptions in the neurophysiologic domain are as follows:

- **The mixtures are linearly mixed:** EEG/MEG is assumed to be a linear summation of electrical/magnetic brain activity from different regions of the brain.
- **The unknown sources are statistically independent:** The activity of inter-

est is assumed to be independent of the ongoing background EEG and the various sources.

- **The number of sources is equal to the number of mixtures:** This is an assumption as in practice there can be either fewer sources or more (in practice it is usually found to be fewer though) than the mixtures and there is no method to ascertain this number prior to, or even after the analysis.
- **The mixing is stationary:** The distribution of the electromagnetic field is assumed to be spatially fixed and only the electrical strength varies [9] (thus the signal sources are assumed to be stationary across the windows being analyzed).
- **Mixing is noiseless:** Noise free mixtures are ideal and sometimes it is very difficult to obtain noise free biomedical signals or to de-noise them prior to use.
- **Mixing is instantaneous:** The propagation delay between the signals is assumed to be negligible and hence the signals can be assumed to be instantaneous.

ICA Implementation The implementations of ICA in the literature are based on the criteria for independence being used or the computational processes involved. The statistical ICA approach is essentially based on non Gaussianity. The independent source is recovered by finding the direction where the data is maximally non Gaussian. Several sources are estimated by finding different directions of maximal non Gaussianity with the constraint of decorrelation. The property of non Gaussianity can be measured by cumulant measures such as kurtosis or by entropy based methods. This estimation can be performed by gradient techniques or faster fixed-point algorithms, referred to as *FastICA*. Apart from statistical techniques, some ICA algorithms are also based on time structure known as *temporal decorrelation* methods. Temporal decorrelation involves joint diagonalization of a set of symmetric matrices that reflect the spatio-temporal covariance structure of the mixture data. The above ICA algorithms can be programmed as deflationary or symmetric. Deflationary refers to estimating one source at a time (using Gram-Schmidt method) while symmetric refers to estimation of all sources in parallel.

Higher Order Statistical ICA The statistical ICA method is based on higher order statistics like Kurtosis or Negentropy. Kurtosis is essentially a normalized version of the fourth moment $E\{\mathbf{x}^4\}$ (first central moment being *mean*, second *variance*, and third *skewness*). Kurtosis is zero for a Gaussian variable. It is a measure of the peakedness of a distribution and is defined as

$$kurt(\mathbf{x}) = E\{\mathbf{x}^4\} - 3(E\{\mathbf{x}^2\})^2, \tag{4.3}$$

The kurtosis is the basis of statistical ICA. The problem is essentially posed as maximizing the (absolute) magnitude of kurtosis which makes the estimated

sources as non-Gaussian as possible and hence 'independent' (Gaussian sources have zero kurtosis).

Negentropy uses the concept of differential entropy. As mentioned in [9], of all the random variables, with different distributions, but having equal variance, Gaussian random variables have the largest entropy. Negentropy: From information theory, differential entropy H of a random vector y with density $p(\mathrm{y})$ is defined as

$$H(y) = -\int p(y) log p(y) dy.$$

Negentropy (J) is a normalized version of entropy, given as:

$$J(y) = H(y_{gauss}) - H(y),$$

Where y_{gauss} is a Gaussian random vector with the same variance as y. For a Gaussian random variable, negentropy is zero and for a non-Gaussian variable, it is positive [135].

This can now be solved as an optimization problem, with ICs as the solution. A common pre-processing step for ICA is whitening or sphering which involves linear decorrelation of the mixtures and scaling to unit variance. This optimization problem is usually solved by a gradient method as described below [135]:

1. Remove mean: mean is removed to make the data centered.
2. Whiten or sphere: Whitening implies linear decorrelation and scaling the data to unit variance. This is usually done by eigenvalue decomposition of the covariance matrix:
   $E\{\mathbf{x}\mathbf{x}^T\} = \mathbf{E}\mathbf{D}\mathbf{E}^T$ where $\mathbf{E}$ is the orthogonal matrix of eigenvectors of $E\{\mathbf{x}\mathbf{x}^T\}$ and $\mathbf{D}$ is the diagonal matrix of its eigenvalues. The whitening transform matrix $\mathbf{V}$ is then given by $\mathbf{V} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T$ and whitened mixture data is obtained by multiplying it with the whitening matrix: $\mathbf{z} = \mathbf{V}\mathbf{x}$. Whitening helps to reduce the number of free parameters and hence improves the quality of separation.
3. Initialize parameters: An initial vector $\mathbf{w}$ with unit norm (random) and an initial value for the learning parameter γ is chosen.
4. Apply the learning rule to update weights at each iteration:

   $$\Delta\mathbf{w} \propto \gamma\mathbf{z}g(\mathbf{w}^T\mathbf{z}),$$

   where g is derivative of the function G and γ is a constant is estimated as:

   $$\Delta\gamma \propto (G(\mathbf{w}^T\mathbf{z}) - E\{G(v)\}) - \gamma$$

The function G is used for the approximation of negentropy or kurtosis. For example, $G(y) = y^4$ gives a kurtosis based approximation, or $G(y) = tanh(y)$ are useful for a good approximation of negentropy

5. Normalize $\mathbf{w} \leftarrow \mathbf{w}/\|\mathbf{w}\|$
6. If it does not converge, iterate step 4 onwards for given number of iterations.

One IC can be estimated by maximizing the negentropy or kurtosis as described above. However to find more than one component, the algorithm could be run many times with different initial points. But this would not be reliable and in practice, to find more than one component, a decorrelation scheme is used. The algorithm is iterated as before, and additionally to prevent the weights converging to the same maxima, the weights are orthogonalized at each iteration. This orthogonalization could be deflationary (finding one source at a time) or symmetric (finding many sources in parallel).

Some commonly used statistical ICA algorithms are described here:

FastICA FastICA [131, 104, 9, 135] is a statistically based fixed point algorithm that is much faster in operation than a gradient based method. The FastICA algorithm is similar to the gradient algorithm with a change in the update rule as shown below [135]:

$$\mathbf{w} \leftarrow E\{\mathbf{z}g(\mathbf{w}_T\mathbf{z})\} - E\{g'(\mathbf{w}^T\mathbf{z})\}\mathbf{w}$$

The gradient method for ICA has a slow convergence and depends on the choice of the learning rate and with a bad choice of learning rate, it may even destroy the convergence [135]. FastICA converges much faster, as its name implies. Its convergence is cubic as compared to the gradient ICA algorithms. Also FastICA does not have adjustable parameters like the learning rate, making it simpler and more reliable [135]. It is freely available as a MATLAB package [136], and is most commonly used ICA algorithm because of its ease of implementation, reliability and fast operation.

Infomax Another commonly used algorithm is Infomax. This algorithm for ICA is based on maximizing the non-Gaussianity of the components by their *negentropy* [132]. Infomax algorithm is a neural network gradient based algorithm that aims to maximize information (hence the name *infomax*) by maximizing the output entropy of the neural network. The learning rule is based on the ICA model (to maximize the non-Gaussianity). However, Infomax faces the problems that are inherent in neural networks, such as over fitting, optimizing the step size and the learning rate or the choice of the nonlinearity to be used.

Joint approximate diagonalization of eigenmatrices (JADE) This approach of solving ICA uses tensorial methods [135]. JADE is based on the fourth

order cumulant tensors, aiming to make it as close to zero as possible, which implies statistical independence. The eigenvalue decomposition of the covariance matrix of data transforms the data such that the second order correlations are zero by jointly diagonalizing (making all other off-diagonal components close to zero) a set of matrices [9]. A limitation of JADE is that it can't be used with high dimensional data due to numerical reasons.

Temporal Decorrelation ICA can also be implemented using a time based approach referred to as temporal decorrelation. Algorithms for temporal decorrelation are LSDIAG$_{TD}$, TDSEP [133] or SOBI [137]. These methods are robust in the presence of Gaussian noise as they do not assume the sources to have non-Gaussian distributions, unlike methods based on higher order statistics. As it is not easy to ascertain *a priori* the non-Gaussianity of the unknown signals in the neurophysiologic domain, temporal decorrelation methods appear more suited for signals like EEG. However, temporal decorrelation assumes stationarity of the signal. This may pose a problem for non-stationary signals like EEG. This is overcome by analyzing windowed EEG segments [138]. Short epochs of EEG can be assumed to be fairly stationary and temporal decorrelation can be reliably used with short segments.

Temporal decorrelation methods exploit the fact that, due to independence of the source signals, the covariance matrix of the source signals would be diagonal for all time lags $\tau = 1, 2, 3, ...$ They try to capture the dependency structure of the signal (mixture) using a set of square matrices (arranged in a stack), and then find the ummixing matrix which is the joint diagonalizer of the stack

$$\mathbf{C}_x^k = \mathbf{A}\mathbf{C}_s^k\mathbf{A}^T, \tag{4.4}$$

where $\mathbf{A}$ is the mixing matrix, $\mathbf{C}_x^k$ is the k^{th} covariance matrix of the data $x(t)$ and $\mathbf{C}_s^k$ is the corresponding covariance matrix of the sources $s(t)$. The source covariance is obtained through inversion as

$$\mathbf{C}_s^k = \mathbf{W}\mathbf{C}_x^k\mathbf{W}^T, \tag{4.5}$$

where $\mathbf{W}$ is the un-mixing matrix. The relation of the two stacks is shown in Figure 4.4. The index $k = 0, 1, 2...L$ is the index into the stack referring to the lag, , where L is the maximum number of lags used. The mixing matrix can be estimated in two ways: *Forward* and *inverse* method. Inverse method is more common and involves estimating the un-mixing matrix $\mathbf{W}$ first. Here as $\mathbf{C}_s$ is supposed to be diagonal, the coefficients of $\mathbf{W}$ are optimized such that the matrix given by

$$\mathbf{W}\mathbf{C}_x^k\mathbf{W}^T,$$

is as diagonal as possible. The diagonality can be measured by sum of the squared

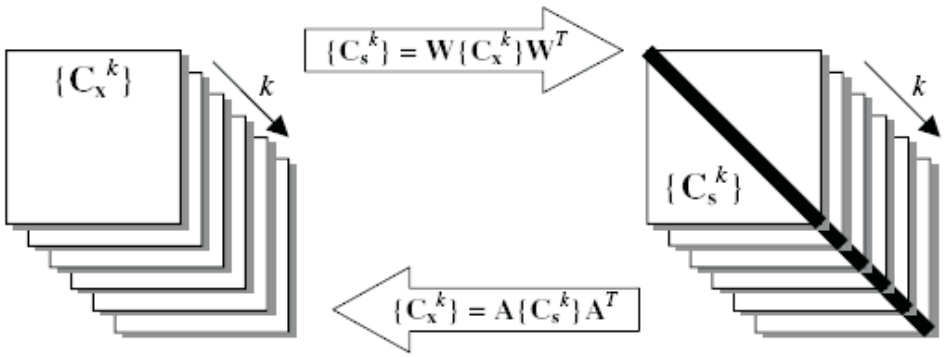


Figure 4.4: Relation of the two stacks $\mathbf{C}_x^k$ and $\mathbf{C}_s^k$ [9]

off-diagonal elements. The mixing matrix is then the inverse of the un-mixing matrix (if it exists).

Constrained ICA The main application of ICA in biomedical signals is usually to unmix the mixed complex multichannel signals and/or de-noise them. Thus, ICA is essentially used to obtain a particular desired source signal(s) from the available signal mixtures. This is done by selecting the component of interest from the set of components obtained from ICA and discarding the rest. As explained above, ICA is applied blindly to the mixture signals, i.e. without any prior knowledge of the source signals, or their number. The selection of components *a posteriori* of interest is not trivial and is carried out subjectively, using contextual knowledge. Also the algorithm is not guaranteed to converge to the global maxima due to random initialization of the weights and and other computational factors. Moreover, when the number of components are not known *a priori* and dimension reduction is not carried out (due to the risk of losing relevant information), then a conventional ICA would produce many components. If the number of underlying sources is actually less, these sources tend to be split among the ICA output signals. This would prevent any automated morphological analysis, as sometimes ICs could be unrecognizable as signals of interest. To assist in finding the source components, prior knowledge, in the form of frequency content, multi channel spatial information, temporal information, etc. can be quite useful. Such prior information may be available from different sources such as time series analysis, visual inspection, patient record, known physiologically relevant patterns, etc. Examples include ECG waveform morphology, rhythmic brain activity like seizures, artifacts as eye blinks, or muscular. It has been shown that such prior knowledge can be integrated within the ICA algorithm by few modifications in the estimation algorithm [9, 139, 10, 108, 140, 141, 107, 142], which helps to steer its spanning vector towards a desired direction. Such *a priori* information is provided in the form of a rough template, which is referred to as a *reference signal*. This is called *constrained ICA*, *cICA* or *ICA with reference*. The reference signal

carries information of the desired signal but is not necessarily identical to it. Such an extraction is termed a constrained optimization problem. The algorithm now introduces a measure of the closeness between the reference signal and the output in the ICA contrast function. However, it should be noted that constrained ICA partially relaxes the independence assumption for constrained sources, which are independent of the unconstrained sources but may be mutually dependent [140].

Constrained ICA can thus be helpful in the following ways:

- To identify the expected IC (source),
- To recover a source which may not have the maximum kurtosis/negentropy but is still an important independent source (based on the application)
- To assist in the convergence and locating a single source of interest.

Some ways of constraining ICA are explained below:

Temporally Constrained ICA The use of temporal constraints has been described in [142, 9]. The closeness constraint is given as

$$g(w) = \varepsilon(w) - \xi \leq 0, \tag{4.6}$$

where w is the un-mixing vector, ξ is a threshold and $\varepsilon(w)$ is the closeness between estimated output source and the reference signal.

The closeness can be measured by many ways such as,

- Mean squared error,
- Correlation,

As used in [9, 10], constrained ICA (cICA) problem can then be written as:

maximize

$$f(w) = \rho[E\{G(w^T x)\} - E\{G(v)\}]^2, \tag{4.7}$$

with constraint

$$g(w) \leq 0, h(w) = E\{y^2\} - 1, \tag{4.8}$$

and

$$E\{r^2\} - 1 = 0, \tag{4.9}$$

where r is the reference signal, E denotes expectation, w is the un-mixing vector, y is the estimated source vector, $g(w)$ is the closeness constraint, $h(w)$ is to constrain the output y to be of unit variance (second cumulant (variance) =1) and the reference is constrained to have unit variance. This problem can be solved using an augmented Lagrangian function.

The above method mentions the use of one constraint, but the number of references can be increased easily. Temporally constrained ICA has been successfully

used for artifact rejection (ocular artifact) in electromagnetic brain signal analysis [10]. Figure 4.5 from [10] gives a pictorial overview of this application.

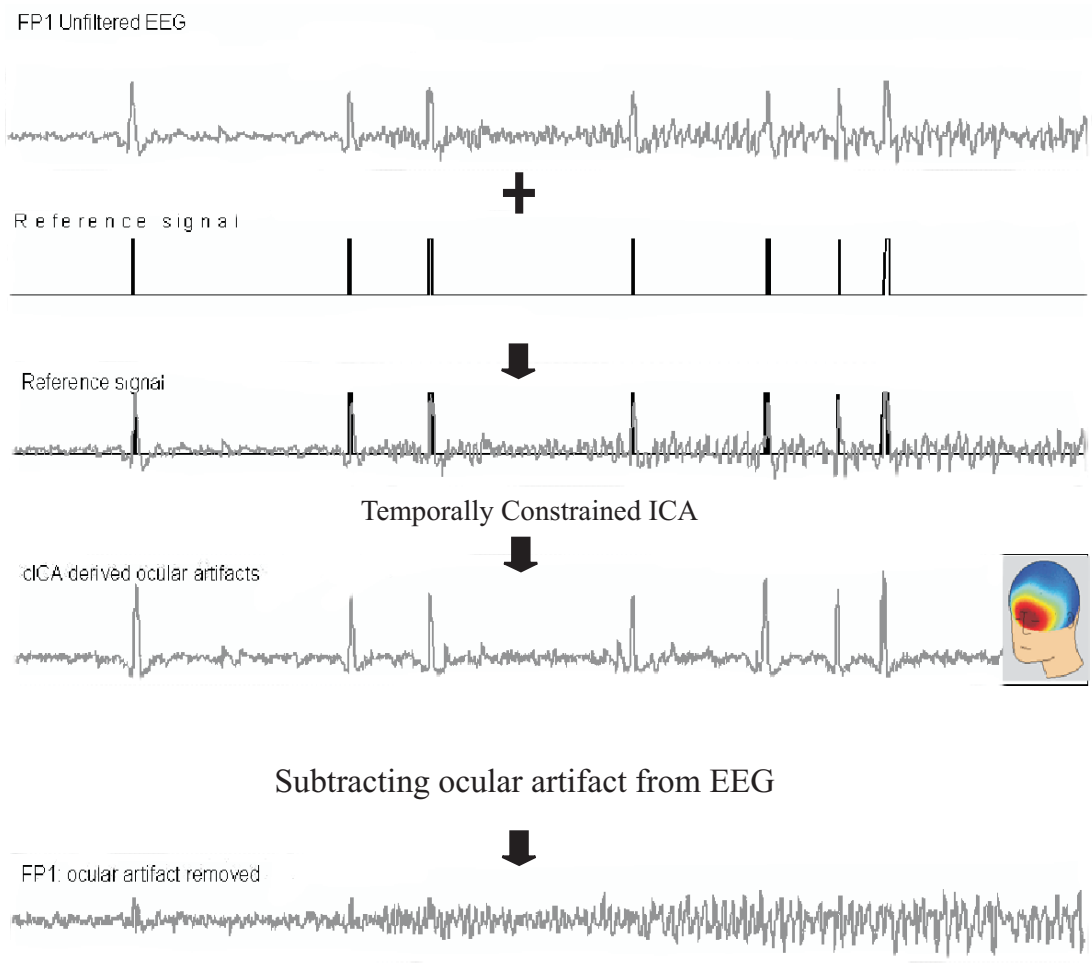


Figure 4.5: Temporally constrained ICA used for ocular artifact removal [10]

Algebraically, for constrained ICA [108] a reference channel can be added as an extra row to the data matrix $x(t)$ such that the new data is

$$\widehat{x}(t) = \begin{pmatrix} x(t) \\ r_1(t) \end{pmatrix}$$

where $r_1(t)$ is a reference vector. The extra row in the data matrix results in an extra column and extra row in the mixing matrix and the un-mixing matrix respectively. For an n channel system this is interpreted as: The first n elements of the extra mixing matrix column show the spatial distribution of the new IC (formed due to the extra row). The extra row in the mixing matrix shows the weighting of the corresponding IC's (given by the columns of the mixing matrix), thus depicting the contribution of topographies of all IC's due to the reference vector.

Spatially constrained ICA An expectation of the spatial information as is viewed by topographic maps are sometimes available in certain biomedical applications. This prior information can be integrated into the ICA model using the columns of the mixing matrix and is referred to as a spatial constraint. Essentially, they provide an initial value for one or more columns of the mixing matrix [9]. ICA which is essentially an optimization problem, can be influenced by changing the initial conditions. However, changing the initial conditions does not guarantee the convergence to a solution (finding the independent sources) but it may improve the solution [9, 140]. A spatial constraint is defined as *hard*, *soft* or *weak*, to reflect the degree of certainty about the accuracy of predetermined source signals and the extent to which these can be modified by the ICA algorithm [143]. A hard constraint column remains fixed, assuming maximum accuracy of the constraint topography. In contrast, a weak constraint reflects a higher uncertainty about the accuracy of the constraint, and is only used as an initial guess of the mixing matrix columns. A soft constraint has an intermediate degree of accuracy and its update is limited to a small region around the reference topography.

The spatial constraint could be obtained from the biophysical or physiological system under study, for example, in case of epileptic EEG, the patient's medical history or scan images can give information about the spatial topography (area of the brain) of interest, the eye blink artifact is known to mostly affect the EEG electrodes near the eyes and hence that topography has successfully been used to extract the ocular artifacts [140, 144]. Spatial constraints can also be obtained in conjunction with the conventional ICA source estimation process. That would involve manual selection of the topography of a source of interest when conventional ICA is applied to a segment of data in which the topography is clearer or easier to select manually. This is saved as a *spatial template* for application of spatially constrained ICA for further data analysis.

Spatially constrained ICA can also be incorporated with FastICA algorithm as shown by Hesse and James [140].

Spectrally constrained ICA Another variant of ICA is based on spectral constraints [9, 109]. It uses a reference signal that reflects neurophysiological prior knowledge of the sources in question, in the form of spectral content. For example, power spectrum of rhythmic EEG components such as alpha activity or epileptic seizures can be added as a reference channel(s), in the form of band pass filtered noise with the desired power spectrum.

Other constraints Other constraints that may prove useful when used with ICA are listed below:

- Time frequency: Changes in time-frequency together might be helpful, as in EEG different/combination of frequency bands pertain to different activities

and may get activated at different times. Observing/constraining the ICA algorithm in the direction of a certain time frequency change might prove helpful to detect a change/seizure or predict a seizure.

- Phase synchronization: Clinical studies have shown before that certain synchronization / de-synchronization keeps taking place when our brain is involved in activities. Observing and utilizing the changes in patterns of synchronization might help to constrain the ICA algorithm to look for a certain source of interest (seizure localization) or to detect/predict specific changes in brain, e.g. a seizure onset.

Constrained Temporal Decorrelation Temporal decorrelation has been found to be useful when used with temporal constraints [108] or spatial constraints [107] as well. Constraints help to provide an appropriate spanning basis which can help to obtain the topography of interest.

Constrained ICA and EEG: Constrained ICA has been used successfully with the EEG for artifact rejection, seizure onset analysis and waveform extraction, using spectral [145, 109], temporal [10, 146, 139] or spatial content [144, 143, 140]. James *et al.* [139] showed that even with a crude temporal reference waveform (derived from raw EEG/ MEG recordings), it was successful in ocular artifact rejection and seizure waveform extraction. They found that the algorithm repeatedly converged to the desired component within a few iterations, resulting in waveforms with expected morphologies and realistic spatial distributions. James and Gibson [10] also showed the use of temporally constrained ICA for artifact rejection in brain signal analysis. They also showed that the morphology of the reference signal did not prove critical in obtaining plausible solution with ICA, which is very useful for practical analysis.

4.3.3 Selection of ICA Algorithm

The many implementations of ICA in the literature have been used on different signals for varying applications. Low dimensional data that fits the ICA assumptions clearly are decomposed into the required components quite efficiently, by statistical algorithms such as FastICA, Infomax or JADE. The algorithms have their own particularities and parameter choices and hence may give slightly different results [147]. As mentioned by Delorme *et al.* [148], JADE is fast and stable for data with low number of channels, as it manipulates fourth order moments which becomes difficult to handle with high number of channels. They also mention that FastICA maybe less robust than Infomax ICA for high dimensional data (see also [147]). But statistical ICA algorithms on the whole have another limitation when applying on EEG data. As also pointed out by Delorme *et al.* [148], statistical algorithms may not give reliable results when less data is presented to them. This is because statistical ICA algorithms require statistically

adequate information about the functionally distinct processes to separate them. More data gives qualitatively consistent decompositions. In order to use longer data epochs of EEG the issue of spatial and temporal non-stationarity becomes more relevant, that would refute the basic ICA assumptions. In comparison, temporal decorrelation algorithms are more suitable for EEG decomposition as they do not require long data epochs.

4.3.4 Testing the Reliability of ICA Separation

The ICA algorithms have certain assumptions which may not hold true for real biomedical data. For instance, ICA assumes that the underlying source signals in the system are independent (and Gaussian). This is an assumption that may be partly true in real signals. Many times Gaussian noise contaminates the signals and is not easy to remove prior to ICA. This may still contaminate the extracted source signals. Another assumption for ICA is the linear mixing of source signals. In real complex biomedical systems such as the brain or the heart, strict linear mixing may not be true and presence of non-linear interactions can show up as remnant dependencies which cannot be removed by ICA [23, 25]. Also, sometimes the ICA algorithm is used with larger number of mixtures than the number of source signals present. This leads to a smaller number of sources being mapped as more output signals. Thus one original signal may be split in more than one source, and would be best described by considering all the split sources. These may be referred to as multi-dimensional sources. However, as the ICA algorithms lack permutation the parts of the multi-dimensional source may be spread apart, which is not trivial to identify or rectify.

Recently some studies have shown the existence of such remnant dependencies [23, 24, 25]. These have been supported by analysis on artificial data (using uniformly distributed noise, purely Gaussian noise, and uniformly distributed noise superimposed with Gaussian noise), where non-linearly mixed signals are shown to have remnant dependencies. The splitting of sources due to overfitting may be one of the causes of such dependencies. The presence of such remnant dependencies motivates the need to assess the quality and reliability of the separation of sources. Recently, Stögbauer, *et al.*, [23, 24], Himberg and Hyvrinen *et al.*, [149] and Meinecke, *et al.*,[150, 25] have described methods to test the reliability of ICA separation by various methods. Stögbauer, *et al.*, and Himberg *et al.*, have described the use of mutual information estimators for testing reliability of decomposition by ICA. They measure the dependency between components to check if they are mutually independent or have remnant dependencies. The variability of the mutual information with a rotational remixing is used as a measure of stability. They have shown that stable components show rapid growth of the mutual information under remixing, while a low variability of mutual information is shown by dependent components [23, 24]. Meinecke, *et al.*,[150, 25] have described

a bootstrap resampling approach where the time structure of the signals is preserved, such that they can be used with the temporal decorrelation algorithms. The resampling is carried out after the ICA separation to form a surrogate data set. This surrogate data is whitened and the source separation is applied again to estimate the rotation required to separate the sources. As the sources are already separated, the estimated rotation matrices are expected to be in the vicinity of the original mixing matrix. The variance of rotation parameters is gives a *separability matrix* [25, 150] which is used as a measure of instability of the separation. In other words, the separability matrix shows how reliably the one dimension subspace can be separated. An example of the separability matrix for a set of extracted sources is shown in Figure 4.6. This bootstrap

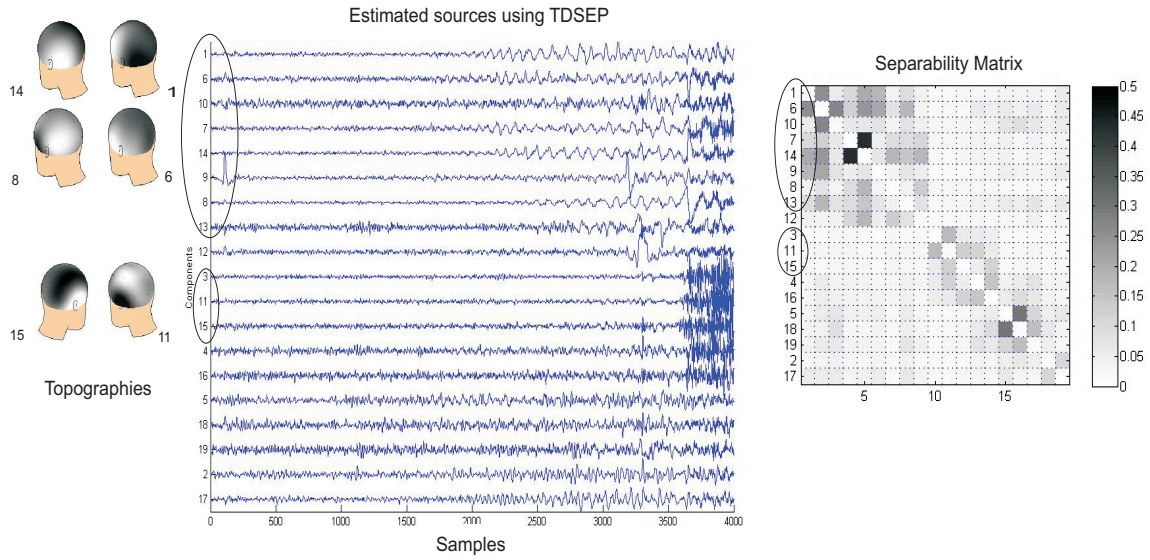


Figure 4.6: An example of the separability matrix for sources extracted by TDSEP is shown. The remnant dependencies are seen in the clusters of the IC's marked by ovals. Their corresponding topographies are also seen to be similar.

resampling approach and the corresponding separability matrix has been used in this thesis as a measure of remanent dependency in separated sources and the sources have been referred to as *least dependent components (LDCs)* as opposed to *independent components (ICs)*.

4.3.5 Perceived Limitations

There are some limitations with ICA. ICA is solvable up to a permutation, sign and power indeterminacy of the source signals. These suggest ambiguities in interpreting the results obtained from ICA rather than the accuracy of the model or source estimation. However, there are ways to overcome these perceived limitations.

Permutation: ICA algorithms output source signals in a random order. The source

signals obtained are not ordered in any specific way. As a result when ICA is applied to consecutive windows of a data segment, the sources recovered in one window segment will, in general, be in a different order than those of the previous window.

There are ways to work around the permutation problem depending on the signals and the application. For example, if the morphology of the source signals is identifiable in consecutive windows, then they can be manually post-ordered. Sometimes ordering can be manually imposed on the extracted sources, using the information of signal power or frequency. ICs from consecutive overlapped windows could also be matched using a statistic based on cross-correlation of temporal or/and topographic information [151, 152]. However, for long-term EEG decomposition, tracking source(s) across time remains a bottleneck in research. Tracking source(s) in long-term EEG, across time using ICA is not trivial, mainly because of the the permutation problem. EEG being a non-stationary signal, is assessed as windowed segments and ICA is applied separately on each epoch. When the morphology of the source signal is known/identifiable it may be possible to isolate the relevant ICs from consecutive window segments. But when the morphology of the source signal is not known or is changing or appearing/disappearing across time, then tracking it is very difficult, if not impossible. Also, in EEG the original source signals may be sparsely distributed across the epochs. Separate decomposition of an epoch may reveal a different set of sources according to their strength. To add to the above, the sources may keep changing with time (remaining stationary in a sub epoch but changing over different epochs), due to which ICA may become sensitive to underestimation (sources are fewer than there are mixtures, hence some of them may split) or overcompletness (sources are more than being estimated, hence some of them may get combined).

Power: The energy or variance of the ICs is not determined, as both source signals $\mathbf{S}$ and mixing matrix $\mathbf{A}$ are initially unknown, therefore multiplying a scalar in any source could get canceled by the corresponding column being divided by the same scalar, α_i [135]:

$$\mathbf{x} = \sum_i (\frac{1}{\alpha_i}\mathbf{a}_i)(s_i\alpha_i) \tag{4.10}$$

To overcome this indeterminacy of power,the magnitudes of the ICs s_i are fixed to unit variance. Then the power of the contributions of the sources to the mixtures can be recovered from the columns of the mixing matrix $\mathbf{A}$ [135, 9] as shown below:

$$p_j = \sqrt{\frac{1}{m}\sum_{i=1}^{n}(\mathbf{a}_i^j)^2} \tag{4.11}$$

where a_{ij} is the ith element of column j of the mixing matrix $\mathbf{A}$ and p_j is the RMS power for independent source $j(1 \leq j \leq n)$

Sign: Similar to the case of power ambiguity, the sign is difficult to be determined from the output of ICA. Usually it does not pose much of a problem and is not insurmountable.

Number of sources: ICA also has the usual problems of overestimation and underestimation when recovering original sources without knowing the number of sources to be recovered, which is usual in biomedical applications. The knowledge of the number of sources to be estimated can impact the accuracy of the ICA solution.

Various statistical techniques have been used to estimate the number of sources, based on determining the dominant eigenvalues of the data covariance matrix [135]. This is referred to as model order selection or dimension reduction.

Model order selection: Model order selection is an attempt to estimate the number of dominant underlying sources in the mixed signals. One of the approaches is based on using the dominant eigenvalues in the covariance matrix of the mixture data. The dominant eigenvalues contribute more than a minimum proportion of variance. For example, Figure 4.7 shows a plot of ordered eigenvalues of the covariance matrix. The shape of the curve can be used to identify the eigenvalues that contribute the least. This gives an initial guess of the number of dominant sources and hence reduces the dimension of the unknowns. These methods are essentially trying to determine the noisy subspace of the eigenvalue spectrum but it can run into problems when there are noise-free signals or signals with Gaussian noise. Especially, in case of recovering weak signals embedded in stronger noisier sources, the dimension reduction methods may even remove the source(s) of interest [108], defeating the purpose of ICA.

Selection of Components: The main challenge with ICA, especially for biomedical signals is, interpreting the sources recovered and relating them to their physiological meanings. This is usually a subjective process, where sources are selected using the temporal, spectral or spatial information of the signal. For example, ocular or cardiac artifacts can be isolated from EEG sources, as their spatial topography (frontal lobe, near the eyes) and temporal morphology is known *a priori*. Nevertheless, selection of components remains highly subjective, as little is usually known about the morphology or topography of most of the underlying source signals that are obtained from EEG.

Nevertheless, ICA is a useful tool to unmix EEG signals and has gained attention in EEG signal processing [9, 94, 139, 109, 104, 130, 152, 153] and in epileptic seizure detection and prediction as well [130, 9, 153]. Although, the interpretation of the recovered sources and tracking them across time remains a challenge. Following sections give a few basic examples of using ICA to unmix and de-noise EEG.

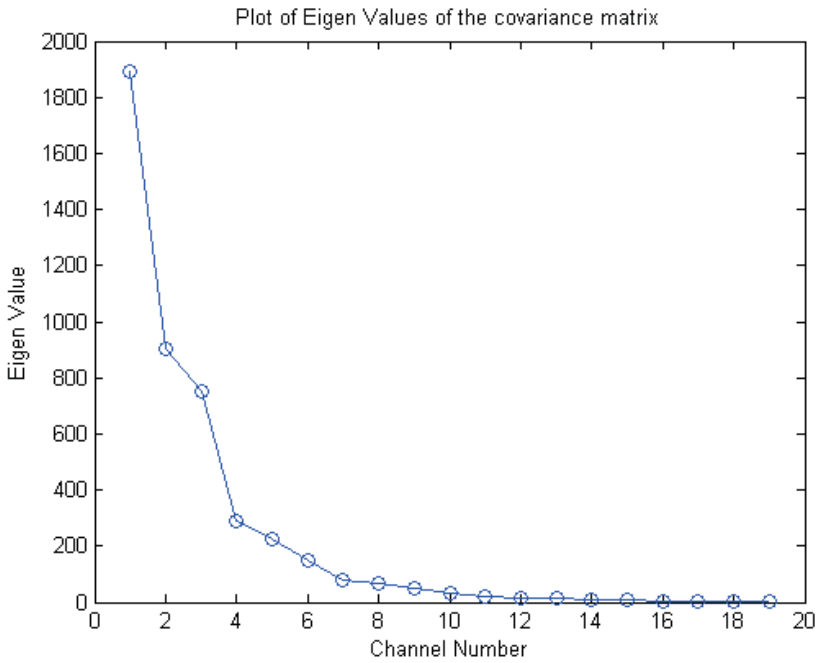


Figure 4.7: Plot of ordered eigenvalues of covariance matrix of a set of 19 EEG signals. The shape of the plot shows the presence of 8-9 dominant sources, while the rest of the sources (9-19) do not have much variance and may be discarded in the ICA application.

4.3.6 ICA for Unmixing Epileptiform EEG

In the literature the application of ICA has been found useful in seizure detection [107], artifact rejection [10, 139] and seizure onset analysis [141]. These studies with conventional / constrained ICA indicate that ICA/cICA is helpful for de-noising pre-seizure EEG and may even unveil some information which may help in the prediction of a seizure. This idea is supported by the fact that when ICA is used for epileptiform EEG analysis, the seizure topography of an identified component (identified subjectively) is seen to reflect the topography as seen during the subsequent seizure, sometime (8-10 seconds) prior to the seizure [9]. This assumes that the seizure focus is spatially fixed or at least varying very slowly.

ICA has been used in the literature for unmixing EEG in order to analyze event related activity of the brain [148, 154, 155], for improving the EEG based classification of features [156] or analyzing the EEG for extracting artifacts or signals of neurophysiological interest [9, 157, 158]. A comparison of statistical ICA and temporal decorrelation applied for unmixing EEG in the context of epileptic source extraction will be discussed using two examples. Examples with two types of epileptic segments are shown next. The first dataset is a 20 second multichannel EEG recording with an ictal episode starting seven seconds into the segment (marked with a vertical line). It is a 19 channel scalp recording from the standard 10-20 electrode placement system, with a sampling frequency of 200 Hz. The seizure has a left temporal focus which can be seen as the

onset of rhythmic high frequency waves at EEG electrodes F7, T3, T5 (see Figure 4.8). The ICs obtained with conventional ICA and Temporal Decorrelation are shown in Figures 4.9 and 4.10. The artifact source signals have been identified subjectively using temporal and spatial information. The ocular artifacts have a typical morphology and have a frontal topography. The muscle artifacts have a high frequency waveform with a temporal lobe focussed topography, similarly the seizure sources have been identified manually by looking for the source with a rhythmic onset around seven seconds into the signals. The seizure is of left temporal lobe origin and hence the expected seizure source topographies are left temporal lobe as well. The seizure and the ocular artifact (eye blink) are observed to be temporally overlapping at the area marked. Here, it can be seen that FastICA is able to separate out one prominent seizure component as spatially distinct irrespective of the rhythms involved in that seizure IC, whilst TD-SEP is able to describe the seizure source with more detail with respect to temporal distinctions, even if they share similar spatial patterns. The topographic plots show that ICs 1, 6, 10, 7, 14 have similar spatial distribution.

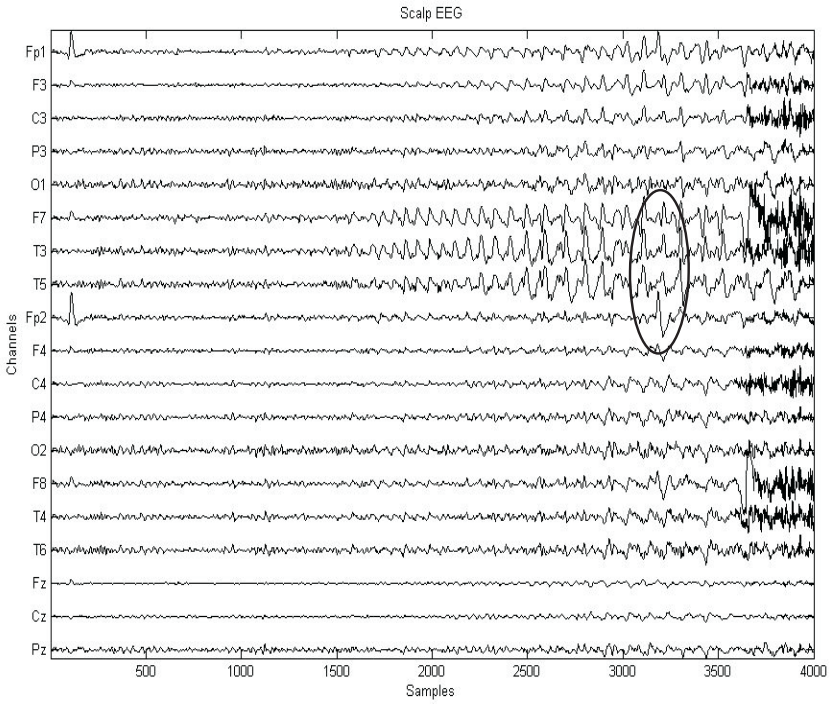


Figure 4.8: Example 1: The figure shows multichannel scalp EEG of 20 seconds, with seizure onset is at sample 1500. Seizure is observed to be prominent in the left temporal lobe at onset with clear rhythmic waves with increasing amplitude at channel F7, T3 and T5. Ocular artifacts such as eye blinks are also observed to have a transiently similar morphology as the seizure waves, but concentrated at the frontal channels Fp1 and Fp2. The encircled segment of the EEG shows temporal similarity in the ocular artifact and seizure signal.

Example 2 is for an EEG with frontal seizure (see Figure 4.11). In this case the seizure

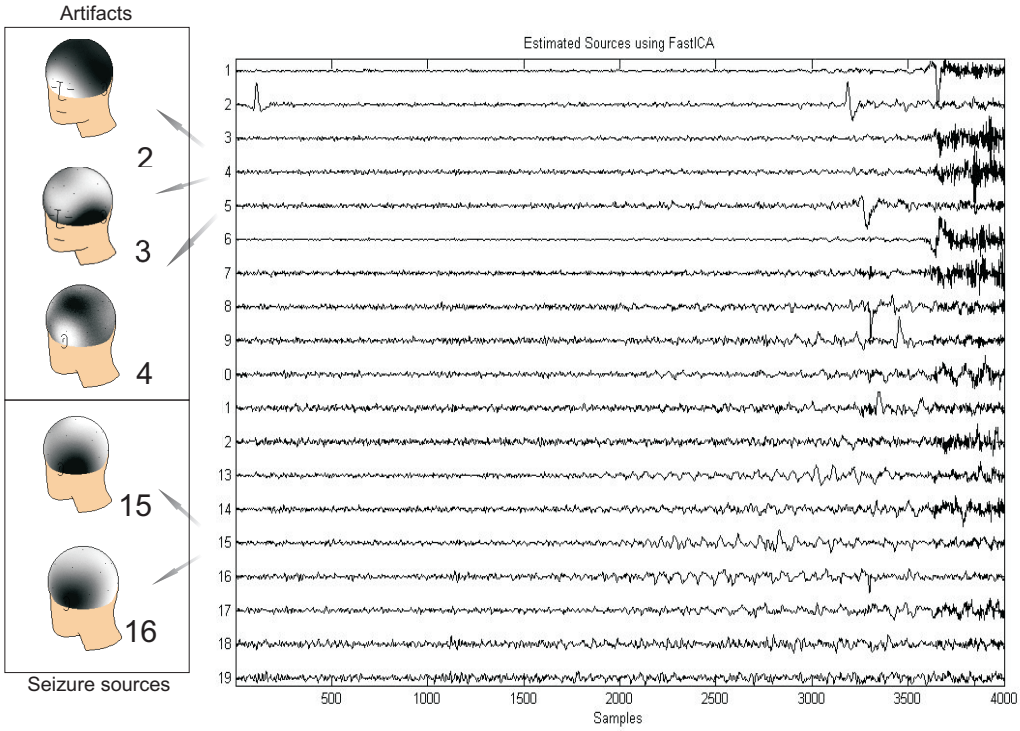


Figure 4.9: Sources estimated using FastICA for EEG seizure example 1 and topographic maps for seizure source and artifact sources (ocular/muscular)

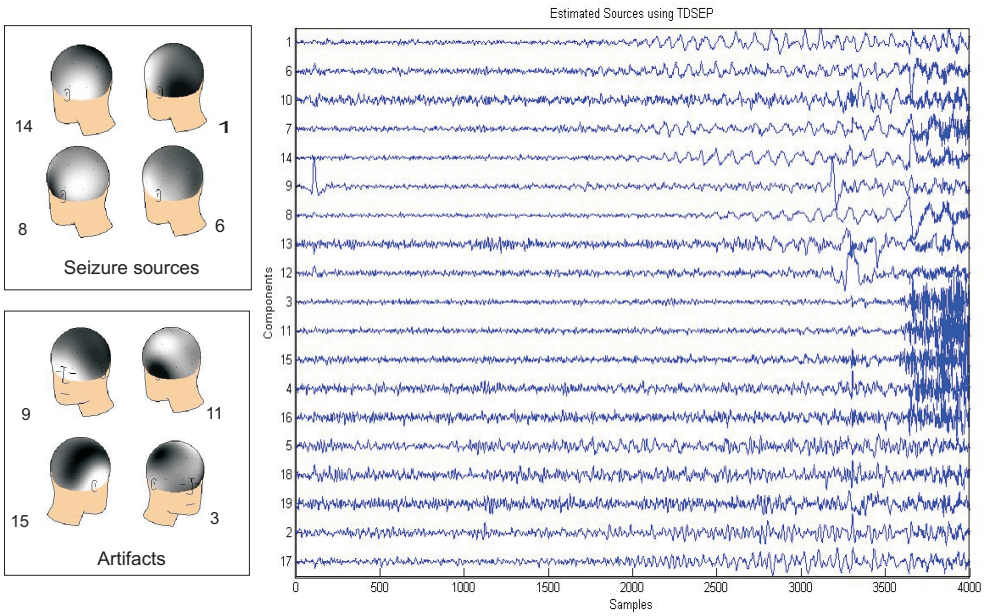


Figure 4.10: Sources estimated using TDSEP for EEG seizure example 1 and and topographic maps for seizure source and artifact sources (ocular/muscular)

source topography will be expected to be frontal, along with a rhythmic temporal waveform. This EEG segment, also has ocular artifacts that will have frontal topography as well. Hence the data has, apart from other sources, two prominent physiologically independent sources, with spatial overlapping but temporal distinction. The extent of separability of the ocular artifact and the frontal seizure source is shown here using conventional ICA and temporal decorrelation. It can be seen in Figures 4.11 and 4.12 that conventional ICA as well as temporal decorrelation are both equally efficient in extracting the seizure source and isolating it from the ocular artifact. The seizure source has been isolated quite well, when it was not even visible in the raw EEG. The IC's obtained with FastICA and TDSEP are been shown in Figure 4.11 and Figure 4.12. The seizure and the ocular artifact source and its topography are shown along with the IC's as well.

4.3.7 Application 2: ICA for De-noising EEG

The application of ICA to EEG helps to unmix EEG and in turn helps to isolate the artifacts and noise signals. Removing such signals from the multichannel EEG is not trivial, as the noise and artifacts are mixed with the EEG in a complex manner over different frequency bands and across different channels at various times. Simply filtering the EEG with linear filters may also remove information that is useful for feature extraction and further analysis. ICA is very useful for de-noising the EEG suited to the requisite analysis. For de-noising EEG, the noise components are selected either subjectively or objectively after the application of ICA. The columns corresponding to the noisy components/sources are zeroed, giving a new mixing matrix. The source signals are then multiplied by the new mixing matrix to obtain the de-noised mixed signals (EEG). An example of raw EEG and the de-noising is shown in Figure 4.13(a). The multi-channel EEG has a right parietal focus. It is de-noised with respect to epileptic sources. As can be seen, the de-noised EEG has prominent epileptic activity and the artifacts and noise have been removed. The cross-correlation between the de-noised and the original EEG has been plotted on the scalp topography 4.13(b). It shows that the de-noising has been more prominent in the areas away from the seizure focus area. The seizure focus areas have been least affected by de-noising.

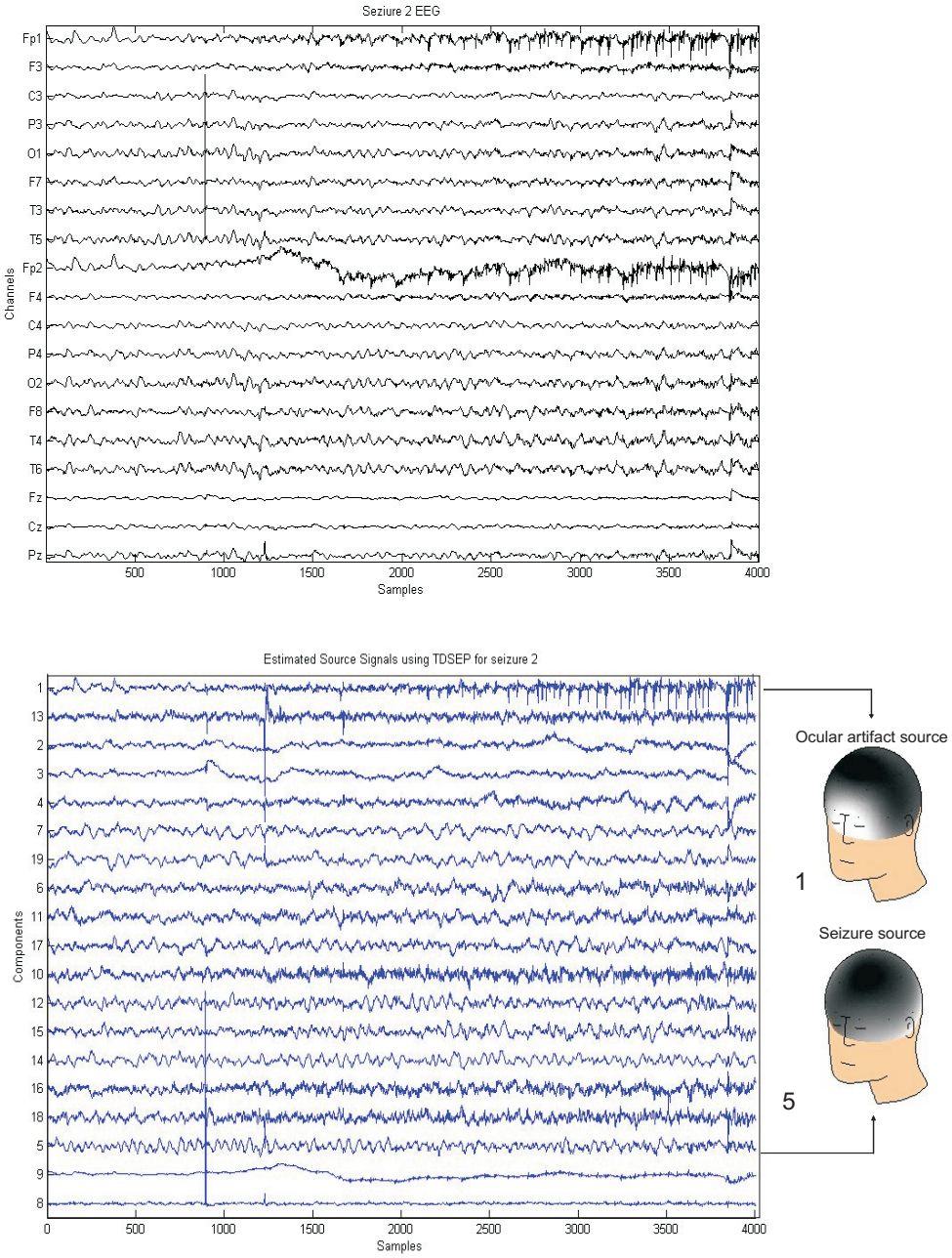


Figure 4.11: Sources estimated using TDSEP for EEG with frontal seizure and topographic maps of the seizure source and the ocular artifact

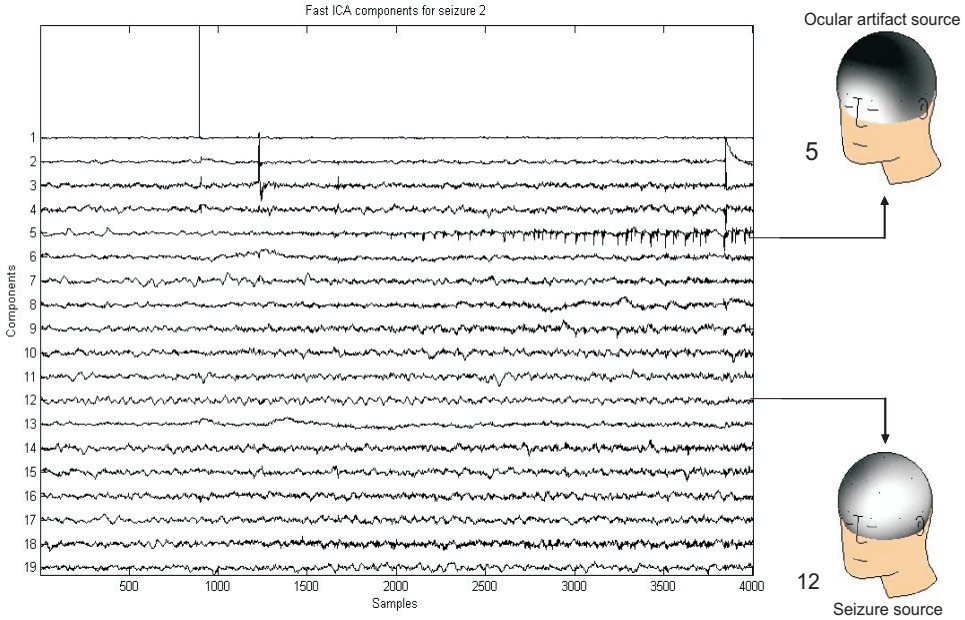


Figure 4.12: Sources estimated using FastICA for EEG with frontal seizure and topographic maps of the seizure source and the ocular artifact

4.4 Preprocessing 2: Spatial Filtering of the EEG

Chapter 1 mentions, that for seizure prediction, it is necessary to investigate dynamics of continuous long term EEG (of a few days), as opposed to short stand-alone segments (of a few minutes) each. To analyze such continuous EEG, the need to assess stationarity is required for statistical measures. The EEG is essentially a non-stationary signal and hence to analyze continuous EEG a windowed approach is used. EEG is segmented into short segments varying from a few seconds or minutes (depending on the stationarity of the aspect being analyzed in the EEG) and the statistical measures are calculated on these segments.

4.4.1 Tracking a Seizure Source across Time

Thus, for analyzing dynamics of the EEG across time, the pre-processing and extraction of a source signal is carried out on windowed segments. It is to be remembered that ICA will be applied independently on each segment and as has already been stated earlier, the sources extracted will appear in a non-specific order, and a source of interest will have to be manually/automatically identified for each segment being analyzed. The de-noising and source extraction is possible when the data window contains ictal EEG (EEG with a seizure), as described in the previous section. As in such a windowed segment, the source (in this case the seizure signal) has a known morphology

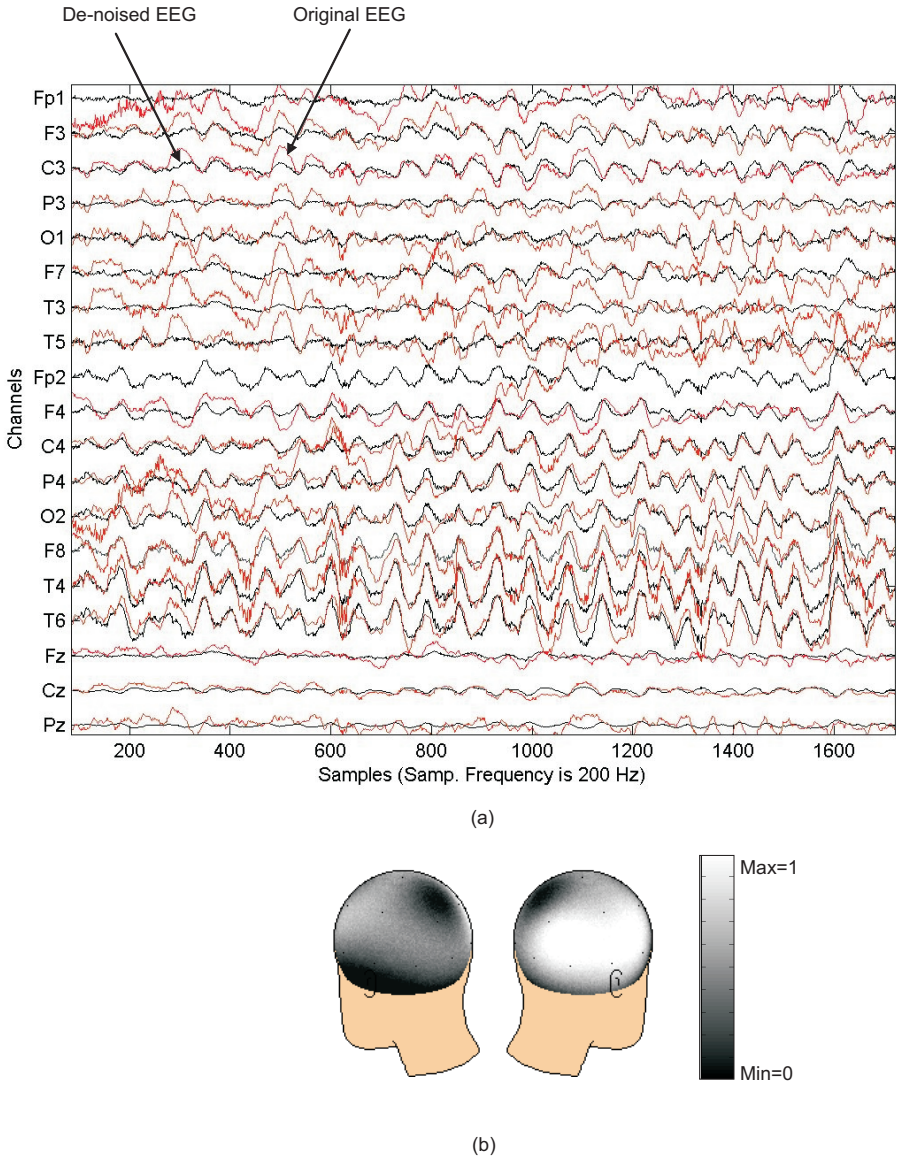


Figure 4.13: (a)A 20 second segment of ictal EEG (with a right parietal focus) de-noised by ICA, (b) Correlation between original and de-noised EEG is shown topographically

(rhythmic/spike and wave/sudden high frequency onset) and as at the time of seizure, the signal gets stronger, it becomes more probable for it to be estimated as a separate source by ICA, making it easier to be identified manually. The problem arises when the segments away from the seizure (pre/post) are being analyzed, in order to track a seizure source signal across time.

This has been illustrated with an example in Figures 4.14, 4.15, 4.16, 4.17, 4.18. Figure 4.14 shows an EEG segment with a *right parietal* seizure onset marked by a vertical line. It is a 19 channel scalp EEG of epileptic patients undergoing pre-surgical evaluation for possible surgery for epilepsy. The data is sampled at 200 Hz, and is present in the form of two minute segments with a gap of a quarter of an hour. It is completely anonymized and lacks annotations and medical reports. An EEG segment of two minutes that

included a visually apparent focal ictal episode (seizure) was further segmented into non-overlapping 5 second windows (1000 samples) appropriate for stationarity. The data was pre-processed before analysis by removing the mean and making the variance of signal segment equal to one. It was then filtered with a low pass filter (0-50 Hz). ICA was first run on the window with the seizure. The topographies (Figure 4.15 (b)) were then analyzed visually along with the temporal source estimates (Figure 4.15 (a)), to identify the seizure source. The seizure source signals and topographies identified in this example are for IC 8 and IC 13.

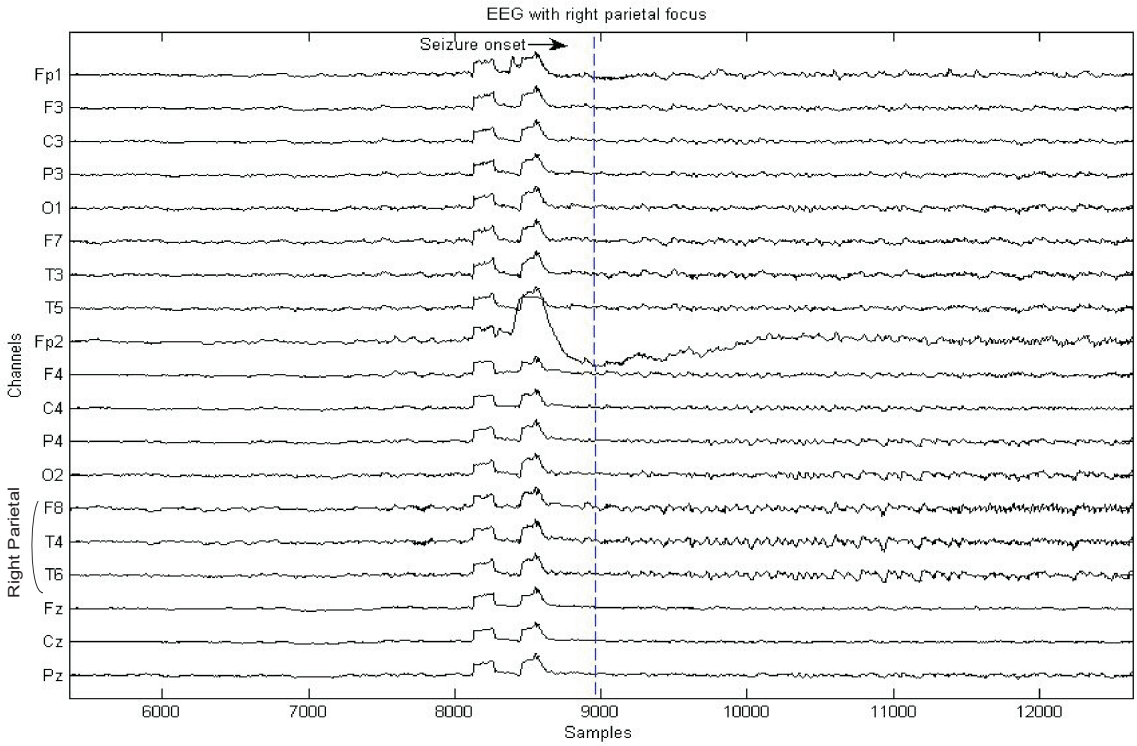


Figure 4.14: Tracking seizure source across time: Scalp EEG segment with 19 channels with a right parietal seizure (seizure onset marked by the vertical line),

The ICA algorithm was then run on the preceding window segments without dimensional reduction. This gave 19 LDC estimates for each preceding window. A separability matrix is also formed, using the resampling approach, between the window sources to check how well the sources are separated. Any residual dependance among sources can be seen from the groups formed in the separability matrix. The topographies and source signals obtained for preceding windows were analyzed visually to find any similarity or pattern that may enable to track the seizure source. As can be observed in Figure 4.16 and 4.17, 4.18 it is not trivial to identify the seizure source preceding the seizure. As the temporal morphology of seizure source is no longer identifiable, identifying the source by topography alone can be quite misleading. This is because, firstly there is no assurance that the seizure source remains fixed spatially over longer time

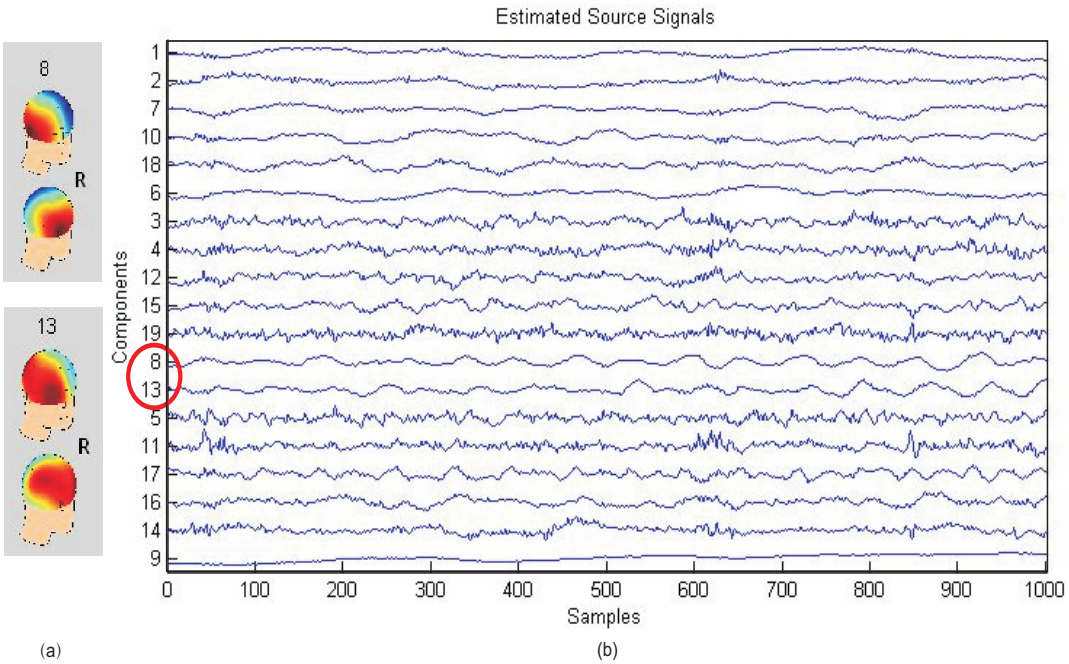


Figure 4.15: Tracking seizure source across time: ICA estimates for the ictal segment (samples 9000-10000) and their respective topographies

spans, and secondly, there are other artifacts/signals that may show similar spatial distribution.

The above was studied for 12 patients with generalized and focal seizures, on short EEG segments, and the difficulty in tracking the seizure source across time, remained a challenge. In the absence of a known signal morphology, it is difficult to identify the seizure source because of the following reasons:

- **The temporal morphology of seizure source** signal before the clinical seizure is unknown. This is not known prior to analysis from either the engineering or medical literature. The topography can be useful in such as case, for example, if the spatial topography of the seizure signal is known, it might be possible to manually extract a source signal that correlates a target topography. But as the process will need to be automated due to the vastness of the data (3-4 days of continuous multichannel EEG (sampled at 200 Hz) segmented with 2 minute windows), such correlation will face further obstacles as mentioned next.
- **The seizure source may be very weak** in segments prior or post seizure, such that sometimes it might not be strong enough to be extracted as a separate source signal by conventional ICA, making its morphological/topographical identification impossible.
- **The over-fitting** occurring in the independent application of ICA on segments prior to seizure, can also lead to splitting of source signals inconsistently for all

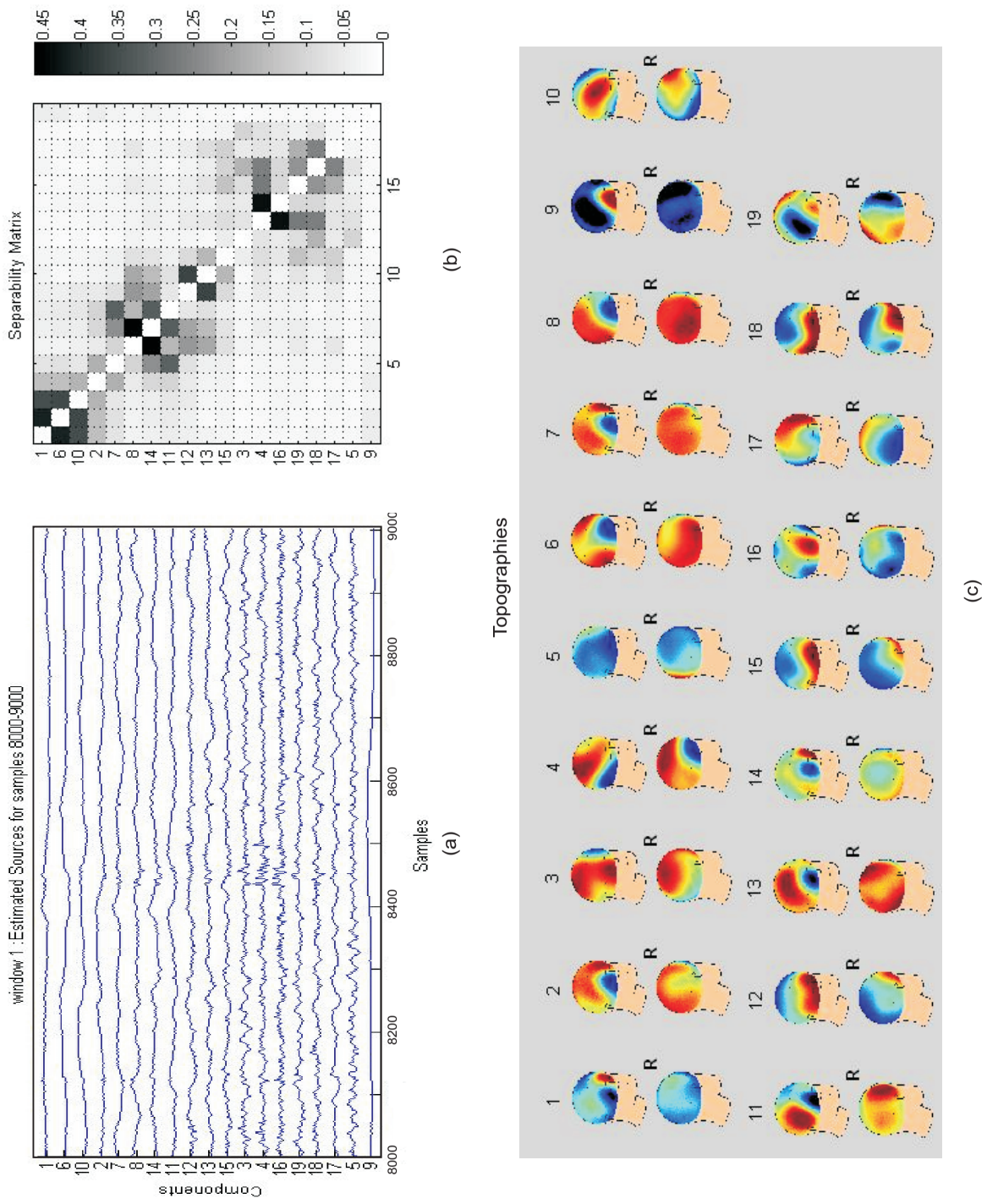


Figure 4.16: (a) Estimated sources for preceding window 1 (samples 8000-9000), (b) Separability matrix and (c) Source topographies

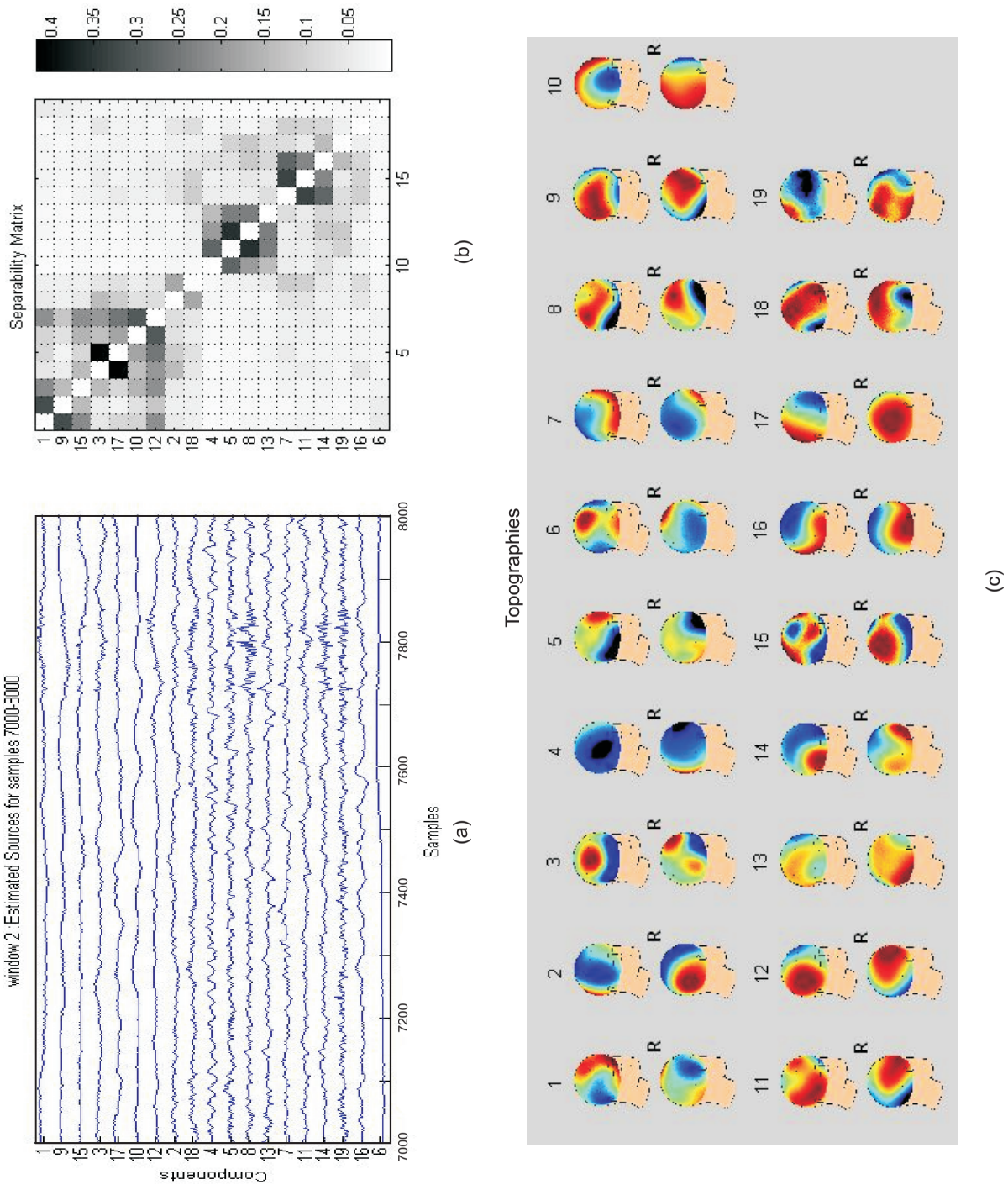


Figure 4.17: (a) Estimated sources for preceding window 2 (samples 7000-8000), (b) Separability matrix and (c) Source topographies

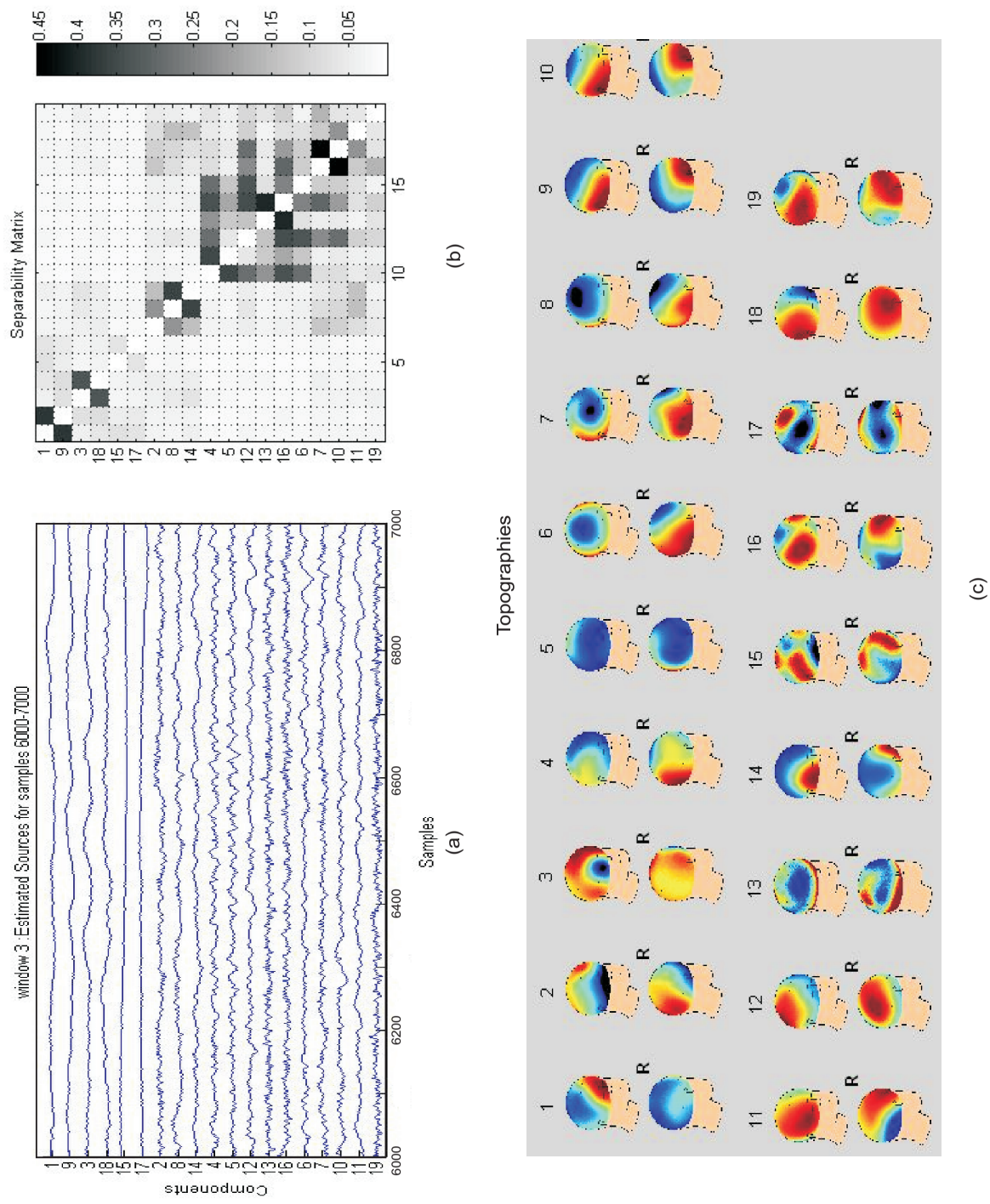


Figure 4.18: (a) Estimated sources for preceding window 3 (samples 6000-7000), (b) Separability matrix and (c) Source topographies

windows making it difficult to identify and extract a signal of interest by either temporal or topographical means.

- **The seizure source may be moving spatially** over longer periods of time. It could be appearing / disappearing over time.
- **The number of sources may be changing** over longer periods of time, for example, in one day artifacts such as chewing, electrode pop, twitching, and other brain activity may occur and stop in few windows, making sources appear and disappear over time.

Considering these problems with extracting seizure sources from continuous EEG, it is felt that a constrained ICA solution would be more appropriate (as proposed by Hesse and James [144]). The possible type of constraints known to be used with ICA are temporal, spectral and spatial. Prior to a seizure, the temporal and spectral content of the seizure source may still be arbitrary. However, a spatial constraint would be useful as it could localize the search for the seizure source to the area around the suspected seizure focus. Essentially, in this case the dynamics and sources in a particular area of the brain will be tracked. In spatially constrained ICA, the indices of a column of the mixing matrix are fixed (with a choice of rigidity and span of search). This prevents the source signal being split up. Also, most importantly, spatially constrained ICA identifies the source signal and separates it automatically, overcoming the problem of identifying it from a non-specific set of extracted sources in each data segment. As the name implies, spatially constrained ICA has in-built aspect of correlating target topographies when searching for the source signals, while decorrelating the topography with other extracted sources as well.

4.4.2 Creating a Spatial Template

To apply spatially constrained ICA, there is a need for a spatial template. A spatial template can be obtained by using the prior information of the biophysical system, in this case the epileptic brain. Also, it is possible to obtain it through conventional ICA. A selected segment of data is taken, where a source signal is known to exist. Conventional ICA is applied to this segment of data and the source of interest is identified; its topography (column of the mixing matrix) is then retained as a spatial template to be used for other segments of data. In the case of epileptiform EEG analysis, an EEG segment of a patient with a clearly identifiable seizure onset is used. Conventional ICA is used on this segment without dimensional reduction (to prevent any potential loss of information). The seizure source is identified by manually correlating the temporal changes in the source signals with the seizure onset time or/and correlating the topography with the knowledge of the type of seizure, such as left temporal/right parietal, etc. The corresponding column of the mixing matrix is then used as the spatial

template. This template can be used to run spatially constrained ICA for other data segments of the same patient, to obtain the source signals possessing a similar spatial distribution (i.e. in the same area of the brain).

4.4.3 Spatial Filtering

The spatial template obtained above is used for spatially constrained ICA on further segments. Essentially such a procedure is equivalent to spatially filtering the source signals. It is clear that the spatial information in the EEG is useful in epileptic EEG analysis. This is still a subjective process and dependent on external knowledge from patient history (used in source selection).

4.4.4 Tracking Temporal Dynamics with Spatially Constrained ICs

Spatially constrained ICs obtained as described in the previous section can now be analyzed for patterns or dynamics that may indicate an approaching seizure onset. One way to analyze the ICs is by tracking the power of the sources. It should be noted, that in raw EEG, the seizure source prior to a seizure may be quite weak (low power), but as ICA extracts source signals irrespective of the power, the low power signal is well preserved. It should also be noted that ICA has a limitation of being unable to extract the actual power information of the signals. However, there is possibility to calculate the power of extracted sources as has been mentioned in Section 4.3.5.

The tracking of power for spatially constrained ICs is illustrated by a few examples here. It is whitened and low pass filtered (0-50 Hz). ICA is applied on an ictal segment having a visually apparent focal seizure. The topographies are analyzed visually along with the ICs to isolate the seizure source. The normalized topography of this identified seizure source is then saved as the *spatial constraint template* for the particular patient. Spatially constrained ICA is then run on the preceding EEG segments (without dimensional reduction) to obtain a spatially constrained source for each segment. The topography is normalized and the power information is transferred to the source signal as mentioned above. The source signals are then concatenated together to give a continuous temporal view of the constrained source. The results from two seizures of one patient have been shown here (see Figures 4.19 and 4.20).

In seizure 1 it can be observed that there is a sudden increase in the power of the constrained source signal at 7800 samples. The source also shows some rhythmicity from about 7800 samples before seizure onset i.e 6 seconds prior to any electrographical change is seen on the EEG. Such an increase in power and rhythmicity is also seen in seizure 2 of the same patient at about 12.5 seconds prior to seizure onset.

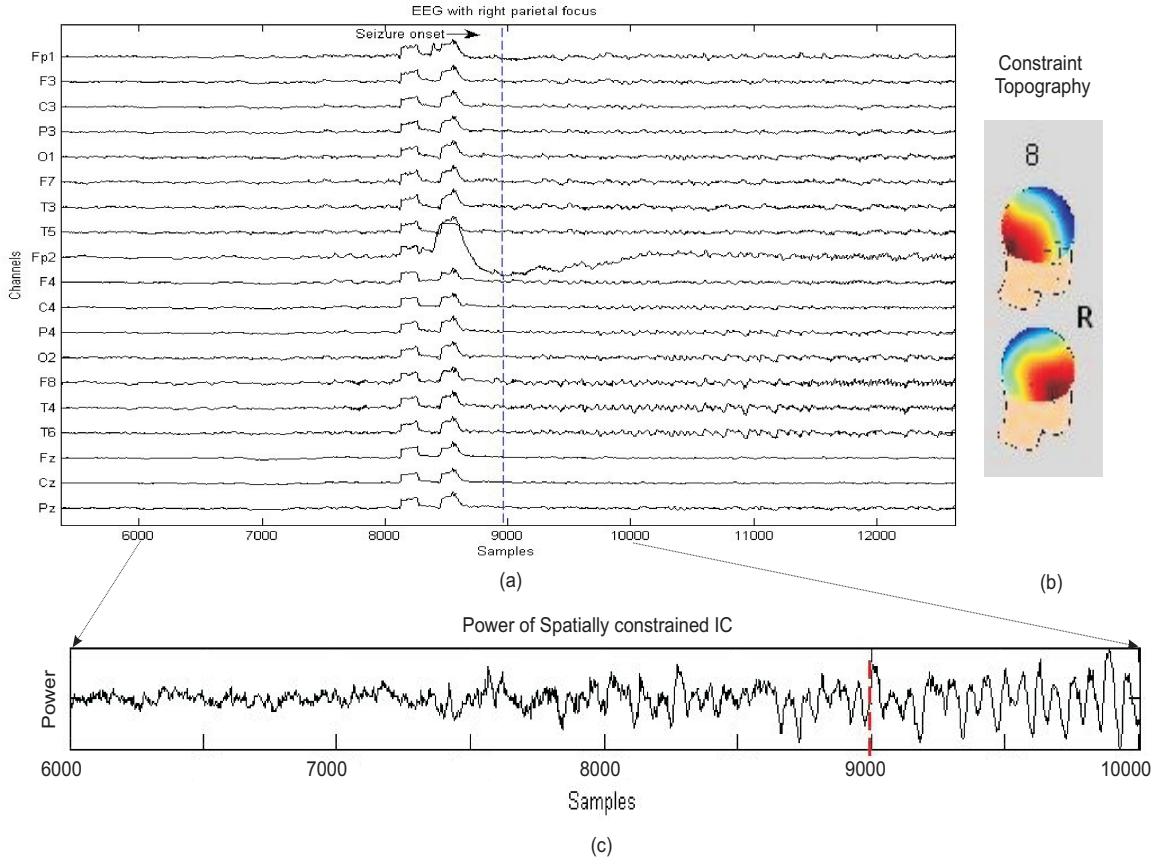


Figure 4.19: Seizure 1 Patient 1: (a) EEG with the seizure onset marked,(b) Topography of the selected seizure source to be used as constraint,(c) Power of the spatially constrained source from three windows preceding the seizure onset. A rhythmic signal can be observed 6 seconds prior to seizure onset. Patterns prior to that are difficult to ascertain visually.

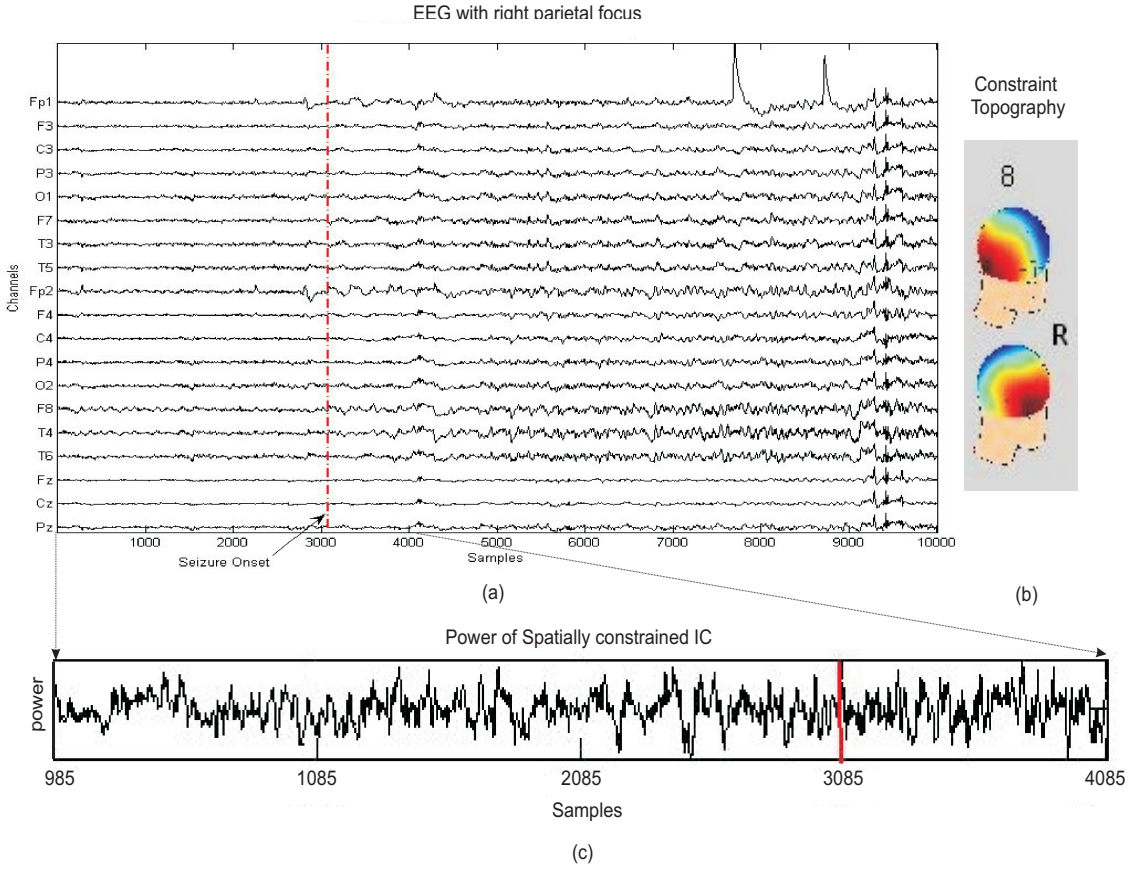


Figure 4.20: Seizure 2 Patient 1: (a) EEG with the seizure onset marked,(b) Topography of the selected seizure source to be used as constraint,(c) Power of the spatially constrained source from three windows preceding the seizure onset. Rhythmic signal can be observed 12.5 seconds prior to seizure onset, however patterns prior to that are difficult to ascertain.

4.4.5 Tracking Spatial Dynamics with De-Noised EEG

In contrast to the above method where the power of one source is tracked, another method to track changes in the EEG is by looking at the change in spatial distribution of a combination of many possible seizure sources across time. The EEG is first unmixed with ICA. The artifact sources are then identified and removed from the mixing matrix. Re-mixing the sources using the updated mixing matrix gives de-noised EEG signals. These signals now mainly consist of the seizure signals. Projecting their distribution at any point in time gives the spatial distribution of the possible seizure sources across time. This is shown with the help of an example here. Using the same data as above, a 50 second of 19 channel scalp EEG is used (see Figure 4.21). The data is pre-processed before analysis by removing mean and making variance of signal segment equal to one. It is then filtered with a low pass filter (0-50 Hz). ICA is run on the ictal segment (50 seconds, 10000 samples) having a visually apparent focal seizure. The segment selected is quite long in order to capture the maximum types of artifacts. In terms of stationarity of the segment it is assumed that the artifacts can be considered stationary over this time length. The sources for the artifacts/noise are selected using temporal or topography information. The mixing matrix $\mathbf{A}$ is updated by zeroing the columns corresponding to the artifact topographies. This retains the possible seizure sources only. The updated mixing matrix is then multiplied with the ICs to obtain a new set of mixtures in which the artifacts are filtered out. This de-noises the EEG. The topographies of this new mixture are observed at different points in time to observe how the distribution of the sources changes with time.

The ICs, artifact/noise sources identified (ocular, electrode movement, EMG, line noise), and the separability matrix are shown in Figure 4.22. The separability matrix shows the groups of estimated sources that have residual dependence in some form. The artifacts isolated are sources 1 and 9 for ocular artifact, sources 2 and 3 for EMG artifact, sources 11, 13, 4 for artifacts (transients perhaps due to movement or interference) and source 8 for line noise. These selected artifacts are then used to de-noise the EEG. The potentials across the de-noised EEG are then used to observe changes in the spatial distribution of the source potentials, shown in Figure 4.23. A gradual change in potentials is observed in the region (spatial) of seizure origin (right parietal) as the seizure approaches, but it is difficult to ascertain and quite subjective. It is clear that an objective method is needed to standardize and measure such changes reliably.

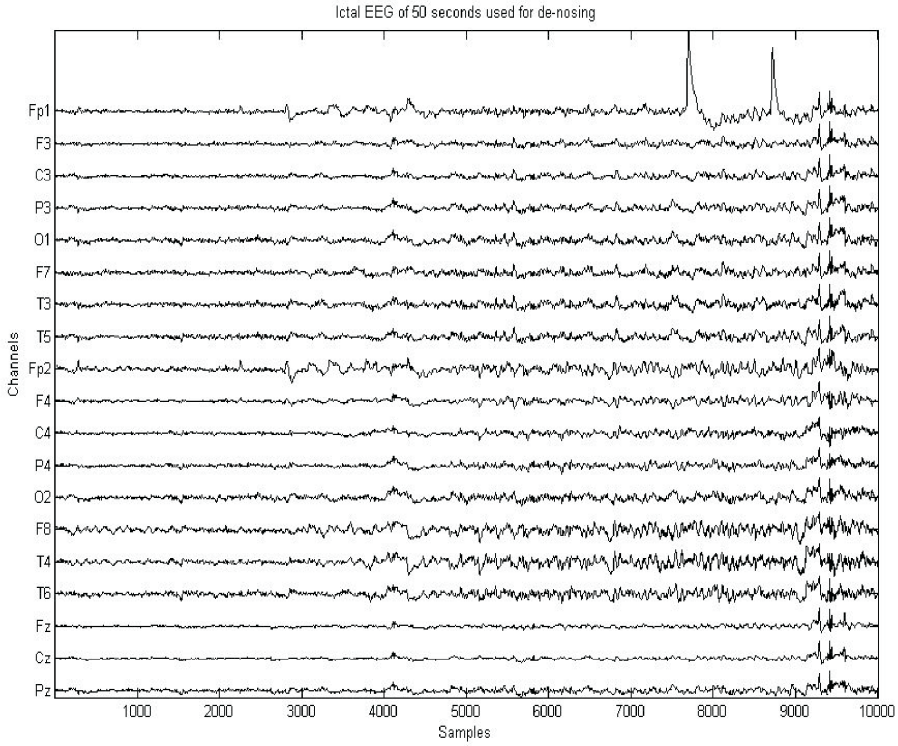


Figure 4.21: EEG segment used for finding artifacts for de-noising EEG

4.5 Discussion

The above review shows that using blind source separation techniques - ICA, cICA and Temporal Decorrelation, can be useful for de-noising multichannel EEG and extracting sources of interest. In the context of seizure prediction, ICA will be useful for multichannel continuous epileptiform EEG, in the following ways:

- De-noising epileptic EEG to isolate artifacts (muscle, eye, electrode movement, etc.) and noise that contaminate the epileptic EEG during seizure/pre-seizure and are mixed in a complex manner across channels and frequencies. These artifacts are not easily filtered using traditional filtering methods.
- Estimating underlying sources by decomposition of multichannel EEG to extract a source of interest from a linear instantaneous mixture of brain signals. ICA helps to decompose the signal mixtures (EEG), to estimate the underlying sources. This is a non-trivial problem as the inverse problem of EEG (acquiring the sources from EEG mixtures) does not have a unique solution.
- Localizing a source of interest spatially (on the scalp) is an additional advantage of using ICA. This in turn can help to localize the seizure focus when coupled with other techniques like beamforming. ICA provides spatial information of the

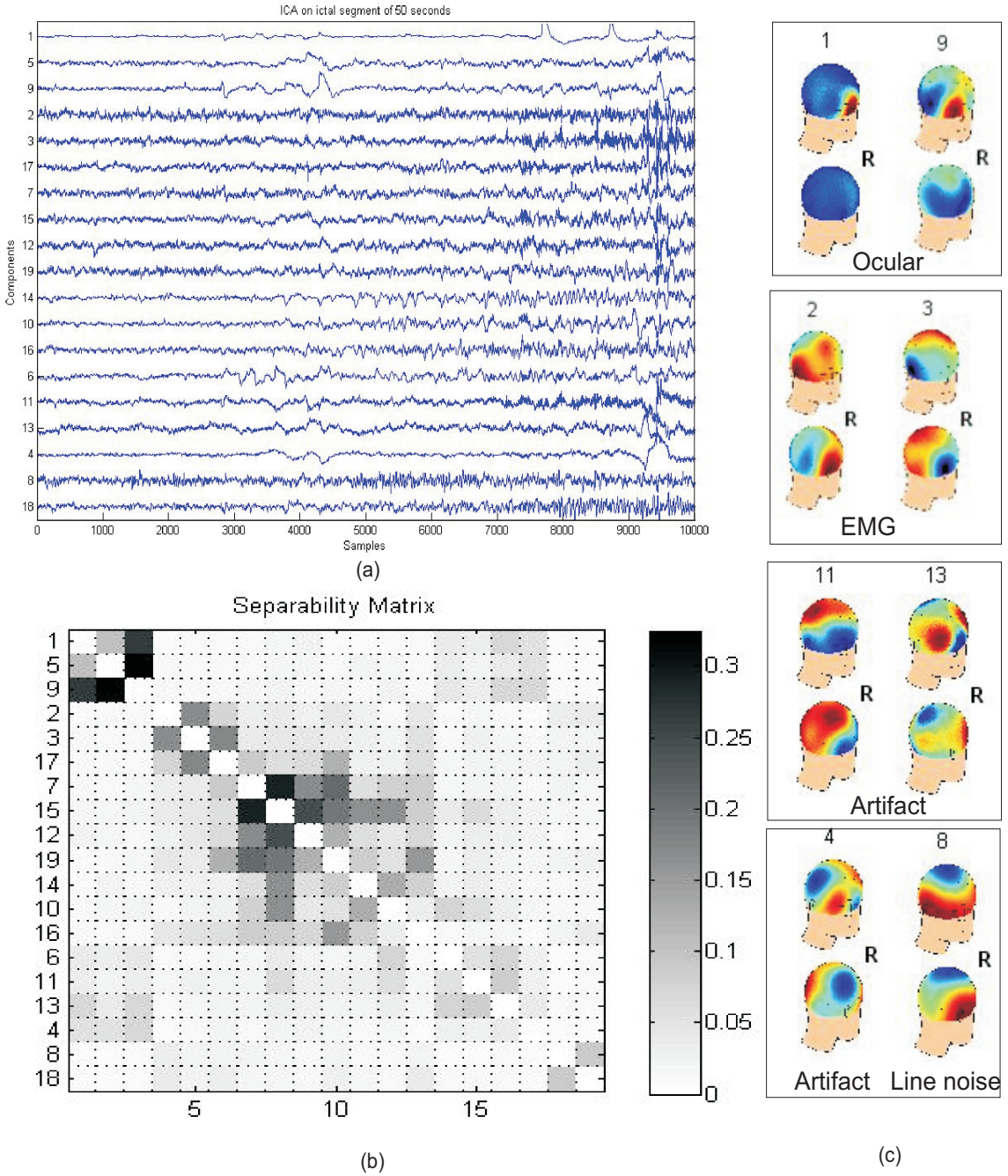


Figure 4.22: (a) The ICs obtained on applying ICA on EEG segment for de-noising EEG. (b) Separability matrix for the estimated sources (c) Topographies of selected artifact sources

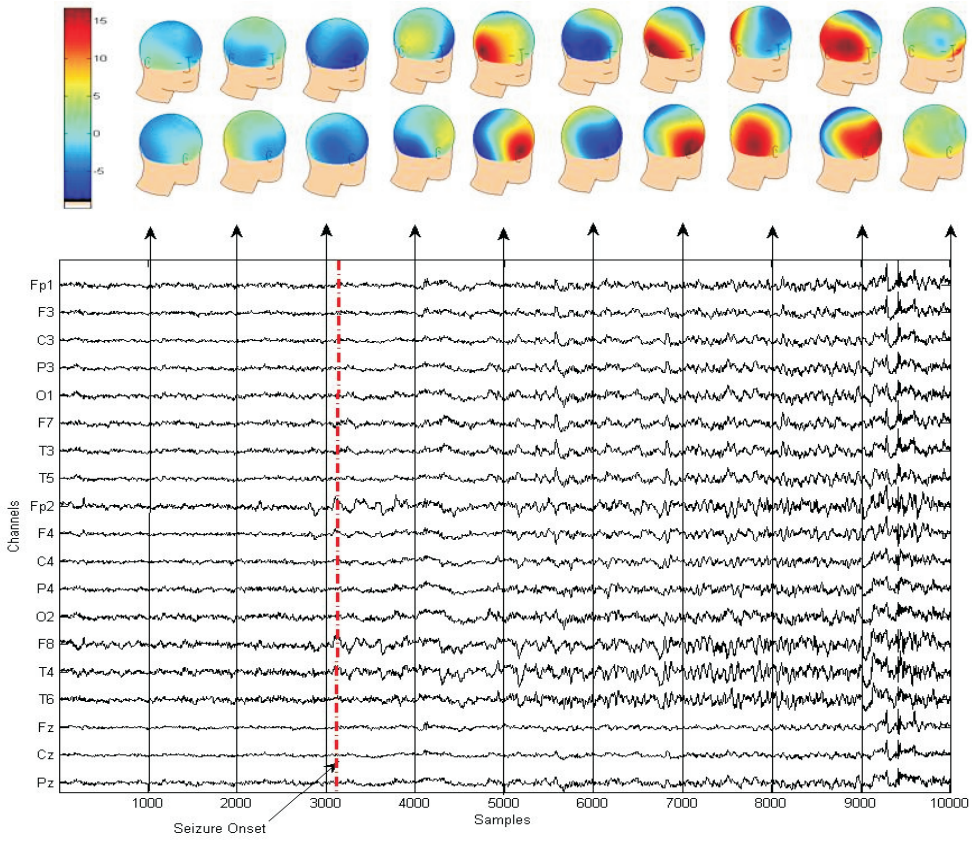


Figure 4.23: The EEG de-noised by ICA artifact removal is shown here. Spatial changes potential distribution is also shown at different time points across the segment.

sources along with temporal information which cannot easily be obtained by other statistical methods.

- Extracting a precursor for seizure prediction is the main aim of this analysis. It is being explored to search for some underlying structural/dynamic change(s) in the continuous multichannel EEG that will indicate the onset of a seizure.

The selection of the ICA algorithm has been identified as an important first step. As described in this chapter, temporal decorrelation aims to find independence across time therefore it gives robust and better estimates of sources even with short data windows, whilst statistical methods such as FastICA based on mutual information or entropy need more data for a better estimation of sources and hence may not be suitable for short windows of data. It is also seen in the examples of ICA applied to EEG data, for spatially or temporally overlapping sources, temporal decorrelation is more suited to the separation of sources as it isolates subtle temporally distinctive sources that may not be identified by conventional ICA which is based on spatial separation.

Thus, the algorithms that will be best for the current analysis is temporal decorrelation for step one: template creation. This is because, for template creation the segments

used are shorter and the time based separation would be more suited as the temporal rhythmic seizure onset would be more distinct than its spatial existence at that time. For example, a temporal lobe seizure source would have a temporal lobe topography, and the usual accompanying muscle artifact (teeth clenching, etc. usually at the time of seizure) generally has a temporal distribution too. Temporal decorrelation would provide a cleaner template. However, for the step two: applying spatially constrained ICA on prior and post EEG segments, the algorithm used will be spatially constrained FastICA. This is because, statistically sufficient long segments can be used at this step. The spatial constraint will already prevent splitting of the signal even if the source signal is weak and the focus will be on spatial decorrelation rather than temporal.

Using ICA with real EEG signals poses certain challenges, such as the need to process windowed data. This prevents the tracking of a source of interest, as with ICA sources may move, or split in unpredictable ways. Constrained ICA, especially spatially constrained ICA is a more suitable choice in this case. Analysis of tracking temporal and spatial changes in the ICs has shown the possibility of existence of measurable dynamics that might be useful for prediction. These and similar observations on the rest of the 12 patients illustrate the idea of obtaining useful information from the underlying signals. It supports the concept of tracking changes in dynamics of an underlying source signal as opposed to the raw EEG. However, there is a need to find an objective method to identify the existence of such patterns. Especially, because this procedure will need to be automated in order to analyze long term continuous EEG (72 hours or more continuous signal), which is necessary for such a predictor analysis. This also motivates exploring additional measures of tracking dynamics of the spatially constrained ICs. Nevertheless, there are certain challenges that will remain with spatially constrained IC analysis, such as the following:

1. **Artifact interference:** Spatial constraint also allows artifacts in an area (spatially) to be picked up as an independent source. For example, a constraint of temporal lobe area of the brain can pick up artifacts like chewing that are independent but also affect the temporal electrodes (sharing spatial distribution with the source).

2. **Generalized seizure data:** If the data has generalized seizures (seizure spreads to more than one focus), they lack a spatial focus or are not contained in one spatial location. This makes the creation of spatial template difficult as it would be difficult to ascertain the seizure source even in the ictal EEG.

In order to automate template creation and to make the feature search an objective process, additional measures will be required along with conventional ICA. The medical and engineering literature on seizure prediction has highlighted the idea of synchronization in many ways, in the context of epilepsy. Many studies independently were able

to show signs of hyper-synchronization happening during a seizure and deterministic changes in synchrony occurring before this hyper-synchronous state. This motivates the exploration of phase synchrony with ICs and will be described in the following chapters.

4.6 Summary

This chapter describes the EEG as a measurement of a set of mixed signals and emphasized the need to pre-process it. A blind source separation technique called Independent Component Analysis has been used for pre-processing. ICA was found useful for unmixing the EEG, extracting sources of interest and de-noising the signals. EEG is a multichannel signal that gives temporal information of the brain to some extent, along with it, it however contains spatial information which is overlooked by many research studies. ICA utilizes this spatial information inherent in the EEG. Also, much previous work is based on single channel or two channel analysis of EEG, which restricts information because EEG is a multichannel signal.

The concept of ICA has been briefly described along with the type of ICA algorithms and their implementations. The general ICA assumptions and perceived limitations are also described. Time structure based decomposition through temporal decorrelation has been introduced and compared with standard ICA implementations. Thereafter, the concept of using ICA with prior knowledge (constrained ICA) has been shown. ICA has been found to be suitable for decomposing the EEG into the underlying independent/least dependent signals as the assumptions for ICA are found to be fairly satisfied for EEG signals. Previous research indicates that ICA can be a useful tool for seizure detection, EEG analysis and de-noising epileptiform EEG, especially to extract information which may not be uncovered by other signal processing approaches in isolation. This further motivates the use of ICA for pre-seizure analysis. The application of ICA to de-noise multichannel EEG and to track a seizure source across time, in epileptiform EEG has been described in the context of seizure prediction. The next step was to measure a certain aspect of the seizure source to assess the changes in the dynamics of the source across time. Especially as the seizure approaches feature(s) may be found that might help in predicting seizure onset. The power of the constrained seizure source was tracked and was found to highlight the increase in power and rhythmicity, 6-10 seconds prior to electrographical changes as seen in the EEG signal segment. These observations support the concept of using underlying signals (ICs) as opposed to the raw EEG for analyzing dynamics of the brain activity.

However, the application of ICA described in this chapter involves manual identification of seizure sources (for spatial template) and subjective evaluation of the changes in seizure sources. To automate/objectify this process and to better measure the dynamics

of a seizure source, *synchronization* has been identified in addition and will be described in more detail in the next chapter.

Chapter 5

Assessment of Synchrony Dynamics

5.1 Introduction

Synchronization is a phenomenon that was discovered by Huygens [159] and is described as 'an adjustment of rhythms of self-sustained oscillating objects due to their weak interaction'. When two oscillating bodies, that are oscillating at their independent periods, are brought together and coupled weakly, they start to adjust their rhythm and begin to oscillate with a common period. This phenomenon is described as *frequency entrainment* and *locking*. These oscillations then tend to get synchronized depending on the following:

1. Frequency Detuning: The difference in the independent autonomous frequencies of the oscillators is called frequency detuning. $\Delta f = f_1 - f_2$. The extent of dependence can be seen by measuring the independent frequencies of the oscillators, and then measuring their frequencies when coupled. The plot of the difference in the independent frequencies or *mismatch* and the coupled frequencies is of the form shown in Figure 5.1. For a certain range of independent frequencies, the coupled oscillators tend to be identical ($\Delta f = 0$) and are said to be synchronized.

   The synchronization region also depends on the coupling strength and will increase by increasing strength of coupling.

2. Rhythm: The 'rhythm' is characterized by the period and frequency of the oscillation.

3. Adjustment of frequencies of the interacting oscillators takes place depending on the frequencies of the drive and the driven oscillator. If the natural frequencies of the drive is ω_1 and of the driven oscillator is ω_2 then if the coupling is in one direction then the frequency of the driven oscillator (ω_2) is pulled towards the frequency of the drive (ω_1) (see Figure 5.2(a)). This is similar to an application of external force on an oscillator to synchronize it. If the interaction is two way

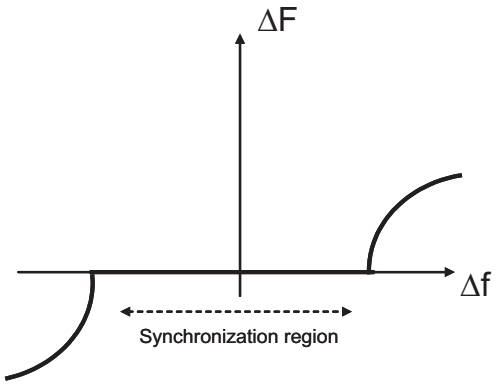


Figure 5.1: Plot of difference of autonomous frequencies (detuning) of the oscillators and the difference of the coupled frequencies of those oscillators. For a certain range of detuning, the coupled systems become synchronized ($\Delta f = 0$)

or bidirectional then the frequencies of both the interacting systems changes and becomes equal to a common stable frequency Ω in between ω_1 and ω_2 (see Figure 5.2 (b)).

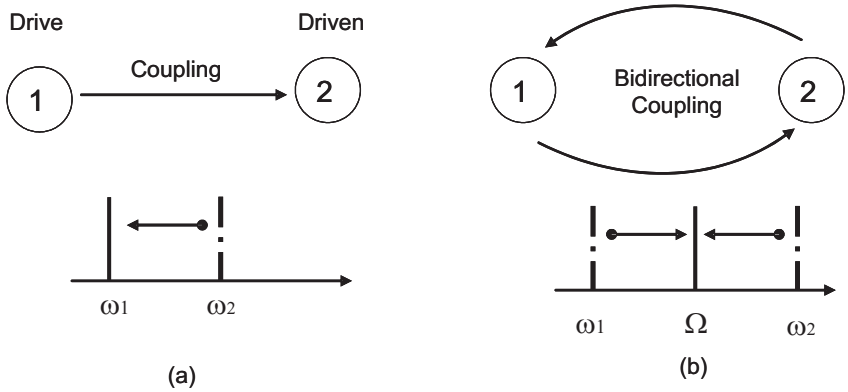


Figure 5.2: (a) Unidirectional coupling: A drive frequency or external force ω_2 tries to entrain the frequency of an oscillator to its own frequency ω_1 (b) Bidirectional coupling: Natural frequencies of both the interacting systems changes to a stable frequency Ω, in between ω_1 and ω_2 [11]

Synchronization was first introduced for the description of two coupled harmonic oscillators. A few examples of naturally occurring synchronization phenomenon are: fireflies with independent rhythms flash light synchronously when grouped together, pendulum clocks hanging on a common beam get synchronized, crickets chirp in unison (synchrony) when together, and some biomedical applications are: normalizing the abnormal respiration of a patient by synchronizing it by a forced frequency of a ventilator [160], restoring the irregular cardiac rhythms by making them synchronous to external pulses [161], synchronization of oscillation of human insulin secretion and glucose infusion [162].

5.2 Measures of Synchronization

Synchronization has been measured in different ways including:

1. **Cross-correlation:** is the simplest and most used measure synchronization. It is used to measure similarity between two signals. The cross-correlation coefficient of two signals $x(t)$ and $y(t)$ is given by

$$\rho(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

where, $\bar{x}$ and $\bar{y}$ are the mean values of the respective signals. Cross-correlation varies from -1 to 1, where -1 indicates anti-synchronization and 1 indicates complete synchronization. A value of or near to 0 indicates linearly independent signals. Some indices used for measuring interdependence are cross-correlation at zero lag ($C(0) = c_{x,y}(0)$). This index is good for linear systems for estimating linear synchrony or interdependence only and is not able to pick non-linear synchrony information nor is it able to give any information about the frequency bands in which the synchronization is taking place. This has been supported by an example using 2 nonlinear Rossler systems as shown next.

The Rossler systems used are described by the following equations

$$x_{1,2} = \omega_{1,2}y_{1,2} - z_{1,2} + \varepsilon(x_{2,1} - x_{1,2}),$$

$$y_{1,2} = \omega_{1,2}x_{1,2} + ay_{1,2},$$

$$z_{1,2} = b + z_{1,2}(x_{1,2} - c),$$

where ω_1, ω_2 are the natural frequencies , $\omega_0 \pm \omega, a, b, c$ are constants and the parameter ε is the coupling coefficient. The values used are as follows

$$\omega_0 = 1, \Delta\omega = 0.035, a = 0.15, b = 0.2, c = 10,$$

x, y and z coordinate signals of 2 Rossler systems have been generated. They have been coupled over a range of coupling strengths from 0.01 to 0.1. From the above equations, it can be seen that the x coordinates of the 2 systems are linearly coupled, while the z coordinates involve nonlinear coupling. These dependencies and the effect of coupling is shown in Figure 5.3 with the help of phase portraits of x_1 vs. x_2 and z_1 vs. z_2.

