Neural Anomalies Monitoring: Applications to Epileptic Seizure Detection and Prediction

by

Walid Juffali

September 2012

A thesis submitted for the degree of
Doctor of Philosophy of Imperial College London

Department of Electrical and Electronic Engineering
Imperial College London

Me, Myself and I.

Acknowledgements

Albert Einstein said; "There are two ways to live your life. One is as though nothing is a miracle. The other is as though everything is a miracle"

Looking at how I lived my life, I experienced both ways. The miracle way started during my visit to Thailand after I met Professor Chirs Toumazou, by chance, where he was working on rice gene modification. It was then, at the age of 53, I decided to go back to University and do something completely different to what I studied and trained. Something in the life sciences not in the world of business, my familiar domain.

How exciting it was the first day I walked into Imperial college to attend my first lecture after decades of being away from formal education.

The loss of my son Mohammed this year has shattered my core and the will of God and the support of my family and friends is helping me deal with this unimaginable loss. This work is dedicated to you, my beloved Mohammed. Your memory will remain deep in my heart forever. My love to you is immeasurable.

First and foremost, I would like to thank my loving family for their support and for being there for me throughout my life.

When it comes to acknowledgement, I would like to express my sincere gratitude to my advisor and mentor Professor Chris Toumazou for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance and enthusiasm have been invaluable to me. He not only kept me focused on the tasks ahead but also kept my passion in this subject alive. I am extremely fortunate to have had Professor Toumazou as my mentor.

I wish to also thank Dr Amir Eftekhar for the tremendous effort he put in helping me with the research and the construct of the thesis. I am eternally grateful for his contribution and for sharing his in-depth knowledge on the subject with me. Thank you Amir very much.

My thanks and gratitude goes to Dr. Belinda Nedjai, Dr. Pantelis Georgiou and Dr. Timothy Constandinou for inspiring me and for helping me get started and stay on course. Their energy, youth and intellect is a shining example for all to see. My admiration to the respective work they are doing in bio-medical engineering under the guidance of Professor Toumazou is immense.

My thanks also goes to Dr Jamil El-Imad, my lifetime friend, who shared my enthusiasm from the very start and together over many late nights of out-of-the-box thinking, in Gstaad and London, we developed the core logic. Jamil and I fuelled each other's enthusiasm during the research as we approached the problem from a completely different way and from two different mindsets. we had a lot of fun diving into the unknown.

Beautiful Gstaad was the place where we came up with this unusual signal processing method in the early hours of one winter's day so thank you Gstaad for the inspiration and for the tranquility that I find every time I visit.

I started with an Einstein quote and will end with another having reflected on my period of studies.

"When I examine myself and my methods of thought, I come to the conclusion that the gift of fantasy has meant more to me than the talent for absorbing positive knowledge."

Abstract

There have been numerous efforts in the field of electronics with the aim of merging the areas of healthcare and technology in the form of low power, more efficient hardware. However one area of development that can aid in the bridge of healthcare and emerging technology is in Information and Communication Technology (ICT). Here, databasing and analysis systems can help bridge the wealth of information available (blood tests, genetic information, neural data) into a common framework of analysis. Also, ICT systems can integrate real-time processing from emerging technological solutions, such as developed low-power electronics.

This work is based on this idea, merging technological solutions in the form of ICT with the need in healthcare to identify normality in a patients' health profile. In this work we develop this idea and explain the concept more thoroughly. We then go on to explore two applications under development.

The first is a system designed around monitoring neural activity and identifying, through a processing algorithm, what is normal activity, such that we can identify anomalies, or abnormalities in the signal. We explore Epilespy with seizure detection and prediction as an application case study to show the potential of this method. The motivation being that current methods of prediction have proven to be unsuccessful. We show that using our algorithm we can achieve significant success in seizure prediction and detection, above and beyond current methods.

The second application explores the link between genetic information and standard tests (blood, urine etc...) and how they link in together to define a personalised benchmark. We show how this could work and the steps that have been made towards developing such a database.