The cross-correlation coefficient and phase synchrony between the linear and non-

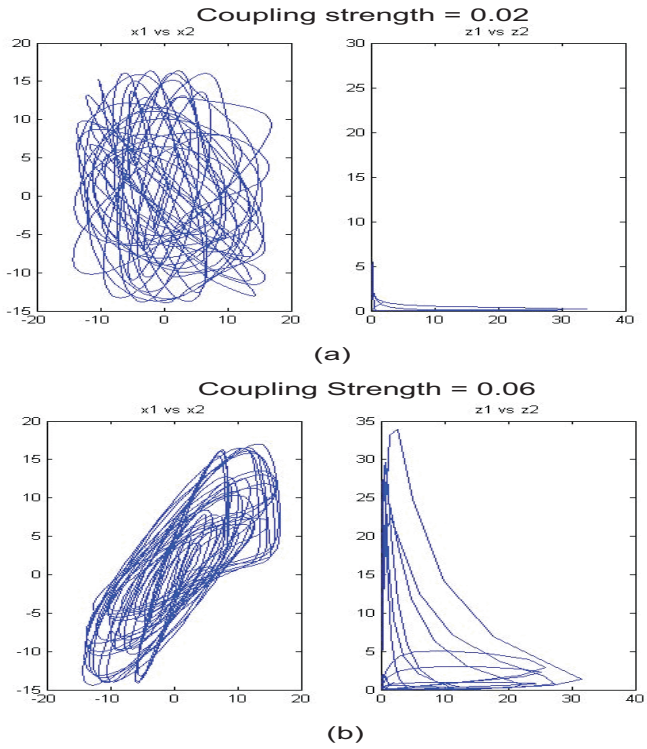


Figure 5.3: Two non-linear Rossler systems are coupled linearly in x coordinates and non-linearly in z coordinates. (a) The phase portraits of x_1 vs. x_2 and z_1 vs. z_2 do not show any structure at 0.02 coupling (b) The phase portraits of x_1 vs. x_2 and z_1 vs. z_2 at coupling strength of 0.06 start to show structure. A linear relationship (points spaced closely and aligned across a line) is observed in x_1 vs. x_2 while a nonlinear relationship (points positioned across a nonlinear curve) is observed in z_1 vs. z_2.

linear series for increasing coupling strengths are shown in Figures 5.4 and 5.5.

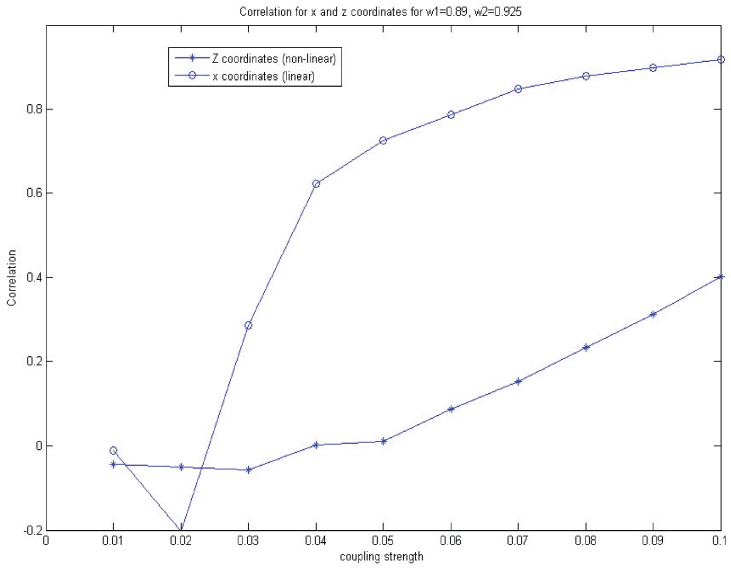


Figure 5.4: Two non-linear Rossler systems are coupled linearly in x coordinates and nonlinearly in z coordinates. The plot of cross-correlation coefficient vs. coupling strength shows the expected increasing cross-correlation trend for x coordinate series as the coupling strength is increased. The cross-correlation between the z coordinate series is not able to show a high correlation even though the coupling is increased by the same amount because the coupling in z coordinate series is nonlinearly related.

2. **Coherence:** helps to find linear interdependence at different frequencies. Thus it provides additional information compared to cross-correlation. It is a function of the power spectral densities $P_{x,x}, P_{y,y}$ and the cross power spectral density $P_{x,y}$ of signals x and y:

$$C_{x,y}(f) = \frac{|P_{x,y}(f)|^2}{P_{x,x}(f)P_{y,y}(f)}$$

But like correlation it is only a good estimator for linear synchronization. Also in coherence, the measure is, in a way, dependent on the amplitude synchronization or dependence and if there is underlying synchronization with chaotic amplitudes, it will not be able to estimate that.

3. **Phase synchronization:** is a method of estimating synchrony of signals using the frequency and phase information. Two signals are said to be phase synchronized when their phase difference $(\phi_1 - \phi_2)$ becomes constant over time $(m\phi_1 - n\phi_2 = constant)$. The advantage of this measure is that it is applicable for chaotic signals as well and can estimate synchronization even when the amplitudes remain uncorrelated [11]. (An example has already been shown in section 4.4). Thus this measure can help to estimate synchrony of signals even when they are non-linearly dependent. This method has been used in this work and is explained in more detail in the following sections.

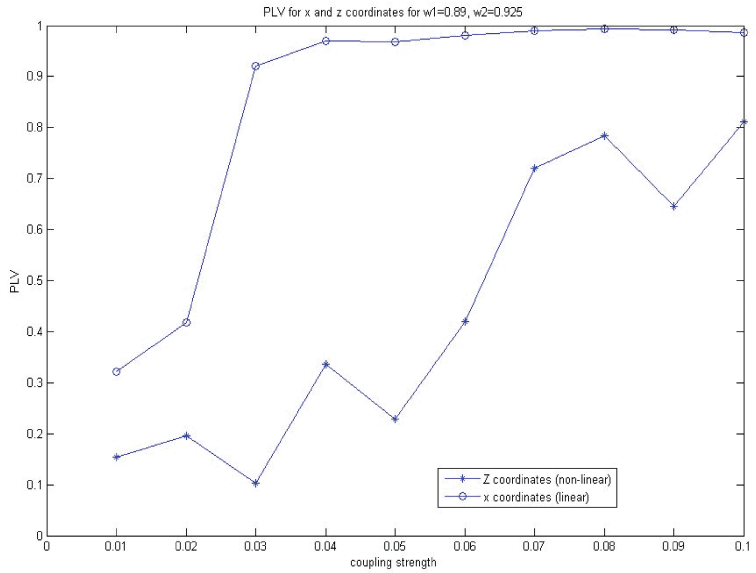


Figure 5.5: The phase synchrony in x coordinate series shows a smoothly increasing synchrony with increasing coupling strength and it shows the same trend in z coordinates with the synchrony levels corresponding to the strength of coupling, albeit not to the same extent as the x coordinates.

5.3 Synchronization in Chaotic systems

Recently the concept of synchronization for chaotic oscillators was introduced [84, 163, 11]. The challenge in measuring synchronization for chaotic systems is in characterizing the rhythm which unlike a periodic oscillator may not be trivial to assess. Synchronization in such cases may be seen as an onset of a structure or constancy in phase as the mean frequencies start to coincide. It should be noted however for chaotic signals coinciding of mean frequencies does not imply coinciding of amplitudes of the signals as well [11]. Thus for weak coupling, the signal amplitudes may remain irregular and uncorrelated while the mean frequencies adjust, as shown in Figure 5.6 and the signals are said to be *phase synchronized*. In the case of strong coupling, along with the mean frequencies, the amplitudes may also start to coincide and this leads to complete synchronization.

5.3.1 The Effect of Noise

Noise affects synchronization adversely and the presence of strong noise can even destroy synchronization. The outcome depends on the strength of noise. If the noise is weak then it is not capable of displacing the phase above the critical threshold and hence the phase only fluctuates around a stable equilibrium point as shown in Figure 5.7(a).

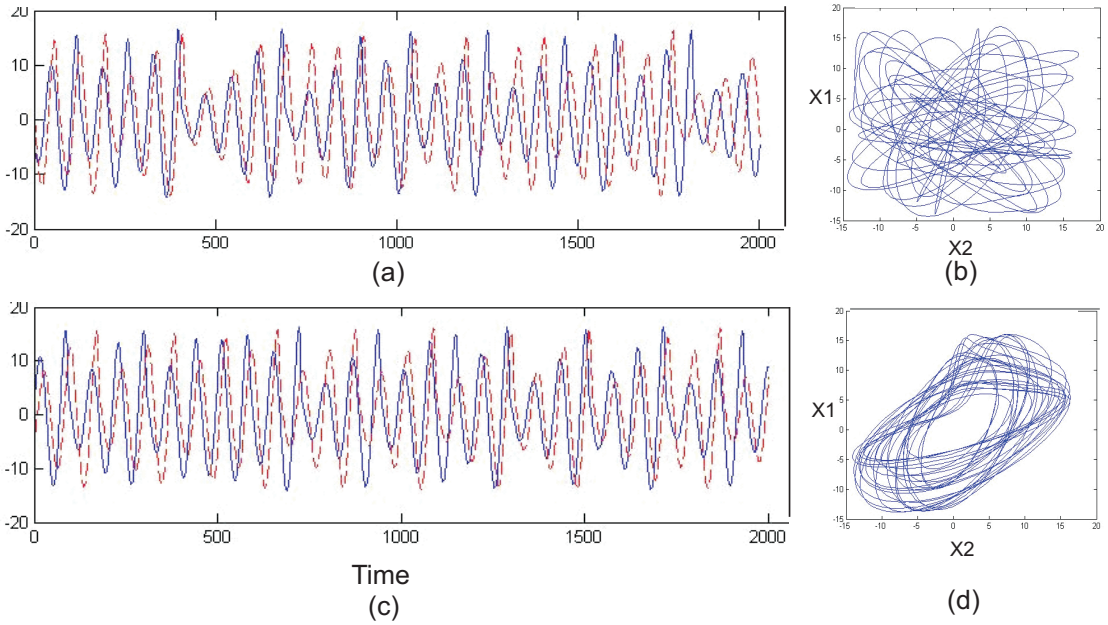


Figure 5.6: (a) Signals from two chaotic (nonlinear Rossler systems) systems having different mean frequencies, (b) Plot of x1 vs. x2, no structure is visible (c) Coupling tends to coincide the mean frequencies even though the amplitude remain irregular (d) Plot of coupled x1 vs. x2 shows the onset of some structure.

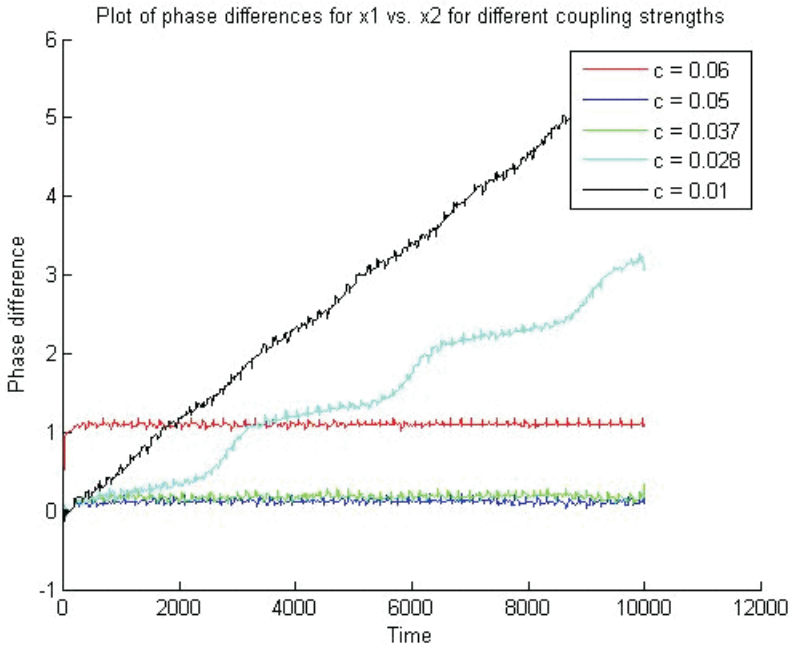


Figure 5.7: (a) Phase difference fluctuates around a constant level in the presence of weak bounded noise and varies with coupling strength like a noise free oscillator (synchronization sets in with increasing coupling strength).

As biomedical signals such as EEG may not be periodic, this study on generalization of synchronization for chaotic signals with noise, has spurred interest in analyzing synchrony in EEG signals. The next section will concentrate on measuring phase synchronization, with examples based on EEG signals.

5.4 Measuring Phase Synchronization

To measure phase synchronization the first step would be to find the phase or frequency of the signal. In case of a periodic signal this may be quite trivial but in case of chaotic, non-periodic signals this may not be straightforward as the frequency may vary continuously. A few methods to measure the frequency or period of oscillation for a chaotic signal (signal with varying amplitude and frequency) are enumerated below:

- **Mean frequency:** If the signal appears to have a pattern of similar cycles, then the frequency can be measured as *mean frequency*, by counting the number of cycles in a fixed large time interval. $< f >= N_\tau/\tau$, where N_τ is number of cycles and τ is the time interval in which the N_τ cycles occur.
- **Instantaneous frequency/phase:** For chaotic signals, another approach to find the frequency or phase is to compute the instantaneous frequency/phase. This can be done by at least two methods: the Hilbert Transform and the Wavelet Transform. For both the cases, the original signal is transformed to a complex valued signal using an auxiliary function, and the instantaneous phase or frequency can then be found from his complex signal. The details of extracting the instantaneous phases using the two methods are as follows:
  1. **Hilbert Transform:**[11, 88, 164] can be used to estimate the instantaneous frequency of a signal, in the time domain or frequency domain, as explained below:
     - **Computing the Hilbert Transform in the time domain:** In the time domain the Hilbert Transform of a signal $x(t)$ helps to form an analytical signal:
       $$A(t) = x(t) + i\widetilde{x}(t),$$
       where the imaginary part $\widetilde{x}(t)$ is the Hilbert transform of $x(t)$
       $$\widetilde{x}(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau.$$
       (P being the Cauchy principle value).
     - **Computing the Hilbert Transform in the frequency domain:** To compute the Hilbert Transform in the frequency domain, the FFT of the

signal is computed, and then the phase of every frequency component is shifted by $-\pi/2$ (equivalent to swapping imaginary and real parts of the FFT) followed by the inverse FFT. To reduce boundary effects 10% of the values at each end can be discarded.

In essence the Hilbert Transform removes the negative frequencies, and doubles the positive frequencies which is also called the analytical form of the signal. The instantaneous phase can then be found from the *unfolded angle* from the analytic signal, by

$$\phi(t) = \arctan \frac{\widetilde{x}(t)}{x(t)i}.$$

that gives the cyclic phase in the interval $[-\pi, \pi]$. The relative phase or the phase difference can then be found as

$$\phi_{1,2} = \phi_1(t) - \phi_2(t).$$

(Unfolding the angle is required to make the signal continuous by removing the 2π discontinuities)

2. **Wavelet Transform:** [39] With the Wavelet Transform the auxiliary function used is a complex wavelet such as Morlet wavelet or Gabor wavelet:

$$\psi(t) = (e^{i\omega_0 t - e^{-\omega_0 \sigma_t^2}})e^{-t^2/2\sigma_t^2},$$

where ω_0 is the center frequency; $\omega_0\sigma_t^2$ is the number of cycles of the corresponding frequency needs to be included; σ_t is the bandwidth parameter that determines the width of the spectral peak (greater σ_t gives a wavelet with a better frequency resolution but poorer time localization and vice versa).

Convolving this complex wavelet with the signal helps to estimate the instantaneous phase $\phi_w(t)$

$$W(t) = \int \psi(t')x(t-t')dt' = A^W(t)e^{i\phi_W(t)}.$$

A comparison of Hilbert Transform and Wavelet Transform for estimating instantaneous phase/frequency: The two methods of estimating instantaneous frequency have been compared by Quiroga *et al.* [82], and they have essentially found that both the methods are equally efficient in performance. The main difference is in the pre-processing required. The Hilbert Transform being a filter with unit gain at each frequency, it takes into account the whole range of frequencies to estimate the instantaneous phase. This method was initially described for narrow frequency range signals and extending it to broad frequency range data can give spurious results [100]. Thus using the Hilbert Transform

requires pre-filtering of data in the frequency band of interest. With the Wavelet Transform, $W(t)$ is non-zero for frequencies close to central frequency ω_0, which is like band-pass filtering of the signal prior to analysis.

After finding the instantaneous phase or frequency, the next step is to quantify the change in phase of two signals over time, and to estimate the onset of synchrony or phase locking from the distribution of this change in phase difference across time. The phase realizations of the signals across time can be characterized by a probability distribution, and in essence the strength of phase synchronization corresponds to the peakedness of this distribution. This measure or quantification has been done by various phase synchronization indices, a few of them are described in the following sections as applied to EEG signals.

5.4.1 Phase synchronization indices

Phase synchronization is thus a measure of the probability of the phase differences, aiming to locate a preferred value of phase difference that occurs with the onset of a dynamic dependence of the oscillator signals. It is important to note that, the phase is defined on a circular scale and its values whose difference is an integral multiple of 2π are regarded the same. Standard linear statistical measures of mean and variance are not applicable to such random circular variables as they can yield different values if a period is added or subtracted from some values, even though the physical meaning of these changed values remains the same [165]. The onset of phase synchrony is expected to be marked by an onset of constant phase difference across time

$$n\phi_a(t) - m\phi_b(t) = constant \tag{5.1}$$

The variables n and m denote the phase relationship of the oscillators and hence the $n : m$ ratio is called the *phase locking ratio*. They may both be equal to one for a system or may be of a higher order [85]. Another example is the study of Parkinsonian tremor and study on the synchronization between the brain (MEG) and muscle activity (EMG), the EMG frequency was about 6Hz and MEG signal frequency was about 12Hz and the synchronization was of the order of 1:2 [11] (Chapter 4). In the present study, as we aim to measure the synchronization between signals from within the same physiological system (i.e. the brain), we assume the phase locking ratio of $n : m = 1 : 1$ [72, 91, 82, 87].

Different indices have been used for measuring this phase synchronization, such as:

1. **Phase Locking Value (PLV):** [82, 72, 39, 88, 164, 90] The term PLV was coined by Lachaux *et al.*, [39] and is also referred to as the *Mean Phase Coherence* [72].

PLV makes use of relative phases and can be best explained with the help of directional statistics which are useful for circular variables. Directional Statistics use trigonometric functions θ: $\alpha_p = \langle \cos p\theta \rangle$ and $\beta_p = \langle \sin p\theta \rangle$, where p is the order of the moment and $\langle \cdot \rangle$ denote the average over time. Their combined complex form is given by : $\chi = \alpha_p + \iota\beta_p$. On a sample of θ_j $(j = 1, 2, ...n)$ the corresponding estimator quantities are given by

$$a_p = \frac{1}{n} \sum \cos p\theta_j,$$

and

$$b_p = \frac{1}{n} \sum \sin p\theta_j.$$

The components of the first empirical moment are given by

$$\bar{C} = \frac{1}{n} \sum \cos \theta_j,$$

and

$$\bar{S} = \frac{1}{n} \sum \sin \theta_j,$$

and in polar representation as

$$\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2},$$

and

$$\bar{\theta} = \arctan \frac{\bar{S}}{\bar{C}}.$$

Since $\bar{R}$ is the length of the mean of the unit vectors corresponding to the sample values, it is called the mean resultant length and here as the unit vectors are the phase differences, it is used as a measure of the strength of synchronization.

This is equivalent to the Euler form of

$$\gamma_{1,2} = \sqrt{|\langle e^{\iota\phi_{1,2}(t)} \rangle|} = \sqrt{\langle \cos \phi_{1,2}(t) \rangle^2 + \langle \sin \phi_{1,2}(t) \rangle^2}.$$

In essence $\gamma_{1,2}$ measures the relative phase distribution over a unit circle. If the two signals are phase synchronized, the relative phase will occupy a small portion of the circle and mean phase coherence will be high, while a lack of synchrony will have relative phases that spread over the entire unit circle giving a lower mean phase coherence. PLV varies from 0 to 1, 0 being no synchronization and 1 being perfect synchronization.

Effect of Wrapping/Unwrapping Phase: An important point to note here is the need and use of wrapping/unwrapping of the cyclic phase. The method using Hilbert Transform for estimating the instantaneous frequency gives a wrapped

phase between $[-\pi, \pi]$. For estimating the instantaneous phase, it is required to make the signal continuous, removing the discontinuities that occur in a wrapped phase after every 2π, thus the phase is found from the *unfolded angle* instead. But as has been mentioned above, the presence of noise can lead to arbitrary phase slips of 2π which may effect the measure of synchronization as even though the angle may be the same before and after a slip (in the physical sense), it may add to or subtract from the index being used. Thus to assess the phase wrapped in the interval $[0, 2\pi)$ the Euler form of the PLV index is useful as it preserves the directional statistics.

2. **Information Theory:** Another index is based on information theory and uses conditional probability [166]. The aim is to find any common information shared by the two signals that shows their synchrony. In this method, the phases are wrapped individually for both signals into the intervals of $[0, 2\pi)$. These are then divided into L bins, and the index measures the probability of one of the phases to belong to a certain bin provided the other has a certain value.

3. **Shannon Entropy:** [82, 111, 85] This index quantifies the distribution of the phase difference by partitioning the $[0, 2\pi)$ interval into L bins and comparing it with the distribution of the cyclic relative phases for the two signals independently [85]. The Shannon Entropy is calculated as:

$$I_X = -\sum_{i=1}^{M} p_i log p_i$$

where p_i is the probability density function of the original phases (I_ϕ) and for independent phases (I_{max}). In the case of independent phases as there is no preferred value of $\phi_{1,2}(t)$ expected, so the distribution of phases should be uniform and giving a maximal value of I_{max}. While in the case of synchronized phases the phase distribution would show a sharp peak and have a low entropy. Due to finite size of windowed data the phase distribution may not be completely uniform for uncorrelated signals [111], hence the value of I_{max} is calculated by creating independent phases (by random shuffling) and then calculating a normalized index as

$$\rho_{1,2} = \frac{I_{max} - I_\phi}{I_{max}}.$$

These indices have been compared by some studies on simulated data and on EEG [111, 85, 167]. They have the same range of variation (0 to 1). For 1:1 synchronization the Shannon Entropy index strongly depends on the number of bins used for creating the distribution of phase differences and it can indicate low synchrony even in case of perfect synchronization [167] and the information theory index was also found to show non-zero coupling where there was none present [166]. PLV and Shannon Entropy

indexes were found to be most sensitive to transitions from weak to strong coupling.

5.5 Parameters in Estimating Phase Synchronization

The analysis of phase synchronization for chaotic and noisy signals as given by Rosenblum *et al.,*[84], Pikovsky *et al.,* [11] and Tass *et al.*, [85] are recognized as proof of principle for using the phase synchronization analysis for biomedical signals such as EEG. But there are certain parameters which need attention before the phase synchronization can be applied to biomedical data, these are discussed next.

5.5.1 Effect of Signal Bandwidth

As shown in previous sections, phase synchronization analysis involves the estimation of instantaneous phase using Hilbert Transform or Wavelet Transform. The theory of estimating phase synchronization was based on narrowband signals (with a narrow frequency spectrum) and it is important to acknowledge that biomedical signals are not necessarily narrowband, especially EEG signals (see Figure 5.5.1). Various studies have extended this theory of Phase Synchronization to broadband data [92, 77, 72, 89, 82]. However, applying the Hilbert Transform to broadband signals gives meaningless estimations of instantaneous phase (see Figure 5.8) which may lead to incorrect analysis.

Some studies [100, 92, 168, 169] have shown the effect of broadband data on phase synchronization analysis and it is recommended to filter the data into *appropriate* overlapping bands before estimating instantaneous phase/frequency from it. It is also essential to use an appropriate filter, bandwidth of the filter and the number of samples to prevent detection of spurious synchronization.

5.5.2 Type of Filter

The type of filter that would be suitable for phase synchronization analysis would need to be such that is does not affect the phase of the original signal. The effect of passband and stopband ripples has also been assessed in a study [12], and it was found that a filter with bigger ripples produces variation in the instantaneous phase of the filtered signals. It was seen that bigger ripples led to lower values of synchrony for low synchrony pairs (as expected) and the high synchrony pair did not exhibit any dependence on the ripples. The choice of an appropriate filter in such a context has been to be a finite impulse response (FIR) filter.

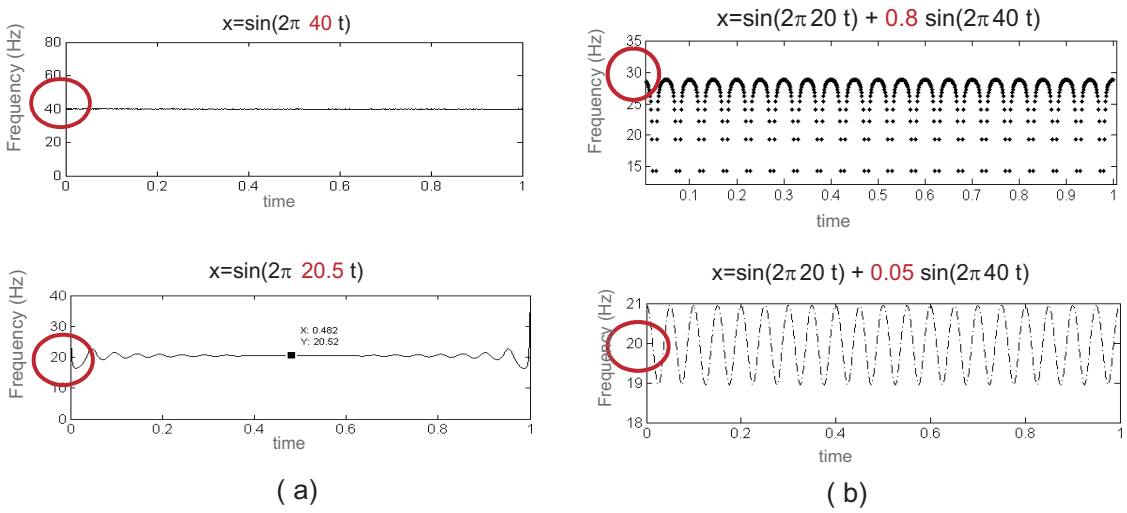


Figure 5.8: Estimating instantaneous frequency using the Hilbert Transform on narrowband and broadband data. (a) Two examples of signals with narrow frequency spectrum are shown. The frequency estimated with HT is found to be as expected (40 Hz and 20.5 Hz) in both the cases. (b) Two examples of signals with broad frequency spectrum are shown. The frequency estimated with HT is not as expected and is much fluctuating and difficult to interpret. The upper plot shows that equally strong frequency components shows an instantaneous frequency that appears to be an average of the two components (original being 20 Hz and 40 Hz, is estimated as 30 Hz). The lower plot shows that with one strong component the frequency of the component with higher amplitude may appear to dominate.

5.5.3 Effect of Filter Bandwidth

It has been shown by a few studies that the bandwidth of the filter has a major affect on the phase synchronization estimates [12, 168, 169, 170]. Too narrow a filter bandwidth may lead to spuriously high synchronization leading to an over estimation of synchrony where the coupling strength is not known *a priori*. Xu *et al.* [168], have shown that a narrow bandwidth of $2\pi\delta f < 1$ (where δf is the normalized width of the bandpass filter) can lead to a rapid artificial increase in apparent synchrony while the coupling remains unchanged. A bandwidth of $2\pi\delta f > 1$ leaves the synchrony estimation unchanged as compared to synchrony estimated without filtering (for artificially coupled Rossler systems). Chavez *et al.*, [170] have described this effect of bandwidth applied to biomedical signals. They have concluded that bandpass filtering is an essential step for phase synchronization analysis of biomedical signals and it depends on the trade off between the accuracy of the analytic signal model and the temporal resolution. If the signals have a dominant clear spectral peak then an accurate and meaningful model may be obtained from those filtered oscillations with the main rhythmic component, even when they are filtered with a relatively large bandwidth. Whilst if the signal does not show a clear oscillatory behavior then a meaningful instantaneous phase may be found only after narrow band filtering. Tcheslavski [12] has also shown that the narrower the filter bandwidth, the higher is the likelihood of error in low synchrony

signal pairs.

5.5.4 Effect of number of samples used

Similar to the effect of filter bandwidth, the sample window length is also seen to affect the synchronization analysis [12, 169]. Tcheslavski [12] has shown a dependence of the expected phase synchrony on the sample length as

$$E\{\gamma\} \approx \frac{1}{\sqrt{N}},$$

where $E\{\gamma\}$ is the expectation operator, γ is the phase synchronization measure and N is the analysis time window. It was found that the shorter the time window, the higher was the likelihood of phase synchrony error for any low synchrony pair. The graph of Figure 5.9 shows the effect of filter bandwidth and sample length for a low synchrony pair in terms of fractional frequency [12].

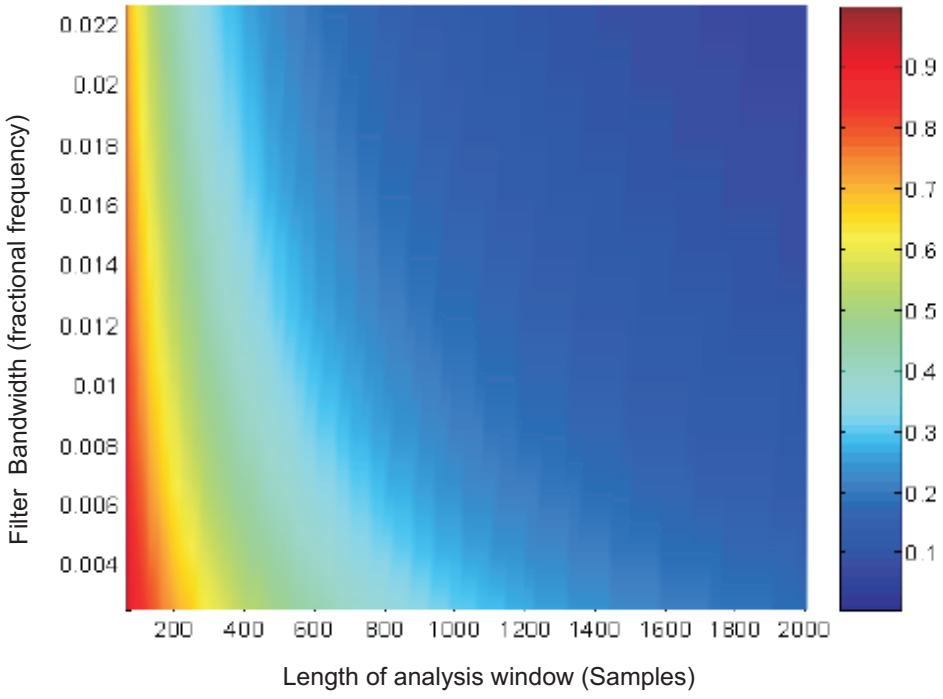


Figure 5.9: The plot shows the dependence of phase synchrony of a pair of signals with low synchrony, on the filter band width and sample size of data. It can be seen that the low synchrony pair can show spuriously high synchrony if short window lengths of 200-600 are used, especially with a narrow filter bandwidth as well [12].

5.5.5 Effect of Signal Mixing

An important issue that has been raised recently is the effect of mixing of signals on phase synchronization analysis [98]. This is an important issue to be considered for biomedical signals such as EEG as they are usually mixtures of underlying signals. Meinecke *et al.* [98] have shown that using linearly mixed signals instead of independent signals for phase synchronization analysis can show spurious synchrony where there is none. They describe that using phase synchronization on linear superposition of signals generally lead to incorrect results. A basic example is shown below: Let two signals be given by $S_1(t) = cos(\omega_1(t))$ and $S_2(t) = cos(\omega_2(t))$ where $\omega_1 \neq \omega_2$. These signals are independent and hence their phase synchronization should be 0. If we apply a linear transformation to them such that $x_i(t) = \sum_{j=1,2} A_{i,j} S_j(t)$ then these show non-zero synchronization ($PLV > 0$) for a particular mixing matrix where $A_{11} = A_{22}$, $A_{12} = 0$ and $A_{21} = \alpha$ and $\alpha > 1$. Thus linear mixing can show synchronization even when there is no synchronization at all. Ideally, the mixed EEG signals should be resolved to the source signals first before applying phase synchronization analysis on them.

5.6 Significance Testing: Phase Locking Statistics

5.6.1 Introduction

As the phase locking results for EEG signals have no gold standard results to compare to, it becomes necessary to evaluate their significance. This step helps to check that the values presented by the synchronization indices depict true synchronization and are not reflecting synchrony as a result of some feature of the time series such as the non-stationarity [171]. To check the validity of the index results, the method of surrogate data has been found useful [166, 172, 39, 87, 72, 85]. This tests the hypothesis that the signals are not synchronized. The negation of this null hypothesis shows that the PLV obtained for the signals is significant. A significant PLV found in such a way is called the *Phase Locking Statistic* (PLS). A comparative study was also carried out by Allefeld, *et al.* [165] comparing different parametric (based on directional statistics) and non-parametric (based on Bootstrap theory) methods of testing for phase synchronization with regards to generality, complexity, precision and applicability. It was concluded that the non-parametric tests based on surrogate data and bootstrap theory were the best suited for phase synchronization, though they may be computationally intensive.

5.6.2 Hypothesis Testing, Surrogate Data and Bootstrapping

The underlying principle of this statistical testing is hypothesis testing. As, due to lack of a gold standard in biomedical signal analysis, it is not possible to prove or disprove a

null hypothesis, the attempt is made to reject a null hypothesis and a default conclusion is made by showing lack of contrary evidence. This is done by calculating a statistic from the original data and checking if this value is within the range of values that would be expected if the null hypothesis was true. This range of values is found by using the *Bootstrap* and *Surrogate Data* methods. Surrogate data is a set of data series that is created such that it shares the statistical characteristics (such as the mean, variance, power spectrum) of the original data and only differs in the property whose effect on the statistic needs to be validated. Then the bivariate statistic is calculated repeatedly from the pairs of the surrogate series, which gives a distribution of the estimated indices that is referred to as a *Bootstrap Distribution*. The test is to check the probability of the original statistic (estimated from the original data series) not belonging to the surrogate distribution. Thus the surrogate test can help to provide a significance threshold, beyond which the statistic can be regarded as significant with a certain level of statistical confidence. The surrogate series have been constructed with phase randomization. Figure 5.10 summarises the concept of surrogate data.

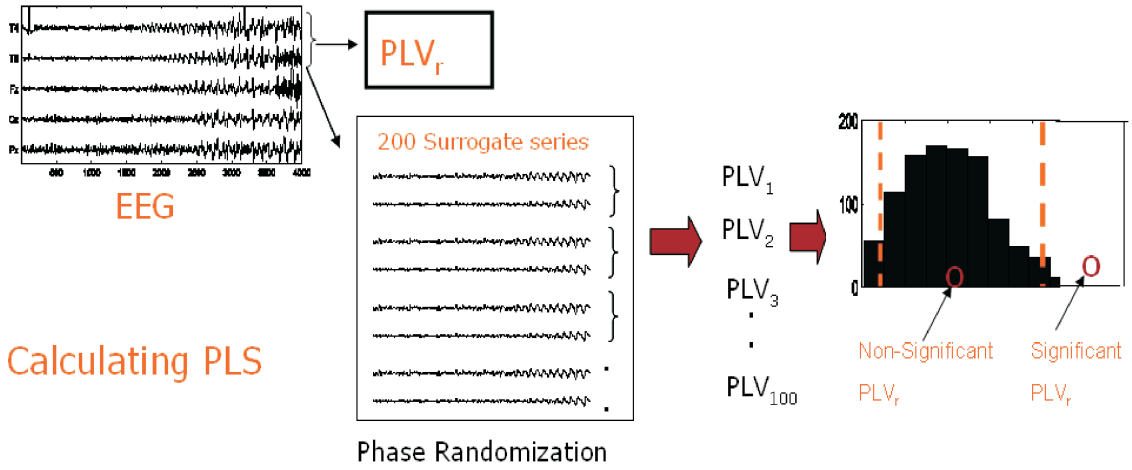


Figure 5.10: Statistical analysis of phase locking index. Firstly, the original PLVr is calculated for the pair of EEG signals. Then a set of 100 phase randomized surrogate signals each are created for the pair of EEG signals. The PLVs of these pairs of surrogate signals are then calculated to obtain a bootstrap distribution of the surrogate PLVs. Using the percentile method, the PLVr is compared with the distribution. A PLVr that lies outside the 95 percentile of the distribution is considered statistically significant and the EEG pair is significantly synchronized.

Phase Randomized Surrogate Data: Phase randomization is a technique to create surrogate series by randomizing the phases of the series in order to destroy any phase relationships inherent in the signals. The requirement is to maintain the autocorrelation of the signals and as it is linked to the power spectrum by Wiener's theorem these surrogate signals are created by preserving the power spectrum in the frequency domain. To implement this, the FFT of the original signal is computed and then the phase is randomized by multiplying the complex spectrum at each frequency by $e^{j\phi}$, where

$\phi \in [0, 2\pi)$ is a uniformly distributed random variable and is independently chosen at each frequency. The symmetry in phases is maintained to enable the computation of the inverse FFT. The inverse FFT then gives the surrogate series. Thus by keeping the magnitude of the Fourier coefficients fixed the spectrum remains unchanged, while the phases get randomized.

These surrogate series help to realize the null hypothesis of two processes which oscillate asynchronously with the same frequencies (power spectra) as the original series. With the bootstrap distribution generated, the next step is to check the significance of the original index with a high level of confidence. Ideally this can be done by using the 95% cut off of the Bootstrap distribution that is found by finding the mean of the distribution and the standard deviation of the distribution and then the 95% cut off is given by $(mean + 2 \times std)$. It should be noted that this is valid for normal distributions only.

Here, it was observed that the distribution of PLV's from surrogate series of EEG signals did not form normal distributions but were skewed. In such a case if the threshold of significance is found as mentioned above using the mean and standard deviation then it may show an under-estimated upper cut off level and a negative lower cut off level. The mean and median of a skewed distribution may be estimated quite accurately using a bootstrap distribution but the variance may be grossly under-estimated [173]. Therefore when the distribution is anything other than normal it is recommended to use the percentile method [173] for finding the 90% confidence intervals. With this method, the end points of the bootstrap percentile intervals are defined as:

1. First sort the statistic (PLVs of the surrogate series) in ascending order.
2. Find the 95% and 5% percentiles from this bootstrap distribution.
3. These give the upper and lower limits of the 90% confidence interval.

It was also observed that with higher resampling (increasing the number of surrogates to 1000), the bootstrap distribution tends towards a normal distribution. However, a trade off is required between the computational time and tolerated accuracy, as generating 1000 surrogate series for each time window for multivariate data is quite computationally expensive.

5.7 Synchronization applied to Epileptic EEG

In recent years there has been a momentum in studies based on analyzing the phase synchronization in epileptic EEG signals for seizure prediction. A number of studies analyzing brain synchrony have been performed on invasive and non-invasive EEG

signals [8, 86, 39, 87, 72, 77, 76, 88, 76, 89, 90, 91]. The phase synchrony analysis of EEG signals has been generally approached in the following ways:

- **Bivariate or Multivariate:** Bivariate analysis uses two signals (two EEG channels) showing focal synchronization [99, 39], while multivariate analysis aims to use more than two signals (more EEG channels) to obtain a global synchronization measure as shown by Tcheslavski [12]. When using bivariate synchronization, the results of the analysis can be *subjective to the EEG channels* used and selecting the appropriate channels becomes a crucial decision of the analysis.
- **Local or Distant:** Synchronization is also described as local or distant [39, 90]. Local synchronization analysis is generally based on electrodes that are nearby (approximately 2 cm) while distant synchronization analysis is based on synchrony of contralateral electrode signals, spatially farther away from each other.

5.8 Spurious Synchrony in EEG

An important discussion that is now being considered in synchrony research studies, is the spurious synchrony that can appear in linearly mixed signals. EEG signals being mixed signals (due to overlapping as an effect of volume conduction), have also been shown to display synchrony where there is none [98, 99]. This can result in misleading interpretations. One of the causes of spurious synchrony in linearly mixed signals could be attributed to the presence of similar signals in the two mixtures being analyzed. Thus in analyzing the two mixed signals, one signal shows synchrony with itself (see Figure 5.11). Therefore, it is important that the EEG signals should be unmixed before further phase synchronization analysis is carried out. ICA would be very helpful for unmixing the EEG and is used as an important pre-processing step in the present research.

The synchrony of spatially local and distant scalp EEG channels compared to their unmixed signals gives a clearer view of the spurious synchrony obtained in EEG (mixed) signals (see Figures 5.12, 5.13 and 5.14). They show the phase synchrony of two EEG signals across a two minute segment and that of their respective least dependent components (LDCs) unmixed by ICA. It should be noted, that ICA is applied here on bivariate signals, which may not be an efficient unmixing, resulting in LDCs as opposed to independent components (ICs). The scalp EEG is from an epileptic patient and has a seizure at 11.6 seconds into the segment (marked by electroencephalographers). The EEG channel signals have been pre-processed by removing the mean and filtering high frequency noise. These signals are used for calculating the phase synchrony using a non overlapping window of 600 samples. PLV is used as the index to obtain the measure of phase synchrony across time for the EEG segment. The unfiltered EEG signals are

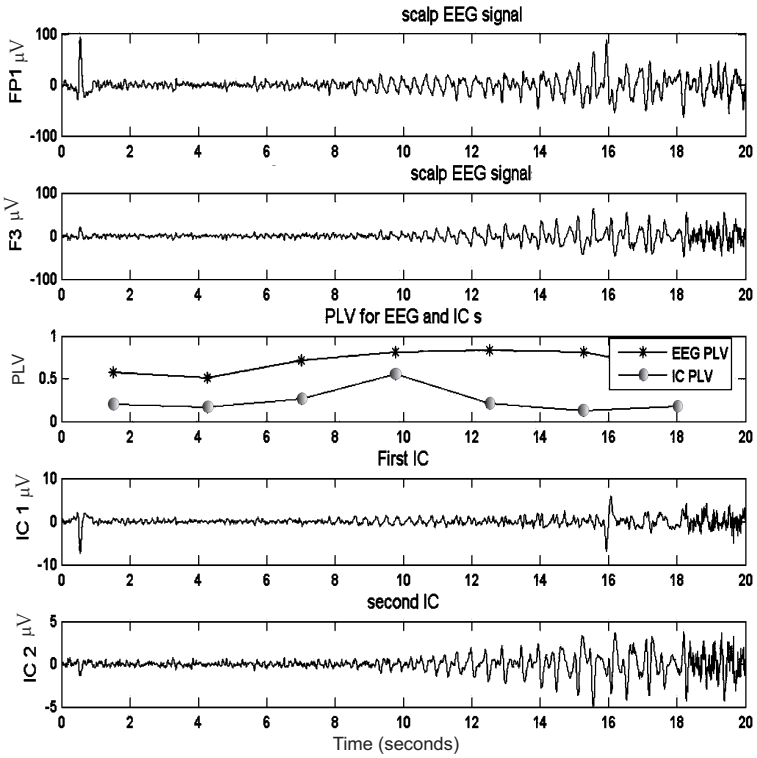


Figure 5.11: The effect of signal mixing on phase synchrony analysis is shown here. A set of EEG signals are observed to show high phase synchrony (marked with asterisk). The set of EEG signals are then unmixed using ICA. The phase synchrony of ICs can be observed to be much less than phase synchrony of the original EEG signals (marked with circles). The spurious synchrony of EEG signals can be attributed to a strong underlying signal being linearly mixed in both the EEG signals.

temporally decorrelated using TDSEP to obtain two LDCs of all EEG signal pairs. These LDCs are then filtered using a bandpass filter of bandwidth 2-8 Hz. The PLV of the filtered ICs is similarly calculated using a non-overlapping window of 600 samples. The PLV of the EEG signals (bold red line) and the PLV of their corresponding LDCs (thin blue line) have been shown here. The bivariate combinations used for analysis involve a randomly picked channel (enclosed by a bold small circle), e.g. Fp1 in Figure 5.12 with the rest of the EEG channels. The spatially neighboring or local electrodes are marked by a dotted circle around Fp1 and the rest of the channels are considered distant to it. The synchrony results of the local electrodes with Fp1 are also enclosed in a dotted circle. The phase synchrony for the **distant** channel combinations of EEG signals and their LDCs is observed to be similar. In case of all the **local** channel combinations EEG synchrony is seen to be much higher than their LDC synchrony. The cause of this higher synchrony seen in the local EEG signals is most likely a consequence of spurious synchrony expected in mixed signals. Local EEG channels are the most overlapped and mixed due to the effect of volume conduction.

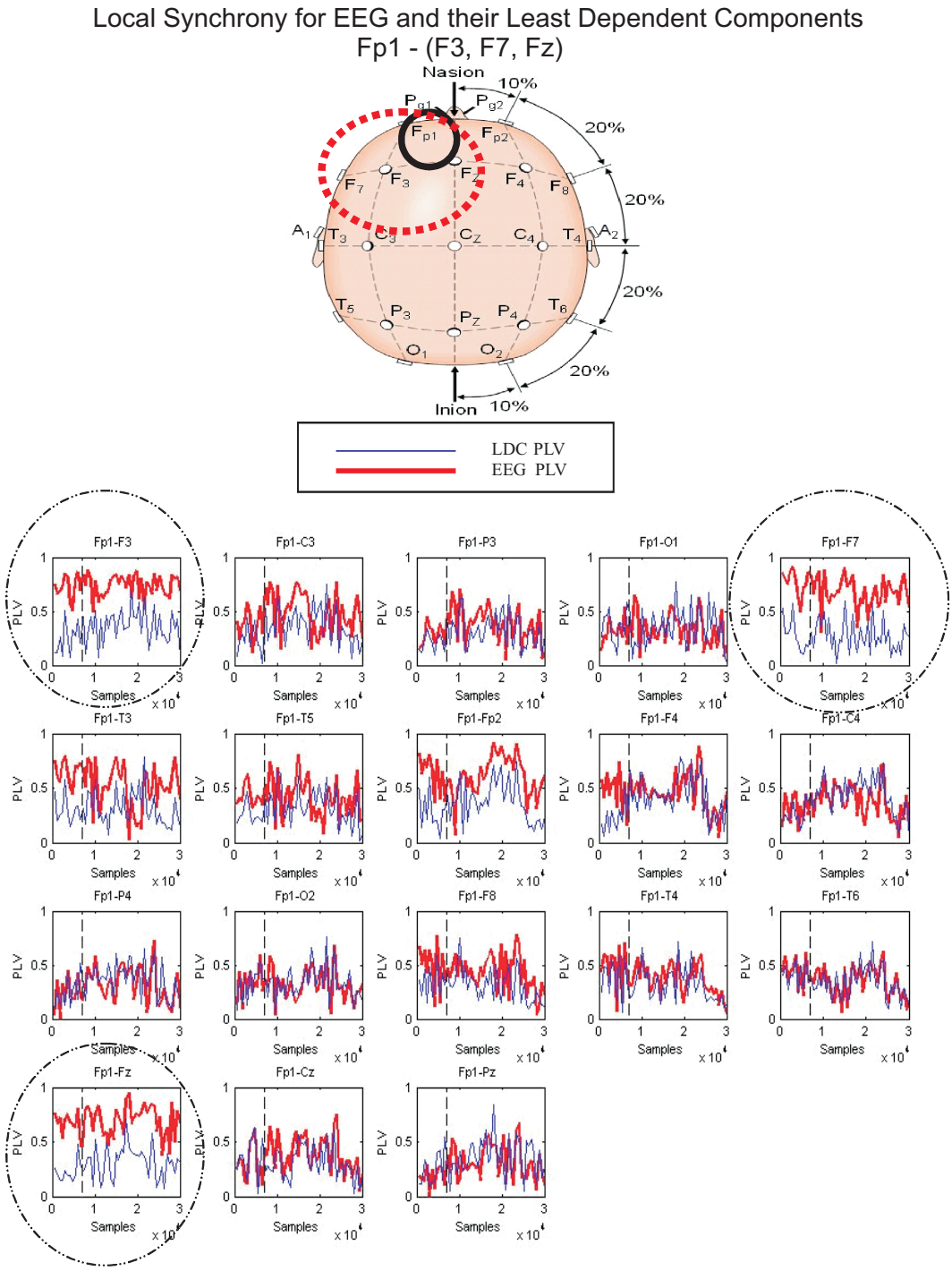


Figure 5.12: The upper image shows the 10-20 EEG electrode placement system. The electrode being analyzed is Fp1. The spatially neighboring electrodes are F3, F7 and Fz (enclosed by the dotted red line). The lower plots show the PLV of two EEG signals across a two minute segment (shown by bold red line) and the PLV of the respective least dependent components (shown by thin blue line), for all combinations of EEG electrodes with Fp1.

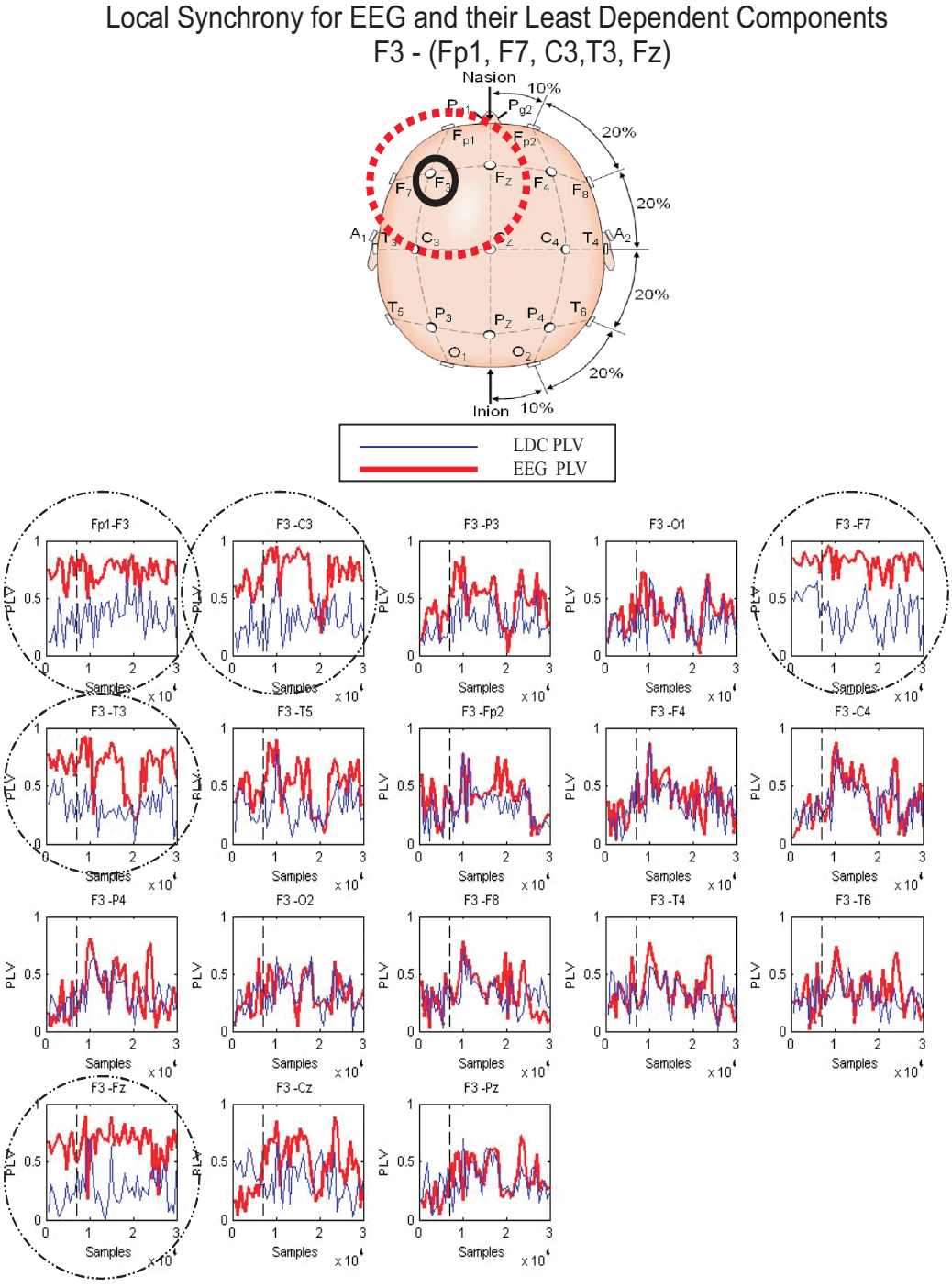


Figure 5.13: The upper image shows the 10-20 EEG electrode placement system. The electrode being analyzed is F3. The spatially neighboring electrodes are Fp1, F7, C3, T3 and Fz (enclosed by the dotted red line). The lower plots show the PLV of two EEG signals across a two minute segment (shown by bold red line) and the PLV of the respective least dependent components (shown by thin blue line), for all combinations of EEG electrodes with F3.

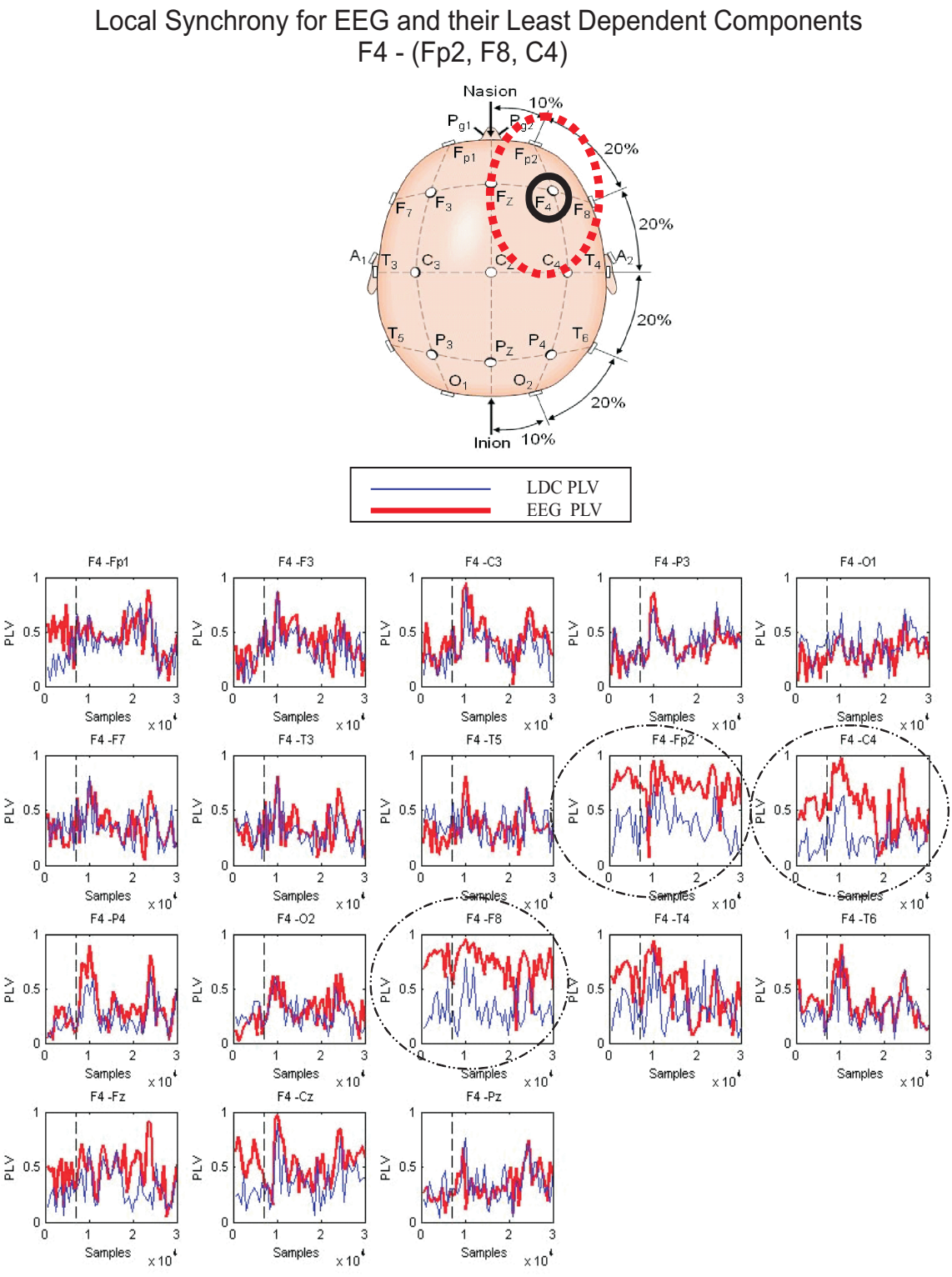


Figure 5.14: The upper image shows the 10-20 EEG electrode placement system. The electrode being analyzed is F4. The spatially neighboring electrodes are Fp2, F8 and C4 (enclosed by the dotted red line). The lower plots show the PLV of two EEG signals across a two minute segment (shown by bold red line) and the PLV of the respective least dependent components (shown by thin blue line), for all combinations of EEG electrodes with F4.

5.9 Summary

The concept of synchrony, especially phase synchronization was described in this chapter. Phase synchronization is a preferred measure for the analysis of biomedical signals as it can estimate non-linear interactions which may not be measurable by linear techniques such as cross-correlation or coherence. The underlying principle of phase synchronization has been described, and the effect of detuning or difference in frequencies shown using the limit cycle. Subsequently the effect of noise and chaos on synchrony analysis has been shown. It was seen that chaotic and noisy signals behave similarly, and with bounded low levels of noise, synchrony may get a bit diffused but may still be estimated, such as with the probability distribution of preferred phases. The measurement of phase for non-periodic signals has been explained, along with various indices used in the literature. The effect of various parameters (such as signal bandwidth, filter type, filter bandwidth, sample size) on the estimation of synchrony has also been seen. The effect of signal mixing on phase synchrony analysis was introduced here. It is important to consider when working with biomedical signals such as EEG that these are inherently mixtures of underlying signals. Thus such signals should ideally be separated prior to the application of phase synchronization analysis. Lastly the Phase Locking Statistics (PLS) were discussed to check the significance of the measured synchrony. Various surrogate methods were analyzed and phase randomization method of surrogate series and bootstrap method was found suitable for current work for significance testing. The next chapter describes the analysis and results specifically of using phase synchronization and ICA unmixing for tracking synchrony dynamics in epileptic EEG.

Chapter 6

Analyzing Epileptic EEG with Phase Synchrony and ICA: Preliminary Analysis

6.1 Overview

In Chapter 4 it was seen that ICA alone was not sufficient to objectively track changes in the brain activity, although it is good at de-noising EEG. As an additional measure, synchrony is explored because it has long been associated with epilepsy. The phase synchrony measurement methods, challenges, significance tests, and application to EEG have been described in Chapter 5. Previous studies for epileptic EEG analysis based on synchrony [77, 76, 88, 89, 72, 90, 91] have used with intracranial EEG. Synchrony was assessed as a bivariate measure, involving subjective channel selection in raw broadband EEG (usually with manual artifact removal). However, it has been recently highlighted that phase synchrony measures applied on linearly mixed signals can show spuriously high synchrony where there is none [98, 99]. Examples shown in Chapter 5, support these views. Therefore, it is important to unmix the EEG signals before phase synchrony is measured. ICA has been found to be a suitable unmixing algorithm for EEG signals, especially the temporal decorrelation techniques. It is a powerful multivariate analysis technique that helps in unmixing and de-noising multichannel signals. James *et al.* [174] have given a proof of principle for the use of ICA as a preprocessing tool for EEG seizure segments. In addition, ICA can also be used to filter the data in a spatio-temporal way by identifying and removing irrelevant patterns such as artifacts or noise. Thus ICA provides spatial information about the unmixed signals, which is quite useful in case of EEG. ICA can also be used with prior information in which case it is referred to as constrained ICA (cICA). In constrained ICA prior information in the form of spatial, temporal, or spectral cues are provided to the ICA algorithm to tune the search for independent components biased by the

	ICA	Phase Synchrony
Provides	• Analysis of multi-channel information • Unmixing overlapping signals • De-noising by spatio-temporal filtering - extracting and isolating the artifacts • Extract underlying signals with prior information removes subjective assessment	• An additional measure for assessing dynamics of brain activity • Statistical powerful technique for analyzing nonlinear couplings of signals • Use of a concept that has a strong association with epilepsy
Requires	• Additional measure to select signals of interest • A measure to track changes in the extracted signals • Additional feature to make the procedure objective	• Unmixed signals • De-noised signals • Additional multivariate analysis

Table 6.1: Table of advantages and requirements of ICA and Phase synchrony techniques

prior information. It is important to note here, that in real biomedical signals, noise and other transient non-linearities may not allow the retrieval of the true independent sources, but components with remanent dependence. Thus ICA extracted signals are better referred to as least dependent components (LDCs) here. Using phase synchrony analysis to measure the non-linear couplings in these LDCs will remove the effect of spurious detection of synchrony which is observed when EEG (mixed signals) is used instead. In this way, ICA and phase synchrony analysis complement each other and help in overcoming the challenges of the techniques for epileptic seizure prediction, summarized in Table 6.1: This shows that ICA and phase synchrony would essentially complement each other if used in combination. ICA unmixes and de-noises the EEG and also uses multivariate information present in the signals. Spatial constraints on ICA will remove the ambiguity of channel/component selection required in phase synchrony analysis. On the other hand, phase synchronization is able to assess a dynamic

for seizure prediction, closely associated with epilepsy, with EEG as the data. ICA would help to invariably include multivariate information in phase synchrony analysis as opposed to using information of only two signals. Phase synchrony along with ICA is also useful for making the analysis for seizure prediction more objective.

However, using ICA and phase synchrony in combination poses a contradiction in theory. ICA essentially aims to extract signals that are independent while phase synchrony analyzes the coupling or dependence of the independent signals. In theory, if signals are truly independent, they cannot be synchronized and if they are synchronized, they cannot be independent. Nevertheless, the 'independent' signals obtained after applying ICA on EEG, show a high synchrony amongst sets of signals. This high synchrony is seen to exist near a seizure, which correlates with the literature on hyper synchronized state of seizure. This contradiction can be explained as follows:

Remnant Dependence: Ideally ICA aims to extract signals that are independent. However, with real signals, especially biomedical signals, it is difficult to ascertain the number of signals to be extracted, which leads to overfitting. In order to extract more sources than present in the mixtures, a few extracted sources can get split into more than one source. These split sources have remnant dependencies which ICA has characteristically failed to remove. Such remnant dependencies have recently been shown to exist in EEG analysis by Stögbauer *et al.*, [23, 24, 25]. Stögbauer, *et al.*, and Himberg and Hyvärinen, have described the use of mutual information estimators for testing reliability of decomposition by ICA. The variability of the mutual information with a rotational remixing is used as a measure of stability of ICs. Meinecke, *et al.*,[150, 25] have described a bootstrap resampling approach using surrogate data. They have described the use of a *separability matrix* that shows how reliably the one dimension subspace can be separated. The surrogate data is obtained from the block of data points of the independent components by randomly drawing a set of say p data points. Then a separating matrix is computed on the original block of data points and the surrogate sets of data points. On the surrogate data, the source separation algorithm is used again to estimate a rotation that separates the surrogate data. The rotation matrices are then compared. Independent sources (one dimensional) have rotation matrices in the vicinity of the identity matrix. The separability matrix reinforces the concept of existing remnant dependencies between ICs. These remnant dependencies may coincidentally be exhibited in the form of phase couplings.

Nonlinear Coupling: Another cause for ICs showing phase synchrony can be reasoned as ICA being unable to process the nonlinear coupling/dependence that phase synchrony assesses. Hence the dependence that ICA fails to remove is essentially useful for phase synchrony analysis to reveal dependencies of interest in underlying brain activity. At this stage this is just an hypothesis though.

Nonetheless, there is still value in performing ICA, as apart from de-noising, it provides an unmixing of components, albeit still with residual dependencies, but which tend to be free from volume conduction effects and suitable for direct analysis with phase synchronization. Henceforth, ICA and Phase synchrony analysis (using the Phase Locking Value, PLV) are used in combination, and are referred to as ICA-PLV analysis.

6.2 Data set used for the preliminary analysis

The data set consists of anonymized multi-channel scalp EEG recordings of 10 patients who were undergoing continuous scalp EEG monitoring for possible epileptic surgery. The EEG was in the form of segments of two minutes with gaps of fifteen minutes, throughout the day and night. The data was recorded using twenty-five electrodes placed on the scalp according to the international 10-20 electrode placement system, with a reference FCz. It was sampled at 200 Hz and digitally stored at 12 bit resolution. The inter-ictal recordings consist of 120 seconds of non-ictal EEG, away from the seizures. The seizure recordings consist of some pre-ictal periods and are of variable lengths from 120 seconds to 300 seconds. Random ictal and interictal segments have been taken from the entire data set per patient, for the preliminary analysis. It includes 4-5 seizures and 20 interictal recordings (distributed across the day) per patient. It has been tabulated in Table 6.2.

6.3 Objective Selection of Seizure Sources

The first step in ICA-PLV analysis is the ICA unmixing and selection of source signals of interest from the extracted set of sources. After un-mixing the signals the main task is to find the physiological explanation of the estimated sources. Presently in research, this is usually done by subjective analysis and remains a bottleneck in ICA applications. The components are selected using temporal or spatial information (topographic maps) or other *a priori* information from the patient history [174, 104, 9]. This involves subjectivity and may not be the most efficient process when long term continuous data is being assessed and/or for automated software. Thus it is desirable to identify the components of interest **objectively** based on other underlying signal information. Interestingly, the concept of phase synchronization has been found to overcome this subjectivity of selecting seizure sources in epileptic EEG, in a way described below.

It is known that at the time of seizure onset the synchrony of neurons close to seizure focus increase dramatically (termed as hyper-synchronous). This should be perceivable at the 'seizure sources' when phase synchrony is applied on pairs of unmixed EEG signals from a short ictal segment. The synchrony of artifacts during this phase may

easily get overshadowed by the hyper-synchrony of the seizure signals and these in effect make the seizure source synchrony appear strongest. This strong synchrony can then be used to objectively select the seizure components based on the underlying principle of hyper-synchrony expected at seizure onset. This has been explained with the help of examples shown below.

The selection of sources is performed on shorter EEG segments of 20 seconds (4000 samples) for each patient (Table 6.2). The segment is first pre-processed by removing the mean across each channel and the line noise (50/60 Hz) is filtered out. This multichannel EEG is temporally decorrelated to estimate LDCs using TDSEP. The sources obtained are then filtered with a bandpass filter between 2-8 Hz (as the seizure rhythm of interest is usually within these bands). The filtered sources are used to track significant phase synchrony across time for all combinations of the sources for the segment, using a non-overlapping moving window of 600 samples. The significance of the PLV values are tested using phase randomized 100 surrogate signals. The average of PLS (statistically significant PLV) for each combination of sources is estimated for the length of the segment. Topographies of the two source combinations with the highest

S. No.	**Patient**	**No. of Seizures**	**Seizure onset time (Date)**	**No. of Inter ictal (Date)**
1	Patient 1	5	06:35 a.m. (13 March) 05:29 a.m. (13 March) 04:08 a.m. (13 March) 02:08 a.m. (13 March) 00:41 a.m. (13 March)	20 (12 March)
2	Patient 2	5	02:00 a.m. (13 April) 06:05 a.m. (13 April) 14:56 p.m. (14 April) 16:06 p.m. (14 April) 22:45 p.m. (14 April)	20 (12 April)
3	Patient 3	4	00:07 a.m. (8 April) 02:14 a.m. (8 April) 18:44 p.m. (9 April) 23:18 p.m. (9 April)	20 (8 April)
4	Patient 4	4	20:14 p.m. (3 Feb) 16:23 p.m. (7 Feb) 18:44 p.m. (9 Feb) 18:53 p.m. (31 Jan)	20 (3, 7 Feb)
5	Patient 5	4	18:44 p.m. (3 May) 23:52 p.m. (3 May) 05:14 a.m. (5 May) 15:05 p.m. (5 May)	20 (3, 4 May)
6	Patient 6	4	19:10 p.m. (1 April) 19:12 p.m. (2 April) 17:13 p.m. (3 April) 19:14 p.m. (30 March)	20 (30 March)

S. No.	Patient	No. of Seizures	Seizure onset time (Date)	No. of Inter ictal (Date)
7	Patient 7	5	09:03 a.m. (8 April) 10:12 a.m. (17 April) 12:02 p.m. (17 April) 14:59 p.m. (17 April) 17:44 p.m. (17 April)	20 (17 April)
8	Patient 8	5	13:27 p.m. (10 March) 20:51 p.m. (17 March) 18:25 p.m. (20 March) 08:02 a.m. (22 March) 01:54 a.m. (23 March)	20 (10 March)
9	Patient 9	5	18:12 p.m. (17 Feb) 05:24 a.m. (20 Feb) 05:29 a.m. (20 Feb) 13:38 p.m. (16 Feb) 01:53 a.m. (17 Feb)	20 (15 Feb)
10	Patient 10	4	15:24 p.m. (3 May) 07:59 a.m. (3 May) 05:59 a.m. (3 May) 08:29 a.m. (3 May)	20 (3 May)

Table 6.2: Data set of short 2 minute multichannel scalp EEG segments, used in the preliminary analysis. It has been recorded with the 10-20 EEG electrode international system and is sampled at 200 Hz. The set includes 4-5 seizures and 20 interictal recordings (distributed across the day) per patient. The dates and times of the seizures and interictal recordings are shown to provide an idea of the nearness of the interictal recordings to the seizure onset times.

average significant phase synchrony in the segment are then selected and shown here.

They are seen to coincide with the expected seizure sources. Figure 6.1 shows the results for few of the patients. The seizure topographies for different seizures of the same patient, obtained from the above procedure with ICA and phase synchrony are shown. The expected topographical area of interest for the patients type of epilepsy is also shown along with it. It is observed that the procedure applied to different seizures of a patient shows the same topographies for that patient. These results are promising as they indicate a robustness in the objective selection of seizure topographies using ICA and phase synchrony. This selection of sources of interest has been a bottleneck of research in ICA.

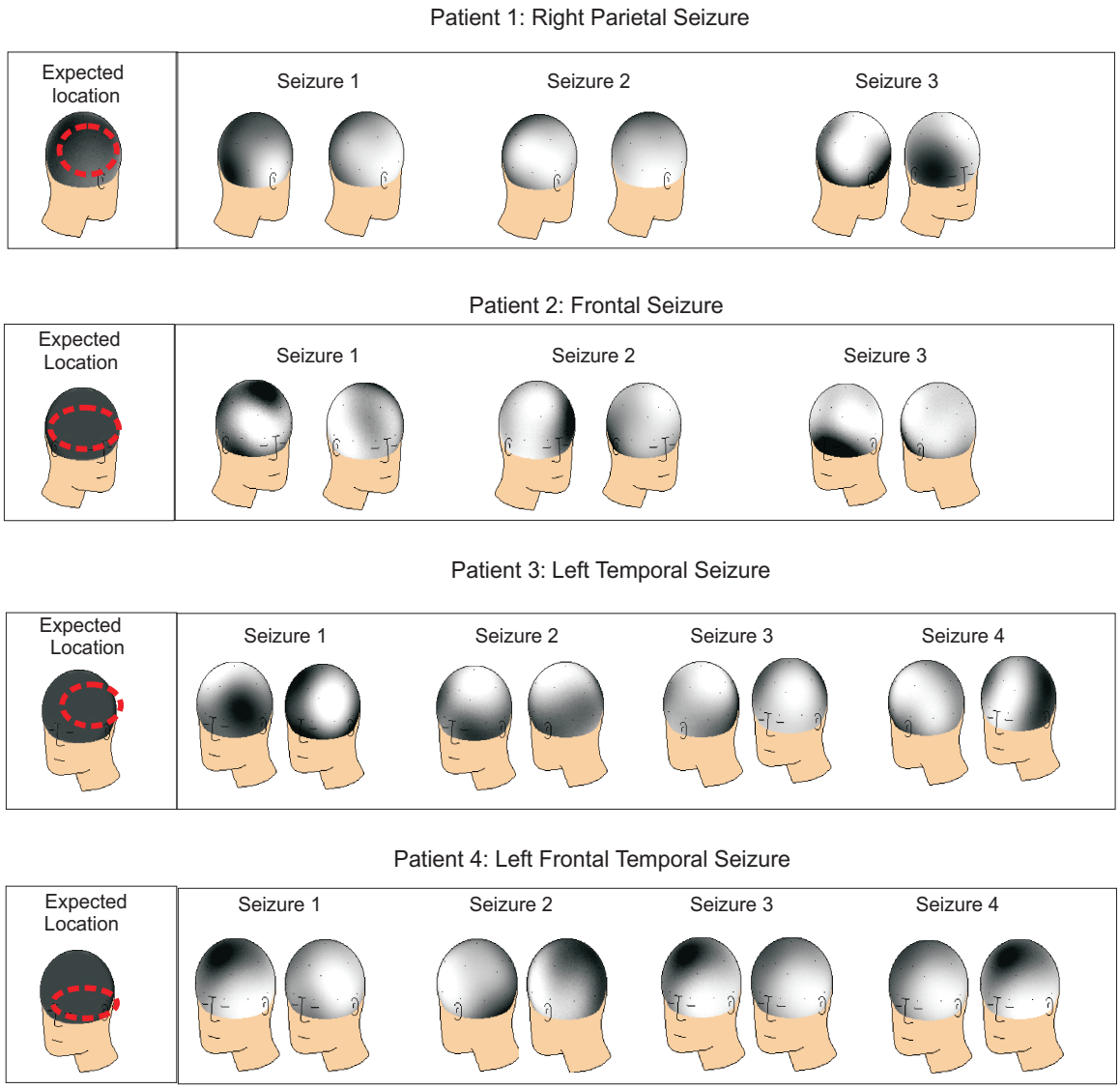


Figure 6.1: Objective selection of seizure sources using ICA and phase synchrony performed on ictal EEG segments of epileptic patients. The epileptic EEG used are: Patient 1: Right Parietal Seizure, Patient 2: Frontal Seizure, Patient 3: Temporal Seizure, Patient 4: Left Frontal Temporal Seizure. The topographical area of interest for the type of epilepsy of each patient is marked in the left figures (dashed circle). The pair of source topographies having highest average phase synchrony are shown on the right. At least one of the pair of objectively selected topographies are seen to coincide with the expected topographies, across seizures for each patient.

6.4 Objective De-Noising Epileptic EEG

Objective selection of seizure sources can be applied for de-noising multichannel EEG as well. In epileptic EEG, as the seizure approaches, associated activity like teeth clenching, muscle spasms, electrode movement adds a lot of noise and artifacts. The ictal information can easily get overshadowed by the noise and artifacts, for these reasons it is generally desirable to de-noise the EEG. Filtering is the usual method of de-noising a signal, but designing a filter for physiological signals like the multichannel EEG is not trivial as the information and noise is generally spread over different channels of the recording, and across time, in a very complex manner. In the context of epileptic EEG analysis, phase synchronization has been found useful for de-nosing EEG objectively.

As mentioned in the previous section, isolation of seizure components has been found to be possible by phase synchronization and this can be useful for de-noising the epileptic EEG as shown in the following examples.

Ictal EEG segments of 20 seconds from epileptic patients are used (see Table 6.2). The sampling frequency of the EEG is 200 Hz. The segments shown here are from patient 7 (temporal seizures) and patient 3 (right parietal seizure). The EEG has been pre-processed by removing mean and filtering line noise (50 Hz). ICA is applied on this pre-processed segment to estimate underlying signals and the distribution of those source signals (mixing matrix). The sources are then filtered from 2-8 Hz. The PLV is then estimated across time with non-overlapping windows of 600 samples, for all combinations of the sources. The significance of the PLV values are checked using phase randomized 100 surrogate data signals. For each source combination the average of significant PLV values are found for that segment. The average significant phase synchrony across the ictal segment for all combination of sources is then plotted on a grid. The source combinations showing the maximum average significant phase synchrony is then used to de-noise the EEG *objectivly*.

This has been retrospectively found to coincide with the subjective seizure source selection.

To de-noise the EEG, all the columns of the mixing matrix ($\mathbf{A}$) that do not correspond to the above seizure components are replaced by zeros giving a new mixing matrix $\mathbf{A}_{new}$. The mixing matrix $\mathbf{A}_{new}$ is then used to back-project the EEG mixtures to get de-noised EEG. The original EEG, the seizure sources identified, de-noised EEG and the correlation of the original and de-noised EEG is shown in the following 2 examples:

Temporal Seizure (Patient 1, seizure 1)

The EEG used here (Figure 6.2), indicates seizure topography to be left temporal at seizure onset. Involvement of left frontal-temporal and right parietal areas is also seen

after some seconds following seizure onset. (The seizure onset for this data segment was subjectively determined to be 5-7 seconds into the data segment.)

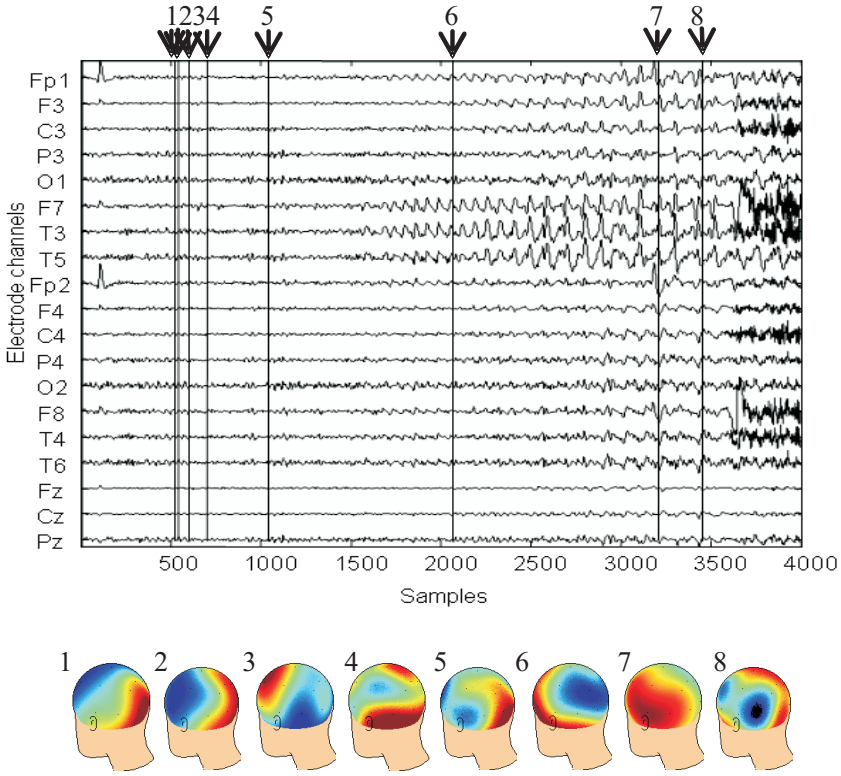


Figure 6.2: A 20 sec segment of raw ictal EEG (left temporal focus) along with topographic maps at various intervals (marked by straight lines)

The ICs obtained by using Fast ICA on raw EEG are shown along with their corresponding topographic maps, in Figure 6.3 the variation of average PLV values for the component combinations have been shown in the grid in Figure 6.4. The axes of the grid correspond to the ICs and the color in the grid squares shows the strength of phase synchronization for the combination of the corresponding source components. The components having the highest phase synchrony were found to be IC9 and IC13, IC1 and IC3, IC16 and IC17, and IC4 and IC7.

These were used for de-noising raw EEG as they were subsequently found to represent the seizure as well. To see the effect of de-noising, the correlation between the raw EEG and the de-noised EEG is plotted as a topographic map (see Figure 6.5) and it is seen that the EEG has been selectively de-noised at channels away from the seizure channels (without loss of seizure related information), even though it is not discernable to the naked eye.

De-noising in a similar fashion with fixed filters may be hard or near impossible. The topographies of the de-noised EEG are shown at various intervals of the ictal period. It can be seen that the underlying topography remains stationary which is expected as it was fixed with the use of ICA. However the changing powers of the potential map

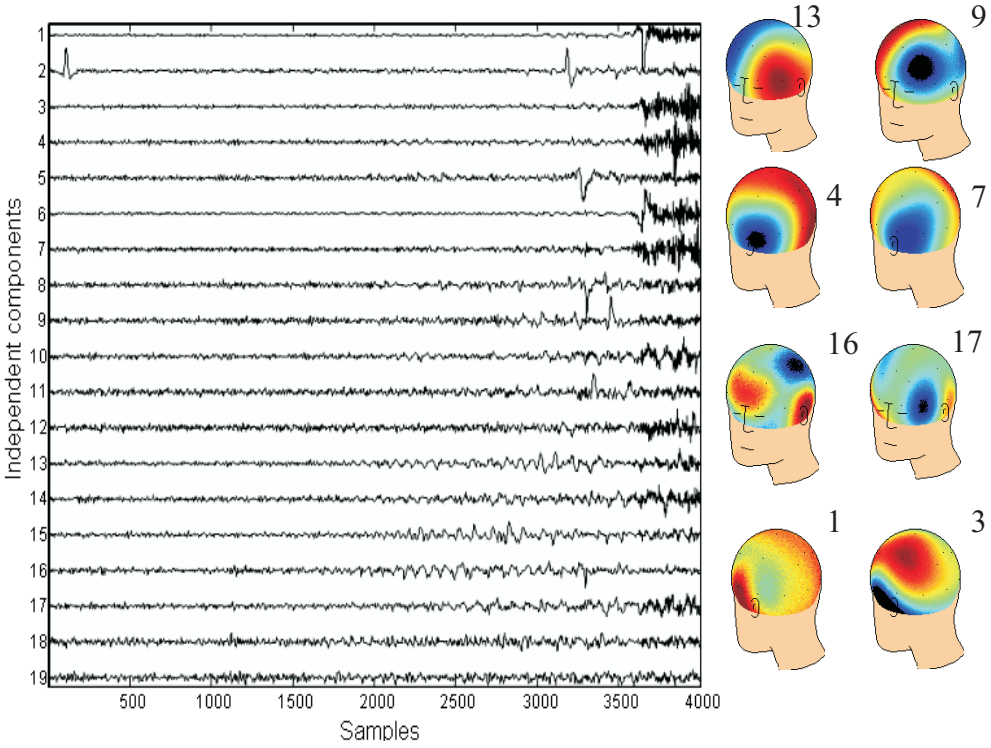


Figure 6.3: Source signal (obtained by using Fast ICA on raw EEG) and corresponding topographic maps of the Independent components.)

focussed at the seizure space can be viewed from the re-mixed spatially filtered EEG (see Figure 6.6). This is usually lost in the raw EEG due to the eclipsing effects of the artifacts (see topographies shown in Figure 6.2), and it usually requires subjective analysis or cues from visually discernible changes in rhythmic structure of time series, clinical changes in patient, known ictal patterns, a priori knowledge from patient's history, etc. in order to effect a diagnosis. The technique is in effect a form of spatial filtering where the spatial filter has been identified using the concept of phase synchronization and the knowledge about the relation of synchrony and the ictal phase. This has allowed to make the de-noising by ICA an objective process for epileptic EEG.

Right parietal focus: (Patient 2, seizure 1) Another ictal segment of 20 seconds is shown in Figure 6.7 with original and de-noised EEG and the topographies of the estimated sources that were selected using the phase synchronization. The correlation between the original and de-noised EEG is plotted as a topography in Figure 6.7(c).

In the above examples it was observed retrospectively that the topographies selected by the phase synchrony algorithm correspond to the seizure topography for the patient, useful for de-noising EEG objectively.

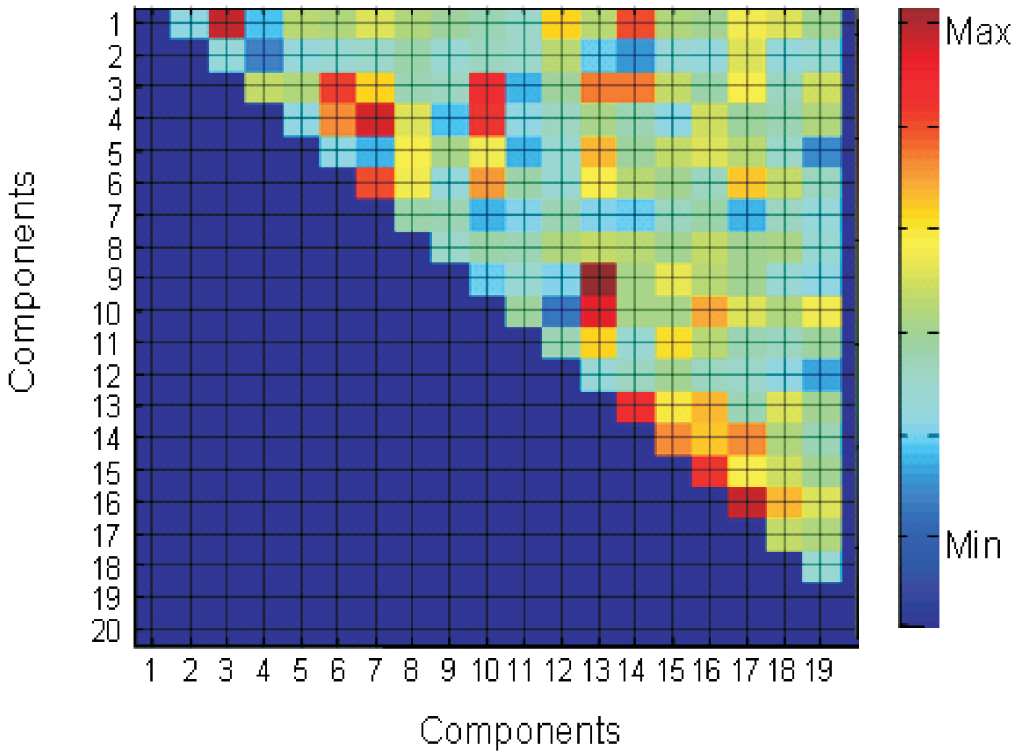


Figure 6.4: Variation of averaged PLV values for all IC combinations where the axes correspond to ICs and color corresponds to the strength of synchronization. As the conjugate pairs were left out of PLV calculations due to symmetry, the left lower section in the grid, below the diagonal should be ignored.

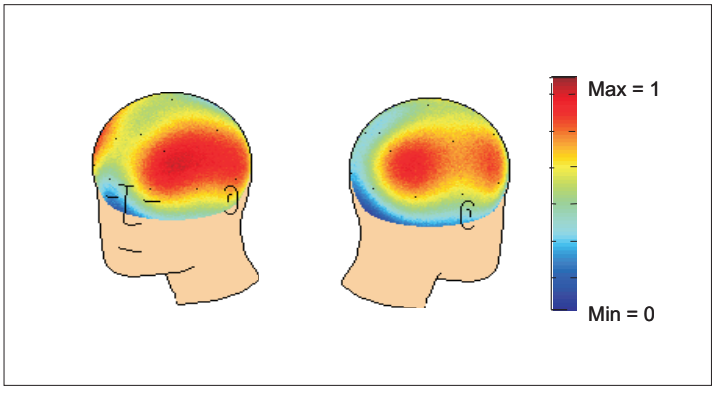


Figure 6.5: Topographic maps of correlation coefficient of raw ictal EEG and de-noised EEG (de-noised by ICA and PLV)

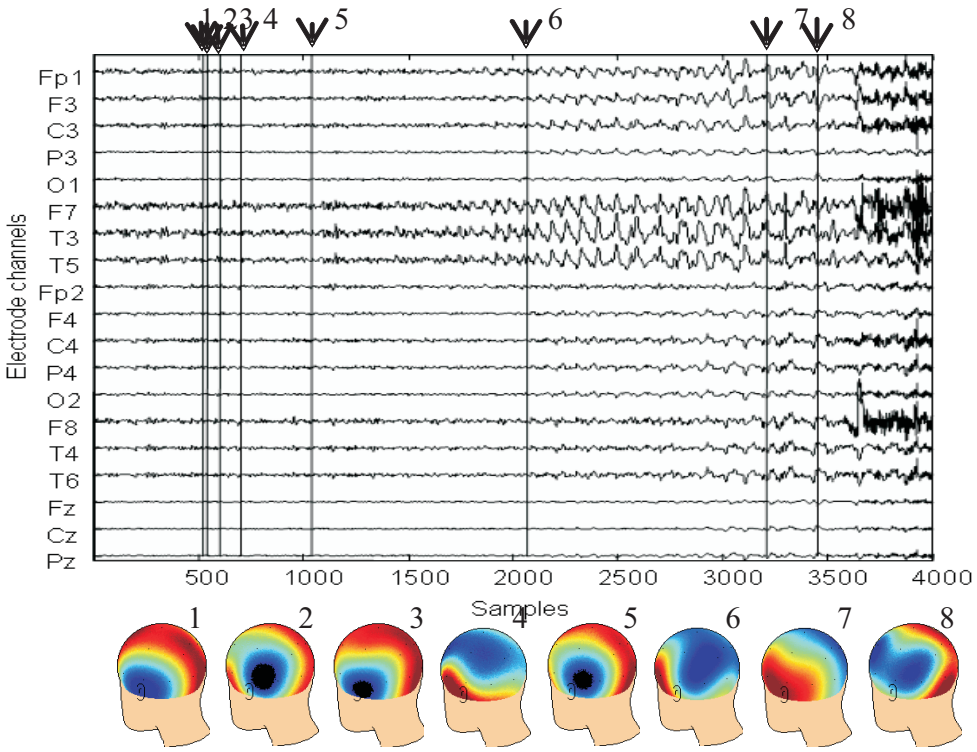


Figure 6.6: A 20 sec segment of ictal EEG (left temporal focus) de-noised by ICA and phase synchrony, along with topographic maps at various intervals (marked by straight lines). The topographies reflect the changing power in the spatial locations fixed by ICA.

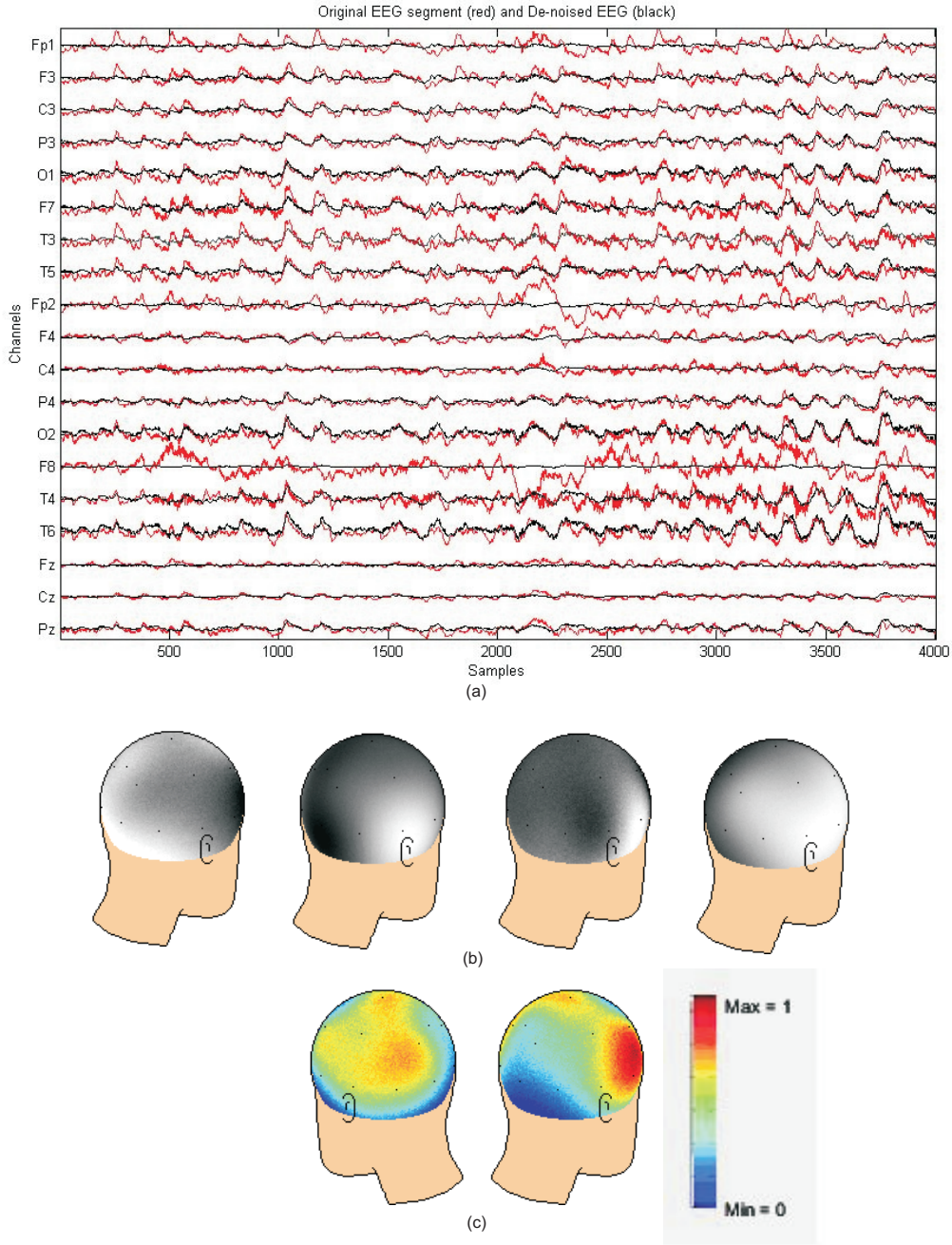


Figure 6.7: (a)A 20 second segment of ictal EEG (red plot) (right parietal focus) de-noised (black plot) by ICA and phase synchrony, (b) Topographic plots of sources that were isolated objectively using phase synchrony and were used to de-noise the EEG, (c) Correlation between original and de-noised EEG is shown topographically

6.5 ICA-PLV Analysis for Bivariate vs Multivariate ICA

The ICA unmixing and phase synchrony analysis of the unmixed LDCs can be carried out as bivariate or with a multivariate element. Phase synchrony estimation is bivariate as it is based on two signals at a time. The two signals for phase synchrony can be obtained either from the ICA unmixing of two pre-selected EEG channels, or from two LDCs obtained from ICA unmixing of multichannel EEG. The bivariate unmixing involves subjective selection of signals, while multichannel unmixing brings in the multivariate information of the entire EEG in the phase synchrony estimation as well. These two approaches for ICA-PLV are discussed here.

Hitherto, literature on phase synchrony analysis of epileptic EEG, has usually used two channels of EEG, selected subjectively (contralateral or focal) without ascribing any criteria for selection. However, using ICA-PLV as proposed here for bivariate EEG data has certain disadvantages, such as:

1. With the use of only two channels of EEG for analysis, the spatial information provided by the EEG signals is not utilized fully. Usually all the underlying signals from the brain may not be able to reach the scalp electrodes close to them because the effects of volume conduction. The EEG channels being placed in a particular systematic way on the scalp allow to capture signals from the brain from various locations. For a complete analysis, the multichannel EEG can provide more information than bivariate analysis.

2. Bivariate analysis is restricted to two channels only, requiring prior selection of 'good' channels from the multichannel EEG (19 channels). This brings subjectivity in the analysis, especially when the explanations of a 'good' or 'possibly useful' channel is still unclear in epilepsy.

3. ICA will not be efficient in un-mixing bivariate signals. It may be able to find independent sources better when more information is provided to the algorithm. ICA on bivariate signals will output a maximum of 2 extracted sources (due to the assumption of square mixing).

A comparison of bivariate and multivariate analysis of EEG phase synchrony with ICA unmixing is shown here to support these views.

The analysis is carried out using 150 seconds long multichannel EEG segment. The multichannel EEG segment used here is an ictal EEG segment from an epileptic patient (Patient 3 from the data set described in Table 6.2). A zoomed segment of the EEG is shown in Figure 6.8. The seizures for this patient originate in the right parietal area/frontal areas. This involves the area around the EEG electrodes T6, P4, T4, C4 and F8 (established from the knowledge of clinical seizures of this patient).

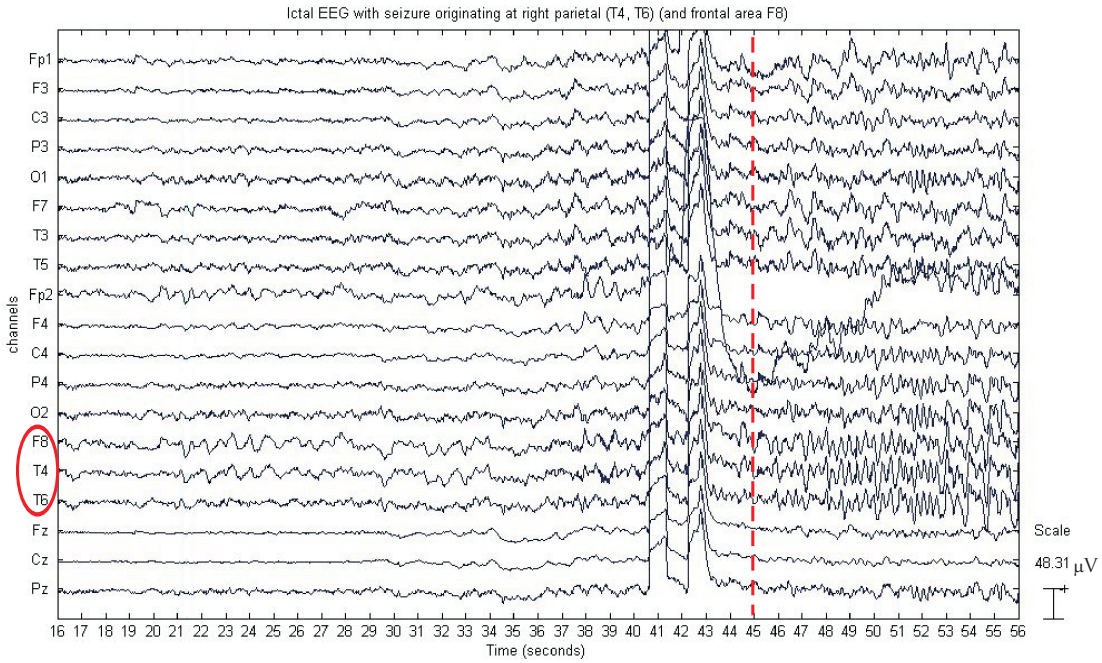


Figure 6.8: An EEG section of a 19 channel 150 seconds ictal EEG segment from an epileptic patient. The seizure onsets at 45 seconds into the recording (marked by a vertical line) originating in right parietal and right frontal area.

1. Bivariate analysis: a bivariate analysis only two channels of EEG are used for computation of two components at a time. All possible combinations of the 19 EEG channels, taken two at a time are used. They are pre-processed by removing the mean and are then unmixed with ICA (temporal decorrelation) as signal pairs to obtain two least dependent components. The unmixed components are then filtered using a bandpass filter of 2-8 Hz. The PLV is then computed across the two minute components using a non-overlapping moving windows of 600 samples. The PLV estimated for bivariate components for a pair of *focal* electrode signals (T6 and F8) is shown in Figure 6.9. No demarcated change can be observed in the phase synchrony along with seizure onset (known) as is expected as in previous examples.

2. Multivariate analysis: In the multivariate analysis the entire 19 channel EEG is used for unmixing with ICA. It is temporally decorrelated using TDSEP to obtain 19 independent components. The components are then filtered using the bandpass filter of 2-8 Hz. The filtered components are used to estimate PLV across the two minutes, using non-overlapping moving windows of 600 samples. The topographies of the sources are also obtained from the mixing matrix. The PLV estimated for multivariate components for sources with topographies matching the focal areas T6 and F8 is shown in Figure 6.10(a). The topographies of these estimate from sources is shown in Figure 6.10(b). A demarcated increase in phase synchrony can be clearly observed at seizure onset (known) with the multivariate

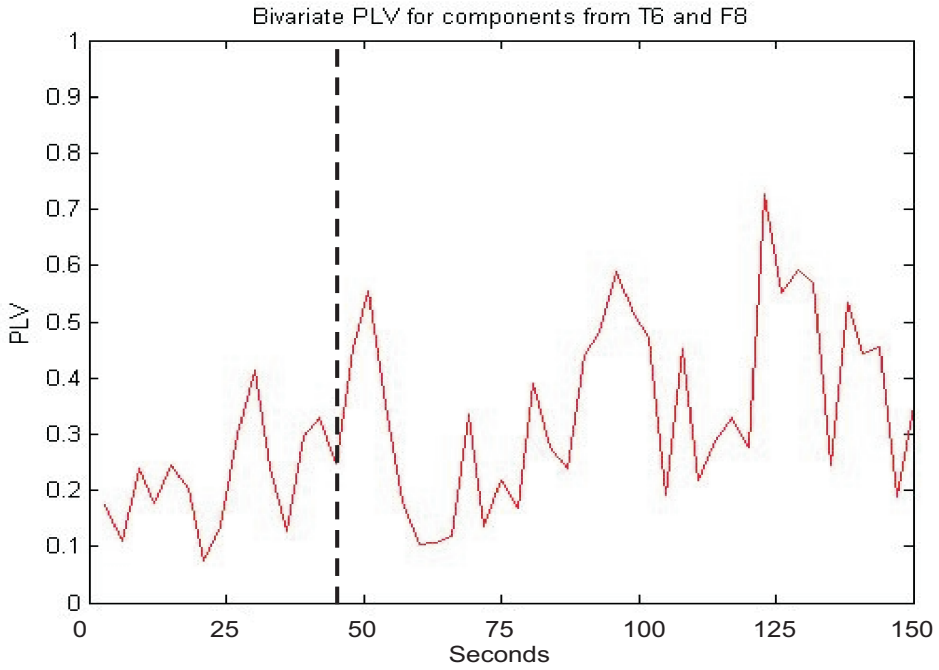


Figure 6.9: PLV for bivariate sources obtained from ICA-PLV analysis of a 19 channel 150 second ictal EEG segment. The channels selected were those focal to the seizure onset T6 and F8. No specific pattern is apparent in the PLV plot of the bivariate ICA-PLV analysis. The vertical line indicates the seizure onset.

analysis.

A difference can be observed between the bivariate and multivariate ICA-PLV analysis. The multivariate plots show a low background phase synchrony and an increase of synchrony at seizure onset. This pattern is not apparent with bivariate analysis. Such patterns recur when applied to other seizures of the same patient. These results reinforce the need to use multivariate analysis as opposed to bivariate analysis, as multivariate analysis provides information that will help in the context of seizure detection or prediction.

6.6 Synchrony Dynamics with an Approaching Seizure

Application of ICA-PLV for objective selection of seizure sources and de-noising has been found to be very useful for epileptic EEG analysis. However, it is important to note that this procedure is highly dependent on the dramatic changes the signal undergoes as it transitions from an interictal to an ictal episode. Extracting the seizure sources objectively in data windows much prior to ictus remains challenging, because then the phase synchrony of other artifacts such as eye blinks can take precedence. As continuous EEG needs to be assessed using short window segments only, for stationarity

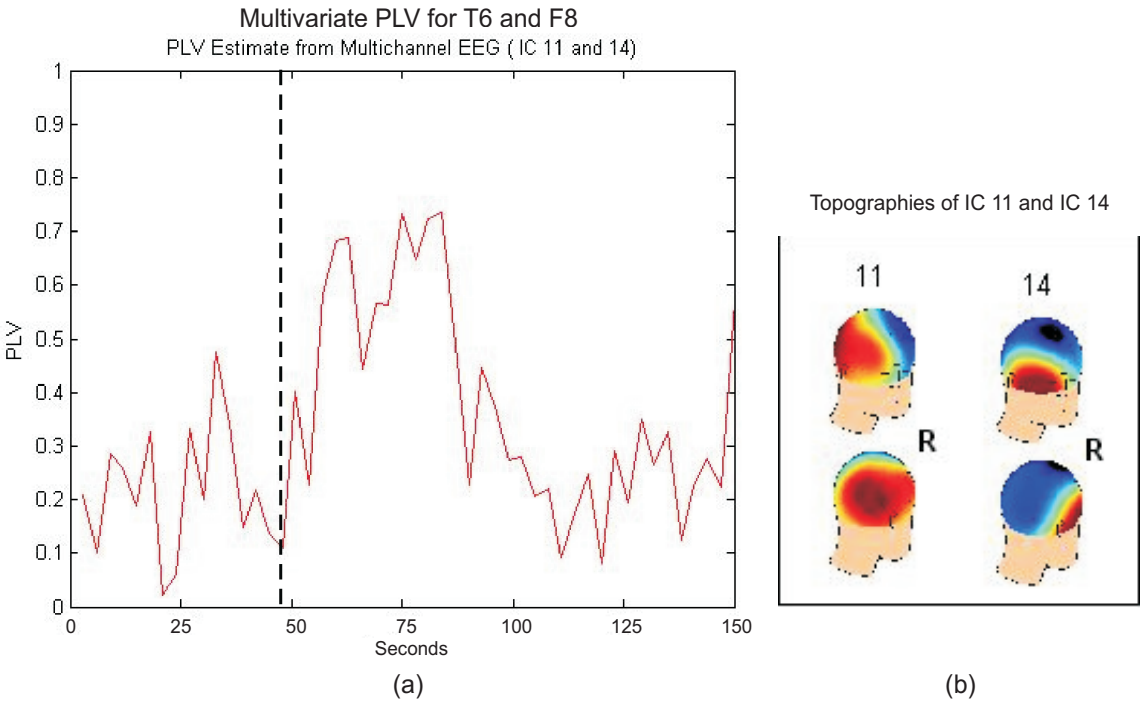


Figure 6.10: (a) PLV for independent seizure sources 11 and 14 of multichannel ictal EEG with seizure originating in right parietal and right frontal area. Multivariate ICA-PLV analysis shows an increase in synchrony at seizure onset, as opposed to bivariate analysis. The vertical line indicates the seizure onset. (b) Topographies of the ICs 11 and 14 also show the right parietal and frontal area as expected.

purposes, the prominent synchrony information prior to ictus is unavailable in the segments. The lack of known morphological, spectral, spatial changes in the signal, and the unknown nature of development of the seizure source signal in particular, makes it impossible to track a source of interest across time. Also, as mentioned before in Chapter 4, with the changing number of underlying sources, and change in sources because of change in dominant artifacts, can make the sorting of sources across windows quite impossible. In an analysis of continuous EEG of 1 patient for 3 days i.e. approximately 5.1×10^7 samples (sampling frequency being 200 Hz), and a non-overlapping window segment of about 4 seconds (800 samples) or at most 2 minutes (24000 samples), will require analysis of 64800 or 2160 segments respectively (more if overlapping windows are used). It is not a trivial task to manually select seizure component signals from the disordered components obtained from each EEG segment (especially in case of segmented long term data). Nonetheless, for a comprehensive analysis of seizure prediction, long term interictal and ictal synchrony dynamics of a seizure signal need to be assessed and for statistical reasons this shall be for more than 1-2 patients. Thus it requires a technique that is able to select a seizure source from the varied set of sources obtained in each segment, away from the time of seizure onset, in the absence of any morphological, spectral or other visually apparent feature.

To resolve this challenge, the objective ICA-PLV based selection of seizure source signals in an interictal-ictal segment is used as feedback to find seizure sources in segments prior to ictus. This is done by, first using the ICA-PLV technique on a segment that has sufficient interictal and ictal activity, that allows the extraction of the pair of seizure sources that show outstanding change in phase synchrony at seizure. Thereafter, using the spatial information of the seizure sources obtained, spatially constrained ICA is performed for source extraction in the remaining segments. The PLV of these extracted sources then allows tracking the dynamics of synchrony of LDCs in a particular area of the brain. It should be noted that this area of the brain (spatial constraint) will be found *objectively* using interictal-ictal segment. It may or may not coincide with the seizure focus determined through the clinical manifestation of the patients epilepsy.

This procedure was applied to short term EEG segments of 2 minutes each, as at this point the long term EEG was still unavailable and was being collected. This algorithm was found to be useful for seizure onset detection. It was applied to multichannel scalp EEG data set of 10 patients with 4-5 seizure each and found to show significant results. Statistical tests for synchrony and bandwidth of LDC signals were carried out and these along with the seizure onset detection results are discussed in following sections.

6.7 Seizure Onset Detection

The ICA-PLV was found useful for seizure onset detection. The algorithm used is detailed below:

1. First a multichannel template EEG segment is formed for the the patient. The template EEG used is such that it exhibits sufficient pre-ictal and ictal periods (about 100 seconds of each).
2. ICs are estimated for this segment using TDSEP and the topographies are plotted from the mixing matrix **A**.
3. The separability matrix is formed for these sources. This separability is based on the resampling approach for determining IC reliability (discussed in Chapter 4).
4. Topographies of the sources obtained are also generated.
5. Using the information from the time domain signal, topography plots and the separability matrix, the sources that show dependence are clustered together manually. These cluster topographies are then used as **constraints** for further analysis of any segment (ictal or inter-ictal) of the same patient.
6. The next step is to select two spatial constraints from these multiple constraints found above. For selection of constraints, all combinations of these constraints are used in pairs as two spatially constrained sources.
7. These sources are filtered over 2-8 Hz and their PLV is obtained and tracked across the EEG sample data. The pair of constraint topographies that show the most outstanding PLV change (determined subjectively) at seizure onset is **selected as the constraint template** for further analysis. This indicates the projected areas of the brain where the maximal change in synchrony takes place at the transition towards a seizure.
8. This constraint is saved as template for application to the rest of the interictal, pre-ictal, ictal continuous EEG analysis of the patient. Signals of interest in a particular spatial area (may or may not be near the epileptogenic focal area) are then partially isolated, with the help of spatially constrained ICA. Spatially constrained ICA has been found to be better suited for biomedical signals than spatial filtering [140].
9. The seizure sources extracted are filtered between 2-8 Hz. The evolution of phase synchrony across time is then assessed with these seizure sources.

The data used here is tabulated in Table 6.2 and described in Section 6.2. The data of each patient was analyzed separately using a moving window technique with a window

length of 600 samples (3 seconds). Also the patient data spans a variety of seizure types, such as frontal, left temporal and right parietal. Due to space constraints only, the PLV results for 4 patients are shown here in detail, describing the source signal selection, few exemplary PLV and PLS (statistically significant PLV) plots for the ictal and interictal data. A few resulting PLS plots for the rest of the patients are also shown, to describe the consistency in results, while the statistics have been tabulated in Table 6.7.2.

6.7.1 Tracking Synchrony across time with ICA-PLV

In this section, the above analysis is performed on ictal segments, to obtain constraints and then the constraint templates are used to obtain spatially constrained sources. The PLV of those sources are then tracked across time on other ictal segments of the same patient. Th results of the 4 patients are shown in Figures 6.11, 6.12, 6.13 and 6.14. Each plot shows the PLV obtained for the spatially constrained seizure sources across 4 ictal segments of one patient.