Table of Contents

Acknowledgements **3**

Abstract **5**

List of Figures **9**

List of Tables **12**

List of Abbreviations **13**

Chapter 1 – Introduction **14**

1.1 Motivation **14**

1.2 Aims **16**

1.3 Report Outline **17**

Chapter 2 - Healthcare and Technology **18**

2.1 Example: Kidney Function monitoring **18**

2.1.1 ISFET Basics 18

2.1.2 Application to Urea:Creatinine 19

2.2 Algorithm of biological system **21**

2.1.1 Initial concept 22

2.3 References **24**

Chapter 3 - A State of the Art Review: *Neural Signal Analysis* **25**

3.1 Introduction **25**

3.2 Neurological Disorders **26**

3.3 Epilepsy **27**

3.4 Analysing EEGs **29**

3.4.1 Seizure Prediction and Detection 30

3.4.2 Clinical EEG 32

3.4.3 Seizure Prediction 34

3.4.4 Linear Analysis 35

3.4.5 Nonlinear Analysis 38

3.4.6 Statistical Measures 47

3.4.7 Results 49

3.4.8 Seizure Prediction: Conclusions 51

3.5 Conclusions **52**

3.6 References **53**

Chapter 4 - WiNAM: A novel tool for analyzing biosignals 58
4.1 Pattern Recognition 58
4.2 NGRAMs 58
4.2.1 Symbolic Representation 59
4.2.2 Multiresolution Ngram 62
4.2.3 System Implementation 66
4.2.4 Data Requirements 69
4.2.5 Data Input 71
4.2.6 Data Analysis 72
4.2.7 WiNAM Parameters 72
4.3 Summary/Conlusions 74
4.4 References 75
Chapter 5 - Clinical Studies and Results 76
5.1 EPI Data 76
5.2 Freiburg Data 78
5.3 Test 1 - *Initial Results* 79
5.3.1 Towards Real-time detection 80
5.4 Test 2 - *Detection and Prediction* 82
5.4.1 Detection 83
5.4.2 Interference Removal 86
5.4.3 Prediction 93
5.4.4 Prediction Validation 99
As described in [6] for a FPR of 0.15 / hour (which is clinically relevant) and an SOP of 10 minutes, this yields values for each patient of their minimal and upper critical sensivity value as outlined in **Table 5.4**. **100**
5.5 Discussions/Conclusions 102
5.6 References 103
Chapter 6 - Conclusions and Future Work 104
6.1 Results - Discussions/Conclusions 104
6.2 Thesis Contributions 106
6.3 Future work 106
6.4 WiNORM 107
Genetic disorders 109
Disease Markers 111
System 112

To create this database we are using the same strategy of system biology. The aim is to understand the body as a whole. **112**
Database design 113
6.5 Final Comments **118**
Appendix A – Matlab Code **119**
DATA ADC **119**
HEX CONVERSION **119**
SIGNED DECIMAL to HEX **121**
NW ALGORITHM **121**
BASIC SINGLE PATTERN NGRAM **122**
BASIC MULTIRESOLUTION NGRAM IMPLEMENTATION **123**
PROCESS NGRAM WITH WINDOWED ANALYSIS **125**
Appendix B - Urea:Creatinine Schematics **132**
Appendix C - Conference Paper **133**

List of Figures

Figure 1.1 – Me, myself and I – what it means. 16

Figure 2.1 - Figure of a CMOS FET (a) with modification to ISFET (b) and its model (c). 19

Figure 2.2 – Urea: Creatinine sensor using weak inversion transistor [1]. 21

Figure 2.3 – Algorithm of life 23

Figure 3.1 – A list of the most common and affecting neurological disorders. 26

Figure 3.2– The estimated DALY for 2005, 2015 and 2030. 27

Figure 3.3 - Table from Strogatz, [31] illustrating the classification of various natural mechanisms, including biological 39

Figure 3.4 - Figure highlighting the seizure prediction and occurrence areas. 48

Figure 4.1 – Figure showing a sample EEG data set re-quantised from its original 16 bit to 6 bit using our outlined methodology. 61

Figure 4.2 – Overview of how the Ngram technique is used with real-data 61

Figure 4.3 - Figure showing the pattern counts for three ngram sizes with three different quantisation levels: (a) 8, (b) 6 and (c) 4. 63

Figure 4.4 – Figure showing the results of the implementation of the unique multiresolution ngram approach. 66

Figure 4.5 - General system overview envisioned for implementing a signal processing method for biological data. 67

Figure 4.6 – The main user interface for accessing the database and analysis. 68

Figure 4.7 – Image of the hierarchy employed in the WiNAM database. 69

Figure 4.8 - The uploading file interface (top right) and the viewed data in its raw ascii form (engineering format) and processed (after weighting). 71

Figure 4.9 – Graphical representation of the seizure (blue) and nonseizure (red) area that can be viewed from the interface. 72

Figure 4.10 – Showing the ngram pattern count. 73

Figure 4.11 – The summary of the runs ratio (expressed as anomalies percentage) for a case. 73

Figure 5.1 – Figure showing the results using WiNAM at different low-pass filtering frequencies 76