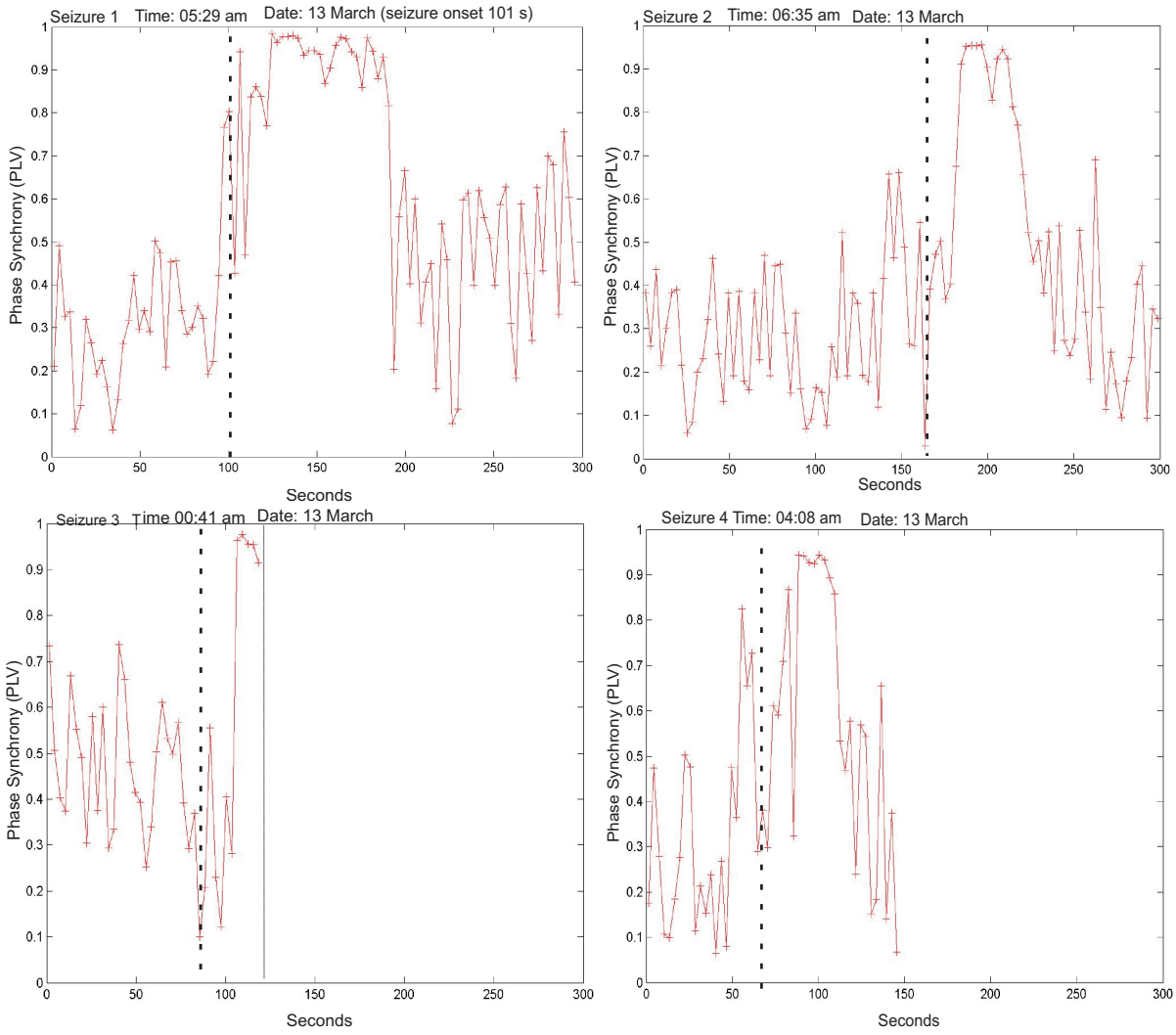


Figure 6.11: **Patient 1: Temporal Seizures**: PLV tracked across time for four seizure segments of patient 1. Vertical lines show the seizure onset times. Seizure 4: data available is only till that marked by a solid vertical line.

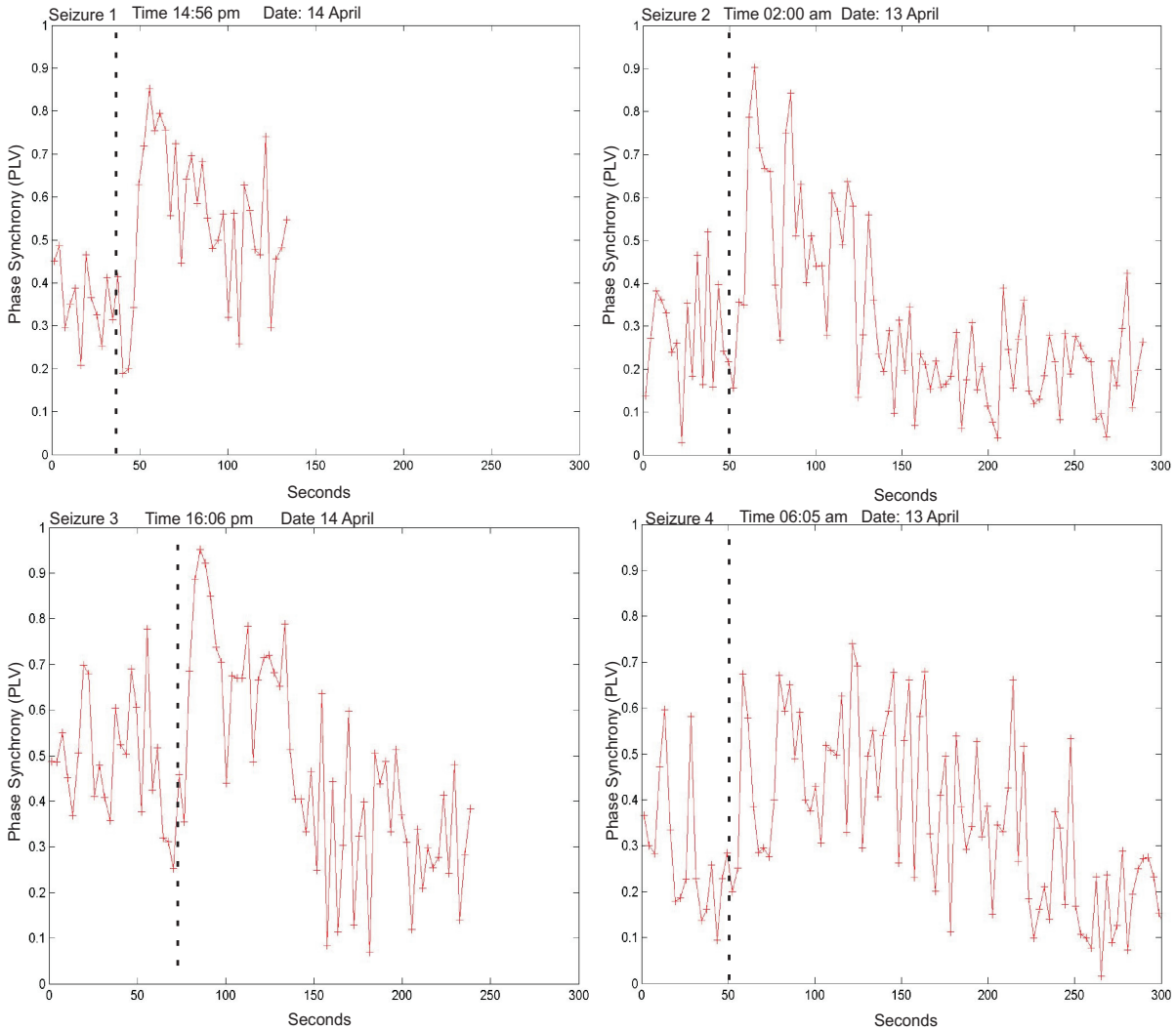


Figure 6.12: **Patient 2: Left Fronto Temporal Seizures**: PLV tracked across time for four seizure segments of patient 2. Vertical lines show the seizure onset times.

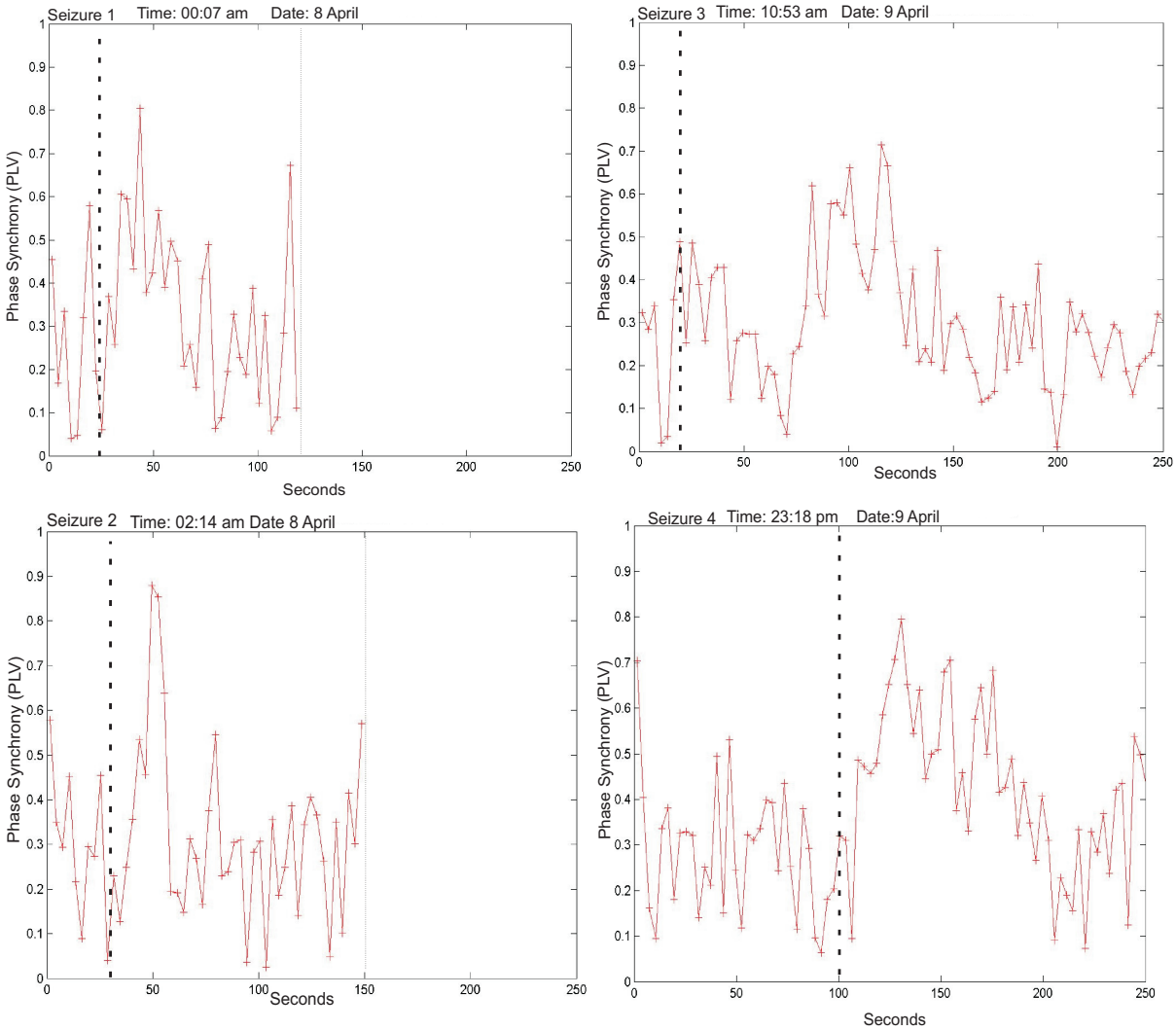


Figure 6.13: **Patient 3: Right Parietal Seizures**: PLV tracked across time for four seizure segments of patient 3. Vertical lines show the seizure onset times. Seizure 1,2: data available is only till that marked by a solid vertical line.

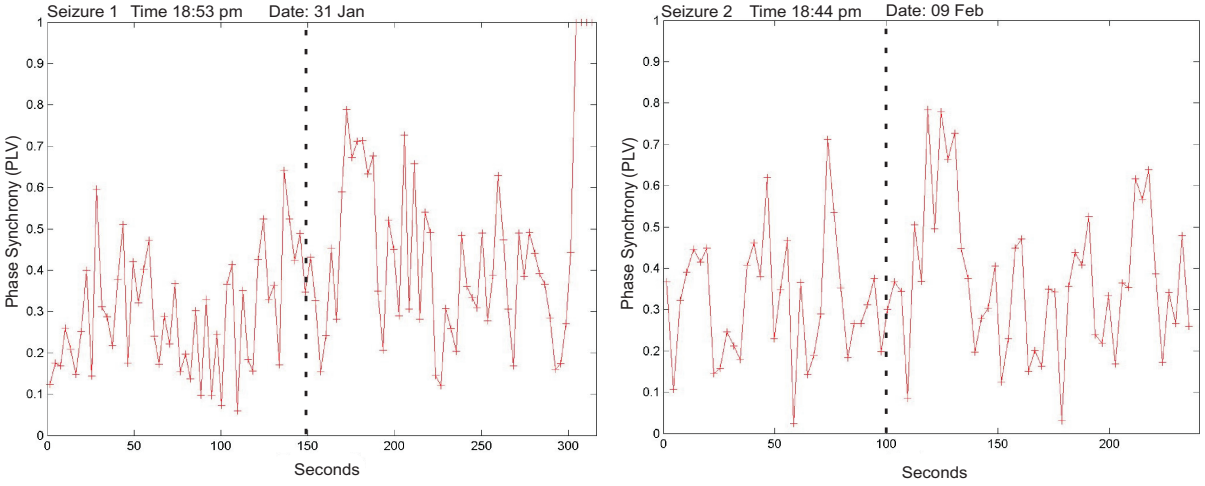


Figure 6.14: **Patient 4: Left Temporal Seizures** :PLS tracked across time for two seizure segments of patient 4. Vertical lines show the seizure onset times.

6.7.2 Tracking Synchrony Across Time with ICA-PLS

In the above procedure, the PLV values have not been checked for significance yet. A statistical method is used here to check the statistical significance of synchrony at each calculation. As the PLV values are changing in level, a single threshold value is not appropriate and hence a test that allows the significance threshold to vary across time is used. Significance testing is performed using 100 phase randomized surrogate series as described in Chapter 4 to obtain the phase locking statistics (PLS). Only the significant PLV (low and high) i.e. the PLS values are retained while the non-significant PLV values are made zero.

In addition, the algorithm is also run on inter-ictal segments of the same patients in order to compare the dynamics of synchrony in the spatial area of interest close-to and far-away from a seizure. Here the same four patient examples as in the previous section are shown in figures 6.15, 6.16, 6.17, 6.18, 6.19, 6.20, 6.21, 6.22, 6.23, 6.24, 6.25. Each patient results shows :

- A segment of the ictal EEG segment used for obtaining the 2 constraints
- The sources estimated using ICA (TDSEP).
- The separability matrix of estimated sources(dark colors showing a higher degree of dependence). Clusters are marked subjectively (marked by dotted lines). The source clusters marked with circles (red, not dotted) are selected as constraints.
- Topographies of the constraints selected for further analysis.
- PLS (statistically significant (low and high) PLV) across time for different ictal segments using the constraints selected above,
- PLS across time for inter-ictal segments using the constraints selected above.

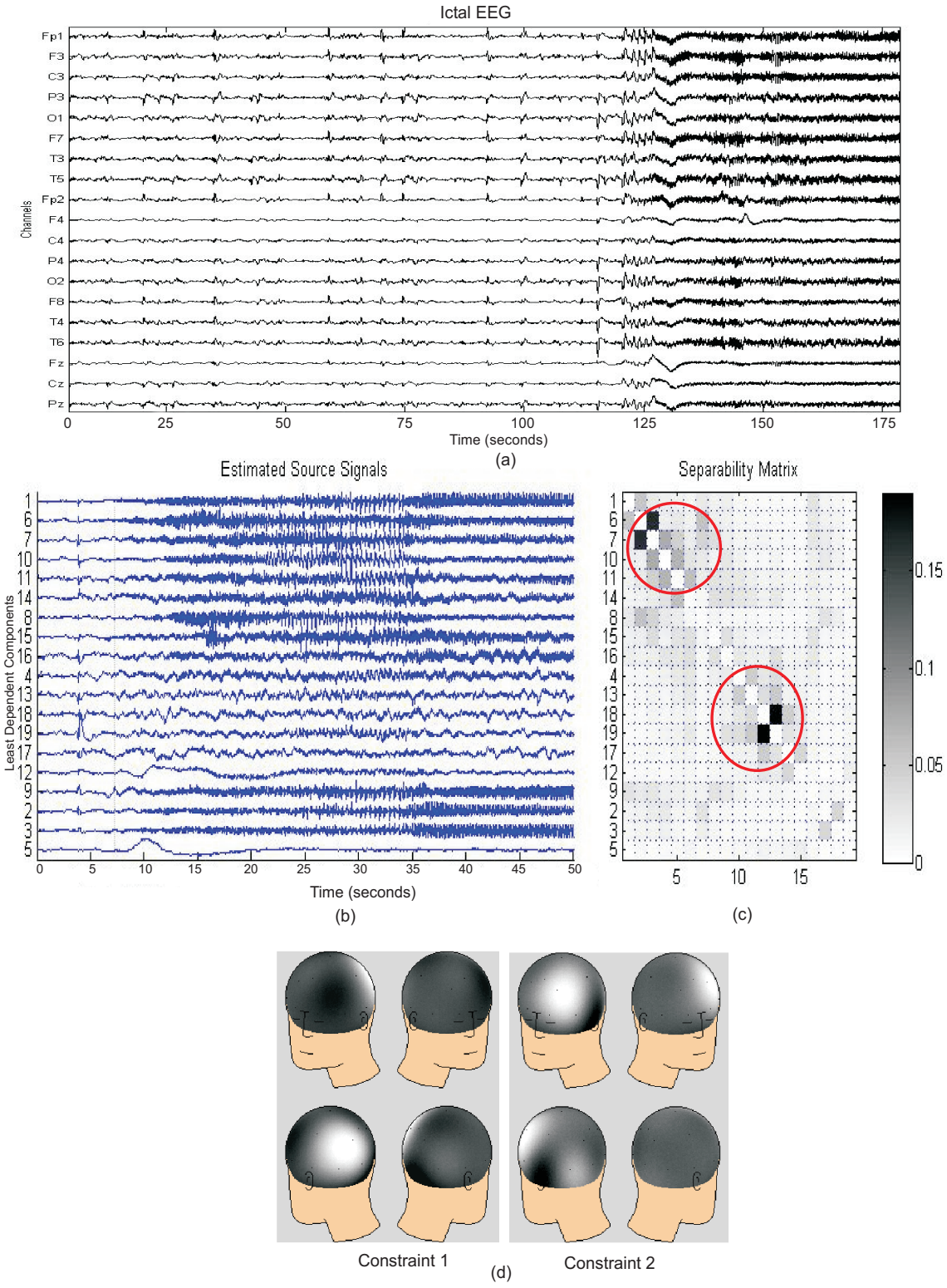


Figure 6.15: **Patient 1: Temporal Seizures**:(a) EEG with seizure (b) A zoomed in plot of the estimated sources (50 seconds), the vertical line indicates seizure onset (c) The separability matrix (d) Topographic plots of constraints selected (The sources in each marked cluster are combined to give one topography and these two combined topographies are used as constraints here)

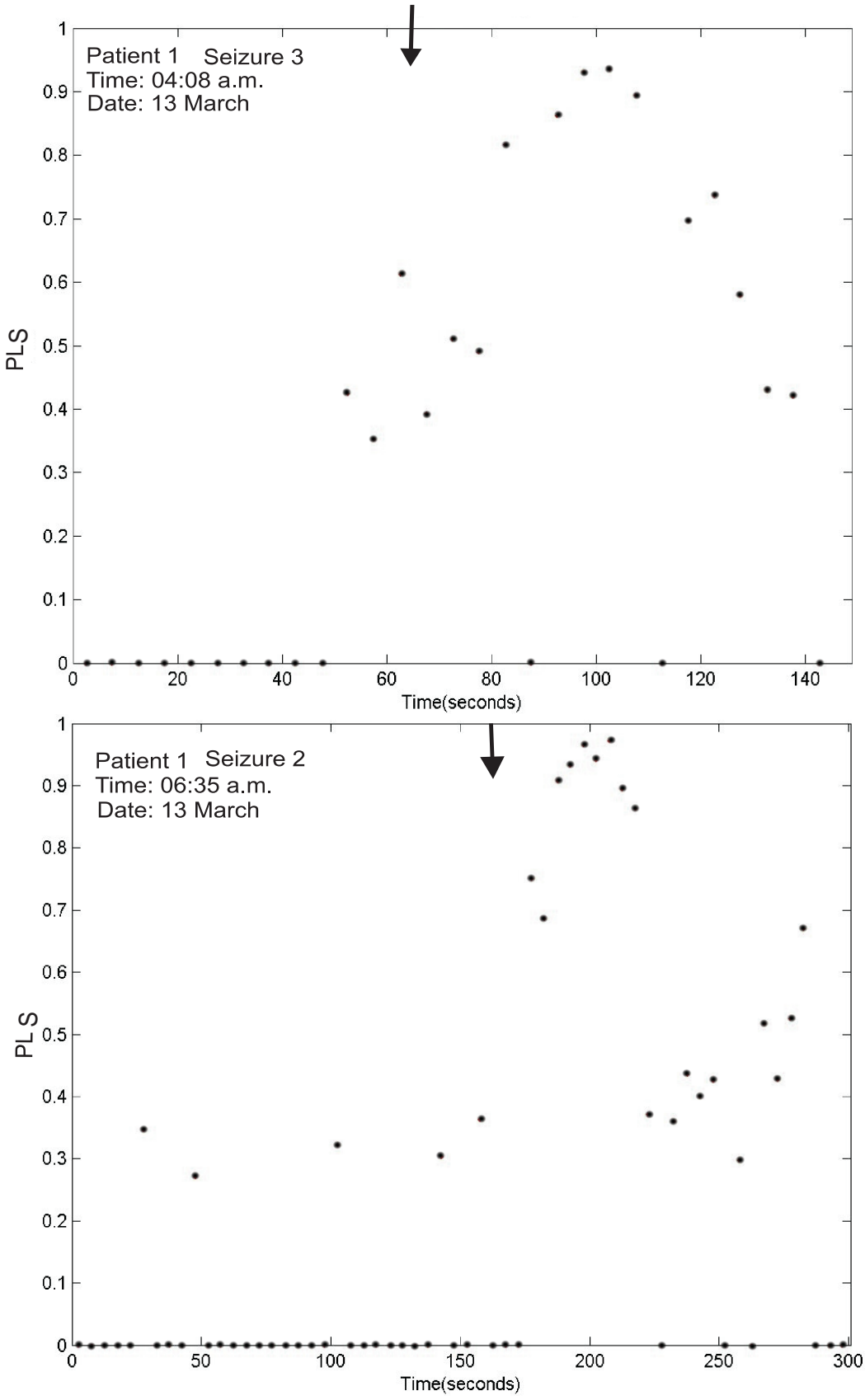


Figure 6.16: **Patient 1: Temporal Seizures**: PLS tracked across time for two seizure segments of patient 1. The arrow indicates the seizure onset.

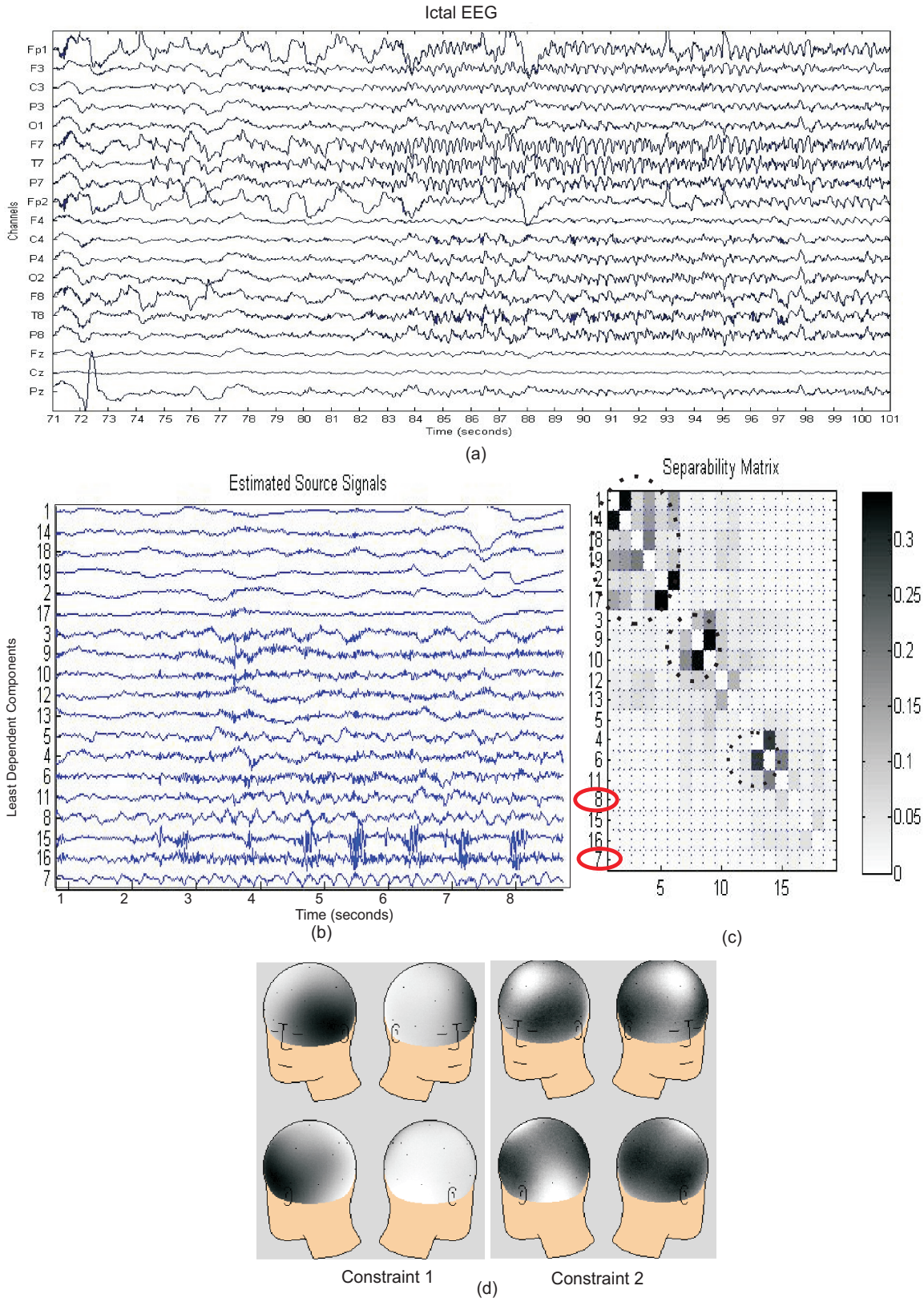


Figure 6.17: **Patient 2: Left Fronto Temporal Seizures**: (a) EEG with seizure, (b) The estimated sources (c) The separability matrix. The separability matrix shows some clusters of dependent sources. These have been marked by dotted circles. The corresponding component signals show either artifacts or they do not correspond to the seizure sources of interest. Hence in this case the sources selected for constraint topographies are independent components marked by red solid circles as they correspond to the seizure sources in the time domain. (d) Topographic plots of constraints selected.

S. No.	Patient	No. of Seizures	No. of Interictal Segments	False Detections	Correct Detections
1	Patient 1	5	20	2	5
2	Patient 2	5	20	2	5
3	Patient 3	4	20	none	4
4	Patient 4	4	20	none	4
5	Patient 5	4	20	4	4
6	Patient 6	4	20	2	4
7	Patient 7	5	20	none	4
8	Patient 8	5	20	none	4
9	Patient 9	5	20	1	4
10	Patient 10	4	20	2	4

Table 6.3: Statistics for seizure onset detection with ICA-PLV as applied to short term data set of 10 patients. The data has been described in Table 6.2.

Two seizure segments of each of the patients 5-10 have been shown in Figures 6.26 and 6.27.

The statistics of seizure onset detections are given below.

Statistics: For Statistical validation, 10 patients' data were used, with 4-5 seizures for each, occurring at different times of the day (morning, afternoon, evening, night), on different days. The inter-ictal data used was 20 segments of EEG of each patient covering most of the day (different times of the day).

Using a PLS threshold in the range of 0.7 - 0.8 (adapted to each patient), 42 out of 45 seizures in total from all patients are detected correctly within a time window of 10 seconds from the seizure onset. The number of false detections in total were 13. The sensitivity and specificity of seizure onset detection is computed as

$$\text{Sensitivity } (\%) = \frac{\text{Number of correctly detected seizure onsets}}{\text{Total number of seizure onsets}} \times 100$$

$$\text{Specificity } (\%) = \frac{\text{Number of correctly detected seizure onsets}}{\text{Total number of detections}} \times 100$$

Sensitivity of the seizure onset has been found to be 93% and the specificity as 76%.

The above analysis uses ICA (temporal decorrelation)- phase synchrony and spatially constrained ICA for seizure detection. The results show that this scheme is quite efficient but it still needs further exploration with regards to the questions as to how and why these areas synchronize so dramatically at seizure onset and if these areas are associated with any dynamics before seizure onset, if they are to be useful for seizure onset prediction.

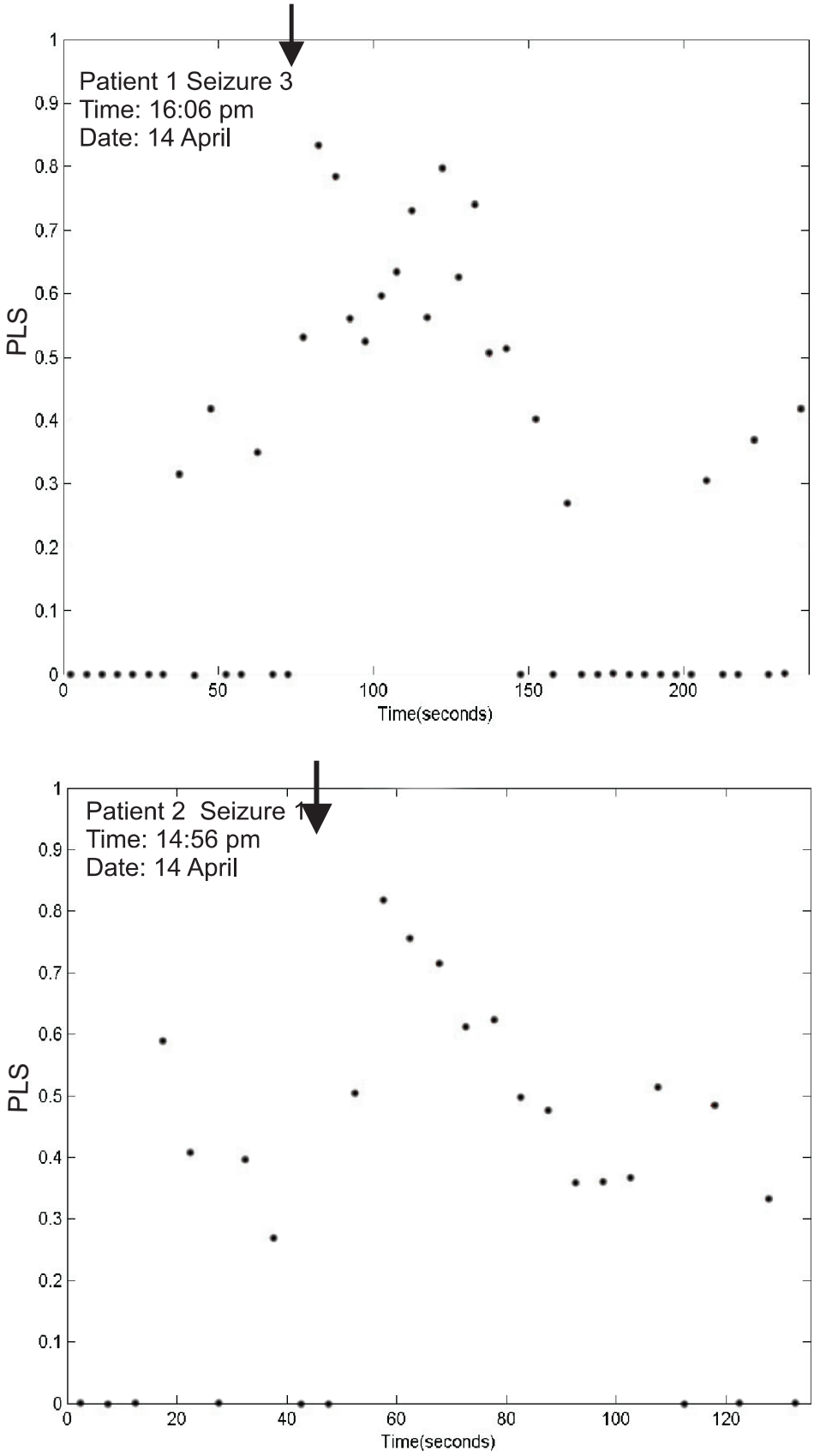


Figure 6.18: **Patient 2: Left Fronto Temporal Seizures**: PLS tracked across time for two seizure segments of patient 2. The crosses show significant low synchrony.

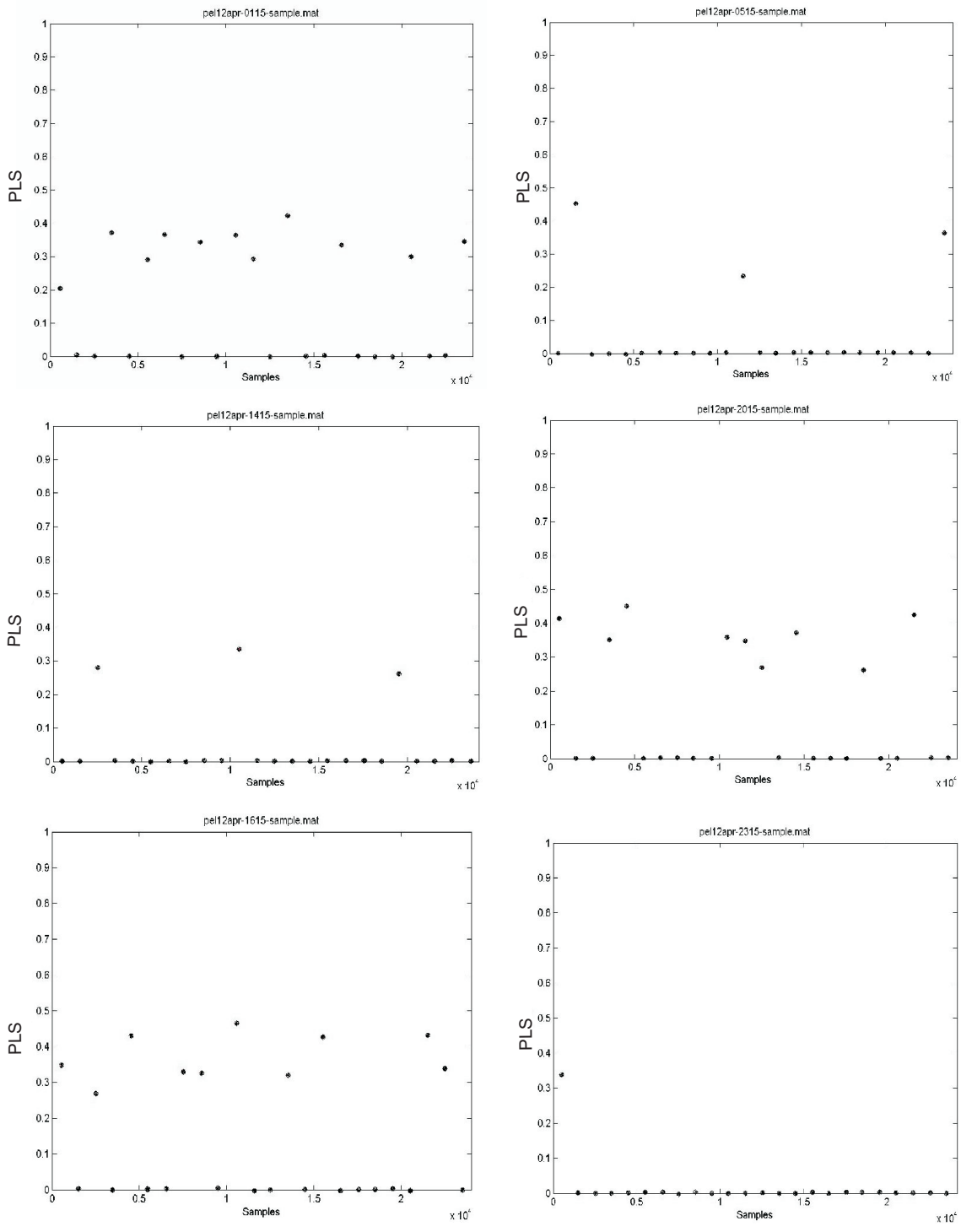


Figure 6.19: **Patient 2: Left Fronto Temporal Seizures**: PLS tracked across time for six inter-ictal segments from different times of the day for patient 2. The crosses show significant low synchrony.

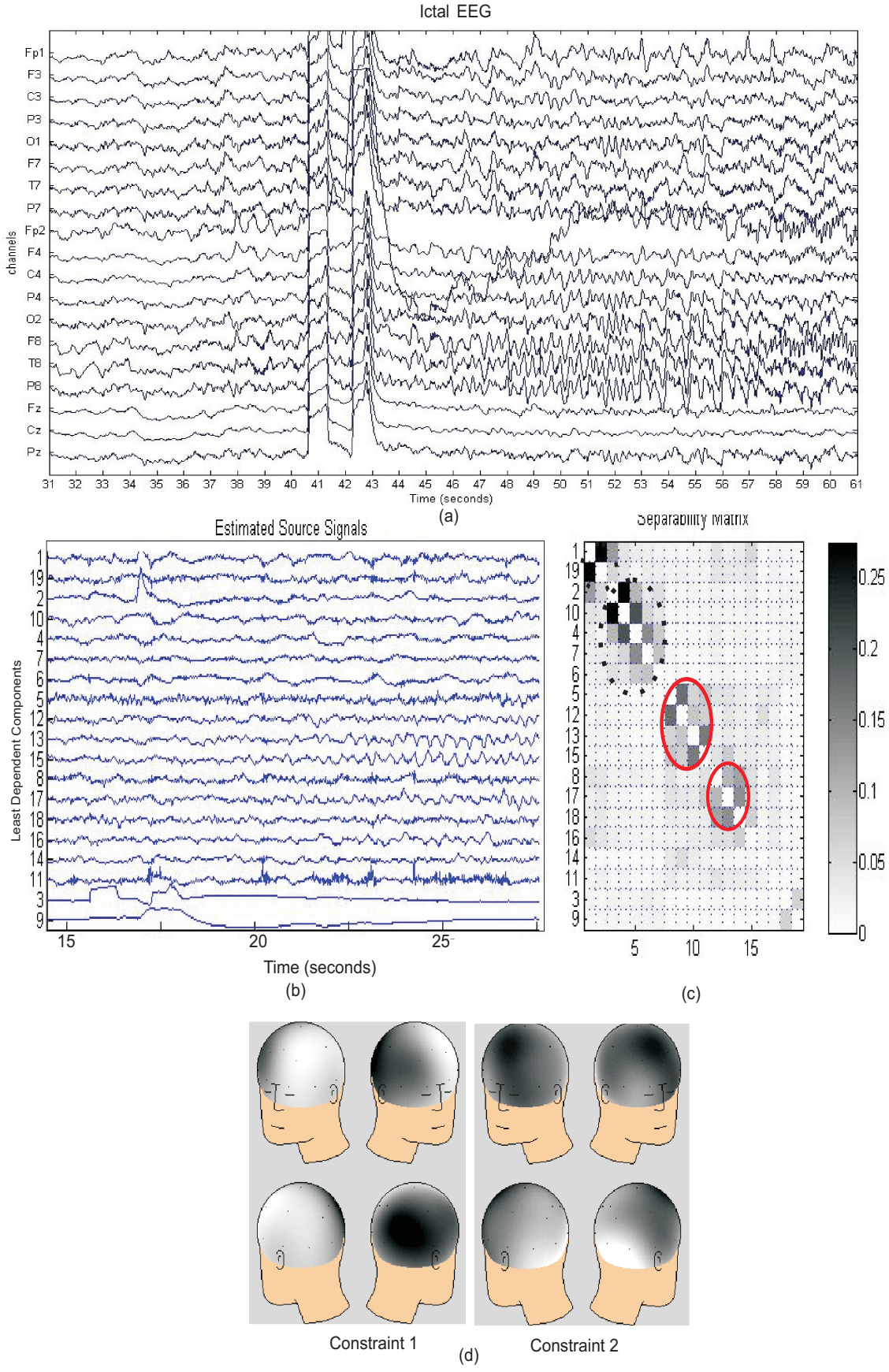


Figure 6.20: **Patient 3: Right Parietal Seizures**:(a) EEG with seizure, (b) The estimated sources (c) The separability matrix (d) Topographic plots of constraints selected

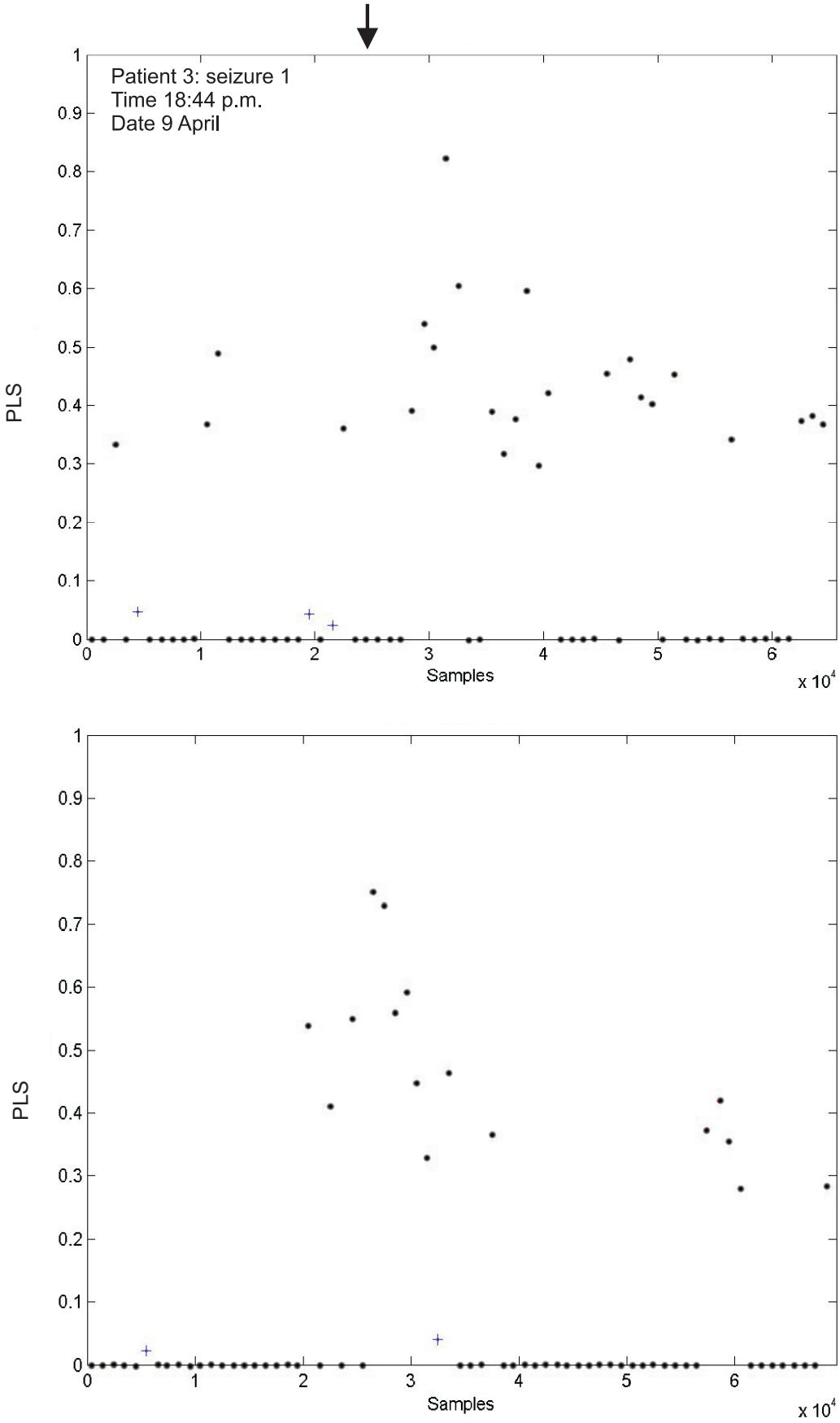


Figure 6.21: **Patient 3: Right Parietal Seizures**: PLS tracked across time for two seizure segments of patient 3. The crosses show significant low synchrony.

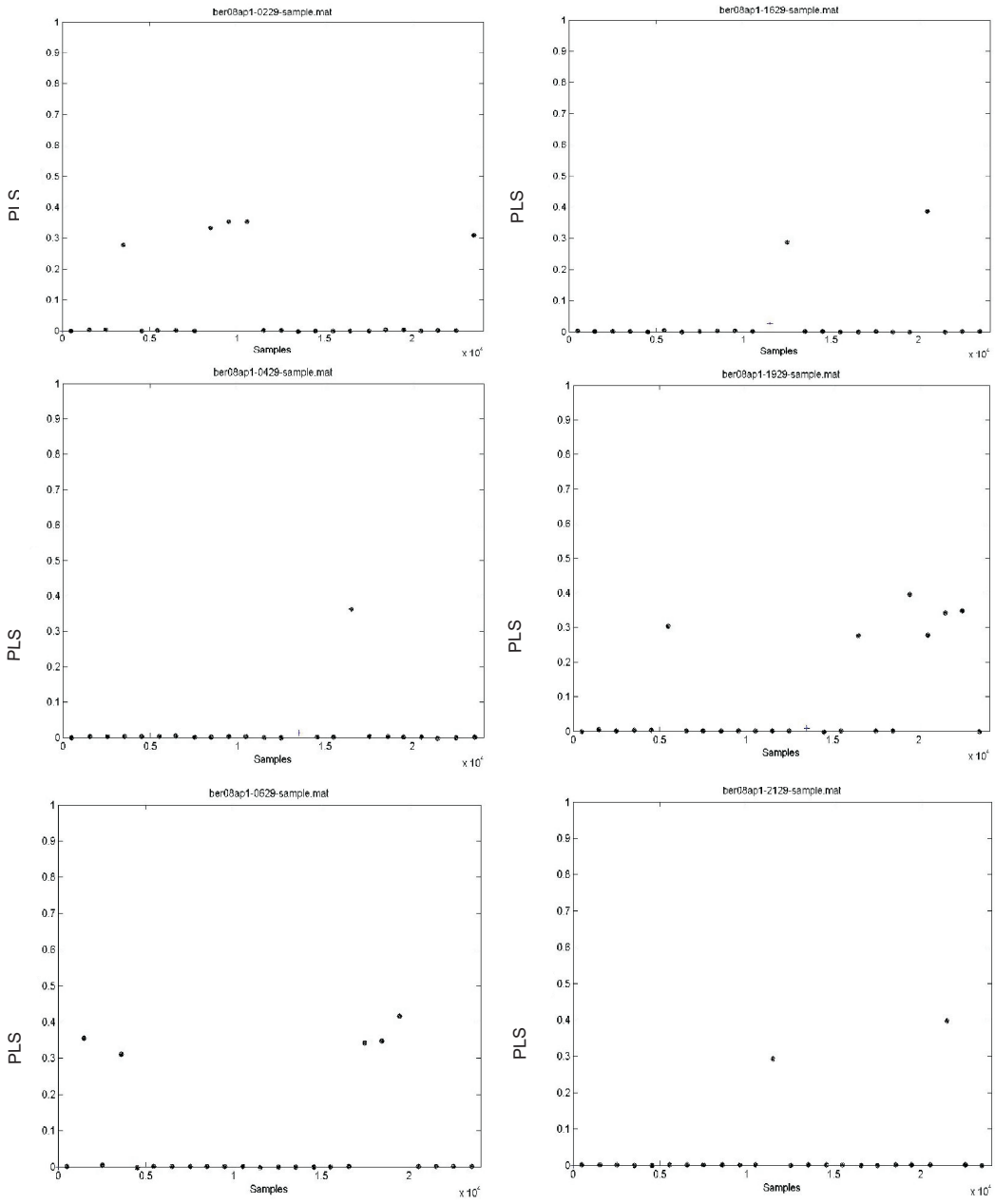


Figure 6.22: **Patient 3: Right Parietal Seizures**: PLS tracked across time for six interictal segments from different times of the day for patient 3. The crosses show significant low synchrony.

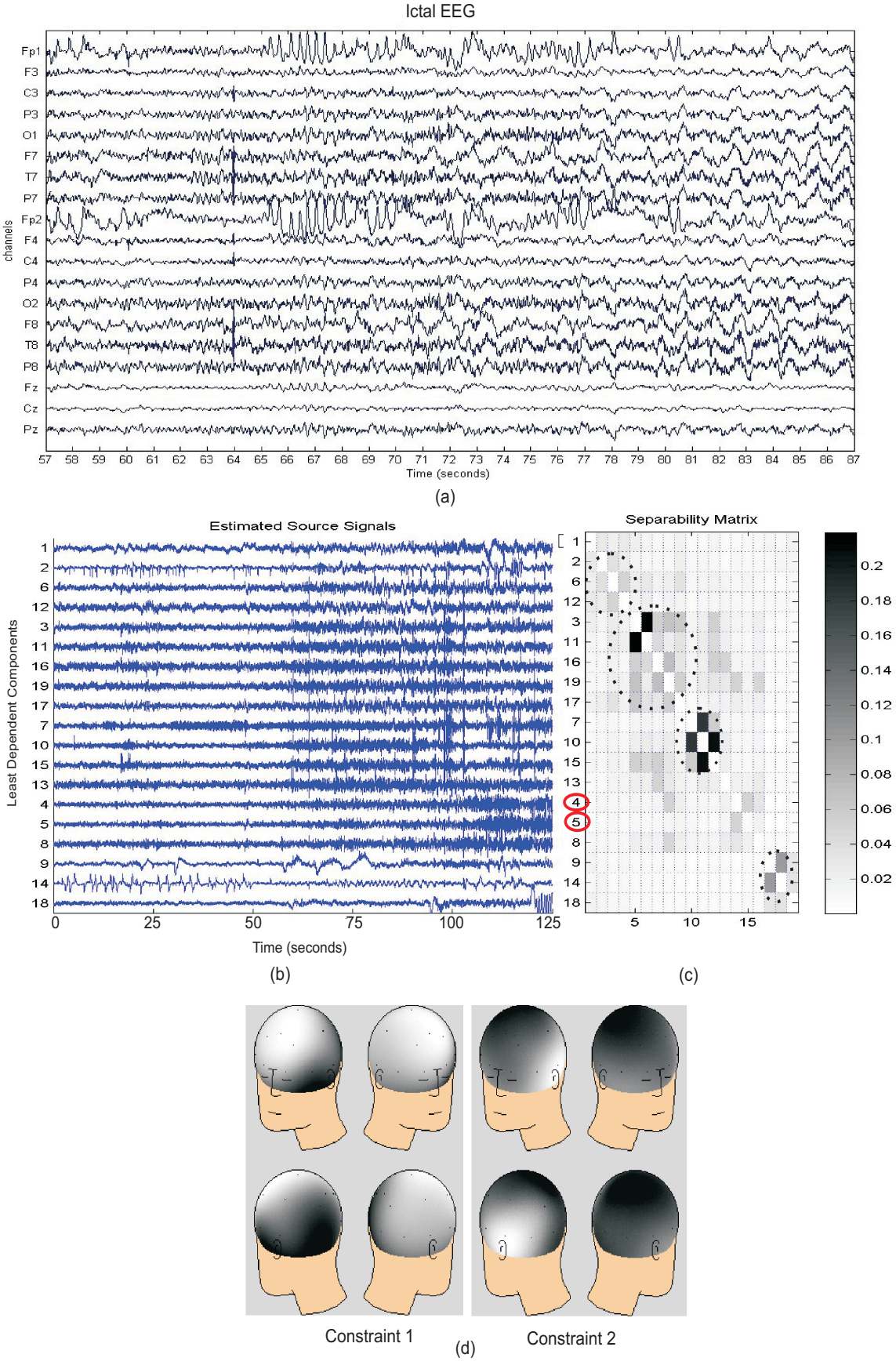


Figure 6.23: **Patient 4: Left Temporal Seizures**: (a) EEG with seizure, (b) The estimated sources (c) The separability matrix (d) Topographic plots of constraints selected (constraints selected here are IC topographies 4 and 5)

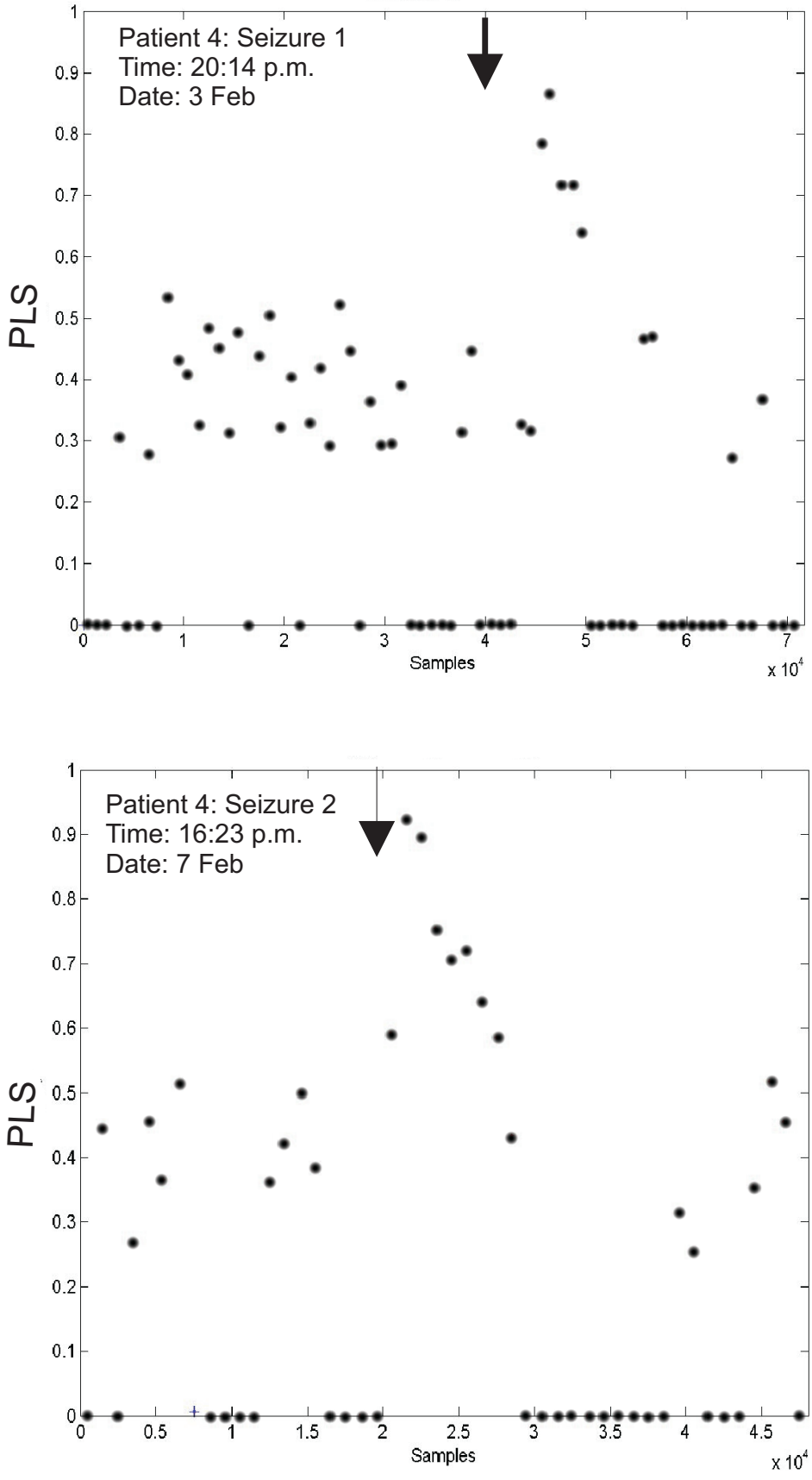


Figure 6.24: **Patient 4: Left Temporal Seizures**: PLS tracked across time for two seizure segments of patient 4. The crosses show significant low synchrony.

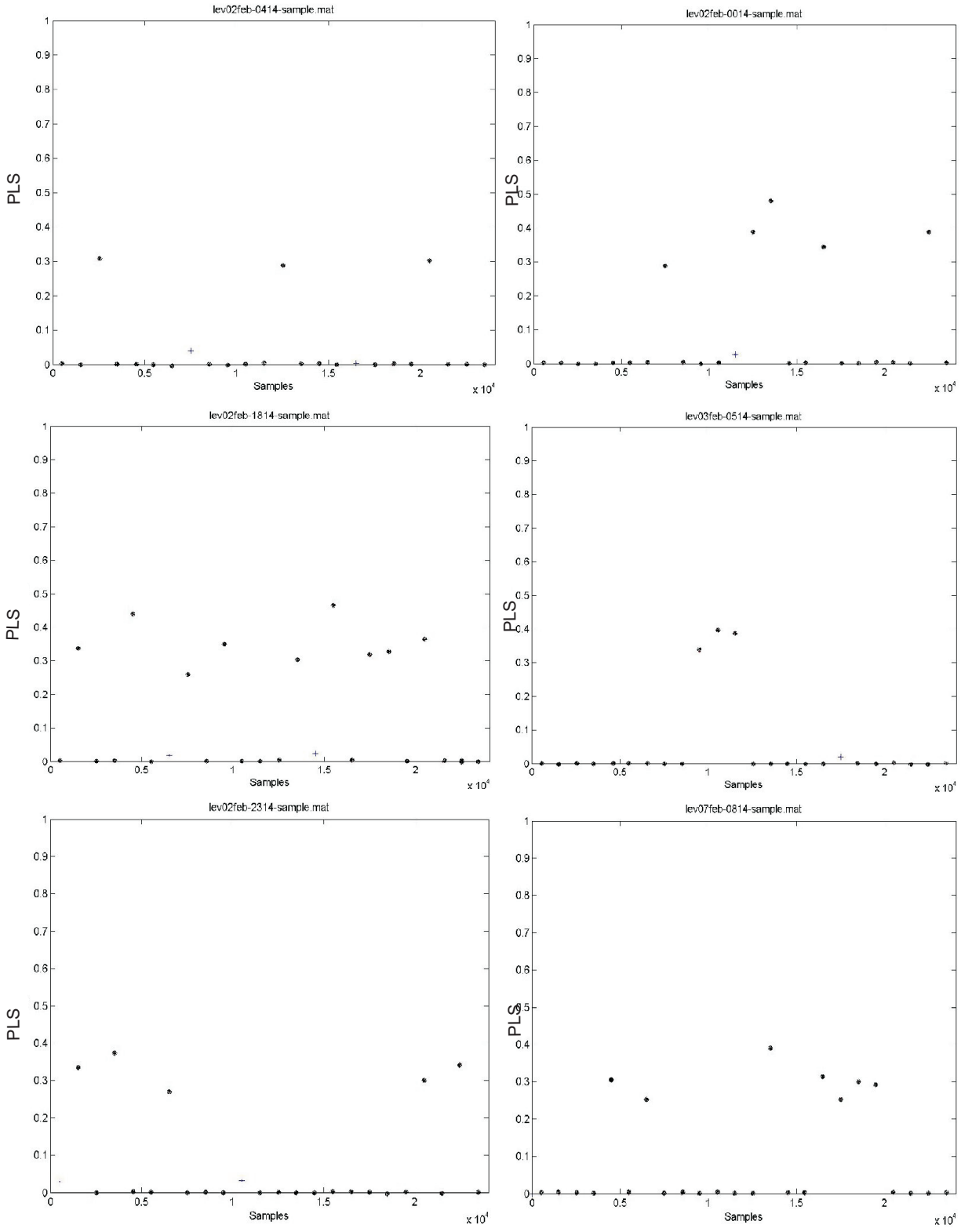


Figure 6.25: **Patient 4: Left Temporal Seizures**: PLS tracked across time for six interictal segments from different times of the day for patient 4. The crosses show significant low synchrony.

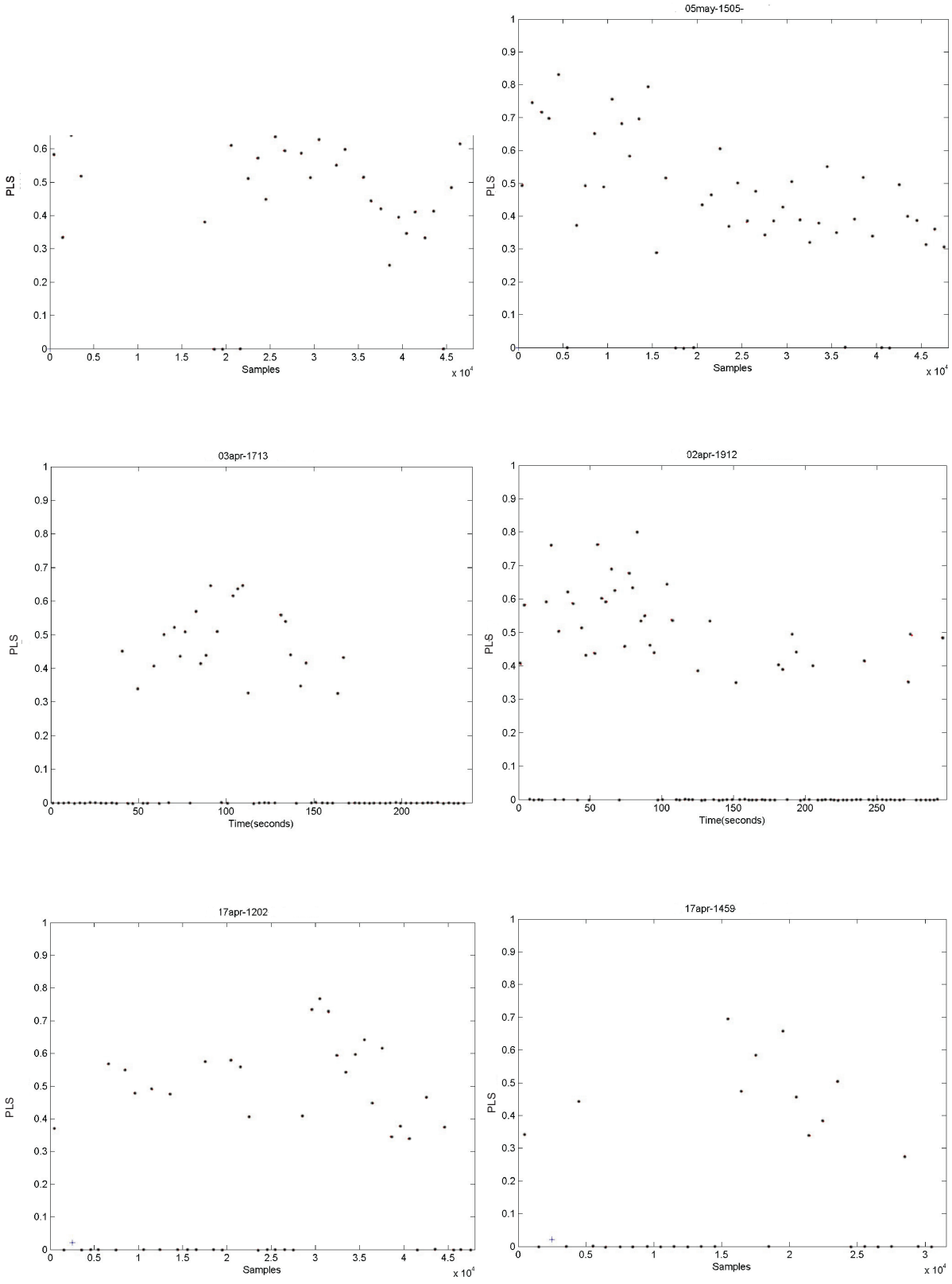


Figure 6.26: **Patient 5-7: Seizure segments**: PLS tracked across time for two seizure segments from different times of the day for patient 5,6 and 7. Seizure onset times have been marked by vertical arrows. The crosses show significant low synchrony.

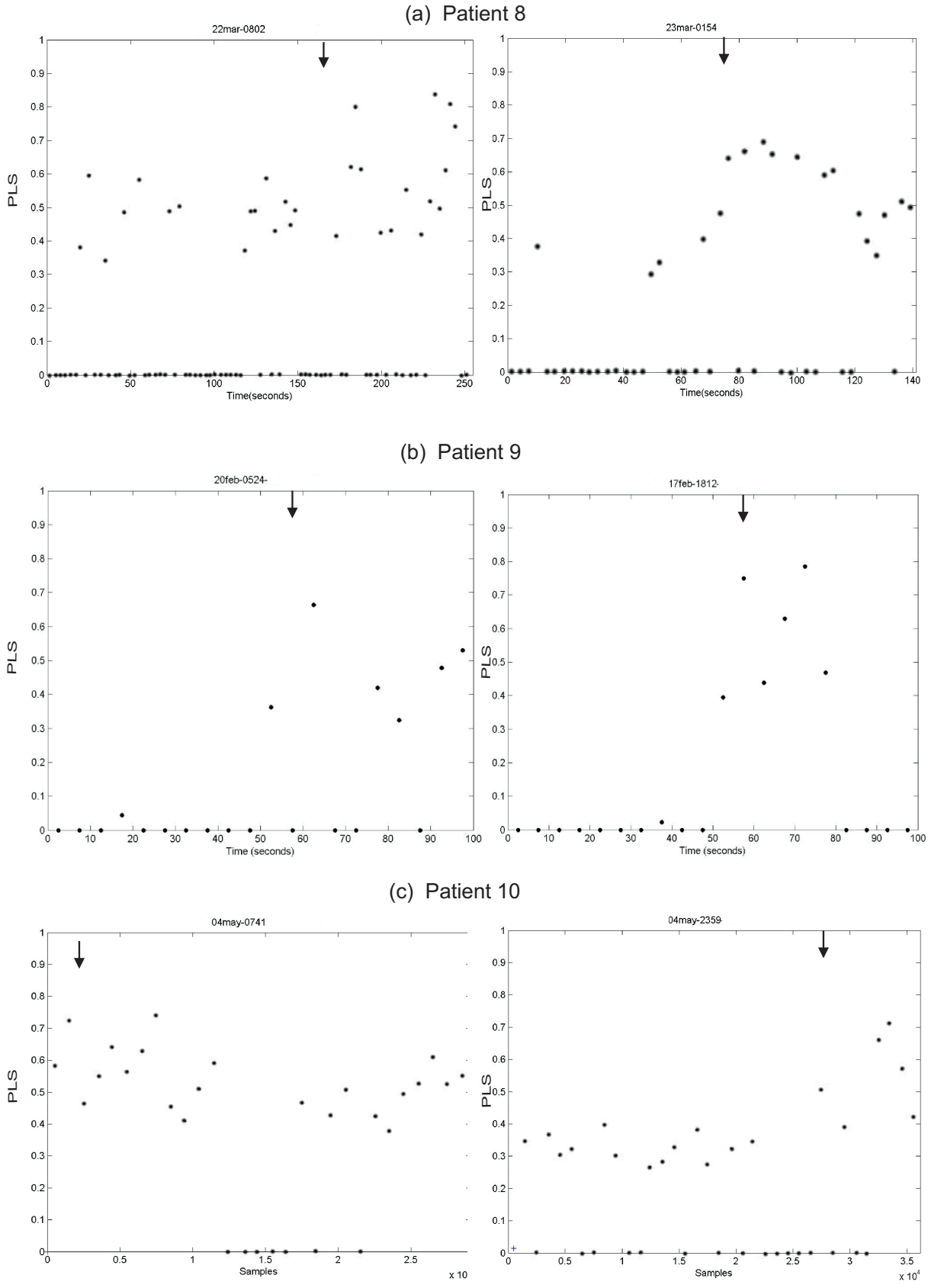


Figure 6.27: **Patient 8-10: Seizure segments**: PLS tracked across time for two seizure segments from different times of the day for patient 8,9 and 10. Seizure onset times have been marked by vertical arrows. The crosses show significant low synchrony.

6.7.3 ICA-PLV with Narrowband vs Broadband Signals for Seizure Detection

There have been doubts expressed in the literature about using broadband data vs narrowband data for phase synchrony analysis. It has been seen in previous studies that usually close examination of the bandwidth of the epileptic EEG signals is ignored. This comparison was also highlighted by Netoff and Schiff [92] but was not conclusive. Ideally, based on the understanding of the basic principles of phase synchrony, especially with the use of the analytical signal computed via the Hilbert Transform, broadband signals would give meaningless results. The theory about the phase synchrony being applied to chaotic signals also assumes the use of narrowband signals. Too narrow a bandwidth can affect the true synchrony as well as it would cause an overestimation of the synchrony, shown by [12, 168, 169, 170]. EEG signals are broadband signals and their LDCs are also broadband (as ICA is used here in a way that decorrelates in space and time rather than in frequency). Thus, before proceeding, in order to clarify the doubt about the need to filter signals prior to phase synchrony analysis, an assessment was carried out in the context of seizure onset detection/prediction.

The multichannel ictal EEG segments of 2 patients (with 3 seizures each at different times of the day and on different days), two to five minutes in length each were used. The EEG was mean-corrected and then temporally decorrelated using TDSEP to generate LDCs of the multivariate EEG signals. The EEG signals and the LDCs were filtered in different frequency bands (1-20 Hz, 20-40 Hz, 40-60 Hz and 60-80 Hz) using a FIR filter. The time-frequency evolution of phase synchronization was investigated on the multivariate EEG signals and the corresponding LDCs by calculating the PLV for each frequency band across time for all combinations of EEG electrodes and LDCs. The PLV was calculated using non-overlapping moving windows of 600 samples each. It was then compared with the PLV obtained from the use of broadband LDCs (with no filtering) and with a narrower bandwidth of 2-8 Hz as was used in previous cases, being relevant to the context of epileptic activity.

General Observations:

1. It was observed that there was inconsistency between the amount of phase synchronization seen at different frequency bands between the EEG electrodes and the corresponding LDCs (to compare combinations of LDCs and EEG electrode signals, spatial information was used, i.e. LDCs having the topographies similar to that of the position of EEG electrodes). An example is shown in Figure 6.28, where it can be seen that the phase synchronization between EEG electrodes F3 and T3 for an EEG segment shows high phase synchrony in low frequency bands and the amount of phase synchrony in high frequency bands is comparatively lower. But the corresponding combination of LDCs (selected by their similar-

ity of topography to the positions of F3 and T3 electrodes) shows high phase synchrony in low frequency bands as well as high phase synchrony in high frequency bands. This discrepancy may be attributed to the overlapping of signals in EEG channels which is reduced/removed in LDCs. This observation further strengthens the concept that ICA de-mixes the scalp recorded data bringing the analysis a step closer to the underlying sources and removing the effect of spatial overlapping of signals.

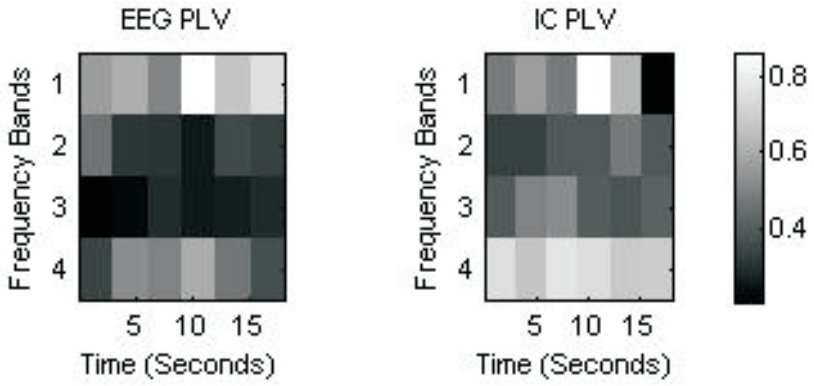


Figure 6.28: Plots of Time Frequency evolution of EEG (F3, T3) and the corresponding ICs (matched to the location of EEG electrodes by their corresponding topographies

2. The time-frequency evolution of the phase synchronization for the estimated LDCs alone was assessed for comparing the PLV obtained from broadband vs narrowband LDC signals. It was observed that some of the narrowband phase synchrony results do not match the broadband phase synchrony results. One example is shown in Figure 6.29 where the upper plot shows the PLV obtained from narrowband LDC signals and the lower plot is found by computing PLV from broadband LDC signals.

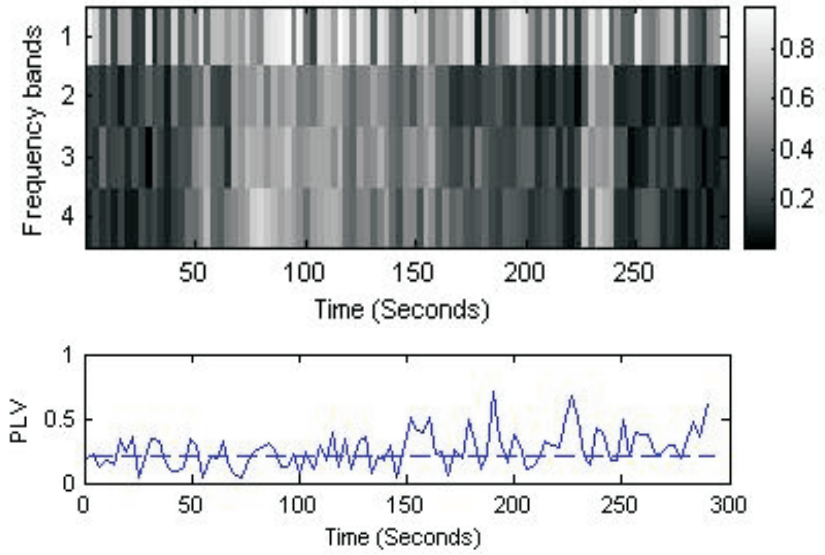


Figure 6.29: Theoretically broadband analysis of PLV can lead to misinterpretations and such differences are shown with the help of PLV analysis of an ictal segment. The synchrony of the ictal segment is expected to be high. A plot of time-frequency evolution of phase synchrony for spatially constrained LDCs of ictal data (upper) and that for the broad band data (lower) is shown. The horizontal line on lower plot shows the PLV significance threshold. Differences in the estimated PLV can be observed in the narrowband and broadband analysis. For example: in the time window of 50-100s, the PLV on broadband data shows a non significant value (below significance threshold) while the PLV in the frequency bands 1 and 4 of the narrowband analysis data shows a significantly high values, with maxima of 0.9 and 0.7 respectively.

The above analysis is then applied to a patient's ictal broadband and narrowband data in the context of seizure detection. The template used for estimating the spatial constraints and the selection of constraint topographies is as mentioned in the previous sections. The narrowband and broadband PLV for the LDCs was estimated for 2 patients (patient 1 and patient 7 from Table 6.2). The seizure segments are at various times of the day and from various days per patient. This variation was considered in order to assess temporal robustness of the algorithm. The signals were filtered in three frequency bands of 1-20 Hz, 20-40 Hz and 40-60 Hz. The band 2-8 Hz is also used as it is the most relevant bandwidth for epileptic seizure analysis.

The topographies of the selected sources were then used as spatial constraints for estimating two spatially constrained sources on the patient's inter-ictal and ictal data. In Figures 6.30, 6.31, 6.32 and 6.33, 6.34, 6.35, the phase synchronization on broadband and various narrow frequency bands in ictal data are compared. In general, it is observed that the narrowband PLV remains usually at a stable low value of PLV at times away from the seizure onset, while the broadband PLV for the same data is either at a higher level and sometimes becomes highly unreliable. The narrowband phase synchrony is seen to reduce to quite low levels prior to the seizure onset and thereafter increase to higher levels at seizure onset. The effect can be seen across segments taken from different times of the day. The most significant observation is the quite consistent (all seizures, both patients) demarcated increase of phase synchrony observed with seizure onset in the 2-8 Hz band. This shows much potential of the technique for a *seizure detection*. Such a demarcated change is not generally observed with PLV of broadband LDCs or even with all narrowband LDCs, when evolving from inter-ictal to ictal periods.

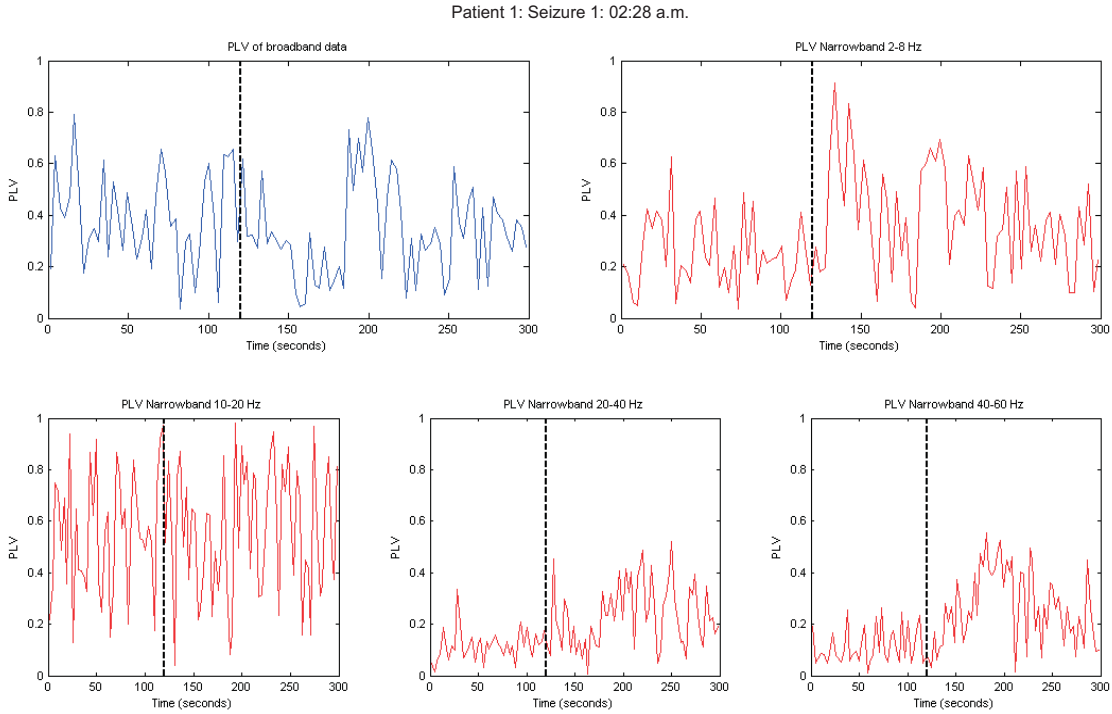


Figure 6.30: Patient 1: Seizure 1: 2:28 a.m. PLV of broadband data (upper left plot), narrowband 2-8 Hz (upper right plot), narrowbands 1-20 Hz, 20-40 Hz, and 40-60 Hz (lower plots)

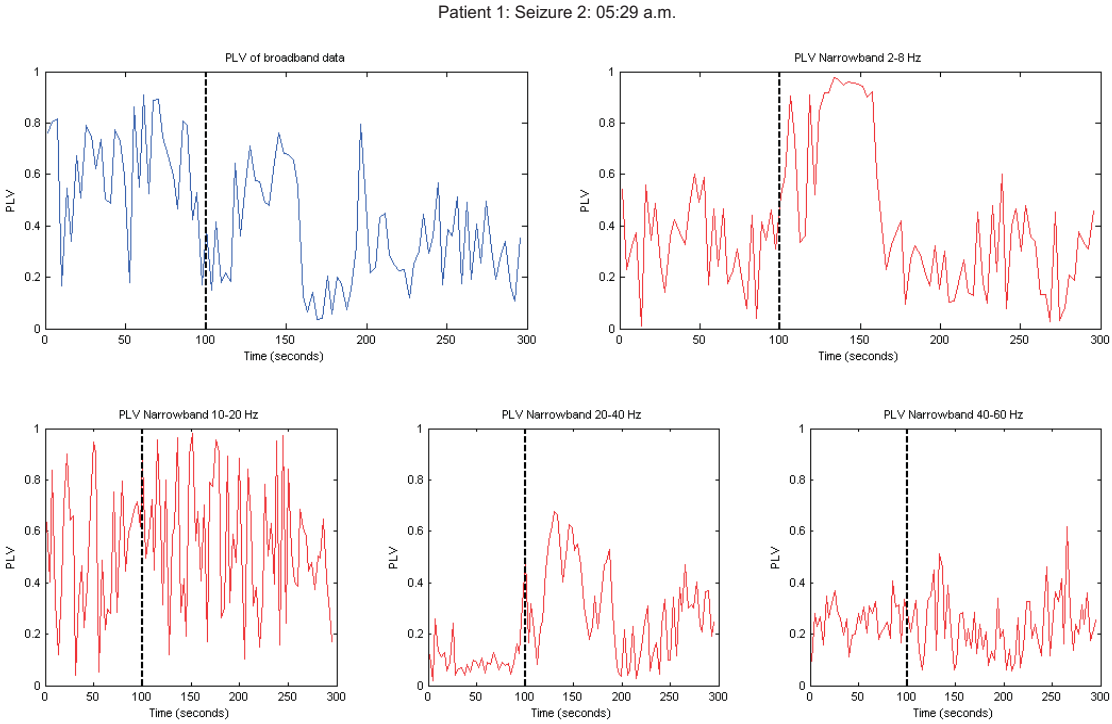


Figure 6.31: Patient 1: Seizure 2: 5:2 a.m. PLV of broadband data (upper left plot), narrowband 2-8 Hz (upper right plot), narrowbands 1-20 Hz, 20-40 Hz, and 40-60 Hz (lower plots)

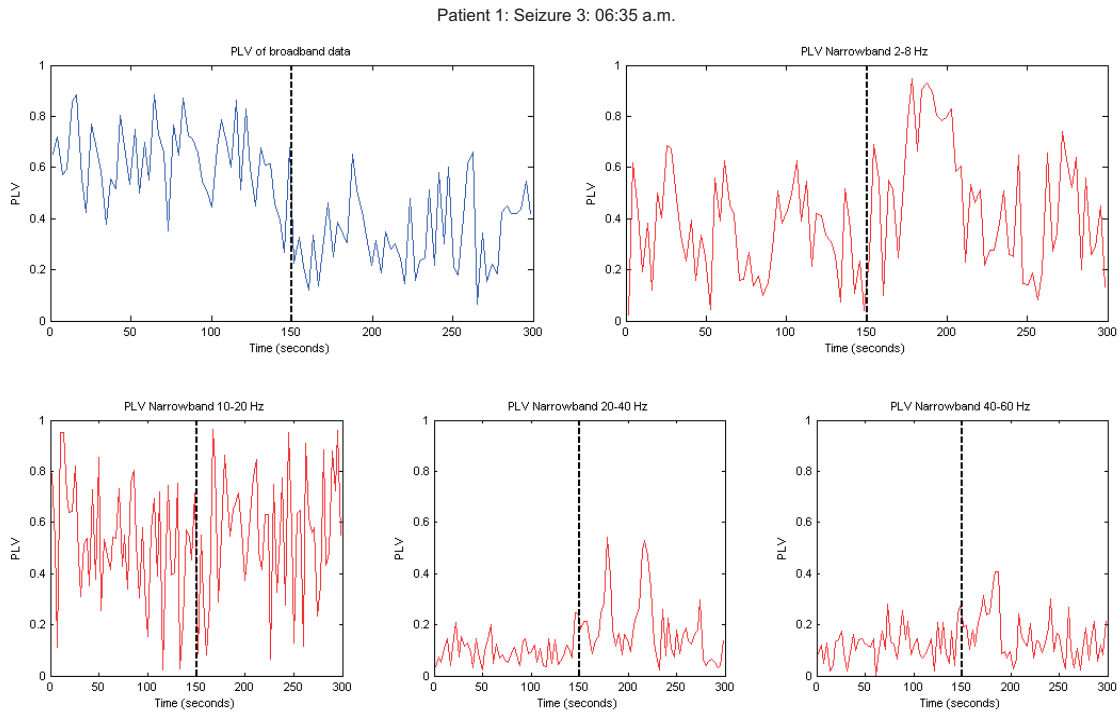


Figure 6.32: Patient 1: Seizure 3: 6:35 a.m. PLV of broadband data (upper left plot), narrowband 2-8 Hz (upper right plot), narrowbands 1-20 Hz, 20-40 Hz, and 40-60 Hz (lower plots)

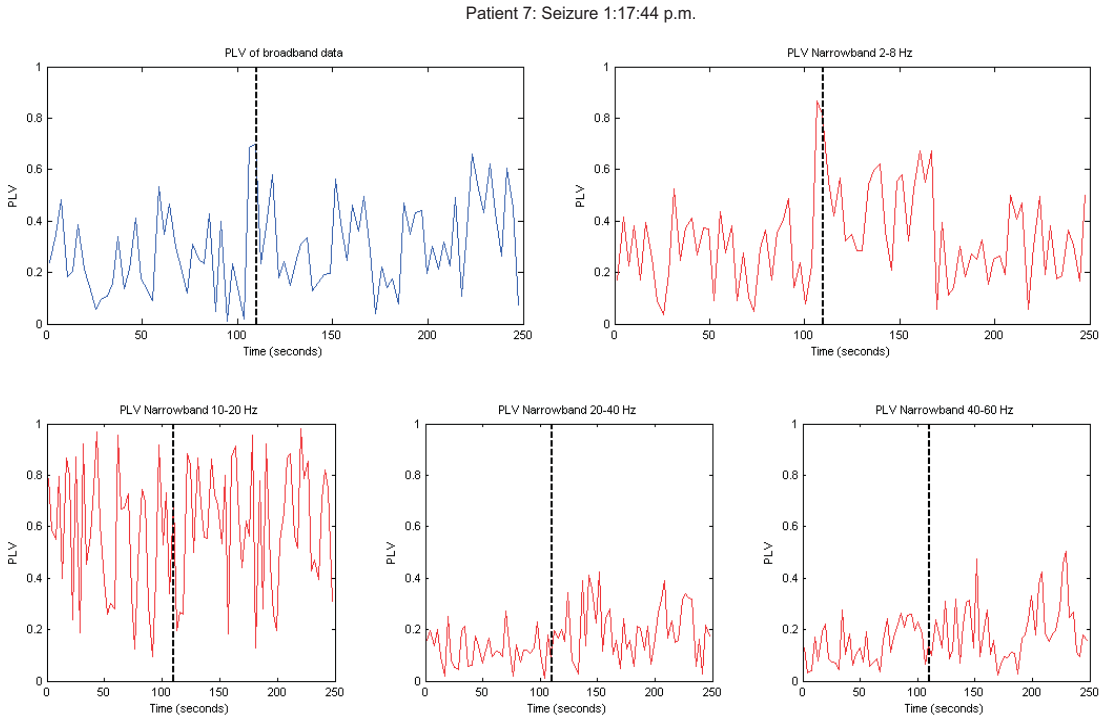


Figure 6.33: Patient 7: Seizure 1: 17:44 p.m. PLV of broadband data (upper left plot), narrowband 2-8 Hz (upper right plot), narrowbands 1-20 Hz, 20-40 Hz, and 40-60 Hz (lower plots)

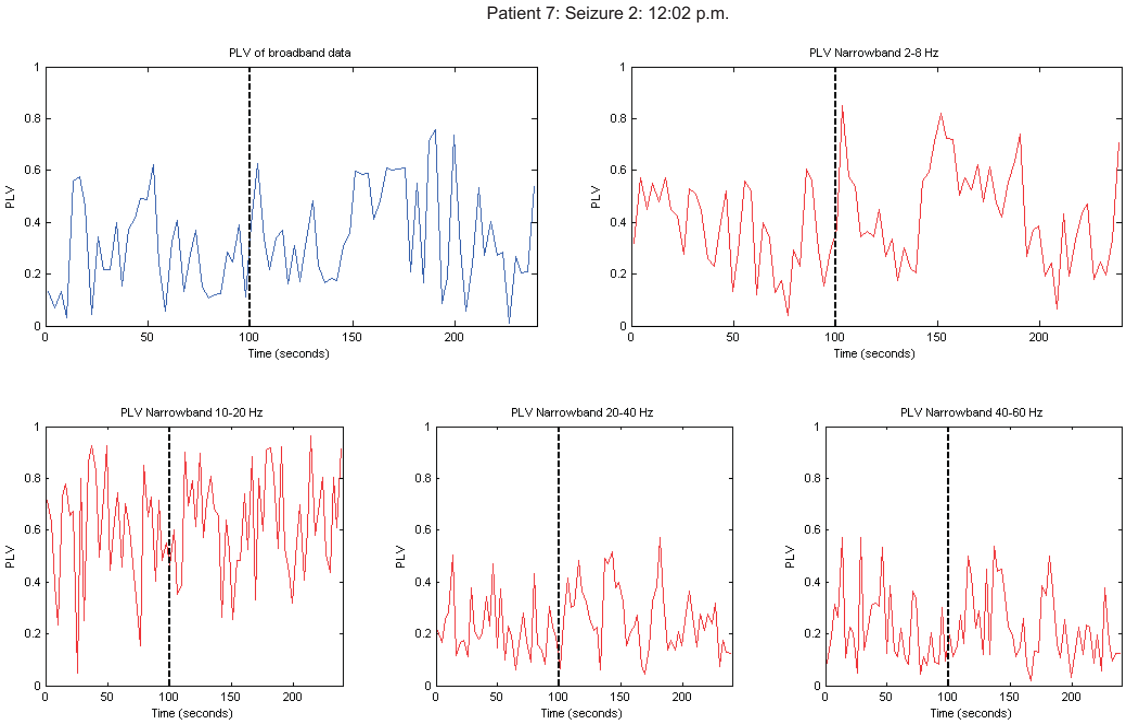


Figure 6.34: Patient 7: Seizure 2: 12:02 p.m. PLV of broadband data (upper left plot), narrowband 2-8 Hz (upper right plot), narrowbands 1-20 Hz, 20-40 Hz, and 40-60 Hz (lower plots)

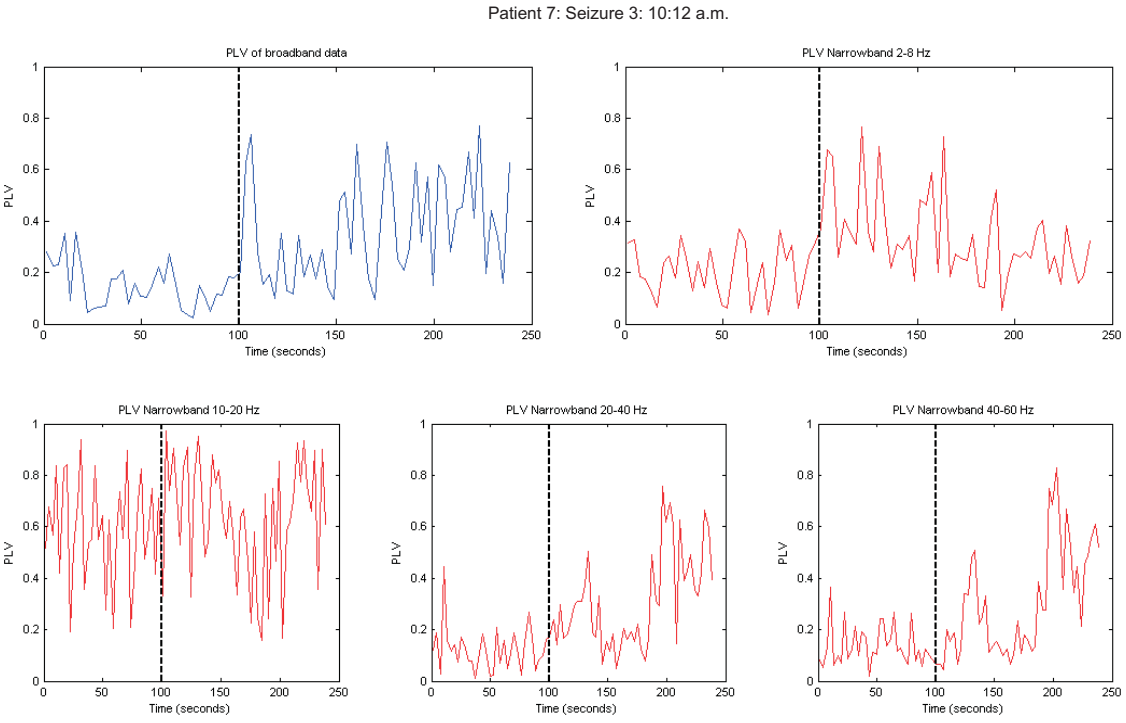


Figure 6.35: Patient 7: Seizure 3: 10:12 a.m. PLV of broadband data (upper left plot), narrowband 2-8 Hz (upper right plot), narrowbands 1-20 Hz, 20-40 Hz, and 40-60 Hz (lower plots)

In this analysis, a discrepancy was seen between the phase synchrony results for broadband and narrowband spatially constrained LDCs in case of ictal data. Such discrepancies can lead to incorrect interpretations in certain applications like seizure onset detection, when relying on broadband phase synchrony analysis; hence it is concluded that the use of phase synchronization analysis on broadband signals should be performed with great caution. Henceforth only narrowband signals are used for phase synchrony analysis in this thesis. Also as the bandwidth of 2-8 Hz appears to be the most significantly useful results in the context of seizure detection, only 2-8 Hz will be used in this thesis herein for epileptic EEG analysis.

6.8 Seizure Onset Prediction

The analysis shown in this chapter using a combination of ICA and phase synchronization to seizure onset *detection* motivates the application of the technique for seizure onset *prediction*. The observation of the existence of a prominent trend of decreasing and increasing phase synchronization at seizure onset, between LDCs found near the epileptogenic focal area might be helpful for seizure prediction. Confirmation about the change in this frequency band synchrony prior to a seizure onset, (that may help to predict a seizure) will require prospective statistical evidence, which was currently not possible due to lack of continuous scalp EEG recordings. The changes in narrow band synchrony prior to a seizure (including the effect of change in behavior or the sleep and wake state on the phase synchrony levels prior to the seizure onset) will be explored further, in the context of seizure prediction.

6.9 Summary

This chapter described the use of ICA and phase synchrony in combination, referred to as ICA-PLV and explained how they complement each other. ICA is useful for unmixing and extracting underlying source signals, but the selection of sources of interest from the set of IC's (or rather LDCs) isolated by ICA remains challenging in research applications. The ICA-PLV was found useful for objective 'seizure source' selection from the disordered set of source signals extracted by ICA from ictal segments. This technique was then used for objective de-noising of the ictal EEG segments, retaining only the information about the seizure. In order to apply the same for seizure prediction analysis, segments prior to a seizure need to be analyzed, which is not trivial as the seizure sources, changes across time as identifying seizure sources without a morphological, spectral, or other visually apparent feature is not trivial. This challenge was overcome by the use of feedback from the ICA-PLV applied to an ictal segment, for application to other segments of the same patients. This algorithm

has been described in this chapter. This technique was used on short EEG segments (due to lack of longer continuous EEG at the time) it was found useful for seizure onset detection.

Chapter 7

Seizure Detection and Prediction using ICA and Phase Synchrony: Long Term EEG analysis

The analysis undertaken so far in previous chapters motivates the use of phase synchrony analysis in combination with ICA for epileptic seizure detection and prediction. The change in phase synchrony observed in spatially constrained LDCs, shows that phase synchrony increases in certain spatial areas in epileptic patients at seizure onset. These spatial areas most likely are specific areas to each patient dependent on the type of seizures (temporal, frontal, parietal). Previous analysis so far has been performed only on short two minute segments of ictal and inter-ictal EEG data. The short ictal segments have shown a sudden increase in phase synchrony at seizure onset and the inter-ictal segments show a low level and stable or no significant phase synchrony for the two spatially constrained source signals.

The next step of this analysis is to check if these patterns, observed in the selected spatial areas, are specific to seizure onset or are sensitive to other changes in the state of consciousness or brain activity as well. This is performed by obtaining the spatial constraints for a given patient and then tracking the phase synchrony across time (through moving windows) using spatially constrained ICA and phase synchrony analysis. The significance of measured phase synchrony is assessed with the help of phase randomized surrogate series and only significant phase synchrony information is used for further analysis. The continuous EEG (scalp and intracranial) data used for this analysis has been collected from The Southampton General Hospital, Southampton, U.K. after ethical approval was obtained.

In essence, spatially constrained ICA helps to focus the phase synchrony analysis, utilizing multichannel information to de-noise the EEG and isolate activity of interest using the criteria of independence. In effect ICA provides a set of LDCs where phase

synchrony can be examined free of noise and volume conduction effects for specific brain areas.

7.1 Epileptic EEG Data Collected

The continuous EEG recordings have been collected from the Southampton General Hospital (SGH), Wessex Neurological Center, Southampton, UK. The data collection has been granted the approval by the Local Research Ethics Committee (Ethical Approval Number: 06/Q1701/132) and the ISVR ethics committee (ISVR Ethics approval number:817). The study has been registered with the R&D Southampton University as Project Number: RHMNEU0098. The EEG recordings are acquired at the hospital using XLTEK products (EEG 32 and EMU 128)[175] and are read by Neuroworks EEG software. Technical specifications of the EEG recording system, data format, conversion of data from European Data Format to ASCII and storage are given in Appendix A.

The data collected is multi-channel, continuous scalp and intracranial EEG of 10 epileptic patients being monitored for pre-surgical evaluation. It is continuous 24 hour scalp/intracranial EEG recordings of patients for 4-5 days per patient. The data has been anonymized with respect to the identifying details of the patient and only the technical information has been retained for analysis. It is stored and used as per the data protection guidelines and restrictions on use of patient data. No video recordings have been used or stored for research analysis.

The present analysis is mainly restricted to focal epilepsies (or focal epilepsy that later becomes generalized) as the patients that undergo pre-surgical evaluation are usually suspected to have focal epilepsy that may be helped by surgery. The pre-diagnosis by clinicians is carried out based on the patient's seizure history, clinical manifestations of the seizures and the ongoing EEG recordings. The aim of pre-surgical evaluation is to confirm the focus of seizure for the possibility of surgery. The patient stays at the hospital for 5-6 days, with scalp electrodes firmly attached to the scalp which are connected to an EEG recording system. They are allowed to carry on with the daily tasks as much as possible. Usually their anti-epileptic drugs (if being used) are withdrawn or reduced, in order to capture a maximal number of seizures in the short evaluation period of 4-5 days. They also undergo imaging tests such as PET scans and MRI scans that may help to diagnose and localize any lesions, tumors, etc. possibly causing the seizures. For intracranial measurements depth or grid electrodes are surgically inserted inside or placed outside the dura matter, respectively. The EEG data that has been collected additionally includes the following information:

- Information on relevant patient medication

- Electronic and/or paper annotations (if present) by clinical neurophysiologist.
- Any imaging studies relevant to this monitoring data
- Anonymized Consultant's diagnosis about the possible seizure onset site, type of seizure, spread of seizure
- Post surgery data/anonymized post surgery Consultant's notes (if present)
- The seizure onset timings (marked by the experts using video recordings and temporal recordings)

The current system of data storage involved storage of segments of relevant data as opposed to continuous EEG. Initially only a few segments of EEG recordings were taken but they had very limited pre-seizure data. They were not useful for the current prediction analysis as it was based on longer pre-seizure data evaluation. The scalp and intracranial data that has been analyzed is summarized in Table 7.1 and 7.2. In some cases, there was no conclusion drawn by the clinicians as to the focus of the seizure. Further investigation and/or intracranial investigation was suggested. The most likely seizure focus was mentioned based on correlations of interictal abnormalities and EEG discharges with neuropsychology results, imaging tests and clinical manifestation at the time of seizure. This highlights one of the major challenges in this study. There is much subjectivity in the diagnosis of the seizure focus itself and as there is no gold standard to validate the results, only estimations can be made. At present the only possibility of a proof of a correctly diagnosed seizure focus in epilepsy is when a focus is surgically removed from the brain and the patient is monitored for a few years and is found to be seizure free.

It is to be noted that the best possible data has been collected in the limited time, with the constraint to the patients taking appointment in that hospital for the pre-surgical evaluation. Each patient has a different and complicated manifestation of epilepsy, at times showing variations in seizures even within the same patient. Therefore, the results for this study are better shown as patient specific because it is very difficult to make generalities with this (limited) data set, for the moment.

7.2 Algorithm for Seizure Prediction and Detection

Seizure Prediction: The aim in seizure prediction is to look for indicators that can predict an impending seizure. Such indicators are required to be distinct from the background activity and should not be present at normal interictal periods away from a seizure. The application of ICA and phase synchronization on short EEG segments has previously (Chapter 6) shown that the additional information of phase synchrony

		Scalp EEG Data			
Patient	**Age**	**Length of Recording**	**Days**	**Seizure Type**	**Seizures**
C scalp *	25	Full	4	Temporal Lobe	3
D scalp	55	Full	4	Right MTLE	5
E scalp *	35	Full	4	Possibly Right Temporal	8
L scalp	56	Full	4	Left Temporal	3
O scalp *	42	Full	4	Right Frontal Lobe	14
P scalp	12	Full	4	Not Known	none
Q scalp	47	Full	4	Not Known	none
R scalp	46	Full	4	Not Known	none
S scalp	37	Full	3	Not Known	none
T scalp	15	Full	2	Left Temporal Frontal	9
W scalp		Full		Not Known	none
X scalp		Full	3	Possibly Occipital propagating to Frontal or possibly non-epileptic attacks	12
Y scalp		Full	4	Possibly Left Frontal	4
Z scalp		Full	2	Right Temporal	4
AA scalp		Full		Not Known	none
BB scalp	24	Full	2	Frontal Lobe	4

Table 7.1: EEG Scalp Data collected from Southampton General Hospital, UK. Recordings for patients with medication reduced are marked with an asterisk (*)

		Intracranial EEG Data			
Patient	**Age**	**Length of Recording**	**Days**	**Seizure Type**	**Seizures**
K IC	26	Full	2	Left Frontal Temporal	
N IC	19 month	Full	2	Lesion	
U IC *	52	Full	9	Left Frontal (Lesion)	10

Table 7.2: EEG Intracranial Data collected from Southampton General Hospital, UK. Recordings for patients with medication reduced are marked with an asterisk(*).

was able to detect seizure onset in short segments. This was seen as a sudden increase in the significant PLV values coinciding with seizure onset. This detection was seen in different seizures occurring on the same day as well as at different days for the same patient. Similar detection was observed across different patients with similar seizure focus and for different seizure foci as well. In comparison, the interictal data did not show such changes in phase synchrony in those spatial areas. The phase synchrony during interictal period was seen to remain at a low level of 0.2-0.5 (where 0 PLV corresponds to no phase synchrony and PLV of value 1 corresponds to complete phase synchrony). The same pattern was observed in interictal data segments taken at different times of the day and the night as well, across multiple patients. Essentially, the interictal segments show a low PLV while the pre-seizure/seizure segments show a sudden increased phase synchrony. The increased synchrony is sometimes observed in

few segments preceding the seizure. With the lack of continuous data, it is difficult to find a general pattern of variation in the the phase synchrony, especially to assess its usefulness as a predictor. This motivates the need for assessment of longer segments of continuous EEG. The main steps of the algorithm are depicted in Figure 7.1 and are explained next. The patient specific results are shown in the sections that follows.

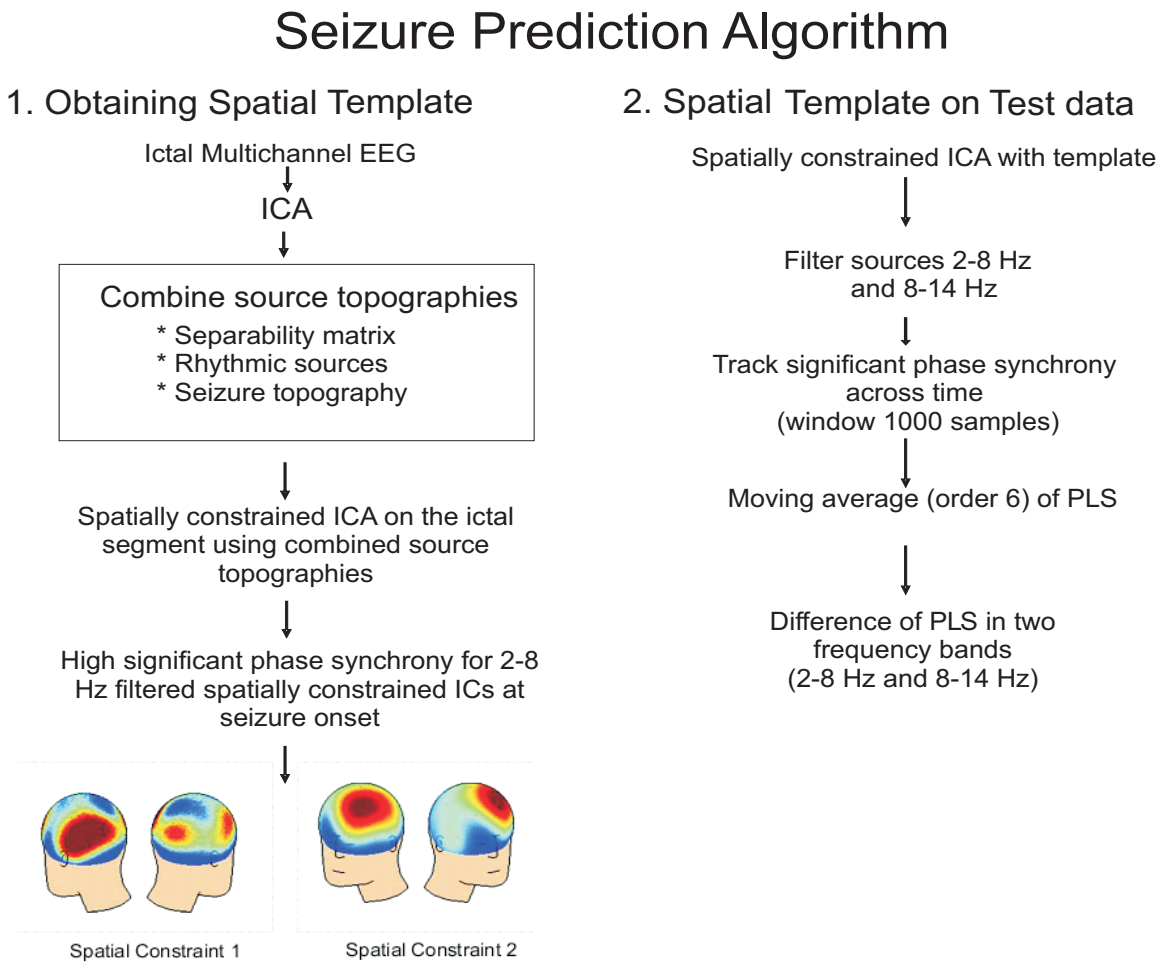


Figure 7.1: A flow diagram of the seizure prediction algorithm. The procedure is divided into two main steps: (1) Obtaining a spatial template (performed once per patient) and (2) Applying spatial template to obtain the synchrony dynamics of the source signals (performed on each segment separately).

1. **Spatial and Temporal Epileptic EEG Analysis:** The first step is the spatial and temporal analysis of multichannel EEG to preprocess the signals and obtain cleaner, fewer signals carrying the information of interest. The signal processing requires stationary signals, therefore the EEG is first segmented into non-overlapping windows of 2 minutes each. Each segment is then mean-corrected. ICA is used to unmix and de-noise these segments. However, each multichannel EEG segment gives an unordered set of multiple source signals. Even though a source signal can be selected from ictal segments, it is not trivial to track the

seizure signals further away from the ictus, due to lack of known features. In order to track seizure sources across these segments, the use of spatially constrained ICA and Phase synchronization has already been suggested (Chapter 6). The ICA-PLV technique requires a patient specific spatial template. The procedure of obtaining a spatial template is an important first step followed by spatial filtering of the entire EEG of that patient. These steps are explained as follows:

2. **Obtaining the Spatial Template:**

   (a) A multichannel template EEG segment is formed for the patient. The template EEG used is such that it exhibits sufficient pre-ictal and ictal periods (about 100 seconds of each).

   (b) ICs are estimated for this segment using TDSEP and the topographies were plotted from the mixing matrix **A**.

   (c) The separability matrix is formed for these sources, with such a threshold as to obtain a good demarcation between source clusters.

   (d) The topography of the sources obtained are also generated.

   (e) Using the information from the time domain signal, topography plots, separability matrix and patient reports, dependent source clusters are formed.

   (f) The clusters sources (dependent sources) are combined together to form a common source.

   (g) These cluster topographies are then used as **constraints** for further analysis of any segment (ictal or inter-ictal) of the same patient.

   (h) For comparison, a pair of **contralateral** (away from seizure focus) and another pair of **ipsilateral** (near seizure focus) **constraints** are formed as above for each patient. This provides a set of measures both within and outside the seizure focus.

3. **Spatial Filtering:** The spatial templates (contralateral and ipsilateral) are then used to apply spatially constrained ICA on all ictal and interictal EEG segments of the same patient. The pair of seizure sources obtained from contralateral and ipsilateral templates for each segment are retained along with their topographies. This helps in searching the LDCs in an area of the brain that has a possibility of some epileptic activity (and the comparison, away from the active area).

4. **Synchrony Dynamics across Time:** The next step is the tracking of the phase synchrony, in specific relevant bands (2-8 Hz, 8-14 Hz), using spatially constrained sources. The sources obtained above are representative of continuous multichannel scalp EEG having a distribution as the spatial template. The seizure segment used for obtaining the spatial constraint is not included in the prediction analysis. Non-overlapping moving windows are used to track the sources and phase synchrony. The window length used for PLS is 5 seconds (1000 samples)

long. **Significance Tests:** The significance of phase synchrony is tested by creating 100 phase randomized surrogate series for pairs of LDC signals. The PLV of each pair of surrogate series is obtained, giving a bootstrap distribution. If the PLV between the LDCs is higher than 90% of the sorted surrogate PLVs then it is considered significant. **Moving Average Filter:** A moving average filter of length 300 samples is used to obtain a smoother variation of the PLV across time. This number was obtained through empirical observation. Taking the PLV windows into account, 300 samples of PLVs correspond to 1500 seconds (25 minutes) of data. **Squared significant PLV difference in two frequency bands or PLV-d:** The similarity in significant PLV (or PLS) of neighboring frequency bands was analyzed by finding the difference between the (significant) PLVs in 2-8 Hz and 8-14 Hz bands. The hypothesis for this analysis is that when a seizure source becomes active it should entrain signals and oscillators in certain spatial locations in the brain (focal or contralateral). Thus neighboring frequency bands might show significant PLVs in the same time period and hence showing a small difference.

The difference of these weighted mean values should show:

(a) A zero difference when the two frequency bands coincide in their respective PLV behavior i.e both frequency bands are synchronizing at the same instant,

(b) A higher difference when any one band significant PLV is higher i.e the synchrony in one band is stronger compared to the synchrony in the other band.

(c) A lower difference when the PLS in both bands is tending to occur at the same instant i.e PLS behavior is becoming similar in the two bands. So if we hypothesize that at seizure the synchrony should be wide spread, the focus might (acting as the force) cause nearby frequency bands to synchronize with its own rhythm then we should see the value of the squared difference of the means become zero at seizure.

The bands 2-8 Hz and 8-14 Hz were selected after analyzing the patterns observed in PLV-d for various combinations of the eight bands: 2-8 Hz, 8-14 Hz, 14-20 Hz, 20-26 Hz, 26-32 Hz, 32-38 Hz, 38-44 Hz, 44-50 Hz. The bands that hold relevance to the analysis in question, i.e. epileptic seizures, as well as those that showed a distinguishable pattern were the difference of 2-8 Hz and 8-14 Hz. Hence, these were selected for the rest of the analysis.

Seizure Detection: Although, previously seizure detection has been performed on short (two minute) discontinuous data segments, it showed promising results. Nevertheless, it is vital to validate it with long term EEG data. Therefore, similar analysis for seizure onset has been performed here on long term continuous data. Analysis on a

longer data set allows one to validate if such patterns are exclusive to seizure onset or are arbitrarily spread across the EEG because of the brain activity through the entire day. Such validation on longer data sets is important to assess the specificity and the sensitivity of the algorithm. The algorithm involves the following steps. The details have been previously covered in chapter six.

1. **Pre-processing multichannel EEG**
2. **Creating spatial template**
3. **Spatial filtering**
4. **Tracking phase synchrony changes**
5. **Sensitivity analysis**

Phase synchrony tracking and sensitivity analysis was also performed for raw EEG signals (with no unmixing or artifact removal, and subjective selection of spatially contralateral and ipsilateral electrode signals) for the sake of comparison. Some results for the seizure detection have been shown here with each patients results in the following sections. They include seizure detection performance on contralateral and ipsilateral LDC PLVs and on contralateral and ipsilateral EEG PLVs for comparison.

7.3 Seizure Analysis for Patient 1 (Patient filename: T Scalp)

The algorithm is now applied and discussed for a continuous EEG dataset of **Patient T scalp**.

Patient History: The patient has recordings for two continuous days. The patient underwent no drug reduction during the evaluation. His diagnosis reports suggested a left temporal lobe onset. A left temporal abnormality is also mentioned as seen in the MRI scans. The electrographical onset of seizures show rhythmic theta (12 Hz) over the left temporal and fronto-polar regions. This activity evolves into high amplitude rhythmic theta, maximal at temporal-frontal region. The ictal EEG for this patient is similar for all nine seizures.

Synchrony Dynamics for Prediction: In order to track the synchrony dynamics, the PLV-d curve for patient has been computed for the two frequency bands 2-8 Hz and 8-14 Hz, shown in Figure 7.2. The PLV and the PLV-d curves were obtained for a pair of contralateral source signals as well as ipsilateral source signals. The one that shows the most prominent pattern is shown here. The figure shows the seizure onset times (marked by vertical lines). The start time of the data is 13:00 hours in the afternoon (of day one). **Observations:** A striking cyclical pattern is apparent in general in

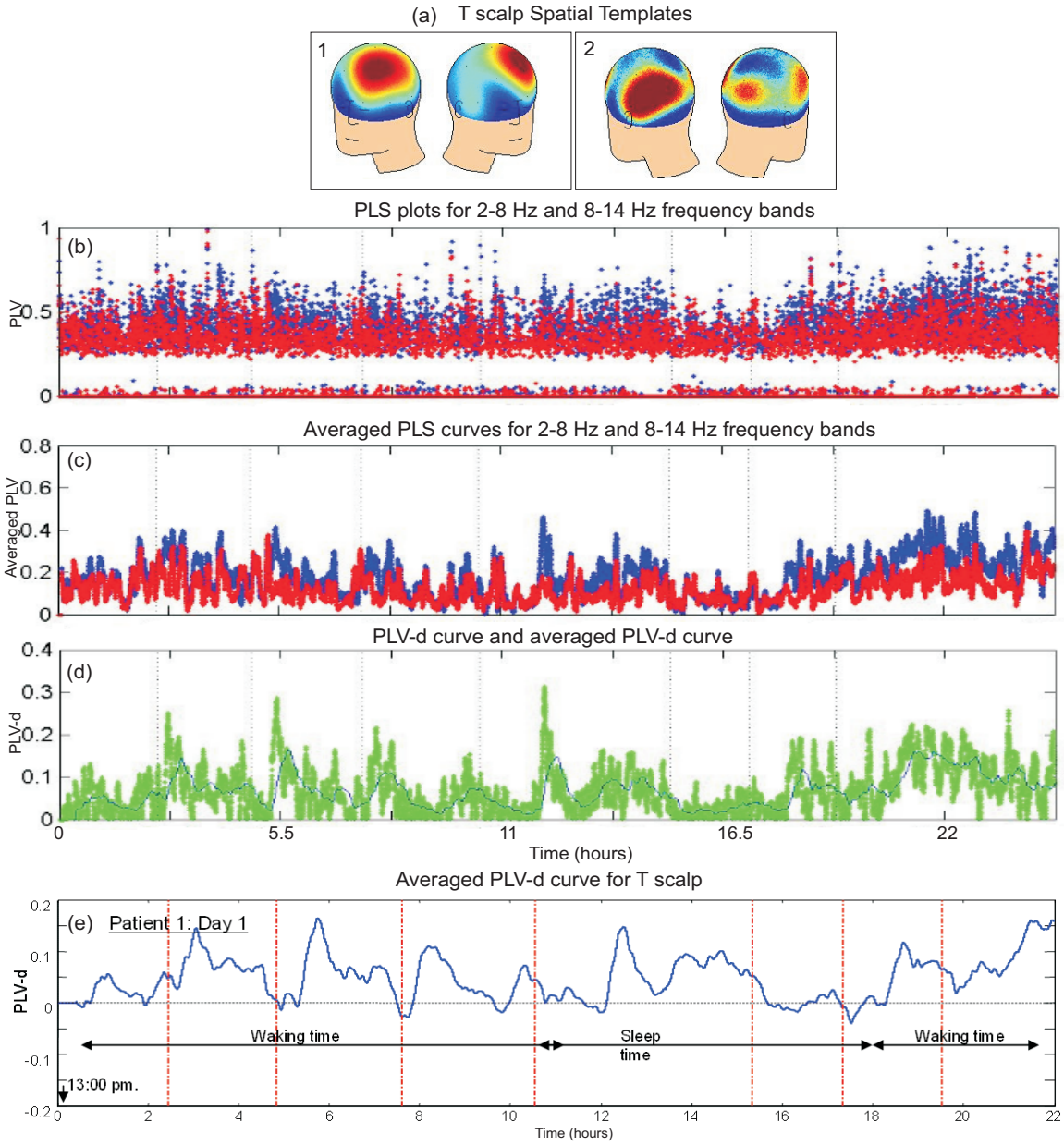


Figure 7.2: The plot shows the transition of PLV to PLV-d curves for Patient T scalp. Vertical lines indicate seizure onset. The start time of the data is 13:00 hours in the afternoon. (a) The spatial template used for Patient T scalp. (b) The significant PLV points plotted for 2-8 Hz and 8-14 Hz frequency bands. (c) The averaged PLV curves obtained by moving average filter applied on (a) above. (d) The PLV-d curve obtained as the difference of weighted significant PLV values of two frequency bands 2-8 and 8-14 Hz over a continuous EEG segment that is 24 hours long. (e) The averaged PLV-d curve, as used in further analysis for all patients.

both the PLV and the PLV-d curve. It shows the possibility of a certain underlying pattern in the synchrony in the brain, irrespective of the activities (such as reading, eating, talking, moving, etc.) that it may be involved in. Perhaps such a pattern might hold clues to predict some major changes in the brain, such an oncoming seizure. On further observation, the cyclical pattern appears to be in some sync with the seizure

onset timings as if indicating a subtle synchrony ***'set and reset'*** mechanism. In figure 7.3, showing the **PLV curve** for day one for this patient, a prominent two peak pattern can be observed (marked by arrows) leading on to a seizure. The second peak may be a possible indicator of an impending seizure. The time differences from the second peak to the seizure onset for the patient is also shown on the figure. The peak times: 11.4 min, 52.8 min, 73.8 min, 107.4 min (sleep time), 6 min, 33 min show an increasing prediction horizon till the sleep time, after which it reduces and starts increasing again.

The spatial template selected for this patient is shown in Figure 7.2 (a) along with the significant PLV points plotted for two frequency bands: 2-8 Hz and 8-14 Hz (Figure 7.2 (b)). The PLV curves are then averaged (using a moving average filter) to highlight the underlying patterns (Figure 7.2 (c)). Figure 7.2 (d) shows the **PLV-d curve** and the final averaged form of PLV-d used in further analysis (Figure 7.2 (e)). The cyclical pattern is seen to be retained in the PLV-d curve and appears to be visually clearer than in the PLV curve. A seizure appears to pull the synchrony difference curve towards zero level, which would imply that the synchrony is becoming the same in the the two frequency bands i.e. both bands are getting synchronized to the same frequency. Then as the seizure wanes, the phase synchrony in the two frequency bands starts to move apart to their different frequencies, showing an increase in the PLV-d. Let us assume this phase is a ***'set'*** phase. As an oncoming seizure starts to synchronize the neurons again in the same frequency, it must pull the synchronies in the two bands together towards zero level again. If the brain is able to break this synchrony in time to prevent a seizure onset, the synchrony in the bands again moves apart, though to a lesser degree (smaller peak). Lets assume this is a ***'reset'*** phase. In succession, say a stronger attempt by the seizure focus brings the neurons to complete synchronization, causing the two frequency bands to completely synchronize, bringing the PLV-d to zero level at seizure onset.

These patterns can be important in indicating an oncoming seizure, or a stage when a seizure may perhaps be stopped by breaking the synchronization. Nevertheless, it is realized that although this subjective analysis is encouraging, it needs to be supported by an objective feature extraction procedure, which is covered in a later part of this thesis.

Synchrony Dynamics for Seizure Detection The detection algorithm described before is used to track the significant phase synchrony across time, in band 2-8 Hz, for spatially constrained sources. The main difference here, being that instead of the PLV-d curves, the filtered PLV curves are analyzed.

The spatial constraint is obtained by pre-analysis of a different seizure segment which is not included in the detection analysis here. Non-overlapping moving windows are used to track the sources and calculate the phase synchrony. The window length used

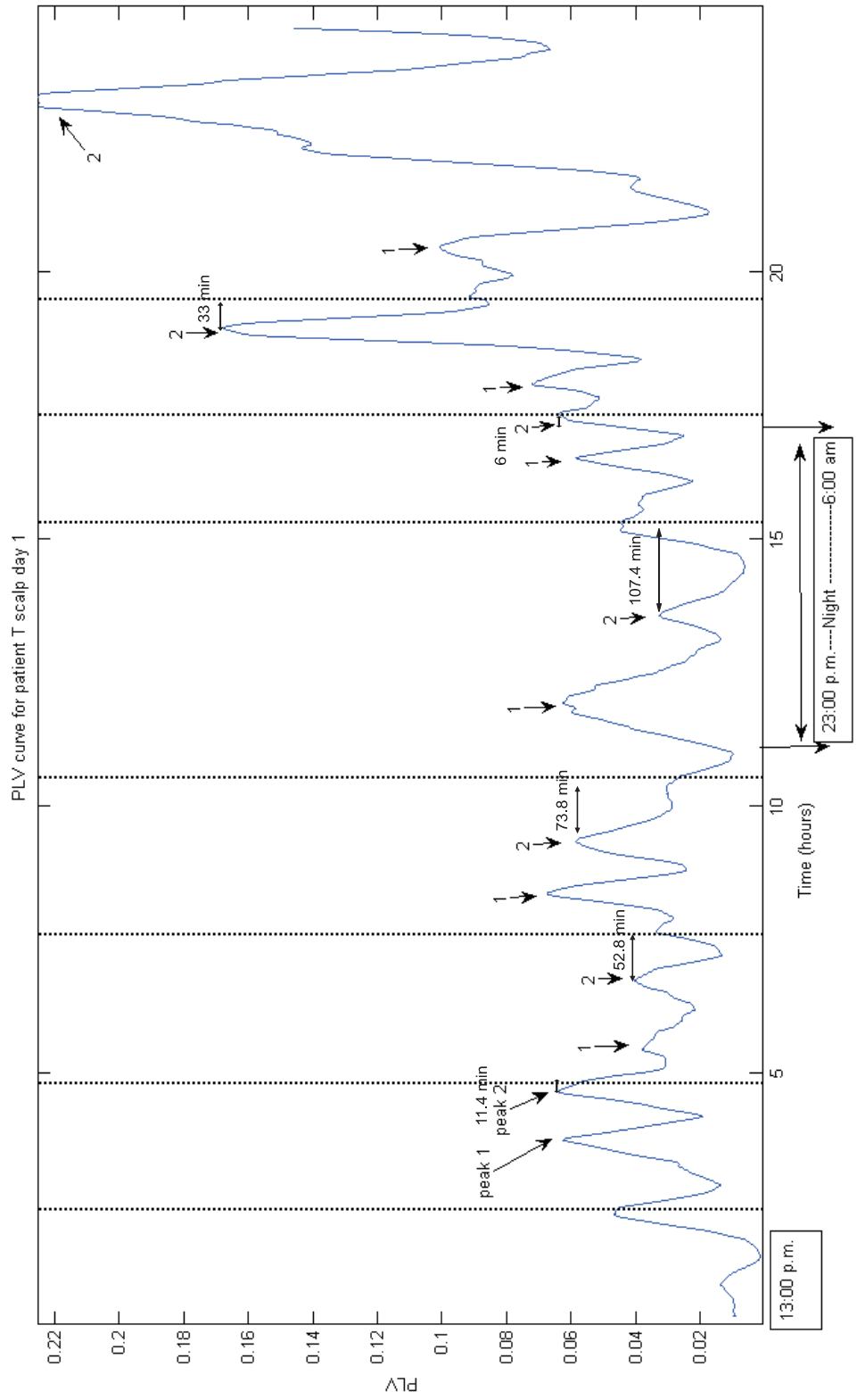


Figure 7.3: The plot shows the PLV curve for the Patient T scalp, day 1, over a segment that is 24 hours long. Vertical lines indicate seizure onset. The start time of the data is 13:00 hours in the afternoon. A prominent two peak pattern can be observed (marked by arrows) leading to a seizure. The time from second peak to the seizure is also shown.

for significant PLV calculation is 5 seconds (1000 samples). The significance of phase synchrony (PLS) is tested by using 100 phase randomized surrogate series. To obtain a smoother curve of the PLV variation across the long time period time, a moving average filter is used i.e. the PLV curve low pass filtered with a moving window of 300 PLV values. The results for ictal activity segments are shown separately in 30 minute segments for a clearer view, in Figures 7.4, 7.5, 7.6 and 7.7. The results are then shown for 15 hours of continuous EEG, see Figure 7.8. From Figure 7.8, it can be seen

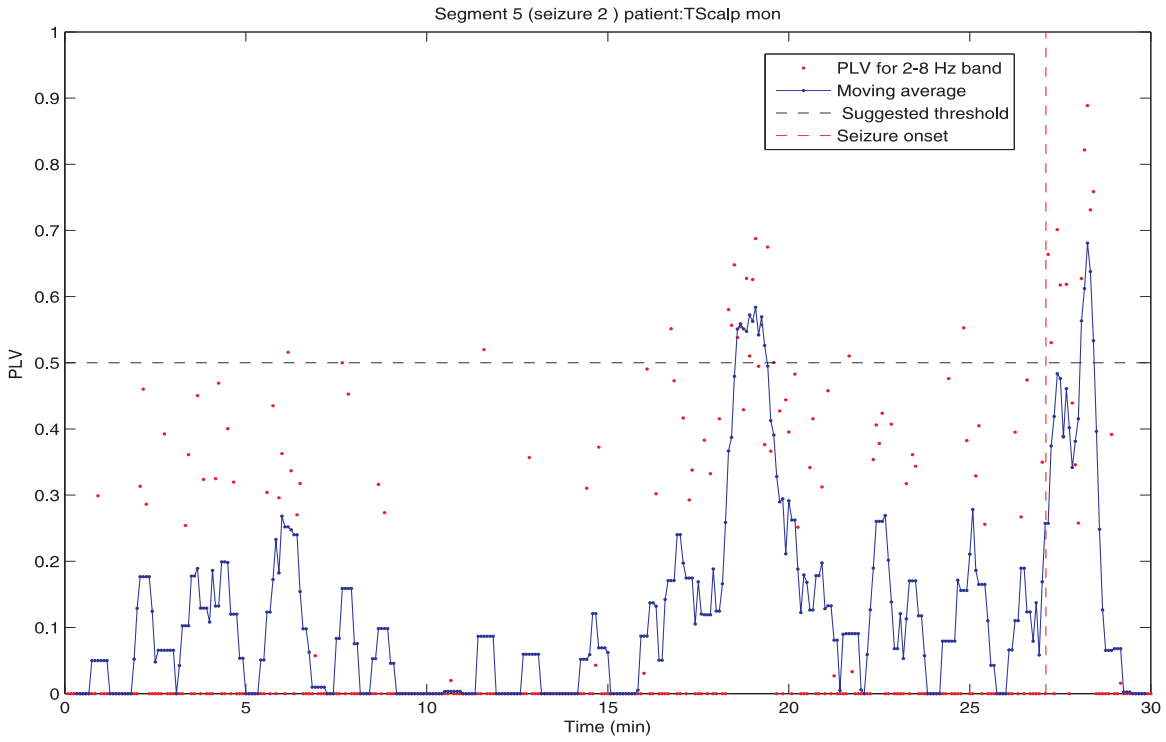


Figure 7.4: The plot shows the PLS (with moving average) for patient T scalp tracked across 30 minutes of two spatially constrained sources for seizure detection. The seizure 1 onset is marked by the vertical line.

that with a particular threshold four out of five seizures are detected (80% sensitivity) and nine false positives occur over a period of 13 hours giving a False Positive Rate (FPR) as 0.6 per hour. This is quite a high FPR and so in order to analyze what led to these false positives, the EEG at those times was observed and some examples have been shown in Figure 7.9. In this instance the false positives were found to occur due to the following causes:

1. **Strong Artifacts:** An artifact such as EEG electrode apparatus being disconnected shows a sudden synchronized sharp wave complex in all channels and hence is found to be retained in sources obtained by spatial ICA.

2. **Sleep Spindles:** Sleep spindles are rhythmic burst activity that is seen to be present in the EEG at the time of sleep (specifically in the stage 2 of sleep), see

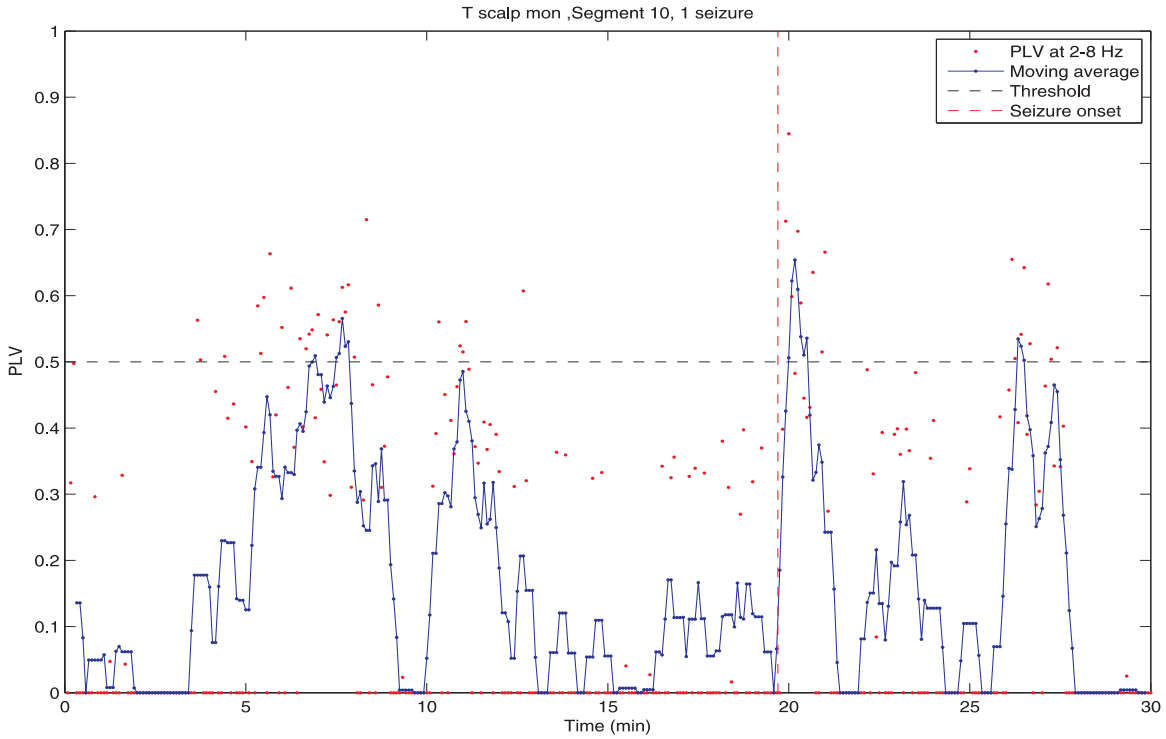


Figure 7.5: The plot shows the PLS (with moving average) for patient T scalp tracked across 30 minutes of two spatially constrained sources for seizure detection. The seizure 2 onset is marked by the vertical line.

Figure 7.9. These bursts are of the alpha frequency (2-14 Hz) and may occur over the central area of the scalp. Thus if the spatial constraints are near the central head region and the band of interest is 2-8 Hz, then this activity is most likely to be picked up as an independent source by spatially constrained ICA during interictal periods. These may or may not be phase synchronized, but as we see here when it does show synchrony then it leads to false positives. Perhaps if it is possible to remove such sleep spindles prior to seizure detection analysis, it might remove 90% of false positives occurring at night time.

3. **Focal Rhythmic Waves:** These apparent random occurrences of synchronized focal waves in EEG are not yet explained. They have similar frequency content as the band being analyzed and occur in the spatial area being analyzed, see Figure 7.10. Therefore they are quite likely found as independent sources during interictal periods with spatially constrained ICA. These focal waves seem to synchronize 10 min - 1 hour prior to a seizure onset. This may indicate a measure for prediction and will be analyzed further with analysis on different patients data in the next sections.

The algorithm for seizure detection was similarly applied to the patient's contralateral and ipsilateral LDCs and corresponding EEG signals for comparison. The results are shown in Figure 7.11. The seizure detection plots for LDC contralateral and LDC

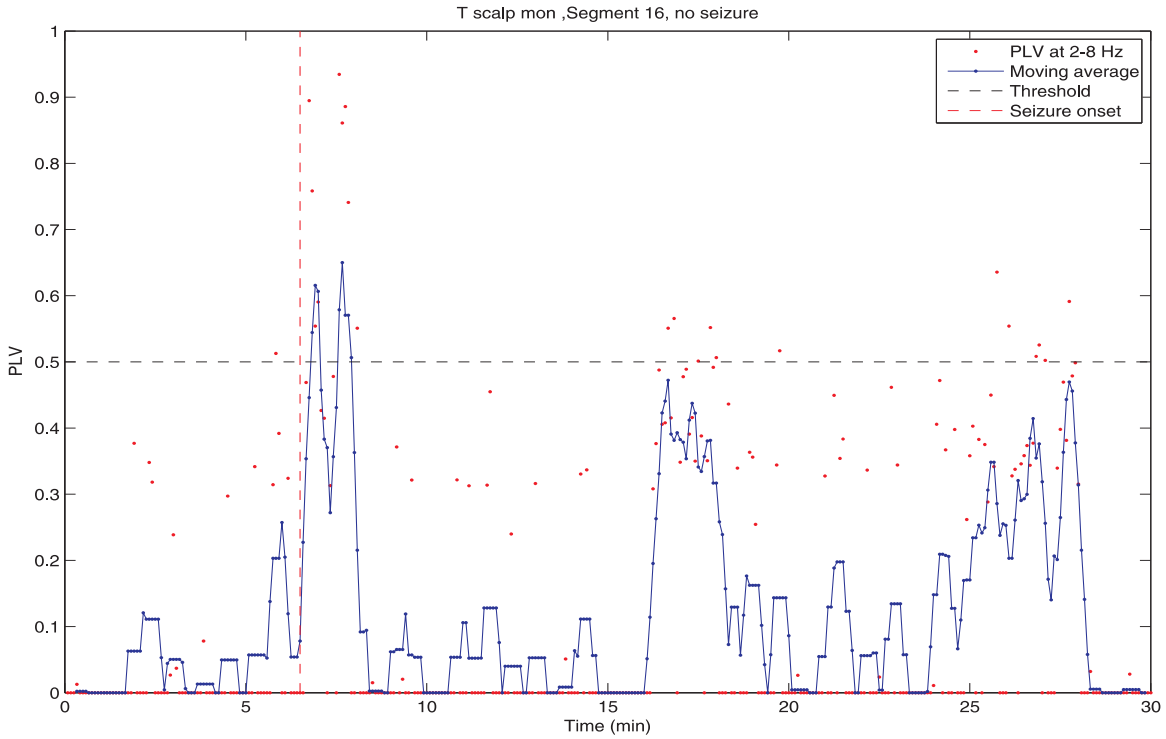


Figure 7.6: The plot shows the PLS (with moving average) for patient T scalp tracked across 30 minutes of two spatially constrained sources for seizure detection. The seizure 3 onset is marked by the vertical line.

ipsilateral source signals show a sensitivity of 84% and 100 % with corresponding specificity of 5.5% and 42.8%, respectively. The LDC ipsilateral results are quite promising. Comparatively, when EEG signals are used without any unmixing, the sensitivity for EEG contralateral and ipsilateral data drops to 67% each and specificity reduces drastically less than 1%. The reason being, that the synchrony of EEG signals appears to be quite high most of the time, leading to the failure of the algorithm to distinguish the increased synchrony due to seizure.

This analysis is further performed for seizure prediction and detection on the scalp EEG data set of nine patients collected as shown in Table 7.1.

7.3.1 Patient 2 (Patient filename: E Scalp)

Patient History: Patient E scalp is an adult patient, with continuous scalp EEG recorded for four days. He underwent drug reduction and had several stereotype seizures during this pre-surgical evaluation. The report indicates abnormal background EEG, with abnormal frequent sharpened slow waves and occasional low amplitude spikes in the right hemisphere, especially the anterior temporal region of C4 and Fz. The ictal EEG is said to be abnormal but the ictal discharge seems bilateral and is not lateralising. The interictal findings and seizure semiology suggest a right temporal abnormality. Neuropsychometry results showed non verbal memory deficit. Seizure onsets are found to be accompanied by some behavior arrests and manual automatisms with right hand. Frequent sniffing, touching mouth and rubbing nose with right hand, all suggest right sided focus. No lesion was found in the MRI scans. With the present information, a mesial temporal lobe epilepsy (mTLE) with right sided onset was suggested (see Figure 7.12).

Selecting Spatial Constraint Templates: The selection of spatial constraints was performed by using the above ictal EEG segment. Temporal Decorrelation was applied to the segment and the source signals obtained are shown in Figure 7.12. The separability matrix for the sources showed apparent clusters as marked on the Figure 7.13. The topographies of source signals within a cluster were found to be similar,

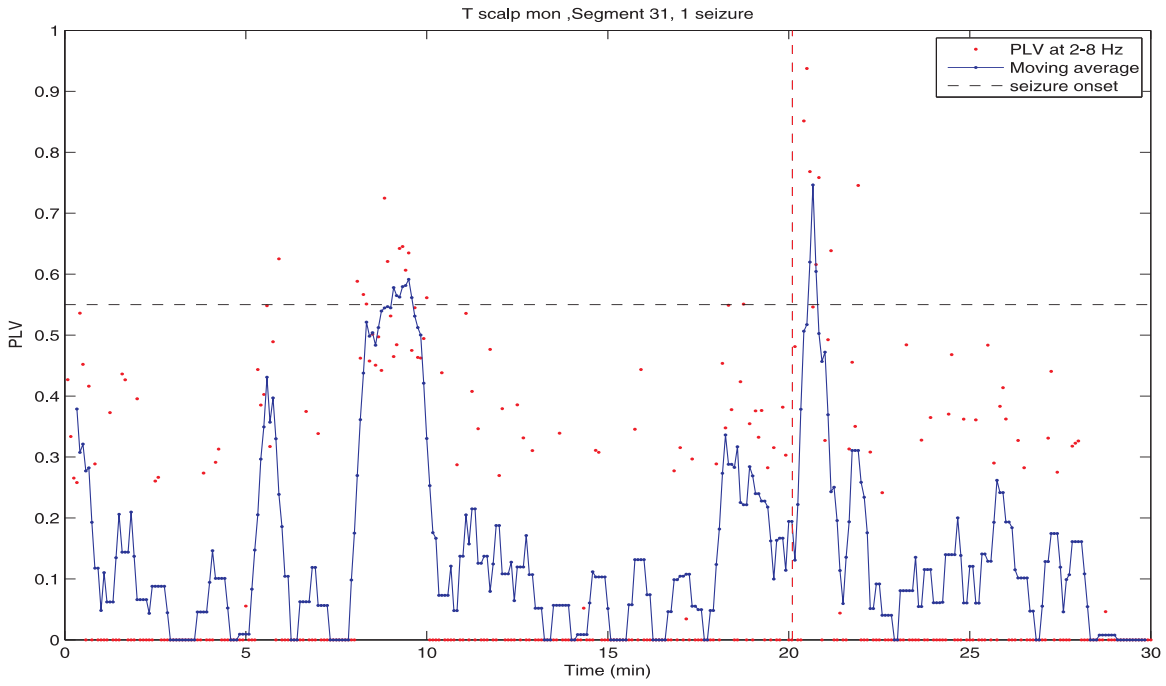


Figure 7.7: The plot shows the PLS (with moving average) for patient T scalp tracked across 30 minutes of two spatially constrained sources for seizure detection. The seizure 5 onset is marked by the vertical line.

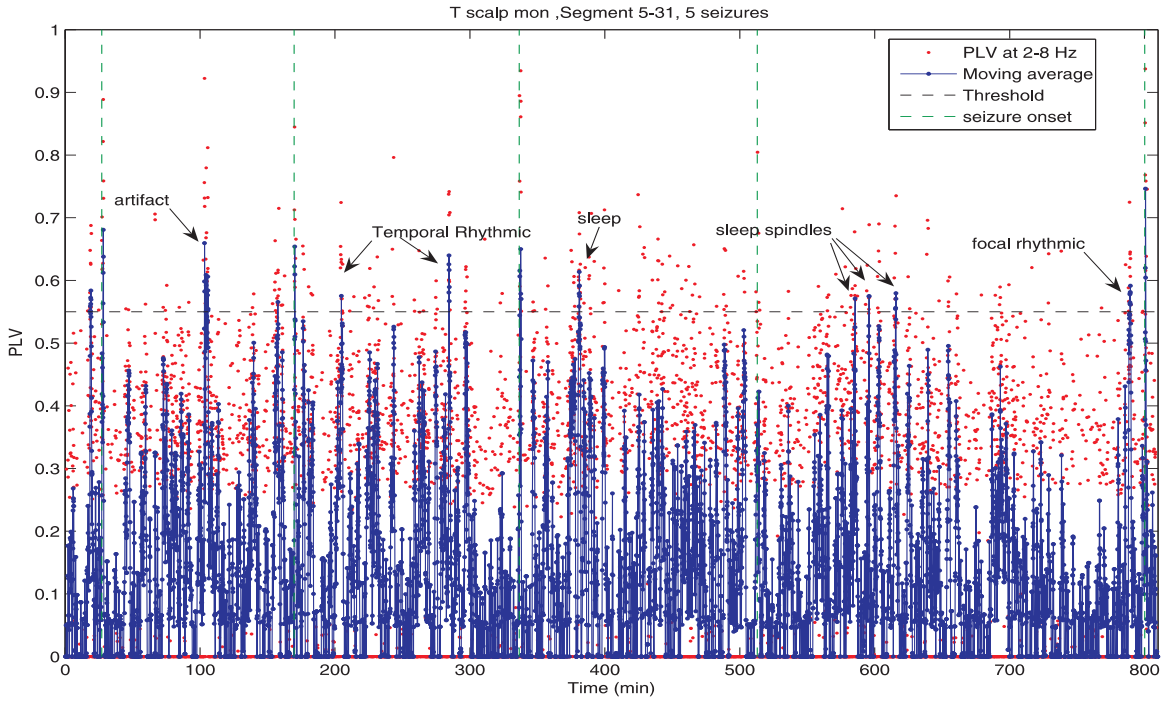


Figure 7.8: The plot shows PLS in 2-8 Hz band of spatially constrained sources of 13 hours continuous scalp EEG. There are 5 seizures marked by vertical dotted lines. A moving average is taken over the PLS plot (solid line).

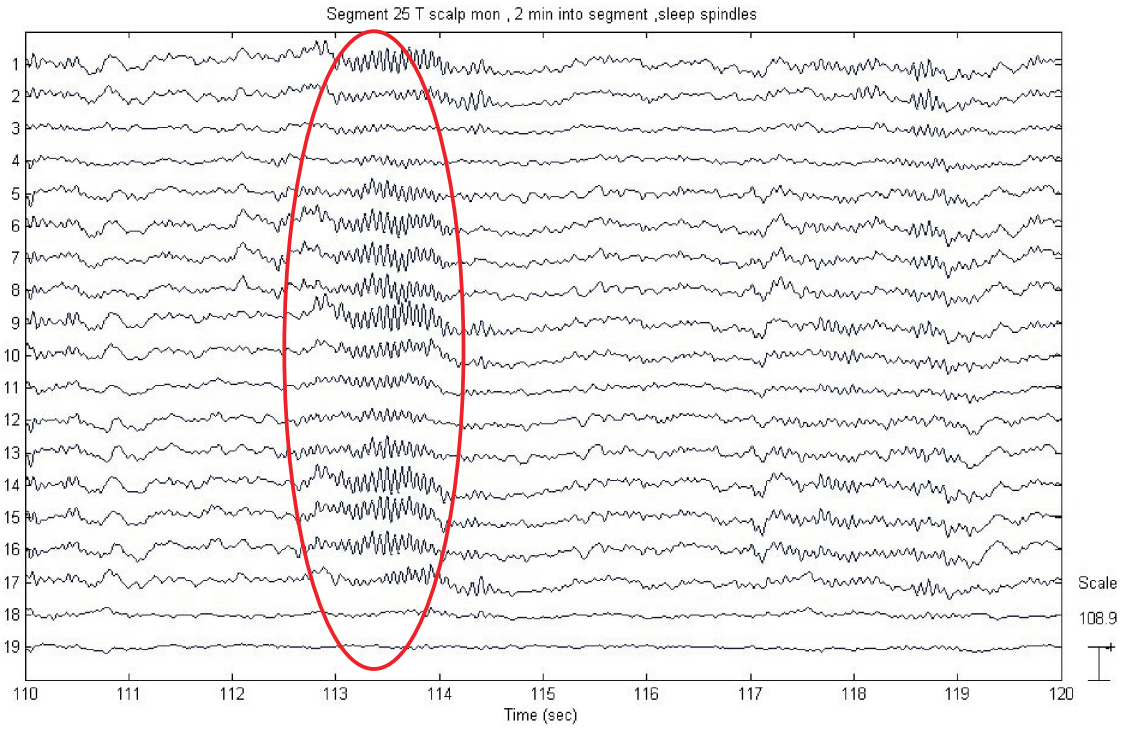


Figure 7.9: Sleep spindles at 600 minutes in the Figure 7.8

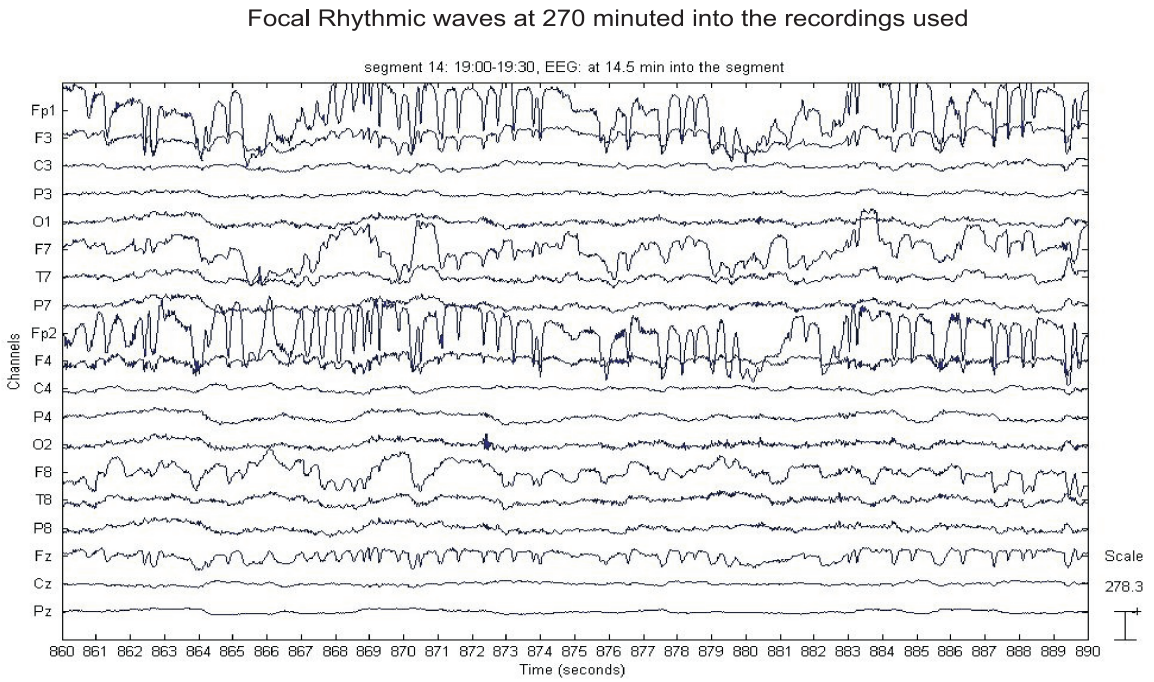


Figure 7.10: Focal rhythmic waves seen to occur at 270 minutes into the EEG recording used in Figure 7.8.

indicating that the separability matrix was essentially estimating a topographical or spatial separation/dependence. If this was the case then it was expected that the contralateral synchrony information would be best observed in intra-cluster source signals as opposed to inter-cluster sources used for ipsilateral constraints. Additionally the temporal information from the source signals, the patient history, the phase synchrony patterns, mutual information were assessed to select the ipsilateral and contralateral spatial constraints for this patient, as shown in Figure 7.13.

ICA-PLV for Seizure Detection: The seizure detection for E scalp has been shown in Figure 7.14, where the detection with ICA-PLV has been estimated for a combination of contralateral and ipsilateral spatially constrained least dependent source signals as well as the corresponding EEG signals. The figures indicate that even though the specificity is low but the maximum possible specificity with a good sensitivity is possible with the source signals as opposed to the EEG signals. The ipsilateral source signals were seen to produce the best results, and the corresponding ipsilateral EEG signals, the worst. The ipsilateral EEG signals show an exorbitant high synchrony for a continuous 22 hours. The spurious synchrony in the ipsilateral EEG signals can be observed to be overwhelming to any possible patterns of synchrony lying in the data. ICA is able to remove this spurious synchrony.

Synchrony Dynamics for Prediction: The PLV-d curves for this patients' LDCs are shown in Figure 7.15. Some of the general and specific observations that were made from studying the PLV-d curves of the four day continuous data of this patient, have

been enumerated here.

Observations

1. *Distinct cyclical pattern of synchrony*: There is a general cyclical pattern observed in all the four days of PLV-d curves of this patient. A free hand curve is drawn to aid the visual appearance of this pattern. On first thought, this cyclical pattern could be attributed to daily chores, such as eating, talking, chewing and other artifacts, but it should be noted that with the averaging and smoothing of the curves, the PLV-d curve is expected to indicate changes in synchrony on longer term data, as long as half an hour. In such a case, artificial synchrony effects would easily be smoothed away. The general cyclical pattern is more likely to be linked to the circadian-like rhythm of the body (brain). The PLV-d curve is specifically observed to undergo a transition at onset of sleep (usually marked as '*onset of drowsiness*' in data annotations by the electroencephalographers). The physical interpretation of this transition would be that at waking time, the synchrony in one of frequency bands is dominant, hence the PLV difference is more than zero. At the onset of drowsiness, the synchrony dominance sharply shifts to the other frequency band, and remains in the same mode until sleep time. At waking time the dominance shifts back to the previous frequency band

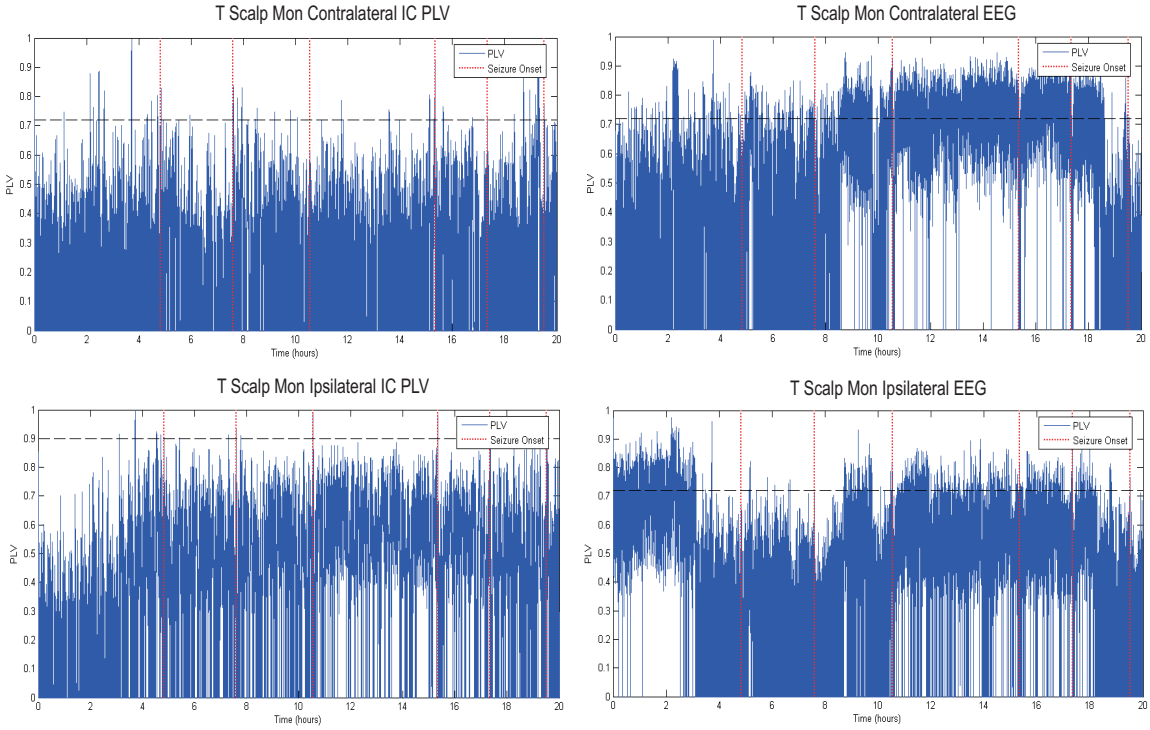


Figure 7.11: The significant PLV for contralateral and ipsilateral source signals and EEG signals. The sensitivity and specificity of seizure detection is estimated using a threshold (marked by horizontal line on the Figures). The vertical dotted line indicates the seizure onset times marked by expert electroencephalographers at the hospital.

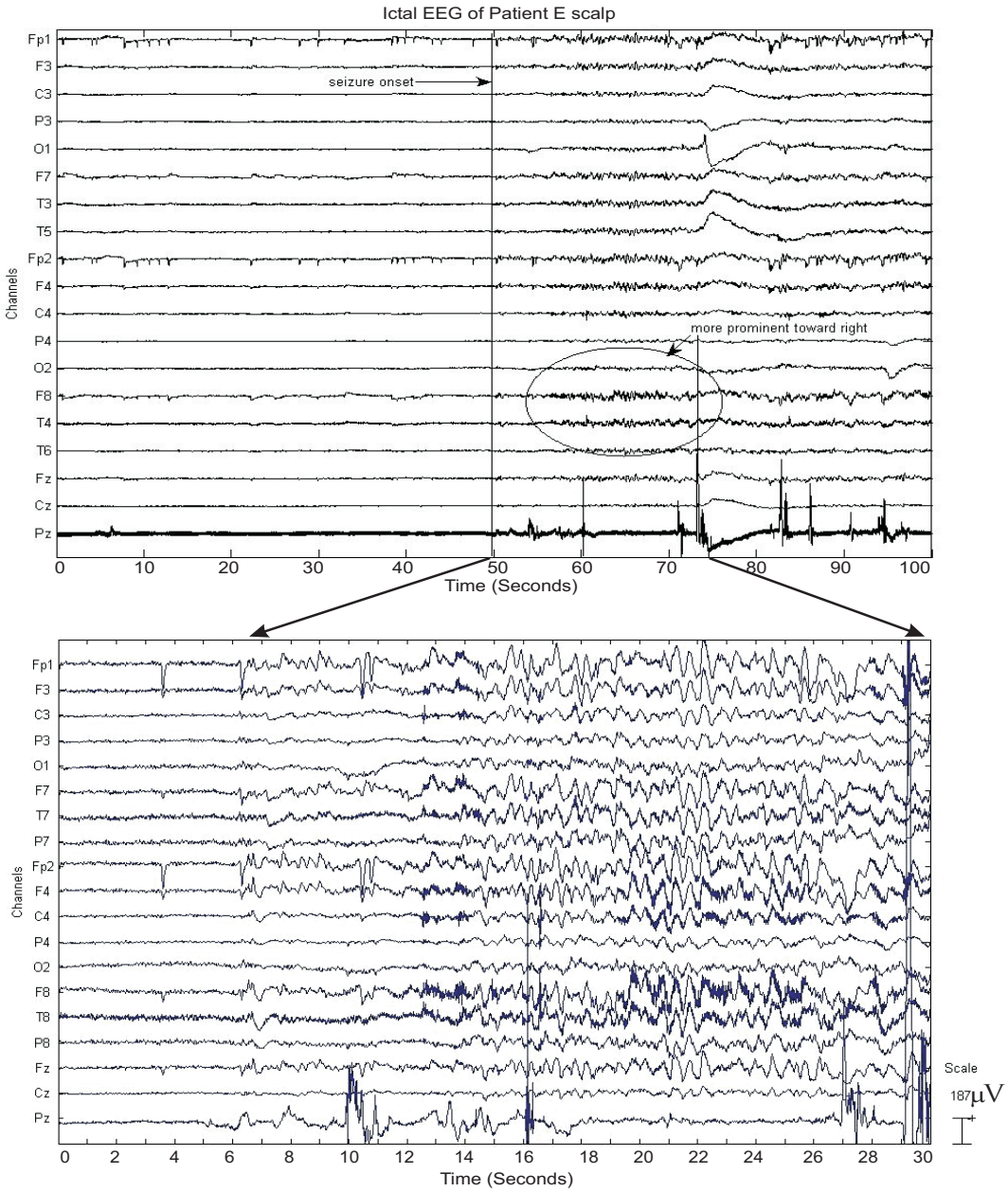


Figure 7.12: Ictal EEG of Patient 2 E scalp.

again.

2. *A seizure appears to pull the PLV-d curve against the normal cyclical pattern*: If the above cyclical pattern is taken to be the norm then from Figure 7.15 it appears that a seizure onset pulls the PLV-d curve away from the normal cyclical pattern. A seizure during the day pulls the PLV-d curves below zero level, implying that at seizure onset the frequency of the 'normal brain synchrony' gets arrested or entrained by a different frequency (or frequencies) for a short while.

However, the last two days that are free of seizure do show some patterns that are difficult to explain. These synchrony peaks during seizure free times might represent

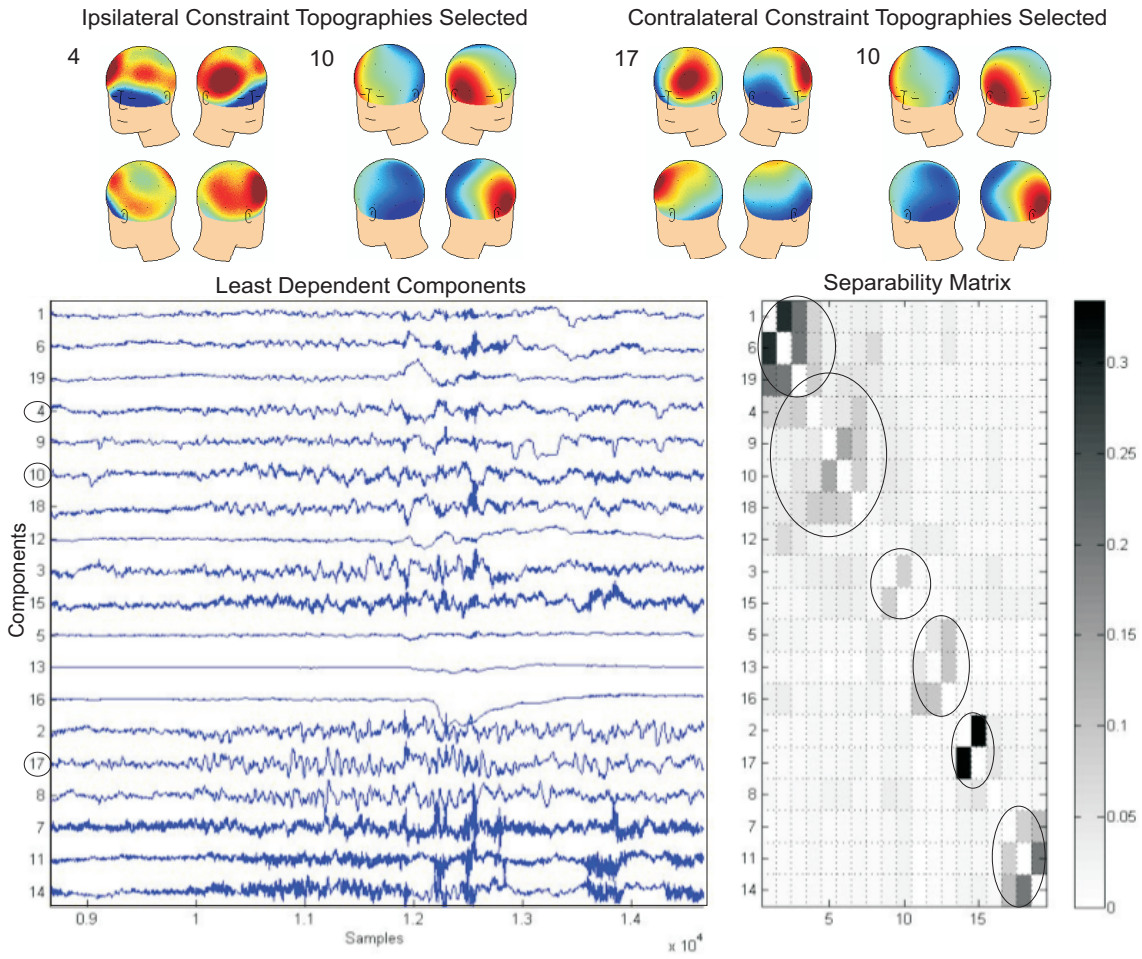


Figure 7.13: Source signals obtained for patient E scalp, using temporal decorrelation. The separability matrix shows the extent of separation of the source signals. Some visually apparent clusters have been encircled. The contralateral and ipsilateral spatial constraint topographies selected for this patient have also been shown.

the times when the synchrony starts to build up but does not reach the threshold to lead to a seizure onset. It may also involve complex dynamics of the brain as it is undergoing a reduction in drugs, sleep deprivation, change in environment and activities, etc. which is very much an integral part of a pre-surgical evaluation EEG data set. This requires a comparative study with non-epileptic patients EEG data.

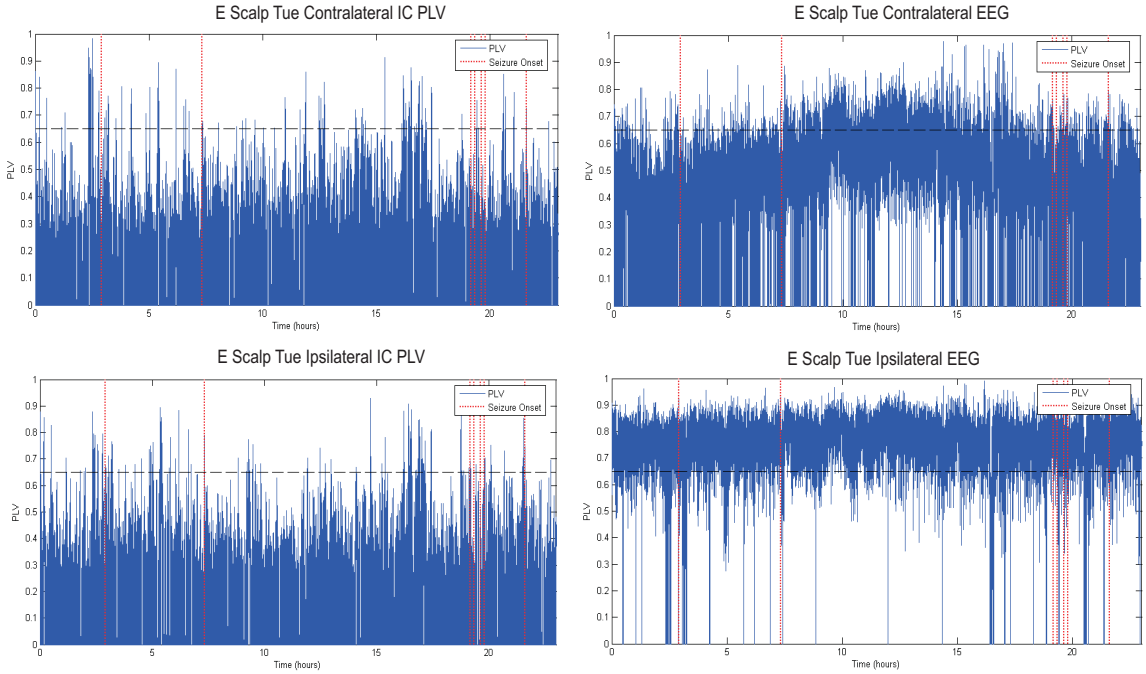
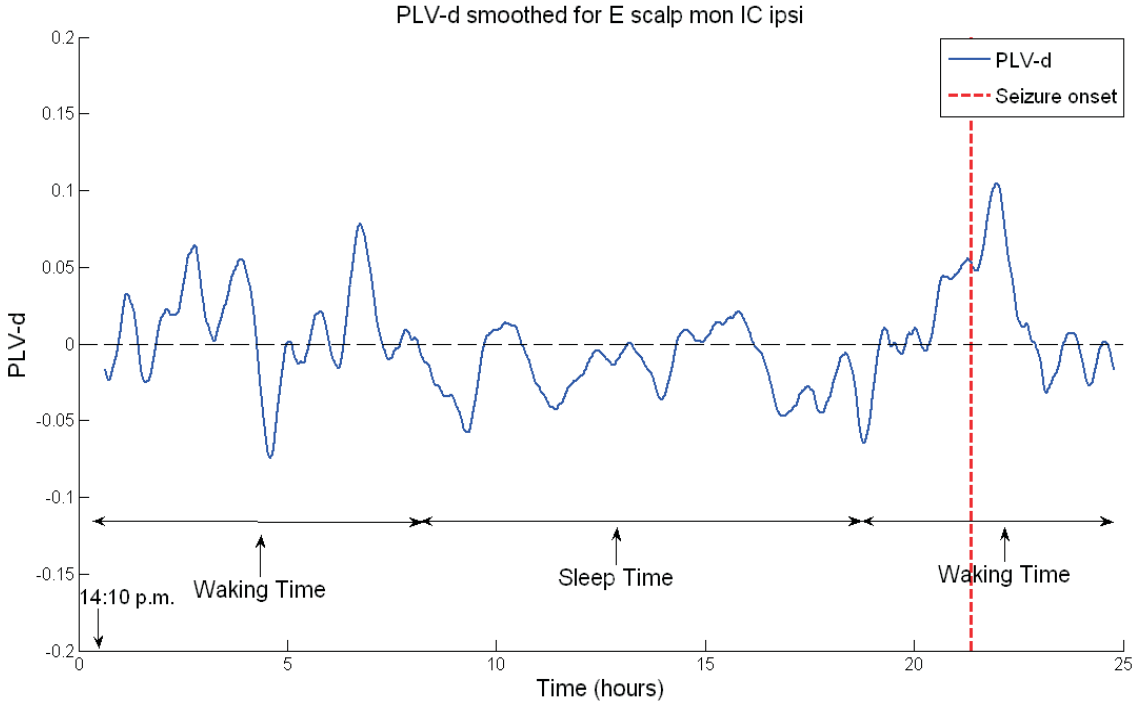


Figure 7.14: The two left figures show the PLV for contralateral and ipsilateral source signals. The two right figures show the PLV for the corresponding contralateral and ipsilateral EEG signals. The vertical dotted line indicates the seizure onset times marked by expert electroencephalographers at the hospital.



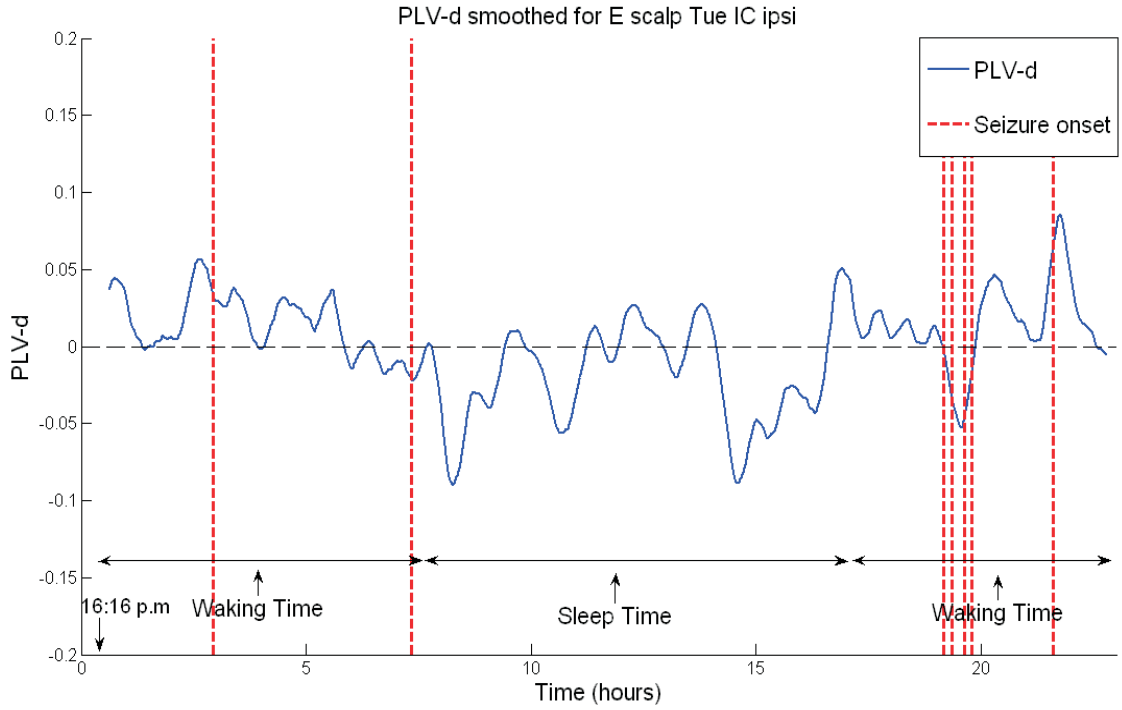
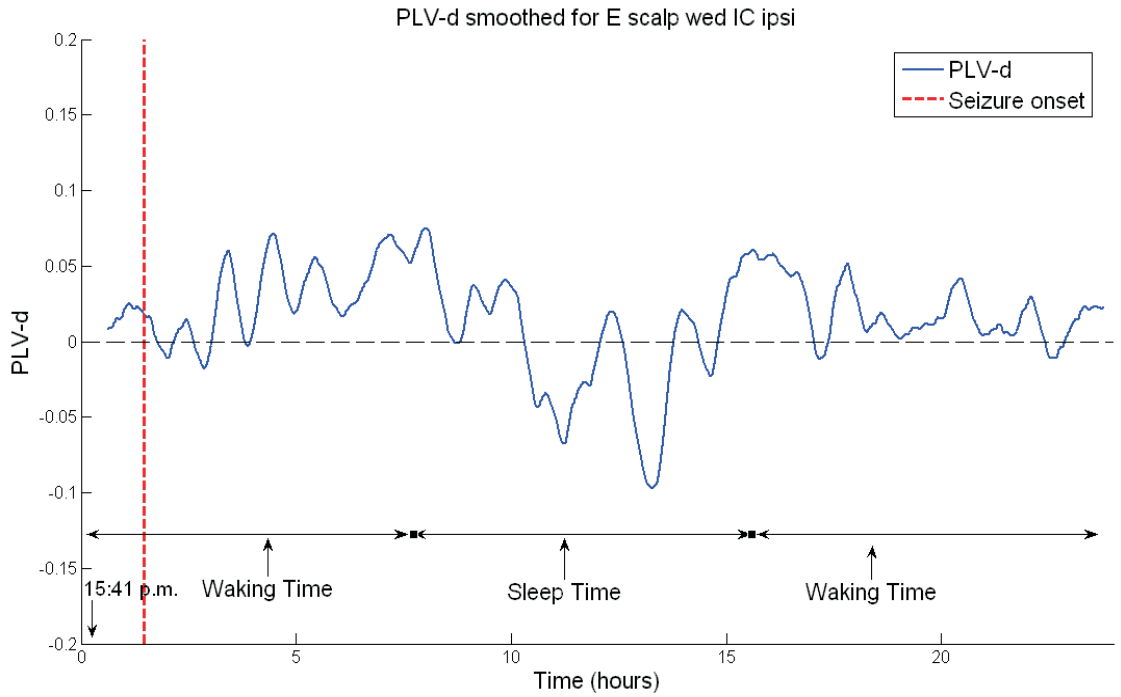


Figure 7.15: The two figures show the PLV-d curves of two continuous days of scalp EEG each, of patient E scalp. The vertical lines mark the seizure onsets. A free hand arbitrary curve (broken line curve) is drawn to aid the visual detection of a circadian-like rhythm that can be observed to be underlying the synchrony patterns. A sharp transition is specially observed at the onset of drowsiness ('Sleep Time').



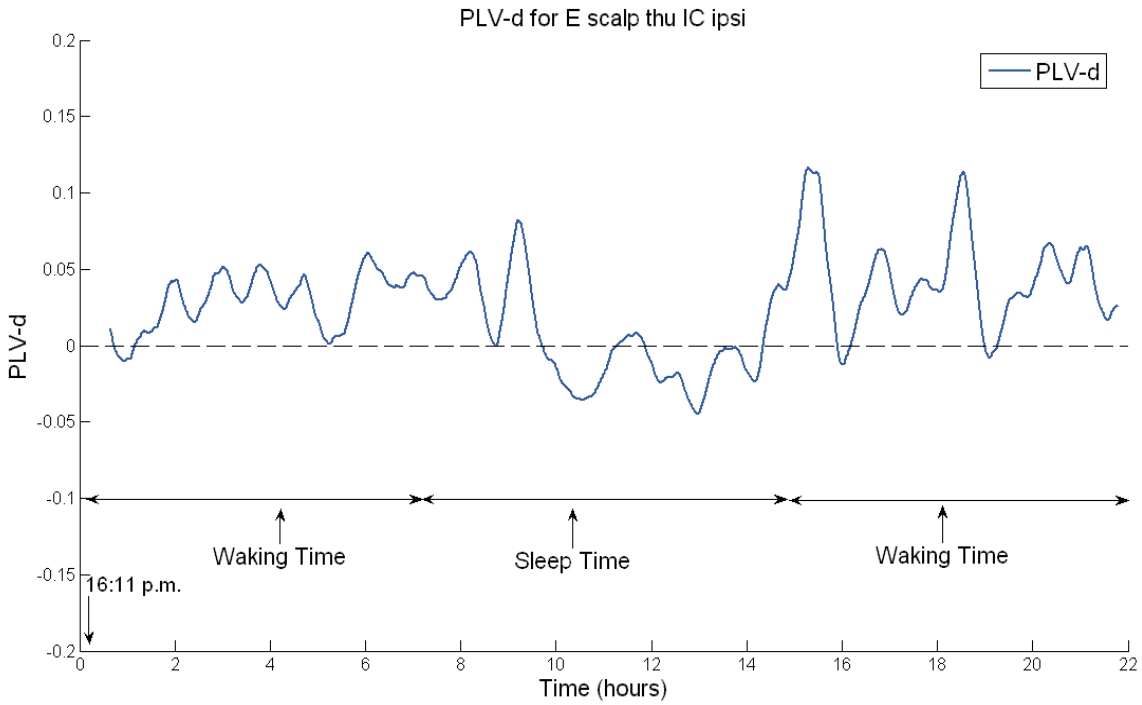


Figure 7.16: these two figures show the PLV-d curves of the next two days of scalp EEG recordings of the patient E scalp. The cyclical pattern is observed in these two days as well. Vertical line marks the seizure onset and the free hand curve shows the observed circadian-like rhythm.

7.3.2 Patient 3 (Patient filename: C Scalp)

Patient History: Patient 3: C scalp is an adult patient with four days continuous scalp EEG records. This patient is on standard drug reduction. Ictal discharge is seen over the left temporal area for all three seizures. Interictal spikes and sharp waves are seen over the left temporal/frontal area during sleep (and rarely during the awake period). This patient has an aura before seizure in the form of a strange pain in the stomach. From the interictal and ictal findings a left sided mTLE is suggested in the patient report and is said to be concordant with imaging abnormality. The patient had all together three seizures on day one and day three, they have been marked on the figures with dotted vertical lines.

Spatial Constraint Templates: The spatial templates used for contralateral and ipsilateral spatial filtering have been shown in Figure 7.17

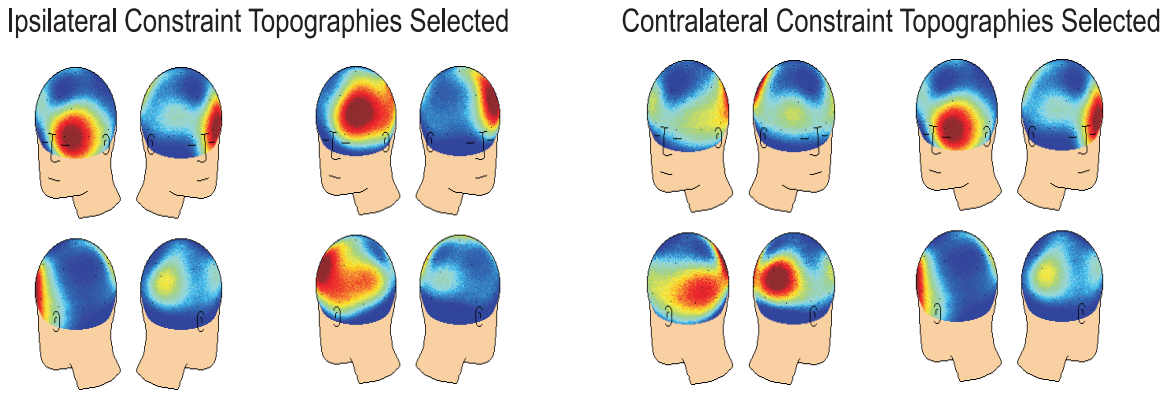


Figure 7.17: These are the ipsilateral and the contralateral pair of topographies that were selected using an ictal segment for spatial filtering for this patients all scalp EEG data.

ICA-PLV for Seizure Detection: The results of seizure detection for one day of C scalp contralateral and ipsilateral LDCs and their corresponding EEG signals have been shown in Figure 7.18. The vertical lines mark the seizure onsets. The plots show no significant patterns with ICA or with EEG signals. The ipsilateral EEG signals, show a constantly high synchrony irrespective of a seizure onset. The specificity may be improved with a trade off with the sensitivity but as the number of seizure recorded for this patient are quite few it is difficult to make conclusions if LDC was helpful in detection for C scalp. Nevertheless, ICA does appear to have removed spurious synchrony in case of ipsilateral signals.

Synchrony Dynamics for Prediction: The PLV-d curves for C scalp were then obtained using the spatial template for contralateral and ipsilateral LDC as well as EEG signals. The PLV-d curves of only the contralateral LDCs have been shown here Figure 7.19, 7.20, 7.21 and 7.22.

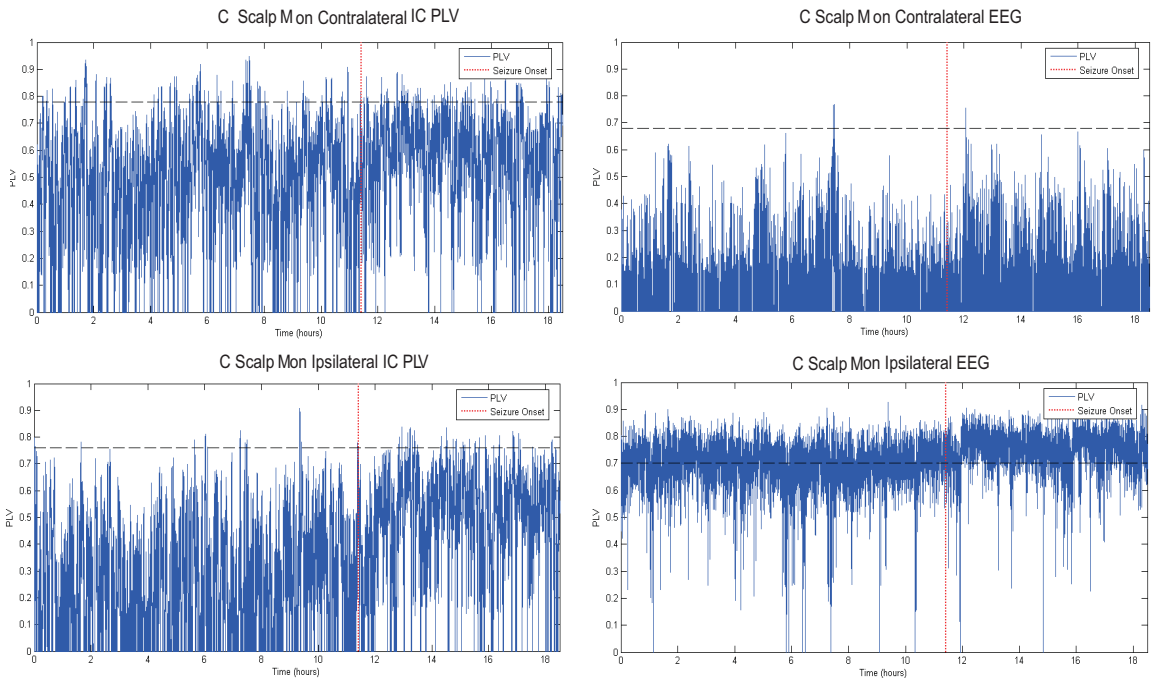


Figure 7.18: The PLS for contralateral and ipsilateral source signals and EEG signals for patient C scalp. The sensitivity (and specificity) of seizure detection may be estimated using a threshold (marked by a horizontal line on the Figures) but it would be arbitarary. The vertical dotted line indicates the seizure onset times marked by expert electroencephalographers at the hospital.

Observations

1. *Circadian-like Rhythm*: The PLV difference curves of LDC signals of patient C scalp shows an interesting underlying cyclical pattern in all four days of this patients data, coincident with the circadian rhythm. These have been marked by free hand curves on the PLV-d graphs to assist in visualizing the rhythm. These are interesting and novel and have not been previously demonstrated. Further analysis of these rhythms may prove beneficial for uncovering more patterns in the EEG of epileptics.

2. *Day and Night Trend*: General observation of the four days PLV-d curves of this patient shows a general trend, which can be interpreted as one frequency dominating in the day, while the trend switches at night time, when the other frequency band synchrony starts to dominate.

3. *Opposing Trend at Seizure Onset*: It also appears that a seizure onset drives the trend away from the 'normal' (positive at day and negative at night) trend (see Figure 7.19 and 7.21). But there are a few other times (marked with a question mark) when such a opposition to normal trend takes place that cannot be explained.

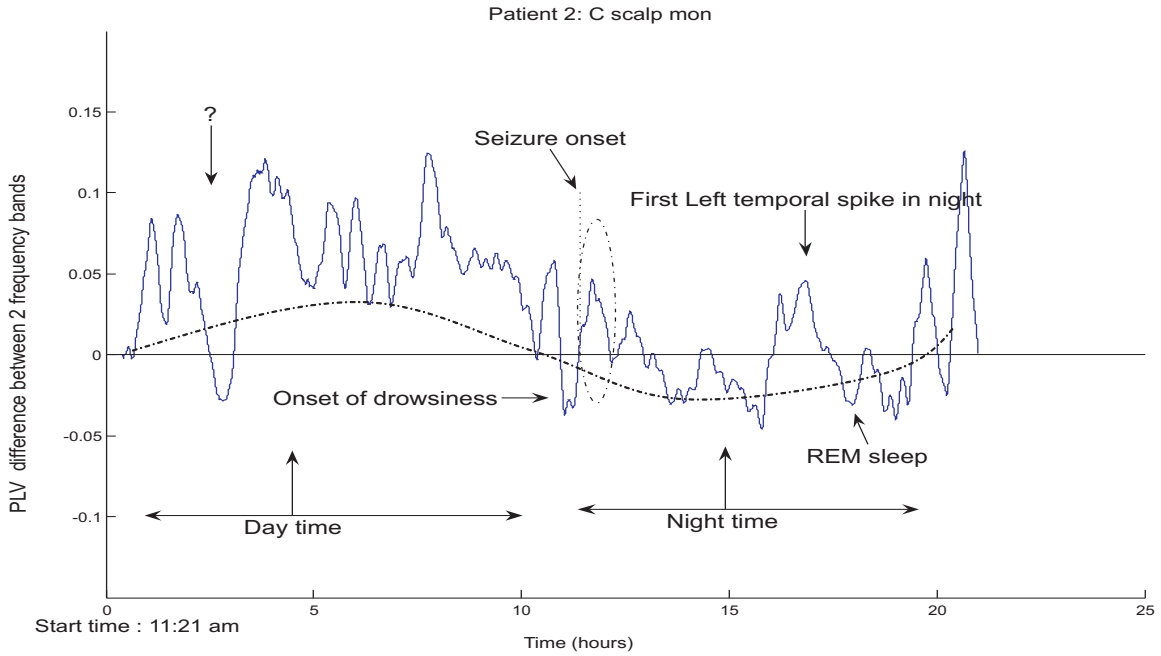


Figure 7.19: **Patient 3, Day 1:** The plot shows the difference of weighted PLV values of two frequency bands 2-8 and 8-14 Hz. The seizure is marked by vertical dotted line. A cyclical pattern seen over 24 hours is marked by a roughly sketched wave. A period when the first focal waves are seen in the night are also marked.

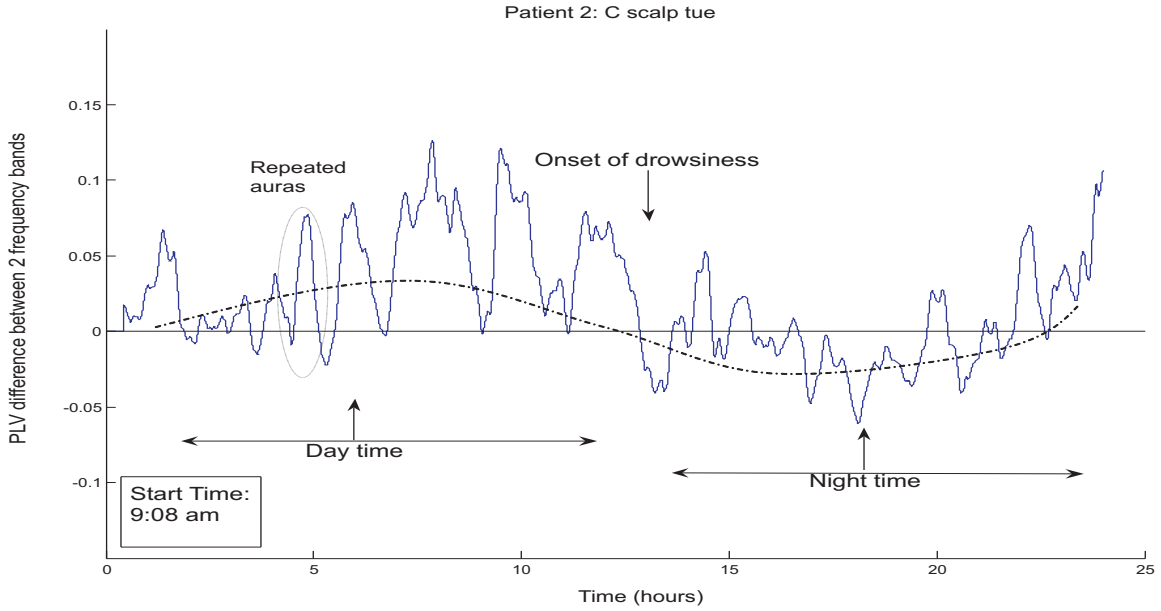


Figure 7.20: **Patient 3, Day 2:** The plot shows the difference of weighted PLV values of two frequency bands 2-8 and 8-14 Hz. The cyclical trend of day and night can be seen here as well with the PLV dominant frequency changing at night time. A period when the patient experienced repeated auras for half an hour has been marked.

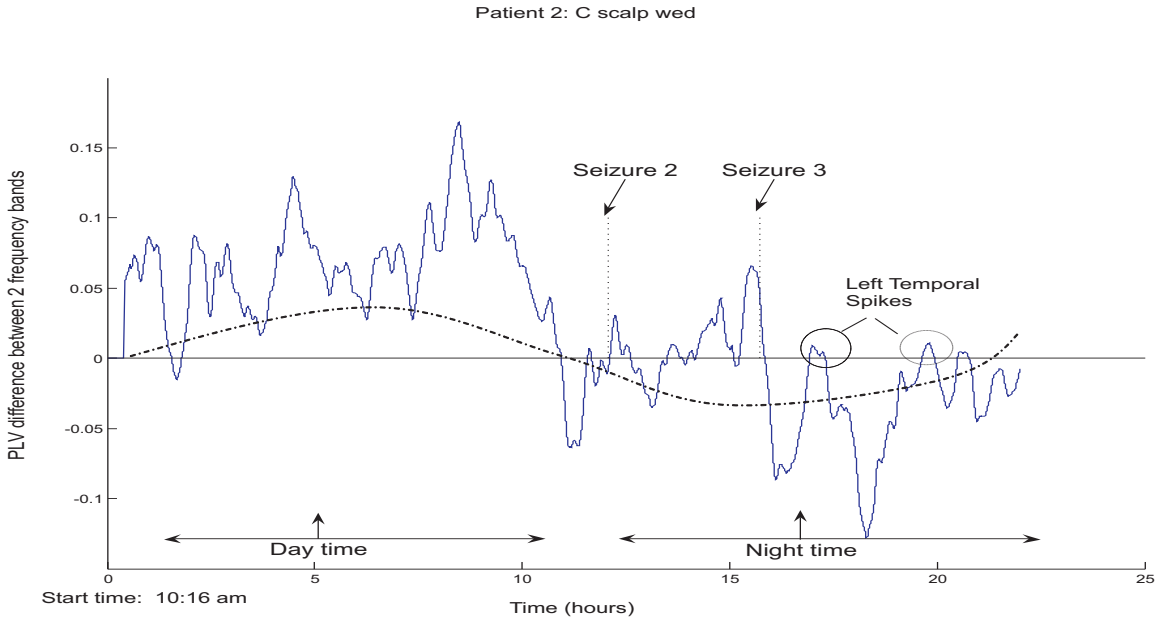


Figure 7.21: **Patient 3, Day 3:** The plot shows the difference of weighted PLV values of two frequency bands 2-8 and 8-14 Hz. The vertical dotted lines shows the seizure onsets. The cyclical trend changing over day and night is seen here as well. Seizure 3 shows the similar pattern of driving the trend to the opposing frequency at seizure onset. It also shows some pre-seizure similar activity.

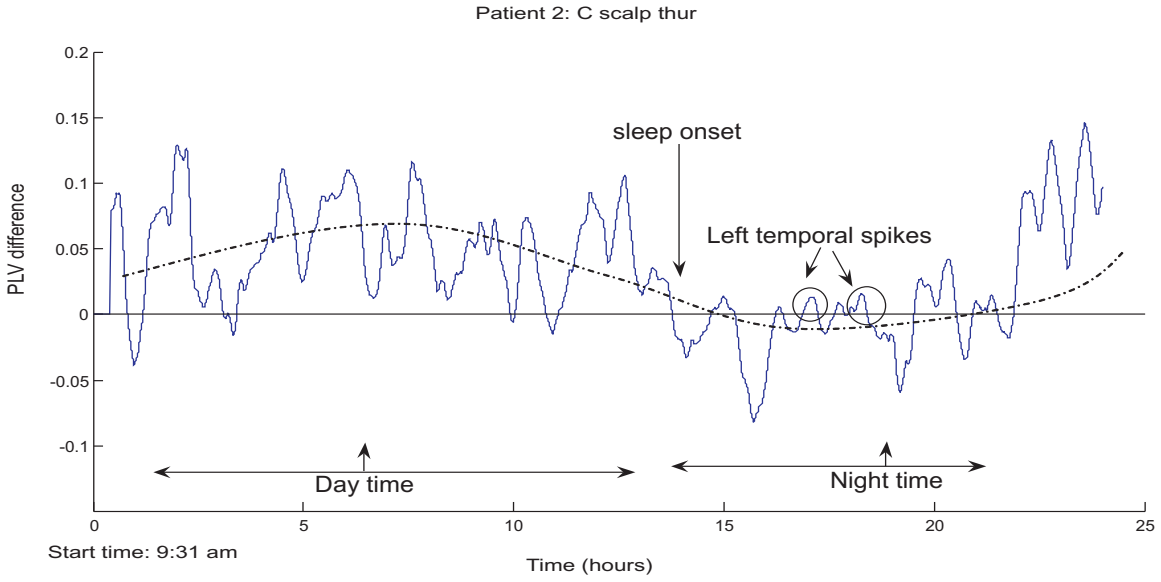


Figure 7.22: **Patient 3, Day 4:** The plot shows the difference of weighted PLV values of two frequency bands 2-8 and 8-14 Hz. No seizure occurred on this day. The day/night cyclical pattern can be seen on this day for the same patient.

However, it is to be noted that this analysis is based on the observation of very few seizures for this patient. It is difficult to make any generalizations in such a case.

7.3.3 Patient 4 (Patient filename: L Scalp)

Patient History: The patient data analyzed here is stored as L Scalp. This patient is an adult with four days of continuous recordings. The patient experiences three seizures on day one. The seizure onset shows rhythmic 3-4 Hz sharp wave activity in the left temporal area which increases in frequency and spreads to frontotemporal areas. Imaging shows abnormality in the right side. The diagnosis is not confirmed.

ICA-PLV for Seizure Detection: The seizure detection for L Scalp data is shown in Figure 7.23. The plots for contralateral and ipsilateral source signals show a sensitivity (and specificity) of 67% (4.1%) and 67% (3.4%) respectively. Whist the corresponding contralateral and ipsilateral EEG signals show a sensitivity (and specificity) of 34% (0.68%) and 67% (0.04%). The EEG ipsilateral signal PLV shows a lack of pattern with a consistently high synchrony over 22 hours. ICA is able to remove that spurious synchrony to some extent by showing a specificity of about 3%, which is quite low but better than that recorded by the EEG signals. The EEG contralateral signals do show a different pattern compared to the ipsilateral signals, but the ICA is able to improve the sensitivity from 34% to 67% and specificity from 0.68% to 4.1%.

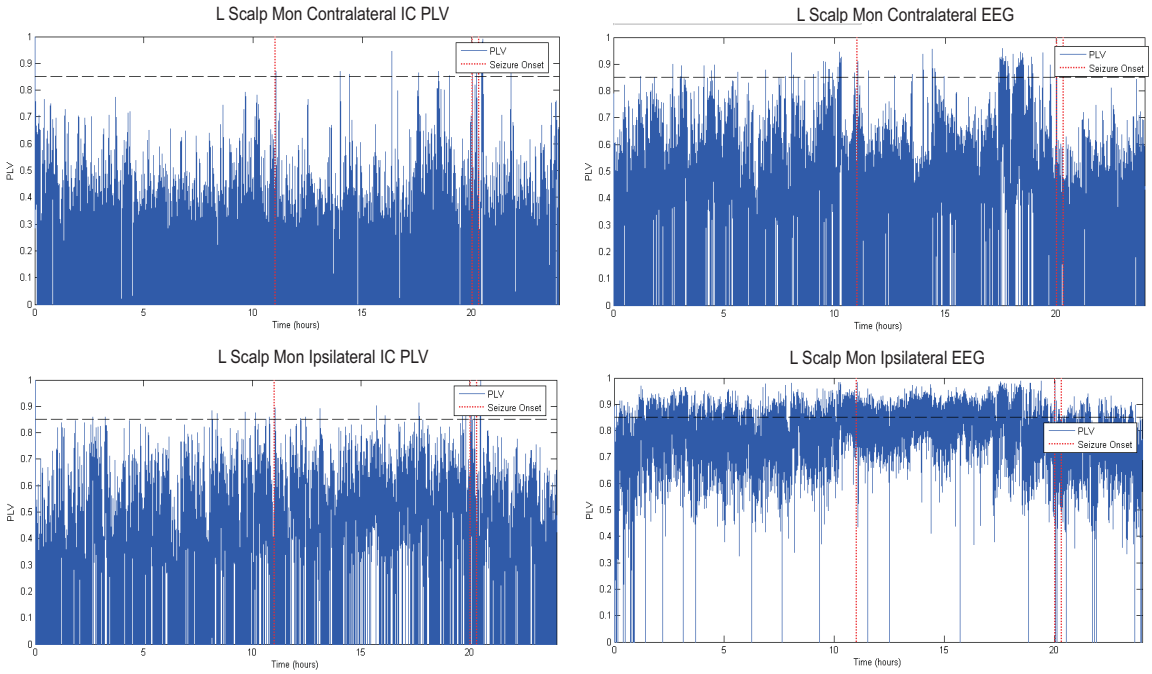


Figure 7.23: The significant PLS for contralateral and ipsilateral source signals and EEG signals is plotted here. The vertical dotted line indicates the seizure onset times marked by expert electroencephalographers at the hospital.

Synchrony Dynamics for Prediction: This patients' data analysis also shows a cyclical trend varying between day and night, with a sudden change in dominant PLV frequency at the onset of drowsiness. The seizures on day one show a similar observation as in previous patients, that the seizure onset is seen to drive the synchrony in the dominant frequency band away from the normal trend.

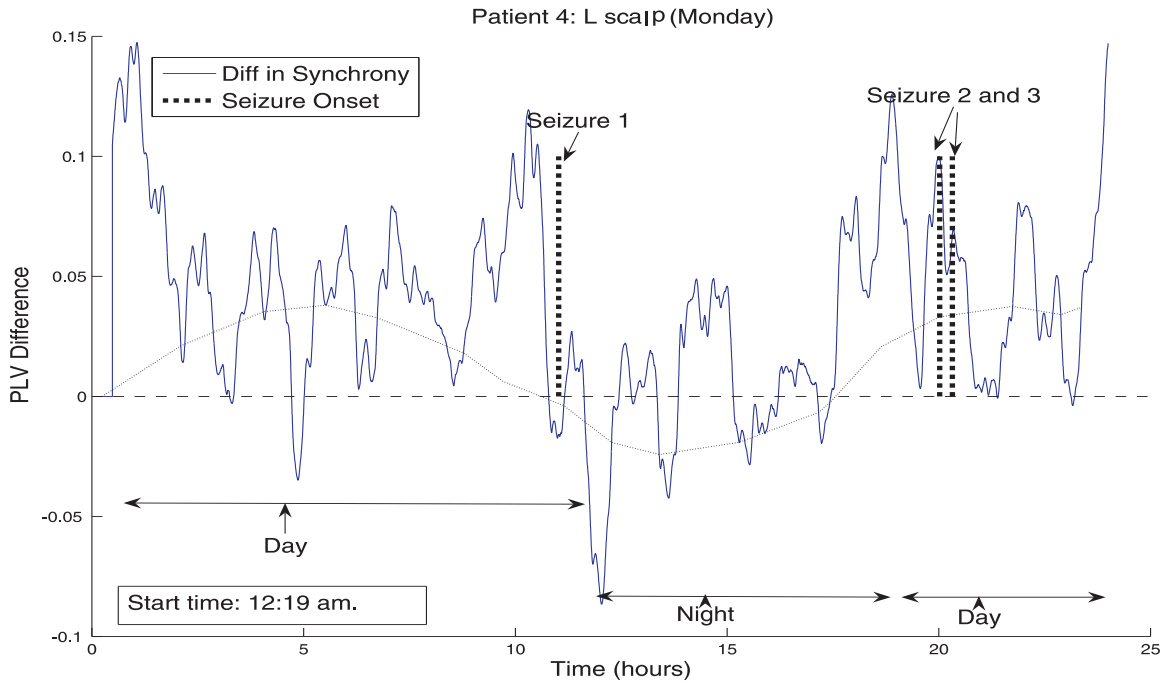


Figure 7.24: **Patient 4, Day 1:** The plot shows the difference of weighted PLV values of two frequency bands 2-8 and 8-14 Hz. The dotted lines indicate seizure onsets. A cyclical pattern over day and night can also be observed in this data.

7.3.4 Patient 5 (Patient filename: X Scalp)

Patient History: Patient 5 is X scalp, an adult patient, with two days continuous scalp EEG recorded for epileptic pre-surgical evaluation. The interictal data did not show any epileptiform activity. The seizures lasted several minutes each with a 'off balance' or a feeling unable to describe. Occasional runs of rhythmic theta (5-6 Hz) were observed in the left fronto temporal area, but were found to occur during or prior to drowsiness. A peaked waveform was seen in Left mid fronto temporal area on one occasion. The patient did not undergo any drug reduction or sleep deprivation during the data collection.

Selecting Spatial Constraints: The patient history did not highlight any specific area of the brain that might be of interest for further analysis. Therefore, the selection of constraints was solely based on the source signal temporal, topographical, mutual information, and phase synchronization information. The source signals selected and the corresponding topographies have been shown in Figure 7.25.

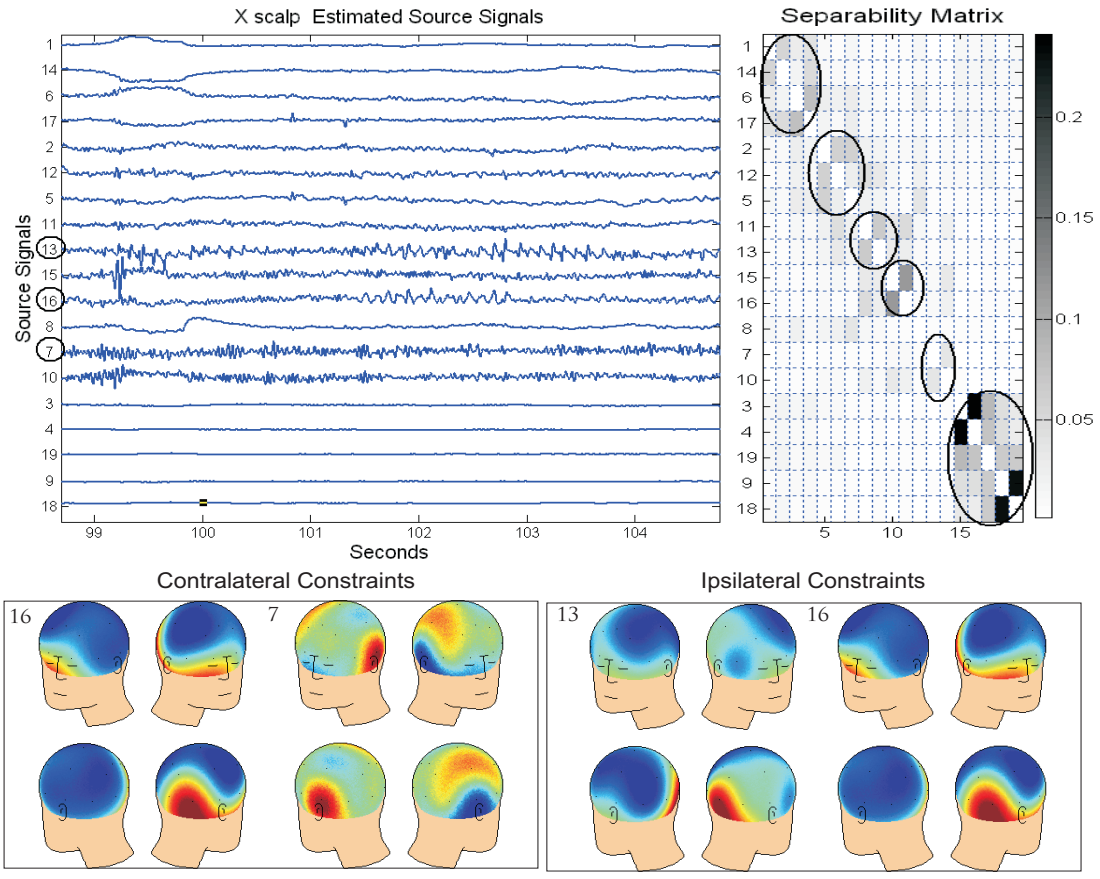


Figure 7.25: The source signals for an ictal segment of X scalp has been shown here. Two sets of rhythmic waves are observed to have been isolated and their topographies are used in combination as constraints. The topographies mapped on the scalp are the contralateral and ipsilateral constraints obtained from the columns of the ICA mixing matrix.

ICA-PLV for Seizure Detection: The X scalp detection results shown in Figure 7.26 based on contralateral and ipsilateral source signals show a sensitivity (and specificity) of 42% (1.6%) and 42% (0.56%) respectively. The corresponding contralateral and ipsilateral EEG signals show sensitivity (and specificity) of 72% (0.19%) and 100% (0.04%) respectively. The specificity is quite low for a useful detection system, nevertheless it highlights that the application of ICA is able to improve the detection of seizures by suppressing spurious synchrony. EEG signals show a much higher synchrony at all times, hiding any existing useful patterns of synchronization.

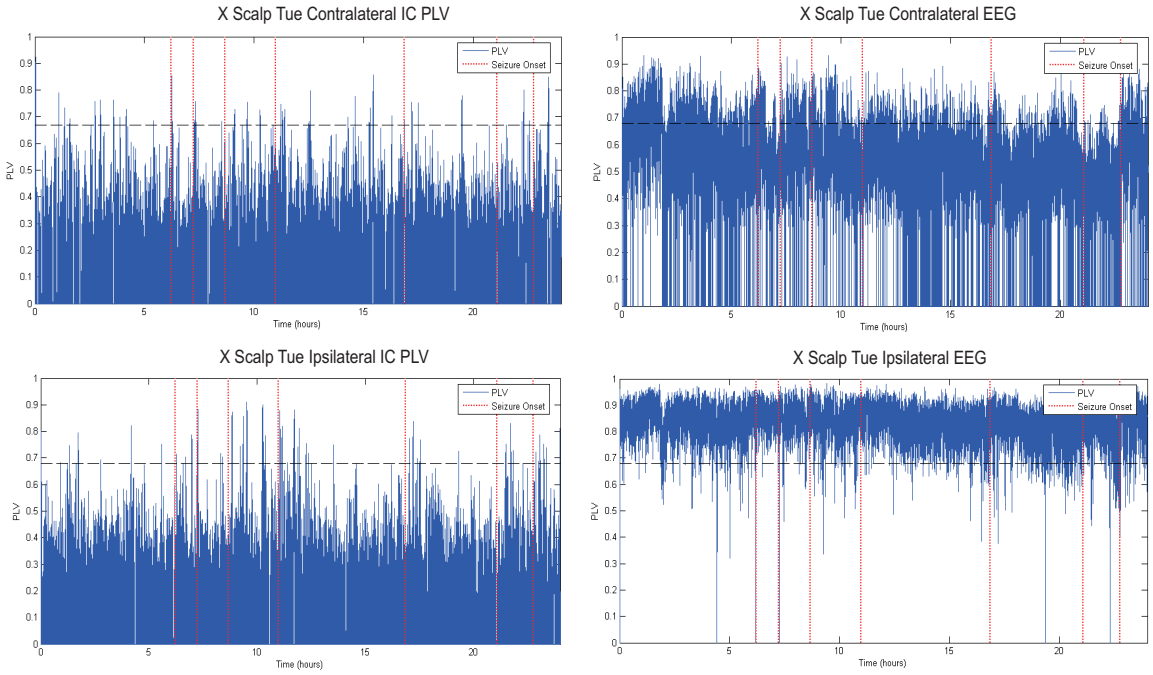


Figure 7.26: The PLS for contralateral and ipsilateral source signals and EEG signals. The vertical dotted line indicates the seizure onset times marked by expert electroencephalographers at the hospital.

Synchrony Dynamics for Prediction: This patient was not confirmed to be having epileptic attacks, based on the lack of evidence from EEG recordings. Nonetheless his EEG and seizures were analyzed as other epileptic data and a similar cyclical pattern and seizures occurring against the trend was observed. The (supposed) circadian-like rhythm is observed to be more prominent on day one as opposed to day two where the seizures appear to have reduced the difference. A sharp transition is observed at the onset of drowsiness on day one which is less marked on day two. Seizures can be observed to pull the PLV-d curve away from the normal circadian-like pattern.

This patient represents a typical case in such research studies. The epileptic disorder being quite broad in its manifestations, it is faced with a very subjective diagnosis, based on a multitude of disciplines and expertise.

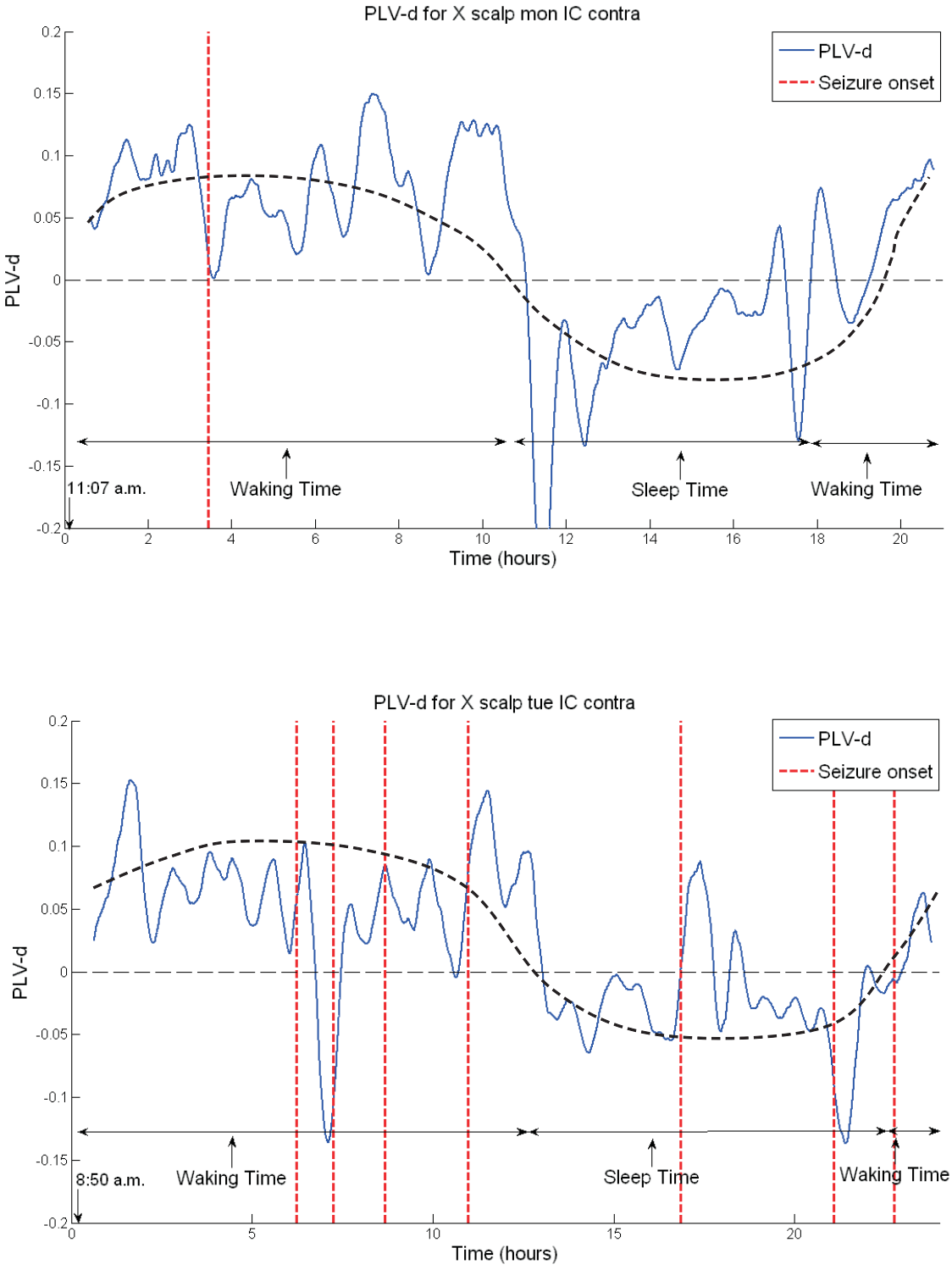


Figure 7.27: The PLV-d curves for the two days of continuous scalp EEG recordings of patient 5 X scalp have been shown here. The seizure onsets have been marked by vertical lines and a free hand curve (black broken) is shown to aid visual appearance of an underlying circadian-like rhythm. The cyclical pattern following a circadian-like rhythm can be observed on both the days. The transitions are less marked on day two that has many seizures. Seizures can be observed to pull the PLV-d curve away from the norm circadian-like pattern.

7.3.5 Patient 6 (Patient filename: Y Scalp)

Patient History: Patient 6 (Y scalp) is an adult patient with four days of continuous scalp EEG data. The patient is not subjected to drug reduction or sleep deprivation. The interictal EEG is mentioned to be abnormal. Focal left frontal epileptiform discharges such as short runs of polyspike, theta and at one time spike/slow wave are seen at F3. The seizures last about 1-5 minutes with odd posturing of arms.

Spatial Constraints: The contralateral and ipsilateral spatial constraints used for Y scalp have been shown in Figure 7.28.

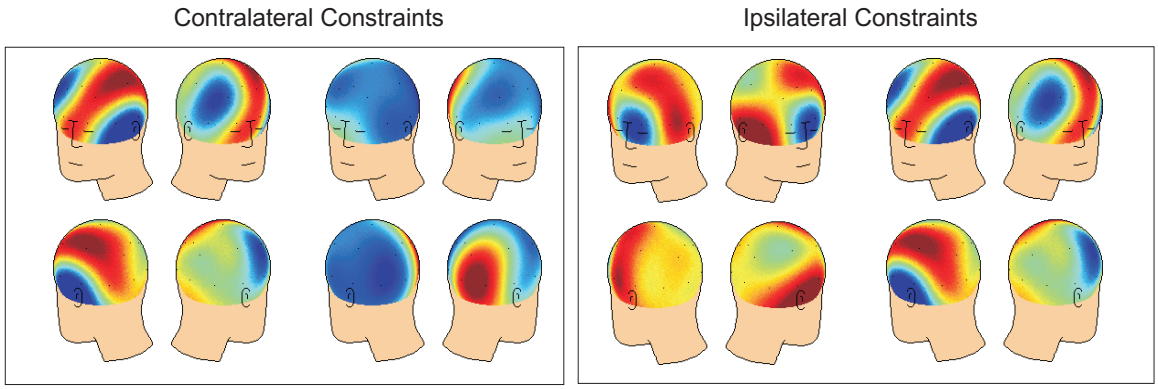


Figure 7.28: The Contralateral and ipsilateral spatial templates used for Y scalp spatially constrained ICA are shown here. These were extracted using an ictal-interictal segment of the patients EEG.

ICA-PLV for Seizure Detection: The seizure detection using contralateral and ipsilateral source signals show a sensitivity (and specificity) of 0% (0%) (no detections) and 67% (1.58%), whilst the contralateral and ipsilateral EEG signals show 34% (0.15%) and 100% (0.04%) sensitivity (specificity) respectively. The PLV plots for these are shown in Figure 7.29. The vertical lines mark the seizure onsets and a threshold is used to estimate the best possible sensitivity and specificity. Clearly, the ipsilateral EEG signals show a high synchrony for the entire day. If any synchrony patterns exist in the data it gets easily masked by spurious synchrony that occurs due to overlapping of signals. This effect is removed by the application of ICA as seen in the ipsilateral plots of PLV of source signals. The detection is not found to be very effective in this patient as the specificity is quite low. A trade off with sensitivity may help improve the specificity or perhaps more seizures and more days data may show an emerging pattern with a better sensitivity and specificity. Nonetheless a pattern appears in the ipsilateral source signal PLV plots and these will be analyzed further for seizure prediction.

Synchrony Dynamics for Prediction: In the PLV-d curves of this patient shown in Figures 7.30 and 7.31 show a prominent cyclical pattern with sharp peaks. An underlying circadian-like rhythm can be observed to exist in the PLV-d curves. Interestingly,

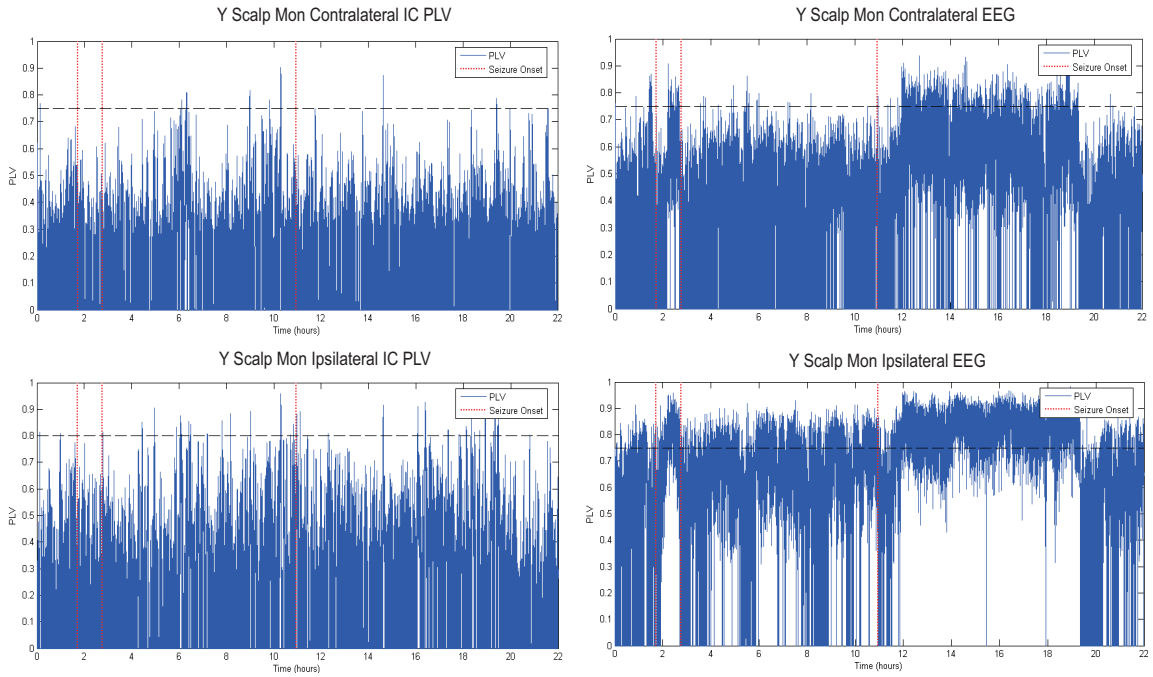


Figure 7.29: The PLS for contralateral and ipsilateral source signals and EEG signals. The vertical dotted line indicates the seizure onset times marked by expert electroencephalographers at the hospital.

in this patients data, the circadian-like rhythm seems to have been upset on the days the seizures occurs. The seizure days have a more undulating curve, with sharp transitions, implying sharp transitions in the dominant frequency in the observed (spatial) area of the brain. A more general rhythm is observed on the last two days, where one frequency dominates during the day time and transitions to the other frequency during the night time. Sharp troughs and peaks in those last two days PLV-d curves are not yet explained. A possibility is a long term artifact spread over an hour such as talking, or some mental task, or perhaps the effect of regular anti epileptic drugs at that time of the day. The first two days also show that an oncoming seizure seems to pull the curve away from the normal trend. It is noted that there are many other similar occurrences with the PLV-d curve moving against the norm, but these are yet to be explained.

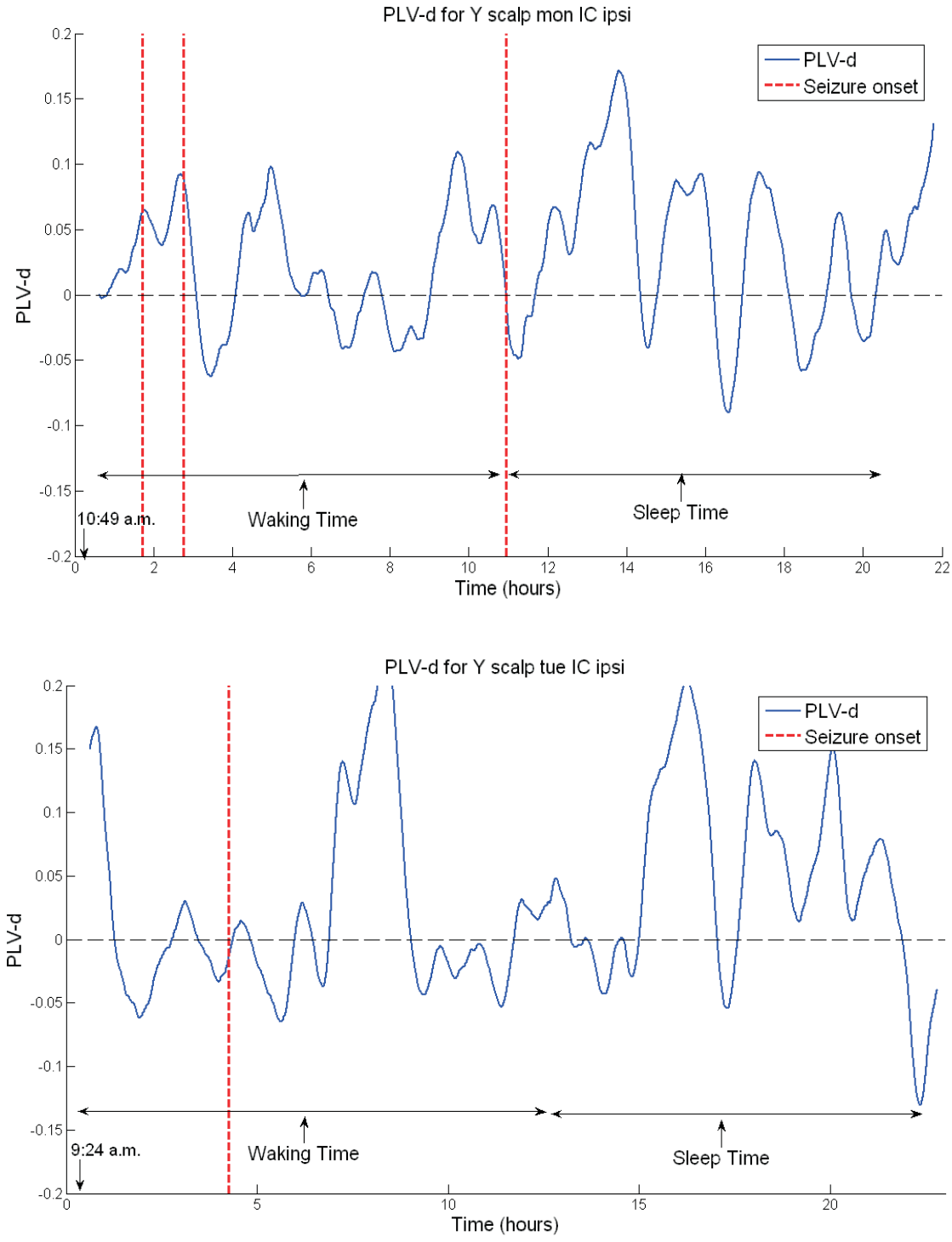


Figure 7.30: The PLV-d curve of one day continuous scalp EEG for patient Y scalp is shown here. The vertical line show the seizure onset. The possible circadian-like rhythm that appears to exist in all PLV-d curves is poorly evident in this particular data. The circadian-like rhythm that was usually seen in other data, seems to have been destroyed in these two days of this patients data, coincidentally the days when the seizures occur.

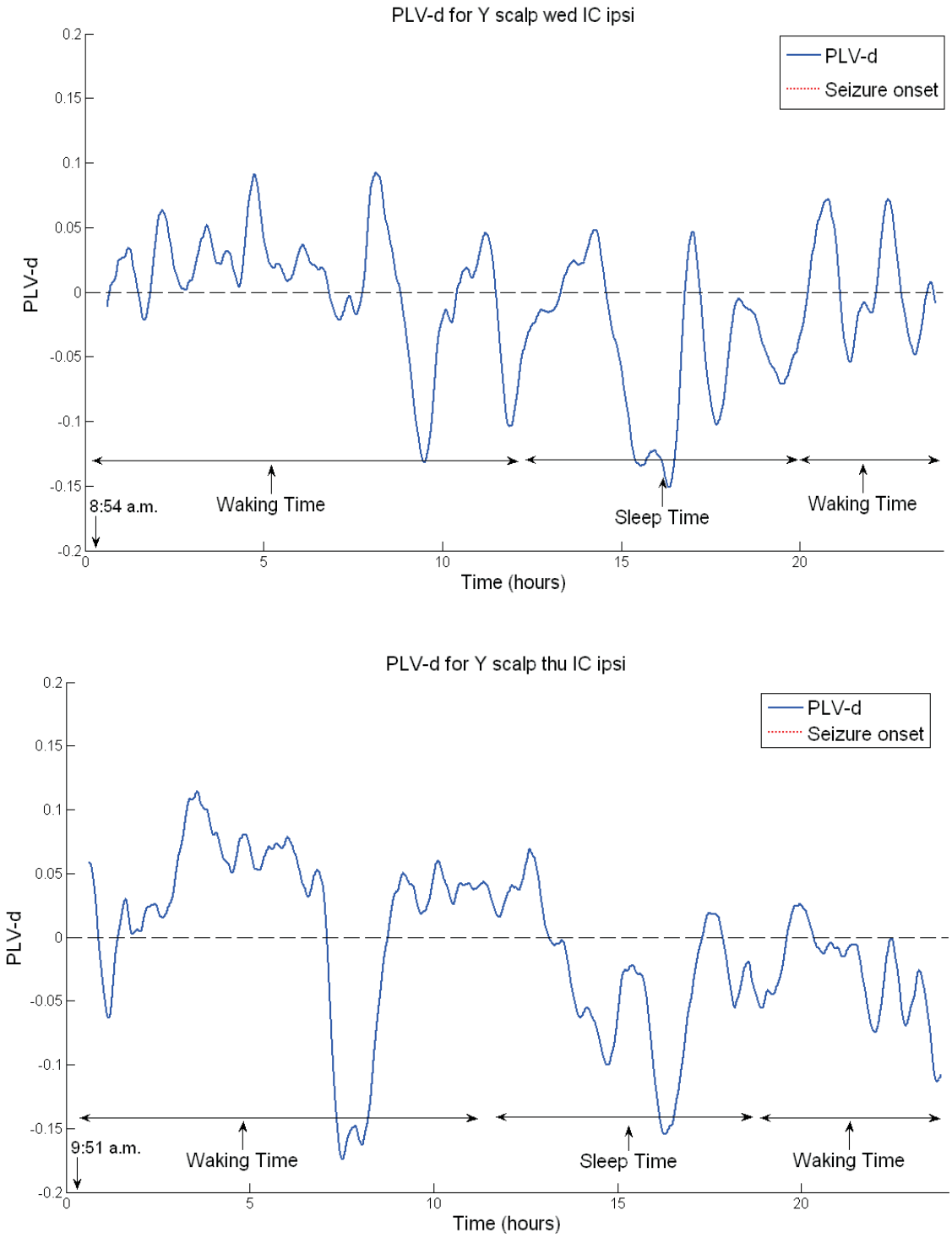


Figure 7.31: The PLV-d curves of the next two days of patient Y scalp are shown here. The seizure onsets are marked by vertical lines. These two days of Y scalp EEG show a circadian-like rhythm as was seen in previous patients, especially the last day data of Y scalp.

7.3.6 Patient 7 (Patient filename: D Scalp)

Patient History: Patient 7 or D scalp has four days continuous scalp EEG recordings, undergoing no drug reduction. The patient is subjected to sleep deprivation. Interictal

EEG is abnormal with frequent right temporal spike/sharp and slow wave epileptiform discharge occurring maximally mid-temporally. Intermittent right temporal slowing of EEG is also seen. The initial ictal EEG has brief bilateral rhythmic theta, then fast activity and then bilateral irregular delta, more in right than left hemisphere. The interictal EEG is in keeping with right mesial Temporal Lobe Epilepsy (mTLE). Ictal EEG is atypical for mTLE showing fast activity at onset. MRI shows right hippocampal sclerosis.

Constraint Selection: The constraint template selection was difficult for this patient due to the few seizure segments. The ictal activity is also not helpful in identification of the seizure source, nor is phase synchrony in this case. It was the combination of morphology, history and mainly mutual information based clustering that was used here to select the spatial template of interest. The constraint templates used have been shown in Figure 7.32. The spatial constraints found do not correlate well with the patient history.

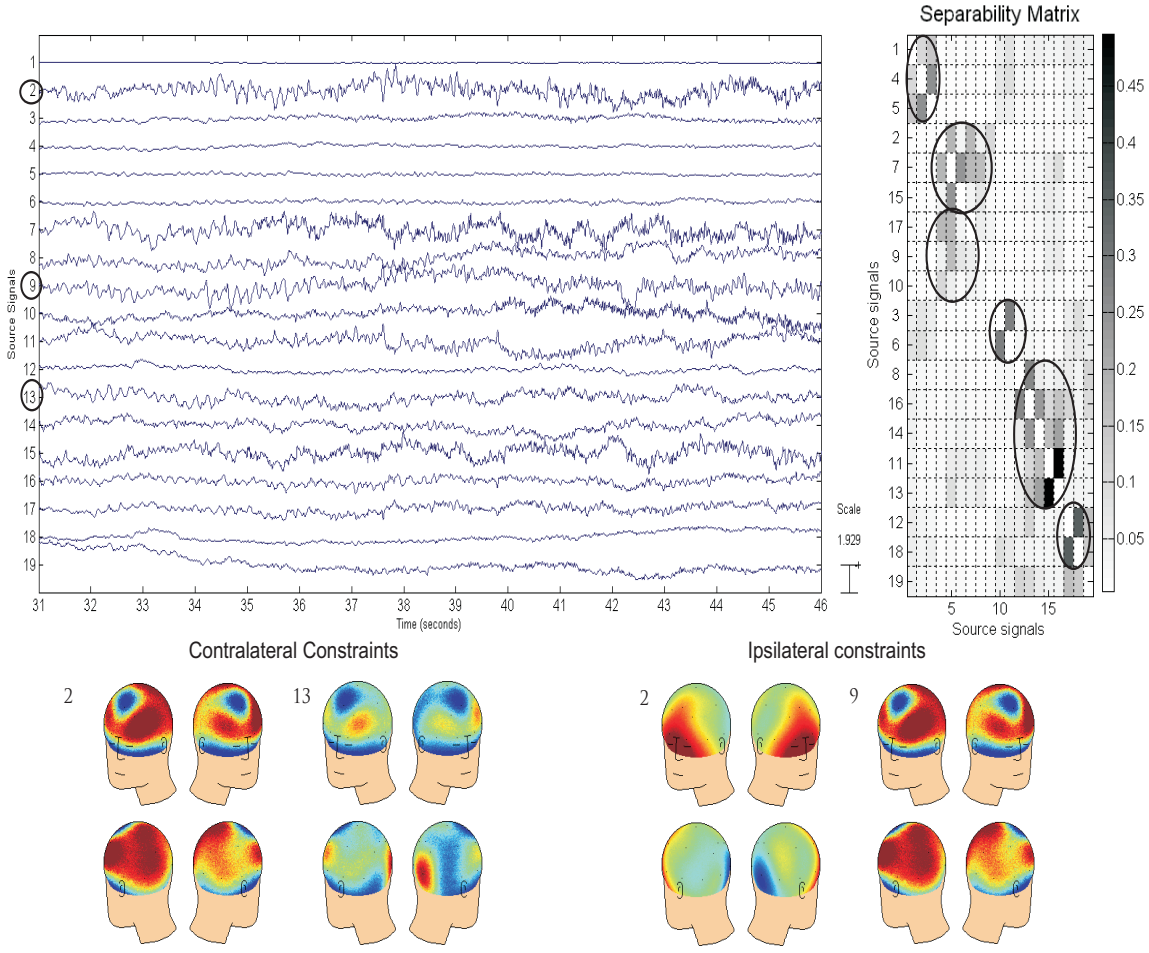


Figure 7.32: The contralateral and the ipsilateral spatial templates for D scalp have been shown here.

ICA-PLV for Seizure Detection: The PLV used for seizure detection with con-

tralateral and ipsilateral source signals for D scalp shows a sensitivity (and specificity) of 100%(0.35%) and 100% (0.82%), whilst with corresponding EEG signals they show nil% (nil%) and 100% (0.92%) respectively. The EEG ipsilateral signals again show a constantly high synchrony that ICA is able to remove to a certain extent as shown in Figure 7.33.

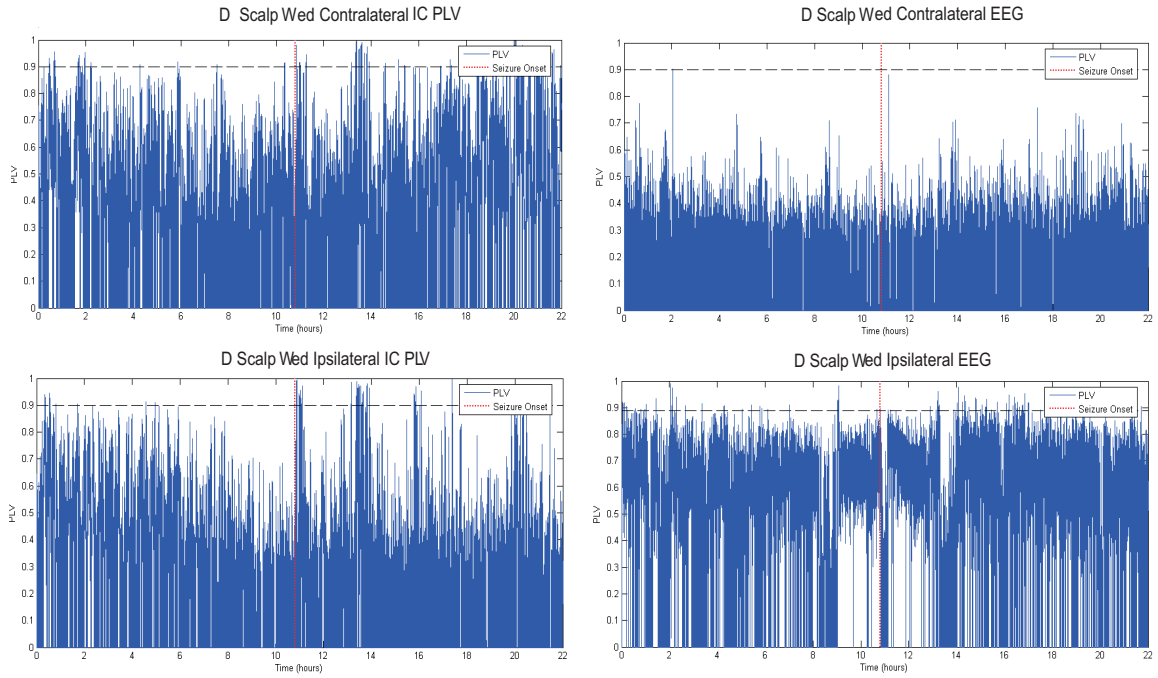
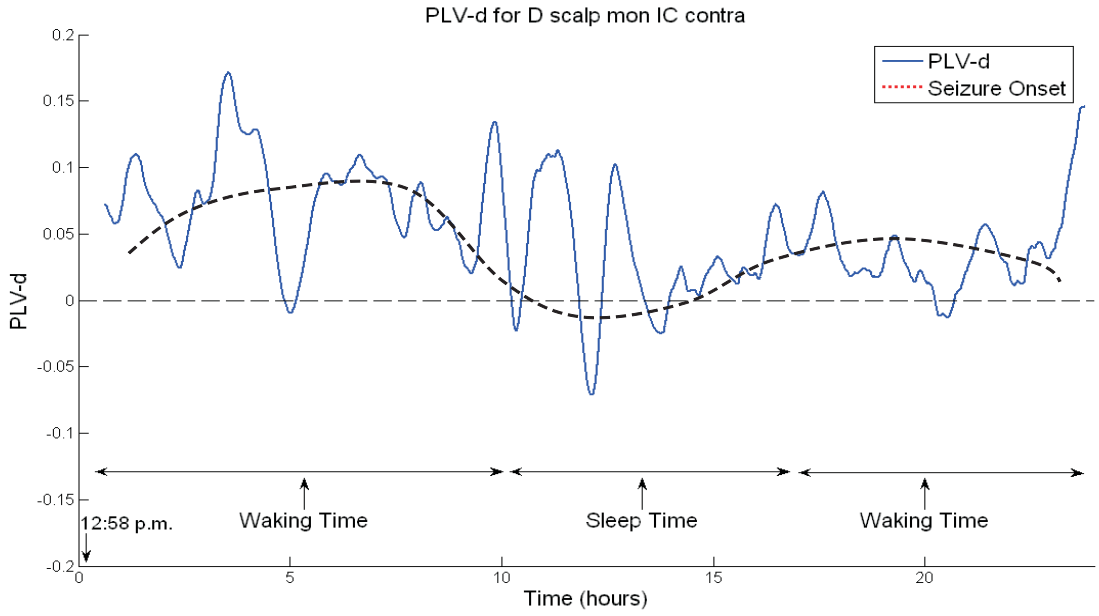


Figure 7.33: The PLS for contralateral and ipsilateral source signals and EEG signals. The vertical dotted line indicates the seizure onset times marked by expert electroencephalographers at the hospital.

Synchrony Dynamics for Prediction: The PLV-d curves for D scalp have been shown for the four days of recordings in Figures 7.3.6, 7.34, 7.3.6 and 7.35. This patients' PLV-d curves are also seen to follow a circadian-like rhythm and a cyclical pattern with peaks and troughs. The circadian-like rhythm is more prominent on days when no seizure occurs. It is also noted that this patient undergoes sleep deprivation. In sleep deprivation the patient is forced to keep awake for more than the usual waking time, depriving him from sleep. The effects are interestingly visible with changes observed in the circadian-like rhythm of day 2, 3 and 4. As was previously seen the usual time of onset of drowsiness is marked by a sharp transition of the dominant frequency on the PLV-d curve. In this case an interesting result is observed: Day 2 shows a general transition in the rhythm at a usual time of drowsiness, but as the patient is forced to keep awake the rhythm is forced back in the other direction and dips again when the patient is finally allowed to sleep at 2:30 a.m., changing again at waking time. On day 3, when the same sleep deprivation is followed, the rhythm does not undergo such sharp transitions and changes when the patient finally is allowed to sleep. On

day 4, once again the rhythm does not fluctuate as much as day 2 and transitions with the new sleeping pattern. It appears the rhythm was getting accustomed to the new pattern on day 2 and has learned the pattern from day 3 onwards. Day 4 EEG data has annotations marking the light and deep sleep that correlate to peaks on the PLV-d curve.



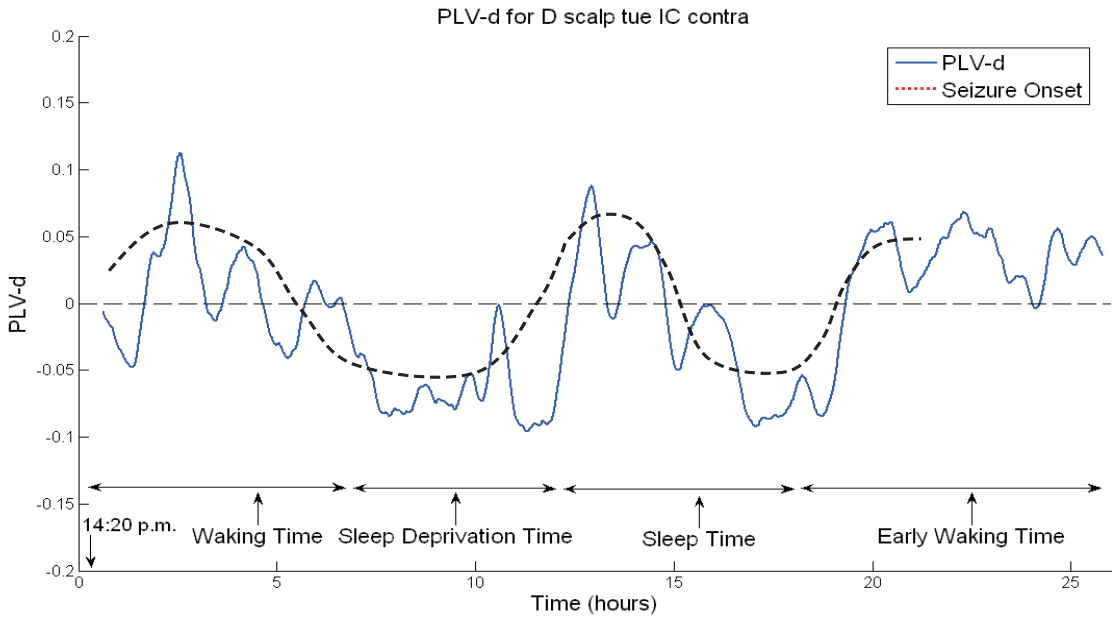
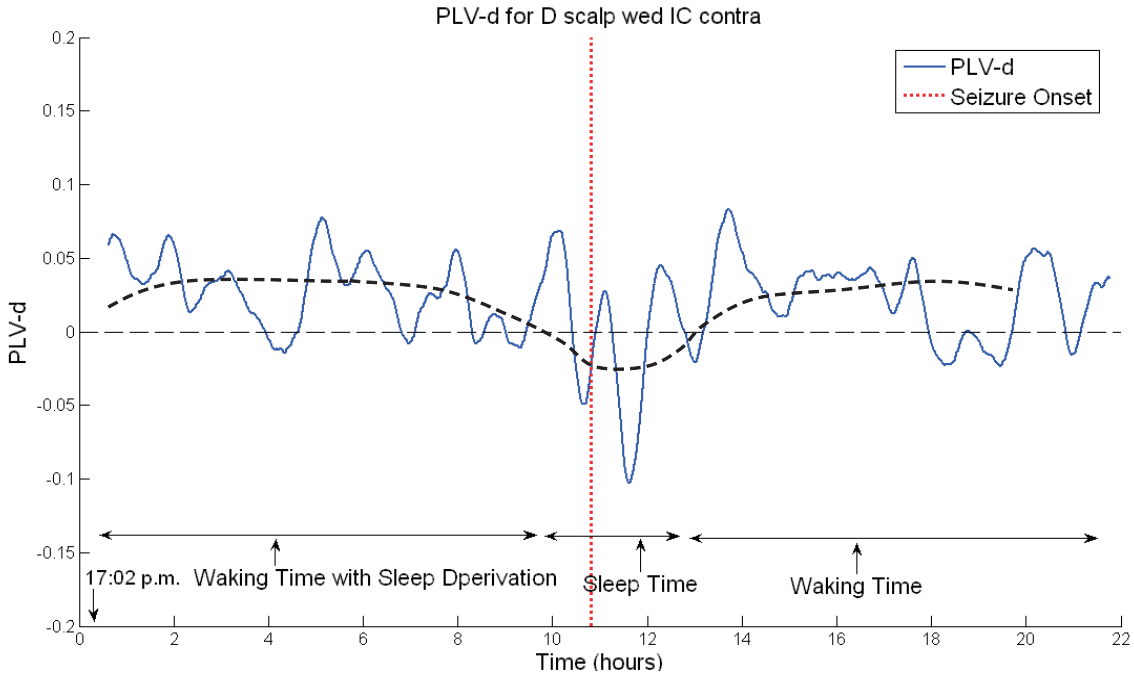


Figure 7.34: The PLV-d curves for two days of source signals are shown for patient D scalp. the vertical lines mark the seizure onset times. A free hand curve indicates the circadian-like rhythm that appears to be present in PLV-d curves. It is more prominent on the days when no seizure occurs. The second figure shows the effect of sleep deprivation. The rhythm is seen to transition at the usual sleep time but as the patient is forced to stay awake it rises and dips again when patient finally sleeps. It appears as if the rhythm is trying to learn the new sleep pattern. The sleep and wake times have been marked using the clinicians annotations in the EEG recordings.



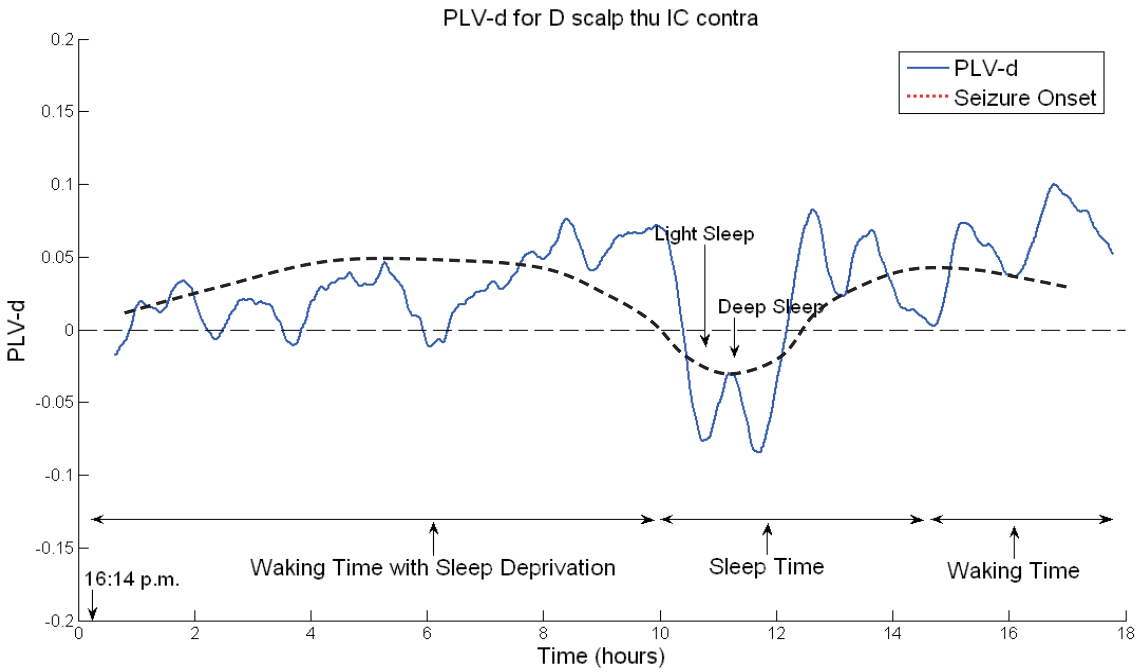


Figure 7.35: The PLV-d curves for the next two days of patient D scalp are shown here. The vertical lines mark the seizure onset and the free hand curve shows an underlying circadian-like rhythm. The figure on top shows the rhythm following the new sleep pattern unlike day 2 and day 3 shows the rhythm appears to be better adjusted to the new sleep pattern. The day with the seizure (top figure) shows a less prominent rhythm. The sleep (deep and light sleep) and wake times have been marked using the clinicians annotations in the EEG recordings.

7.3.7 Patient 8 (Patient filename: O Scalp)

Patient History: This is an adult patient saved as O scalp with four days of continuous scalp EEG recordings. This patient experienced three types of seizures, these have been briefly described below (as per patients report):

Type 1: In type one, a sudden brief extension of left arm followed by patient grabbing left wrist with right hand is observed. The attack lasts only a few seconds. It shows generalized attenuation of the EEG at clinical onset. The type one attack with sudden very brief left arm extension suggests a frontal lobe seizure.

Type 2: This type of seizure involved automatism of both arms, mainly distally with wrist and fine finger movements. It often ended with rapid shaking of left hand. The head turns gently to the right at seizure onset and the patient appears vacant. This type of seizure lasts about 45-60 seconds. It shows generalized attenuation of the EEG at clinical onset and then generalized slowing then occurs with repetitive midline focal epileptiform discharges over vertex. These attacks, with a non-violent manual automatism and mild head turning to right, resemble temporal lobe seizures more than typical frontal lobe attacks but they are briefer than is usual in TLE and the ictal EEG does not suggest TLE.

Type 3: In type 3 there was a stereotyped event observed (patient sits on edge of bed, yawns and legs get briefly extended) followed by a run of epileptiform discharges at the vertex (similar to type 2 seizure discharge). The type 3 attacks are non specific.

This patients' seizures are suggested to be of frontal lobe origin and most likely mesial and from the right, although neither localization or even lateralization is certain from the EEG.

Spatial Constraint Selection: The constraint used for this patient data analysis are shown in Figure 7.36. It should be noted, that the constraint 1 shows a left temporal area and corresponds to a rhythmic wave observed as the estimated source from ICA. It is noted that the patient report does not suggest a left temporal lobe onset.

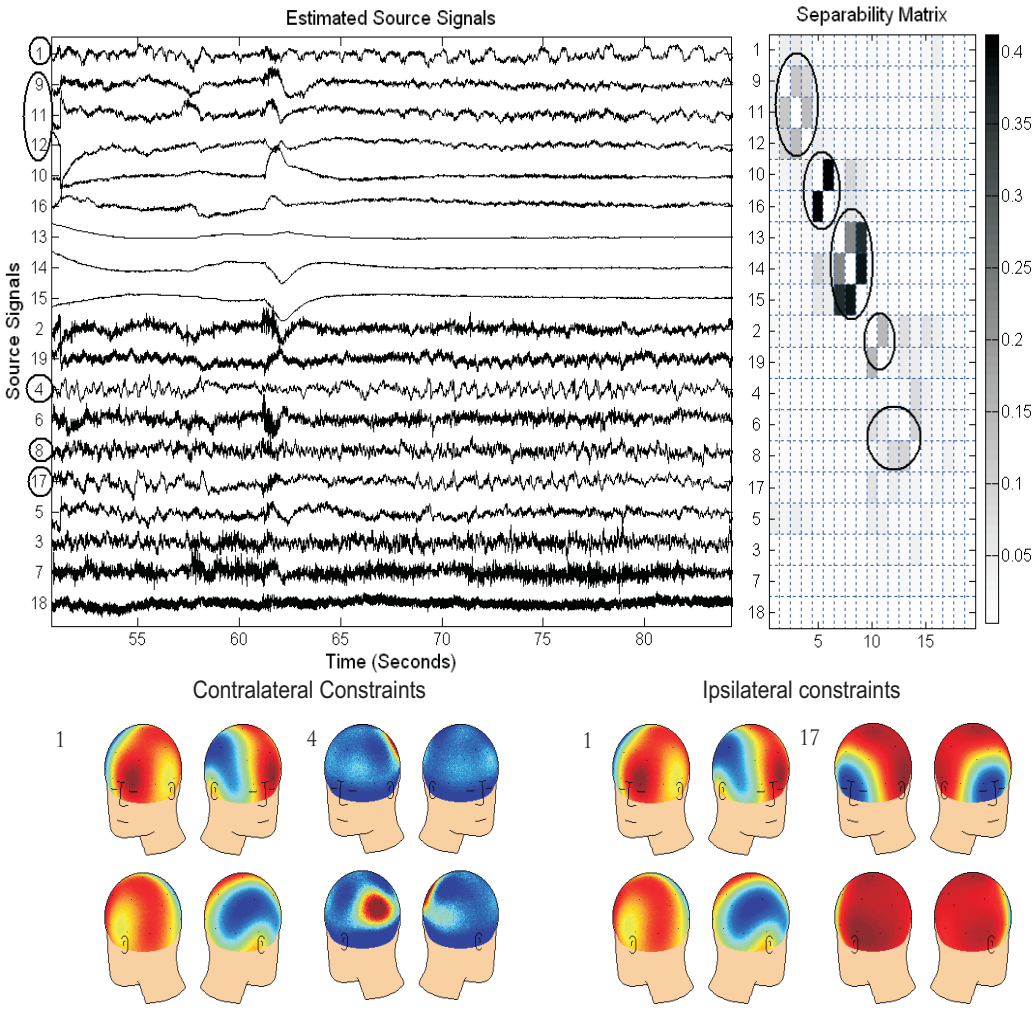


Figure 7.36: **Patient 3** (a) The figure shows the sources estimated by temporal decorrelation for obtaining spatial templates. The encircled sources were selected as probable constraints using information from separability matrix, rhythmic content in the source signals and their spatial topographies. (b) The 2 spatial templates selected to be used as constraints were obtained by the phase synchrony information between the signals.

ICA-PLV for Seizure Detection: The PLV curves for O scalp with contralateral and ipsilateral source signals shown in Figure 7.37 give a sensitivity (and specificity) of 40% (4%) and 30% (2.17%), whilst the contralateral and ipsilateral corresponding EEG signals show 100% (0.13%) and 40% (0.8%) sensitivity (and specificity) respectively. These results are a good example that show that ICA is able to remove the spurious synchronies that overwhelm the EEG signals and are able to improve the specificity (with a trade of sensitivity) in context of seizure detection. Here even the contralateral EEG signals show high synchrony constantly, and the ICA is able to remove by effectively unmixing the signals.

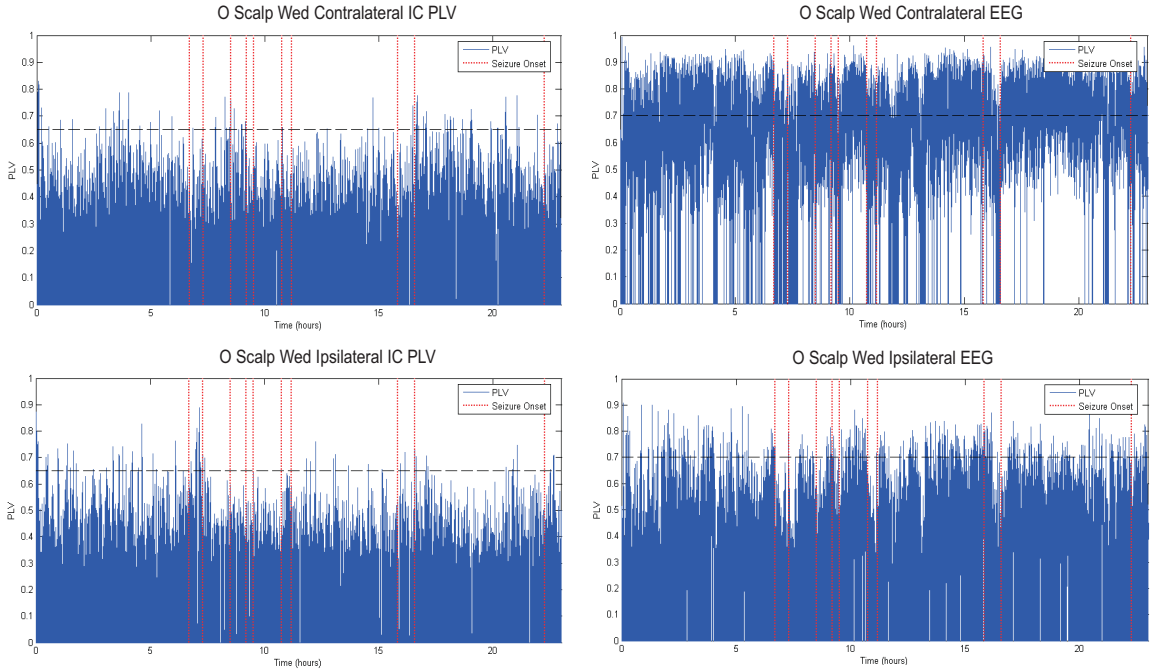
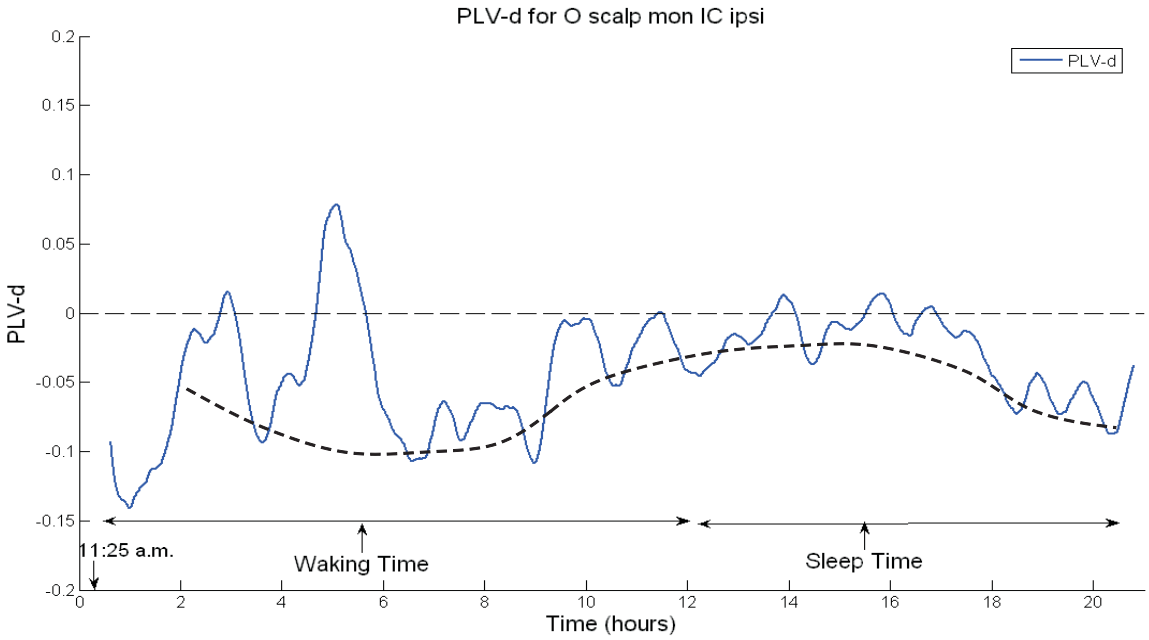


Figure 7.37: The PLS for contralateral and ipsilateral source signals and EEG signals. The vertical dotted line indicates the seizure onset times marked by expert electroencephalographers at the hospital.

Synchrony Dynamics for Prediction: The O scalp patient has a complicated manifestation of epilepsy, having three types of seizures even in the same day. Each type of seizure indicates a unique seizure focus, making it difficult to select one spatial template to represent the seizure focus of the patient. This is overcome by selecting the spatial templates objectively, irrespective of the patient history and only based on the temporal, topographical, phase synchrony and mutual information only. The aim is to locate an area (spatial) of the brain that might indicate early transitions as it prepares itself for an oncoming seizure. These may or may not correlate with the patient history as the clinical diagnosis is based mainly on the ictus, the changes that occur at the time of seizure or the clinical changes that are observed as the seizure onsets. It is quite possible that a completely different area of the brain is involved and it triggers one/many seizure focus(focii) at ictus.

Prominent underlying rhythms similar to circadian-like rhythm are observed in all the days of this patients' PLV-d curves. The first two days shows a very similar, prominent rhythm, but shifted below the zero level unlike previous observations. This implies that the dominant frequency remains within the same frequency band throughout the day and shifts within the band at waking and sleep times. This is unusual as compared to other patients' PLV-d curves. It may perhaps be linked to effects of strong anti-epileptic drugs.

However, Day 3 shows transitions to the above zero level, implying that the dominant frequency is shifting. Co-incidently, this is seen to occur with oncoming seizures on day 3, but not on day 4. Day 4, again shows a similar underlying rhythm as day 1 and day 2 and the seizures recede as well.



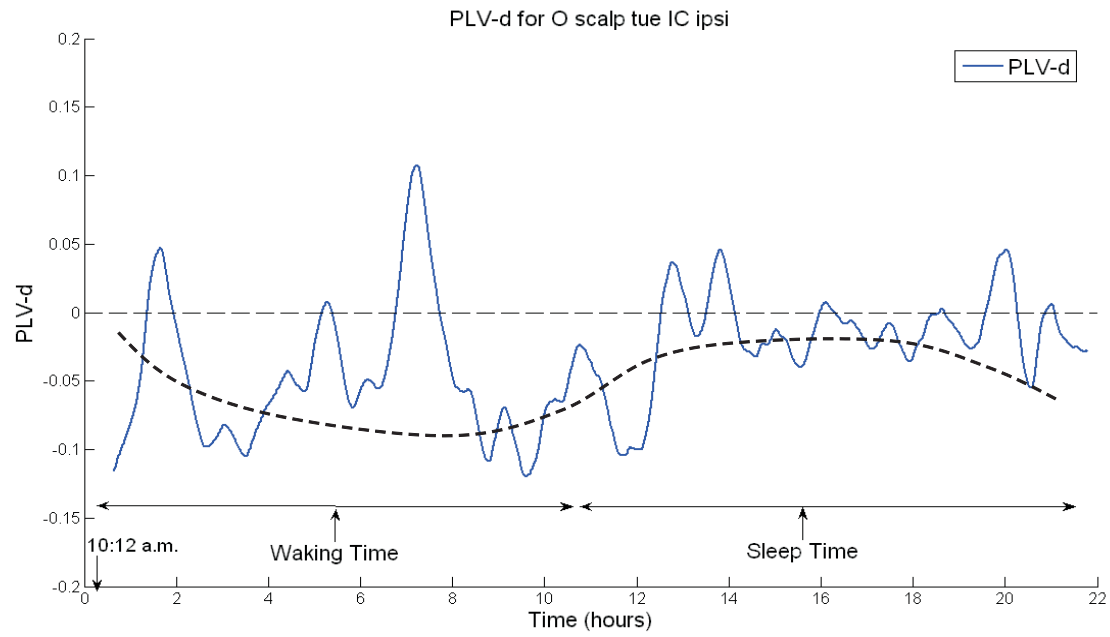
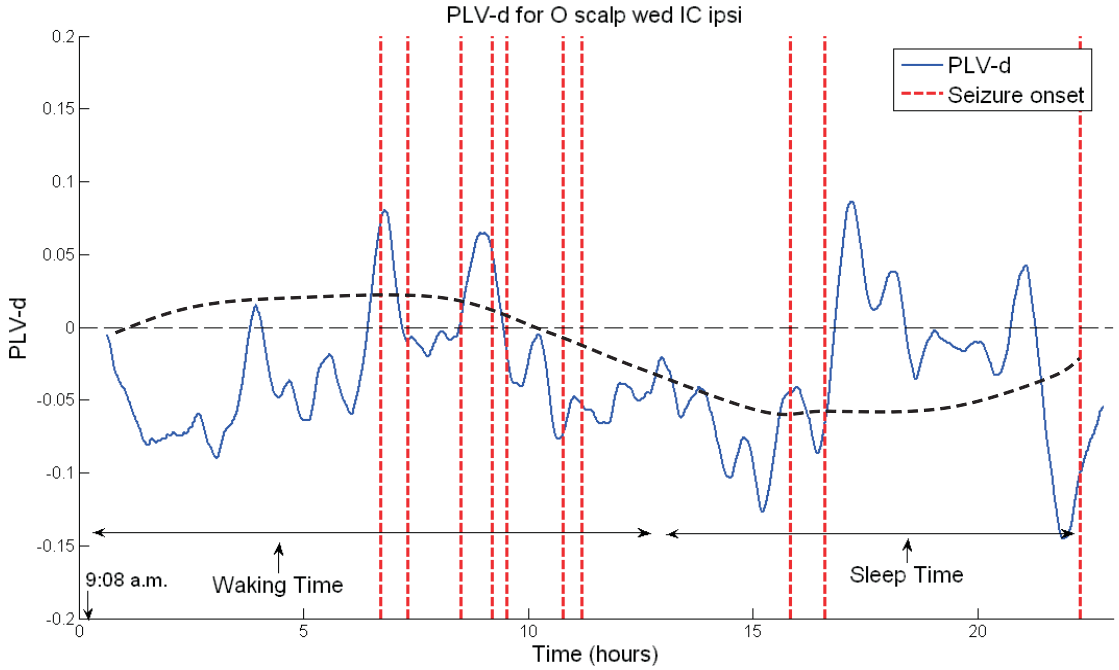


Figure 7.38: **Patient 3, Day 1 and Day 2:** The above two figures show the difference of PLV values of two frequency bands 2-8 and 8-14 Hz for patient O scalp. No seizure occurred on these two days. A free hand curve is drawn to highlight a probable underlying circadian-like rhythm that has been repeatedly observed in previous patients PLV-d curves. The rhythm is seen to remain below zero level unlike previous observations. On day 3 the PLV-d curve is observed to be against the 'normal' probable circadian-like rhythm during sleep time.



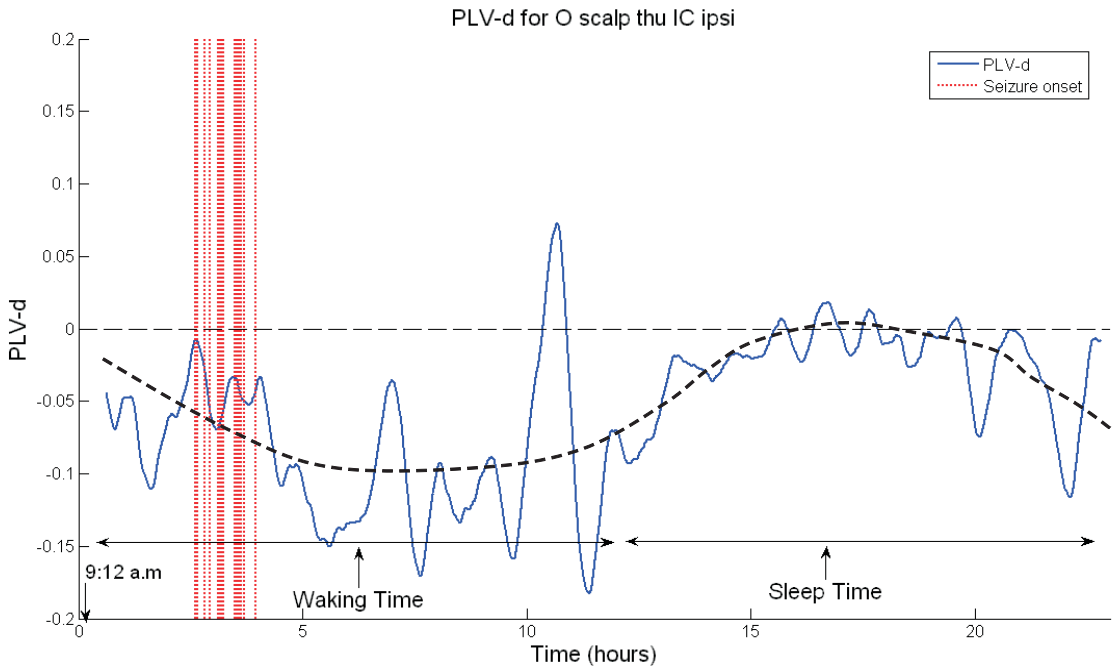


Figure 7.39: **Patient 3, Day 3 and Day 4:** The plot shows the difference of weighted PLV values of two frequency bands 2-8 and 8-14 Hz for patient O scalp. The dotted vertical lines indicate seizure onsets. A free hand curve indicates a probable circadian-like rhythm. Day 3 shows the rhythm moving above zero level apparently with oncoming seizures. As the seizures recede on Day 4, the rhythm similar to that seen on day 1 and day 2 is observed again.

7.3.8 Patient 9 (Patient filename: BB Scalp)

BB scalp **Patient History:** Patient 9 or BB scalp is an adult patient, with two days continuous scalp EEG data. The patient does not undergo any drug reduction or sleep deprivation. Four definite seizures are observed during the first day with 24 hours seizure free period. Clinically the seizures are consistent with frontal lobe seizures and in particular the semiology is suggestive of mesial frontal lobe involvement. However, scalp EEG has not revealed much with interictal/ictal abnormalities, which is common in frontal lobe seizures. There are couple of interictal sharp waves on the right but are not conclusive. Extensive muscle and movement artifacts occur with all four attacks.

Selection of spatial constraint: The selection of constraints in this patients case was not trivial as the ictal and interictal EEG did not show much abnormalities. The selection was based on phase synchronization, mutual information, patient history, topographical and minimally on temporal information. The constraint templates are shown in Figure 7.40.

ICA-PLV for Seizure Detection: The PLV curves for BB scalp for seizure detection are shown in Figure 7.41. The contralateral and ipsilateral source signals show a sensitivity (and specificity) of 75% (2%) and 100% (2.58%) respectively. The corre-

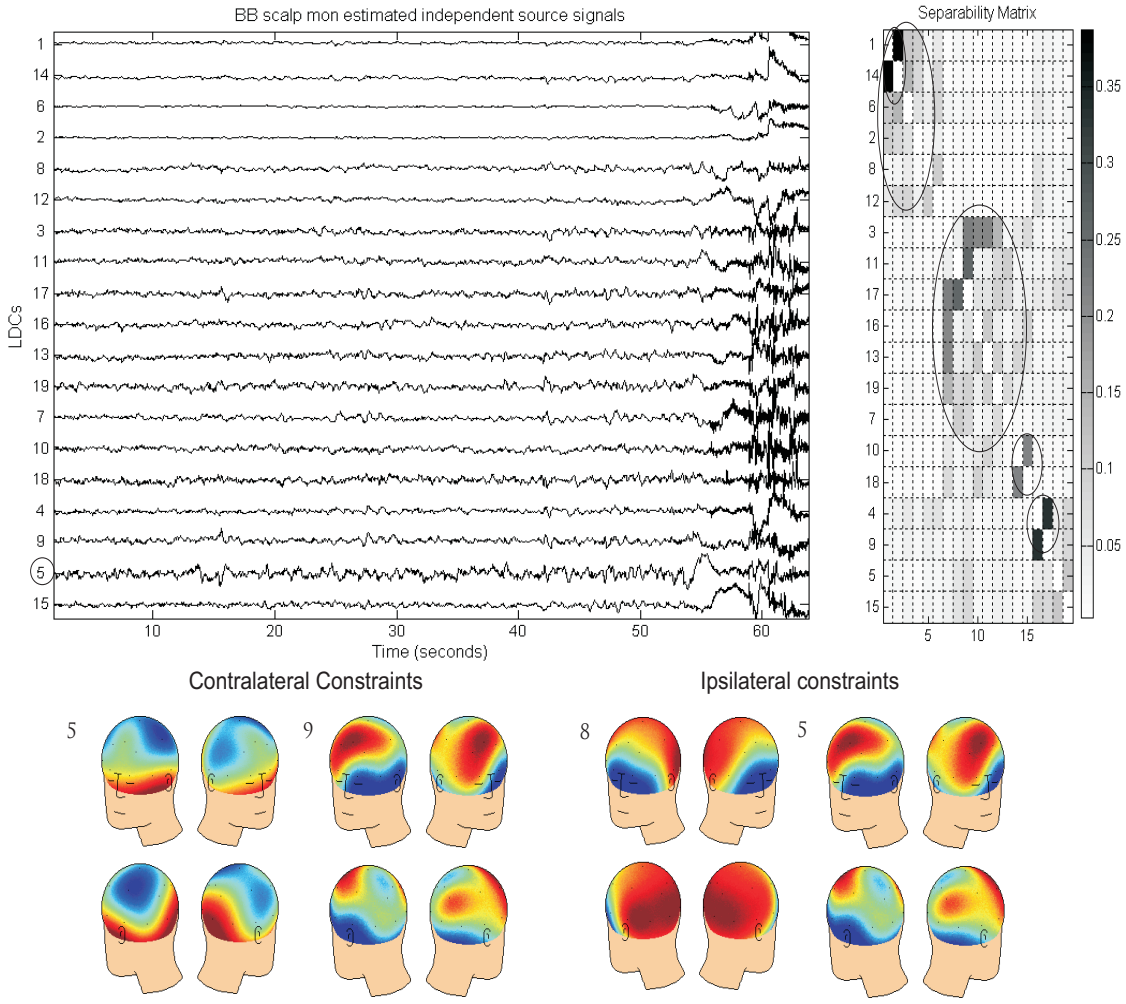


Figure 7.40: The figure shows the sources estimated by temporal decorrelation for obtaining spatial templates. It can be seen that the temporal information is not very useful in this case as the typical rhythmic abnormalities of a seizure can not be seen in the signals. The constraint sources were selected as probable constraints using information from separability matrix, patient history, phase synchronization and their spatial topographies. The spatial templates selected are shown as the topographies.

sponding EEG contralateral and ipsilateral signals show a sensitivity (specificity) of nil (nil) and 100%(0.06%). Clearly the ICA has improved the seizure detection capability in this case. The high spurious synchrony in ipsilateral EEG signals has been removed, improving the specificity.

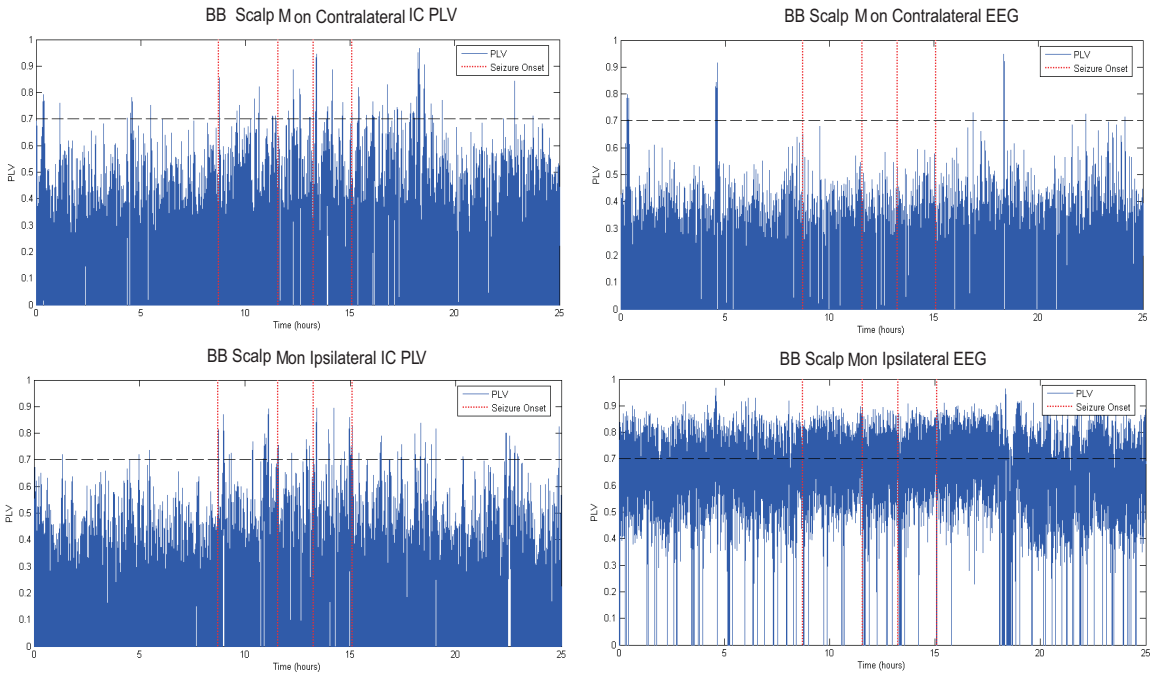


Figure 7.41: The PLS for contralateral and ipsilateral source signals and EEG signals. The vertical dotted line indicates the seizure onset times marked by expert electroencephalographers at the hospital.

Synchrony Dynamics for Prediction: The PLV-d curves for BB scalp also show an underlying circadian-like rhythm and a cyclical pattern. Day 1 shows a rhythm and also it appears that the seizure tends to pull the trend away from the norm. Day 2 PLV-d curve shows a prominent cyclical pattern in absence of a seizure, though it remains on one side of the zero level. This implies that the dominant frequency is in one frequency band most of the time.

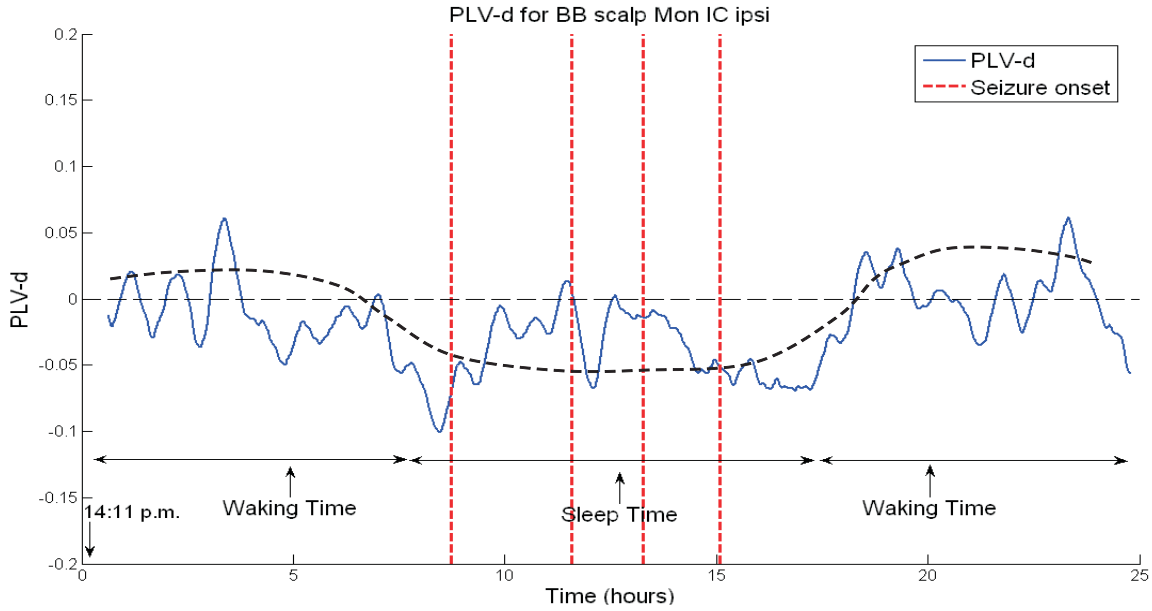


Figure 7.42: The PLV-d curves for day 2 of continuous data of BB scalp is shown here. The seizure onsets are indicated by vertical lines. A free-hand curve is drawn to show a probable circadian-like rhythm underlying the synchrony patterns. It is seen here that the rhythm is coincident with the waking and sleep time, and the oncoming seizures appear to pull the trend away from the norm.

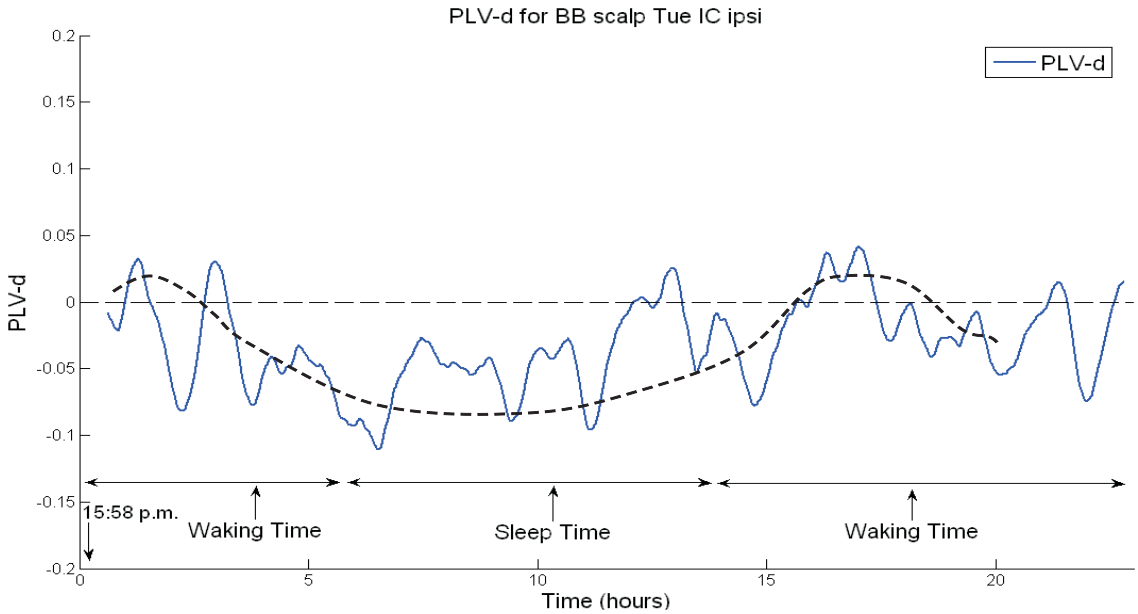


Figure 7.43: The PLV-d curves for day 2 of continuous data of BB scalp is shown here. The vertical lines mark the seizure onsets. A free-hand curve is drawn to show a probable circadian-like rhythm underlying the synchrony patterns. It can be seen here that a general circadian-like trend appears to exist but remains mainly below zero level.

7.4 Discussion: Seizure Detection in Long Term Continuous EEG

The seizure detection results were previously found promising when used on source signals of short EEG segments, in terms of sensitivity. However with short segments it is not possible to analyze the specificity of the detection. The ICA-PLV techniques were then applied to much longer realistic segments for detection. They were applied to source signals and EEG signals. The source signals were obtained from spatially constrained ICA applied to multichannel EEG. The spatial constraints used were obtained by using phase synchronization, separability matrices, temporal and topographic information of the signals. In this way a pair of contralateral and a pair of ipsilateral source signals were obtained. The corresponding spatial signals of the EEG were used for EEG PLV analysis.

Thus, four sets of data for each patient that were used were:

1. Contralateral Source Signals
2. Ipsilateral Source Signals
3. Contralateral EEG Signals
4. Ipsilateral EEG Signals

It was seen that in general for all four data types across patients, long data sets show good sensitivity but a poor specificity. The sensitivity and specificity are improved when source signals are used instead of the EEG signals. Especially for ipsilateral signals, where EEG signals are observed to show constantly high synchrony, which can be explained as the effect of overlapping of signals due to effect of volume conduction. ICA removes the spurious synchrony form ipsilateral EEG signals, improving the specificity many times. For contralateral signals, ICA shows improvement in sensitivity and specificity, though not significant enough for a clinical application.

Low Specificity: The low specificity of seizure detection in source signals is in other words a high false positive rate. A general positive alarm on a ICA-PLV curve occurs when the significant phase synchrony in the analyzed spatial region(s) of the brain, in the frequency band being investigated, crosses an arbitrary threshold. It is considered a detection if it matches the seizure onset time. Some of the reasons that can lead to increased false positive alarms, are mentioned below:

- *Detection window:* It should be noted that a PLV is calculated as an average change of synchrony over a window of 1000 samples which is equivalent to 5 seconds on the EEG time series. If the synchrony change related to a seizure

occurs much slower, it will cause false positives over many ICA-PLV analyzed windows for detection. These can be possibly handled by having a window/period of detection instead of a single time point. This essentially motivates the use of ICA-PLV for seizure prediction.

- *Artifacts:* A positive value on a ICA-PLV curve occurs in a particular spatial region in a particular frequency band, because of increased phase synchrony over 5 seconds. This change takes place at the time of a seizure as shown by the high sensitivity rates across patients. Nevertheless, such an increased phase synchrony for 5 seconds, in a specific frequency band, may also take place due to artifacts such as

  - Chewing: A very rhythmic muscle artifact that dominates the EEG signals, especially the temporal channels. The frequency of chewing lies within the frequency band being analyzed.
  - Sleep Spindles: Sleep spindles occur when a person is in deep sleep. These are very rhythmic waves that dominate and spread in all channels over 1-2 seconds. An example is shown in Figure 7.9. Such waves are easily picked up by the PLV analysis as synchronized activity causing false alarms.
  - Focal Rhythmic waves: Focal rhythmic waves were also commonly observed in the EEG data of epileptic patients. The reason for these waves is not clearly known to clinicians and these phenomenon seemed to surge and disappear abruptly. Many times they lie in the spatial area of interest and hence interfere with the ICA-PLV analysis, causing false alarms.
  - Physical artifacts: A disruption in the apparatus link, electrode pop, etc. is common when a patient has the EEG electrodes attached for 4-5 days continuously. This sudden disruption can introduce a sudden synchrony as all channels get affected, for example in the case of a referential electrode montage.

- *Data Collected:* The increased number of false positives in this data set may also be due to the way the data is collected. It should be noted that the EEG data is collected from patients undergoing pre-surgical evaluation. In a short time of 4-5 days the clinicians aim to collect the maximum information, for the benefit of the patient. In this time, the patient is allowed to have maximum seizures, sometimes by reducing their anti-epileptic drugs over a short period of time, or subjecting them to sleep deprivation, etc. These can lead to altered dynamics in the brain activity as the brain adjusts to the changes. These dynamics, even though not explicable in all instances, may easily be caught by the PLV detector in this pre-surgical evaluation data as a false alarm, possibly absent in normal routine EEG of the epileptic patient.

The above effects were analyzed in detail based on two patients data, one example was shown previously in this Chapter. Although these false alarms jeopardize the detection analysis, they raise an important question, that should all these false alarms be ignored or do they sometimes relay useful information, that may help in seizure prediction. Barring the artifactual false alarms, the focal discharges, spikes, spindles may be a small window that allows us a glimpse into the synchrony dynamics that the brain experiences at all times, even when it is assumed to be at rest. These may indicate the low level formations of synchrony that precede a seizure. This also motivated the detailed analysis of ICA-PLV for seizure prediction.

The PLV-d is observed to remarkably improve the sensitivity and specificity of the detection. The PLV-d for T scalp is shown in Figure 7.44.

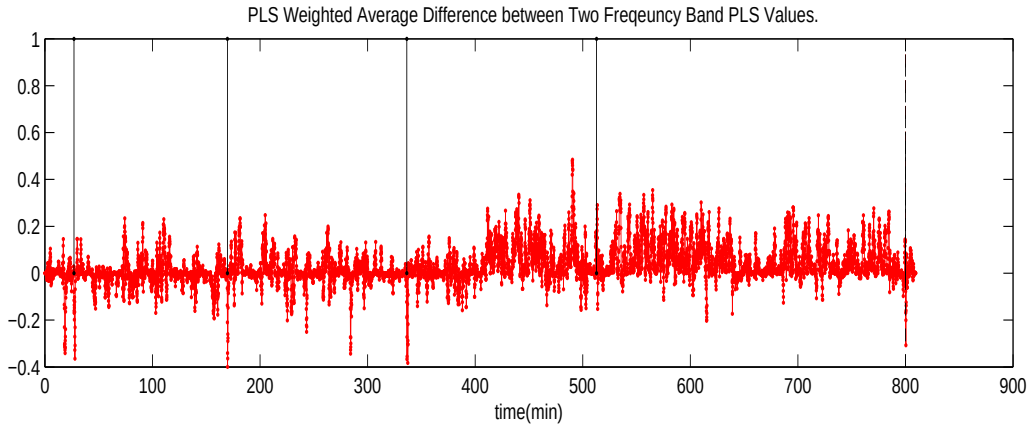


Figure 7.44: The plot shows the PLV-d curve: the difference in significant PLV values in 2-8 and 8-14 Hz bands of spatially constrained sources of 13 hours continuous scalp EEG. There are 5 seizures marked by vertical lines

7.5 Discussion: Seizure Prediction in Continuous EEG

Seizure prediction with ICA-PLV / PLV-d curves was performed here based on long term continuous EEG. The algorithm is essentially similar to the one used for seizure detection. It has an additional step of finding the PLV difference, termed as the PLV-d.

7.5.1 Similarity of PLV in Two Frequency Bands: *PLV-d*

The PLV-d is an important step that helps to analyze the similarity of phase synchrony in two different frequency bands at the same time. It helps to analyze the changes in significant phase synchrony across time, and to seek patterns that might help in seizure prediction. PLV-d essentially helps to compare the similarity of PLV variations in two

frequency bands of 2-8 Hz and 8-14 Hz. PLV-d also helps in canceling the spurious effects in one frequency band by strengthening the underlying pattern as opposed to any specific pattern. The PLV-d curve is averaged or smoothed with a low pass filter to highlight the slowly changing synchrony patterns (conditions out spurious peaks).

As previously stated, the hypothesis for this analysis is that, when a seizure source becomes active it may entrain signals and oscillators in certain spatial locations in the brain (focal or contralateral) within a certain frequency range (chaotic synchronization) causing one dominant frequency to exist. This would show as a peak or a trough on the PLV-d curve. Whilst, at interictal times the varied synchronization networks that co-exist in the brain with different frequencies would cancel or diminish each other.

7.5.2 General Observations

There were some general patterns observed across patients, irrespective of the type of epilepsies. These are listed below:

- *Cyclical Pattern:* In general all PLV-d curves show a low frequency oscillatory pattern. This pattern fluctuates across the zero level. The zero level implies that the difference in the PLV's of the two frequency bands is zero which means that either the PLV's of the two frequency bands are equally strong and cancel each other out or are both simultaneously zero.
- *Circadian-like Rhythm:* The fluctuating rhythm appears to have a pattern that follows the day and night timings, a *circadian-like rhythm*. This rhythm is repeatedly seen across patients irrespective of the type of epilepsy, age, medication, etc.
- *Set-Reset pattern:* The cyclical pattern of the PLV-d curve is also seen to show a 'set-reset' pattern before a seizure onset, which can be useful in indicating an oncoming seizure (see Figure 7.3). The set-reset pattern might be indicative of the dynamics of synchrony leading to a seizure and may provide means for seizure prediction.
- *Seizure pulls the rhythm against the norm:* It is also seen that an impending seizure/discharge is seen to pull the dominant frequency away from the normal trend (a possible 'normal' trend is sketched by hand with dash dot line on the curves, although this is clearly highly subjective).
- *Sharp transition at onset of drowsiness:* A sharp transition is observed as the drowsiness onsets. It is usually seen to follow a pattern of time/habit. Perhaps it shows some link to the biological clock that leads to drowsiness and switching off the brain from an alert state to a drowsy/sleep state. However, this aspect needs

further research before any conclusions can be made. An interesting example was found that could help to validate these correlations: (a) If the drowsiness transitions were correlated with sleep onset and (b) if these transitions were fixed with time or they were flexible. The patient, D scalp underwent sleep deprivation, wherein the patient is forced to stay awake and the sleep wake pattern changes temporarily. The patient sleeps at 2:00 a.m. and wakes at 7:00 a.m. The PLV-d curves of patient D scalp, can be observed for four days. The sleep deprivation setting in on day two. The transitions at the onset of drowsiness is seen on day one (see Figure 7.3.6), though it is gradual then. Day two when the sleep deprivation sets in, the rhythm is observed to dip at the usual sleep time (see Figure 7.34), but as the patient is forced to stay awake, the rhythm appears to adjust itself and then transitions at the time the patient actually sleeps. It is observed to rise at the waking time as usual. The fluctuating wave could indicate a learning pattern of the biological clock. Day three shows that the rhythm follows the new sleep wake pattern much better than day two (see 7.3.6), even though the transition is gradual and small. The day four PLV-d curve shows a perfect synchrony with the new sleep pattern with a sharp transition at the onset of drowsiness (see Figure 7.35). The only peak within the sleep time coincides with the deep sleep, possibly due to sleep spindles. This also shows that few of the peaks/troughs in the cyclical pattern could be attributed to identifiable events.

- *Seizures occur at onset of drowsiness:* Co-incidently many seizures across patients are observed to occur at the onset of drowsiness. This could be related to the sudden sharp shift in the dominant frequency bands at onset of sleep.

- *Seizures dampen the circadian-like transitions:* The transitions seen at the time of drowsiness are usually observed to be sharp and the fluctuations before and after drowsiness much further apart. In terms of synchrony, this implies that the dominant frequency changes much in general. However, the day when a seizure occurs the fluctuations are much less and the transitions are more gradual.

These results are very promising but they are in need of a feature extraction procedure, as it is not trivial to generalize distinctive features across patients. Subjective evaluation of the PLV-d curves shows patterns that may allow prediction but they are in need of an objective procedure for a significant analysis. Thus feature extraction and significance analysis will be the next steps discussed in the following chapter.

7.6 Summary

Previously, it was seen that spatially constrained ICA was useful for seizure onset detection using short data segments. The analysis was then carried out on short segments

and the dynamics of synchrony prior to seizure could not be observed continuously. In order to estimate the specificity along with the sensitivity of seizure detection and prediction, a similar analysis with enhancements was performed here, on long term continuous EEG. The detection was performed on four sets of data per patient, namely (a) contralateral source signals, (b) ipsilateral source signals, (c) contralateral EEG signals and (d) ipsilateral EEG signals. Results for seizure detection have shown a high sensitivity but a low specificity in general. The ICA was seen to improve the specificity as compared to using the EEG signals for PLV analysis as was expected. Various possible causes of the high rate of false alarms were discussed.

The main analysis in this chapter was the seizure prediction analysis based on a similar algorithm as the seizure detection, with an additional step of calculating the PLV-d. PLV-d is the difference of PLVs in two frequency bands 2-8Hz and 8-14Hz. It allows to track the dominant frequency band across time. The PLV-d curves of the pair of spatially constrained source signals showed some general features across patients, irrespective of the type of epilepsy, age, etc. These features included a circadian-like rhythm, a cyclical pattern, a sharp transition at the onset of drowsiness, seizures pulling the rhythm against the normal trend, a possible set-reset mechanism and seizure onset coinciding with the onset of drowsiness.

These features show promise in the possibility of the PLV-d curves being able to warn about an impending seizure. Although, it is clear that these features need to be extracted objectively as they are very difficult to identify across patients. Thus the following chapter introduces a feature extraction procedure applied to these PLV-d curves, along with a sensitivity analysis.

Chapter 8

Multivariate Discriminant Analysis of Long-Term EEG Recording Features

The previous chapter described the use of PLV and PLV-d curves, of contralateral and ipsilateral spatially constrained source signals, obtained from multichannel long term EEG. There were interesting patterns observed, with the filtered PLV and PLV-d curves, useful for seizure onset detection and possibly for seizure prediction. It was seen that the use of source signals through ICA improved the specificity and sensitivity of seizure detection as opposed to using raw EEG signals. The PLV-d curves were found to show cyclical patterns of very low frequency, superimposed with a possible circadian-like rhythm. The seizure onsets appeared to coincide with the rise or fall of peaks, and to pull the cyclical rhythm away from the norm. A 'set-reset' mechanism could also be envisaged with the visual analysis of various PLV-d curves obtained for nine patients (two to three days each, contralateral and ipsilateral). However, with the complexities of the manifestation of the epilepsy in each patient, low number of seizures per patient, it is difficult to visually identify a schema of patterns or features useful for prediction. Thus a multivariate discriminant analysis is required, to determine the features that discriminate between the naturally occurring interictal and pre-ictal groups. The objective of the following visualization techniques is to make the representation of different classes as different as possible in the new feature space. Firstly, a set of features were defined that appeared most likely to be involved in indicating a seizure onset with the PLV-d curves. These are described in the next section. Then a dimensional reduction technique based on Radial Basis Function Neural Networks was used to visualize the multidimensional data in a 2-dimensional feature space. This is followed by a probability density modelling based on Gaussian processes. The features extracted from the complex PLV-d representation are typically multimodal, making Gaussian mixtures with non-uniform weights a good choice for modelling them.

Gaussian Mixture Modelling (GMM) helps to track the probability of occurrence of a predictive event. This algorithm is shown diagrammatically in Figure 8.1.

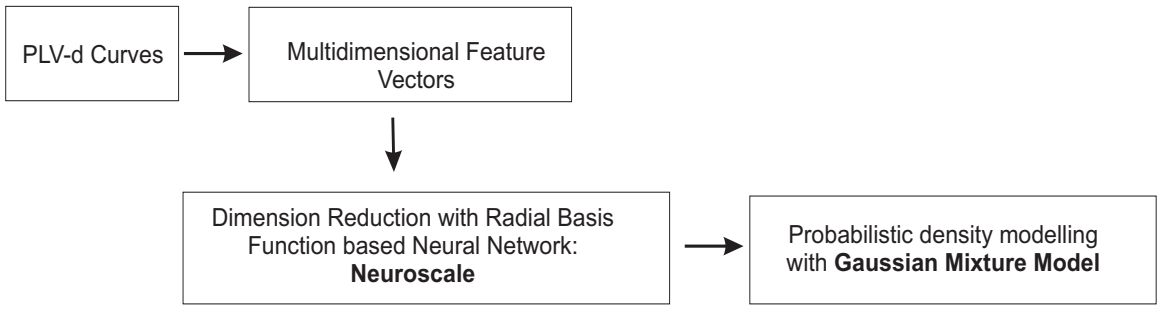


Figure 8.1: An illustration of the sequence of techniques applied for discriminant analysis for seizure prediction.

8.1 Prediction Features

A set of thirteen descriptors or features were defined for the PLV-d curves. It should be noted that the PLV-d curves are filtered to obtain the underlying low frequency curve, and this smooths the curve over longer time spans. Relevant features on PLV-d curve are seen to span hours, hence data windows as large as one hour had to be taken on the smoothed PLV-d curve for feature extraction. Using non-overlapped half hour long windows for a two day data set, gives about 50 data points to be mapped by Neuroscale. These were not considered to be enough for a neural network input. Thus a 75% overlap was used to increase the data points as well as be able to track dynamics across shorter time spans of 15 minutes. The features were selected based on the visual analysis of the PLV-d curves shown in Chapter 7. The PLV-d curves had been obtained for contralateral and ipsilateral source signals of 2-4 days continuous EEG signals of each of nine patients. These are referred to as the *PLV-d features*. Additional features were acquired from the spatially constrained source signals that were used to obtain the PLV-d curves, and are referred to as *IC features*. The PLV-d and IC features extracted are enumerated below:

1. **Standard Deviation:** The standard deviation is used to include the general statistical characteristics of the data window. The PLV-d curve is observed to follow a cyclical rhythm with a general trend being a circadian-like rhythm that appears to follow the sleep and wake times. It appears that the cyclical trend is either due to the normal synchronies of the brain or due to artifacts or rhythmic discharges attempting to energize the brain towards a seizure. An oncoming seizure thus appears to pull this general rhythm away from the normal trend, or with peaks or troughs. This variation may be captured by standard deviation.

2. **Mean:** Mean is used to capture the changes across the time in the cyclical pattern observed on the PLV-d curves.

   - A higher mean in a data window will indicate a wide peak.
   - A low mean in a window will indicate that either the curve remains low in the data segment or the curve may have sudden sharp peaks, both of which are not indicative of a seizure onset. A gradual wider peak has been observed to accompany a seizure onset in most PLV-d curves as shown in Figure 8.2

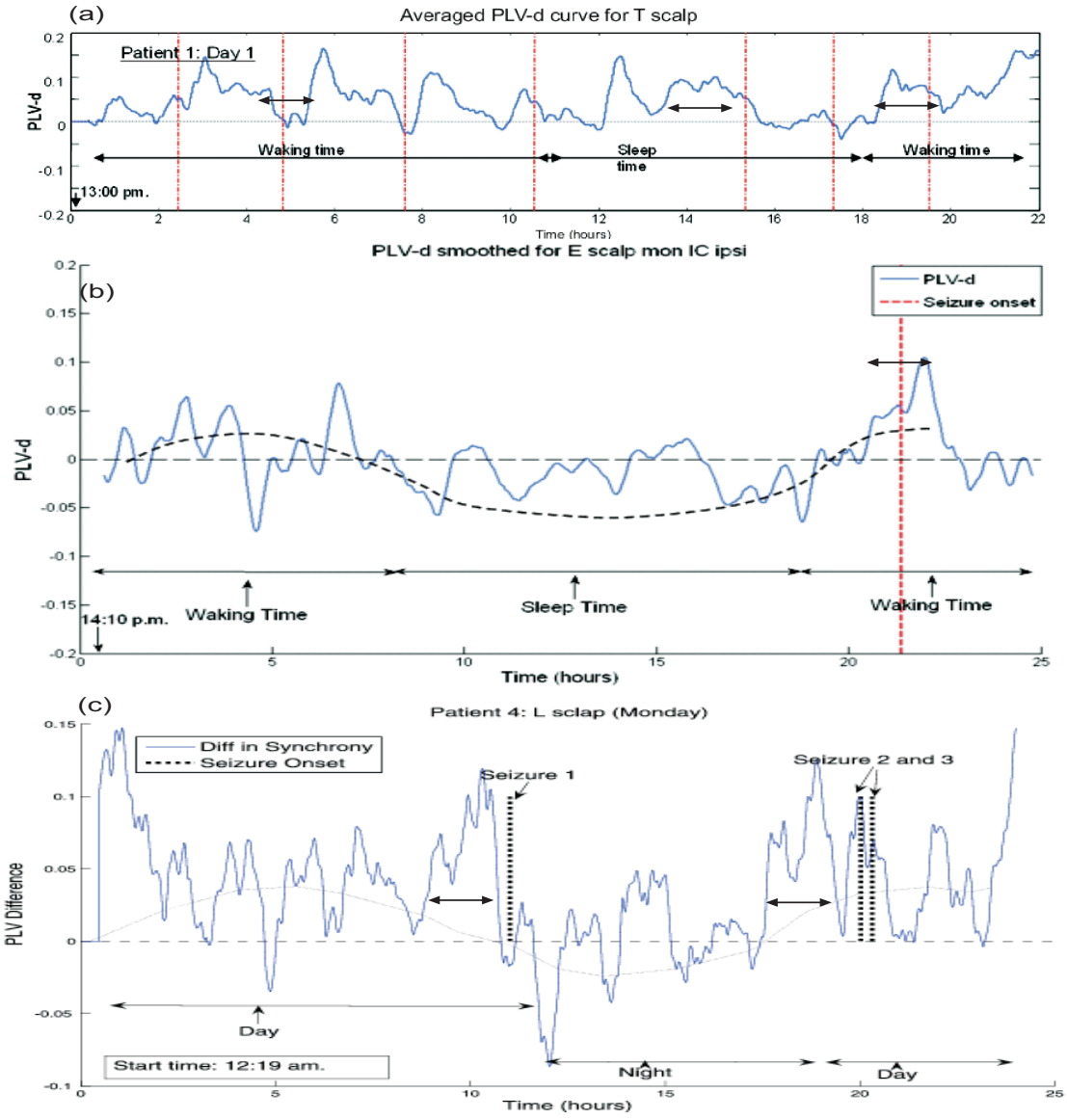


Figure 8.2: An illustration of the PLV-d curve before a seizure onset. These are PLV-d curves of three patients. The vertical lines mark the seizure onsets. The seizure onsets in 3 different patients (a), (b) and (c) are observed to be preceded by wider peaks (indicated by horizontal arrows) or troughs while short sharp peaks or troughs do not appear to indicate a seizure onset.

3. **Ratio of a peak to the preceding mean values:** This feature calculates the

ratio of the maximum amplitude in a data window to the mean of the present and preceding two data windows. This feature tries to capture the changes in the present window as well as a change in trend taking place over a longer time span. It is also helpful in eliminating the spurious peaks that may not lead to a seizure. A high mean value of the preceding three windows combined will indicate the presence of a wide peak spread over three quarters of an hour. This feature is based on the observation of the PLV-d curves that show a wide peak preceding a seizure and a peak immediately before or at a seizure data window. The combination of a wide peak preceding sharp peaks is not common to interictal peaks, where a cycle of peaks are observed. This observation can be seen in Figure 8.2.

4. **Ratio of minimum amplitude to the preceding mean values:** This feature calculates the ratio of the minimum amplitude in a windowed segment to the mean of current and previous two window segments. The minimum amplitude in a data window tries to capture the change in trend over longer time spans in the negative space. The feature is similar to the previous one and tries to indicate a wide trough followed by a deep trough leading to a seizure.

5. **Gradient:** The gradient is calculated for a data segment using the maximum and minimum amplitudes in the data window. It is also possible for a data window to sometimes have more than one slope. Therefore a general slope of the data window is considered to be appropriate and is calculated based on the maximum and minimum amplitude values. Such a gradient value helps to incorporate the information of a general positive or negative trend of the signal within the window. This information can be helpful in case a seizure onset is associated with a general rise or fall of the PLV-d curve (dominant synchrony). An example is shown in Figure 8.3 where all seizures are seen to be on the rising side of the gradient.

6. **Gradient of previous window:** The gradient of the preceding window is used to include the information about the slope transitions before seizure onset. PLV-d curves show a gradual transition of slopes before a seizure onset while at other times the slopes are sharp or vice versa. This feature may be found useful for discriminating the seizure onset.

7. **Mean of previous window**: The mean of the previous window is expected to be significant as changes are observed in the PLV-d curves where the trend starts to change gradually, much before the seizure onset. The feature illustrates the dependence of the present window on the statistics of the preceding window, which may be true only for seizure prediction and may help in eliminate spurious changes observed during interictal times. This gradual change in trend may be found by using the mean of the preceding data window.

8. **Absolute maxima in a window:** The absolute maxima of the signal in a

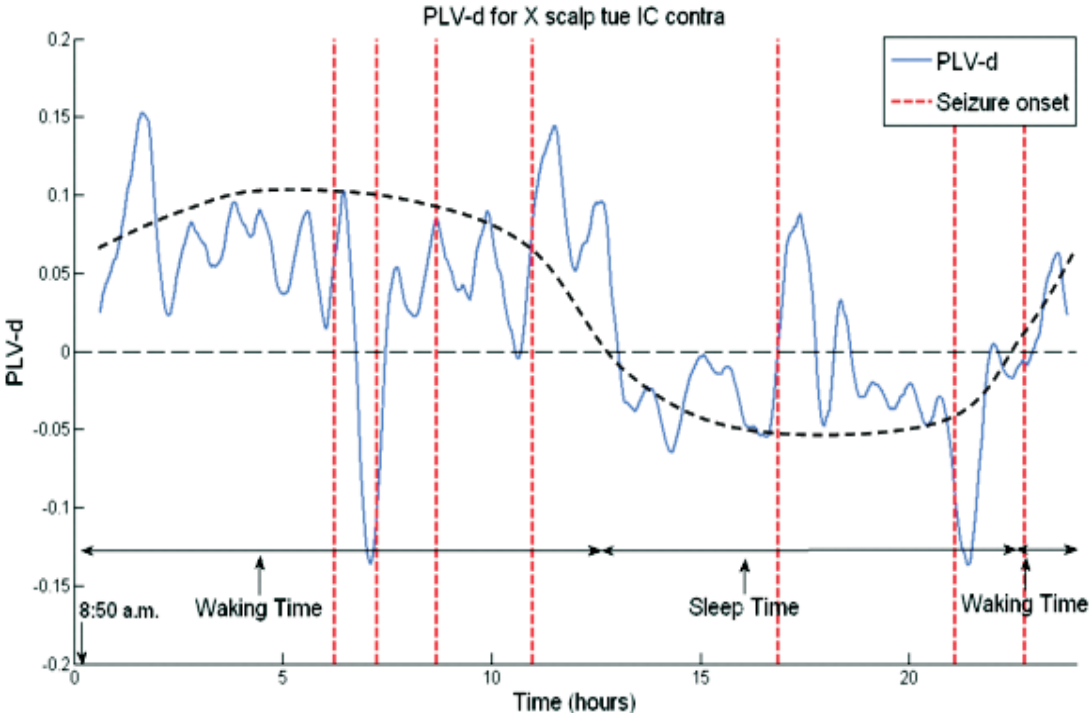


Figure 8.3: An illustration of the gradient of PLV-d at and before a seizure onset. This is a PLV-d curve of a patient. the seizure onsets are marked by vertical lines. The seizures can be observed to be on the rising edge of the slopes of the curve.

data window aims to capture the positive or negative peak. The PLV-d curve appears to be preceded by a sudden peak or trough before a seizure onset. The positive and negative peaks appear to be associated with the day and night trend, hence the absolute maxima is used, ignoring the positive or negative value of the amplitude. Figure 8.4 shows a few segments of PLV-d curves of various patients where such peaks or troughs are observed before/at a seizure onset.

9. **Absolute minima in a window:** The absolute minima is used to capture the feature of the PLV-d curve that indicates interictal activity when having a low amplitude. It may also be used to indicate a transition of the trend towards zero level after a peak before a seizure onset (See Figure 8.4).

10. **Energy of the signal:** The energy of the signal in the data window is captured by using the area under the curve. The energy of the signal is considered as a significant feature as the energy change in underlying signals (source signals derived from EEG signals in this case) is typically associated with seizure onset. The PLV-d curves have been seen to show wider and/or higher peaks or lower troughs, before seizure onset, as compared to interictal times when the peaks are narrower, sharper or generally the curve is smoother or at a low level. This has been illustrated in Figure 8.5, where a series of consecutive data segments show

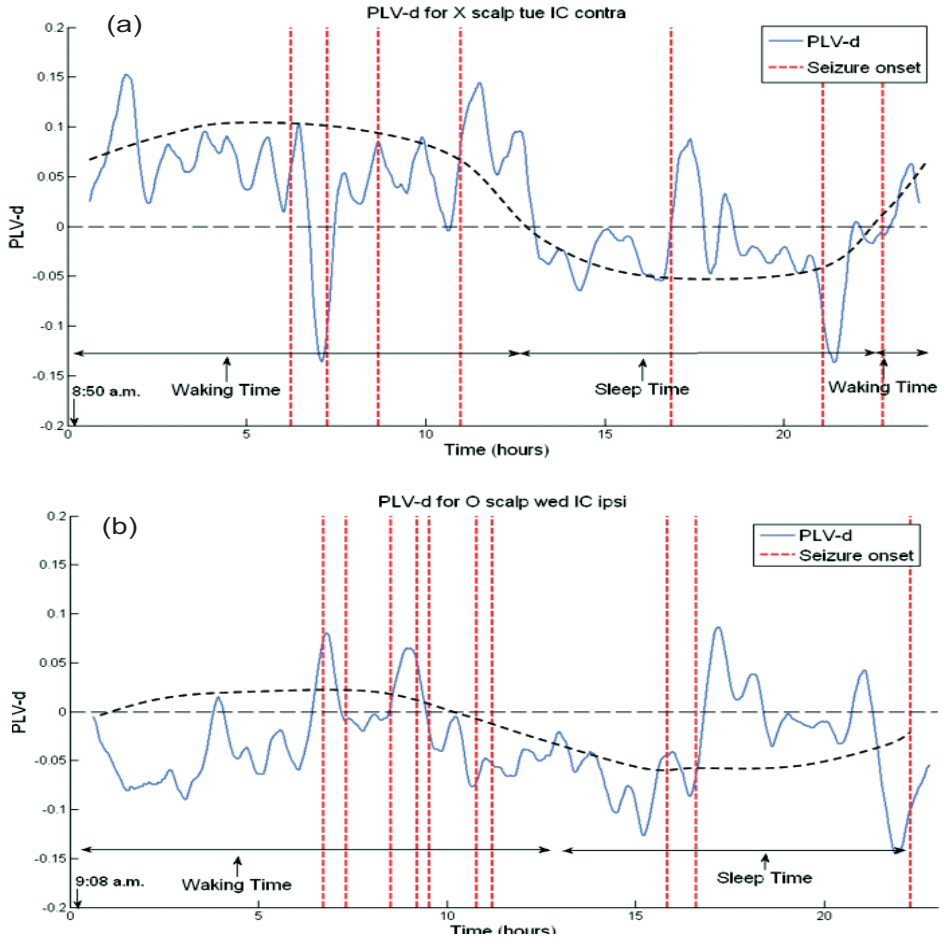


Figure 8.4: An illustration of fluctuation in the amplitude of the PLV-d curve before a seizure onset. These are PLV-d curves of two patients. The vertical lines mark the seizure onsets. The seizure onsets are observed to be preceded by peaks or troughs of high amplitude as compared to the curve at other times.

a seizure onset being preceded by wide and high peaks, while at interictal times the curve remains low.

11. **Fluctuations of PLV-d curve about the mean:** This feature is calculated as the ratio of the duration for which the curve remains higher than the mean of the data segment to the duration for which the curve is lower than the mean. This feature captures the information of fluctuations in trend about the mean of the data segment. It ignores the zero level and identifies the floating rhythm as the zero level with the use of the mean value. The fluctuations of the curve about this level can be useful to capture the changes when the normal trend is getting changed. The use of this feature was motivated by the observation that a seizure onset is observed to pull an underlying rhythm away from its course, which is not the zero level in the case of PLV-d analysis. This is highlighted with an example of a PLV-d curve from a patient in Figure 8.6.

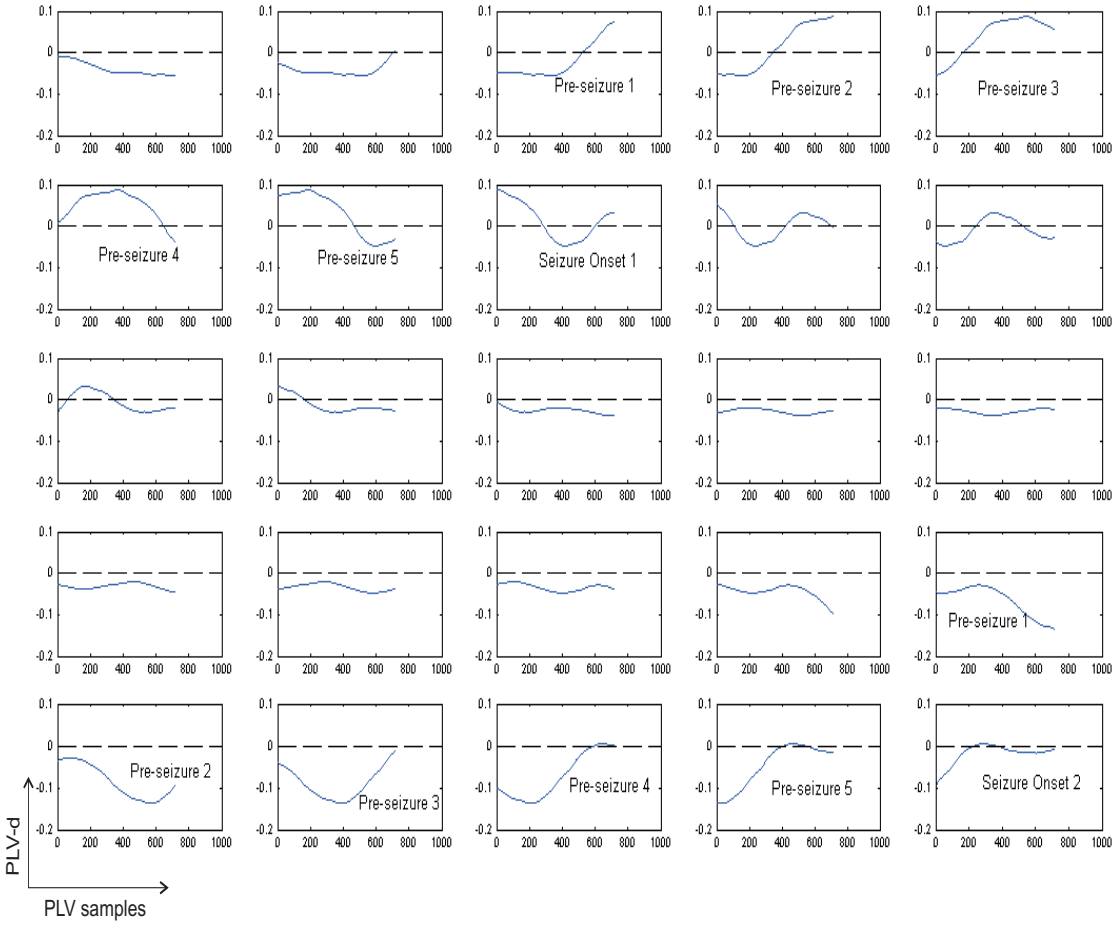


Figure 8.5: An illustration of energy increase before a seizure onset. These are consecutive overlapped data segments each of one hour duration from a PLV-d curve of patient data. The data segment where the seizure occurs is marked and the segments show the increased energy leading to a seizure while it remains low much before and much after the seizure onset.

12. **Dominant frequency trend:** The dominant frequency trend appeared to fluctuate or undergo a transition before or at a seizure onset on the PLV-d curves. The trend being pulled towards zero level at seizure onset and remaining either positive or negative at interictal times. This is shown by a series of consecutive overlapped data segments that show a seizure onset and pre-seizure windows in Figure 8.7. The rhythm can be observed to be pulled towards zero level while it remains mostly positive or negative much before and after the seizure. This feature of change in trend is accounted for by using a ratio of the time the curve remains positive to the time the curves becomes negative.

    - A zero ratio would imply no fluctuation and dominance of band two,
    - A ratio equal to one would imply no fluctuation and dominance of band one,
    - A ratio greater than one would imply a fluctuation and dominance of band one

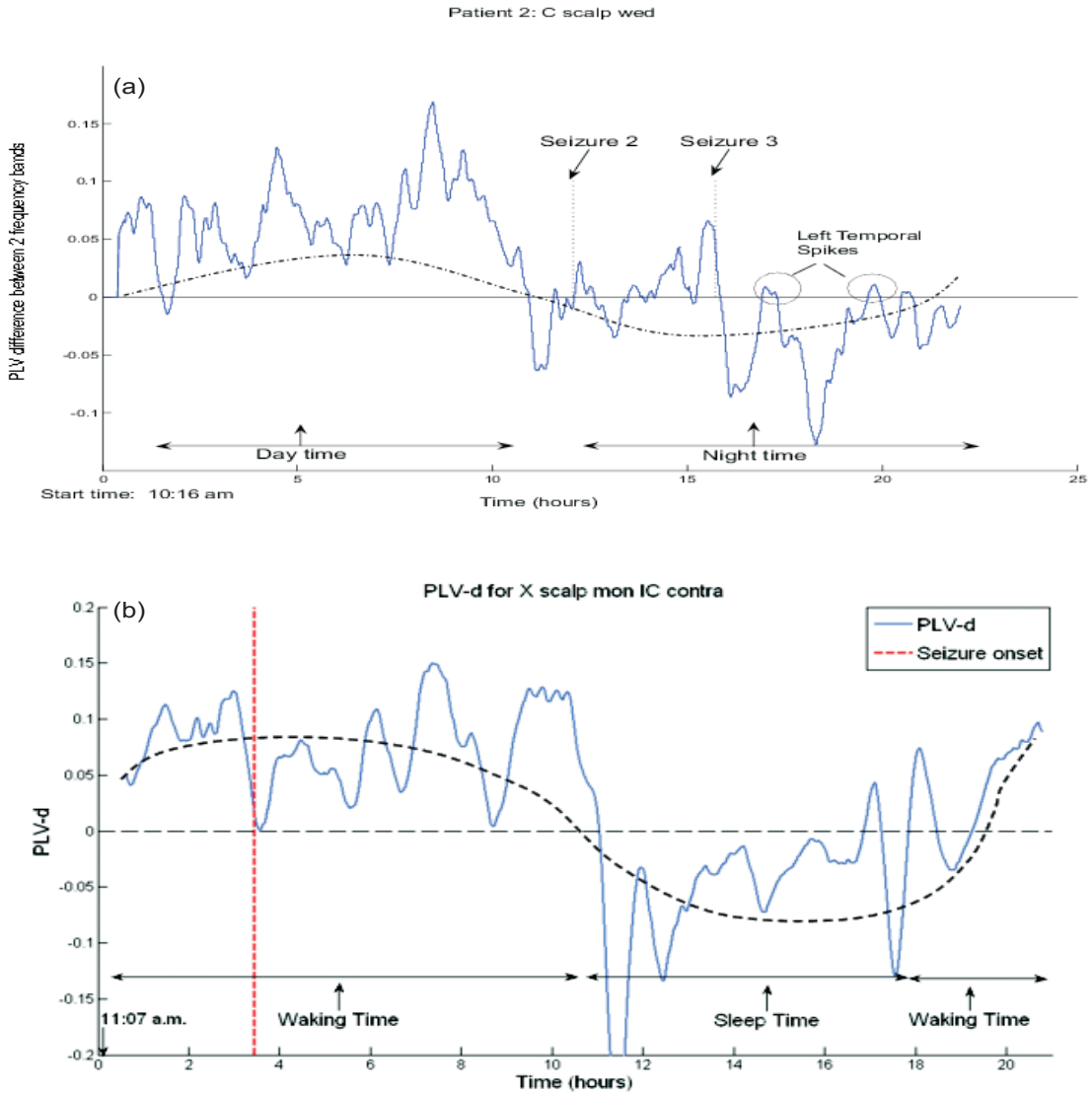


Figure 8.6: An illustration of the circadian-like rhythm away from the zero level is shown on the PLV-d curves. The rhythm appears to follow the day-night pattern, being positive during the wake time with a transition to the negative during the night time. A seizure onset appears to pull the trend away from the normal rhythm (drawn by a rough hand drawn dotted curve over the PLV-d curve). The seizure onsets are marked by vertical lines.

- A ratio smaller than one would imply a fluctuation with dominance of band two.

13. **Zero crossings:** The count of zero crossings in window is another attempt to capture the transition of the trend towards the zero level before a seizure onset.

IC features: In the case of IC features it should be noted that there are two LDCs or source signals used for the calculation of a PLV-d curves. In order to include both the signals in the computation of an IC feature, either the individual feature measure

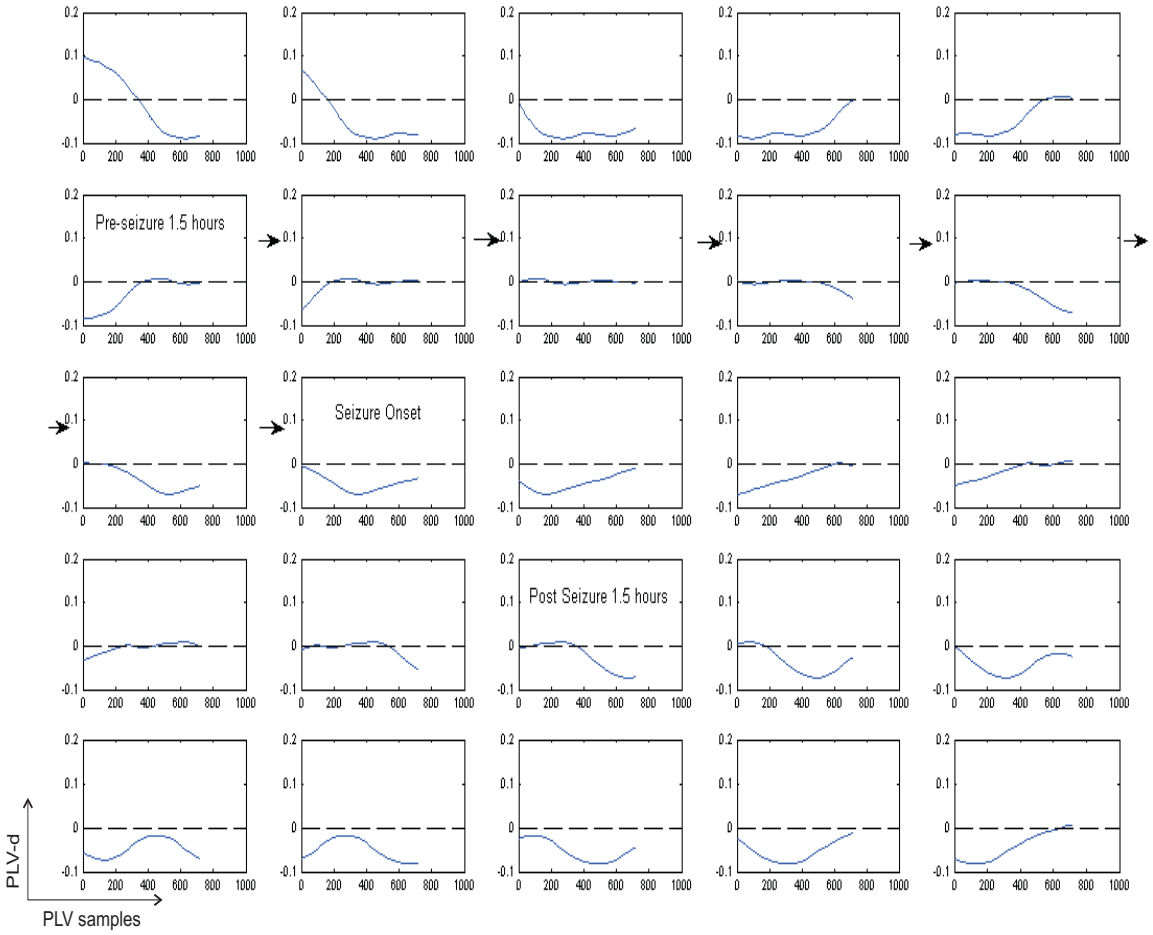


Figure 8.7: An illustration of fluctuation in dominant frequency trend at and before a seizure onset. These are consecutive overlapped data segments each of one hour duration from a PLV-d curve of patient data. The data segment where the seizure occurs is marked and the segments show the trend being pulled towards the zero level before the seizure onset while it remains either mostly positive or negative much before and much after the seizure onset.

of each of the pair of signals or their ratio is used. For example, for the IC feature 'total power' is either computed as two consecutive measures: (1) total power of LDC 1 and (2) total power of LDC 2 or as a ratio of total power of both LDCs such as: $\frac{Total\ power\ of\ IC\ 1}{Total\ Power\ of\ IC\ 2}$.

The IC features used are described as follows:

1. **Total power:** Total power of the two segments of the spatially constrained signals are used as they are likely to show an increased power in a prediction window (preceding a seizure) as opposed to an interictal window.

2. **Rhythmicity:** Rhythmicity is given by the ratio of the peak in the frequency spectrum to the total power. This represents the dominant frequency which is an important characteristic as at seizure onset, a rhythmic discharge with large

amplitude at a certain frequency is significant.

3. **Power ratio in 2-8 Hz:** The power ratio is given by the ratio of the power in 2-8 Hz frequency band to the total power.

4. **Cross correlation:** Cross correlation is given by the cross correlation coefficient that measures the the linear correlation between the two spatially constrained sources. Cross correlation adds the information about the linear relationships in addition to the nonlinear relationships that the phase synchrony considers.

5. **Power in band 2-8 Hz:** Power of the signals in frequency band 2-8 Hz is used as this band typically has a significant change before/at a seizure onset.

These features overlap in the information they are attempting to capture. Also, it is realized that not all features may be useful and may even have a negative impact on the clustering. Some features may also contradict each other causing a loss of important information having an impact on the clustering. Ideally many or all combinations of these features may be the best option for finding the most suitable feature set, but with limitations of time and computing power, a few subjective selections needed to be made. These feature subsets aim to capture useful patterns from PLV-d curves and source signals for an optimal seizure prediction. These feature subsets have been shown in Tables 8.1 and 8.2.

Features	**Set 1**	**Set 2**	**Set 3**	**Set 4**	**Set 5**	**Set 6**	**Set 7**
Total power			✓		✓	✓	✓
Rhythmicity	✓	✓		✓	✓		✓
Power ratio	✓	✓					
Cross correlation	✓		✓	✓		✓	
Power in band 2-8 Hz			✓	✓	✓		
Features	**Set 8**	**Set 9**	**Set 10**	**Set 11**	**Set 12**	**Set 13**	**Set 14**
Total power			✓		✓	✓	✓
Rhythmicity	✓	✓		✓	✓		✓
Power ratio	✓	✓					
Cross correlation	✓		✓	✓		✓	
Power in band 2-8 Hz			✓	✓	✓		

Table 8.1: Feature subsets made of IC features only. The set numbers 1-7 use the features of the pair of source signals in the form of a ratio, where possible. Feature subsets 8-14 use the features from each source signal separately

A combination of PLV-d and IC features was also used after finding the most suitable individual IC and PLV-d feature subsets. The combination of features used was features (1,19), (3,19), (4,19), (11,19), (13,19),(1,20),(3,20),(4,20),(11,20) and (13,20).

Features	Set 15	Set 16	Set 17	Set 18	Set 19	Set 20
Standard deviation	✓	✓	✓			✓
Mean	✓					
Ratio of a peak to the preceding mean values	✓	✓	✓	✓		
Ratio of a minima to the preceding mean values	✓	✓	✓	✓		
Gradient	✓	✓	✓	✓	✓	✓
Gradient of previous window	✓	✓	✓	✓	✓	✓
Mean of previous window	✓					
Abs. max in a window	✓	✓		✓	✓	✓
Abs. min in a window	✓	✓		✓	✓	✓
Energy of the signal	✓	✓	✓	✓	✓	✓
Zero crossings	✓	✓	✓	✓	✓	✓
Fluctuations of PLV-d curve about the mean	✓				✓	
Dominant frequency trend	✓				✓	

Table 8.2: The table shows the PLV-d feature subsets that use only the PLV-d features. The features that are check marked are included in the subset.

These feature subsets were used to obtain a vector of feature values for five patients source signals/PLV-d curves, for both contralateral and ipsilateral pairs. These vectors varied from three to thirteen dimensions. As the features are multidimensional, it is not possible to easily discriminate amongst the features in order to find the best subset of features that will be most relevant and significant for seizure onset prediction. Multivariate discriminant analysis requires visualizing the features in a 2-D space and finding a feature subset that allows a 'good' segmentation of predictive and non-predictive PLV-d feature points.

- *Visualizing multidimensional data:* Visualizing a multidimensional data space requires the transformation of a multidimensional data space to a lower, generally 2-D, feature space. Dimensional reduction helps to visualize the data and to find underlying structures in the data space. Furthermore, the low dimensional data can then be used for data modelling, where the use of 'features' usually improves the modelling performance as compared to using raw data itself.

  Dimensional reduction algorithms can be classified as:

  - Linear or nonlinear: Linear or nonlinear methods utilize linear or nonlinear functional mapping respectively, for data projection.
  - Supervised or unsupervised: supervised methods use *a priori* information about the data to influence the separability of patterns from different classes, in the visualization space. Whilst, the unsupervised algorithms ignore any

class information and use the raw input data for projecting the data on the low dimensional visualization space.

Various techniques preserve the underlying structure in different ways. Most often, the algorithm provides a look up table of data points to be projected for each of the points it has been provided, for example in Sammon Mapping [176]. In such a method, new data points will require re-optimization of the mapping. Other methods, such as Principle Component Analysis, additionally define the transformation mapping between the data space and the output space [177]. This allows the algorithm to project unseen data points without re-optimizing the mapping. The dimension reduction technique used here is a relatively supervised nonlinear method called *Neuroscale* [13, 178, 179, 180, 181, 182]. It is said to be relatively supervised because it inherently uses the topographic structure information as well as it can be provided with *a priori* information for an efficient mapping. A useful quality of the Neuroscale algorithm is that it not only reduces the dimension of the data space, but additionally *learns* the transformation mapping, during the process of dimensionality reduction. This allows Neuroscale to project data unseen by the training phase, saving expensive re-optimization of the parameters. Especially, because in such optimization problems the number of parameters grow with the square of the number of data points, as there are $N(N-1)/2$ distances and weights included in the optimization.

- *Feature Extraction:* The assessment of the suitability of a feature space is analyzed for seizure prediction using GMMs. The use of Gaussian processes can be limited because of one sidedness (non-negative) features, but Neuroscale transforms them to a two sided symmetric feature space of lower dimensionality, which is quite helpful.

These will be briefly described in the following sections, along with their application to the PLV-d curves of patient data.

8.2 Topographic Mapping and Dimension Reduction with Neuroscale

Neuroscale is novel neural network implementation that transforms the data from a p-dimensional input space to a q-dimensional feature space, $(q < p)$. It learns the transformation mapping during dimensionality reduction, allowing the projection of unseen data as well. Neuroscale has been previously used for dimension reduction for epileptic EEG [183], feature extraction for measuring attentiveness with EEG [181], extracting ECG shape descriptors [177], or gas identification from sensor array signals [184]. Neuroscale is a topographic feature extraction method. Topographic extraction

refers to a dimension reducing topographic transformation of the data such that the geometric structure of the data is optimally preserved in the transformed space. This implies that the inter point distance in the feature space closely match the distances in the data space. The algorithm is based on a Radial Basis Function (RBF) Network. An RBF defines each point in feature space y_i as a parameterized nonlinear function of the input: $y_i = f(x_i, w)$, where w is the vector of parameters for the RBF network. The network is then trained by adjusting the network parameters in order to minimize a suitable error measure called the *Sammon Stress metric* which involves the topographic principle.

The algorithm in its basic form does not use class information but uses the data space spatial distribution information. This makes it a relatively supervised algorithm as there is no specific target for the $\mathbf{y}_q$, but a relative measure of separation of $\mathbf{y}_q$ and $\mathbf{y}_p$ is used. However, additional *class information* can also be incorporated with Neuroscale, increasing the supervisory control in the mapping. An illustration of Neuroscale is shown in Figure 8.8 An RBF neural network that forms the basis of the algorithm

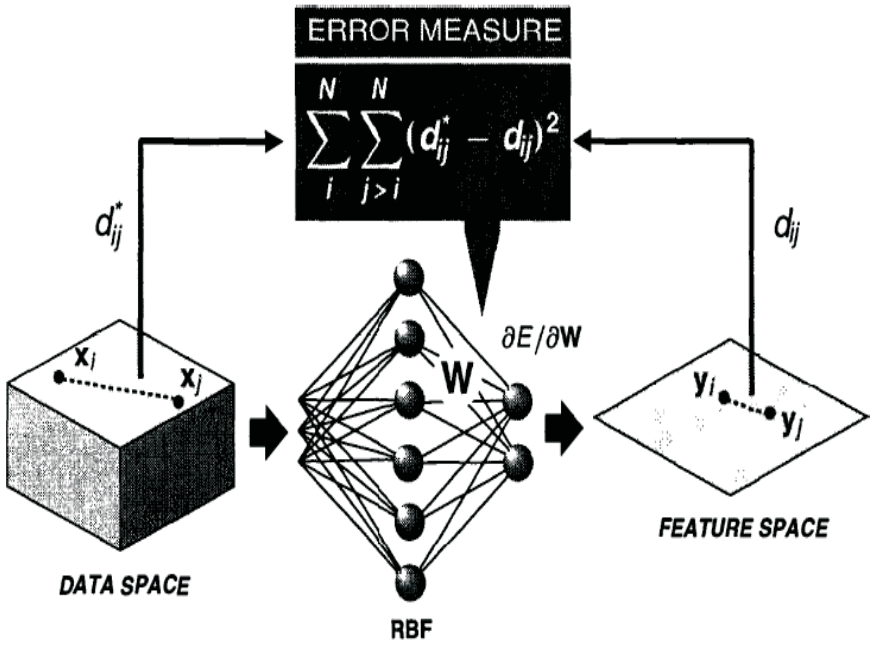


Figure 8.8: An illustration of the Neuroscale model [13]

consists of a three layer network: input layer, hidden layer and the output layer. RBF has a quality of rapid training as compared to the more commonly cited Multilayer Perceptron Neural Network.

The Neuroscale algorithm can be described by the following main steps:

1. *Minimizing the Sammon Stress:* If the N data points of an m dimensional data

space are given by $\mathbf{x}_q$, then the algorithm aims to find an n dimensional feature space of points $\mathbf{y}_q$ such that the relative positions of the feature space points minimize the error called the **Sammon Stress metric**

$$E = \sum_{p}^{N} \sum_{q>p}^{N} (d^*_{qp} - d_{qp})^2, \tag{8.1}$$

where d^*_{qp} are the inter-point Euclidean distances in the data space given by

$$d^*_{qp} = \sqrt{(\mathbf{x}_q - \mathbf{x}_p)^T (\mathbf{x}_q - \mathbf{x}_p)}, \tag{8.2}$$

and d_{qp} are the corresponding distance in the feature space given by

$$d_{qp} = \sqrt{(\mathbf{y}_q - \mathbf{y}_p)^T (\mathbf{y}_q - \mathbf{y}_p)}. \tag{8.3}$$

The topographic nature of the transformation is imposed by the **Stress** term. The Stress term attempts to match the Euclidean distance between points in the feature space with those of input space. A small stress denotes a closer matching of the distances between the visualization space $\mathbf{y}_q$ and the original data space $\mathbf{x}_q$. Therefore, the smaller the stress, the better preserved is structure of the data space in the visualization space. Also, even though the Euclidean distances are used here to measure distances, they may be replaced by any other norm. The advantage of Euclidean distances being, that they assist in the visual discrimination of patterns as the human visual system is more tuned to the Euclidean space [176].

2. *RBF weights training:* The $\mathbf{y}$ points of the low dimensional space are generated by the RBF, given the input data points using a nonlinear transformation

$$\mathbf{y}_q = \mathbf{f}(\mathbf{x}_q; \mathbf{W}), \tag{8.4}$$

where $\mathbf{f}$ is the non linear transformation affected by the $\mathbf{W}$. The distance in the feature space are given by

$$d_{qp} = \|\mathbf{f}(\mathbf{y}_q) - \mathbf{f}(\mathbf{y}_p)\|. \tag{8.5}$$

The quality of projection is measured by the *Sammon Stress metric.*

The model is trained by computing the partial derivatives of the Stress metric $\mathbf{E}$ with respect to the network weights. Then a non-linear optimization algorithm is used to find the optimal weights

$$\frac{\partial E}{\partial w_{kr}} = \sum_{q}^{N} \frac{\partial E}{\partial y_q} \frac{\partial y_q}{\partial w_{kr}}. \tag{8.6}$$

Differentiating Stress metric **E** by y gives

$$\frac{\partial E}{\partial y_q} = -2 \sum_{p \neq q} \left(\frac{d^*_{qp} - d_{qp}}{d_{qp}}\right)(y_q - y_p). \tag{8.7}$$

The gradient of y_q with respect to weights is given by the sum of squares error

$$\frac{\partial y_{qp}}{\partial w_{kr}} = \delta_{qr} z_k, \tag{8.8}$$

where z_k is the output of the kth hidden unit, and δ_{qr} is the Kronecker delta.

The weights are initialized to small random values and the Stress term is differentiated, to calculate the error values. If the error value is outside the tolerance range, the weights are updated and this process of differentiating and weight updating is iterated till maximum iterations or maximum tolerated error is reached.

3. *Using additional class information:* Additional dissimilarity information known about the data can also be included.

Neuroscale is freely available with the NETLAB software package from the Neural Computing Research group at Aston University in U.K. (http://www.ncrg.aston.av.uk/netlab). It has a GNU open License and can be integrated with MATLAB versions 5.0 and above.

Neuroscale was applied to the 30 feature vectors, obtained from five patients' PLV (contralateral and ipsilateral), as well as from the PLV-d (contralateral and ipsilateral) curves. The five patients selected for this analysis were:

- Patient 1 (T scalp)
- Patient 2 (E scalp)
- Patient 3 (C scalp)
- Patient 4 (L scalp)
- Patient 5 (X scalp)

These data sets had slightly more seizures compared to the other data sets and their PLV-d curves appeared (visually) to have some underlying patterns associated with seizure onset, which raised the expectation of extracting useful patterns. The two dimensional feature maps obtained with Neuroscale showed *PLV-d* maps to be having some structure, based on preliminary visual assessment, as compared to the *PLV* maps. Therefore, only the PLV-d maps were used for further analysis.

8.2.1 Supervised vs Unsupervised Training

The Neuroscale mapping was carried out as unsupervised as well as partially supervised.

Unsupervised Neuroscale: In unsupervised training the Neuroscale uses only the information inherent in the distribution of the data points. At this stage it is not given any additional information about the data belonging to any particular class. As compared to Sammon Mapping where a similar one to one projection of data is possible, the Neuroscale additionally *learns* the transformation parameters, during the process of dimensionality reduction. Therefore, it can be used to project data that is not seen by the network before, i.e. data that has not been a part of the 'Training' data set, without computationally expensive re-training of the network for the augmented data set.

Supervised Neuroscale: In supervised training the network being trained is provided additional information about the data points. This can be provided in the form of a class label $\mathbf{t} = t_c^n$ for each vector, where t_c^n is an indicator variable given as

$$t_{qp}^n = \begin{cases} 1 & if x^n \in C_c \\ 0 & otherwise, \end{cases}$$

or as a matrix such as the one used in this analysis. It is a matrix with binary distance metric with the distance between points within the same class is zero while between two different classes is one. For example, if there are 4 data points (each point having seven dimensions), where data points 2 and 3 belong to class 'Predictive' and point 1 and 4 belong to class 'Interictal'. The distance metric will be given by

$$dist = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

This information is incorporated along with the distribution distance metric that the Neuroscale uses. The extent of dependence on this additional information is decided by the parameter α, as the distance being used for training is given by :

$$\delta_{qp} = (1 - \alpha) d_{qp}^* + \alpha \; t_{qp}$$

in place of d_{qp}^* as in unsupervised network. A value of $\alpha = 1$ makes the training completely supervised while a value of $\alpha = 0$ makes it completely unsupervised. A value of α in between 0 and 1 makes the network partially supervised. The additional class information helps to improve the visualization of the feature points showing greater

class separation.

8.2.2 Results: Neuroscale Feature Space

Unsupervised and partially supervised Neuroscale dimension reduction and network training was carried out on five patient data sets (contralateral and ipsilateral each) for 30 feature subsets. The IC features did not show much structure for interictal and predictive windows, but rather overlapping ones. Two examples of Neuroscale maps for the five patients PLV-d curves using IC feature subset number 13 and IC feature subset 10 are shown in Figures 8.9 and 8.10. The PLV features instead showed some distinction that increased with the extent of supervision added to the training. The Neuroscale maps for six PLV feature subsets (varying from 6 to 13 dimensions), for the five patients, trained without supervision and with two levels of supervision ($\alpha = 0.3$ and 0.8) have been shown in Figures 8.11, 8.12, 8.13, 8.14, 8.15 and 8.16.

The above show that feature subset 1, 5, 6 show some distinct structure in the feature points across all patients, for supervised training with $\alpha = 0.8$. These have been put together in Figure 8.17 for a clearer view to observe the extent of overlapping of classes. Subsequently, a technique is used that automatically segments this 2-D feature space, if possible, into *predictive* and *non-predictive interictal* sets. This is performed by GMM [176]. GMMs are used to obtain the probabilistic distributions of interictal and preictal events in the feature map and to track the probability of seizure prediction events. This will be described in the next section.

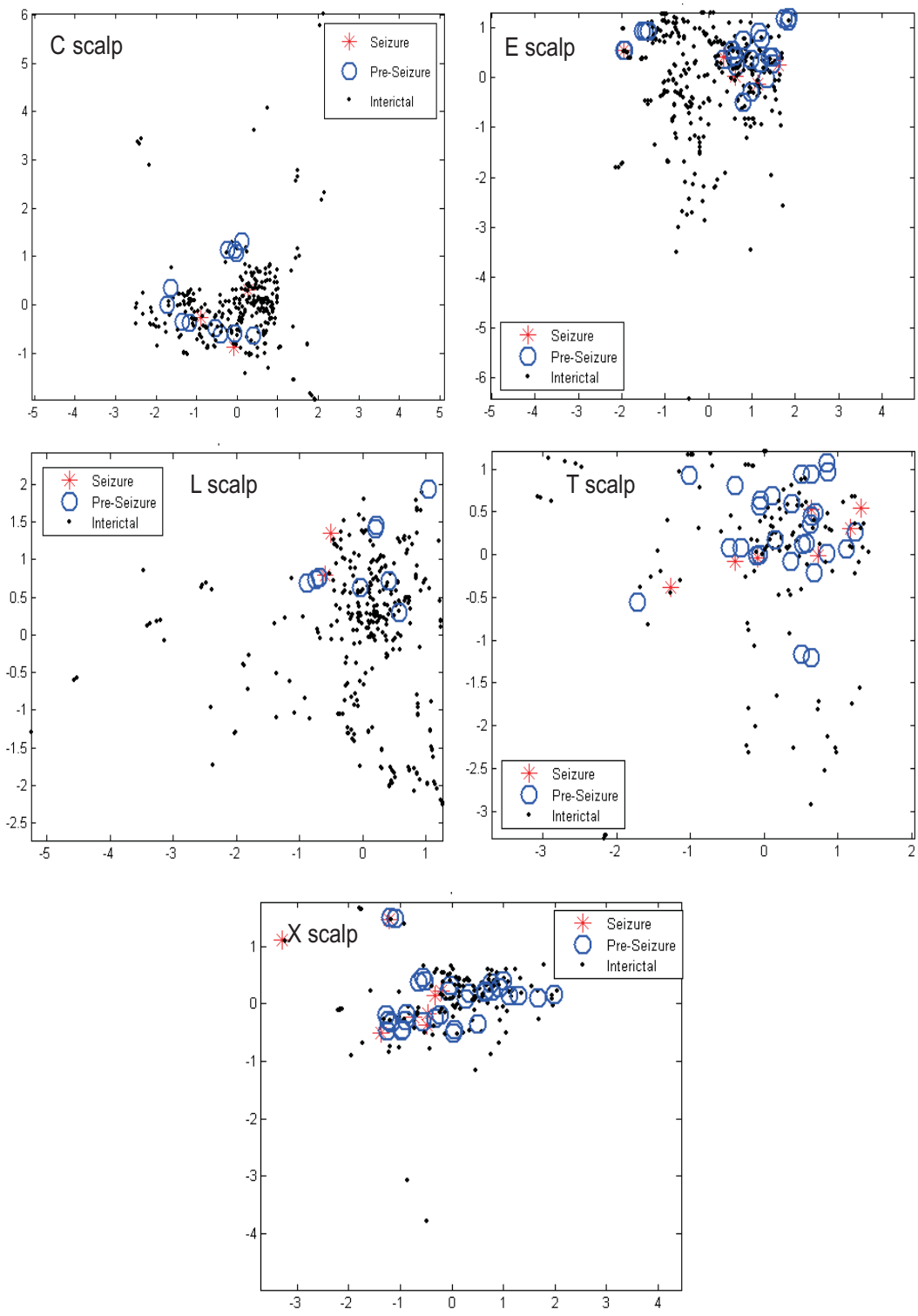


Figure 8.9: The Neuroscale maps for PLV-d curves of five patients have been shown here. The feature subset used is IC subset 13 and the number of hidden nodes used for Neuroscale is nine. The maps show overlapping of clusters of the interictal and pre-ictal data points.

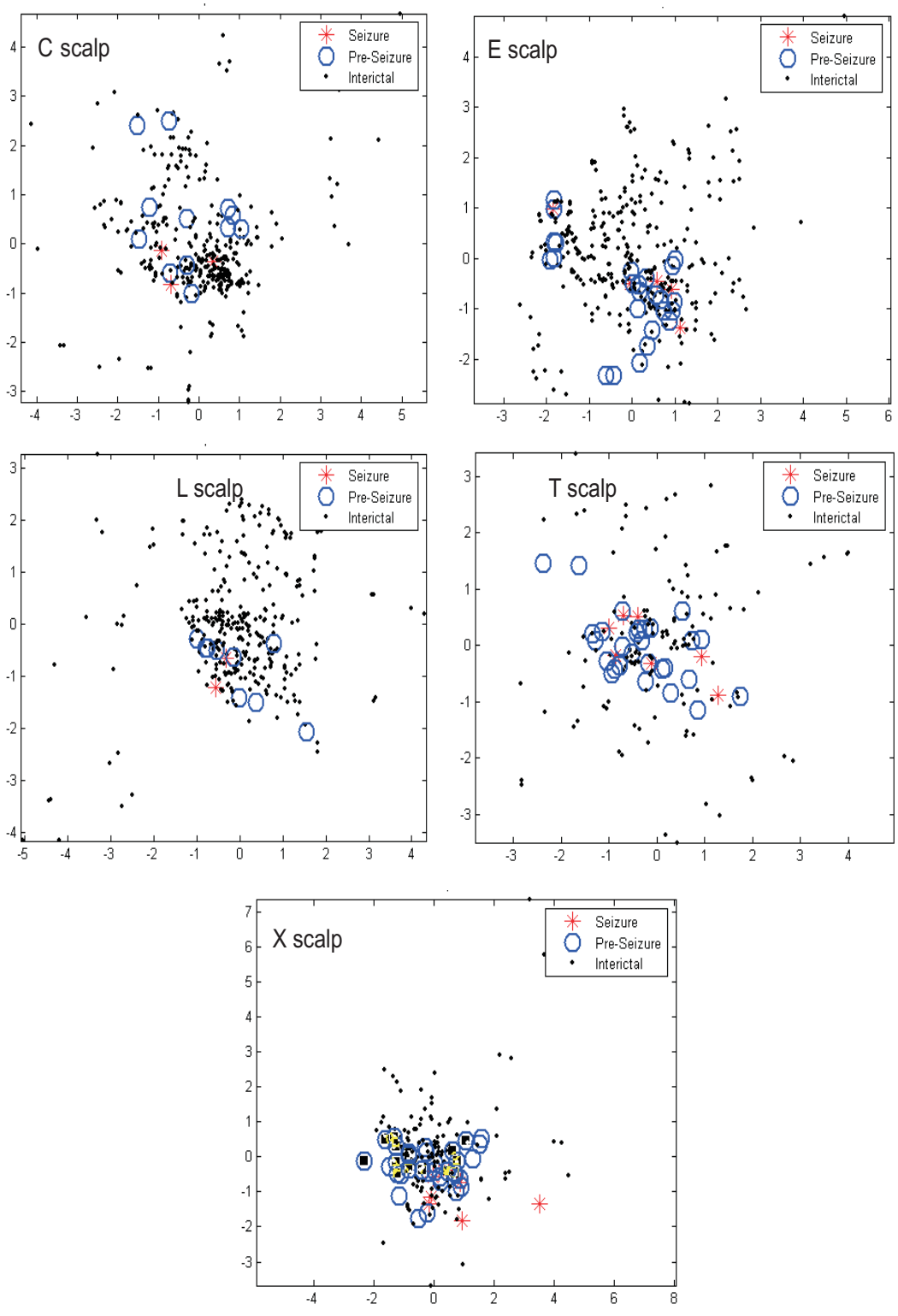


Figure 8.10: The Neuroscale maps for PLV-d curves of five patients are shown here. The feature subset used is PLV feature subset 10 and 13 hidden nodes are used for Neuroscale. These maps also show overlapping of clusters of the interictal and pre-ictal feature points.

8.3 Feature Extraction with Gaussian Mixing Model

The neuroscale feature mapping is followed by a *semi-parametric estimation technique* called Gaussian Mixture Modelling. A parametric model implements a restricted family

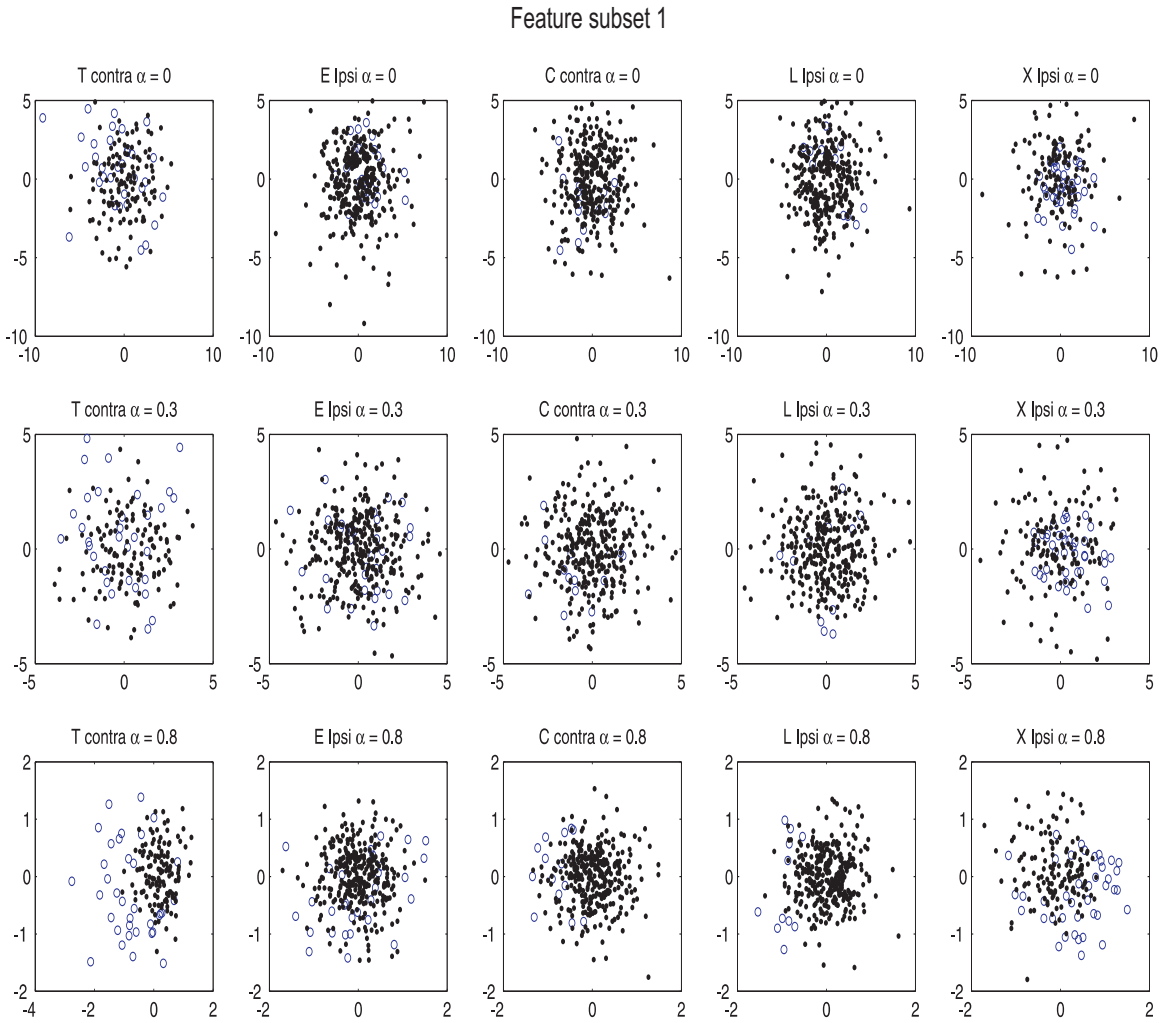


Figure 8.11: Partially supervised training with PLV feature subset 1. Supervised training with $\alpha = 0.8$ shows some distinction in the class clusters for all patients. Unsupervised training shows overlapping clusters for all patients.

of functions, with only a few parameters to be learned, but uses a strong *a priori* information for successful convergence. On the other hand a non-parametric method makes weak assumptions about the data and is constructed directly using the information in the training data, with no or very few parameters to be learned. A semi-parametric method makes use of multiple parametric components, such that components tending to infinity would be capable of fitting any data. Semi-parametric methods can be controlled by the components involved giving them the advantage of having the efficiency of parametric methods and flexibility of non-parametric methods. However, the disadvantage is the need to learn many parameters. Mixture models like the GMM output a weighted sum of their parametric mixture components, estimating the mixture coefficients and the parameters of the individual components. GMM has been used for modelling various statistical distributions, including nonlinear distributions. GMM can be used as unsupervised learning, where no class information is available and unlabelled

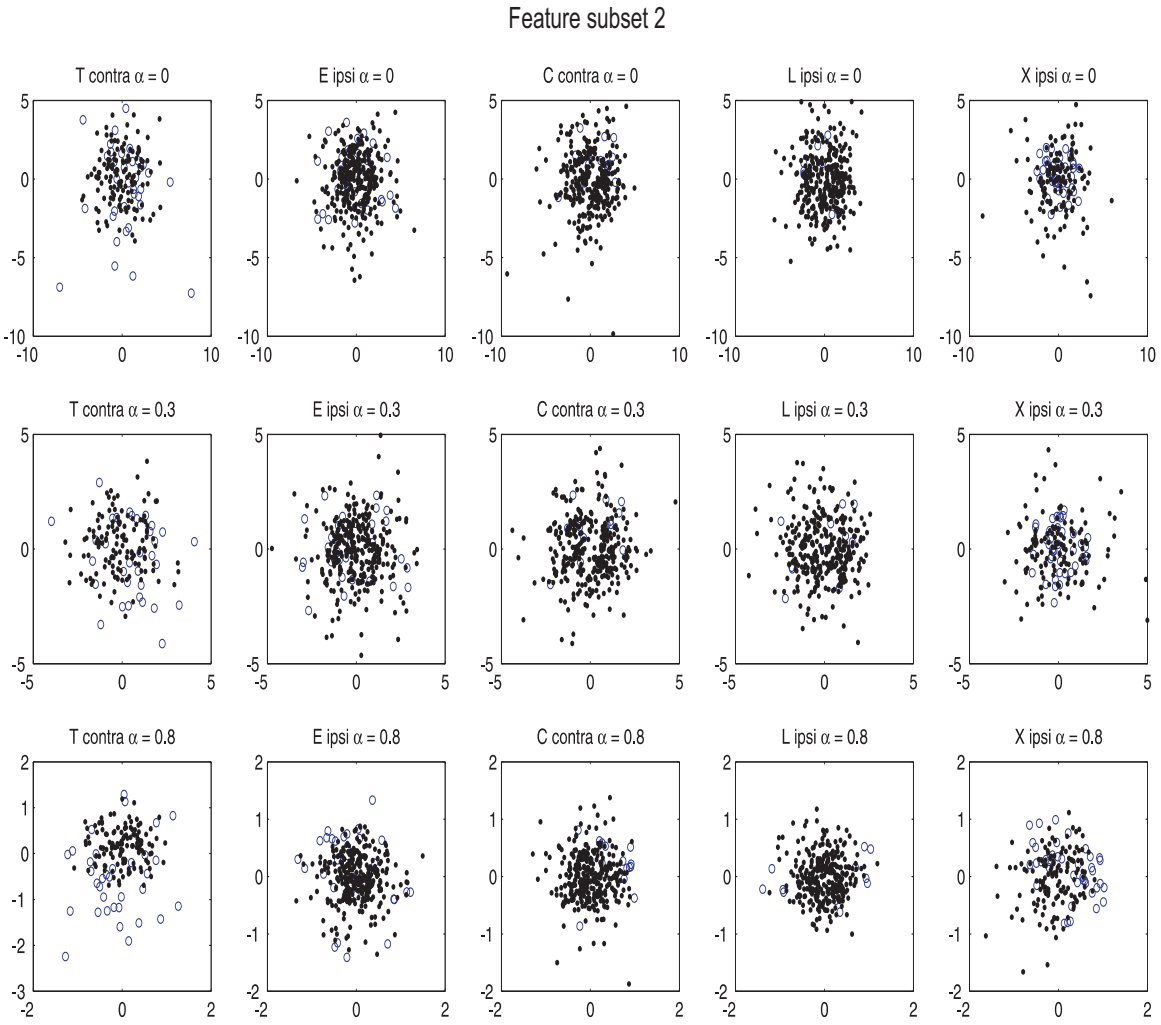


Figure 8.12: Partially supervised training with PLV feature subset 2. Supervised training with $\alpha = 0.8$ shows slight distinction in the class clusters for T, E and L scalp. Unsupervised training and supervised training with $\alpha = 0.3$ shows overlapping clusters for all patients.

data is used to estimate classes in the data, such as in image processing [185, 186] and in gas identification from sensor array signals [184]. Unsupervised learning works best when there is distinction in the distribution of the data points, but in cases when they are overlapping, it is not trivial and almost impossible for GMM to identify the data classes with no prior information. GMM can also be used in a supervised mode, using the class information to ascertain linear/nonlinear decision boundaries in the data points, such as studied for data classification [187, 188] and speech and text recognition by [189]. In this research, GMM will be used as a classifier in two ways:

1. Supervised learning and classification with labelled data,

2. Semi-supervised learning with labelled and unlabelled data.

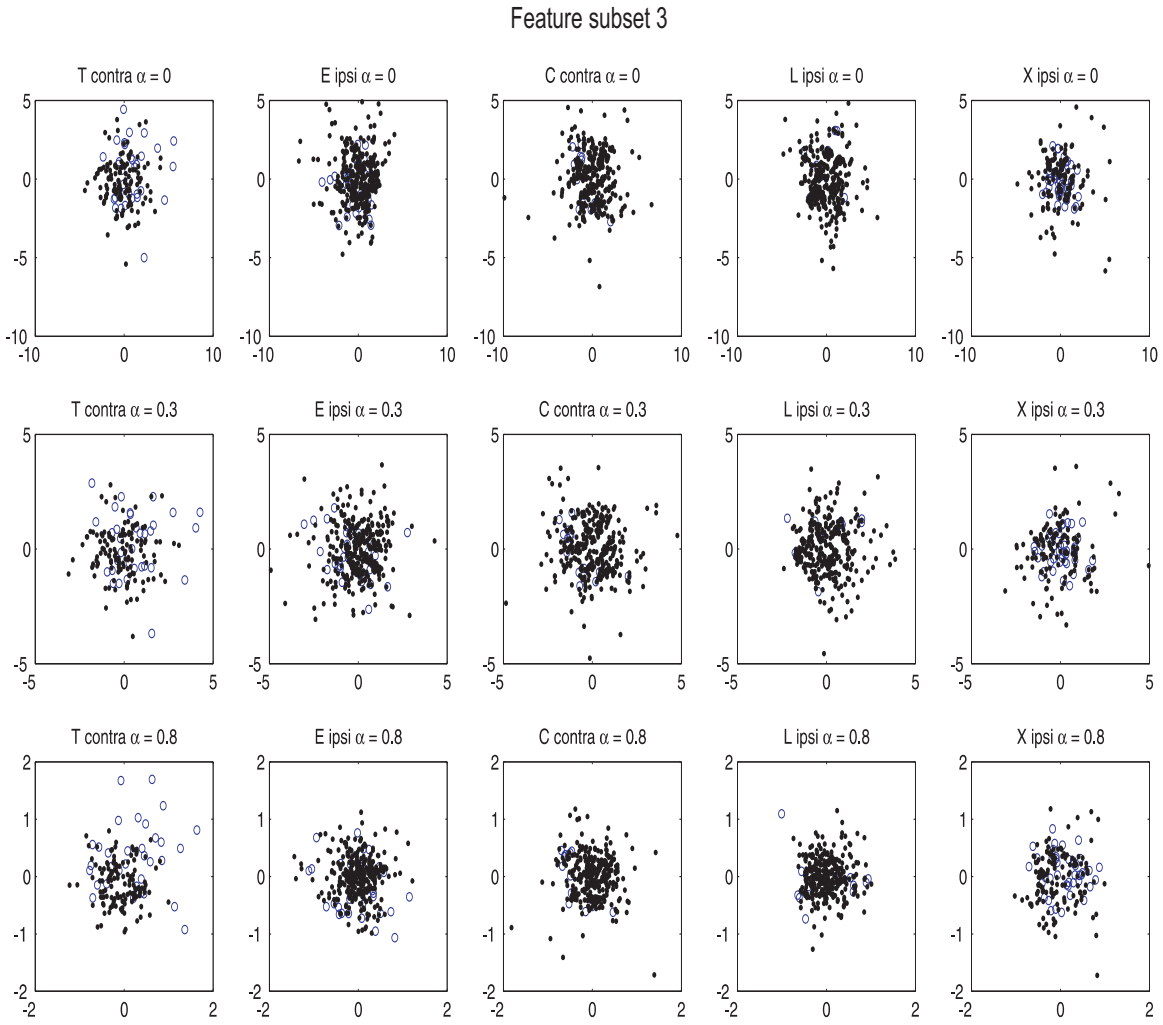


Figure 8.13: Partially supervised training with PLV feature subset 3. Supervised and unsupervised training shows much overlapping in clusters for all patients.

Semi-supervised learning can be thought of as a combination of supervised and unsupervised learning, popular 'machine learning' terms. Essentially, in supervised learning, the learning system is provided with labelled data from a *training* data set. The task of the system is to learn a classification rule that can be used to later classify unlabelled *test* data. However, no learning takes place in such a system and the classifier is only trained on the training data [190]. In the unsupervised learning, the system is given unlabelled data with no class information. The aim of the classifier is to group similar data points together, based on some similarity criteria, usually the Euclidean distance. The system in this case is able to separate the classes but it will not be able to tell between classes as no label information was provided to it. Thus unsupervised learning is usually useful for uncovering structure in the data but requires analysis to interpret it [190].

Semi-supervised learning falls between supervised and unsupervised learning. Such a learning adds to the training of the labelled data with the structure learned from the

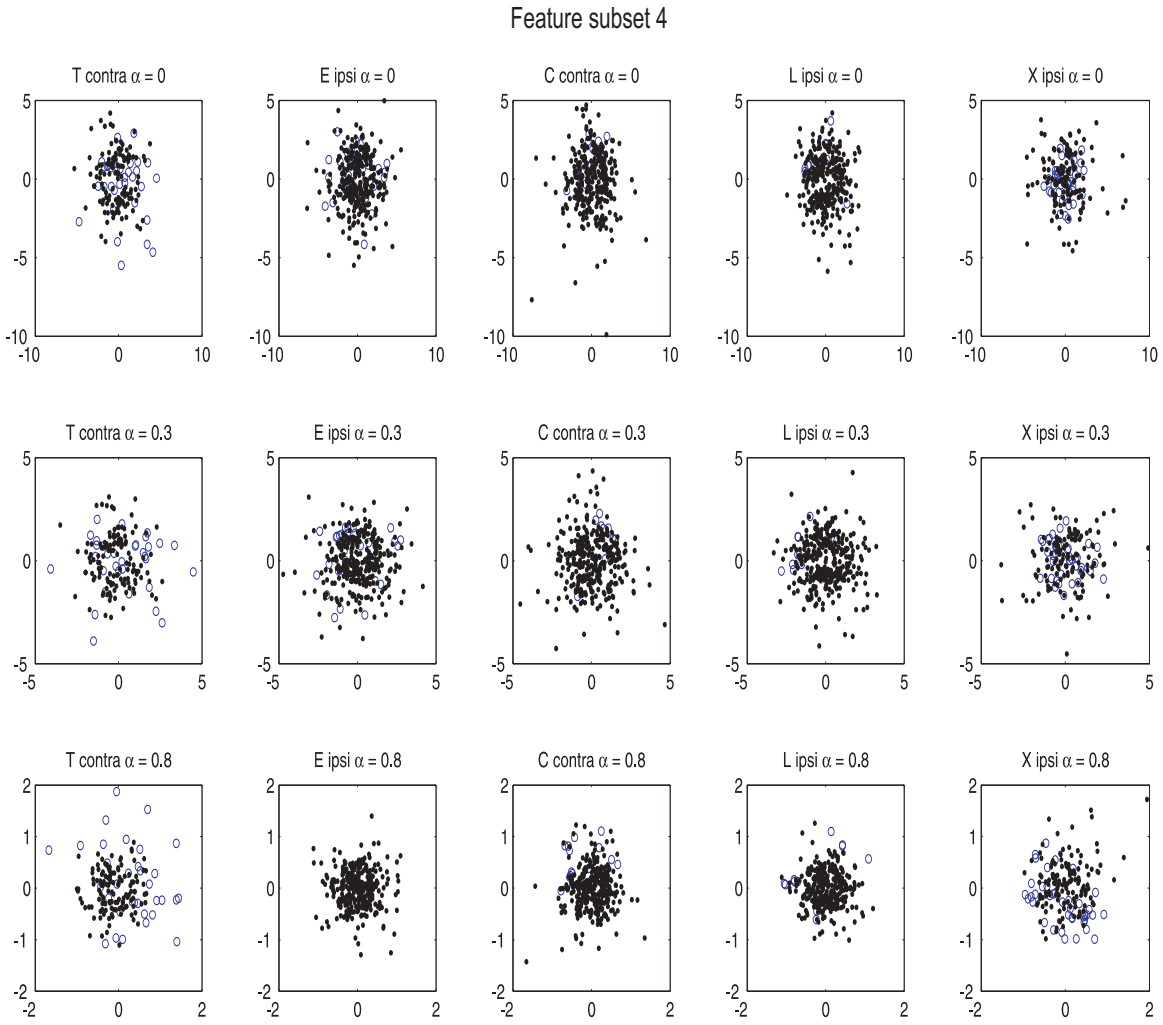


Figure 8.14: Partially supervised training with PLV feature subset 4. Supervised training with $\alpha = 0.8$ shows some distinction in the class clusters for T and C scalp. Unsupervised training and supervised training with $\alpha = 0.3$ shows much overlapped clusters for all patients.

unlabelled data. Therefore the semi-supervised learning essentially trains on both the training set as well as the unlabelled test set. Training on the test data set has been shown to be advantageous (not using the labels of the test set) [190]. If the labels of the test set are known, they can be useful in measuring the quality of the classification.

Formulation of the Gaussian Model

The GMM is based on modelling the unconditional probability density $p(x)$, given a finite set of unlabelled data points $x_1,\ x_2,\ x_3 ... \ x_n$. In a mixture model, the probability density function is expressed as a linear combination of basis functions. A model with M components is written as

$$p(x) = \sum_{j=1}^{M} P(j)\ p(x|j), \tag{8.9}$$

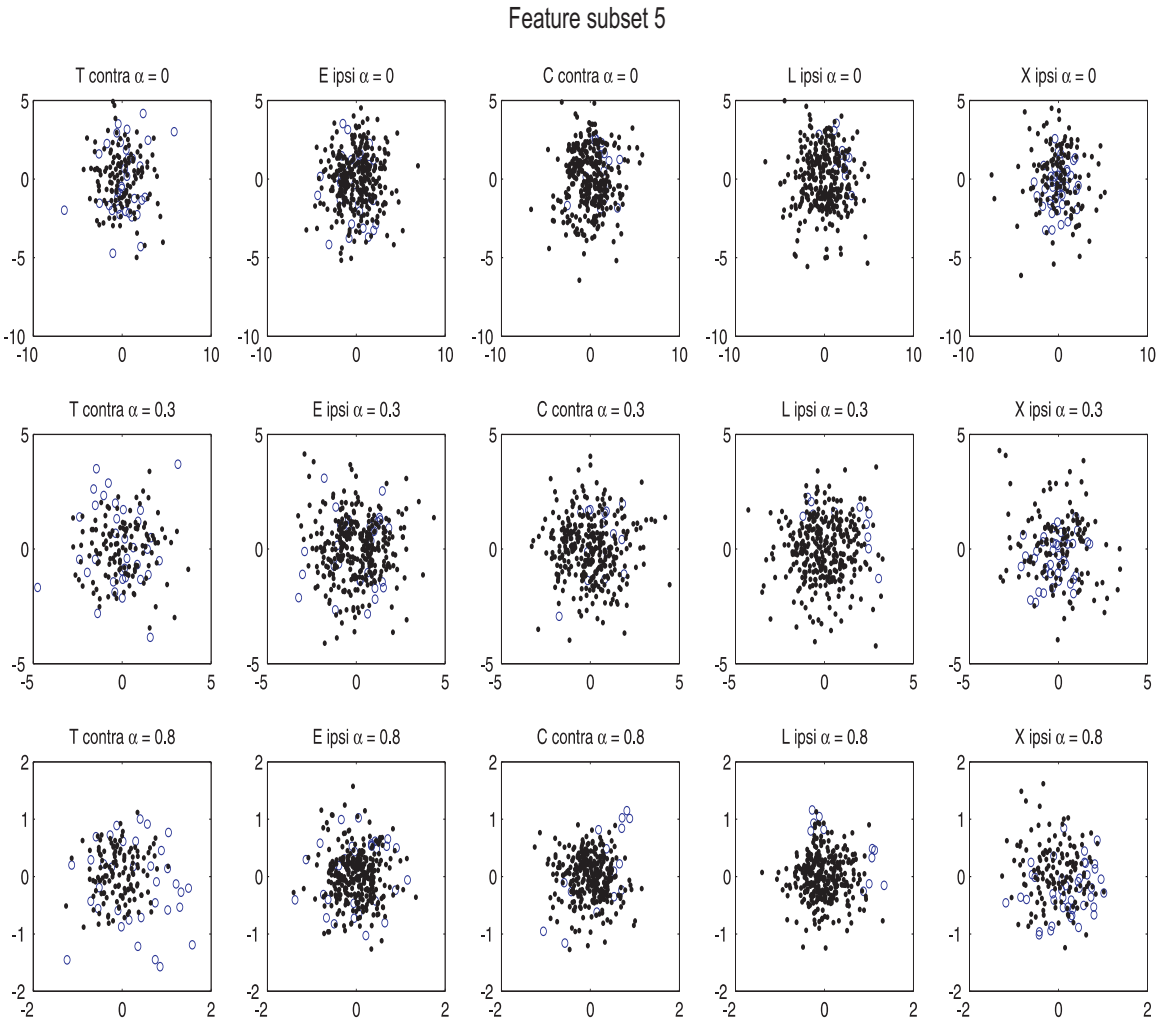


Figure 8.15: Partially supervised training with PLV feature subset 5. Supervised training with $\alpha = 0.8$ shows a slight distinction in the class clusters for T, C and L scalp. Unsupervised training and supervised training with $\alpha = 0.3$ shows much overlapped clusters for all patients.

where the parameters $P(j)$ are called the *mixing coefficients* and the parameters of the component density functions $p(x|j)$ typically vary with j. The mixing coefficients $P(j)$ represent the weights associated with the component functions $P(x|j)$. To be a valid probability density, a function should be non-negative everywhere and integrate to 1 over the whole space. Therefore using the constraints

$$\sum_{j=1}^{M} P(j) = 1 \tag{8.10}$$

$$0 \;\leq\; P(j) \;\leq\; 1 \tag{8.11}$$

and using normalized density functions

$$\int \; p\,(x|j)\; dx \;\; = \;\; 1, \tag{8.12}$$

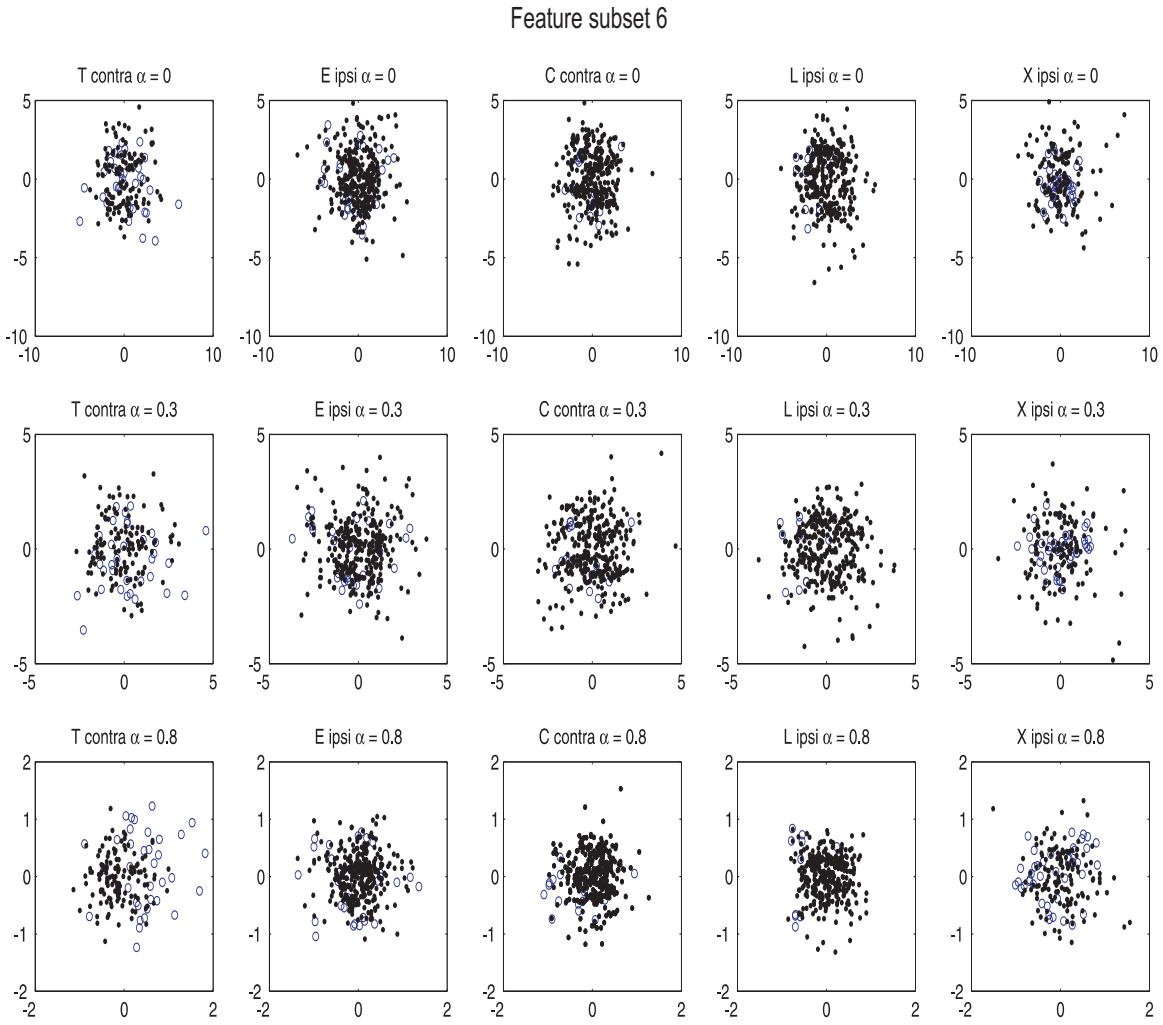


Figure 8.16: Partially supervised training with PLV feature subset 6. Supervised training with $\alpha = 0.8$ shows some distinction in the class clusters for T, C, E and L scalp. Unsupervised training and supervised training with $\alpha = 0.3$ shows much overlapped clusters for all patients.

guarantees that the model represents a density function.

The Gaussian models are parameterized by the mean vector μ of dimension d, a covariance matrix and the weights of all component densities. These parameters are represented by the notation: $\theta = \{\mu_j, \sum_j, P_j\}$, where $j = 1, 2, ...M$.

The covariance matrix used here is of a spherical form, such that it is a scalar multiple of the identity matrix $\sum_j = \sigma_j^2 \mathbf{I}$. The Gaussian distribution is given by

$$p(x|j) = \frac{1}{(2\pi\sigma_j^2)^{d/2}} \; exp \left\{ \; - \frac{\| \; x - \mu_j \; \|^2}{2\sigma_j^2} \right\}. \tag{8.13}$$

The model is a universal approximator and can model any density function arbitrarily closely if they have enough components.

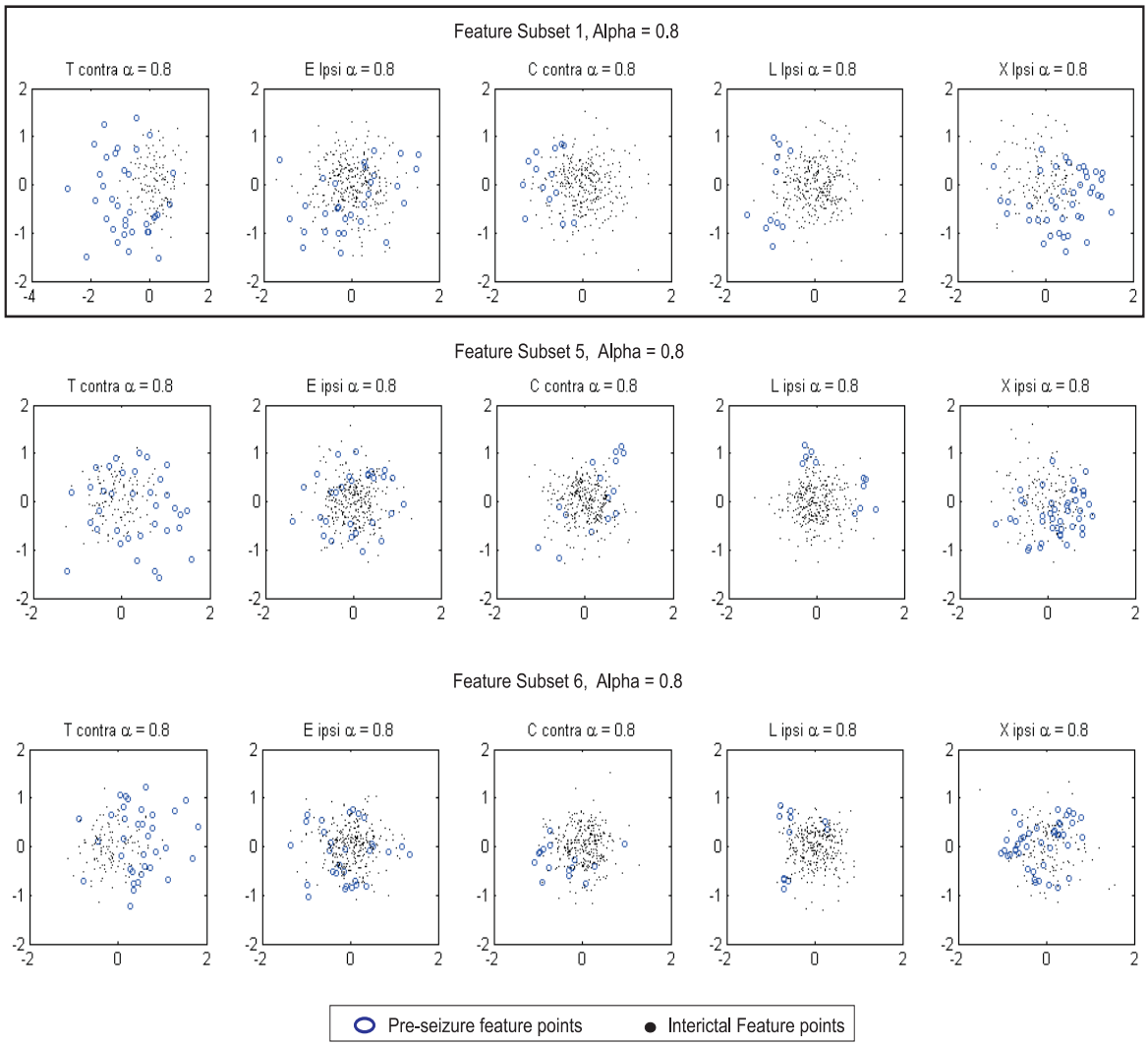


Figure 8.17: 2-D mappings of high dimensional feature space using Neuroscale Topographic projection technique. These are shown for five patients (T, E, C, L, X scalp) using feature subsets 1, 5 and 6. The transformation shown here is partially supervised with $\alpha = 0.8$, making use of topographic information as well as class information. The feature points belong to either class C_1: Pre-seizure (marked by blue circles) or C_2: Interictal (marked by black dots). The cluster patterns obtained from feature subsets 1, 5 and 6 were found to be showing less overlap of the two classes. The feature subset 1 and 6 can be seen to show more visual distinction than subset 5.

8.3.1 Supervised Learning and Classification with Labeled Data

In a supervised learning GMM, suppose a data set contains data points from c classes $C1$, C_2, C_3....C_c. A general approach to classify these data points is to use the posterior probabilities of the class memberships: $P(C_k \mid x)$ [176]. Such a classifier can be developed as a model that estimates the density of each class in a training data set and finds a decision boundary between the classes using Bayes' Theorem. In this case, the density models are estimated for each class in turn. This estimates $p(x \mid C_k)$ by training the model only on data from class C_k. then the Bayes' theorem is used, which

has the form

$$P(C_k \mid x) = \frac{p(x \mid C_k)P(C_k)}{p(x)}, \tag{8.14}$$

where the unconditional density $p(x)$ is given by the sum of the numerator over all the classes as

$$p(x) = \sum_{k=1}^{c} p(x \mid C_k)P(C_k). \tag{8.15}$$

This ensures that the estimated posterior probabilities sum to one. The prior probability $P(C_k)$ is simply given by the fraction of the data points from class C_k in the training set. The trained GMM model is then given a test data set and the probability of a given point occurring in the GMM of a particular class helps to track the probability of the test point being in that class.

1. **GMM Training Phase** The GMM is trained using the training data set and then tested on the test set. The objective of the retrospective GMM Training is the estimation of the component parameters $\theta = (\mu_j, \sum_j, P(j))$ of the multiple components of the Gaussian Mixtures, given the 2-dimensional feature space from Neuroscale. Here, one of the components j is first selected at random with a probability $P(j)$, which makes $P(j)$ as the *prior* probability of the component j. Then a data point is generated from the probability density $p(x|j)$. The posterior probability is then calculated using Bayes' theorem as

$$P(j|x) \;=\; \frac{p(x|j)P(j)}{p(x)}. \tag{8.16}$$

The model parameters of the GMM are determined by maximizing the data likelihood. This is equivalent to minimizing the negative log likelihood of the data set

$$E \;=\; -\sum_{n=1}^{N} log p(x^n). \tag{8.17}$$

The minima of this error function is found by an algorithm called the *expectation maximization* or the EM algorithm. The EM algorithm is faster to converge, is easy to implement and avoids calculation and storage of derivatives [176]. The EM algorithm consists of two main steps: expectation (E-step) and the maximization (M-step). In the E-step, the expected values of the statistics are calculated, given the parameter estimates. The M-step then makes the maximum likelihood estimates of the parameters, given the estimated values of the statistics. Initially a random set of parameters is generated and the E and M-steps are alternated until a predefined maximum number of iterations has been exceeded or the parameter estimates in two consecutive iterations is less than a predefined error tolerance.

It should be noted that the number of components that model the distribution of each class are not known, but the task of the training classifier is to only obtain

a good enough approximation of the distribution of each class. Therefore the number of components need not be guessed accurately, it is just a parameter that defines the complexity of the distribution. Nevertheless, too small a number will prevent the classifier from forming a good approximation for a small training set, while a large number of components may lead to an over-fitted classifier. Also, too large a number can lead to singularities especially with a small training data set [191].

The GMMs for the patient data were modelled with 3-5 components each for predictive feature points and interictal feature points from the training data. The number of components were varied according to the data, the number of seizures it had, the number of predictive feature points that were being modelled, the amount of training data being used, etc. The neuroscale 2-D projections of feature subset 1 and 6 both were analyzed and compared. After the GMMs for interictal and pre-ictal were modelled using the training data, the models were tested for classification.

2. **GMM Testing Phase** Once the parameters of the GMMs have been fitted or the model has been trained, the likelihood ratio is subsequently simulated with an on-line test phase. In the test phase the feature points irrespective of the class are given to the two GMMs and the probability of the points being in the pre-ictal GMMs is acquired for every point. This gives a probability map across time, that shows the probability of a prediction event across time. Irrespective of the amount of probability that leads to prediction, an arbitrary threshold is used in the first instance to simply check when its probability exceeds a certain level. A probability exceeding threshold would imply the occurrence of an event, and it would be called a *prediction event* if it lies within a prediction window. The selection of an appropriate prediction window is not trivial and is discussed in the next section. Nevertheless, the suitability of a series of prediction windows starting from 0 minutes to 4 hours is analyzed, with varying threshold levels, in order to fine tune the predictor to the patients data.

   The results of the training and test phase have been combined and shown in Figures 8.21, 8.23, 8.25, 8.28, 8.32, 8.36 for five patients for feature subset 1 and 6. Each figure shows the training feature map obtained from Neuroscale and the Gaussian mixtures that are fitted to the interictal and pre-ictal classes. It also shows the probability contour maps of the density functions of the trained maps for the two classes. The test points are obtained from the Neuroscale test phase. These points are then projected on these density functions. The probability of a point being projected in the pre-ictal density map shows the probability map of pre-seizure events across time. The following statistical analysis (following section) shows that feature subset 6 gives better classification as compared to subset 1, across all patients.

8.3.2 Semi-Supervised Learning with Labelled and Unlabelled Data

In semi-supervised learning the learner is provided with a set of labelled data from a training set as well as unlabelled data from the test set. The aim of the classifier is to learn the data groupings from the information provided by the labelled data as well the unlabelled data. An attempt to learning from unlabelled data is similar to unsupervised learning. The classifier can learn about the distribution, or structure underlying labelled and unlabelled data. This can be advantageous in modelling the labelled class as it would incorporate information that labelled data could not provide, yet was intrinsic (as an underlying structure) to the class. Alternatively, labelled data may provide guidance for the classifier, but as it would be sparse, the complete picture might only be obtained from the unlabelled data.

Thus, a semi-supervised learning can also be thought of as an unsupervised learning where the classifier uses additional information from the labelled data to assist in its unsupervised learning. Especially in the case of only two classes, when the labelled data is only from one class, it may help to remove ambiguities in cluster correspondence. For example, let a distribution of data from two classes, shown in Figure 8.18 be given to such a classifier, without any labels, for an unsupervised learning task. It may form groupings of the unlabelled data such as those shown by the circles in (a) and (b). The labelled data may assist it to select the more suitable fit (the one shown in (b)) leading to a more efficient training. In a multidimensional learning task, certain other constraints can also be imposed on the labelled data [192]. The semi-supervised

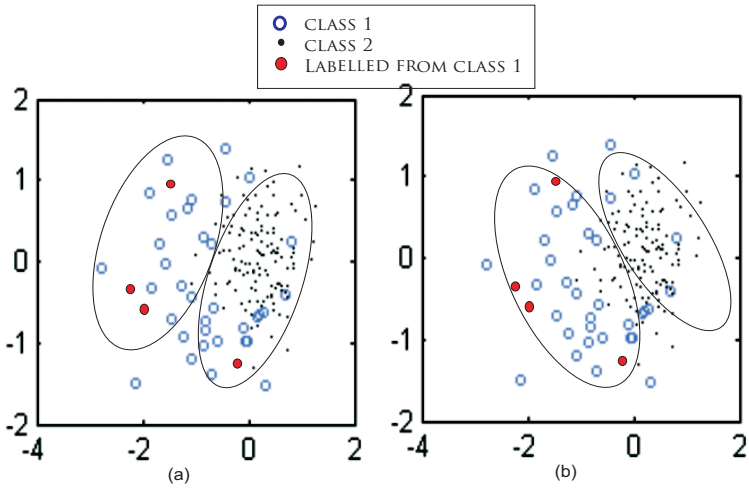


Figure 8.18: Semi-supervised learning task with labelled and unlabelled data (on a patient T scalp Neuroscale map) (a), (b) shows possible groupings of data formed by an unsupervised learning of unlabelled data. (b) a better fit for efficient classification is selected using the information provided by the labelled data (*red filled circles*)

analysis with GMM is carried out on similar lines. The data is known to be distributed in two classes: pre-ictal and interictal. The data is first divided into a training and

a test set, such that the training set has sufficient seizure activity. The labelled data is the pre-ictal feature points obtained from the training set. The rest of the test and training data is used as the unlabelled data. The GMM is first trained on the labelled data and then only on the unlabelled data and finally on the unlabelled data using the information from the GMM of the labelled data. The density maps are computed for the final GMM trained on unlabelled and labelled data. This set of GMMs is then given the entire data set and the classification accuracy is measured using the known label information. Thus few labelled data points can essentially help in a good generalization performance. Compared to supervised learning, it may achieve a low classification error for the same task [190]. Semi-supervised learning with GMM involves the EM algorithm solving for the GMM parameters and alternating between providing soft labels to the unlabelled data.

8.4 Statistical Analysis

The performance of the seizure prediction is evaluated using the most commonly used statistical measures: sensitivity and specificity. The aim of this statistical analysis is to ascertain the efficiency of the features extracted for seizure prediction in (1) raising the alarm in a timely manner, and (2) not raising the alarm unnecessarily. This analysis uses the two dimensional probability curves of event occurrence obtained from GMM and Neuroscale. In order to measure the prediction events the following two parameters need to be defined:

8.4.1 Prediction window

The prediction window is a window of time before a seizure onset when the alarm is raised about the oncoming seizure. Ascertaining the appropriate length of such a prediction window and the *goodness* of sensitivity and specificity depends on certain factors such as application, intervention methods, etc. described as follows:

Therapeutic intervention: The appropriateness of a prediction window would be quite dependent on the therapeutic intervention that is used in conjunction with seizure prediction. Although the following mechanisms are still under research, they help to show the variations tolerable in a 'suitable' prediction window.

- *Deep brain stimulation:* Use of deep brain stimulation [16, 17] for preventing a seizure onset may require longer prediction windows, where the increasing synchrony is curtailed much before a seizure onset to prevent a possible seizure. It may not be useful with short prediction windows. In this case, if extra stimulation is safe and does not cause side effects, a low specificity

may very well be tolerated if it delivers high sensitivity. However, if the stimulation is used only for extreme cases to curtail strong oncoming seizures like a *Grand mal* and leads to subtle side effects, low specificity may not be tolerable, even with high sensitivity as it could perhaps be equally harmful.

- *Injecting fast acting medications:* Another mechanism for intervention is infusion of fast acting drugs before a seizure onset, as opposed to taking regular long term anti epileptic medication [20, 21]. Such a method as mentioned by Smith *et al.* [20] could be used to arrest a seizure focus. The most suitable prediction window for such an intervention would be of the order of seconds. Perhaps the strength of the drugs (with probably some content of anesthetics as in the study [20]) would not allow tolerance of low specificity leading to unnecessary injection of strong drugs. A low sensitivity may be tolerated with high specificity in case the seizures are not too disarming or dangerous.
- *Transcortical cooling:* Another method being proposed for seizure prevention or arrest is transcortical cooling [193, 18, 19]. This method involves cooling the seizure focus to prevent the seizure onset and as mentioned by Burton *et al.* a period of few minutes (3 minutes before a seizure) was used to cool the seizure focus to abort or suppress focal seizures. A lower degree of side effects or dangers involved, allows for a low specificity and high sensitivity in such a case.

Warning mechanism: In the case when the seizure prediction system is applied as a warning mechanism such as warning the patient to move out of danger a very long prediction window of the order of two hours may not be helpful. Even though the patient would be warned, but it would keep him anxious and uncomfortable for two hours. Anxiety might trigger panic attacks similar to epileptic attacks. Then a clinical application may not be tolerated with low specificity or low sensitivity. On the other hand, the uncertainty of seizure occurrence in epilepsy also causes undue anxiety in patients, which might be relieved when a seizure warning is obtained, even with a longer prediction window and low specificity.

Injecting tracers/diagnostic: The prediction system can also be applied as a diagnostic by a clinician to allow him to inject tracers in the patient for improved imaging of the brain. This may tolerate prediction windows as long as 5-20 minutes or even 1-2 hours with a high specificity and sensitivity. Whereas, if the prediction system is being used leading to an epileptic operation for localization of the abnormal tissues, a prediction window of the order of few hours may be useful with a very high specificity and 100% sensitivity.

Injecting slow acting medications: In case the prediction system is used to inject slow acting milder drugs, which may be better than using long term continuous strong dosage of anti epileptic drugs, especially in children, a longer prediction

window may be tolerated with low specificity and high sensitivity.

Frequency of seizures: In general the prediction window would be highly dependent on the frequency of seizure occurrence. A one hour prediction window for a patient having 26 seizures a day would be as good as a random predictor. While, even a ten hour prediction window for a patient having one seizure in six months would be excellent.

The above scenarios show that it is not trivial to ascertain general apposite characteristics of a seizure prediction system such as the length of a prediction window. Nor is it trivial to determine the apt sensitivity and specificity to judge the performance of the prediction system, without the knowledge of the frequency of seizures, application, methods of intervention and the tolerance capacity of the patient. Such a performance would be more suited for a patient specific system. Nevertheless, the sensitivity and specificity of the system researched here are quoted without denoting its *goodness* solely based on the numbers obtained. The prediction windows are also varied from a length as short as five minutes to as long as four hours.

8.4.2 Threshold

An event is considered to have occurred when the probability curve exceeds an (arbitrary) threshold. This event is marked as a prediction when it lies within a prediction window. It is marked as a false positive when it lies outside this prediction window. The threshold to be used can be any arbitrary level from zero to one (as the probability varies from zero to one). The most suitable threshold would be based on obtaining the sensitivity and specificity required.

In order to analyze the performance of the features extracted for seizure prediction, the prediction window is varied from 5 minutes to 4 hours in steps of 5 minutes and the threshold is varied from zero to one in steps of 0.01 (see Figure 8.19). The sensitivity, false positive rate based on the varying prediction window and the varying threshold can be visualized with a three dimensional graph, where the threshold is the x-axis, prediction horizon is y-axis and the sensitivity or false positive rate is denoted by the color map (using a fixed color axis).

8.4.3 Seizure Prediction Performance

In order to obtain the sensitivity and false positive rate the following measures are estimated from the two dimensional probability curves of each patient, for each prediction window and each threshold level.

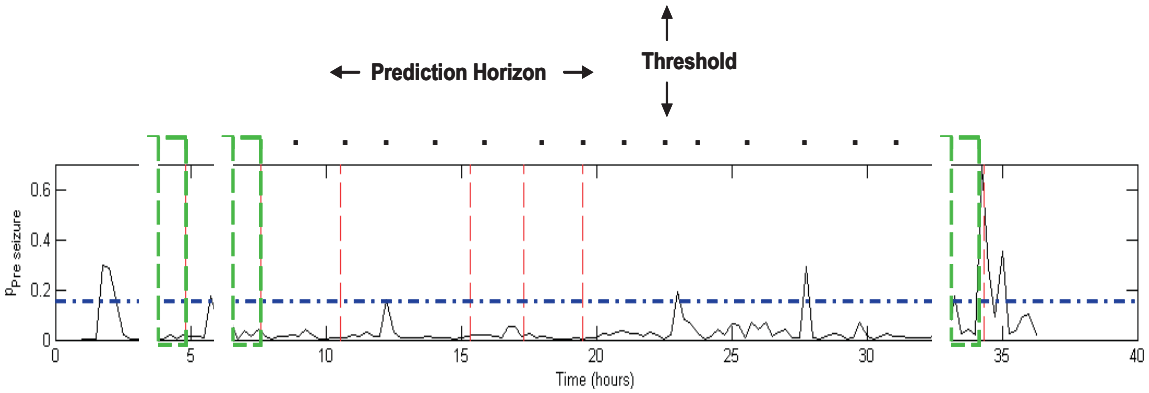


Figure 8.19: Sensitivity analysis is performed for every patient as shown here, for each threshold level (0-1.0) with the prediction window varying from five minutes to four hours in steps of five minutes. The vertical lines indicate seizure onset times.

1. **True Positives (TP):** A Correct prediction within the prediction horizon.
2. **False Positives(FP):** A point exceeding the threshold causing an alarm but not within a prediction window.
3. **False Negatives (FN):** A point lower than the threshold within a prediction horizon (not correctly identified as a

Based on the above values the following statistical measures are hence calculated for each prediction window and all threshold levels.

- Sensitivity
- False Positive Rate
- Specificity
- Receiver Operating Characteristics

The sensitivity, false positive rate, specificity and seizure prediction characteristics for the five patients are shown in the following sections. These are followed by the ROC curves for each patient.

8.4.3.1 Sensitivity

The Sensitivity (S) is the measure of correctly identified events. It is estimated as follows:

$$S = \frac{TP \times 100}{(TP + FN)}, \tag{8.18}$$

equivalent to

$$S = \frac{Number\ of\ seizures\ detected}{Total\ seizures} \times 100\%. \tag{8.19}$$

8.4.3.2 False Positive Rate

The False Positive Rate (FPR) is an indicator of performance and it measures the instances that are incorrectly identified as positive. In the seizure prediction system, a false positive would essentially be an event that is not followed by a seizure within a prediction window. However, considering a clinical application, a false positive is an event that is not followed by a seizure within a prediction window, but in a real application the patient would not know if a positive will not be followed by a seizure and so a *waiting period* of the length of the prediction window will occur. Any positives during this waiting period will have to be ignored by the system, only accounting for the last false positive in that prediction window (in case this last false positive is a predictor). A false positive rate based on this *waiting period* has been used in the statistical analysis that follows

$$FPR = \frac{FP}{number\ of\ total\ hours}. \tag{8.20}$$

8.4.3.3 Specificity

Specificity (Sp) is defined as the measure of correctly identified non-events

$$Sp = \frac{TP\ \times\ 100}{(TP + FP)}, \tag{8.21}$$

equivalent to

$$Sp = \frac{Number\ of\ correct\ detections}{Total\ detections} \times 100\%. \tag{8.22}$$

8.4.3.4 Receiver Operating Characteristics

ROC curves are used to assess the performance or accuracy of test in discriminating the correctly identified events from the incorrectly identified events [194]. They depict a relationship between the sensitivity and specificity of the analysis. the true positives (sensitivity) is plotted against the false positive rate (100-specificity) for different cut off points (in this case threshold levels). Each point on the ROC curve depicts a pair of sensitivity and specificity values corresponding to a threshold. A high sensitivity and specificity is depicted by a curve that passes through the upper left corner (see Figure 8.20). Therefore, the closer the curve is to the upper left corner, the better is the performance of the system. The diagonal line is the line of 'no discrimination'. A point lying on this line would have a 50% accuracy and hence would be as good as a random guess.

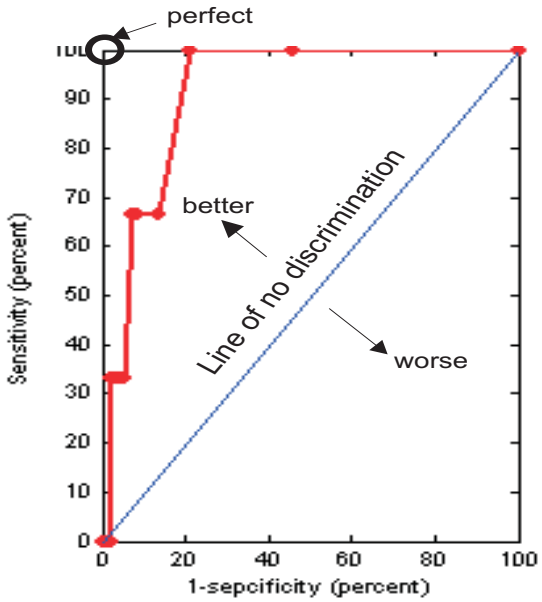


Figure 8.20: A general illustration of an ROC Curve. The best possible prediction method would yield a point at the upper left corner of the ROC space (circled) as it represents 100% sensitivity and 100% specificity. It is also called the perfect classification point. Therefore, the closer the curve is to the upper left corner, the better is the performance the system. The diagonal line divides the ROC space into areas of good and bad classification. The diagonal itself represents the line of no discrimination or the *random guess* line. Points below this line represent bad classification.

8.5 Discriminant Analysis: Results

8.5.1 Results I: Neuroscale and Supervised GMM Classifier

The results of the application of supervised Neuroscale dimension reduction and topographic projection, followed by a supervised GMM classifier, are presented here. The procedure used for the analysis is as follows:

Pre-Processing: The multidimensional feature points obtained from the PLV-d curves are first labelled to be a member of either of the two classes: (1) Pre-ictal or (2) Interictal. The data and their labels are then divided into a *training* and a *test* set. The training set is created such that it includes at least a few seizure events and sufficient interictal data points, in order to have non-zero data points in the pre-ictal class and a good training of the classifier.

Neuroscale Training: The multidimensional feature vectors from the training set are then used to train a Neuroscale network. The Neuroscale is used in a *supervised* mode. The value of α used was 0.8. The Neuroscale training reduced the multidimensional feature vectors to two dimensional as well as learned the mapping parameters. These optimized mapping parameters were stored to be used for the *test phase*.

GMM Density Estimation: A GMM is then modelled for each of the pre-ictal and interictal class separately, but such that the total posterior probabilities sum to one. As the GMM are modelled on each data separately, it is said to be a supervised training. Once the GMMs are trained, their estimated parameters are stored for the *test* phase. The density contours are projected on the data points.

Neuroscale Test Phase: The test phase follows the training phase. The multidimensional test feature vectors are first transformed into two dimensional feature space using the trained Neuroscale network.

GMM Test Phase: The two dimensional feature space obtained from Neuroscale is then given to the trained set of GMM models. The test data points get classified as pre-ictal or interictal points according to the GMM in which they get projected. The probability contours obtained in the training phase help to track the probability of the test data point to exist in the GMM of a certain class. The probability of the given feature points being projected into the *pre-ictal* Gaussian mixtures is hence tracked. This enables tracking the probability of a feature point being predictive. This probability is plotted against time for each data point giving a two dimensional predictive probability curve. An event (non-zero probability) on the probability curve is then termed as a predictive event if it exceeds an arbitrary threshold level.

Statistical Analysis: The statistical analysis is then performed on the probability curve obtained above, to assess a suitable *threshold level* that would ensure a high number of predictions (high sensitivity) and low false positive rate, hence a high specificity. Along with the threshold, another parameter that needs to be optimized in this analysis is the *length of the prediction window*. A suitable prediction window would allow maximum sensitivity and max specificity. It should be noted that it is not trivial to define a 'suitable' length of prediction window as discussed in the previous section. Also, the PLV-d curves created here indicate slower dynamics over hours, making it futile to track events occurring on the timescale of seconds. Therefore, more suited to the present analysis a length of 0 hours to 4 hours is analyzed in steps of 5 minutes. This analysis is followed by creating ROC curves, that show the relationship between sensitivity and specificity for various prediction windows. An ROC curve is created for each prediction window and every point on the curve corresponds to the sensitivity, specificity at a different threshold level for that prediction window. The ROC curve helps to determine the prediction window that would have a high sensitivity and high specificity and the threshold level at which that occurs.

Test with Random Predictor: The last stage of the analysis is a comparison with a random predictor. The performance of the analysis is compared to a random predictor. The random predictor generates 50 surrogate prediction times. These points correspond to the times when the probability curve has exceeded the threshold irrespective of the threshold level. The sensitivity of prediction with these random predictions is then calculated. The intelligent predictor designed above is required to perform better than a random predictor in order to substantiate the intelligence in the predictor being designed.

The feature subset 6 and 1 have been used in the following analysis and the Neuroscale-GMM train and test classifier is developed for each of the five patients. Each of the following figures in this section show:

1. Multidimensional training feature vectors projected by a Neuroscale network on to a two dimensional map. Feature points from different classes are identified. The trained Gaussian centers and mixtures for each class are drawn on the Neuroscale map.

2. The training and the test data is projected on the same Neuroscale map and the points form different classes are identified. The Gaussian density contours are superimposed on the feature space.

3. The probability plot of a data point being in the pre-ictal density map is drawn vs. time. The training and the test boundary is highlighted, showing the interictal data and seizure events used for training and testing.

4. The sensitivity graph is shown for prediction windows varying from 0-4 hours in steps of 5 minutes and threshold levels varying from 0-1 in steps of 0.01.
5. False positives per day are shown for all prediction windows and threshold levels.
6. The specificity graph is shown for all prediction windows and threshold levels.
7. A set of ROC curves are shown for various prediction windows in steps of 30 minutes (due to constraints of space).

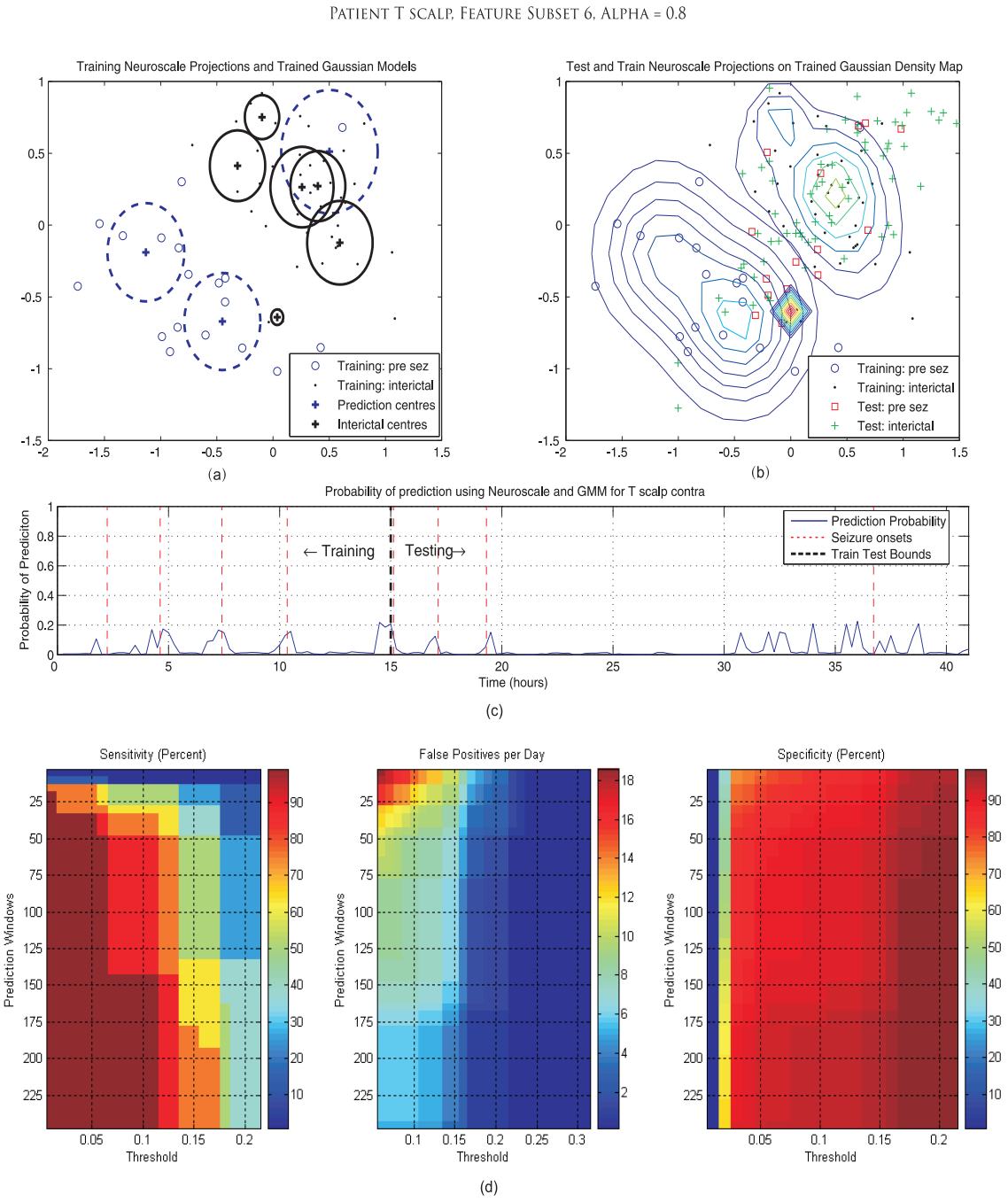


Figure 8.21: The figures show the seizure prediction analysis for patient T scalp. (a) The two dimensional feature map obtained for the training data using supervised Neuroscale, feature subset 6 and $\alpha = 0.8$. Interictal class (*dots*), pre-ictal class (*circles*). Gaussian mixtures trained on both the classes are shown. One standard deviation for interictal GMM (*solid lines*), for pre-ictal GMM (*dashed lines*), centers (*crosses*). (b) Training and test data on the same Neuroscale map. Pre-ictal (*circles*), interictal (*dots*), test pre-ictal (*squares*), test interictal (*crosses*). Gaussian density contours of both classes are shown. (c) Probability plot of a data point being in the pre-ictal density map vs. time. Seizure onsets (*thin vertical dash lines*), Train-Test boundary (*thick dash line*). (d) Sensitivity, false positives per day and specificity for prediction windows varying from 0-4 hours in steps of 5 minutes and threshold levels varying from 0-1 in steps of 0.01.

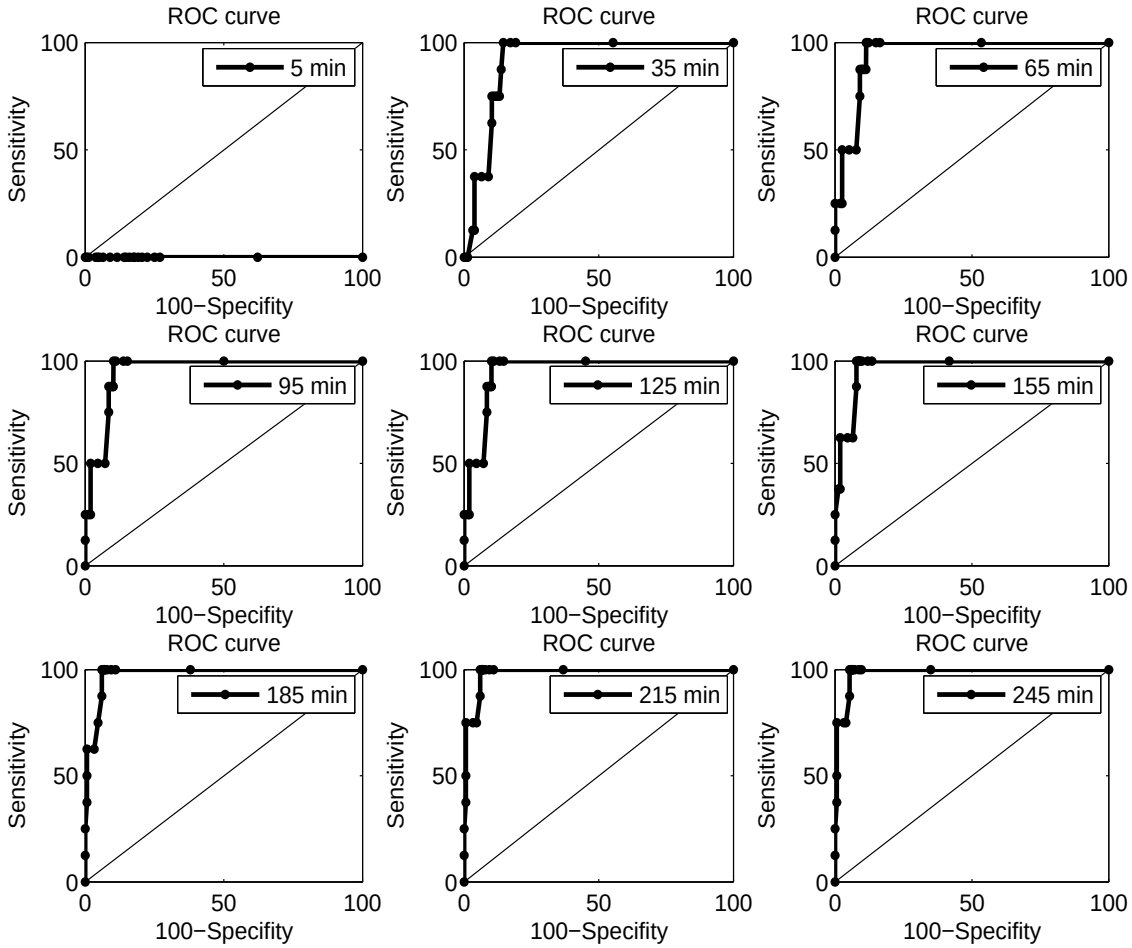


Figure 8.22: ROC curves for **patient T scalp**, for supervised Neuroscale GMM density based seizure predictor using **feature subset 6** and $\alpha = 0.8$. ROC curves are shown for prediction windows varying from 5 minutes to 245 minutes in steps of 30 minutes. Each point on a ROC curve indicates the sensitivity and specificity for every threshold level in that prediction window. The ROC curves show a 100% sensitivity and 80% specificity for the prediction window of **35 minutes**.

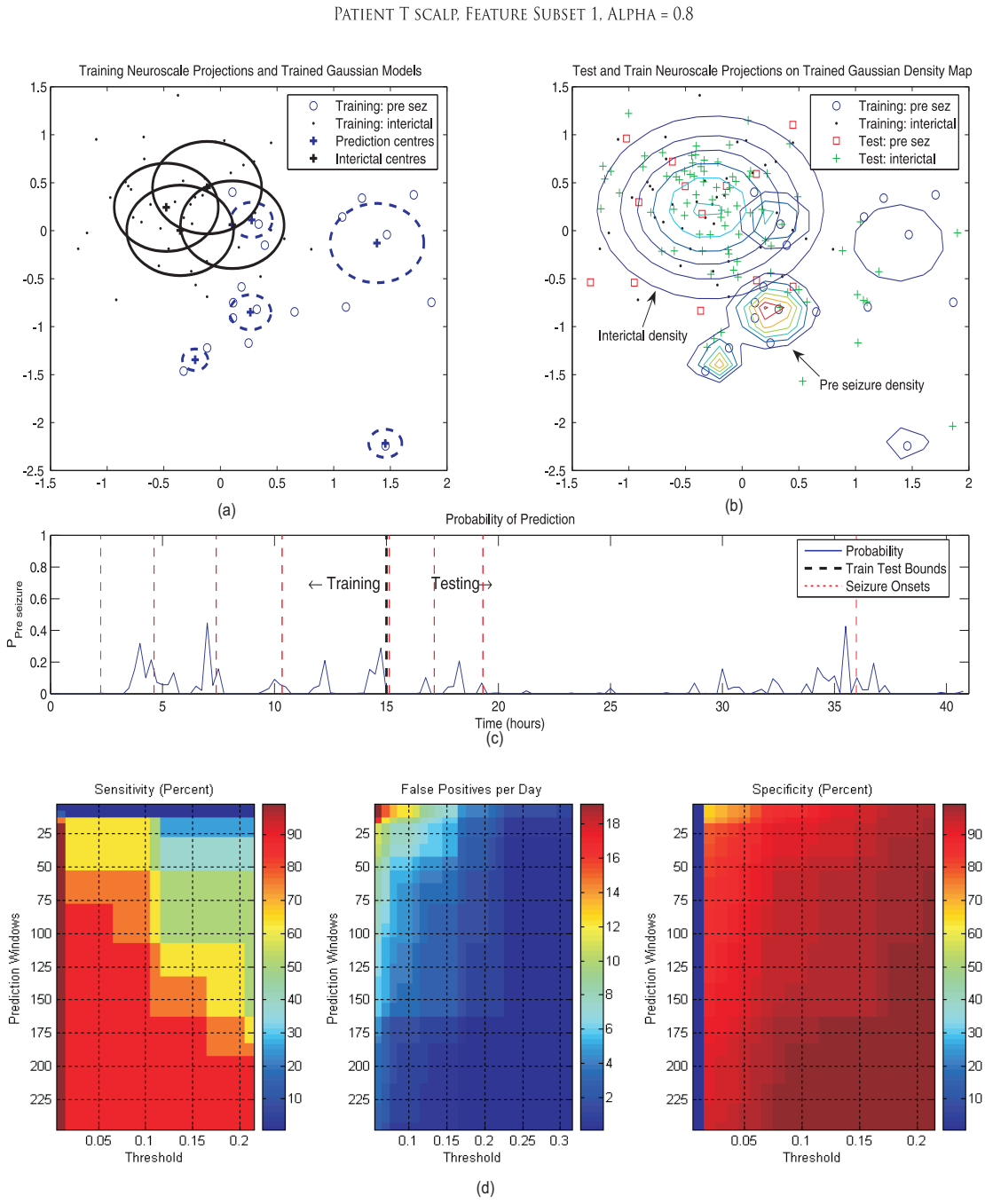


Figure 8.23: The figures show the seizure prediction analysis for patient **T scalp**. (a) The two dimensional feature map obtained for the training data using supervised Neuroscale, **feature subset 1** and $\alpha = 0.8$. Interictal class (*dots*), pre-ictal class (*circles*). Gaussian mixtures trained on both the classes are shown. One standard deviation for interictal GMM (*solid lines*), for pre-ictal GMM (*dashed lines*), centers (*crosses*). (b) Training and test data on the same Neuroscale map. Pre-ictal (*circles*), interictal (*dots*), test pre-ictal (*squares*), test interictal (*crosses*). Gaussian density contours of both classes are shown. (c) Probability plot of a data point being in the pre-ictal density map vs. time. Seizure onsets (*thin vertical dash lines*), Train-Test boundary (*thick dash line*). (d) Sensitivity, (e) False positives per day and (f) Specificity for prediction windows varying from 0-4 hours in steps of 5 minutes and threshold levels varying from 0-1 in steps of 0.01.

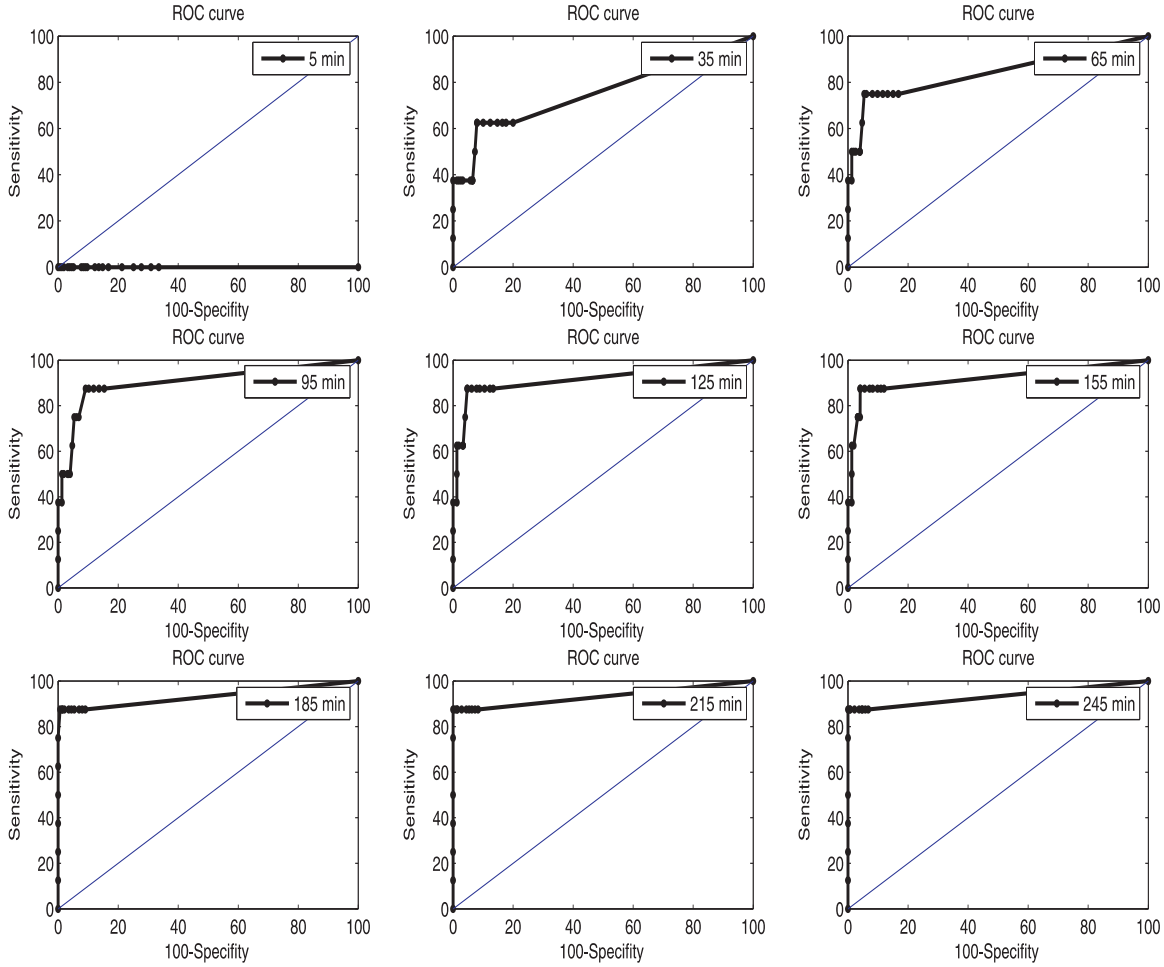


Figure 8.24: ROC curves for **patient T scalp**, for supervised Neuroscale GMM density based seizure predictor using **feature subset 1** and $\alpha = 0.8$. ROC curves are shown for prediction windows varying from 5 minutes to 245 minutes in steps of 30 minutes. Each point on a ROC curve indicates the sensitivity and specificity for every threshold level in that prediction window. The ROC curves show the maximum sensitivity of 90% and 90% specificity for the shortest prediction window of length **95 minutes**.

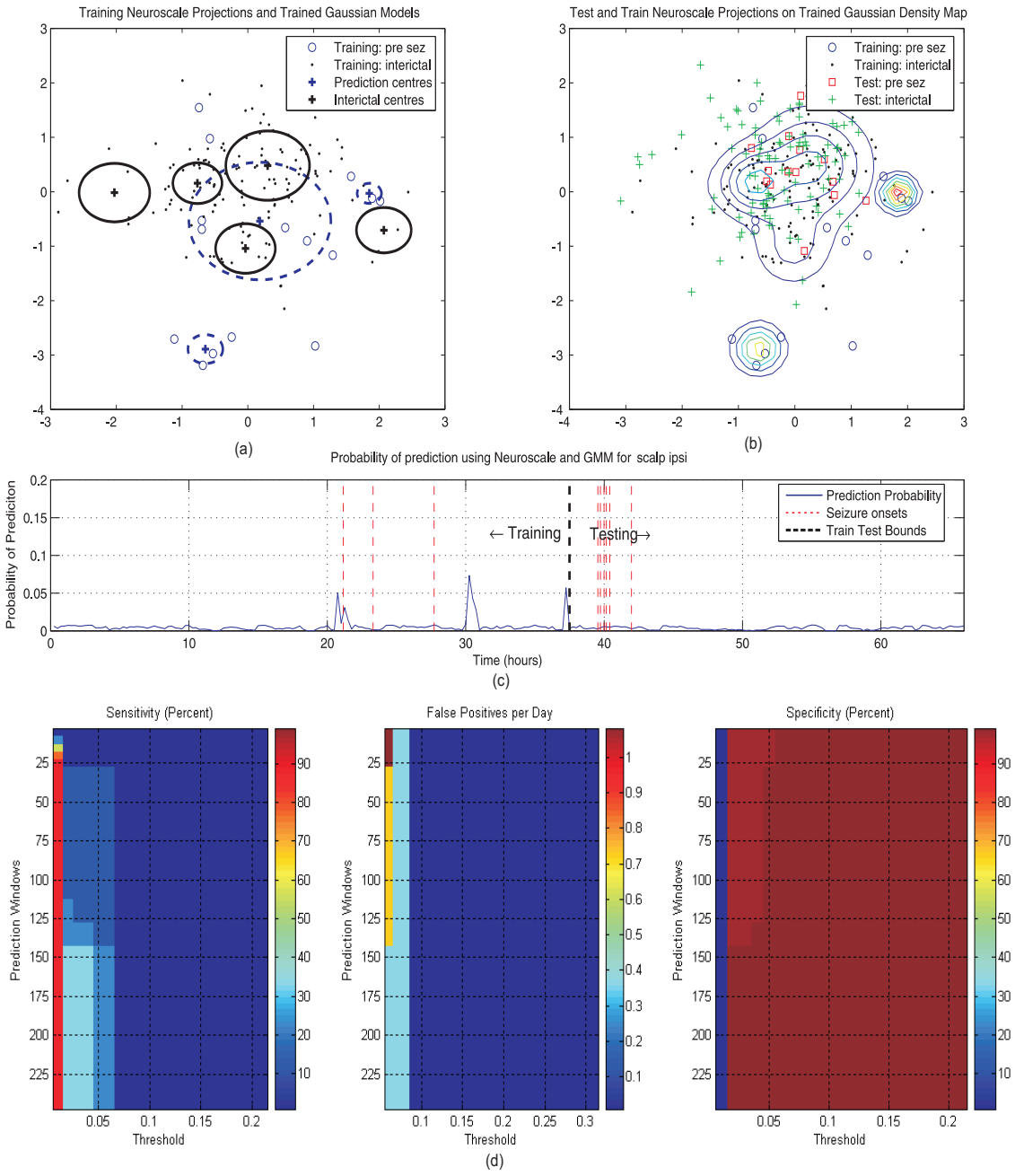


Figure 8.25: The figures show the seizure prediction analysis for patient **E scalp**. (a) The two dimensional feature map obtained for the training data using supervised Neuroscale, **feature subset 6** and $\alpha = 0.8$. Interictal class (*dots*), pre-ictal class (*circles*). Gaussian mixtures trained on both the classes are shown. One standard deviation for interictal GMM (*solid lines*), for pre-ictal GMM (*dashed lines*), centers (*crosses*). (b) Training and test data on the same Neuroscale map. Pre-ictal (*circles*), interictal (*dots*), test pre-ictal (*squares*), test interictal (*crosses*). Gaussian density contours of both classes are shown. (c) Probability plot of a data point being in the pre-ictal density map vs. time. Seizure onsets (*thin vertical dash lines*), Train-Test boundary (*thick dash line*). (d) Sensitivity, (e) False positives per day and (f) Specificity for prediction windows varying from 0-4 hours in steps of 5 minutes and threshold levels varying from 0-1 in steps of 0.01.

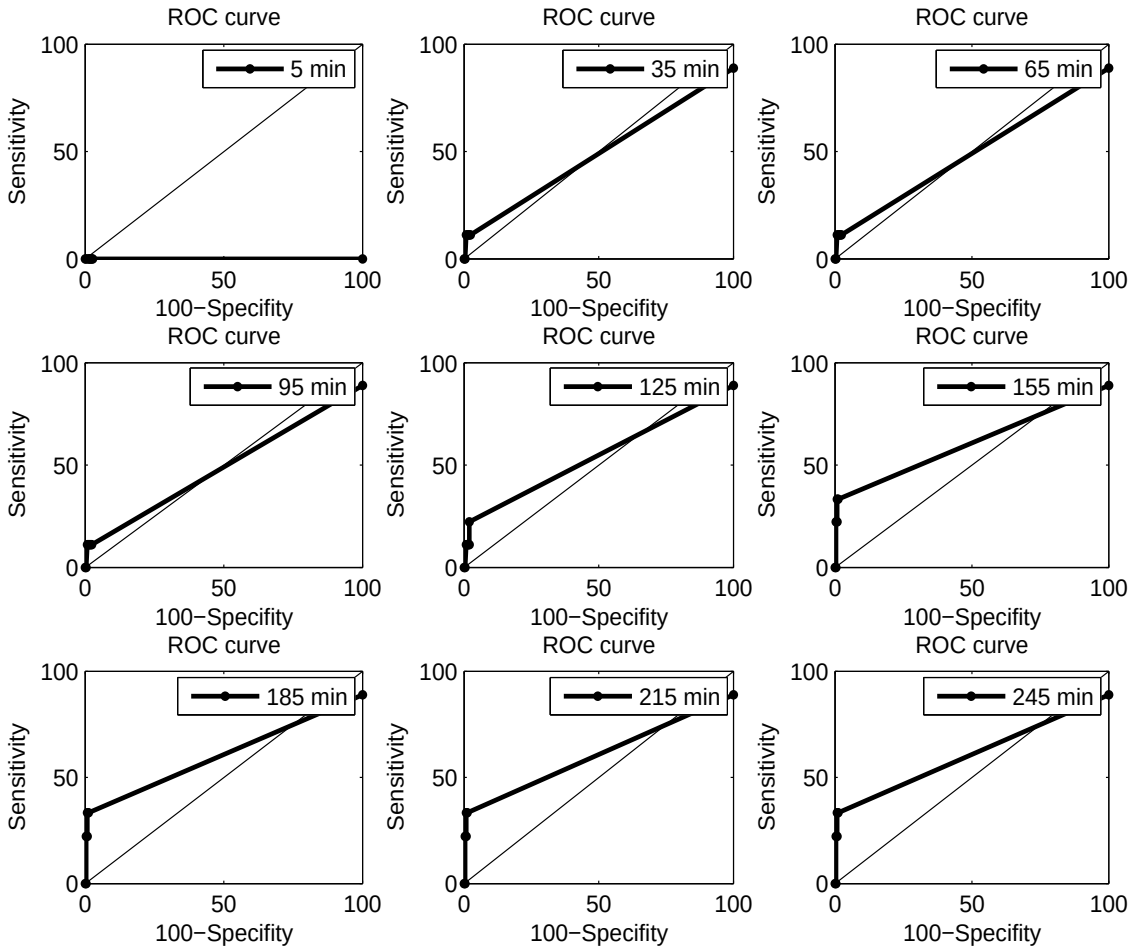


Figure 8.26: ROC curves for patient E scalp, for supervised Neuroscale GMM density based seizure predictor using feature subset 6 and $\alpha = 0.8$. ROC curves are shown for prediction windows varying from 5 minutes to 245 minutes in steps of 30 minutes. Each point on a ROC curve indicates the sensitivity and specificity for every threshold level in that prediction window. The ROC curves show a 40% sensitivity and 98% specificity for a prediction window of length **140 minutes**.

Figure 8.27:

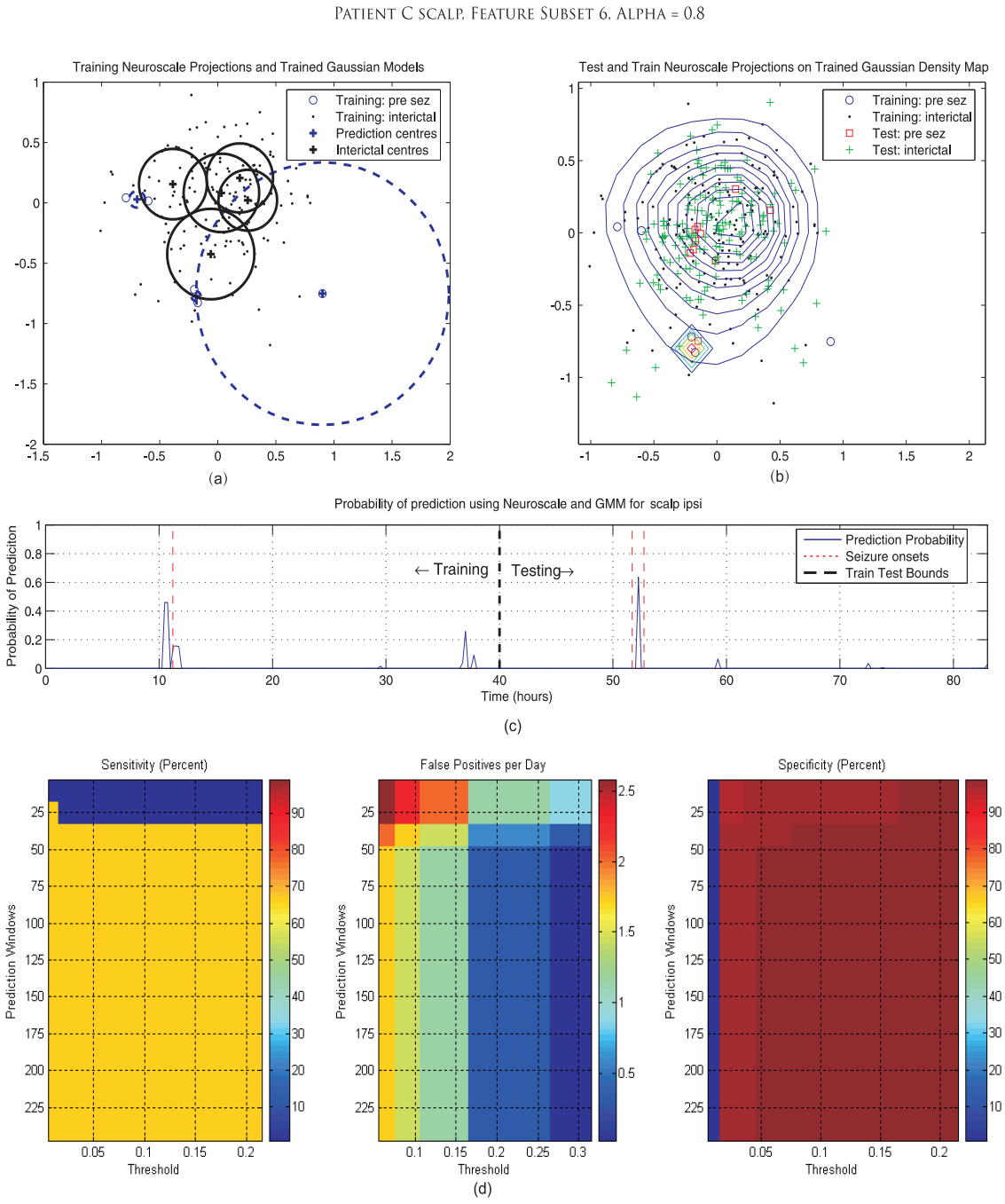


Figure 8.28: The figures show the seizure prediction analysis for patient **C scalp**. (a) The two dimensional feature map obtained for the training data using supervised Neuroscale, **feature subset 6** and $\alpha = 0.8$. Interictal class (*dots*), pre-ictal class (*circles*). Gaussian mixtures trained on both the classes are shown. One standard deviation for interictal GMM (*solid lines*), for pre-ictal GMM (*dashed lines*), centers (*crosses*). (b) Training and test data on the same Neuroscale map. Pre-ictal (*circles*), interictal (*dots*), test pre-ictal (*squares*), test interictal (*crosses*). Gaussian density contours of both classes are shown. (c) Probability plot of a data point being in the pre-ictal density map vs. time. Seizure onsets (*thin vertical dash lines*), Train-Test boundary (*thick dash line*). (d) Sensitivity, (e) False positives per day and (f) Specificity for prediction windows varying from 0-4 hours in steps of 5 minutes and threshold levels varying from 0-1 in steps of 0.01.

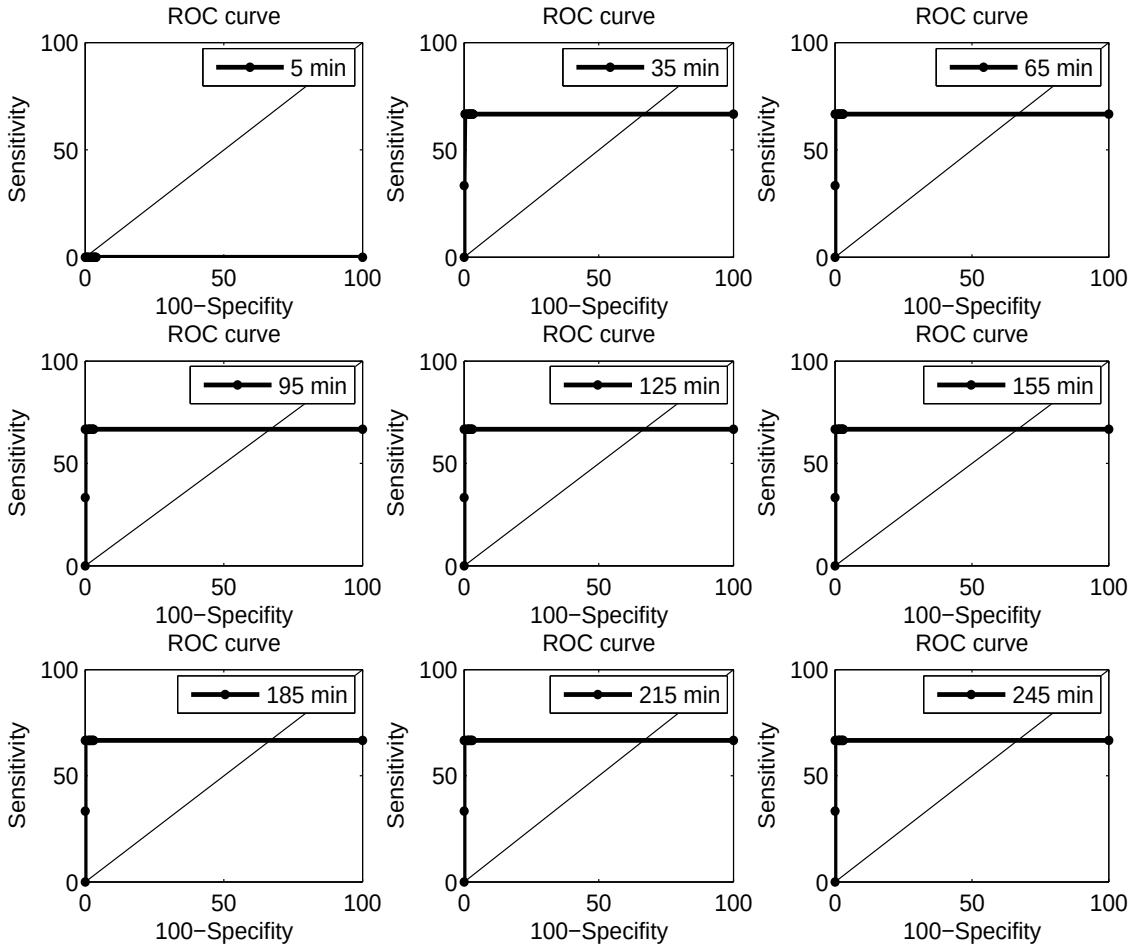


Figure 8.29: ROC curves for **patient C scalp**, for supervised Neuroscale GMM density based seizure predictor using **feature subset 6** and $\alpha = 0.8$. ROC curves are shown for prediction windows varying from 5 minutes to 245 minutes in steps of 30 minutes. Each point on a ROC curve indicates the sensitivity and specificity for every threshold level in that prediction window. The ROC curves show 67% sensitivity and 100% specificity for the shortest prediction window of length **35 minutes**.

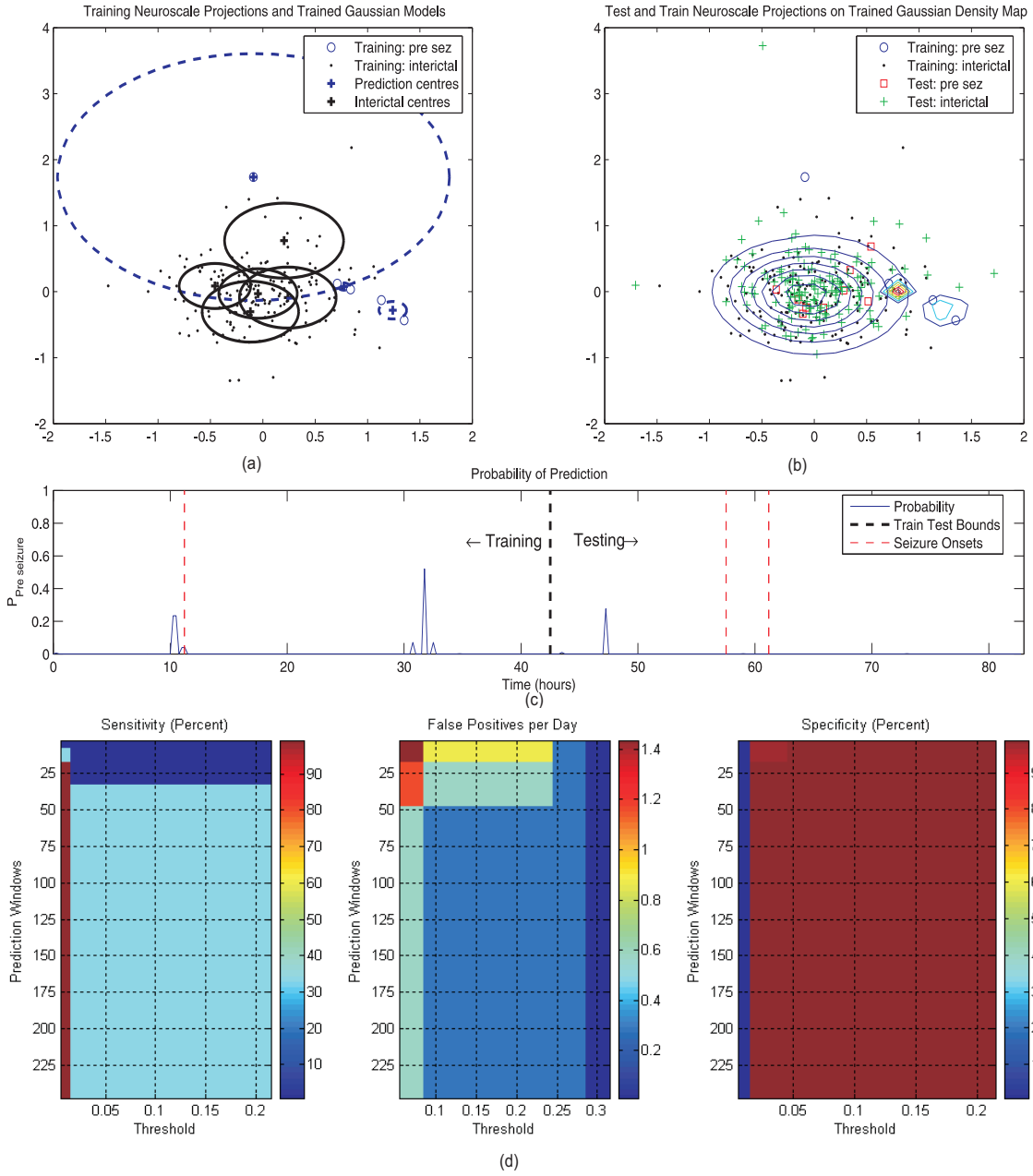


Figure 8.30: The figures show the seizure prediction analysis for patient **C scalp**. (a) The two dimensional feature map obtained for the training data using supervised Neuroscale, **feature subset 1** and $\alpha = 0.8$. Interictal class (*dots*), pre-ictal class (*circles*). Gaussian mixtures trained on both the classes are shown. One standard deviation for interictal GMM (*solid lines*), for pre-ictal GMM (*dashed lines*), centers (*crosses*). (b) Training and test data on the same Neuroscale map. Pre-ictal (*circles*), interictal (*dots*), test pre-ictal (*squares*), test interictal (*crosses*). Gaussian density contours of both classes are shown. (c) Probability plot of a data point being in the pre-ictal density map vs. time. Seizure onsets (*thin vertical dash lines*), Train-Test boundary (*thick dash line*). (d) Sensitivity, (e) False positives per day and (f) Specificity for prediction windows varying from 0-4 hours in steps of 5 minutes and threshold levels varying from 0-1 in steps of 0.01.

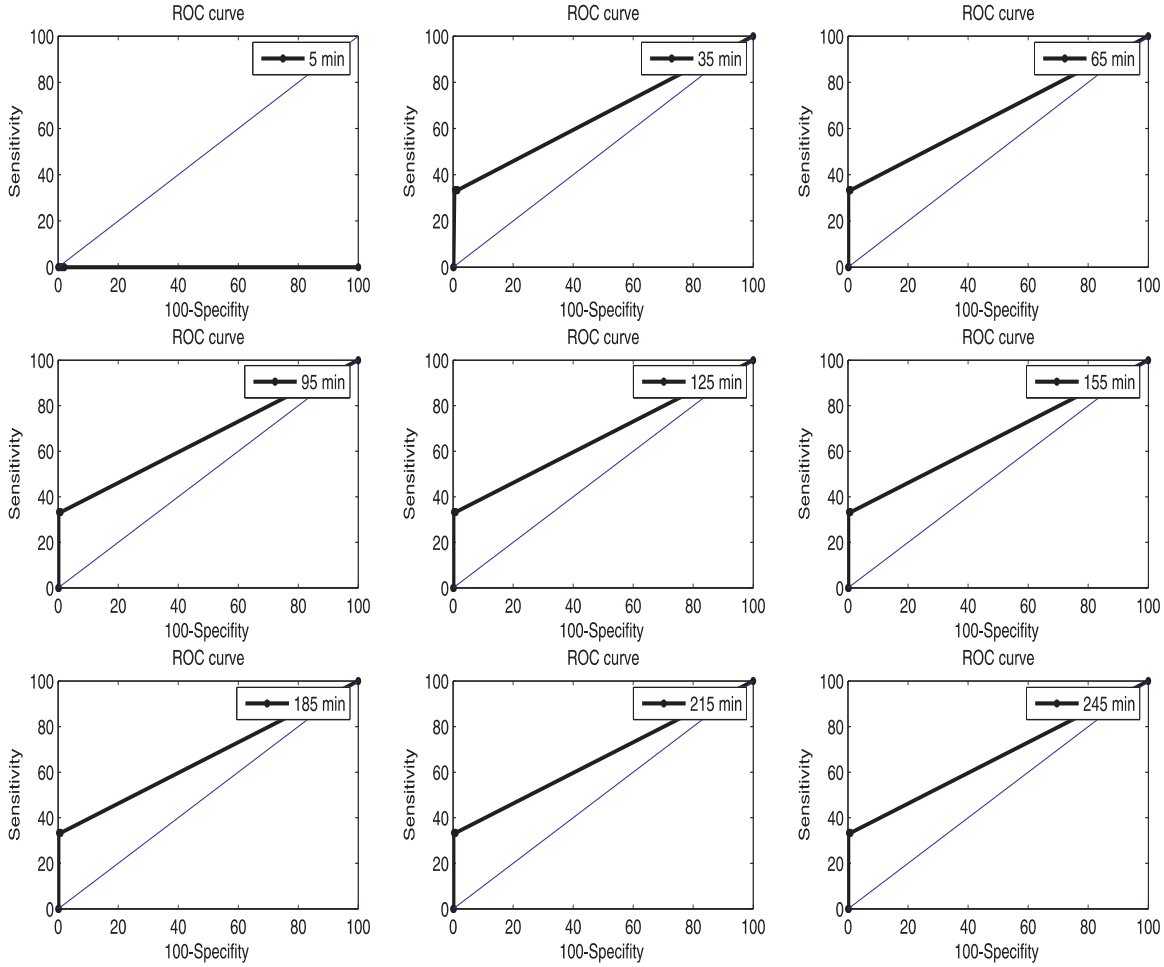


Figure 8.31: ROC curves for **patient C scalp**, for supervised Neuroscale GMM density based seizure predictor using **feature subset 1** and $\alpha = 0.8$. ROC curves are shown for prediction windows varying from 5 minutes to 245 minutes in steps of 30 minutes. Each point on a ROC curve indicates the sensitivity and specificity for every threshold level in that prediction window. The ROC curves show the maximum sensitivity of 35% and 100% specificity for a prediction window of length **35 minutes**.

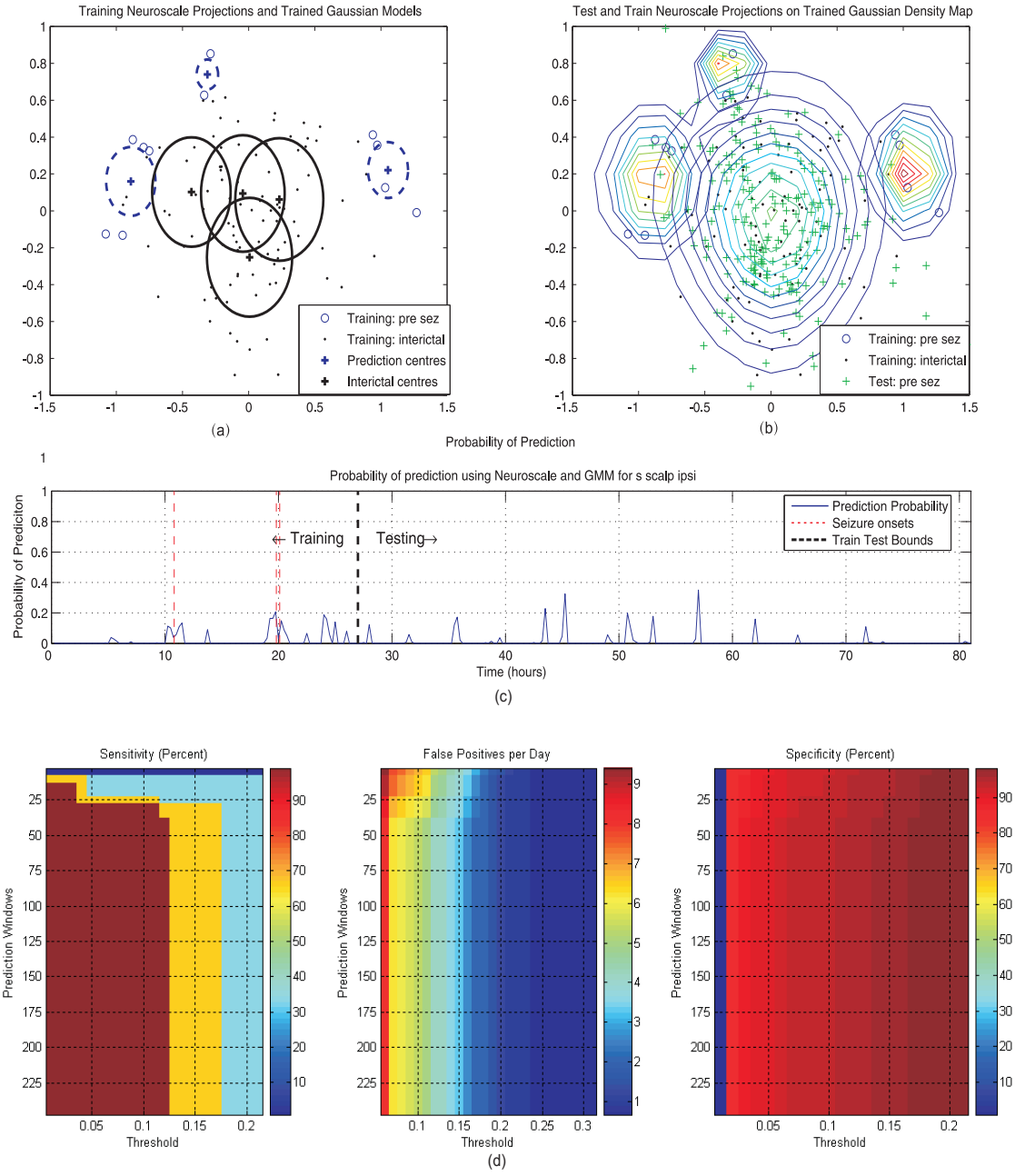


Figure 8.32: The figures show the seizure prediction analysis for patient **L scalp**. (a) The two dimensional feature map obtained for the training data using supervised Neuroscale, **feature subset 6** and $\alpha = 0.8$. Interictal class (*dots*), pre-ictal class (*circles*). Gaussian mixtures trained on both the classes are shown. One standard deviation for interictal GMM (*solid lines*), for pre-ictal GMM (*dashed lines*), centers (*crosses*). (b) Training and test data on the same Neuroscale map. Pre-ictal (*circles*), interictal (*dots*), test pre-ictal (*squares*), test interictal (*crosses*). Gaussian density contours of both classes are shown. (c) Probability plot of a data point being in the pre-ictal density map vs. time. Seizure onsets (*thin vertical dash lines*), Train-Test boundary (*thick dash line*). (d) Sensitivity, (e) False positives per day and (f) Specificity for prediction windows varying from 0-4 hours in steps of 5 minutes and threshold levels varying from 0-1 in steps of 0.01.

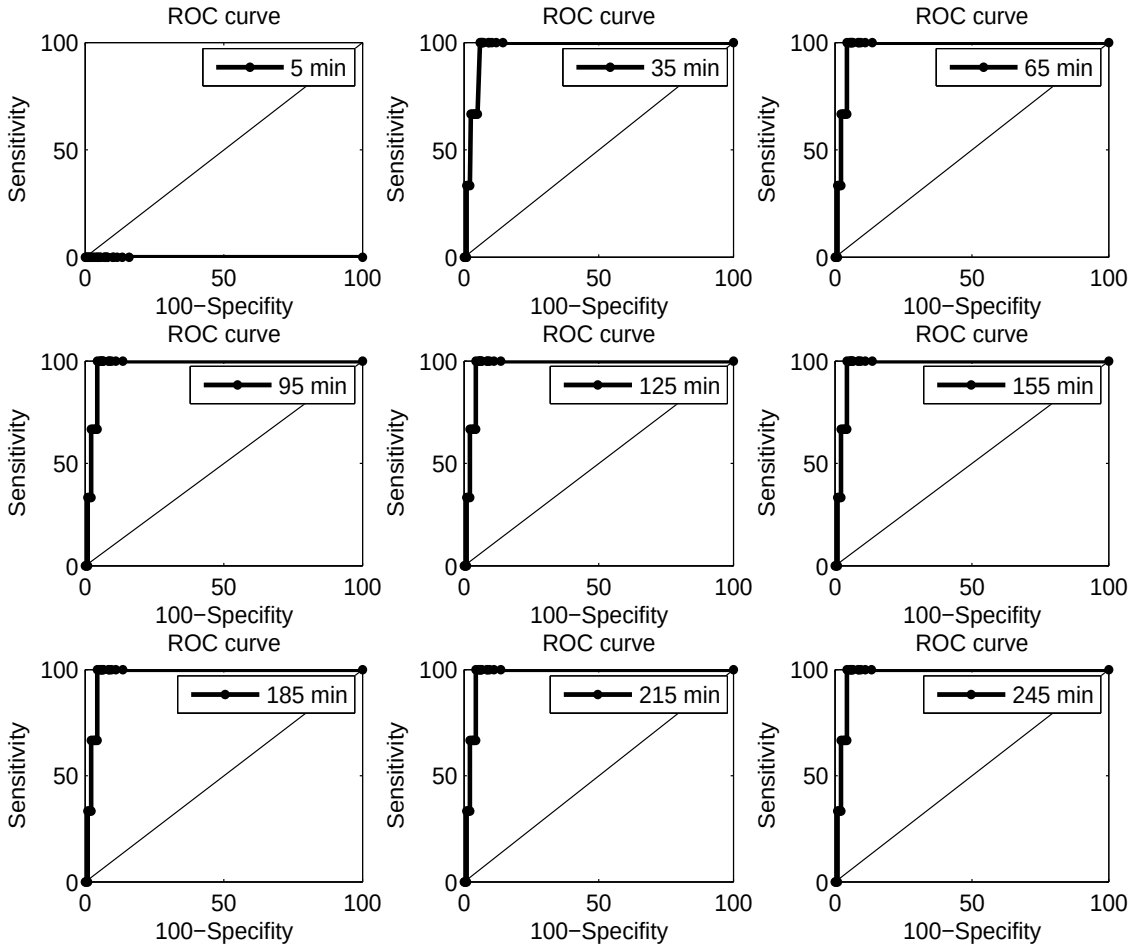


Figure 8.33: ROC curves for patient L scalp, for supervised Neuroscale GMM density based seizure predictor using feature subset 6 and $\alpha = 0.8$. ROC curves are shown for prediction windows varying from 5 minutes to 245 minutes in steps of 30 minutes. Each point on a ROC curve indicates the sensitivity and specificity for every threshold level in that prediction window. The ROC curves show a sensitivity of 100% and specificity of 86% for the prediction window of length **35 minutes**.

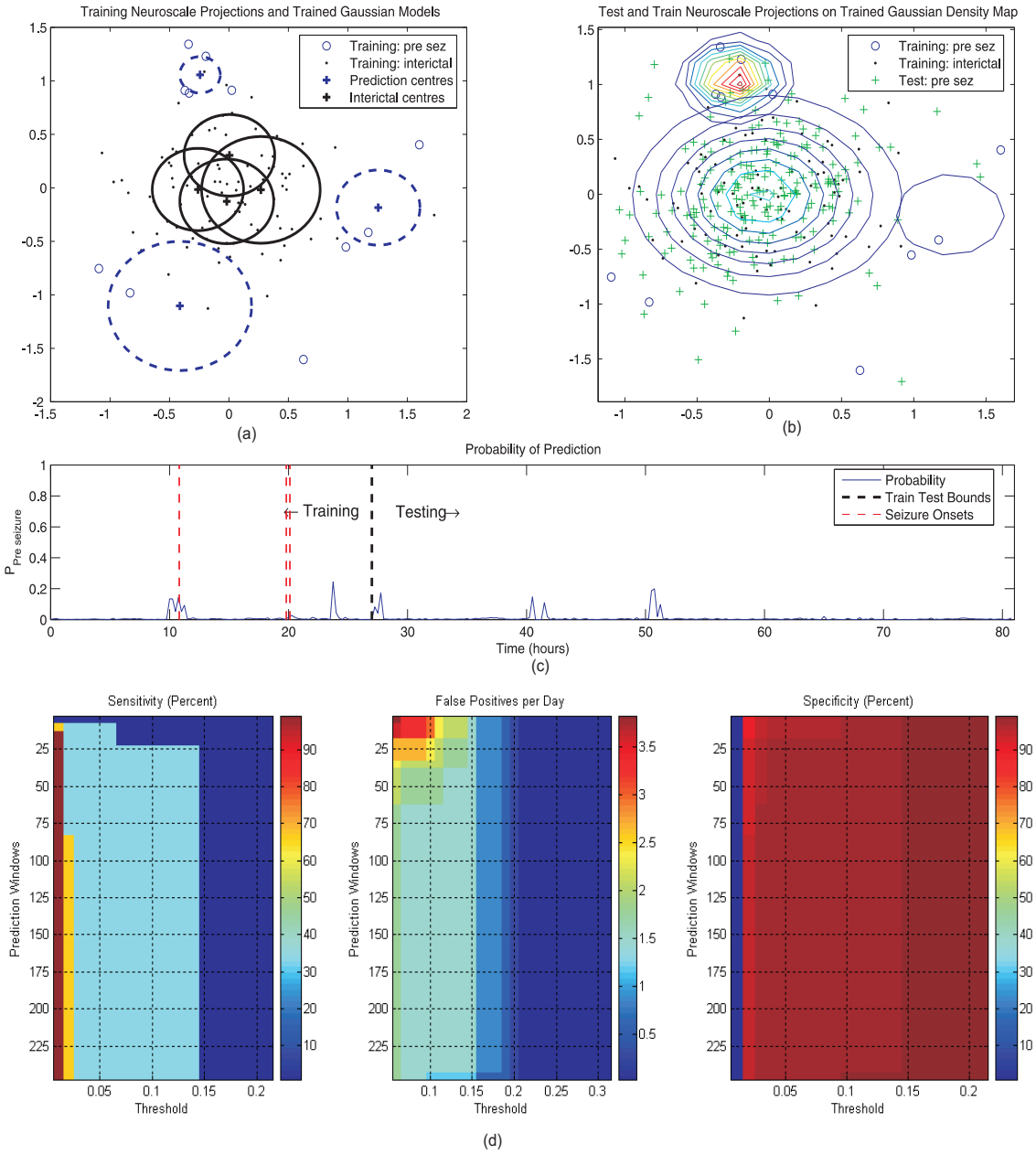


Figure 8.34: The figures show the seizure prediction analysis for patient **L scalp**. (a) The two dimensional feature map obtained for the training data using supervised Neuroscale, **feature subset 1** and $\alpha = 0.8$. Interictal class (*dots*), pre-ictal class (*circles*). Gaussian mixtures trained on both the classes are shown. One standard deviation for interictal GMM (*solid lines*), for pre-ictal GMM (*dashed lines*), centers (*crosses*). (b) Training and test data on the same Neuroscale map. Pre-ictal (*circles*), interictal (*dots*), test pre-ictal (*squares*), test interictal (*crosses*). Gaussian density contours of both classes are shown. (c) Probability plot of a data point being in the pre-ictal density map vs. time. Seizure onsets (*thin vertical dash lines*), Train-Test boundary (*thick dash line*). (d) Sensitivity, (e) False positives per day and (f) Specificity for prediction windows varying from 0-4 hours in steps of 5 minutes and threshold levels varying from 0-1 in steps of 0.01.

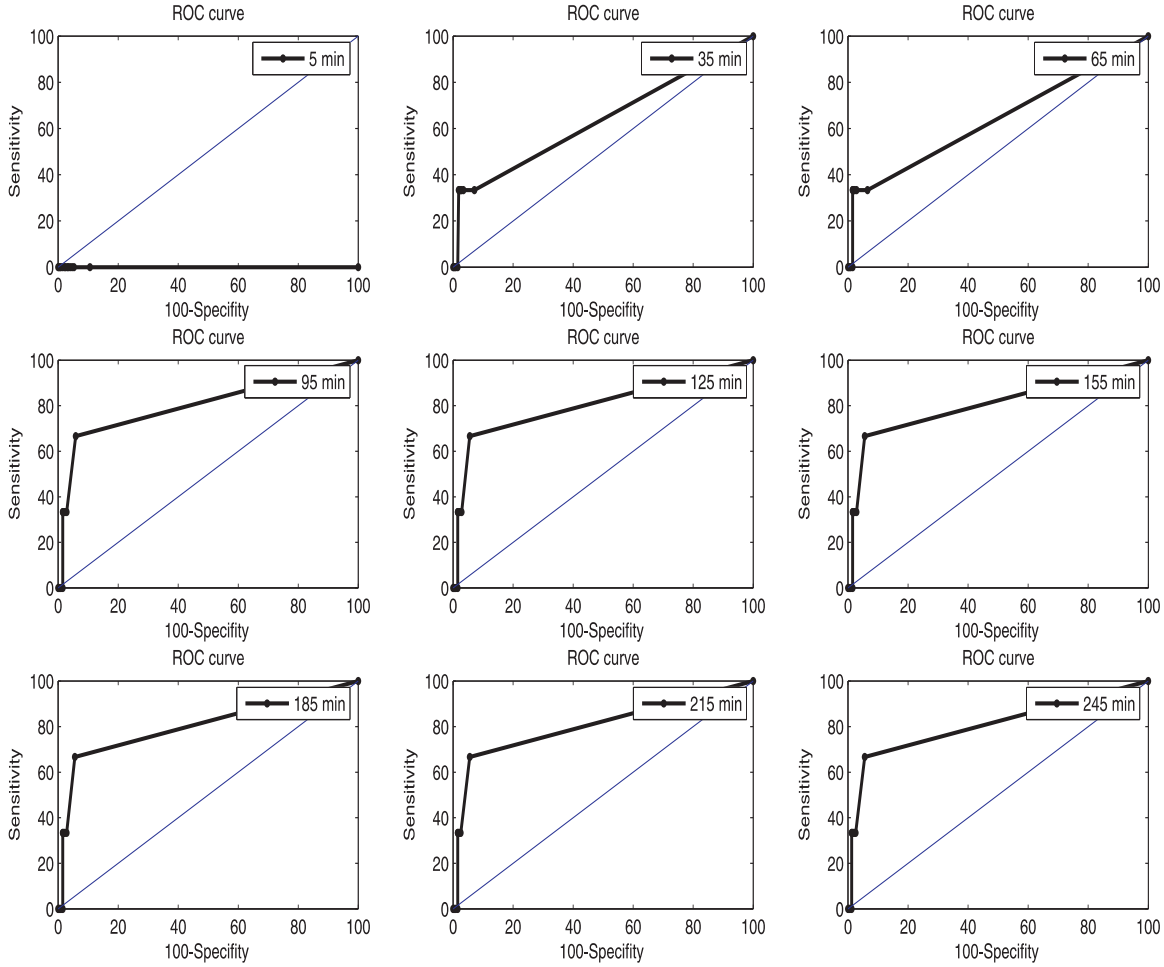


Figure 8.35: ROC curves for patient E scalp, for supervised Neuroscale GMM density based seizure predictor using feature subset 6 and $\alpha = 0.8$. ROC curves are shown for prediction windows varying from 5 minutes to 245 minutes in steps of 30 minutes. Each point on a ROC curve indicates the sensitivity and specificity for every threshold level in that prediction window. The ROC curves show a 67% sensitivity and 90% specificity for a prediction window of length **95 minutes**.

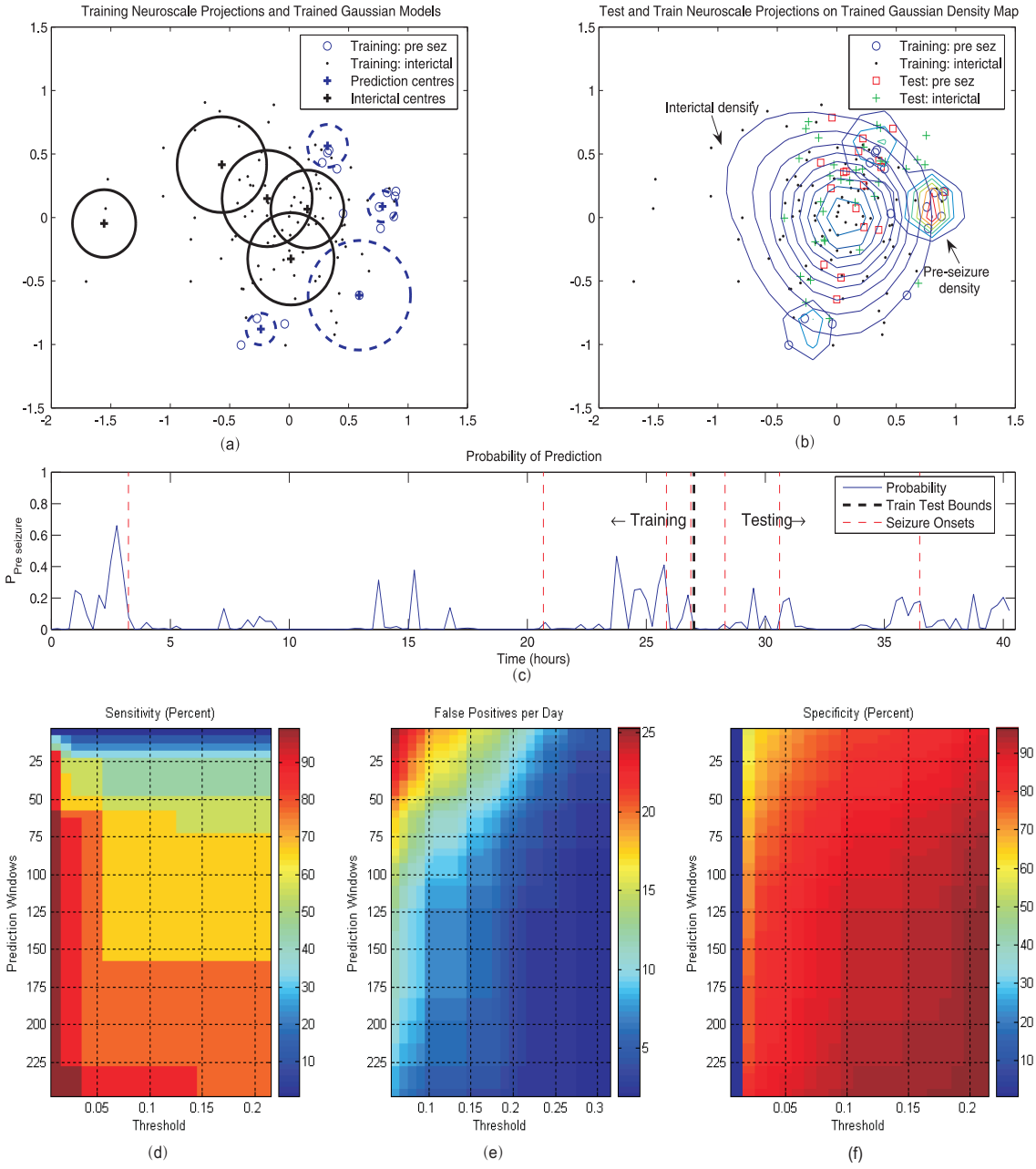


Figure 8.36: The figures show the seizure prediction analysis for patient **X scalp**. (a) The two dimensional feature map obtained for the training data using supervised Neuroscale, **feature subset 6** and $\alpha = 0.8$. Interictal class (*dots*), pre-ictal class (*circles*). Gaussian mixtures trained on both the classes are shown. One standard deviation for interictal GMM (*solid lines*), for pre-ictal GMM (*dashed lines*), centers (*crosses*). (b) Training and test data on the same Neuroscale map. Pre-ictal (*circles*), interictal (*dots*), test pre-ictal (*squares*), test interictal (*crosses*). Gaussian density contours of both classes are shown. (c) Probability plot of a data point being in the pre-ictal density map vs. time. Seizure onsets (*thin vertical dash lines*), Train-Test boundary (*thick dash line*). (d) Sensitivity, (e) False positives per day and (f) Specificity for prediction windows varying from 0-4 hours in steps of 5 minutes and threshold levels varying from 0-1 in steps of 0.01.

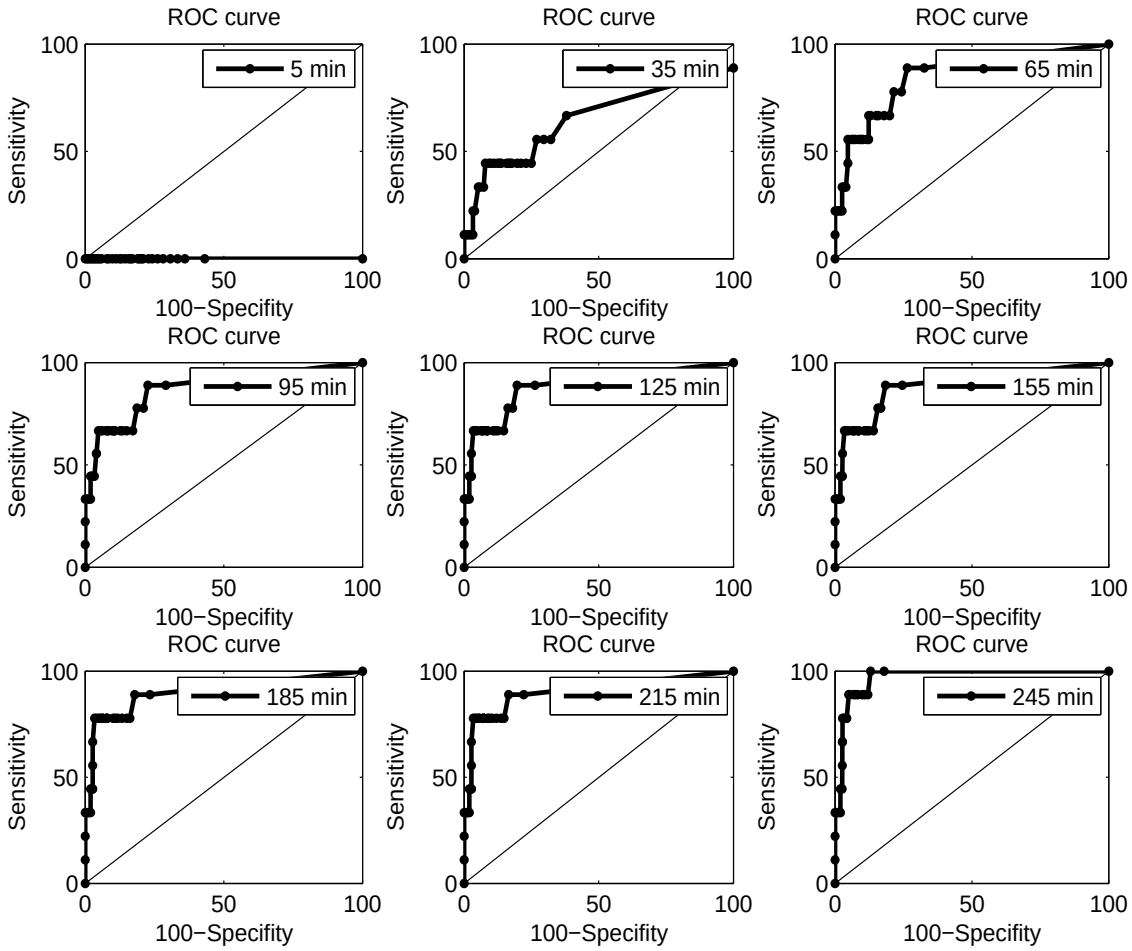


Figure 8.37: ROC curves for patient X scalp, for supervised Neuroscale GMM density based seizure predictor using feature subset 6 and $\alpha = 0.8$. ROC curves are shown for prediction windows varying from 5 minutes to 245 minutes in steps of 30 minutes. Each point on a ROC curve indicates the sensitivity and specificity for every threshold level in that prediction window. The ROC curves show 80% sensitivity and 80% specificity for the prediction window of length of **65 minutes**. With 35 minute prediction window the sensitivity reduces to 65%

Patient T scalp: The data for patient T scalp covers two days of continuous multichannel scalp EEG. It has 8 seizures and the seizures are distributed across the data over the two days. The feature space of T scalp is observed to form distinct clusters in the neuroscale feature training map 8.21 (a). The training feature space is modelled with the GMM quite aptly. The Gaussian density contours shown in Figure 8.21 (b) appear to form distinct density maps with minimal overlap. The test feature points from interictal data can be observed to be projected across the interictal GMM, and most of the pre-ictal test feature points can be seen to be in the pre-ictal GMM. The probability plot in Figure 8.21 (c) shows peaks before a seizure and very few false peaks. The training data four seizures and the test seizures can be observed to have similar peaks, which shows that the trained GMM model was able to classify the test pre-ictal feature points. Figure 8.21 (d) shows the sensitivity, false positives per day and the specificity graphs for all prediction windows and all threshold levels. Figure 8.22 shows the ROC curves for various prediction windows. A 100% sensitivity and 80% specificity can be observed to be for the shortest prediction window being 35 minutes.

Seizure prediction with feature subset 1 for patient T scalp has been shown in Figures 8.23 and 8.24. It produced results slightly inferior to the feature subset 6. The maximum sensitivity of 90% and specificity of 90% was obtained for the short prediction window of length 95 minutes. For the prediction window of 35 minutes as in the case of feature subset 6, the sensitivity is much lesser 65% as compared to the 100% sensitivity with feature subset 6.

Patient E scalp: Patient E scalp had four days of continuous multichannel scalp EEG. Although the patient had nine seizures in total, but most of the seizures were grouped together in the later half of the data, within a short window of one hour. The Neuroscale training data included the first three seizures and the other cluster of seizures were kept in the test data. The Neuroscale map did show some subtle cluster formation 8.25 (a), and the GMM model appeared to encapsulate the distribution of the pre-ictal and interictal data fairly well. The GMM density distributions for the two classes is shown in Figure 8.25 (b). However, the test feature points obtained from the trained Neuroscale network showed most of the pre-ictal feature points to be lying in the interictal density map (see Figure 8.25 (b)). This led to a substandard performance of the predictor as can be observed in the sensitivity, specificity and ROC curves of the patient in Figures 8.25 (d) and 8.26. This can be explained by (1) either the Neuroscale network was not trained adequately, even though many attempts were made with varying the number of hidden nodes and the input data (by varying the train-test boundary), or (2) the pre-ictal features space in the test data (with the cluster of seizures) had some additional underlying features, based on perhaps the distribution of the seizures, that were not aptly captured by the feature subset being used. The patients test data ROC curves shown in Figure 8.26 show a maximum sensitivity for the shortest prediction window to be 40% and specificity being 75% with the prediction

window of 140 minutes.

Patient C scalp: Patient C scalp has data of four days with three seizures. As the two seizures in later part of the data are quite close together, only one seizure event can be included in the training data, making the pre-ictal feature space quite sparse. Nevertheless, the Neuroscale is trained and the Gaussian model created as there are enough inter-ictal feature points. Singularity in GMM model of pre-ictal class is difficult to avoid in this case 8.28 (a). The density maps of the trained GMM are shown in Figure 8.28 (b), with the sparse pre-ictal test points projected on the density maps. The probability of predictive event for the training and the test data is shown in the Figure 8.28 (c). The sensitivity, false positives per day and specificity of this data set is shown in Figure 8.28 (d). The maximum sensitivity of 67% and specificity of 100% for the shortest prediction window of 35 minutes can be observed from the ROC curves in Figure 8.29.

The results for patient C scalp with feature subset 1, shown in Figures 8.30 and 8.31, also show inferior results compared to those of feature subset 6. The maximum sensitivity and specificity for the shortest prediction window occurs at 35 minutes, which is similar to the one with feature subset 6, however, the sensitivity has dropped from 65% to 35% and specificity to 33% for this prediction window and it does not increase for longer prediction window lengths.

Patient L scalp: Patient L scalp has three days continuous multichannel EEG data with three seizures occurring the first day. The closeness of the last two seizures allowed only one seizure to be included in the training data initially, but as all the seizures are in the first half of the data, a train test boundary before the second seizure would leave very less interictal data for adequate training of the GMM. Therefore all the seizures had to be included in the training data and the performance of the test data for prediction would be based on the true negatives (low false positives). The distribution of the training feature points int he neuroscale map shows clusters that would allow clear decision boundaries to be learned by the GMM model (see Figure 8.32 (a)). The density maps with the test points projected on to the GMM density maps have been shown in Figure 8.32 (b). The test interictal feature points can be clearly observed to be concentrated within the interictal density map (see *crosses* marked on the density maps in Figure 8.32 (b)). The probability plot reflects the same in Figure 8.32 (c). The sensitivity, false positives per day and specificity graphs are calculated for prediction windows 0-4 hours in steps of 5 minutes and threshold levels 0-1 in steps of 0.01 8.32 (d). The sensitivity and specificity relationship is shown with the help of ROC curves in Figure 8.33, where a maximum sensitivity of 100% with 86% specificity is observed for the shortest prediction window of 35 minutes.

Feature subset 1 based prediction for patient L scalp show degraded sensitivity compared to those of feature subset 6 (see Figures 8.34 and 8.35). The maximum sensitivity

of 69% is observed with a specificity of 90% for the shortest prediction window of 95 minutes. Compared to the sensitivity of 100% for a prediction window of 35 minutes with feature subset 6, the sensitivity is much reduced to 35% for the same prediction window length, with feature subset 1.

Patient X scalp: Patient X scalp has continuous multichannel EEG data of two days with eight seizures distributed over the data duration. The training data includes four seizures and the neuroscale feature space shows the pre-ictal feature points to be away from the interictal feature points 8.36 (a). The GMM is able to model the Gaussian mixtures for both the classes quite aptly, with least overlap, in this case. The density maps of the GMMs have been shown in Figure 8.36 (b). However, many feature points of pre-ictal events from the test data set can be observed to be present in the interictal density maps as well. This might be hypothesized to be due to this patient experiencing unique sudden changes in EEG before seizure, such that only the immediate pre-ictal data window identifies with the pre-ictus. The probability plot of the same is shown in Figure 8.36 (c). Bursts of activity is observed close to a seizure event in training as well as test phase, with mostly silent periods at other times. The sensitivity and specificity is computed for prediction windows from 0-4 hours and threshold levels 0-1, shown in Figure 8.36 (d). ROC curves in Figure 8.37 show 80% sensitivity with 80% specificity achievable with a prediction window of 65 minutes. A compromise on the sensitivity (65%) can allow a shorter prediction window of 35 minutes as well.

The above analysis shows that ***feature subset 6*** performs the best with training and test using supervised Neuroscale and supervised GMM for seizure prediction. Therefore for all further analysis, feature subset 6 is selected.

8.5.2 Results II: Neuroscale and Semi-Supervised GMM with Labelled and Unlabelled Data

The Neuroscale followed by a semi-supervised GMM with labelled and unlabelled data is applied to the five patients data using the feature subset 6 and $\alpha = 0.8$. The procedure used is outlined here:

Neuroscale Topographic Mapping: The multidimensional feature vectors are first reduced to a two dimensional feature space using supervised Neuroscale. The supervision with $\alpha = 0.8$ is performed with the use of a distance metric informing the network about the class membership. The data is divided into a set of training and test set, making sure that the training set includes sufficient pre-ictal data and the test data has some pre-ictal data as well.

Creating labelled and unlabelled data set: The pre-ictal data points in the training set are marked as the labelled data. The remaining interictal data in the training set along with the rest of the test data set is used as the unlabelled data.

GMM for labelled data: A GMM is first trained on the labelled pre-ictal data. The number of Gaussian components are varied as trial and error for the best fit among five trials. A density map of the same is created.

GMM for unlabelled data: A GMM is then trained on the unlabelled data only. Its number of components are also varied and the best fit GMM is used from among five trials. The density map for the unlabelled data is also drawn.

GMM for unlabelled data knowing GMM of labelled data: The GMM for the unlabelled data is trained but using the GMM for the labelled data. This ensures that the probability of a given data point to be present in the pre-ictal density map and in the interictal density map sums to one.

Probability curve: The probability curve is then obtained as the probability of a given data point to be classified in the pre-ictal density map.

Statistical analysis: The sensitivity, false positives per day, specificity and ROC curves are computed for various prediction windows zero to four hours and threshold levels zero to one.

The result figures 8.38, 8.39, 8.40, 8.41, 8.42, 8.43, 8.44, 8.45, 8.46 and 8.47 for each patient includes the following information:

1. Neuroscale topographic two dimensional map.
2. GMM trained on the labelled pre-ictal data only (from the training set), with its density map.

3. GMM trained on the unlabelled data (from the training and test set) only, with its density map.
4. GMM trained on the unlabelled data, but using the labelled GMM (along with the density map), such that the sum of probabilities remains one.
5. The density maps of labelled pre-ictal training data and the unlabelled data combined.
6. The probability of every data point to be present in the pre-ictal density map.
7. Sensitivity, false positives per day and specificity graphs for prediction windows 0-4 hours and threshold levels 0-1.
8. ROC curves for various prediction windows showing the statistics as a function of sensitivity and specificity.

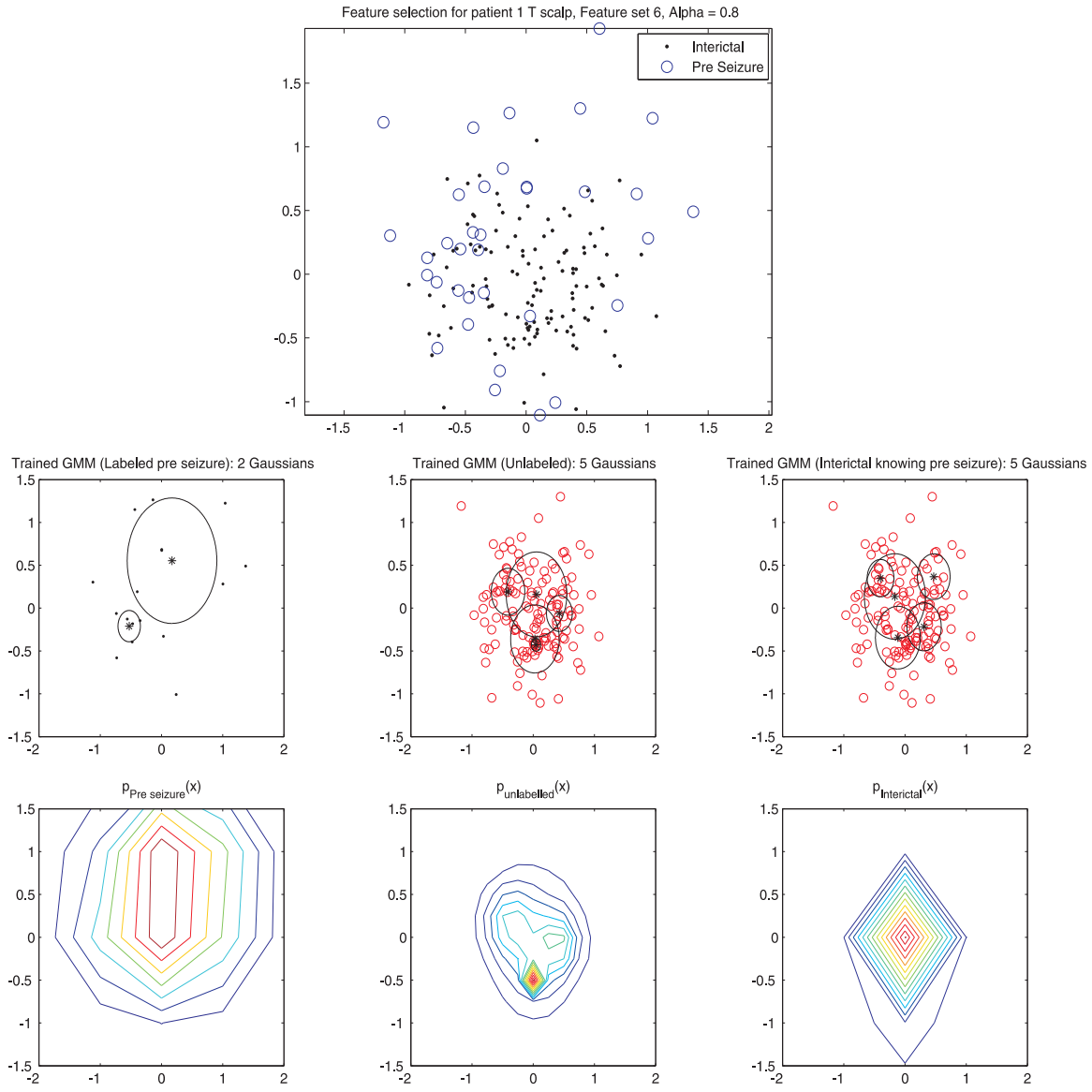


Figure 8.38: Neuroscale two dimensional projection of feature space for **patient T scalp** (*top*), GMMs trained on labelled pre-ictal training data and its density map (*left column*), GMMs trained on unlabelled data and its density map (*middle column*), GMMs trained on unlabelled data knowing the GMM parameters of the labelled data and its density map (*right column*).

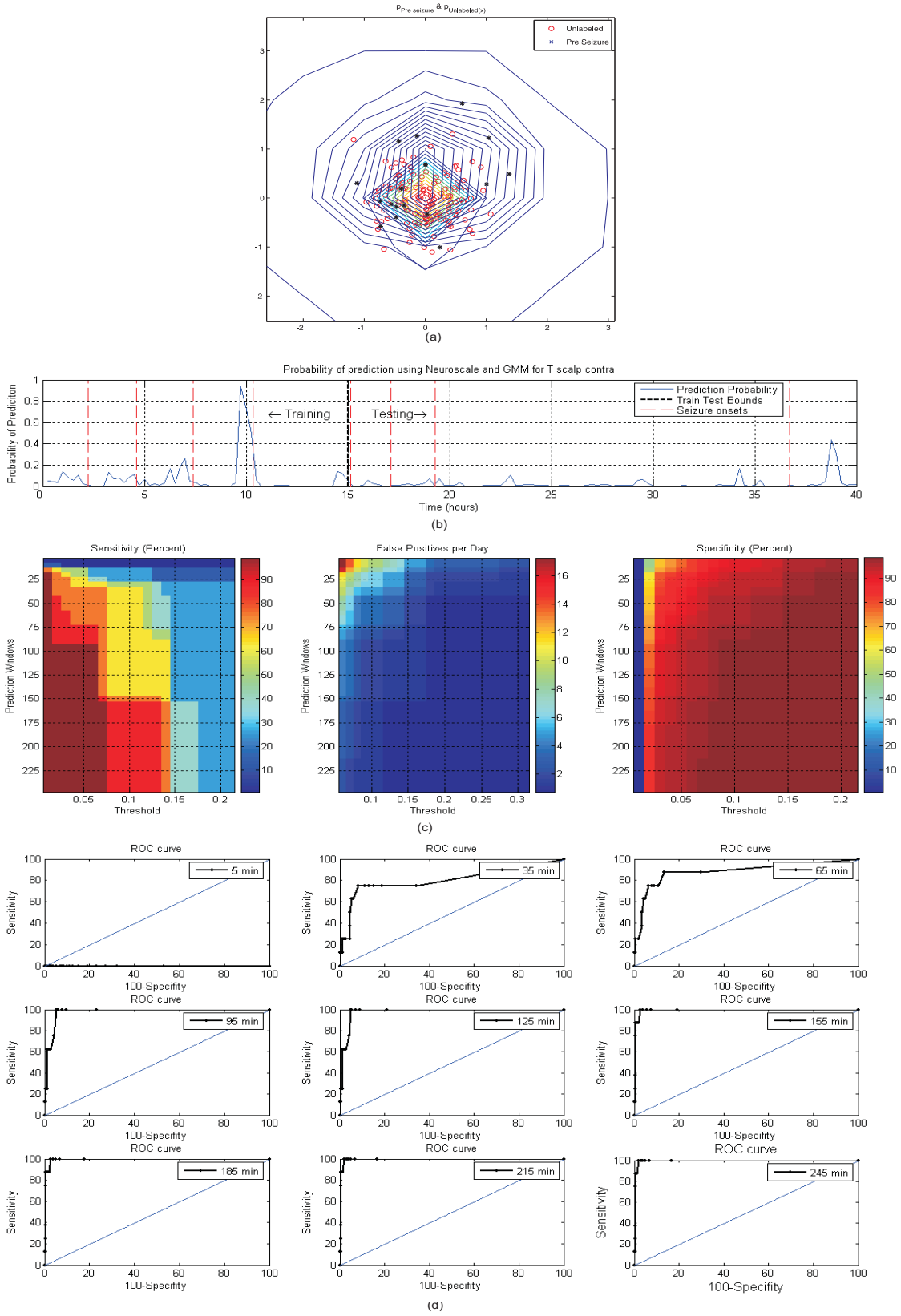


Figure 8.39: **Statistical analysis of semi-supervised GMM for patient T scalp** (a) Combined density contours for the labelled pre-ictal and unlabelled GMMs, (b) Probability curve for a pre-ictal event occurrence, (c) Sensitivity, false positives per day and specificity graphs for prediction windows 0-4 hours and thresholds 0-1, (d) ROC curves for various prediction windows. The shortest prediction window that has maximum sensitivity (78%) and maximum specificity (90%) is 35 minutes.

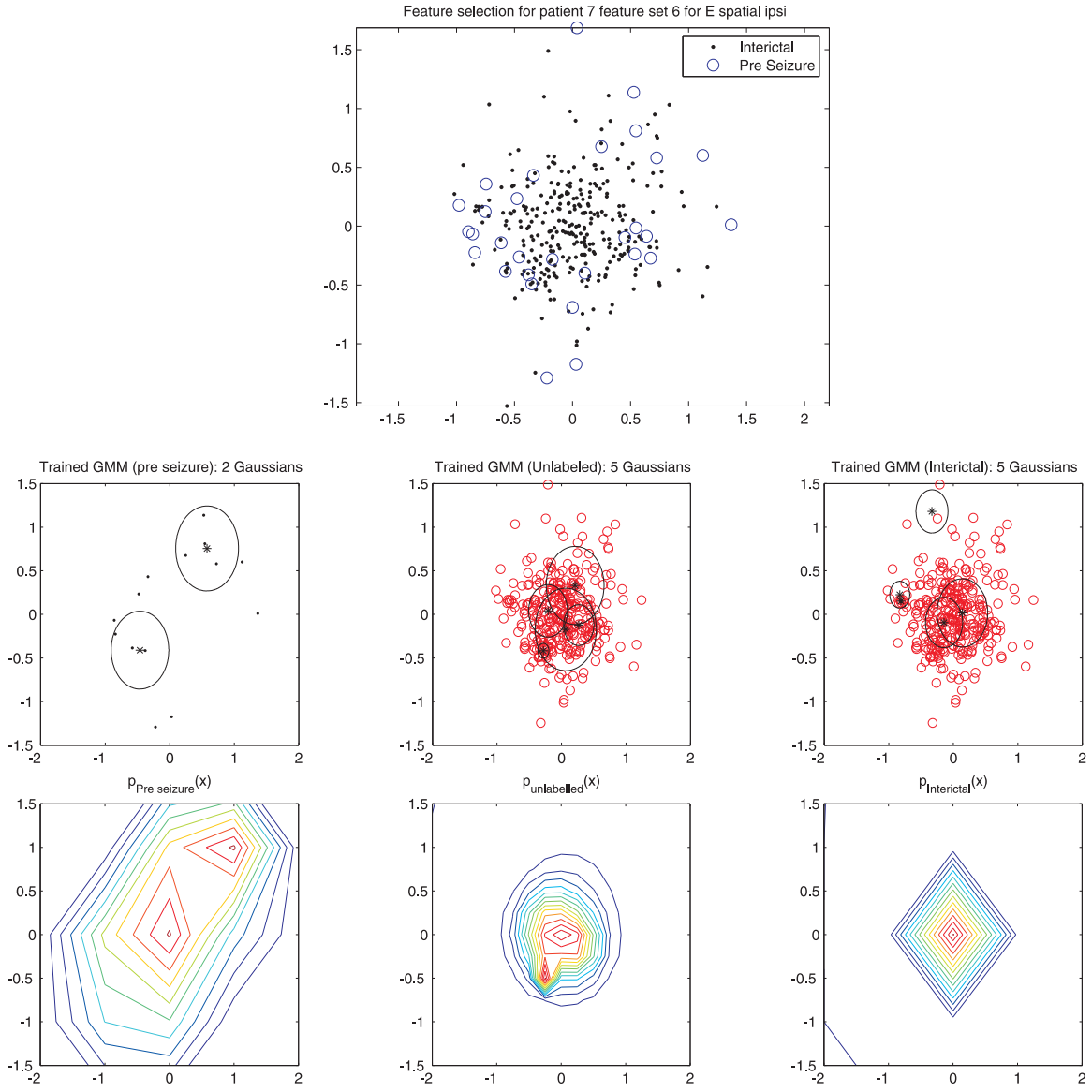


Figure 8.40: Neuroscale two dimensional projection of feature space for **patient E scalp** (*top*), GMMs trained on labelled pre-ictal training data and its density map (*left column*), GMMs trained on unlabelled data and its density map (*middle column*), GMMs trained on unlabelled data knowing the GMM parameters of the labelled data and its density map (*right column*).

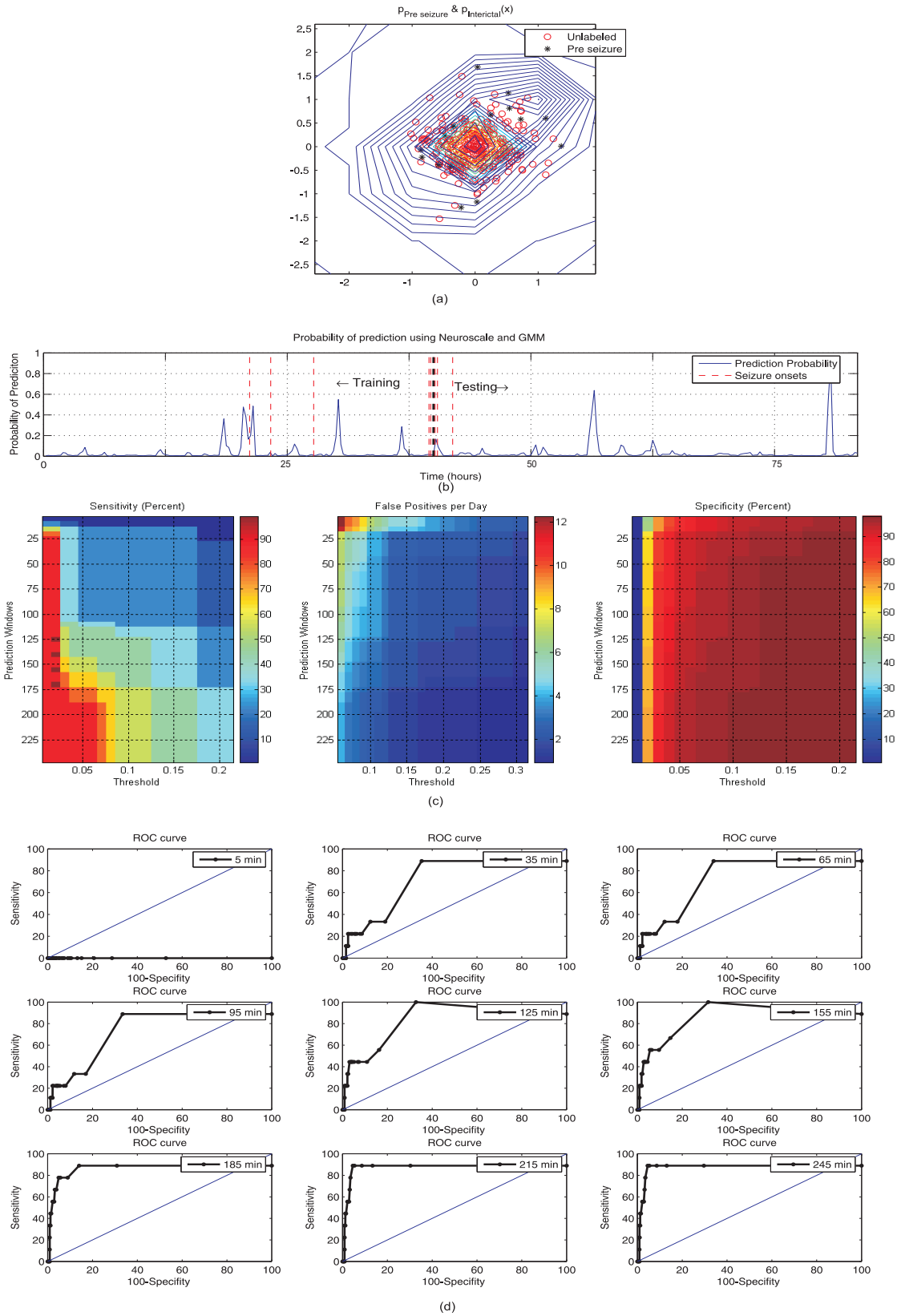


Figure 8.41: **Statistical analysis of semi-supervised GMM for patient E scalp** (a) Combined density contours for the labelled pre-ictal and unlabelled GMMs, (b) Probability curve for a pre-ictal event occurrence, (c) Sensitivity, false positives per day and specificity graphs for prediction windows 0-4 hours and thresholds 0-1, (d) ROC curves for various prediction windows. The shortest prediction window that has maximum sensitivity (90%) and maximum specificity (65%) is 35 minutes.

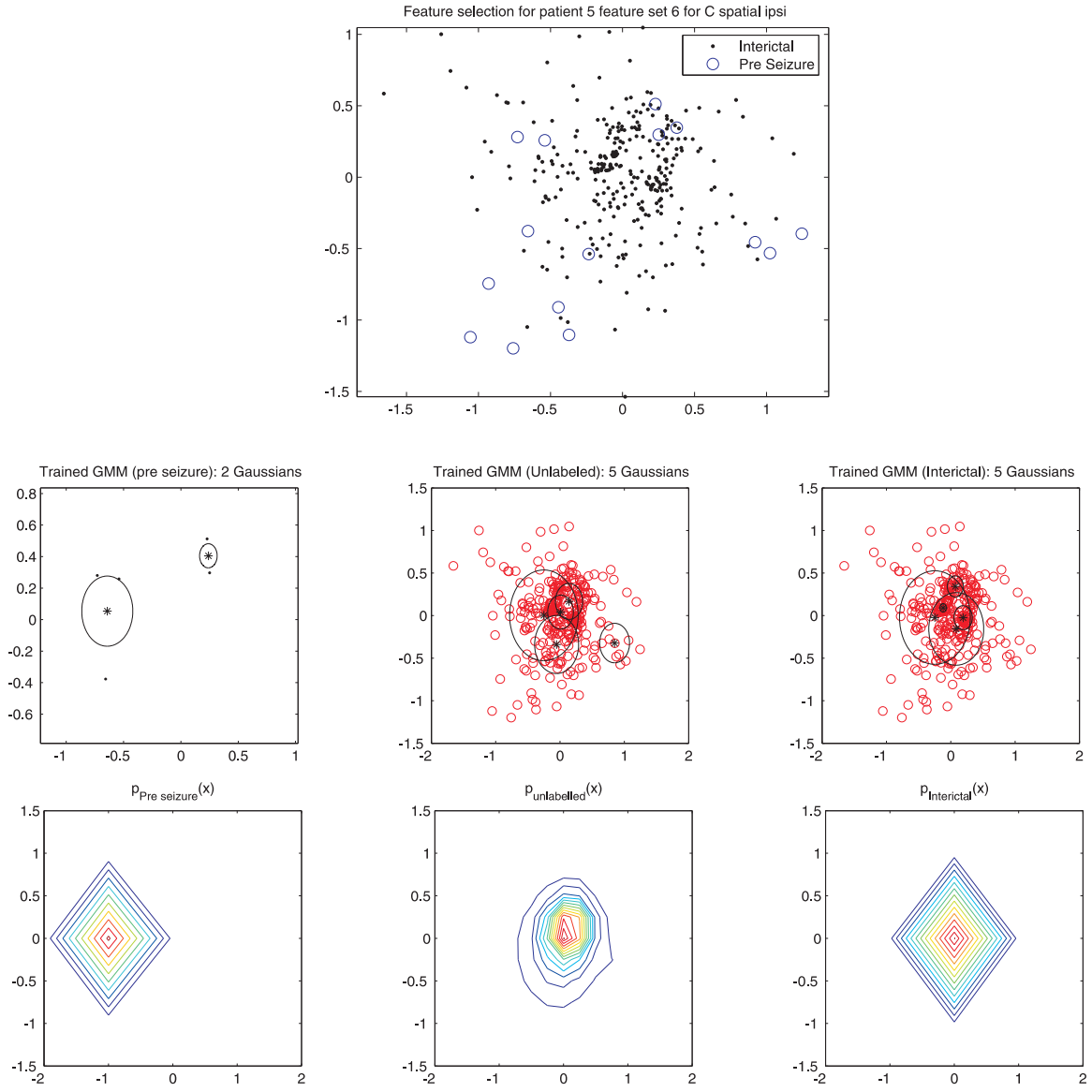


Figure 8.42: Neuroscale two dimensional projection of feature space for **patient C scalp** (*top*), GMMs trained on labelled pre-ictal training data and its density map (*left column*), GMMs trained on unlabelled data and its density map (*middle column*), GMMs trained on unlabelled data knowing the GMM parameters of the labelled data and its density map (*right column*).

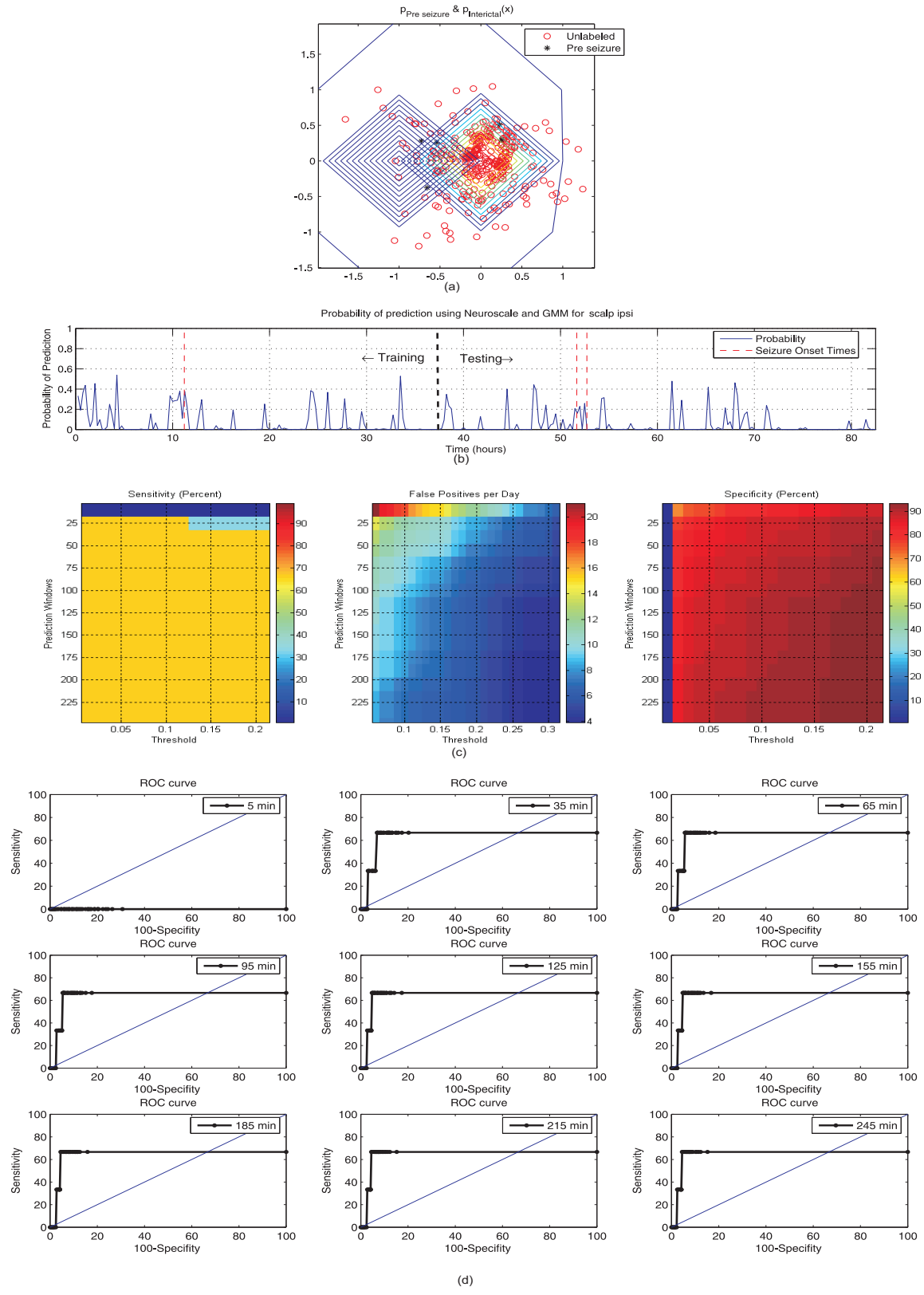


Figure 8.43: **Statistical analysis of semi-supervised GMM for patient C scalp** (a) Combined density contours for the labelled pre-ictal and unlabelled GMMs, (b) Probability curve for a pre-ictal event occurrence, (c) Sensitivity, false positives per day and specificity graphs for prediction windows 0-4 hours and thresholds 0-1, (d) ROC curves for various prediction windows. The shortest prediction window that has maximum sensitivity (65%) and maximum specificity (90%) is 35 minutes.

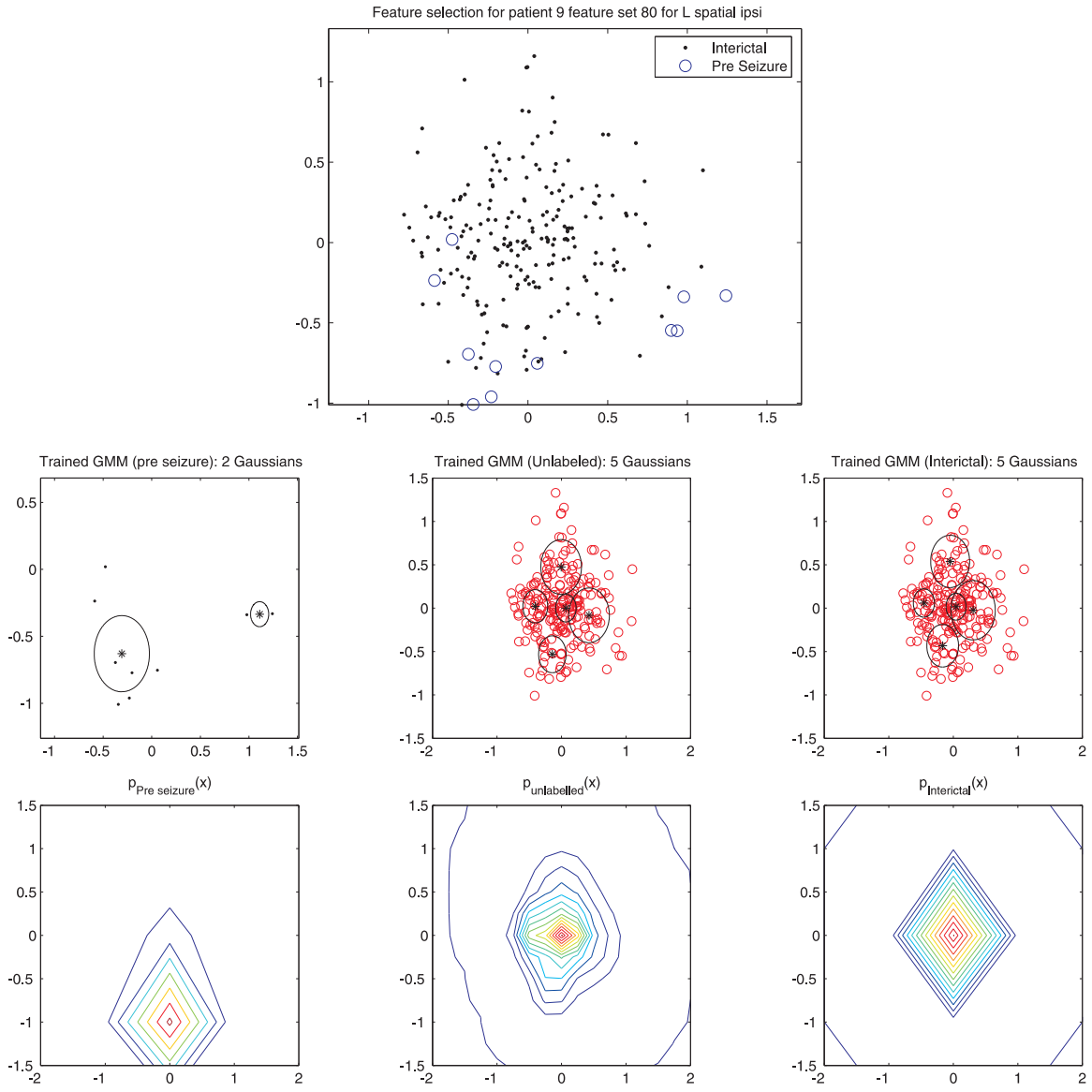


Figure 8.44: Neuroscale two dimensional projection of feature space for **patient L scalp** (*top*), GMMs trained on labelled pre-ictal training data and its density map (*left column*), GMMs trained on unlabelled data and its density map (*middle column*), GMMs trained on unlabelled data knowing the GMM parameters of the labelled data and its density map (*right column*).

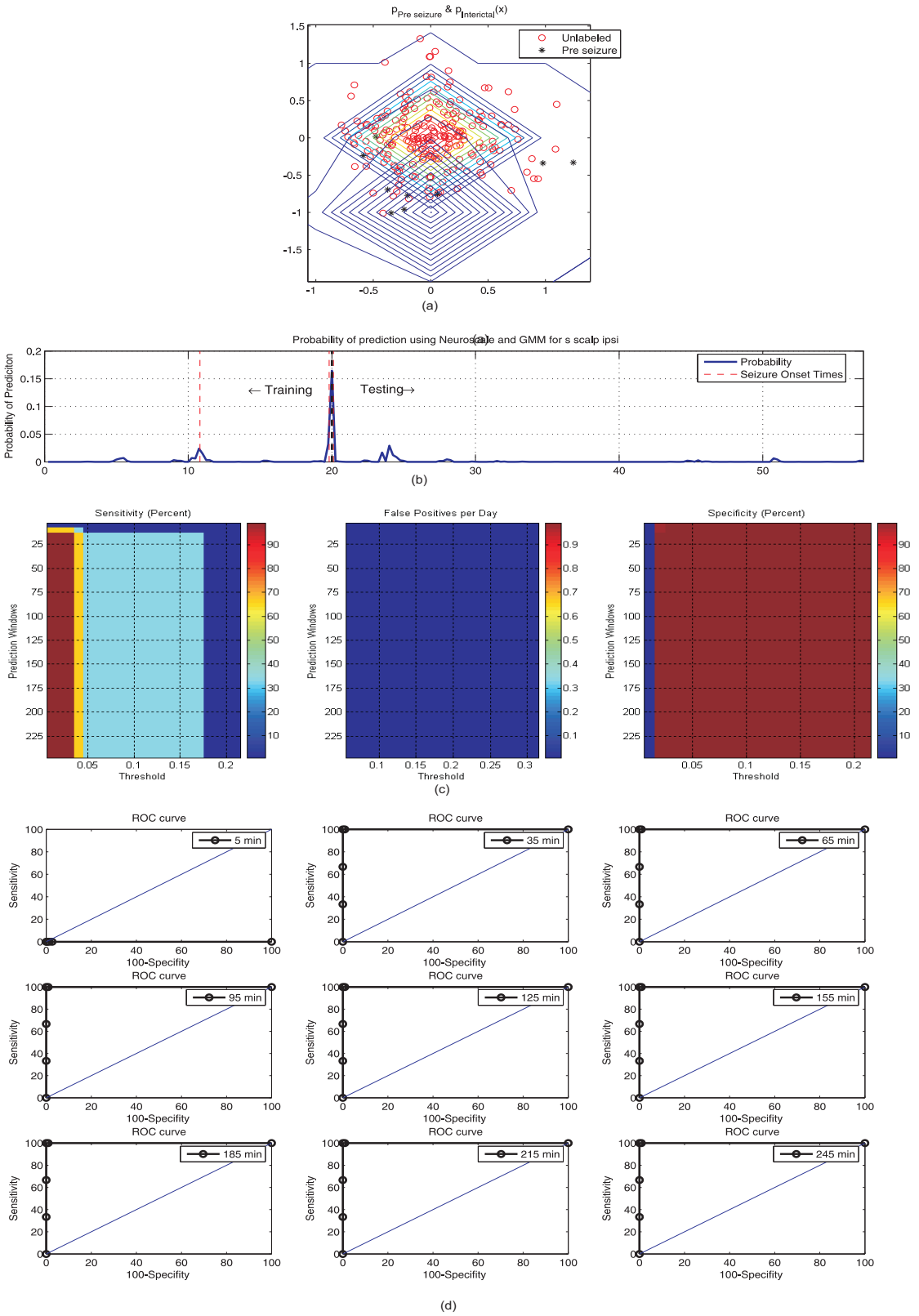


Figure 8.45: **Statistical analysis of semi-supervised GMM for patient L scalp** (a) Combined density contours for the labelled pre-ictal and unlabelled GMMs, (b) Probability curve for a pre-ictal event occurrence, (c) Sensitivity, false positives per day and specificity graphs for prediction windows 0-4 hours and thresholds 0-1, (d) ROC curves for various prediction windows. The shortest prediction window that has maximum sensitivity (100%) and maximum specificity (100%) is 35 minutes.

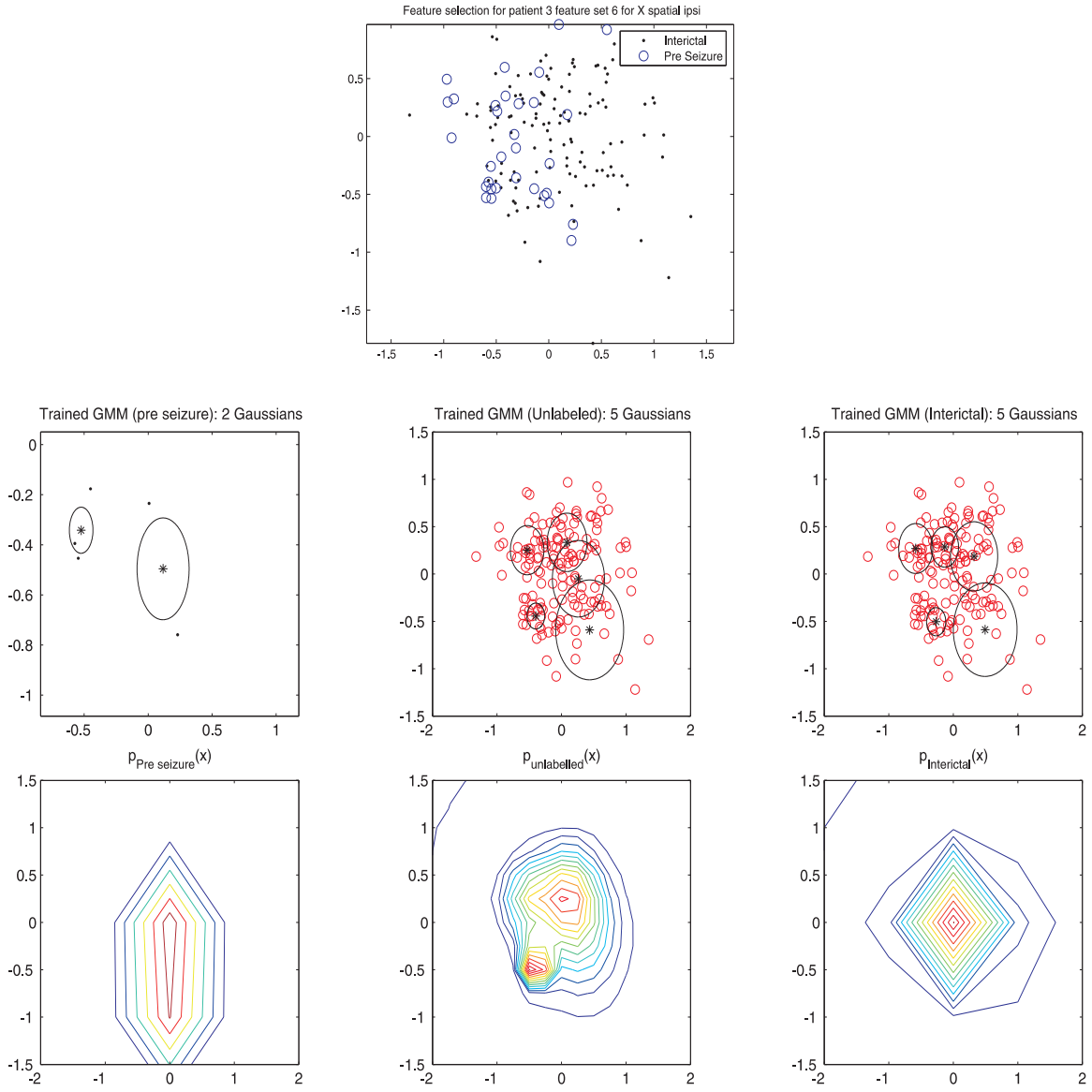


Figure 8.46: Neuroscale two dimensional projection of feature space for **patient X scalp** (*top*), GMMs trained on labelled pre-ictal training data and its density map (*left column*), GMMs trained on unlabelled data and its density map (*middle column*), GMMs trained on unlabelled data knowing the GMM parameters of the labelled data and its density map (*right column*).

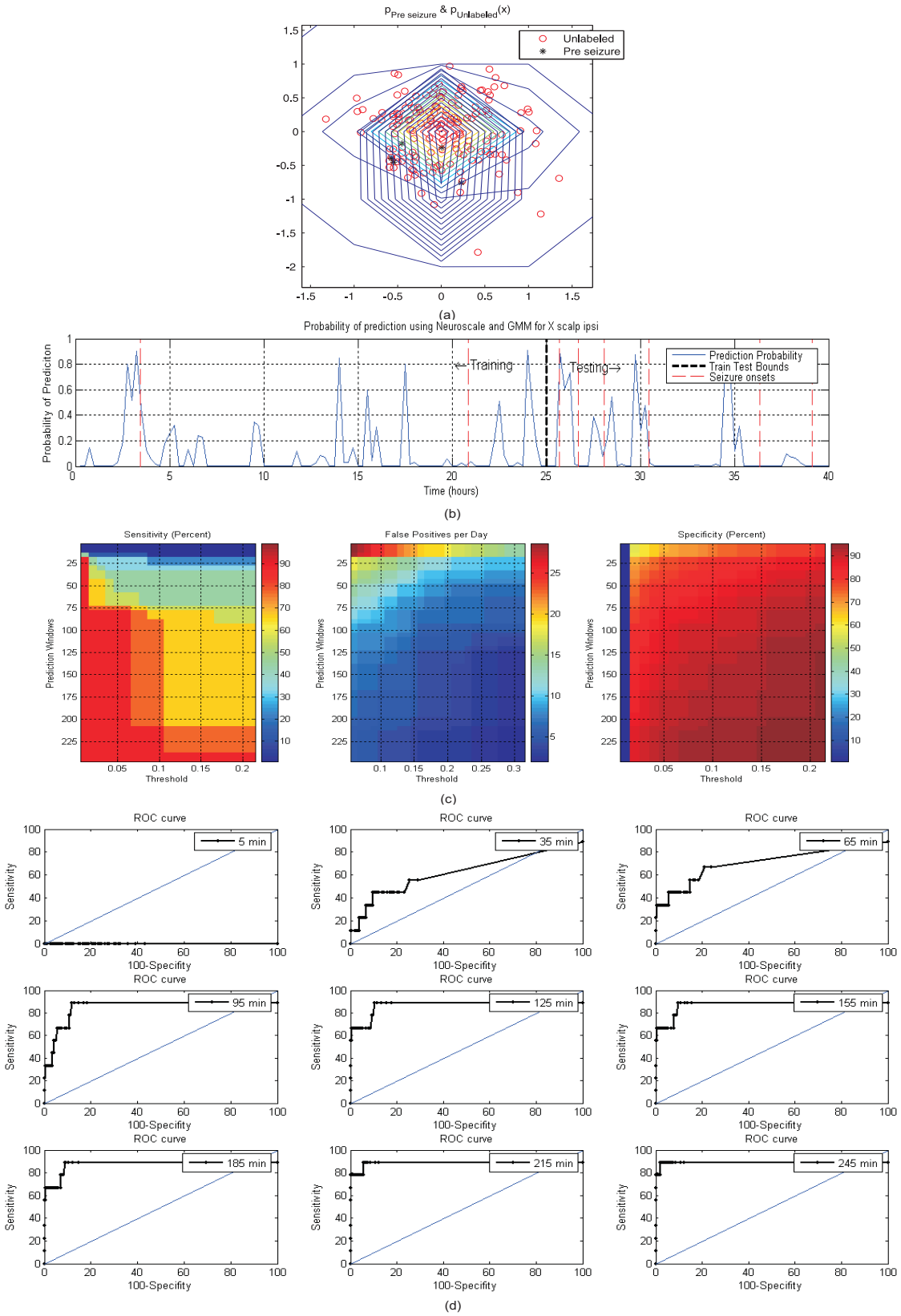


Figure 8.47: **Statistical analysis of semi-supervised GMM for patient X scalp** (a) Combined density contours for the labelled pre-ictal and unlabelled GMMs, (b) Probability curve for a pre-ictal event occurrence, (c) Sensitivity, false positives per day and specificity graphs for prediction windows 0-4 hours and thresholds 0-1, (d) ROC curves for various prediction windows. The shortest prediction window that has maximum sensitivity (65%) and maximum specificity (75%) is 65 minutes.

8.5.3 Supervised GMM vs. Semi-Supervised GMM

The application of supervised GMM as a classifier showed promising results with feature subset six. It was used as a train and test technique, that simultaneously verified the quality of a classification that can be used in real time once the Neuroscale and GMM models have been trained. The retrospective application of the semi-supervised GMM classification was also carried out on the same data with the same feature subset. The difference in the techniques being that the semi-supervised GMM used only the preictal labelled data from the training set and the rest of the data as the unlabelled data, and estimated the decision boundaries for the two classes. The semi-supervised GMM can help to verify the existence of a predictive feature space even with a semi supervised learning. The results of supervised GMM and semi-supervised GMM have been compared as follows:

Patient T scalp showed a poorer learning with the semi-supervised GMM. The predictor was less sensitive and less specific. The sensitivity of 78% with 90% specificity was now achievable for a prediction window of 35 minutes. Whilst, with supervised GMM the sensitivity of 100% and 80% specificity was being achieved for a prediction window of 35 minutes. **Patient E scalp** showed superior prediction results with semi-supervised GMM as compared to supervised GMM classifier. Despite the clustered seizures, a higher sensitivity of 90% with higher specificity of 65% was observed for a much shorted prediction window of 35 minutes. The supervised GMM had shown a 40% sensitivity with a prediction window of 155 minutes. This showed that the feature subset did adequately capture predictive features even for the clustered seizures, contrary to the thought of inadequacy of feature subset in encompassing dynamics prior to a cluster of seizures (as per the results from supervised GMM for patient E scalp). **Patient C scalp** showed increased false positives, but the same sensitivity for the prediction window of 35 minutes as was the case of supervised GMM. For **patient L scalp** the semi-supervised GMM gave excellent improvement, with negligible false positives and 100% prediction for the same prediction window of 35 minutes as was the case with supervised GMM. In the case of **patient X scalp**, a sensitivity of 65% and specificity of 75% can be seen for prediction window of 65 minutes, which is marginally different from the supervised GMM results. The sensitivity was seen to be higher for longer prediction windows and the amount of probability of prediction was stronger overall.

In general, marginal difference in the prediction performance was observed with the two techniques. Only one of the data sets E scalp showed significant improvement. Nevertheless, it showed that underlying information captured by the feature vectors was capable of predicting a seizure with high sensitivity and specificity.

8.5.4 Testing with Random Predictor

The seizure prediction performance is best evaluated with a null hypothesis test with a random predictor. A test using seizure time surrogates is used to assess if the seizures being predicted are better than a random predictor. Seizure time surrogates can be constructed by replacing the original seizure onset times with random onset times. Specific properties of the original onset times can be used as constraints for the surrogate times [112]. However, in the present case this is not taken into account as no specific pattern of seizure onsets can be observed consistently across patients. Instead a random predictor is used with the only constraint being, the total number of seizures equal to the number of random predictions. If a prediction occurs, then the intelligent system designed here should have a higher sensitivity of seizure prediction than a random predictor. Similar analysis are also used in seismology, where the null hypothesis is regarded as inevitable for earthquake predictions [195]. A set of 50 surrogate onset times are constructed for each patient and the sensitivity of the predictor is estimated for all prediction windows from five minutes to 4 hours in steps of 5 minutes. The test with random prediction is shown for the five patients in Figure 8.48. From the figure it can be seen that in general the sensitivity of predicting randomly when the patient has few seizures is quite low, as in case of patient C scalp and L scalp. However, also in case of high number of seizures as in case of E scalp and X scalp the sensitivity of prediction remains low. But in case of T scalp the sensitivity is comparatively higher. This can be explained by the spread of the seizures in each patient. In patient T scalp the number of seizures is higher and most importantly the seizures are well distributed across the entire recording and the probability of a random prediction is higher in that case as any random number might make at least 2-3 correct predictions. In that case the intelligent predictor being designed would have to show a prediction sensitivity higher than this already high random prediction. However, in case of patient E scalp and X scalp, even though the number of seizures is higher than C scalp or L scalp, but most of their seizures are grouped together, such that a random predictor away from this cluster would miss all these seizures. Applying a distribution constraint on the random seizure events correlated to the seizure events distributed on the real data, may bias the test and is avoided here. To analyze the performance of the intelligent predictor designed in this research, compared to the random predictor, the sensitivity of the predictor for a particular prediction window (considered suitable for prediction in patient data) will be compared with the average sensitivity of prediction by the random predictor for that prediction window.

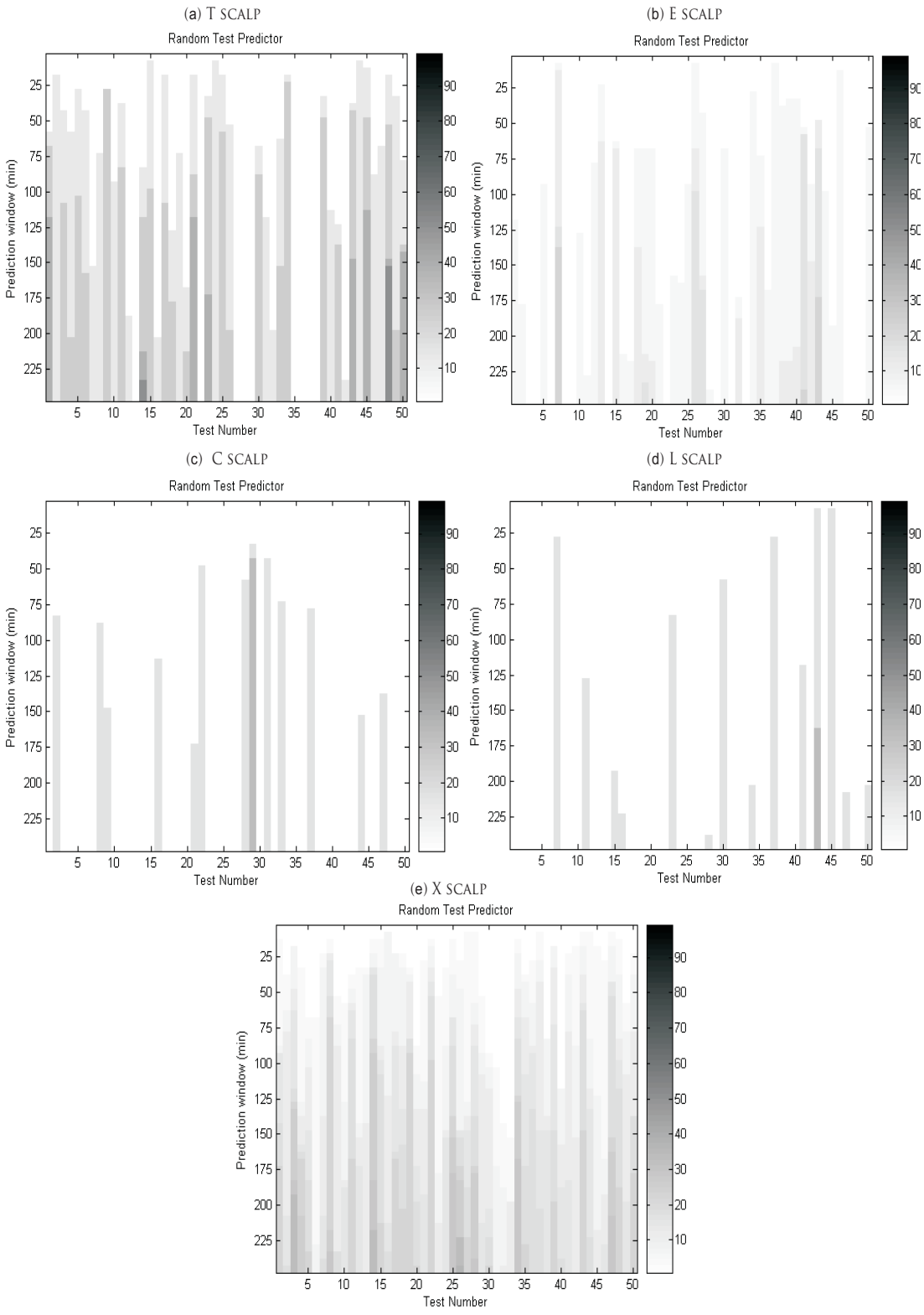


Figure 8.48: The test with a random predictor. Each graph shows the sensitivity of prediction with a random predictor for various prediction windows. The only constraint on the randomness of the seizure onset times is that the total number of seizure onset times is equal to the number of random predictions.

8.5.5 Probabilities of Prediction: Delayed Warning

In the above statistical analysis, a seizure event was said to be predicted if the event probability became higher than a threshold. The probability value that indicated a predictive event was not being analyzed as such. In this section, the probability values of predictive events in each prediction window, and one exemplary threshold, is analyzed for all seizures of two patients. A threshold level is first selected for each patient, that would lead to the highest sensitivity and specificity with the GMM probability curve. At this threshold level, the probability of prediction is assessed for each prediction window and for each seizure, shown in Figures 8.49, and 8.50 for supervised GMM ad Figures 8.51 for semi-supervised GMM. In general a trend is observed that the probability of prediction is higher for a longer prediction window and gradually reduces as the prediction window becomes shorter. This can be observed in seizures 2, 3, 4, 6, 7 and 8 for patient T scalp 8.49, and in seizures 1, 3, 4, 5 and 7 for patient X scalp 8.50. This trend might be useful in raising an initial *soft* warning and then a *stronger* alarm in a seizure prediction application. Such a possibility might be found useful in combining different intervention therapies to prevent a seizure occurrence, such as injecting long acting drugs at the soft warning and then if it is still succeeded by a stronger warning, an electrical stimulation is triggered.

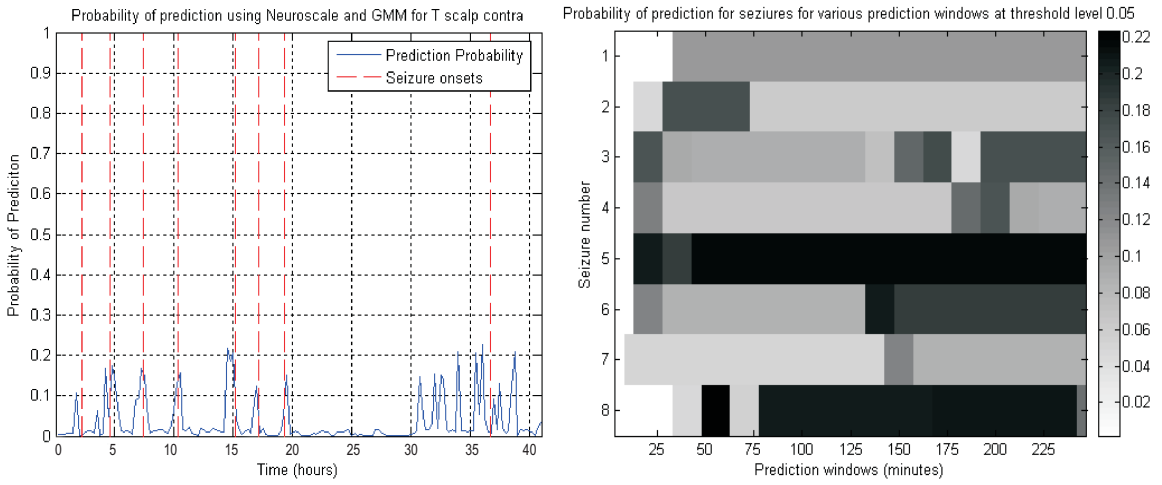


Figure 8.49: The probability curve obtained by GMM for **patient T scalp** is shown on the left. The vertical lines mark the seizure onset. The probability of a predictive event for each prediction window for threshold 0.05 is shown on the right. A general trend can be observed in most seizure prediction probabilities (Seizure 2, 3, 4, 6, 7 and 8), where the probability away from the seizure is higher and gradually reduces to a lower probability as a seizure approaches.

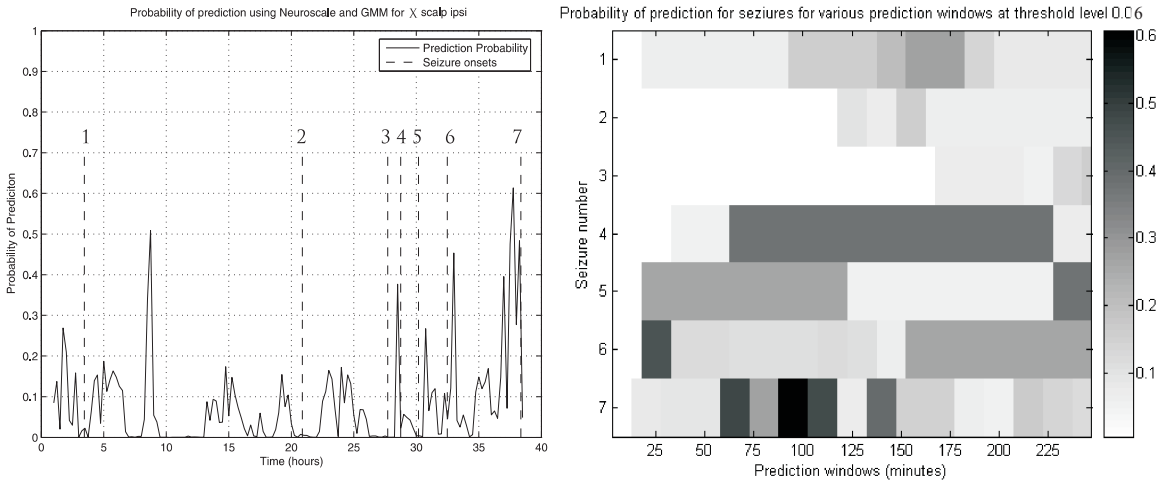


Figure 8.50: The probability curve obtained by GMM for **patient X scalp** is shown on the left. The vertical lines mark the seizure onset. The probability of a predictive event for each prediction window for threshold 0.06 is shown on the right. The general trend observed in the previous patient can again be observed in most seizure prediction probabilities (seizures 1, 3, 4, 5 and 7) for this patient as well, where the probability away from the seizure is higher and gradually reduces to a lower probability as a seizure approaches.

8.5.6 Results III: Seizure Prediction: Train and Test with Unseen Data

The parameters and features selected in above analysis are now applied to data that was held-back and was not seen during the training and discriminant analysis process. This is carried out in order to remove any bias that can easily be formed by feature subsets for the data that is trained and tested on during discriminant analysis. The technique used for seizure prediction in this unseen data set is the ***Neuroscale and semi-supervised GMM with labelled and unlabelled data***. The data consists of a set of four patients, each patient having 3-4 days of continuous multichannel scalp EEG. The EEG is pre-processed by removing mean, filtering high frequencies above 50 Hz. ICA is then applied on an ictal- interictal segment to obtain a spatial template. This template is used to obtain two spatially constrained signals using spatially constrained ICA, from the rest of the multichannel EEG data. Significant phase synchrony is then obtained for the spatially constrained signals (contralateral and ipsilateral). The PLV-d is obtained from the significant PLV output and is used for feature extraction. Feature subset 6 is used to create the multidimensional feature vectors as it was found most suitable in the preceding analysis. The multidimensional feature vectors for the entire data are divided into a training and a test set. The training data is used to train a patient specific radial basis network using a supervised topographic Neuroscale Network, resulting in a 2-D projection map of the feature space. The 2-D feature points are then used in GMMs with labelled and unlabelled data to train Gaussian mixtures such that the total probability of a feature point to exist in pre-seizure and interictal GMM is equal to one. The resulting probability of each data point to exist in

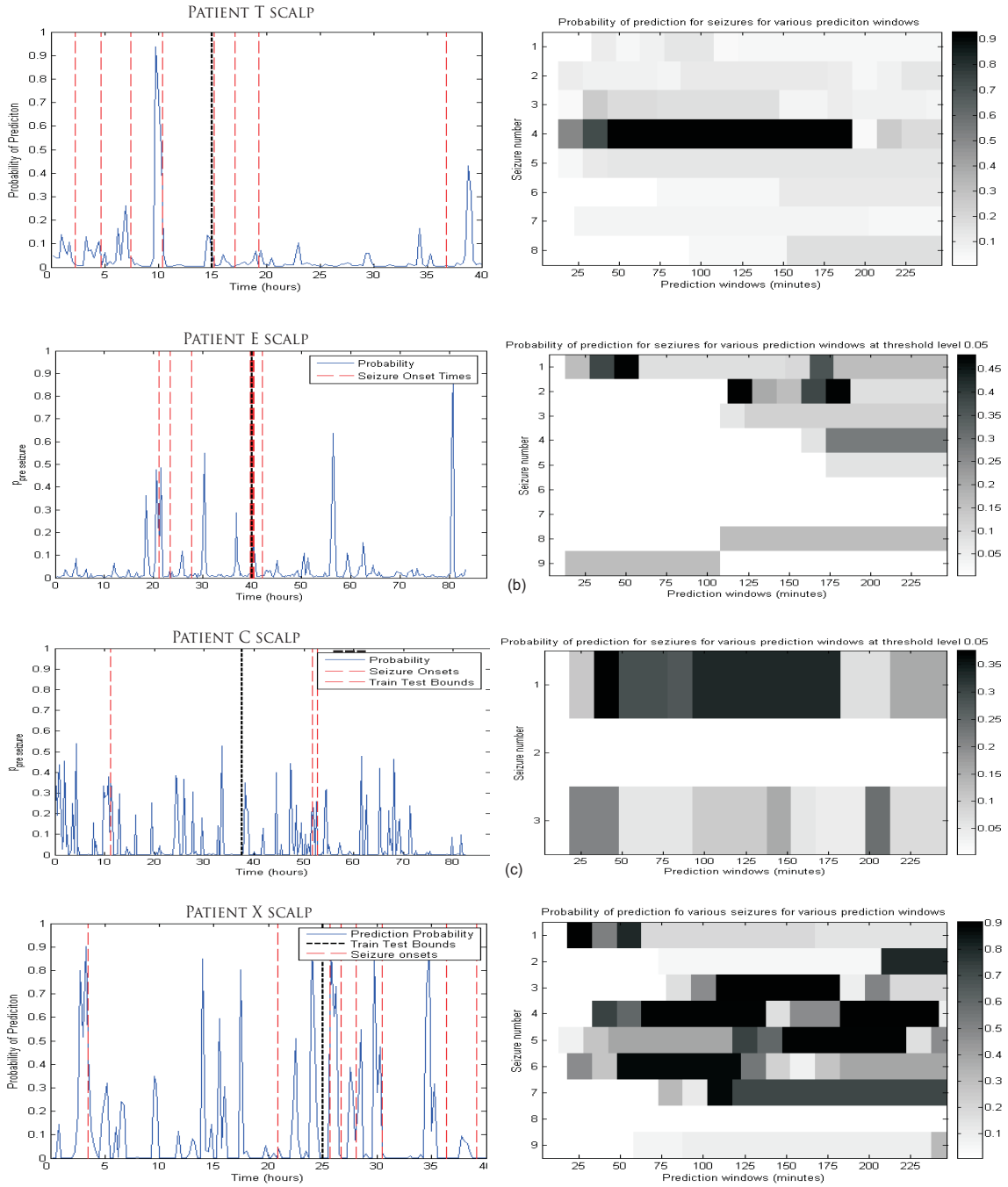


Figure 8.51: The extent of probability of the prediction for various seizures at different prediction window lengths (*right*). The probability curves obtained by semi-supervised GMM for **patients T, E, C and X scalp** (*left*). Seizure onsets (*vertical lines on left side graphs*). The general trend observed is a higher probability away from the seizure and lower probability closer to the seizure, as in (a) seizures 1, 2, 3, 4, 7 and 8 (b) seizures 1, 2, 3, 4, 5 and 8 (c) seizure 1 (d) seizures 5, 6, 7 and 8 .

a pre-seizure GMM is used to track the occurrence of a seizure prediction event. This probability time series is then statistically evaluated for sensitivity and specificity of seizure prediction, for various prediction windows (5 minutes to 4 hours) and threshold levels. ROC curves for the same are also shown. This prediction is then tested against a random predictor. The amount of probability of prediction for various prediction windows is also analyzed for all seizures for each data set.

Test Patient 1: Y scalp This patient had four seizures on day one. The seizures being on the day one and very closely placed, allows very less data to be used as training. The ideal data for training would have at least 3 seizures and 15-20 hours of interictal data points. It would help model a better GMM set. In Y scalp patient data, the data covering the first two seizures and upto the third seizure was used for training. The first two seizures being very closely placed do not possibly give a lot of information about the prediction statistics. However, as the results show (see Figure 8.52 and 8.53), the training data is aptly clustered into seemingly two clusters with Neuroscale (see Figure 8.52 (a)). A few predictive points seen to be overlapping with the interictal points can possibly be the pre-pre seizure points that may have similar features as interictal points. This feature map is then segmented with a label-unlabelled GMM model (see Figure 8.52 (d)). It is observed that even though it may not be efficient for modelling pre-ictal events but well enough for classifying interictal events. The GMM on the unlabelled test data shows that the model does not predict seizure events falsely, having a low false detection rate and is able to predict one of the two test seizures correctly, within a small prediction window.

The algorithm is applied to the contralateral as well as the ipsilateral data space of the patient data. The clusters of Neuroscale map are found to be much distinct in the ipsilateral case (see Figure 8.53 (a)). This leads to a GMM model with a much noisier probability plot (see Figure 8.53 (d)) as the probability of an event being interictal or pre-ictal is not found to be distinct, moreover with the limited training data. However, the probability plot of this data is able to predict higher number of seizures with higher probability (higher threshold). This can be seen from the sensitivity plots in Figures 8.52 (e) and 8.53 (e). The contralateral signals show a higher sensitivity for the shortest prediction window being 60 minutes (see Sensitivity graph and ROC curves in Figure 8.52 (e) and (f)). Whilst in case of ipsilateral data, the shortest prediction window is 80 minutes (see Sensitivity graph and ROC curves in Figure 8.53 (e) and (f)). The false positives per day for the contralateral data with the best possible sensitivity (100%) combination of threshold 0.04 and prediction window 80 minutes is about 3 false predictions per day. Whilst, for the ipsilateral data the most suitable combination for high sensitivity of 100% is a threshold of 0.1 and prediction window of 80 minutes, with 80% specificity (Figure 8.53 (e)). Nevertheless, both the contralateral and ipsilateral models outdo the random predictor in predicting seizures. The random test predictor shows $<$ 10% predictions on average for a prediction window of 60 minutes and 80

minutes for contralateral and ipsilateral data respectively as shown in Figures 8.52 (c) and 8.53 (c). The amount of probability of a predictive event for various prediction windows is also shown for each seizure across time in Figures 8.52 (b) and 8.53 (b). It is not found to be informative in the case of contralateral Y scalp data, as a constant high or a constant low probability with a sudden peak is seen across time. In case of ipsilateral Y scalp data, the amount of probability very close to a seizure is observed to be negligible in all four seizures. Whilst, no other general pattern can be observed. In seizure 1 and 4, the amount of probability is observed to be higher away from the seizure, leading to a reduced probability when approaching the seizure. In seizures 2 and 3, the probability is observed to follow a trend of being lower-higher-lower as the seizure approaches.

Test Patient 2: D scalp This patient has four day continuous multichannel scalp EEG. It was pre-processed and unmixed with spatially constrained ICA, as mentioned before. The multidimensional features were then extracted using feature subset six and these were reduced to a two dimensional map using Neuroscale. The semi-supervised topographic neuroscale projection was divided into a training and test set for further GMM modelling with labelled and unlabelled data. This patient had only one marked seizure and few seizure like events (that have not been highlighted here). The lack of seizures, gives less training and test data to test its capability as a predictor. Nevertheless, it is analyzed for a novelty detector training, where it is trained on much interictal data and few pre-seizure points, and when presented with only interictal data, it should detect them as interictal. To include the seizure, the training data used is much more than 50% of the data. Interestingly, in the short test data, not many false positives are found, showing perhaps a subtle distinction modelled by GMM in the feature space of the pre-seizure and interictal points, see Figure 8.54 (d). It is observed to perform better than a random test predictor as seen in Figure 8.54 (c). The sensitivity is 100% for a threshold limit of 0.1 with short prediction window of 10 minutes. This threshold and prediction window leads to 4 false positives per day, as can be seen in the sensitivity, false positive graphs and ROC curves in Figures 8.54 (e) and (f). The amount of probability of the seizure prediction is observed to be higher away from the seizure, gradually reducing slightly over time as the seizure approaches.

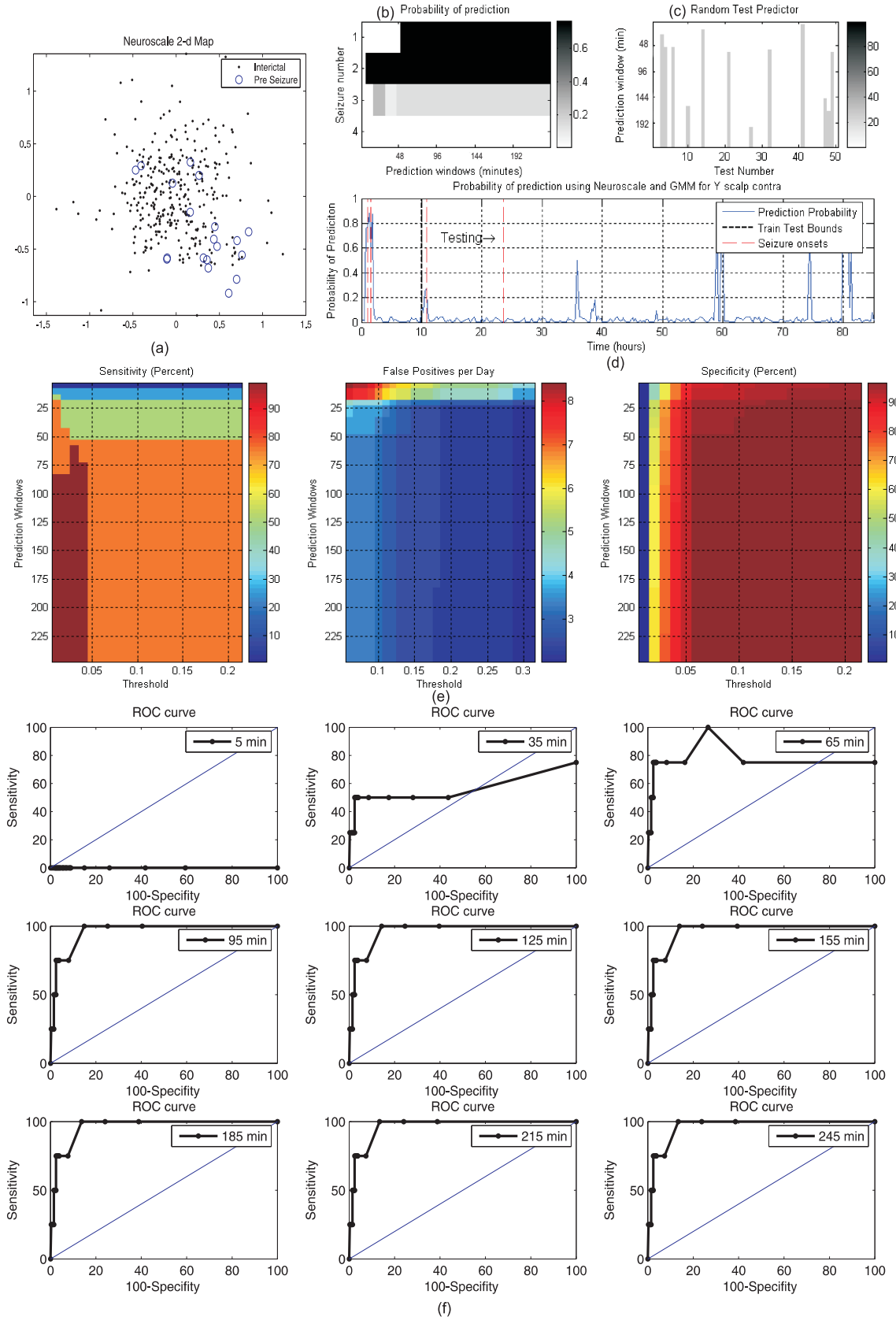


Figure 8.52: Test and Train with Neuroscale and GMM for unseen data: patient **Y scalp Contralateral** using feature subset 6 and alpha 0.8 (a) Neuroscale 2-d projection (b) The probability of the predictive events before a seizure (c) Sensitivity with a random test predictor for various prediction window lengths (d) The probability plot from the GMM for the pre-seizure events (e) Sensitivity, False Positives per day and Specificity (f) ROC curves for various prediction windows and thresholds.

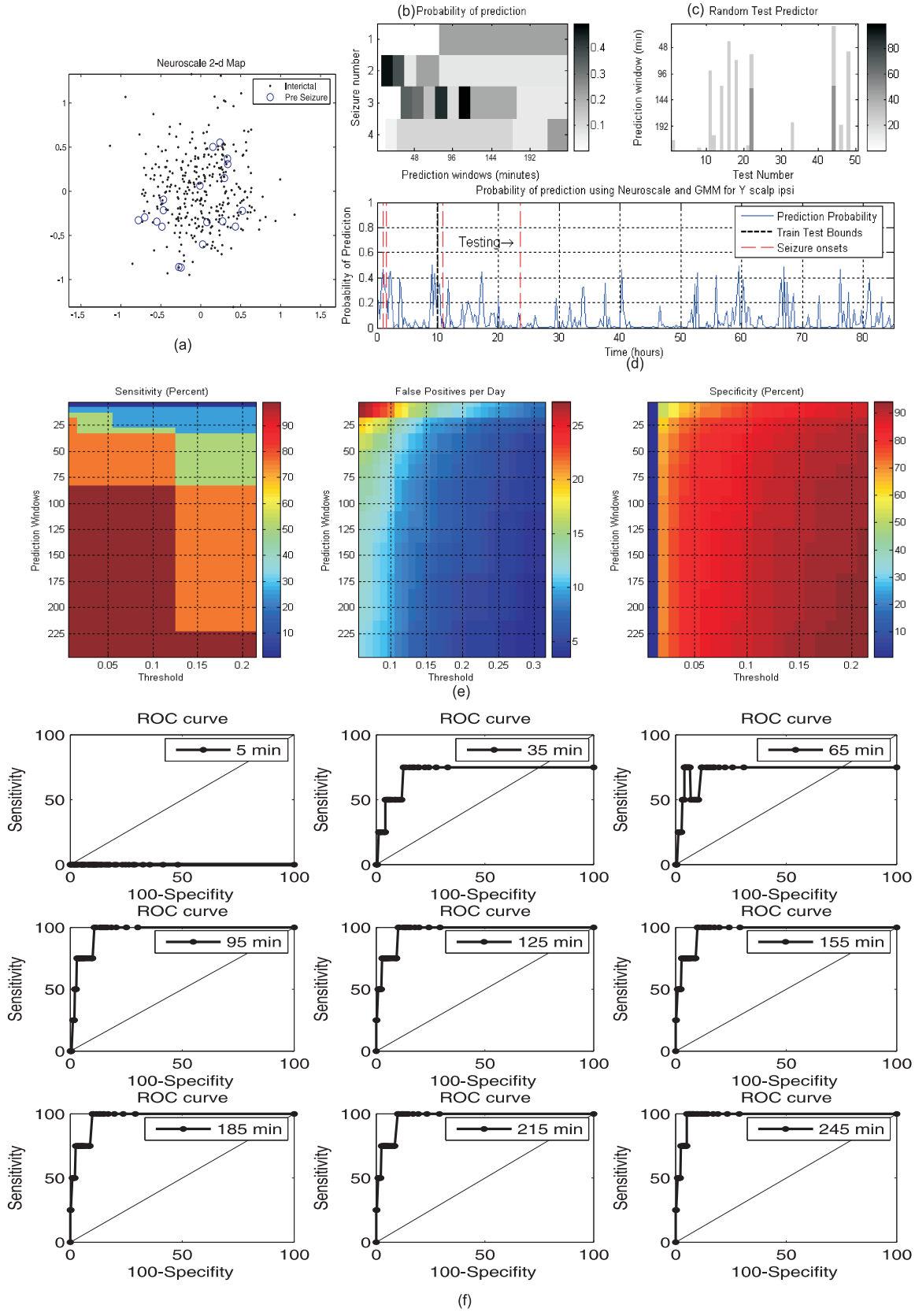


Figure 8.53: Test and Train with Neuroscale and GMM for unseen data: patient **Y scalp Ipsilateral** using feature subset 6 and alpha 0.8 (a) Neuroscale 2-d projection (b) The probability of the predictive events before a seizure (c) Sensitivity with a random test predictor for various prediction window lengths (d) The probability plot from the GMM for the pre-seizure events (e) Sensitivity, False Positives per day and Specificity (f) ROC curves for various prediction windows and thresholds.

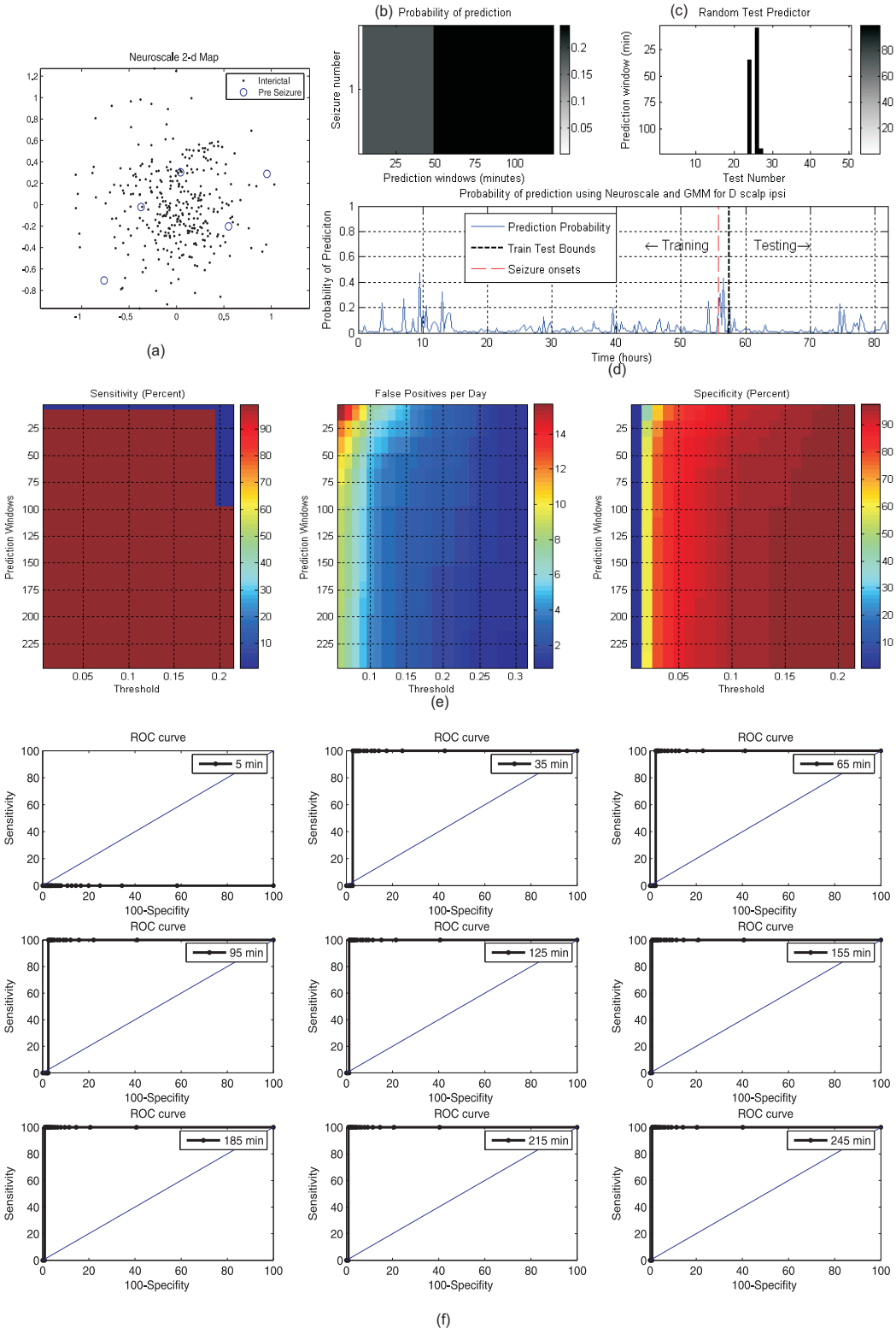


Figure 8.54: Test and Train with Neuroscale and GMM for unseen data: patient **D scalp Ipsilateral** using feature subset 6 and alpha 0.8 (a) Neuroscale 2-D projection (b) The probability of the predictive events before a seizure (c) Sensitivity with a random test predictor for various prediction window lengths (d) The probability plot from the GMM for the pre-seizure events (e) Sensitivity, False Positives per day and Specificity (f) ROC curves for various prediction windows and thresholds.

Test Patient 3: BB scalp The data set for Patient BB scalp is for two days of continuous multichannel scalp EEG. The patient had four seizures on day one. The data set was pre-processed and unmixed with spatially constrained ICA as per the algorithm defined in the previous Chapter. The multidimensional features were extracted from the PLV-d curves of the spatially constrained source signals of the data. The features were projected using a supervised Neuroscale to obtain a two dimensional feature space. This space was divided into a training and test set to model a GMM with labelled and unlabelled test data. The probability of a pre-seizure event was hence tracked using the probability contours obtained from the trained GMM. The seizure events of this patient were in day one, and in order to leave few seizures for the test data, the training data used was 12 hours long, and included 2 seizure events. The test data had two seizure events and 30 hours of interictal data. The first step of Neuroscale topographic projection showed some demarkation between the interictal and pre-ictal feature points. The pre-seizure feature points being the farthest from the center as compared to the interictal points, see Figure 8.55 (a). The resulting probability map from the GMM model is shown in Figure 8.55 (d). It shows bursts of high probability at certain times during the entire recording, few of them correlating with the seizure. The sensitivity is observed to be as high as 100% for short prediction windows of 65 minutes, but at this threshold and prediction window the number of false positives being high, reduce the specificity to 75%. Longer prediction window of length 80-125 minutes shows a 100% sensitivity and 5-10 false positives per day. A compromise on the sensitivity, being 75%, a prediction window of length 35 minutes is also appropriate, with 5 false positives per day (see sensitivity, false positives per day and specificity graphs, and the ROC curves in Figures 8.55 (e) and (f)). It is observed to outdo a random test predictor for prediction windows of 35, 65 or even 125 minutes 8.55 (c), that does not show more than 20% prediction sensitivity on average, at these prediction window lengths. In two cases, the amount of probability of prediction is observed to be higher at times further away from the seizure, gradually reducing as the seizure approaches 8.55 (b). Interestingly, a very low or negligible probability is observed for a short prediction window of 10-15 minutes, supporting a hypothesis of *a gradual build up of synchrony over a long time, rather than a short sudden change at seizure onset with a loss of synchrony 2-3 minutes before the seizure.*

Test Patient 4: O scalp Data for patient O scalp is for four days of continuous multichannel scalp EEG. The patient experiences 29 seizures in last two days. The seizure are quite clustered together in the last day. The data of this patient is also pre-processed and unmixed as for the previous data sets. The source signals are extracted with spatially constrained ICA their PLV-d curves are constructed. The features of the synchrony dynamics are extracted in the form of multidimensional feature vectors using PLV feature subset six. The Neuroscale topographic transformation technique is used in a supervised form to reduce the dimension of the feature space to two dimensional.

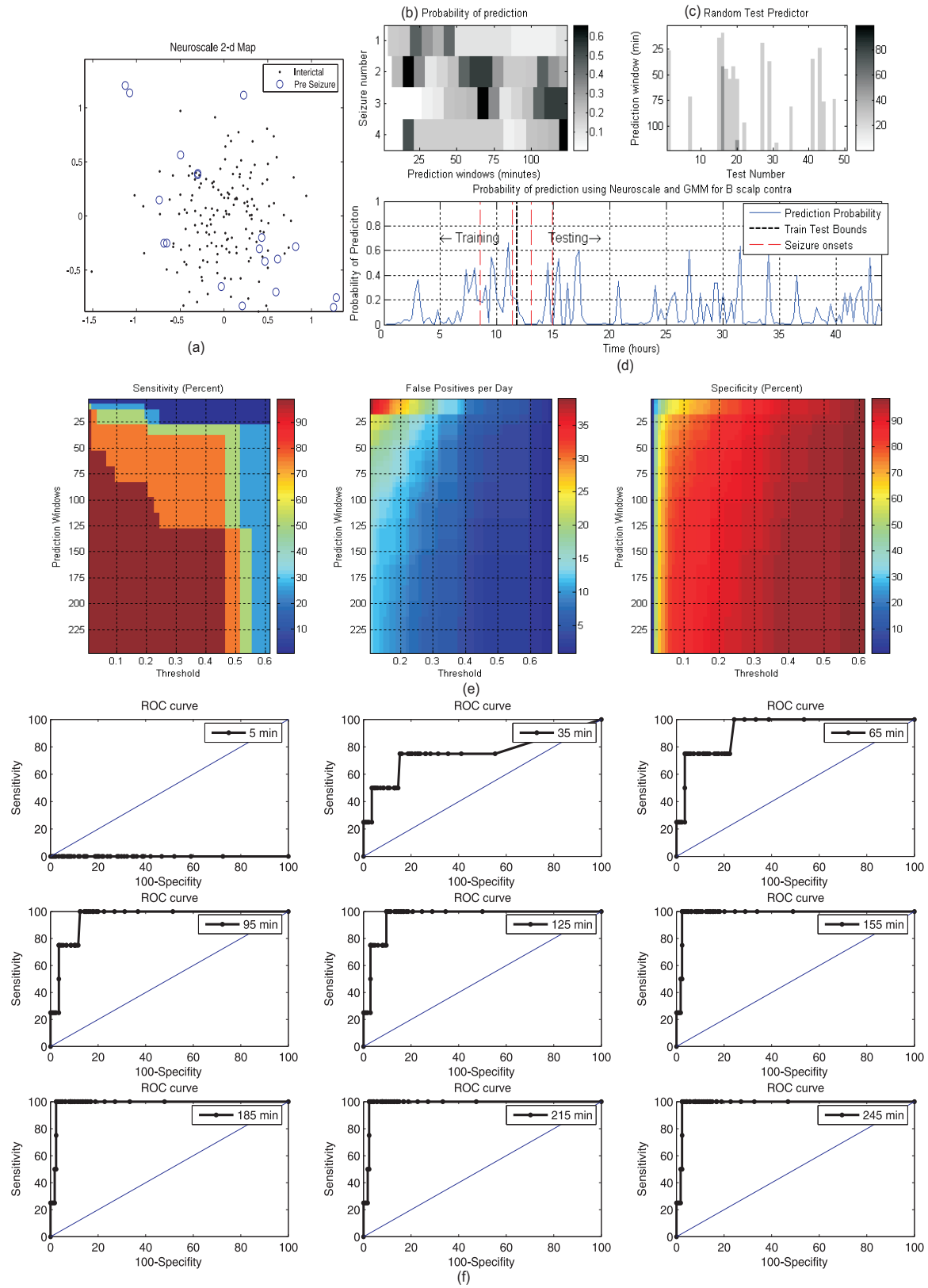


Figure 8.55: Test and Train with Neuroscale and GMM for unseen data: patient **BB scalp Contralateral** using feature subset 6 and alpha 0.8 (a) Neuroscale 2-D projection (b) The probability of the predictive events before a seizure (c) Sensitivity with a random test predictor for various prediction window lengths (d) The probability plot from the GMM for the pre-seizure events (e) Sensitivity, False Positives per day and Specificity (f) ROC curves for various prediction windows and thresholds.

This feature space is then divided into a training and test data set, with the training data set including sufficient ictal activity. The Neuroscale projection of the data set shows some clustering, even though there is much overlap of the pre-ictal and interictal data points 8.56 (a). The probability map from GMM modelling is shown in Figure 8.56 (d). Bursts of high probability can be seen at various times, indicating pre-seizure events. *A big cluster of seizures occurring within a time window of 10-30 minutes is seen in the later part of the data.* The changes in feature space if inadequately modelled or supplied with feature information can easily lack the ability to predict this cluster, apparently affecting the sensitivity analysis noticeably (see sensitivity, false positives per day and specificity curves in Figure 8.56 (e)). With the present predictive capability, the ROC curves in 8.56 (f) show a prediction window of 95 minutes to be more appropriate for a high specificity (55%) and high sensitivity (65%). It should also be recalled that patient O scalp had complex types of seizures. The patient had three types of seizures intermixed in two days of seizure activity. It is quite possible that the spatial template used as per one of the type of seizure activity was not suitable to classify or identify features of the other type of seizures. Nevertheless, prediction is shown to be possible with this test data and better than a random test predictor that has an average sensitivity 6.4% and the maximum sensitivity of 17% for a 95 minute prediction window 8.56 (c). The amount of probability of seizure prediction shown in Figure 8.56 (b), shows in general trend of higher probability of prediction further away from the seizure as compared to being near the seizure. Also, it is to be noted that the probability of prediction is quite low in the 10-15 minutes of prediction window, showing some underlying difference in features of the data points near the seizure as opposed to the interictal or pre-ictal.

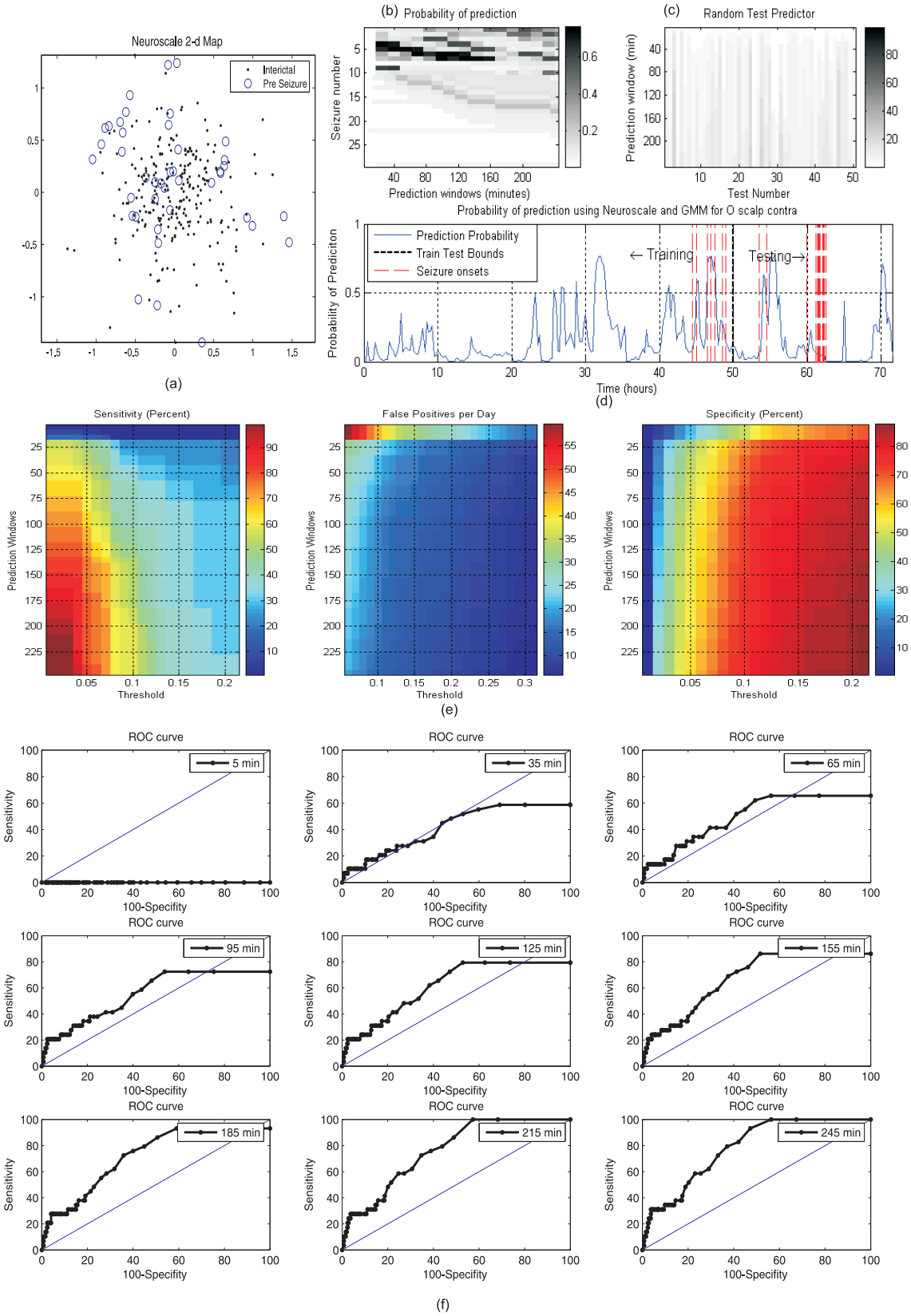


Figure 8.56: Test and Train with Neuroscale and GMM for unseen data: patient **O scalp Contralateral** using feature subset 6 and alpha 0.8 (a) Neuroscale 2-D projection (b) The probability of the predictive events before a seizure (c) Sensitivity with a random test predictor for various prediction window lengths (d) The probability plot from the GMM for the pre-seizure events (e) Sensitivity, False Positives per day and Specificity (f) ROC curves for various prediction windows and thresholds.

8.6 Discussion

In this chapter, multivariate discriminant analysis was introduced and applied. A set of 23 features were identified by observation of the feature time series obtained from spatially constrained ICA and phase synchronization analysis of continuous long term scalp EEG data of nine patients (about 2-4 days each). The features consist of 10 IC features (extracted from the spatially constrained IC's used to obtain the PLV-d curves) and 13 PLV features (obtained from the PLV-d curves of the PLV analysis of spatially constrained ICs of the EEG data). As all combinations of the features cannot be analyzed due to limitation of time, relatively fewer combinations of the features were created (30 feature subsets). These were created by observation of the prominent features that show predictive aspects in the PLV-d curves of the nine patient date sets.

The selection of the the feature subset, most suitable for prediction was carried out in two steps: (1) Dimension reduction of the feature space to two dimensions using supervised Neuroscale network and (2) Probabilistic density modelling of the feature space with supervised and semi-supervised GMMs. These applications helped to track the probability of a seizure prediction event. The dimension reduction performed with the Radial Basis Neural Network-Neuroscale on training data, resulted in a two dimensional feature space and a trained transformation network. The output feature map of Neuroscale was used as a two dimensional input to the supervised GMM. GMM density classification was carried out in a supervised mode, with density maps being trained for both the pre-ictal and interictal classes, using the training data. The trained Neuroscale network and the GMMs were then applied to the test data. The probability of every feature point from the test data was tracked in the pre-ictal density map. This helped to track time events that would be classified as predictive by the trained Neuroscale-GMM system. This probability map was used for further statistical analysis to ascertain a suitable threshold level and prediction window length that would lead to a high sensitivity and high specificity, with low false positives per day. The ROC curves were constructed for various prediction windows and threshold levels, that showed a relationship between the sensitivity and specificity and helped in selecting the most suited prediction window.

Many of the feature subsets were discarded in the first dimension reduction step due to their inability to show any distinction in pre-ictal and interictal classes in the two dimensional maps, across the five patient data sets, even with supervision. The feature subset selected was made of purely PLV features. It was observed that the combination of IC and PLV features was not improving the Neuroscale maps, rather it was many times impairing the mapping. From the six PLV only feature subsets, only feature subsets 6 and 1 were pursued as they both showed better clustering, across maximum patients, in a supervised mode.

The supervised Neuroscale-GMM classifier then helped to finalize the 'feature subset' to be used, based on its performance in seizure prediction. The feature subset 6 with α=0.8 was selected for further analysis.

The prediction windows of the the order of 35 minutes were found to show on average 86% sensitivity and 80% specificity.

The semi-supervised GMM technique based on labelled and unlabelled data was also applied on the five data sets. The multidimensional feature vectors were first transformed to a two dimensional feature space using a supervised Neuroscale network. The supervised Neuroscale used topographic information as well as class information to map the multidimensional data onto a two dimensional space. This was followed by the application of semi-supervised GMM to verify the classification of predictive and interictal features using the PLV feature subset six.

The positive results obtained for all the five data sets verified the existence of a predictive feature space that is accessible through the synchrony dynamics of spatially constrained source signals of the EEG.

The predictions were then tested with a random predictor. 50 surrogates of random event markers were generated for each patient, and these were tested for prediction windows varying from 0-4 hours (in steps of 5 minutes).

Additionally, the amount of the probability that identifies an event as a predictor, is also analyzed for various seizures and prediction windows. It was observed that the probability of prediction was apparently higher further away from the seizure and gradually reduced as the seizure approached, at times even becoming very less immediately before the seizure onset. ***This indicates significant changes taking place in the brain long before the seizure onsets that have a higher predictive capability, compared to the immediate changes that occur just before a seizure onsets.***

The final analysis was performed as a validation of the entire algorithm using previously unseen data. In semi-parametric models of machine learning, it is possible that the feature subset is biased to the data that is being trained and tested on it for feature extraction. Therefore, a data set that is completely unseen during the discriminant analysis is tested with the feature subset selected. The data used was from four patients, each having 2-4 days of continuous multichannel scalp EEG. This data had not been previously used for any training and discriminant analysis. This data was pre-processed and unmixed as per the algorithm designed. Its seizure source signals were obtained using the spatially constrained ICA. The significant synchrony dynamics were tracked for each patient data across time using PLV and the PLV-d. The PLV-d curves were used to extract multidimensional feature vectors, using the feature subset selected in the above discriminant analysis. The feature vectors were divided into train and test sets.

The Neuroscale topographic transformation training was carried out in a supervised mode with α=0.8, followed by a supervised GGMM. The probability of the test data was then obtained across time for the pre-ictal GMM in order to track the occurrence of predictive events. The statistical analysis with sensitivity, specificity and ROC curves were constructed for prediction windows varying from zero to four hours in steps of five minutes, and varying threshold levels (zero to one). The performance of the four test patients was found to be better than a random predictor. **A sensitivity of the order of 65%-100% and specificity of 65%-80%, for a prediction window of 35-65 minutes was achieved with the semi-supervised GMM technique.**

Chapter 9

Discussion and Future Work

The aim of this research has been advancement of knowledge for seizure prediction using long term continuous brain activity. To date, seizure prediction methods have been applied on short term brain activity and are in need of a specificity analysis along with the reported sensitivity for a clinically viable prediction system. It is understood that such a system would ideally be expected to perform in real time, have a low rate of false positives, high sensitivity and have a practically feasible prediction horizon.

This thesis has examined the use of Independent Component Analysis and phase synchronization on multichannel continuous EEG for epileptic seizure onset prediction with a novel approach. The novelty in the technique resides in the use of:

Multichannel EEG: Multichannel EEG has spatial as well temporal information that is not tapped to its full potential in the context of seizure prediction. Most of the statistical techniques used in this context are either based on single channel or two channel data, where subjective selection of EEG channels becomes crucial. We have made use of the spatial as well the temporal information inherent to the EEG signals with the help of temporal decorrelation and spatially constrained ICA.

Long term Continuous EEG: Previous research for seizure onset prediction, lacks the statistical testing based on long term EEG. Much research has been carried out on short EEG ictal/interictal segments or on limited number of patients or types of epilepsies. However, a rigorous seizure predictor would be best modelled on long term data, because statistics on short term data are not able to validate the performance of the system with the ongoing complexities of the normal EEG. It is essential to be demonstrated that the dynamics in the brain ascertained close to a seizure onset are unique against the complex dynamics of the interictal periods, when the brain may be involved in inexplicable unknown linear or nonlinear processes. Therefore, this research has been based on long term EEG of multiple patients having different types of epileptic manifestations. The EEG used is long

term and continuous, 2-4 days of continuous data per patient for nine patients. Furthermore, no segments of data were discarded as unusable.

Unmixed source signals: The use of unmixed source signals from the EEG is not new, but to use it on long term EEG for seizure onset prediction along with phase synchronization is an advancement. It is not trivial to use ICA with long term EEG as it involves the use of ICA on windowed EEG, in order to fulfill the assumption of stationarity. However, windowed ICA is plagued by challenges such as obtaining a new combination of sources in every window, lack of morphological information of a seizure source in windows preceding a seizure, splitting of source signals in uncontrolled ways in every window, and the permutation of sources with ever window. The novel technique used here to overcome such challenges has been to concentrate on brain activity spatially constrained to the probable seizure source. The aim was to track dynamics of a seizure source across time as a seizure approaches, best captured at the moment in the ipsilateral or contralateral areas where the most prominent activity of the seizure is observed. It should be noted that in this research, the ipsilateral spaces were not considered to be the most informative, as is conventional in such studies, but the signals that showed more informative patterns among the contralateral and ipsilateral signals were used. This can be explained by acknowledging that in certain types of epileptic manifestations, it is possible that certain areas of the brain trigger the seizure activity (termed as the true seizure focus), affects of which may be observed contralateraly clinically. Such information may not be captured or perceived with the EEG alone, due to a lack of morphological knowledge/evidence about such subtle changes. Epileptic manifestations are typically identified with rhythmic ictal activity or typical known patterns observed in the EEG of the patient coincident to the clinical seizure, mostly based on experience. Interictal spikes and abnormal discharges are also informative of epilepsy, but are not used for confident lateralising as such as their occurrence is not fully understood or explained. These manifestations of bursts of synchrony may perhaps be more informative than is understood today. Therefore, even though this research is a scratch on the surface of the research required to understand and provide help with epilepsy, it will surely help to broaden the perspective of those involved with it in the medical or biomedical engineering field.

Synchrony dynamics with PLV-d of unmixed signals: Although assessing the synchrony dynamics with the use of phase synchronization is being much studied in context of seizure prediction, yet statistical studies with multiple patient data with long term multichannel scalp EEG has been lacking. Also previous research in this area has concentrated on the use of phase synchrony on a subjective selection of raw EEG signals. Recent literature that shows the estimation of spurious synchrony when estimated with mixed signals, raises the question of its appli-

cation to the EEG signals. EEG signals are understood as being mixed signals that are overlapped as a result of volume conduction, with original independent sources being embedded in the cortex, away from the scalp. Therefore, an essential step to the estimation of phase synchrony on the EEG signals, is their unmixing. This is an advancement in the context of epileptic seizure prediction research. Furthermore, the use of PLV-d is entirely a novel statistic used in this analysis to track the synchrony of two frequency bands in conjunction, uncovering interesting (possibly) circadian-like rhythms. Also, typically, synchrony dynamics over short time scales have been analyzed before, whereas the filtered PLV-d curves used here have highlighted the gradual changes in synchrony over days that hold useful information for seizure prediction.

Discriminant analysis with Neuroscale and GMM: Discrimination based on features of the synchrony dynamics from multichannel EEG is novel in the use of PLV-d, the feature subsets used and the underlying multichannel information. The extraction of multidimensional features from the synchrony dynamics alone has not been performed before, and it has supported the hypothesis that the synchrony dynamics of the brain activity holds much more information than is perceived by the numerics alone. Instead of using solely the information of the increase or decrease of synchrony, the way it develops over time can also be informative, sometimes not apparent to the visual system. The dimensional reduction with Neuroscale has been very much useful for bringing forth underlying structure in the multidimensional features. Neuroscale and GMM have not been used previously for seizure onset prediction. Especially, the use of GMM in the ***semi-supervised*** way, that has not been previously used, helped to validate the existence of distinct interictal and pre-ictal features.

The overall performance of the **ICA-PLV-Neuroscale-GMM** algorithm has been quite useful in demonstrating:

- the existence of pre-ictal features in synchrony dynamics of unmixed scalp EEG;
- the possibility of seizure prediction using intelligent algorithms;
- interesting underlying possible Circadian-like rhythms of brain synchrony;
- informative patterns in the slow dynamics of synchrony in the brain;
- the usefulness of unmixing EEG for phase synchrony analysis in the context of seizure prediction.

The course of this research had to be guided towards a very narrow perspective, as there were many possible divergences equally apt for analysis, typical to any research.

These are reflected upon here, as now they can be helpful in understanding areas that might have polished the analysis, or raise questions that may motivate further research, and perhaps elicit a new aspect of the research problem itself.

The very first step being **the data**, where-in many choices were available: if the data to be used would be scalp, intracranial or both. We preferred to keep it more general and collected both scalp and intracranial data. A more important concern was the type of data to be used, i.e. data specific to a certain type of epilepsy; a certain category such as with drug reduction or without drug reduction; within a certain age limit, etc. The decision was constrained by physical issues of data collection, for example, the amount of data that could finally be collected would be restricted to the number of patients coming in to the Neurophysiology department of the attached Hospital on their own accord without recruitment, due to ethical and financial reasons. This would restrict the amount of data collected on the whole in the limited time, especially when about nine months were spent on getting the ethical committees' approval to collect such data, added to the uncertainty of the research evolvement. Therefore, a general data set was collected, almost in real time (due to memory space constraints at the hospital), and it included patients with different types of epilepsies, various ages (pediatric patients were excluded), and both with and without drug reduction. Data much more than has been analyzed was collected, but it had to be trimmed as some patients did not experience a seizure during the data recording, or the suggested diagnosis would be 'non-epileptic' after a few weeks.

An ***ideal data set*** for this research would have been multichannel scalp and intracranial EEG of at least ten patients each with 'confirmed' or 'most likely' diagnosis of two-three different types of epilepsies, with at least five patients each in drug and without drug reduction category, as drug reduction can bring in unexplained dynamics, sometimes different than the patients typical brain activity. It is hypothesized that, as the anti-epileptic drugs are reduced quickly within 3-4 days, the brain activity may or may not fluctuate in chaotic ways as the restrains on the seizure activity is removed, confusing the signal processing analysis. Each data set would ideally have at least 6-10 seizures or more distributed over a few days, with sufficient quieter interictal activity as well. The annotations of the data would highlight the sleep and wake times, the seizure onsets (clinical and electrographical), major artifacts such as eating food, moving, etc. A data set of similar continuous multichannel scalp or intracranial EEG of normal persons would be beneficial for a comparative analysis.

The data brought other intricacies, such as, if the extraction of the spatial template was exceedingly based on the diagnosis (that lacked a 'confirmation'), it could be possible that the diagnosis was in error and the investigation was carried out in a space of the brain that was off beam, likely to produce uninteresting results. However, a 'confirmation' of the epileptic diagnosis is usually obtained *ad hoc* through the post

operative results, if the patient undergoes epileptic surgery. The confirmation of the seizure focus is possible through the surgical epileptic resection (removal of a part of the brain), if the patient experiences seizure reduction or remains seizure free when observed for six months or a year after the operation. These can be rare and not easy to obtain within the limited research time. In this case a benefit of doubt had to be given to the uninteresting results. Thus in formulating the analysis a general view had to be considered from time to time instead of being strictly patient specific, whereas the results also had to be analyzed patient specifically considering their individual epilepsy manifestations which were difficult to generalize.

Scalp and intracranial EEG signals have been extensively used for seizure onset prediction in their raw form. Ideally EEG is a technique to detect ongoing activity in the neuronal populations close to an EEG electrode. However, due to the complex anatomy of the brain and the distribution of the neurons, and the *effect of volume conduction* source distribution or mixing takes place. Along with such complex mixing and background noise, the signals at the electrodes also get mixed with *physical and biological artifacts* such as ocular, muscular and cardiac signals; power line noise, electrode movement and electrode pop. In addition, as the signals travel through the soft bone and tissue, they *get attenuated and diffused*. Diffusing causes nearby signals to overlap and attenuation can make an expectedly weak signal from a seizure source (prior to a seizure onset) to get overshadowed in the background noise and EEG. This makes the ***EEG signals a set of mixed signals***. Thus influencing the estimation of linear and nonlinear quantities from the EEG, leading to misinterpretations and inconsistent measures. It is not trivial to revert such mixing, due to lack of knowledge about the mixing process or the underlying true signals. It is neither easy to filter out the artifacts or noise from multichannel EEG to improve the signal to noise ratio as the artifacts that are spread across frequency bands, across time and electrode channels in a complex way, difficult to be selectively filtered out. Thus, it is important to pre-process the EEG to improve the signal to noise ratio, remove artifacts and appropriately unmix it, before analysis. This motivates the first pre-processing step for the EEG signals namely: ***unmixing and de-noising***.

The step of de-nosing and unmixing of EEG was faced with much challenge. Previous research had concentrated on the technique of ICA rather than the application. It was based on short data segments, and to use the same techniques on long term continuous data brought problems. To consider the assumption of stationarity for ICA, the data had to be windowed, and analyzed in shorter segments. Non-overlapped data windows were used lest it should increase the data size many times. Applying ICA: temporal decorrelation or FastICA on consecutive windowed data segment posed the following challenges:

- Consecutive data windows can have different set of artifacts, or artifacts of varying

contribution. In such a case ICA can result in source signals that are quite different in the two consecutive windows. In the process it may sometimes overfit some source signals while split others in random forms. This prevents the tracking of source signals across consecutive windows. For example, an ocular artifact revealed in one window (as a result of an eye blink), may be missing in the consecutive window (when no eye blink occurs), causing the same sources to be split to follow the assumption of square mixing. This alienates the sources in consecutive segments, making it difficult for an objective system to track a source(s).

- ICA has a limitation of permutation i.e. the source signals recovered with ICA are not ordered, adding to the difficulty of source signals to be linked across consecutive windows.

Moreover, as the power of a 'seizure source' was expected to be low in segments prior to the seizure, the ICA could not be used with dimension reduction, lest the seizure source of interest gets removed (having a low eigen value). It could also easily get combined with other low power signals in case it had much less structure compared to other sources in the segment.

To overcome these issues, clustering of source signals from consecutive windows was considered but was not found useful. The aim had been to analyze the seizure source signals before a seizure and identify (if existing) some ***seizure predictive features***. The problem of not being able to track seizure sources across time was heightened by the nonexistence of knowledge about the static or dynamic morphology of a seizure source prior to an ictal segment. This analysis would have helped to track seizure sources even if they were moving.

Nevertheless, even if the seizure source signals had been extracted across time, use of some dynamic feature of the signals would be helpful to evaluate predictive features. Previous studies had considered parameters such as signal energy, amplitude variance, rhythmicity, complexity, phase synchrony or similarity, but these were based on single or two channel raw short segment EEG and lacked validation with interictal EEG. For example, a change in energy before or at seizure onset would not be significant till it is proven to be unique for seizure onset on long term EEG. In view of the parameters used, ***phase synchronization*** was considered motivating because apart from statistical formulations based on signals, phase synchronization had an ***abstract link to the context of seizure prediction*** as well. Phase synchronization had long been associated with epilepsy and it was being hypothesized by the medical and biomedical engineering research groups, that neuronal groups near a seizure focus undergo predictive alterations. The human brain is constantly creating and losing synchrony among neuronal groups, with limited energy. In order to amass energy as overpowering as that observed during a seizure, a focus neuronal group could cause neighboring neurons to

lose synchronization with neurons around them. Recruiting or synchronizing with this synchronously silent group, the focus could lead to a seizure with an overpowering synchrony, arresting the senses involved, and muting the rest of the ongoing synchronies of the brain. If such a dynamism existed, a seizure could be predicted and possibly prevented by observing the ***synchrony dynamics*** of a seizure focus. A moving seizure focus can complicate this theory further, therefore for the time being a static seizure focus was assumed. Synchrony dynamics was hence selected as the feature to be used. Phase synchronization was selected as an apt statistic for the task, as it was shown to be useful even for chaotic signals, possibly suited for EEG. Concerns that came with this selection were:

- The knowledge that use of phase synchronization on mixed signals could result in spurious synchrony and misleading results. The EEG being a mixture of signals could result in similar problems,
- Phase synchrony estimated for broadband signals was misleading as the statistic would not have a physical interpretation. EEG being a broadband signal would need to be filtered before being used.
- Phase synchrony was a bivariate statistic, it would involve subjective selection of EEG signals which could bias the results and not fully utilize the spatial and temporal information inherent in the multichannel EEG as a whole.

A solution was the merging of ICA and phase synchronization techniques. Although this was not novel as such, but it was not clearly established in the literature either. The approach remains novel in the context of epileptic seizure prediction. The combined use of ICA and phase synchronization was useful as they complement each other in following ways:

- ICA unmixes the EEG signals required by phase synchrony
- ICA utilizes the multichannel EEG to obtain a more informative source signals, that can now be used with phase synchronization (which now indirectly uses the multichannel EEG).
- Phase synchrony would help by detecting interesting patterns relevant for seizure prediction.

This was a subjective decision, *additional features might have been helpful?* However, the above scheme raised some important questions, discussed later in this section, making us concentrate on phase synchronization as the only feature to be explored.

The challenge of using ICA on long continuous windowed EEG was partially supported by the involvement of phase synchrony. ICA on windowed multichannel EEG was to

be used as spatially constrained ICA. The use of ***spatially constrained ICA*** allowed to constrain the search for seizure sources across time, with a spatial template. The use of spatially constrained ICA was encouraged by the knowledge that synchrony dynamics within particular areas (spatial) of the brain such as the seizure focus, might be informative and useful seizure prediction. Essentially, spatially constrained ICA was constraining the '***observer***' (view), instead of the '***observed***' (signals). Now, instead of following one or two sources of interest in space, we were concentrating on two locations in the brain and observing the synchrony of sources originating there, with time. This would allow to track changes in synchrony around the seizure focus with either the neighboring area (ipsilateral) or an area showing involvement but lying much further away (contralateral). The selection of the ipsilateral and contralateral areas were not based on the EEG but on a more objective criteria. The spatial template for each patient was searched with the help of an interictal-preictal segment of the patient. The pair of sources that were selected were those that showed maximal change in synchrony (observed to be seizure sources previously), or maximal involvement as a seizure source from the temporal and spatial information or the clinicians notes, annotations or source dependencies. The topography of these sources were then used as the ***spatial templates***. Not always, did ***ipsilateral sources*** show maximal involvement, some patients sources from ***contralateral*** locations were also found to be equally involved (such as patient T scalp). It should be noted that the spatial template was used as a ***soft constraint*** allowing slight spatial variations within the given template.

Moreover, spatially constrained ICA relaxed the assumption of sources to be completely independent in that spatial area. The source signals could be mutually dependent giving a higher chance of exploring synchronized yet unmixed sources.

ICA-Phase synchrony was initially used on short ictal segments to analyze its performance. ICA was applied to a segment of EEG having interictal and ictal periods (the total signal length being 20 seconds), and phase synchrony was analyzed among all pairs of source signals obtained. The first interesting results were obtained: ***The pair of ICs that showed a sudden change in synchrony at or slightly before seizure onset coincided spatially and temporally with source signals closest to the seizure focus.*** These results were then applied to a set of 4-5 ictal segments and 20 interictal segments of each of the ten patients. Each segment was two minutes in length. Similar promising results were obtained cross patients, irrespective of the time of the day from when the segments came from. This was used in ***objective identification or selection of seizure source signals from the set of sources obtained from the application of ICA*** (which is not a trivial task and is usually performed subjectively). The objectively selected seizure sources were used to de-noise the multichannel EEG by retaining only the seizure related information and removing artifacts and noise: ***'objective' de-noising of epileptic EEG***. The observation of

the change in synchrony at seizure onset was then tested on slightly longer data *lengths of two minutes* each in the context of ***seizure onset detection*** and similar results were obtained with a high sensitivity across the patients.

The segments were however, not continuous, with a gap of 15 minutes between segments (this was due to how the data was recorded at source). This did not allow the calculation of the specificity of such a seizure onset detection as it lacked a validation with the continuous ictal and interictal data. The promising observation was that significant phase synchrony during the interictal segments away from the seizure, remained at a stable low level (0.2-0.4) slightly increasing as the segments approached seizure onset, while at seizure onset it was seen to rise to 0.6 or 0.8 on a scale of zero to one. It supported the view of a possible seizure detector or predictor, but required long term continuous data (that was still being collected).

Although interesting results were being obtained, a crucial question had to be answered: ***If ICA is used to find independent signals, how can they be synchronized (dependent)? or vice versa?***. Experimentally ICA source signals were being found to be synchronized and there had to be an explanation for it. The answer to this dilemma was found with recent studies on ICA source dependencies. It had been studied that the source signals obtained from ICA can have remnant dependencies. Such remnant dependencies were seen to exist in our source signals obtained from EEG as well, which explains the possibility of synchrony. Phase synchrony captures a non-linear dependency between two signals and the remnant dependencies in ICs could possibly be a non-linear dependency that is seen as phase synchrony. However, not all source signals having remanent dependency show an equally consistent phase synchrony, supporting the fact that phase synchrony captures only a certain type of non-linear dependency. ICA in such cases tries to make the sources as independent as possible but is not able to remove these non-linear dependencies. The source signals obtained in this case have been referred to as ***Least Dependent Components***. This raises an important question: ***Perhaps ICA can be altered to seek source signals based on phase independence?***

Another important step was the filtering of the signals: broadband vs. narrowband analysis. It was understood that phase synchronization would be relevant only when used with narrowband signals, unlike previous studies in this area in the literature. Therefore it could not be used with EEG directly, nor with the LDCs, as FastICA did not assure any frequency independence.

The question was: ***should the filtering of the signals be performed before or after ICA?*** In both cases being before phase synchronization. The pros and cons of filtering before and after ICA were:

- The data-space would be reduced, the signals would become super-Gaussian as

the distribution would become peaky with narrow band filtering. This would effect the ICA source separation. With less structure to discriminate, the source separation would be less efficient.

- Filtering before performing ICA would increase the number of source signals to be handled and comparisons would be difficult because of the permutation problem. For example, if a 19 signal segment is filtered in two bands giving two sets of 19 signals, and ICA is applied to both the sets separately, it would result in two sets of sources in random order. They may even be slightly altered because of the difference in structure ICA finds in the different bandwidth data. This would make it challenging to compare the results of the two bands and increase the number of signals manifold. Alternatively, if ICA is performed on the segment giving 19 source signals, they can be filtered into many bands and compared for their performance with phase synchronization, in order to select the band of interest.

Therefore, the signals were ***first to be unmixed with ICA and then filtered*** into various narrow bandwidths.

The effect of using ***broadband vs. narrowband data for phase synchrony analysis*** was also performed with short term data (used above), in the context of seizure onset detection. Discrepancies were observed between results of broadband vs. narrowband data for phase synchrony analysis, as was expected. Phase synchrony analysis on broadband data showed an apparent averaging effect of the synchronies over different bands. They masked the patterns that were useful for seizure onset detection that used narrowband data. Broadband data was hence cautioned against for phase synchrony analysis. Although, it was kept in mind that too narrow a bandwidth could also insert spurious synchronies, the literature suggested a 4-6 Hz bandwidth to be a suitable preference (with the sampling rate of our data-set).

Phase synchrony or its index PLV (PLS with significance analysis) was hence found useful in ***measuring synchrony dynamics across time*** for epileptic EEG. This was finally ***applied to the long term continuous scalp EEG*** data that had been collected. It was fist applied in the form analyzed above: as a seizure detector. The aim was to observe the changes in synchrony across time in segments preceding the seizure onset in pre-defined spatial areas. The results were disappointing as even though the seizures were observed to being detected or predicted within a prediction window of few seconds, with a high sensitivity, the spatial areas were not entirely silent in terms of synchrony at other times. Artifacts such as chewing or sleep spindles that often shared the spatial areas of the brain with seizure focuses (such as temporal or frontal) also have a *significant* synchronous prevalence. A two minute synchrony silence as was seen before in other data, during interictal period was easily ***altered by the presence of encapsulating strong artifact synchrony*** (that was not seen in the previous

data). A spike, or short synchronous electrical discharge also showed up as a significant high synchrony, reducing the specificity of the seizure detection and prediction.

At this stage, a ***comparison of the PLV estimation with subjectively selected filtered EEG signals was performed with the PLV estimation of the filtered ICA source signals***. The comparison was *not entirely fair* as it is difficult to justify the contextual similarity of the ICs and the EEG signals being selected. Nevertheless, it illustrated the use of unmixing in removing spurious synchronies inherent in the EEG signals and in highlighting patterns of synchrony dynamics that remain hidden with the raw EEG (especially the ipsilateral signals).

The low specificity of detection and prediction found above enthused the concept of *smoothing* the PLV curves by low pass filters, allowing to ***bring forth the underlying gradual synchrony dynamics instead of sharp, short-term dynamics***. This would easily smooth out or ***reduce the effect of short lived artifacts*** such as electrode movement, chewing or even sleep spindles. The averaging was performed in a way that now the ***synchrony was being observed over hours instead of minutes and seconds***.

The PLV-d curves revealed interesting features: A cyclical pattern is observed showing dynamical synchrony of the brain. On further observation, the cyclical pattern appears to be in some sync with the seizure onset timings as if indicating a ***'set and reset'*** mechanism. The observation of a patients' (Chapter 7, section 7.2, patient T scalp) two days PLV curve showed a prominent two peak pattern leading on to a seizure with the possibility of the second peak to be an indicator of an impending seizure. The time differences from the second peak to the seizure onset for the patient also gave clues of a possible underlying pattern. The peak times showed an increasing prediction horizon till the sleep time, after which it appeared to refresh the next day and increased as before (patient T scalp: prediction horizons from day 1 afternoon: 11.4 min, 52.8 min, 73.8 min , 107.4 min (sleep time), day 2 morning: 6 min, 33 min).

At this stage, an important addition to the analysis was the ***innovative use of PLV-d curves obtained by the difference of two PLV curves of two frequency bands***. The filtered PLV of one frequency band (2-8 Hz) illustrates the synchrony dynamics occurring in that frequency band in a pre-defined space and similarly for the 8-14 Hz band. The difference of the PLV of the two frequency bands PLV-d can be viewed as a novel statistic that:

- ***Represses the apparently superfluous synchrony information in both the bands and emphasizes on the distinctive synchrony dynamics in both the bands***. It is vital to remember that the brain under normal conditions (and possibly non-seizure times) would be continuously involved in creation and loss of synchrony with local or distant neuronal groups, ***synchrony being ac-***

knowledged as an important fragment of information retention and exchange in the brain.

- It seeks the instances when the synchrony in one of the frequency band becomes more powerful while the other gets muted. This may be interpreted as the ***emergence of an oscillator or force with a frequency that overpowers and forces other frequencies to be synchronized*** in the same band with it,
- It will ***ignore instances when the synchrony in both bands is equally strong or equally low***, both suggesting periods of normal brain involvement ***rather than an emerging strong entraining force, plausible of a seizure focus/entrainer***.

The resulting ***PLV-d curves per day of each patient formed the basis of the analysis***. Visual analysis of the PLV-d curves revealed some interesting features:

- A ***cyclical pattern*** is observed with peaks and troughs showing a gradual dynamical synchrony prevalent in the brain at all times. ***If this is typical of epileptic EEG or not is still yet to be answered.***
- A prevailing underlying rhythm coinciding with the sleep and waking times: aptly considered perhaps as a ***Circadian-like Rhythm***. ***If this is unique in epileptic EEG or altered compared to normal EEG is now a new research question.***
- The 'Circadian-like Rhythm' appeared to ***demonstrate the conventional synchrony dynamics*** for the patient.
- A ***distinctive peak (or trough) near seizure onsets that appeared to drag the curve against the conventional synchrony dynamics*** towards the zero level of the PLV-d curve. A zero level of the PLV-d curve being the level when the synchronies of both the bands are equally strong or equally nil.
- ***Distinctive peaks or troughs in the PLV-d curves lagging seizure onsets:*** considered as predictive features. They were at times distinctive from other peaks on the same curve in their variance, latency, gradient, preceding peaks, etc. item A ***switch in dominant frequency at onset of drowsiness is observed to coincide with seizure onsets*** many times. Such observations of seizure onsets coinciding with onset of sleep or waking time has separately been made by the medical field as well.

Stretching the line of thought: the possible 'Circadian-like Rhythm' can be interpreted as a change in the dominant frequencies across the day:

- either, ***vigilance or mental tasks keep the brain activity of a certain frequency band dominating for most of the day and the dominant frequency switches with the onset of sleep***,
- or ***one frequency band dominates during the day allowing mental activity and the internal clock switches the dominant frequency at a likely or learned time to impose a resting time for the brain, causing drowsiness*** and most likely reduced vigilance (due to the activation of the other frequency band). ***However, these are simply hypothesis' and require further study research.***

These observations were very promising in the context of seizure prediction and epileptic EEG analysis in general as well. The capability of this analysis for seizure prediction was yet to be assured. The analysis of seizure prediction was yet subjective as it was difficult to generalize the predictive features or demonstrate patterns across patients only with visual analysis. This required a formal and objective feature extraction process.

The ***discriminant analysis for feature extraction*** was the next major step. The observations of the PLV-d curves were used to propose discriminating features (called ***PLV features***). Additional features based on the underlying source signals were also proposed (called ***IC features***), and a combination of the IC and PLV features. As the analysis of all combinations of these 23 features was a time consuming and computationally expensive process, a smaller number of feature subsets were created to be analyzed. The feature subsets were used to construct multidimensional (2-13 dimensions) feature vectors using PLV-d curve segments corresponding to one hour each. The segments were ***labelled to be predictive or interictal*** according to their nearness in time to a seizure onset. The features were formulated with general observations and it can be questioned if these features and feature subsets were sufficient and necessary, if they captured all the aspects appropriately and if the subsets were suitably composed. A trial and error was the plausible path, though, guided by the experience, work of related studies on feature extraction, and general understanding and expectations from the analysis. The analysis of feature extraction was restricted to five patient data sets with sufficient number of seizures each, with their contralateral and ipsilateral PLV-d curves each. Each data set was divided into a ***training set*** and a ***test set***. The training set was made up of the data until the first few seizures were included. The remaining data was used as the test set. Sometimes a cluster of closely spaced seizures made it difficult to make a train-test boundary. And sometimes the seizures being in the very beginning of the data forced the use of a small training set. In addition, each patient had very few seizures, as few as one or three, separating them as training and test data would reduce the information required for training. Nevertheless, in order to acquire relevant features, the data window as large as an hour had to be taken on

the smoothed PLV-d curve. Therefore, *a 75% overlap was used* instead to increase the data points as well as be able to track dynamics across shorter time span of 15 minutes. Despite these challenges, they were all found useful to highlight the exitance of predictive features in the PLV-d curves. *Four patient data sets were held back*to be used for the final testing with the selected features.

The ***multidimensional feature vectors were reduced to 2-D with a Neuroscale topographic technique*** based on a radial basis neural network. Along with dimensional reduction the Neuroscale had an added advantage of ***learning the transformation*** parameters allowing the projection of untrained data as well. It was used in an unsupervised mode as well as with a varied strengths of supervision. The ***supervision*** was performed by providing class information in the form of a distance metric. The resulting clusters obtained with various Neuroscale training instances were compared and the supervised training showed much distinct clusters amongst them. Also, the clusters obtained with various feature subsets were compared, and the feature space obtained by the subsets with IC features alone or IC and PLV features were found to be much overlapped, and hence were discarded. The PLV only feature subsets one and six were observed to be performing the most distinction between classes, and were selected for further analysis. It should be *noted that clean distinction of pre-ictal and interictal feature points was not expected. This was because the pre-ictal data windows had been defined as the windows that were about one hour prior to the seizure onset.* It was quite possible that the data reflected features similar to interictal features when further away from the seizure. A ***trajectory of the pre-ictal points in time*** would have been informative helping to illustrate a transition from the interictal to the pre-ictal state. This observation was reinstated when the prediction horizon of 35 minutes was later found to be most suited in most of the patient data sets.

The parameters such as number of hidden nodes, training iterations, learning rate, etc. can affect the training of the Neuroscale network. They were varied to analyze their affect on the resulting map. The hidden nodes were varied from 10 to 30 nodes and it was found that too few < 10 and too large > 30 were not having much impact on the results. The number of hidden nodes that were found suitable for each patient were set at about one-sixth of the training data points and more than three times the input dimension. The test data was then projected on to the learned feature space using the network that had been trained for each patient.

Next, the feature space obtained by Neuroscale projection had to be segmented, in order to enable the classification of the pre-ictal and interictal feature points. A semi-parametric estimation method, ***Gaussian Mixture Modelling (GMM)***, was used for this task. The GMM estimates a density model where the number of adaptive parameters can be increased in a systematic way (by adding more components to the model), making the model arbitrarily flexible. GMM is based on modelling the

unconditional probability density, given a finite set of unlabelled data points. the probability density function is expressed as a linear combination of basis functions. The objective of the *GMM Training* is the estimation of the component parameters $\theta = (\mu_j, \sum_j, P(j))$ of the multiple components of the Gaussian Mixtures.

The ***GMM was first used in a supervised manner***, making use of the class information. Two GMMs were separately modelled for each of the pre-ictal and the interictal class. The GMMs were separately trained using training data to estimate their respective parameters $\theta = \{\mu_j, \sum_j, P_j\}$. This trained set of GMM was used to obtain the density models for the pre-ictal and interictal classes. It hence finds a decision boundary between the classes using Bayes' Theorem. It was then applied to test data as a classifier to ascertain the probability of a given feature point to exist in a pre-ictal GMM density map. This gave a two dimensional ***curve of probability of pre-seizure event vs. time***. On this curve a prediction event was said to have occurred at a time point when the curve exceeded an arbitrary threshold within a prediction horizon before the seizure onset.

The next task was to determine the prediction horizon and the threshold level most suited for such a predictor. Therefore, a ***statistical analysis*** was performed on these probability curves for both the feature subsets: one and six. It included the calculation of the ***sensitivity*** of prediction, ***false positives per day***, ***specificity*** and finally the ***ROC curves*** that illustrated a function of the sensitivity and specificity. These were calculated ***for various prediction windows varying from 0 to four hours*** in steps of five minutes and ***varying threshold level*** (zero to one). It was found that Neuroscale-GMM demonstrated the most promising classification of pre-ictal features based on the ***feature subset 6***, for seizure prediction. ***The prediction windows of the the order of 35 minutes were found to show on average 86% sensitivity and 80% specificity***.

Interestingly, a very low or negligible probability is also observed for a short prediction window of 10-15 minutes for many patients supporting the hypothesis about *a gradual build up of synchrony over a long time with a loss of synchrony before the seizure* . The loss of synchrony before seizure onset had been explored in previous studies but as was observed in this study, the loss of synchrony fails to show specificity as many other events also lead to a loss of synchrony during the day. Therefore, as this study shows, it would be more useful to consider a more elaborate gradual dynamic of synchrony.

The classification was then ***tested against a random predictor*** as well. A set of random descriptors were used to define prediction times and the sensitivity of them being true predictors (lying within a prediction window) was determined for various prediction windows. ***The ICA-PLVd-Neuroscale-GMM predictor demonstrated an ability to predict that was better than a random predictor***.

The ***GMM was also used in a semi-supervised manner***. A semi-supervised GMM uses labelled and unlabelled data for training. A semi-supervised learning is essentially an unsupervised learning guided for the best performance by the use of additional information from the labelled data to assist in its unsupervised learning. The unlabelled data provides the classifier with a more general information about the underlying data structure, while the sparse labelled data may guide it to select the most appropriate distributions for defining each class. The main difference between the implementation of the supervised and semi-supervised GMM was that in semi-supervised case the data from the pre-ictal class in the training data was used as the labelled data and the entire test data was used as the unlabelled data. The probability curve was then obtained by analyzing the probability of a given point to exist in the pre-ictal density map. As the labels of the entire data were known in this case, the classification accuracy could be checked by observing if the pre-ictal feature points did lie in the pre-ictal density map. The most interesting observation with the semi-supervised GMM was that the data set on which the supervised classifier was failing to be trained aptly, the semi-supervised GMM showed excellent classification and prediction performance (patient E scalp in Chapter 8 shows a reduction of prediction horizon from 155 minutes to 35 minutes with increase sensitivity and specificity with semi-supervised GMM). ***This shows that a semi-supervised GMM is more intuitive in capturing underlying structure in the feature space with a sparse labelled data compared to a fully labelled dataset***.

Usually, a *semi-supervised GMM is expected to perform even better than a supervised learning*, but it was found to be marginally different in the present study. This can be explained with the subtle alterations in the use of both the schemes. The train/test bounds for both the schemes were not exactly the same, nor was the same feature map used in both the cases. The feature map was independently trained in both cases before the GMM was applied. Thus it would *not be fair to compare the supervised and semi-supervised GMM results*. ***Nonetheless, the semi-supervised GMM validated the GMM classification performance and supported the existence of predictive features in the synchrony dynamics***.

The extent of the probability that identifying an event as a predictor, is also analyzed for various prediction windows. It was observed that generally the probability of prediction was higher much before the seizure onset and it gradually reduced as the seizure approached. ***This indicates significant changes taking place in the brain long before the seizure onsets that have a higher predictive capability, compared to the immediate changes*** that occur before a seizure onsets. Such a trend could be useful in raising an initial *soft* warning and then a *stronger* alarm in a seizure prediction application. A combination of intervention therapies aimed to prevent a seizure occurrence might be based on such a predictor, such as, injecting long acting drugs at the soft warning and then triggering electrical stimulation if the

warning is observed to be succeeded by the strong alarm. ***However, the resulting complexity that comes with a forced intervention on the synchronization patterns of the inbrain activity would first need to be studied closely.***

The ***algorithm now fully developed: cICA-PLV-d-supervised Neuroscale-GMM was then applied on the set of four patient data sets*** that had not been involved in the discriminant analysis. This final test would demonstrate the ability of the selected feature set as a predictor in unseen data. The performance of the four test patients was found to be ***better than a random predictor***. ***The sensitivity of the order of 65%-100% and specificity of about 65%-80%, for a prediction window of 35-65 minutes*** was achieved with the semi-supervised GMM technique. The specificity for the same was

Thus the overall performance of the algorithm cICA-PLV-d-supervised Neuroscale-GMM has been very encouraging. It has successfully demonstrated the existence of a predictive space based on synchrony dynamics, more suitably acquired through unmixed multichannel EEG signals. A prediction performance with training and test data was found to show a specificity of 60%-100% and 65-100 % sensitivity with a prediction window of 35-65 minutes in general. It was also found to perform much better than a random predictor.

9.1 Future Work

This study has raised a few questions that motivates further research.

Phase based independence: The detection of phase synchrony in independent signals is quite an interesting topic for further research. A phase based unmixing needs to be explored, although it might not be useful in the context of finding synchronized signals. An algorithm similar to ICA based on phase independence and multichannel data would be an interesting approach.

Frequency based ICA: The use of frequency based ICA in place of filtering might be an interesting step in a similar analysis as was carried out in this research. It will pose challenges by increasing the number of signals enormously, and perhaps in bringing together the independent parts of the narrow band seizure sources.

Moving sources: Tracking moving seizure sources would be an important next step for this field of seizure prediction based on phase synchrony.

ICA with frequency constraint: Using ICA with a frequency constraint will remove the need to filter signals. Perhaps a spatial-frequency constraint might be developed in such a case.

ICA with phase constraint: Using ICA with a phase constraint is another interesting possibility to improve such a seizure prediction analysis.

Circadian-like Rhythm: Further analysis of the possible Circadian-like Rhythm uncovered here would be most interesting. It might unveil other informative phenomenon connected to the ongoing synchronies in the normal or pathological human brain.

Synchrony in normal vs. epileptic EEG: A comparison of synchrony dynamics and especially Circadian-like rhythms in the normal long term EEG and epileptic EEG might elicit information useful in diagnosing pathological brain or brain activity.

Removing the effect of artifacts in the PLV-d curves: The cyclical pattern of the synchrony as was observed in the PLV and PLV-d curves might perhaps be more useful if the challenge of objectively identifying and removing selective artifacts, especially those with enhanced synchrony like sleep spindles, is overcome. If they are responsible for the undulations and cyclical fluctuations, they might make the PLV-d curves less undulating and much cleaner. It might help to identify patterns associated with prediction more effectively, perhaps removing the need of an expensive and complicated discriminant analysis.

Source Localization: An advancement to the present analysis would be the use of source localization techniques to track the synchrony dynamics more closely within the cortex. Beamforming or other localization techniques such as dipole modelling might be suitable for such an analysis.

Apart from these advancements to the current research, the following are reflections on a few questions that intrigued me during the course of the research and might motivate further research in this area:

Informative relation between spikes or electrical discharge and seizure: This research has shown possibilities of existing underlying patterns in the brain that might be useful for seizure prediction. The synchronous activity typical to epileptic EEG such as spikes and short electrical discharges are known to us today. But the cause of such activity and the relation if any of such activity with approaching seizures is not yet understood and are thought to be random phenomenon in the epileptic brain. There is a possibility, as this research shows, that synchrony dynamics (perhaps such as these) might in fact be informative about an approaching seizure, for example, perhaps as the neuronal synchrony increases a protective mechanism in the brain causes a spike in order to break the strengthening synchrony by suddenly bringing in a stronger frequency to entrain all those neurons. Sometimes a few spikes are successful in suppressing such abnormally

increasing synchrony and a seizure is prevented. While at other times a train of spikes occurs, and when the spikes are not able to suppress the increasing synchrony effectively, it rises back again in a short time, leading to a seizure. An analysis of spike activity with the synchrony dynamics shown here would be quite interesting for this field. It might help to advance the understanding of abnormal electrical discharge in the otherwise normal functioning human brain.

Frequency Entrainment: In light of the present research I also came across the information about how stroboscopic lighting is used to invoke seizures or epileptiform activity, typically in a short EEG test for epilepsy diagnosis. The stroboscopic lighting test involves exposing the patient to strong light at various pre-defined frequencies. Different patients epileptiform discharge is seen to be evoked by different frequencies or by none at all. The present research and this observation has raised a question, if perhaps the stroboscopic lighting plays the role of an entrainer that tries to seek the frequencies similar to the seizure focus of the patient perhaps matching with the frequency at which the focus is uniquely tuned to become active causing seizure like activity. Perhaps knowledge of the frequency that causes the focus to become active might be more useful prior knowledge for a seizure prediction algorithm based on narrow band synchrony. As it might define the band that might show the most informative activity for that patient.

In all, this has been an interesting journey across the confluence of two domains of neurophysiology and advanced signal processing. Studying a neurophysiology challenge with an abstract view has surfaced more questions and a curiosity for deeper understanding of the problem. Although, it will still be many years and much research in related areas, before a viable clinical predictor is acceptable for this complex disorder, however, every inch forward brings us closer to that hope..Disha Gupta.

Appendix A

EEG Data Collection

The continuous EEG recordings have been collected from the Southampton General Hospital (SGH), Wessex Neurological Center, Southampton, UK. The data collection has been granted the approval by the Local Research Ethics Committee (Ethical Approval Number: 06/Q1701/132) and the ISVR ethics committee (ISVR Ethics approval number:817). The study has been registered with the R&D Southampton University as Project Number: RHMNEU0098.

The data collected is multi-channel, continuous scalp EEG of epileptic patients being monitored in preparation of possible surgery (pre-surgical evaluation). It is anonymised with respect to the details of the patient and only the technical information is retained for analysis. It is stored and used as per the data protection guidelines and restrictions on use of patient data. No video recordings are taken or stored for research analysis.

A.0.1 EEG Recording System

The EEG is acquired at the hospital using XLTEK products (EEG 32 and EMU 128) and are read by Neuroworks EEG software [175].

Technical specifications of the system are given in table

Channels	32
Sampling Frequency	200 Hz
Bandwidth	0.1-100 Hz
Voltage resolution	Accurate to 0.3 micro volts
Analog/digital converter	22 bits
Reference channels	FCz

Table A.1: Technical specifications of EEG32

Neuroworks is the software used to display, the multichannel biomedical signals. It is a software that allows to visualize the EEG signal in different montages and has in

Figure A.1: XLTEK apparatus to acquire EEG

Channels	128
Sampling Frequency	500 Hz
Bandwidth	0.1-100 Hz
Voltage resolution	Accurate to 0.3 micro volts
Analog/digital converter	22 bits
Reference channels	Varies: contra lateral to the location of the grid, strip or depth electrodes

Table A.2: Technical specifications of EMU128

built filters (high, low, bandpass, notch) and allows visualizing the EEG at different timescales with variable scales. It also synchronizes the patient video with the EEG signals which is helpful in replay analysis and diagnosis. Neuroworks is used to convert the data into EDF format for collection using an in-built function. Neuroworks also displays the annotations on the EEG and any other event markers which have been collected as well along with the data.

A.0.2 EDF Format

The data is exported for analysis to University in the form of European Data Format (EDF) files. The EDF format is a standardized European data format which is used for exchange and storage of biological signals (single or multi-channel) [196, 197]. It is a non-proprietary format that allows archiving, distribution and exchange of data from commercial devices independent of the acquisition system. EDF was published in 1992 in Electroencephalography and Clinical Neurophysiology 82, (pages 391-393). Since then, EDF became the de-facto standard for EEG recordings in commercial equipment and multi-center research projects. This format stores the data information in standard header files and stores the data as continuous fixed duration epochs as a *.edf* or *.rec* file. The header contains technical information about the data like the

sampling frequency, length of data, number of channels, filtering, etc. as well as some general information such as start time, date, etc. The EDF data can then be converted to ASCII format using the header information embedded in the EDF files as per EDF file specifications.

The initial 256 bytes of the header contains information about the record, specifying the version of the format used, general information about the record, identification number, time of start, number of records and the number of signals (ns). The next 256 bytes per signal, specify the type of signal (e.g. EEG), calibration and the number of samples in the record.

Header $= (256 + numberofsignals \times 256)$ bytes

The header is followed by each record starting with the duration of the signal (in seconds). Gains, electrode montages and filters are assumed to have been kept fixed for the entire recording.

The digital format of the header and data record is shown in Tables A.0.2 and A.0.2 as specified by Kemp *et al.* in [196]:

Number of ASCII characters used in Header	**Specification**
8	Version of the data format
80	Local patient identification
80	Local recording identification
8	Start date of recording (dd.mm.yy)
8	Start time of the recording (hh.mm.ss)
8	Number of bytes in header record
44	Reserved
8	Number of data records (-1 if unknown)
8	Duration of a data record (seconds)
4	Number of signals (ns) in data record
ns * 16	ns * label (Channel name)
ns * 80	ns * transducer type
ns * 8	ns * physical dimension
ns * 8	ns * physical minimum
ns * 8	ns * physical maximum
ns * 8	ns * digital minimum
ns * 8	ns * digital maximum
ns * 80	ns * prefiltering
ns * 8	ns * number of samples in each data record
ns * 32	ns * reserved

Table A.3: Digital specification of header in EDF data formatting

Number of ASCII characters used in data record	**Specification**
Number of samples in first signal * integer	First signal in the data record
Number of samples in second signal	Second signal
.	.
.	.
Number of samples in last signal	Last signal

Table A.4: Specification of record data encoding in EDF format.

A.0.3 Calibration

Calibration of the ASCII data is the next important step which is required to convert the integer value stored in the EDF files to the actual physical units of the recorded signal. The EDF header includes information about the calibration in the form of four values: physical minimum, physical maximum, digital minimum and digital maximum. Using these four values it is possible to retrieve the signal in the physical units of micro volts using:

$$Signal_{Physical} = \frac{(ASCIIvalue - Digi_{min}) * (Phys_{max} - Phys_{min})}{(Digi_{max} - Digi_{min})} + Phys_{min}$$

where $Digi$ refers to the digital values of the stored signal and $Phys$ corresponds to the value of the physical (originally recorded) signal.

Appendix B

Publications

The following publications arise from the work undertaken in this thesis:

D. Gupta, C. J. James and W. Gray, *Phase Synchronization with ICA for Epileptic Seizure Onset Prediction in the Long Term EEG*, IET 4th International Conference on Advances in Medical and Signal Processing MEDSIP, Italy, (2008).

D.Gupta and C. J. James, *Narrowband vs. Broadband Phase Synchronization Analysis Applied to Independent Components of Ictal and Interictal EEG*, IEEE-EMBS (Engineering in Medicine and Biology Society) 29th International conference in Lyon, France (Aug 2007). (Finalist for Student Paper Competition)

C. J. James, D. Absolo, D. Gupta, *Space-Time ICA versus Ensemble ICA for Ictal EEG Analysis with Component Differentiation Via Lempel-Ziv Complexity*, Proc. of IEEE-EMBS (Engineering in Medicine and Biology Society) 29th International conference in Lyon, France (2007).

D. Gupta, C. J. James and W. Gray, *De-noising Epileptic EEG using ICA and Phase synchrony*, Proceedings of IEE-MEDSIP2006 - The 3rd International Conference on Advances in Medical , Signal and Information Processing (July 2006). (Awarded William James Memorial Biomedical Engineering Student Award)

M. Pal, D. Gupta, M.G. Edwards and C.J. James, *Flux-continuous schemes for solving EEG source localization problems*, ACME 07 Conference Proceeding, Scotland (2nd - 3rd April 2007).

D.Gupta, C.J.James and W.Gray,*Seizure Onset Prediction through EEG using ICA and Phase Synchrony*, Conference of International League Against Epilepsy (ILAE UK Chapter), Newcastle upon-Tyne, UK (Sept. 2006).

D.Gupta, C.J.James and W.Gray, *De-noising Epileptic EEG using ICA and Phase Synchrony*, (Poster) Faculty of Medicine, Health and Life Sciences Postgraduate Conference 2006, University of Southampton (June 2006).

D. Gupta, C. J. James and W. P. Gray, *Seizure onset prediction in epilepsy through EEG using ICA*, SET for Britain 2006. Early-Stage Research Engineers at the House of Commons, (Parliament) London, UK (12 Dec 2006).

D.Gupta, C.J James and W.Gray, *Seizure Onset Prediction in Epilepsy* (Poster). 2nd Life Sciences Interface Conference, Southampton, UK (2005).

Appendix C

Graphical User Interface for Seizure Onset Detection

A graphical user interface (GUI) was developed for the analysis of seizure onset detection. The GUI allows the loading of contralateral and ipsilateral phase synchrony time series for any patient (one day at a time) that the user invokes. It allows selection of the EEG PLV time series or the LDC PLV time series as well. After the data loading is completed (shown by a statement 'Loading Complete'), the filter and threshold levels can be varied to assess the sensitivity and specificity of seizure detection. The plot is refreshed on resetting and the sensitivity and specificity is displayed as a percentage.

Sensitivity has been calculated as :

$$Sensitivity = \frac{Number\ of\ seizures\ detected}{Total\ seizures} \times 100\%$$

$$Specificity = \frac{Number\ of\ correct\ detections}{Total\ detections} \times 100\%$$

A clip of the gui has been shown in Figure C.1.

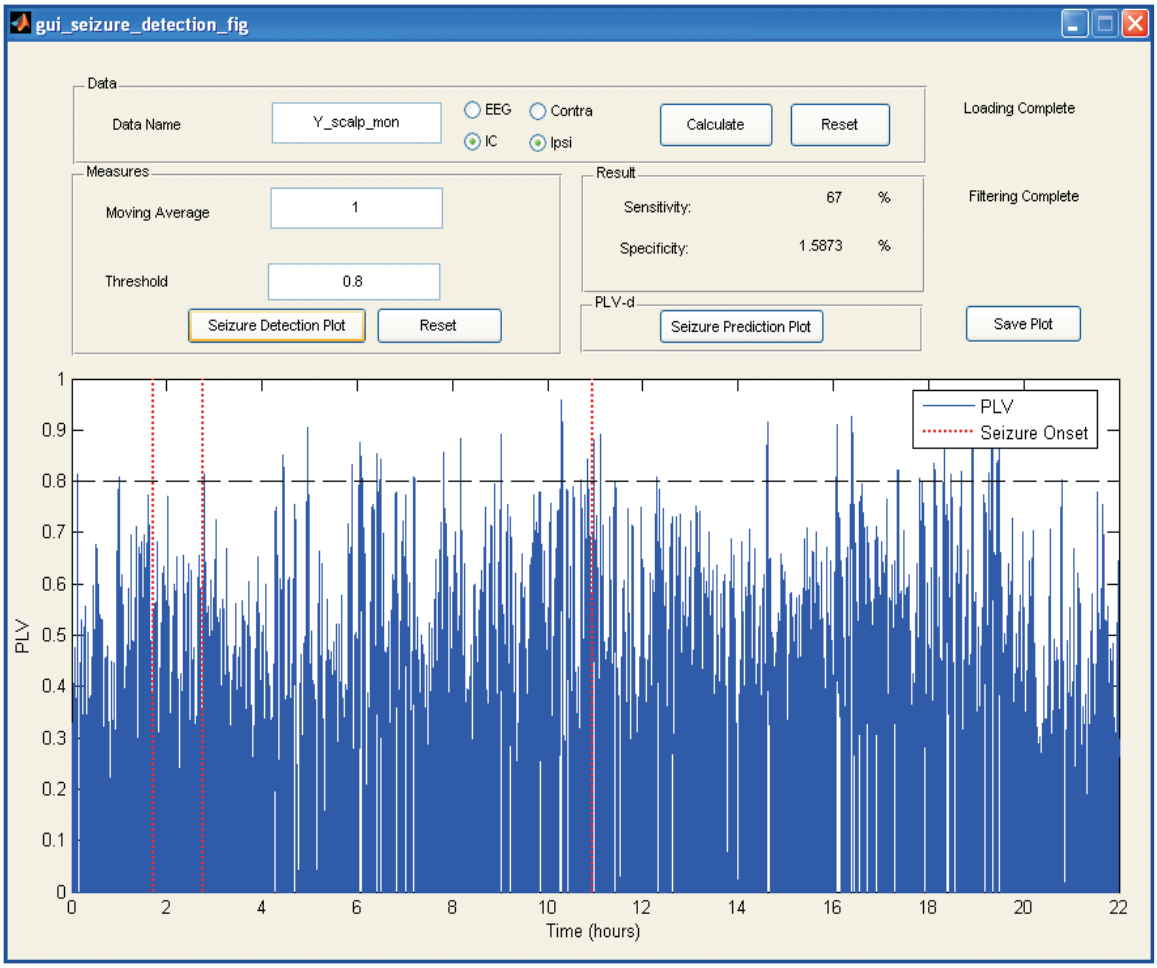


Figure C.1: Graphic User Interface developed for seizure detection. The inputs for this interface are the patient data name and the signals to be used, moving average and threshold. It estimates and plots the PLV (checked for significance) of the signals across time, and the sensitivity and specificity of seizure detection for the selected threshold.

References

[1] H Gray. *Gray's Anatomy of the Human Body*. Philadelphia: Lea & Febiger, Bartleby.com (image copies from the 20th edition of Gray's Anatomy and are not copyrightable as per Bridgeman Art Library v. Corel Corp, lapsed into public domain), 20th u.s. edition edition, 1918, 2000.

[2] Col Jeananda. Enchanted learning., 2009.

[3] J Malmivuo and R Plonsey. *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford University Press, Oxford, New York, 1995.

[4] H H Jasper. Report of the committee on methods of clinical examination in electroencephalography. *Electroenceph. Clin. Neurophysiol.*, 10:370–371, 1958.

[5] R Cooper, J W Osselton, and J C Shaw. *EEG Technology,*. Butterworths, London (2nd ed.), 1969.

[6] F Sharbrough, G-E Chatrian, RP Lesser, H Lders, M Nuwer, and TW Picton TW. American electroencephalographic society guidelines for standard electrode position nomenclature. *J. Clin. Neurophysiol*, 8:200–202, 1991.

[7] Arthur DiPatri and Tord Alden. Surgical treatment of epilepsy in children. The Child's Doctor, Spring 2005.

[8] M Le Van Quyen. Anticipating epileptic seizures: from mathematics to clinical applications. *C R Biol.*, 328(2):187–98, 2005.

[9] C J James and C W Hesse. Independent component analysis for biomedical signals. *Physiological Measurement*, 26:R15–R39, 2005.

[10] C J James and O J Gibson. Temporally constrained ica: an application to artifact rejection in electromagnetic brain signal analysis. *IEEE Trans Biomed Eng.*, 50 (9):1108–16, 2003.

[11] A Pikovsky, M Rosenblum, and J Kurths. *Synchronization : A Universal Concept in Nonlinear Sciences*. Cambridge University Press, 2001.

[12] G V Tcheslavski. *Coherence and Phase Synchrony Analysis of Electroencephalogram*. PhD thesis, Virginia Polytechnic Institute and State University, 2005.

[13] D Lowe. Feed-forward neural networks and topographic mappings for exploratory data analysis. *Neural Computing and Applications*, 4:83–95, 1996.

[14] D Durand. Electrical stimulation can inhibit synchronized neuronal activity. *Brain Research*, 382:139–144, 1986.

[15] D Durand and E Warman. Desynchronization of epileptiform activity by extracellular current pulses in rat hippocampal slices. *Journal of Physiology (1994), 480.3!*, 480(3):527–537, 1994.

[16] M Morrell. Brain stimulation for epilepsy: can scheduled or responsive neurostimulation stop seizures? *Current Opinion in Neurology.*, 19(2):164–168, 2006.

[17] W H Theodore and R S Fisher. Brain stimulation for epilepsy. *The Lancet Neurology*, 3, Issue 2:111–118, 2004.

[18] S Rothman and X-F Yang. Local cooling: A therapy for intractable neocortical epilepsy. *Epilepsy Currents*, 3(5):153–156, 2003.

[19] M W Hill, M Wong, A Amarakone, and S M Rothman. Rapid cooling aborts seizure-like activity in rodent hippocampal-entorhinal slices. *Epilepsia*, 41(10):1241–1248, 2007.

[20] D Smith, S Krahl, R Browning, and E Barea. Rapid cessation of focally induced generalized seizures in rats through microinfusion of lidocaine hydrochloride into the focus. *Epilepsia*, 34(1):43–53, 1993.

[21] A Depaulis, O C Snead, C Marescaux, and M Vergnes. Suppressive effects of intranigral injection of muscimol in three models of generalized non-convulsive epilepsy induced by chemical agents. *Brain Research*, 498:64–72, 1989.

[22] B Litt, M D'Alessandro, R Esteller, J Echauz, and G Vachtsevanos. Translating seizure detection, prediction and brain stimulation into implantable devices for epilepsy. In *First International IEEE EMBS Conference on Neural Engineering, Italy*, 2003.

[23] H Stgbauer, R G Andrzejak, A Kraskov, and P Grassberger. Reliability of ica estimates with mutual information. *ICA 2004 : international conference on independent component analysis and blind signal separation*, 3195:209–216, 2004.

[24] H Stgbauer, A Kraskov, S A Astakhov, and P Grassberger. Least dependent component analysis based on mutual information. *Phys. Rev. E*, 70, 066123, 2004.

[25] F Meinecke, A Ziehe, M Kawanabe, and K-R Muller. A resampling approach to estimate te stability of one dimensioanl or multidimensional independent components. *IEEE Transactions of Biomedical Engineering*, 49(12):1514–1525, 2002.

[26] WHO. World health organization. http://www.who.int/en/.

[27] E Niedermeyer and F L Da Silva. *Electroencephalo graphy: Basic Principles, Clinical Applications, and Related Fields*. Baltimore:Williams Wilkins 4th Ed, 1993.

[28] *International League Against Epilepsy , online access: http://www.epilepsy.org/ (updated 2006)*.

[29] *Epilepsy foundation, online access at: http://www.epilepsyfoundation.org*.

[30] *Vagus Nerve Stimulation Therapy from Cyberonics, Inc. (2005), online access at http://www.vnstherapy.com/*.

[31] *Cyberonics Inc. A commoercial product for VNS, online access at http://www.cyberonics.com/*.

[32] *International League Against Epilepsy, surgery; online access : http://www.epilepsy.org.uk/info/surgery.html*.

[33] S Baillet, J C Mosher, and R M Leahy. Electromagnetic brain mapping. *Signal Processing Magazine, IEEE*, 18(6):14–30, 2001.

[34] M Teplan. Fundamentals of eeg measurement. *Measurement Science Review*, 2 (section 2), 2002.

[35] S V Notley. *Prediction of Epileptic seizures from depth EEG signals*. PhD thesis, University of Southampton, UK, 2001.

[36] J Gotman, P Gloor, and J R Ives. Long-term monitoring in epilepsy. *Electroenceph. Clin. Neurophysiol.*, Suppl. 37:341–55.

[37] C J James. *Detection of epileptiform activity in the electroencephalogram using artificial neural networks*. PhD thesis, University of Canterbury, Christchurch, New Zealand,, 1997.

[38] Raymond Cooper, Colin Binnie, and Richard Billings, editors. *Techniques in Clinical Neurophysiology: A Practice Manual*. Elsevier, 2005.

[39] J P Lachaux, E Rodriguez, J Martinerie, and F J Varela. Measuring phase synchrony in brain signals. *Human Brain Mapping*, 8(4):194–208, 1999.

[40] P Rajna, B Clemens, E Csibr, E Dobos, A Geregely, M Gottschal, I Gyrgy, A Horvth, F Horvth, L Mezfi, I Velkey, J Veres, and E Wagner. Hungarian

multicentre epidemiologic study of the warning and initial symptoms (prodrome, aura) of epileptic seizures. *Seizure*, 6(5):361–368, 1997.

[41] P Federico, D F Abbott, R S Briellmann, A S Harvey, and G D Jackson. Functional mri of the pre-ictal state. *Brain*, 128 (8):1811–1817, 2005.

[42] C Baumgartner, W Serles, F Leutmezer, E Pataraia, S Aull, T Czech, U Pietrzyk, A Relic, and I Podreka. Preictal spect in temporal lobe epilepsy : Regional cerebral blood flow is increased prior to electroencephalography-seizure onset. *The Journal of nuclear medicine*, 39(6):978–982, 1998.

[43] M E Weinand, L P Carter, W F el Saadany, P J Sioutos, D M Labiner, and K J Oommen. Cerebral blood flow and temporal lobe epileptogenicity. *J Neurosur.*, 86(2):226–232, 1997.

[44] V Novak, A L Reeves, P Novak, P A Low, and F W Sharbrough. Time-frequency mapping of r-r interval during complex partial seizures of temporal lobe origin. *J Auton Nerv Syst.*, 77(2-3):195–202, 1999.

[45] R S Delamont, P O O Julu, and G A Jamal. Changes in a measure of cardiac vagal activity before and after epileptic seizures. *Epilepsy Research*, 35 (2):87–94, 1999.

[46] S Viglione and G Walsh. Epileptic seizure prediction. *Electroencephalogr. Clin. Neurophysiol.*, 39:435436., 1975.

[47] L Iasemidis and J Sackellares. The evolution with time of the spatial distribution of the largest lyapunov exponent on the human epileptic cortex. *Measuring Chaos in the Human Brain*, pages 49–82, 1991.

[48] L D Iasemidis and J C Sackellares. Chaos theory and epilepsy. *The Neuroscientist*, 2:118–126, 1996.

[49] Y-C Lai, M A F Harrison, M G Frei, and I Osorio. Inability of lyapunov exponents to predict epileptic seizures. *Phys. Rev. Lett.*, 91 (6):068102, 2003.

[50] Y-C Lai, M A F Harrison, M G Frei, and I Osorio. Controlled test for predictive power of lyapunov exponents: their inability to predict epileptic seizures. *Chaos*, 14:630–642, 2004.

[51] R Aschenbrenner-Scheibe, T Maiwald, M Winterhalder, H U Voss, J Timmer, and A Schulze-Bonhage. How well can epileptic seizures be predicted? an evaluation of a nonlinear method. *Brain*, 126:2616–2626, 2003.

[52] H Kantz and T Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, 2002.

[53] L D Iasemidis, Deng-Shan Shiau, P M Pardalos, W Chaovalitwongse, K Narayanan, A Prasad, K Tsakalis, P Carney, and J C Sackellares. Long-term prospective on-line real-time seizure prediction. *J. Clin Neurophysiol.*, 116:532–544, 2005.

[54] K Lehnertz and C E Elger. Spatio-temporal dynamics of the primary epileptogenic area in temporal lobe epilepsy characterized by neuronal complexity loss. *Electroencephalography and Clinical Neurophysiology*, 95(2):108–117, 1995.

[55] P Grassberger and I Procaccia. Characerizationof strange attractors. *Physical Rev. Letters*, 50(5):346–349, 1983.

[56] K Lehnertz and C E Elger. Can epileptic seizures be predicted? evidence from nonlinear time series analysis of brain electrical activity. *Phys. Rev. Lett.*, 80:5019–5022, 1998.

[57] K Lehnertz, R G Andrzejak, J Arnhold, T Kreuz, F Mormann, , C Rieke, P David, and C E Elger CE. Nonlinear eeg analysis in epilepsy: Its possible use for interical focus localization, seizure anticipation, and prevention. *J. Clin. Neurophysiol.*, 18:209–222, 2001.

[58] M A Harrison, I Osorio, M G frei, S Asuri, and Y C Lai. Correlation dimension and integral do not predict epileptic seizures. *Chaos*, 15, 2005.

[59] F Mormann, T Kreuz, C Rieke, R G Andrzejak, A Kraskov, P David abd C E Elger, and K Lehnertz. On the predictability of epileptic seizures. *Clin Neuro.*, 116(3):569–87, 2005.

[60] J Martinerie, C Adam, M Le Van Quyen, M Baulac, S Clemenceau, B Renault, and F Varela. Can epileptic seizure be anticipated by nonlinear analysis? *Nature Medicine*, 4(10):1173–1176, 1998.

[61] M Le Van Quyen, J Martinerie, M Baulac, and F Varela. Anticipating epileptic seizures in real time by a non-linear analysis of similarity between eeg recordings. *Neuroreport.*, 10(10):2149–55, 1999.

[62] M Le Van Quyen, C Adam, J Martinerie, M Baulac, S Clemenceau, and F Varela. Spatio-temporal characterizations of nonlinear changes in intracranial activities prior to human temporal lobe seizures. *Eur. J. of Neurosci.*, 12:2124–2134, 2000.

[63] M Le Van Quyen, J Martinerie, V Navarro, P Boon, M D'Have, C Adam, B Renault, F Varela, and M Baulac. Anticipation of epileptic seizures from standard surface eeg recordings. the lancet, 357: 183-188 . *The Lancet*, 357:183–188, 2001.

[64] W De Clercq, P Lemmerling, W Van Paesschen, and S Van Huffel. Characterization of interictal and ictal scalp eeg signals with the hilbert transform. In *25th Annual International Conference of the IEEE EMBS, Cancun, Mexico*, 2003.

[65] M Winterhalder, T Maiwald, H U Voss, R Aschenbrenner-Scheibe, J Timmer, and A Schulze-Bonhage. The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods. *Epilepsy & Behavior*, 4:318–325, 2003.

[66] P E McSharry, T He, L A Smith, and L Tarassenko. Linear and non-linear methods for automatic seizure detection in scalp electro-encephalogram recordings. *Med. Biol. Eng. Comput.*, 40:447–461, 2002.

[67] B Litt, R Estellerand J Echauz, M D'Alessandro, R Shor, T Henry, P Pennell, C Epstein, R Bakay, M Dichter, and G Vachtsevanos. Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. *Neuron*, 30(1):51–64, 2001.

[68] M A Harrison, M G Frei, and I Osorio. Accumulated energy revisited. *Clin Neuro*, 116(3):527–31, 2005.

[69] T Maiwald, M Winterhalder, R A-Scheibe, H U. Voss, A S-Bonhage, and J Timmer. Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic. *Physica D*, 194:357368, 2004.

[70] P E McSharry, L A Smith, and L Tarassenko. Prediction of epileptic seizures: are nonlinear methods relevant? *Nat Med.*, 9(3):241–242, 2003.

[71] O A Rosso, S Blanco, and A Rabinowicz. Wavelet analysis of generalized tonic-clonic epileptic seizures. *Signal Processing*, 83(6):1275 – 1289, 2003.

[72] F Mormann, K Lehnertz, P David, and C E Elger. Mean phase coherence as a measure for phase synchronization and its application to the eeg of epilepsy patients. *Physica D*, 114:358–369, 2000.

[73] V Navarro, J Martinerie, M Le Van Quyen, S Clemenceau, C Adam, M Baulac, and F Varela. Seizure anticipation in human neocortical partial epilepsy. *Brain*, 125 (3):640–655, 2002.

[74] M Le Van Quyen, J Martinerie, V Navarro, M Baulac, and F Varela. Characterizing the neuro-dynamical changes prior to seizures. *Journal of Clinical Neurophysiology*, 18:191–208, 2001.

[75] L D Iasemidis, P Pardalos, J C Sackellares, and D S Shiau. Quadratic binary programming and dynamical system approach to determine the predictability of epileptic seizures. *Journal of Combinatorial Optimization*, 5(1):9–26, 2001.

[76] F Mormann, R G Andrzejak, T Kreuz, C Rieke, P David, C E Elger, and K Lehnertz. Automated detection of a pre-seizure state based on a decrease in synchronization in intracranial eeg recordings from epilepsy patient. *Phys. Rev. E*, 67, 2003.

[77] F Mormann, T Kreuz, R G Andrzejak, P David, K Lehnertz, and C E Elger. Epileptic seizures are preceded by a decrease in synchronization. *Epilepsy Res.*, 53:173–185, 2003.

[78] M E Saab and J Gotman. A system to detect the onset of epileptic seizures in scalp eeg. *Clin Neurophysiol.*, 116(2):427–42, 2005.

[79] L D Iasemidis, Deng-Shan Shiau W Chaovalitwongse, J C Sackellares, P M Pardalos, J C Principe, P R Carney, A Prasad, B Veeramani, and K Tsakalis. Adaptive epileptic seizure prediction system. *IEEE Transactions on Biomedical Engineering*, 50(5):616–627, 2003.

[80] M D'Alessandro, G Vachtsevanos, R Esteller, J Echauz, S Cranstoun, G Worrell, L Parish, and B Litt. A multi-feature and multi-channel univariate selection process for seizure prediction. *Clinical Neurophysiology*, 116(3):506–16, 2005.

[81] H Petsche and Mary A B Brazier et al. *Synchronization of EEG Activity in Epilepsies: A Symposium Organized by the Austrian Academy of Sciences, Vienna, Austria, September 12-13, 1971.* Springer-Verlag/Wein, New York, 1972.

[82] R Quian Quiroga, A Kraskov, T Kreuz, and P Grassberger. Performance of different synchronization measures in real data:a case study on electroencephalographic signals. *Phys Rev E Stat Nonlin Soft Matter Phys*, 65(4), 2002.

[83] C C Jouny, P J Franaszczuk, and G K Bergey. Signal complexity and synchrony of epileptic seizures: is there an identifiable preictal period? *Clin Neuro*, 116(3):552–558, 2005.

[84] M G Rosenblum, A S Pikovsky, and J Kurths. Phase synchronization of chaotic oscillators. *Phys. Rev. Lett.*, 76:1804–1807, 1996.

[85] P Tass, M Rosenblum, J Weule, J Kurths, A Pikovsky, J Volkmann, A Schnitzler, and H Freund. Detection of n:m phase locking from noisy data: Application to magneto encephalography. *Physical Review Letters*, 81(15):3291–3294, 1998.

[86] J Matias Palva, Satu Palva, and Kai Kaila. Phase synchrony among neuronal oscillations in the human cortex. *The Journal of Neuroscience*, 25(15):3962–3972, 2005.

[87] J Bhattacharya. Reduced degree of long-range phase synchrony in pathological human brain. *Acta Neurobiol. Exp.*, 61:309–318, 2001.

[88] M Le Van Quyen, J Soss, V Navarro, R Robertson, M Chavez, M Baulac, and J Martinerie. Preictal state identification by synchronization changes in long-term intracranial eeg recordings. *Clin.Neuro.*, 116(3):559–68, 2005.

[89] K K Jerger, S L Weinstein, T Sauer, and S J Schiff. Multivariate linear discrimination of seizures. *Clinical Neurophysiology*, 116:545–551, 2005.

[90] L Garcia Dominguez, R A Wennberg, W Gaetz, D Cheyne, O C Snead, and J L Perez Velazquez. Enhanced synchrony in epileptiform activity? local versus distant phase synchronization in generalized seizures. *J Neurosci*, 25(35):8077–84, 2005.

[91] M Le Van Quyen, J Foucher, J P Lachaux, E Rodriguez, A Lutz, J Martinerie, and F J Varela. Comparison of hilbert transform and wavelet methods for the analysis of neuronal synchrony. *Journal of Neuroscience Methods*, 111(2):83–98, 2001.

[92] TI Netoff and SJ Schiff. Decreased neuronal synchronization during experimental seizures. *J Neurosci.*, 22:7297–307., 2002.

[93] C E Elger and K Lehnertz. Seizure prediction by non-linear time series analysis of brain electrical activity. *Eur J Neurosci.*, 10(2):786–789, 1998.

[94] C J James, D Absolo, and D Gupta. Space-time ica versus ensemble ica for ictal eeg analysis with component differentiation via lempel-ziv complexity. In *Proc. of IEEE-EMBS (Engineering in Medicine and Biology Society) 29th International conference in Lyon, France*, 2007.

[95] C J James, R D Jones, P J Bones, and G J Carroll. Detection of epileptiform discharges in the eeg by a hybrid system comprising mimetic, self-organized artificial neural network, and fuzzy logic stages. *Clinical Neurophysiology*, 110:2049–2063., 1999.

[96] H Firpi, E Goodman, and J Echauz. Genetic programming artificial features with applications to epileptic seizure prediction. In *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, China*, 2005.

[97] K Kobayashi, C J James, H Yoshinaga, Y Otsuka, and J Gotman. The electroencephalogram through a software microscope : Non-invasive localization and visualization of epileptic seizure activity from inside the brain. *Clinical Neurophysiology*, 111:134–149, 2000.

[98] F C Meinecke, A Ziehe, J Kurths, and K R Muller. Measuring phase synchronization of superimposed signals. *Phys Rev Lett.*, 94(8): 084102, 2005.

[99] B Hong, S Acharya, S Gao, and N V Thakor. Transient phase synchrony of independent cognitive components underlying scalp eeg. In *IEEE Engineering in Medicine and Biology 27th Annual Conference, China*, 2005.

[100] D Gupta and C J James. Narrowband vs. broadband phase synchronization analysis applied to independent components of ictal and interictal eeg. In *Proc.*

of IEEE-EMBS (Engineering in Medicine and Biology Society) 29th International conference in Lyon, France, 2007.

[101] T-P Jung, S Makeig, M J Mckeown, A J Bell, T-W Lee, and T J Sejnowski. Imaging brain dynamics using independent component analysis. *Proceedings of the IEEE*, 89(7):1107–22, 2001.

[102] D Gupta, C J James, and W P Gray. De-noising epileptic eeg using ica and phase synchrony. In *Proceedings of IEE-MEDSIP2006 - The 3rd International Conference on Advances in Medical , Signal and Information Processing , Glasgow*, 2006.

[103] C Sweeney-Reed. *Analysis of Synchronisation in the EEG*. PhD thesis, University of Reading, 2007.

[104] C J James and C W Hesse. A comparison of time structure and statistically based bss methods in the context of long-term epileptiform eeg recordings. In *5th Int. Conf. on ICA and BSS (ICA 2004), Granada, Spain*, 2004.

[105] N Nicolaou and S J Nasuto. Comparison of temporal and standard independent component analysis (ica) algorithms for eeg analysis. In *ICANN/ICONIP Constantinople, Turkey.*, pages 157–160, 2003.

[106] M Jing, S Sanei, J Corsini, and G Alarcon. Incorporating bss to epileptic seizure predictability measure from scalp eeg. In *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, China*, 2005.

[107] C W Hesse and C J James. Seizure tracking and detection in ictal eeg using time-structure based blind source separation methods and prior spatial topographical information. In *Proceedings of Third European Medical and Biological Engineering Conference: EMBEC05,Prague*, 2005.

[108] C J James and C W Hesse. Mapping scalp topographies of rhythmic eeg activity using temporal decorrelation based constrained ica. In *26th Annual International Conference of the IEEE EMBS, San Francisco, CA, USA*, 2004.

[109] C J James and C W Hesse. On the use of spectrally constrained ica applied to single-channel ictal eeg recordings within a dynamical embedding framework. In *27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, China*, 2005.

[110] J Corsini, L Shoker, S Sanei, and G Alarcn. Epileptic seizure predictability from scalp eeg incorporating constrained blind source separation. *IEEE Trans Biomed Eng.*, 53(5):790–9, 2006.

[111] T Kreuz, R G Andrzejak, F Mormann, A Kraskov, H Stgbauer, C E Elger, K Lehnertz, and P Grassberger. Measure profile surrogates: a method to validate

the performance of epileptic seizure prediction algorithms. *Phys Rev E Stat Nonlin Soft Matter Phys.*, 69, 2004.

[112] R G Andrzejak, T Kreuz, F Mormann, C Rieke, A Kraskov, C E Elger, and K Lehnert. Testing the null hypothesis of the non-existence of a pre-seizure state. *Phys. Rev. E*, 67, 2003.

[113] B Schelter, M Winterhalder, T Maiwald, A Brandt, A Schad, Andreas Schulze-Bonhage, and Jens Timmer. Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction. *Chaos*, 16, 2006.

[114] P E McSharry, L A Smith, and L Tarassenko. Comparison of predictability of epileptic seizures by a linear and a nonlinear method. *IEEE Trans Biomed Eng.*, 50(5):628–33, 2003.

[115] M Chvez, M Le Van Quyen, V Navarro, M Baulac, and J Martinerie. Spatio-temporal dynamics prior to neocortical seizures: Amplitude versus phase couplings. *IEEE Transactions of Biomedical Engineering*, 50(5):571–583, 2003.

[116] M Le Van Quyen, J Martinerie, V Navarro, M Baulac, and F Varela. Characterizing neurodynamic changes before seizures. *J Clin Neurophysiol.*, 18(3):191–208, 2001.

[117] K K Jerger, T I Netoff, J T Francis, T Sauer, L Pecora, S L Weinstein, and S J Schiff. Early seizure detection. *Journal of Clinical Neurophysiology*, 18(3):259–268, 2001.

[118] M DAlessandro, R Esteller, G Vachtsevanos, A Hinson, J Echauz, and B Litt. Epileptic seizure prediction using hybrid feature selection over multiple intracranial eeg electrode contacts: A report of four patients. *IEEE Trans. Biomed. Eng.*, 50:603–615, 2003.

[119] R Esteller, J Echauz, M D'Alessandro, G Worrell, S Cranstoun, G Vachtsevanos, and B Litt. Continuous energy variation during the seizure cycle: towards an on-line accumulated energy. *Clin Neuro*, 116(3):517–26, 2005.

[120] V Navarro, J Martinerie, M Le Van Quyen, M Baulac, F Dubeau, and J Gotman. Seizure anticipation: Do mathematical measures correlate with video-eeg evaluation? *Epilepsia*, 46(3):385–96, 2005.

[121] L M Hively, V A Protopopescu, and P C Gailey. Timely detection of dynamical change in scalp eeg signals. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 10(4):864–875, 2000.

[122] K Schindler, R Wiest, M Kollar, and F Donati. Eeg analysis with simulated neuronal cell models helps to detect pre-seizure changes. *Clinical Neurophysiology*, 113(4):604–614, 2002.

[123] W De Clercq, P Lemmerling, S Van Huffel, and W Van Paesschen. Anticipation of epileptic seizures from standard eeg recordings. *The Lancet*, 361(9361):970, 2003.

[124] L M Hively and V A Protopopescu. Channel-consistent forewarning of epileptic events from scalp eeg. *IEEE Transactions on Biomedical Engineering*, 50(5):584–593, 2003.

[125] B Litt and K Lehnertz. seizure prediction and the preseizure period. *Current Opinion in Neurology*, 15:173–177, 2002.

[126] F Mormann, R G Andrzejak, C E Elger, and K Lehnertz. Seizure prediction: the long and winding road. *Brain 2007*, 130(2):314–333, 2007.

[127] L D Iasemidis. Epileptic seizure prediction and control. *IEEE Transactions on Biomedical Engineering*, 50(5):549–558, 2003.

[128] Y Yaari and H Beck. 'epileptic neurons' in temporal lobe epilepsy. *Brain Pathol.*, 12:234–239, 2002.

[129] F Perrin, O Bertrand, and J Pernier. Scalp current density mapping: value and estimation from potential data. *IEEE Trans. of Biomed. Engg*, 34:283–288, 1987.

[130] C J James and D Lowe. Ica in electromagnetic brain signal analysis. In *Fourth International Conference on Neural Networks and Expert Systems in Medicine and Healthcare, Greece.*, pages 197–202, 2001.

[131] A Hyvrinen and E Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9:1483–1492, 1997.

[132] A J Bell and T J Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, 7:1129–1159, 1995.

[133] A Ziehe and K-R Muller. Tdsep-an effecient algorithm for blind separationusing time structure. *ICANN'98 proceedings*, pages 675–680, 1998.

[134] A Belouchrani and M G Amin. Blind source separation based on time-frequency signal representations. *EEE transactions on signal processing*, 46(11):2888–2897, 1998.

[135] A Hyvrinen, J Karhunen, and E Oja. *Independent Component Analysis*. John Wiley Sons, 2001.

[136] *The FastICA MATLAB Package online at http://www.cis.hut.fi/projects/ica/fastica, (1998).*

[137] A Belouchrani, K Abed-Meraim, J F Cardoso, and E Moulines. A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing*, 45(2):434–444, 1997.

[138] S Choi, A Cichocki, and A Belouchrani. Second order nonstationary source separation. *Journal of VLSI Signal Processing*, 32:93–104, 2002.

[139] C J James and O Gibson. Ica with a reference: extracting desired electromagnetic brain signals. In *IEE Colloquium on Medical Applications of Signal Processing, London*, 2002.

[140] C W Hesse and C J James. The fast ica algorithm with spatial constraints. *IEEE Signal Processing Letters,*, 12 (11), 2005.

[141] O Gibson and C J James. On the analysis of seizure onset in the eeg: the application of constrained ica. In *Proceedings of Second European Medical and Biological Engineering Conference: EMBEC02,*, 2006.

[142] W Lu and J C Rajapakse. Ica with reference. In *1st International Conference on Brain Inspired Cognitive Systems (BICS 2004)*, 2004.

[143] C J James and C W Hesse. Semi-blind source separation in em brain signal processing. In *Proceedings of 3rd IEE International Seminar on Medical Applications of Signal Processing*, 2005.

[144] C W Hesse and C J James. On semi-blind source separation using spatial constraints with applications in eeg analysis. *IEE Transactions on Biomedical Engineering*, 53(12):2525–2534, 2006.

[145] O J Gibson and C J James. On the analysis of seizure onset in the eeg: the application of constrained ica. In *Proceedings of 2nd European Medical and Biological Engineering Conference (EMBEC'02)*, 2002.

[146] C J James and O J Gibson. Electromagnetic brain signal analysis using constrained ica. In *Proceedings of 2nd European Medical and Biological Engineering Conference (EMBEC'02)*, 2002.

[147] F Esposito, E Formisano, E Seifritz, R Goebel, R Morrone, G Tedeschi, and F Di Salle. Spatial independent component analysis of functional mri time-series: to what extent do results depend on the algorithm used? *Human Brain Mapping*, 16:146–157, 2002.

[148] A Delorme and S Makeig. Eeglab: an open source toolbox for analysis of single tial eeg dynamics including independent component analysis. *Jounal Neuroscience Methods*, 134:9–21, 2004.

[149] J Himberg and A Hyvrinen. Icasso: software for investigating the reliability of ica estimates by clustering and visualization. In *Proc. 2003 IEEE workshop on neural networks for signal processing (NNSP2003)*, pages 259–268, 2003.

[150] F Meinecke, A Ziehe, M Kawanabe, and K-R Muller. Estimating the reliability of ica projections. *In Advances in Neural Information Processing Systems*, 14, 2002.

[151] W Addison and S Roberts. Blind source separation with non-stationary mixing using wavelets. In *6th International Conference on Independent Component Analysis and Blind Source Separation*, 2006.

[152] S Makeig, S Enghoff, T P Jung, and T J Sejnowski. Moving-window ica decomposition of eeg data reveals event-related changes in oscillatory brain activity. In *Proc. 2nd Int. Workshop on Independent Component Analysis and Blind Source Separation (ICA2000)*, pages 627–632, 2000.

[153] O Gibson and C J James. Constrained ica for seizure onset analysis in the eeg. In *IEEE EMBSS PG Conference in Biomedical Engineering and Medical Physics, Birmingham,*, 2002.

[154] S Makeig, A Delorme, M Westerfield, T-P Jung, J Townsend, E Courchesne, and TJ Sejnowski. Electroencephalographic brain dynamics following manually responded visual targets. *PLoS Biology*, 2:747–762, 2004.

[155] B Hong, S Acharya, Y Ku, S Gao, and N V Thakor. Measurement of dynamic coupling of independent eeg components from cognitive tasks. *International Journal of Bioelectromagnetism*, 8(1):1–7, 2006.

[156] A Cichocki, S L Shishkin, T Musha, Z Leonowicz, T Asada, and T Kurachi. Eeg filtering based on blind source separation (bss) for early detection of alzheimers's disease. *Clinical Neurophysiology*, pages 1–9, 2004.

[157] R N Vigario. Extraction of ocular artefacts from eeg using independent component analysis. *Electroenceph. Clin. Neurophysiol.*, 103:395–404, 1997.

[158] J C Rajapakse, A Cichocki, and AV D Sanchez. Independent component analysis and beyond in brain imaging: Eeg, meg, fmri, and pet. *Neural Information Processing, ICONIP '02. Proceedings of the 9th International Conference on*, 1:404– 412, 2002.

[159] C HHuygens. *Horologium Oscillatorium*. Apud F Muguet, Parisiis, France, English Translation : Iowa State Univeristy Presss,1986, 1673.

[160] G A Petrillo and L Glass. A theory for phase locking of respiration in cats to a mechanical ventilator. *American Journal of Physiology*, 246:R311–R320, 1984.

[161] V S Anishchenko, A G Balanov, N B Janson, N B Igosheva, and G V Bordyugov. Entrainment between heart rate and weak noninvasive forcing. *Int. J. Bifurcation Chaos*, 10:2339, 2000.

[162] J Sturis, C Knudsen, N M O'Meara, J S Thomsen, E Mosekilde, E Van Cauter, and K S Polonsky. Phase-locking regions in a forced model of slow insulin and glucose concentrations. *Chaos*, 5:193–199, 1995.

[163] M G Rosenblum and J Kurths. *Nonlinear Analysis of Physiological Data*. Springer, Berlin,, 1998.

[164] W De Clercq, P Lemmerling, W Van Paesschen, and S Van Huffel. Characterization of interictal and ictal scalp eeg signals with the hilbert transform. Technical report, 03-81, ESAT-SISTA, K.U.Leuven ,Belgium,, 2003.

[165] C Allefeld and J Kurths. Testing for phase synchronization. *International Journal of Bifurcation and Chaos*, 14(2):405–416, 2004.

[166] E Pereda, R Q Quiroga, and J Bhattacharya. Nonlinear multivariate analysis of neurophysiological signals. *Progress in Neurobiology*, 77:1–37, 2005.

[167] M G Rosenblum, A S Pikovsky, and J Kurths. Synchronization approach to analysis of biological systems. *Fluctuation and Noise Letters*, 4(1):L53–L62, 2004.

[168] L X, Z Chen, K Hu, H E Stanley, and P Ch Ivanov. Spurious detection of phase synchronization in coupled nonlinear oscillators. *Physical Review E*, 73(6), 2006.

[169] A V Sazonov, C K Ho, J W M Bergmans, J B A M Arends, P A M Griep, E A Verbitskiy, P J M Cluitmans, and P A J M Boon. Analysis of the phase locking index for measuring of interdependency of cortical signals recorded in the eeg. In *Proceedings of the 29th Annual International Conference of the IEEE EMBS, France*, 2007.

[170] M Chavez, M Besserve, C Adam, and J Martinerie. Towards a proper estimation of phase synchronization from time series. *J Neurosci Methods.*, 154(1-2):149–160, 2006.

[171] J Bhattacharya, E Pereda, and H Petsche. Effective detection of coupling in short and noisy bivariate data. *IEEE Trans. on Systems, Man and Cybernetics*, 33 (1), 2003.

[172] M Palus and D Hoyer. Detecting nonlinearity and phase synchronization with surrogate data. *IEEE Eng Med Biol Mag.*, 17(6):40–45, 1998.

[173] A C Davidson and D V Hinkley. *Bootstrap Methods and their Application*. Cambridge University Press, UK, 1997.

[174] C J James and D Lowe. Using dynamical embedding to isolate seizure components in the ictal eeg. *IEE Proc: Sc., Meas & Tech*, 147(6):315–320, 2000.

[175] Xltek neuro products.

[176] I T Nabney. *NETLAB Algorithms for Pattern Recognition.* Springer-Verlag London, 2004.

[177] P Nicholas Hughes and Lionel Tarassenko. Novel signal shape descriptors through wavelet transforms and dimensionality reduction. *Proceedings of the SPIE: The International Society for Optical Engineering*, 5207:763–773, 2003.

[178] M E Tipping. *TopographicMappings and Feed-Forward Neural Networks.* PhD thesis, Aston University, Birmingham, U.K., 1996.

[179] D Lowe and M E Tipping. A novel neural net-work technique for exploratory data analysis. In *ICANN 95, Paris*, 1995.

[180] D DAlimonte, D Lowe, I T Nabney, and M Sivaraksa. Visualising uncertain data. In *European Conference on Emergent Aspects in Clinical Data Analysis (EACDA'2005), Italy*, 2005.

[181] N C Noel. Dynamical embedding and feature extraction of electroencephalographic data. Master's thesis, Aston University, 1998.

[182] D Lowe. Feature space embeddings for extracting structure from singlechannel wake eeg using rbf networks. In *Neural Networks for Signal Processing VIII, IEEE Signal Processing Society Workshop*, volume 31, pages 428–437, 1998.

[183] C J James and D D'Alimonte. Tracking multisource brain activity in multichannel eeg with a probabilistic model through ica. In *Mediterranean Conference on Medical and Biological Engineering - MEDICON2004, Italy*, 2004.

[184] S Brahim-Belhouari and A Bermak. Gas identification using density models. *Pattern Recognition Letters*, 26:699–706, 2005.

[185] H Caillol, W Pieczynski, and A Hillon. Estimation of fuzzy gaussian mixture and unsupervised statistical image segmentation. *IEEE Trans. on Image Proces.*, 6:425–440, 1977.

[186] W Pieczynski, J Bouvrais, and C Michel. Estimation of generalized mixture in the case of correlated sensors,. *IEEE Trans. on Image Proces.*, 9:308–311, 2000.

[187] A C Oliveira de Melo, Ri M de Moraes, and L dos S Machado. Gaussian mixture models for supervised classification of remote sensing multispectral images. In *CIARP, Progress in pattern recognition, speech and image analysis , Cuba*, 2003.

[188] M H Yang and N Ahuja. Gaussian mixture model for human skin color and its application in image and video databases. In *Proc. of the SPIE, vol. 3656: Conf. on Storage and Retrieval for Image and Video Databases, San Jose*, 1999.

[189] D Tran, T Pham, and M Wagner. Speaker recognition using gaussian mixture models and relaxation labeling. In *Proceedings of the 3rdWorld Multiconference on Systemetics, Cybernetics and Informatics/ The 5th Int. Conf. Information Systems Analysis and Synthesis (SCI/ISAS99), USA*, 1999.

[190] M Olof Szummer. *Learning from Partially Labeled Data*. PhD thesis, Massachusetts Institute of Technology, 2002.

[191] L Menga, M G. Frei, I Osorio, G Strang, and T Q. Nguyen. Gaussian mixture models of ecog signal features for improved detection of epileptic seizures. *Medical Engineering & Physics*, 26:379–393, 2004.

[192] D Klein, S Kamvar, and C Manning. From instancelevel constraints to space-level constraints: Making the most of prior knowledge in data clustering. In *Proc. 19th Intl. Conf. on Machine Learning (ICML)*, 2002.

[193] J M Burton, G A Peebles, D K Binder, S M Rothman, and M D Smyth. Transcortical cooling inhibits hippocampal-kindled seizures in the rat. *Epilepsia*, 46 (12):1881–1887, 2005.

[194] M H Zweig and G Campbell. Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem.*, 39(4):561–77, 1993.

[195] Y Y Kagan. Seismic moment-frequency relation for shallow earthquakes: Regional comparison. *Int. Journal of Geophys.*, 131:505, 1997.

[196] B Kemp, A Vrri, A C Rosa, Nielsen Kim Dremstrup, and J Gade. A simple format for exchange of digitized polygraphic recordings. *Electroencephalography and Clinical Neurophysiology*, 82:391–393, 1992.

[197] B Kemp and J Olivan. European data format plus (edf+), an edf alike standard format for the exchange of physiological data. *Clinical Neurophysiology*, 114:1755–61, 2003.