Figure 5.2 – Figure showing a 5Hz filtered version of the EPI data with voltage reference levels set to the maximum and minimum of the signal. 77

Figure 5.3 – Figure showing the patient details breakdown, taken from https://epilepsy.uni-freiburg.de/freiburg-seizure-prediction-project/eeg-database..... 78

Figure 5.4 - Example of how seizure and non-seizure areas are selected and compared..... 79

Figure 5.5 – Shows a typical ECoG signal (top) with seizure occurrence (shaded area) and the pattern count (bottom) for a 10 second window 81

Figure 5.6 – Shows the quick increase and slow decrease seen in some sample data..... 82

Figure 5.7 - Example of non-optimised ngram analysis of EEG signal (top) and ngram pattern counts (bottom). Seizure occurs at approx. t = 65mins. 84

Figure 5.8 - Figure showing the thresholds as a function of signal moving mean..... 84

Figure 5.9 - Bar chart of the sensitivity (top) and false detections (bottom) of each Patient over all possible seizures used to determine the optimal threshold ("All") and for the best case scenario ("Best"). 85

Figure 5.10 - Showing EEG plagued with artefacts (all large amplitude spikes), including at seizure onset (~60 minutes). Enclosed area indicates artifact shown in Figure 5.11.. 86

Figure 5.11 - A zoomed artifact of Figure 5.10. 86

Figure 5.12 - Image showing (top) original EEG artifact signal, (middle) SG filter estimate of artifact and (bottom) EEG with artifact removed. 87

Figure 5.13 -Bar chart of the 60 second windows with artifact removed..... 88

Figure 5.14 - Bar chart of the 30 second windows with artifact removed..... 89

Figure 5.15 - Optimal results for combined 30 and 60 second windows. 89

Figure 5.16 - Optimal results using the combination of multiple pattern detections to determine if a detection should be considered. 92

Figure 5.17 - Example of Case 1, between (a) Interictal period and (b) the ictal period with seizure occurring at interval 53 (where the histogram initially rises). The thick (green) line shows the standard deviation threshold previously defined..... 94

Figure 5.18 - Patient 1, all seizures with each of 4 seizures (red), the 4 preictal periods (in ascending order, green) and interictal (blue). The cross indicated the mean, with minimum and maximum indicated with the error bars. Each 7 plots are Ngram size 16 (top) to 4 (bottom) with intervals of 2..... 95

Figure 5.19 - Patient 10, all seizures with each of 4 seizures (red), the 4 preictal periods (in ascending order, green) and interictal (blue). The cross indicated the mean, with minimum and maximum indicated with the error bars. Each 7 plots are Ngram size 16 (top) to 4 (bottom) with intervals of 2..... 96

Figure 5.20 - Case 1, seizures 1 to 4, and indication of inteictal mean, maxima and minima. 97

Figure 5.21 - Sensitivity and FDR for an IT of 40 minutes and SOP of 10. 98

Figure 5.22 - Sensitivity and FDR for an IT of 20 minutes and SOP of 10. 99

Figure 5.23 - Optimal sensitivity for each patient as a combination of all processed IT periods. Noted with an asterix (*) for the upper and circle (o) for the lower critical sensitivity values of Table 5.4, SOP of 10 minutes. 101

Figure 5.24 - FPR for the optimal sensitivity described in Figure 5.23 and the 0.15 FPR.... 101

Figure 6.1 - WiNorm Database concept 113

Figure 6.2 - Database design concept 117

Figure 6.3 - Diagram of WiNorm database 117

List of Tables

Table 2.1 – Significance of particular BUN to creatinine ratios 20

Table 3.1 – Frequency Measures used in analyzing EEGs. 29

Table 3.2- Interpreting the EEG 33

Table 3.3 - Typical seizure classifications 33

Table 3.4 - Table describing the typical metrics used in prediction results. 47

Table 4.1 - Example of transition from sampled value to HEX value 62

Table 5.1 - Taken from [4] showing the initial results aiming to quantify the ratio of patterns between seizure and non-seizure. 80

Table 5.2 – Showing the qualified results of full analysis using single pattern counts over the full preictal period available. 81

Table 5.3 - Depiction of the number of detections need shared by multiple patterns to warrant a detection of seizure and/or false detection. 91

Table 5.4 - Lower and upper sensitivity (%) for two different SOP values. The significance of a seizure prediction result should be ideally place above the upper value. 100

List of Abbreviations

ARMA	AutoRegressive Moving Average
BUN	Blood Urea Nitrogen
CBIT	Centre for Bio-Inspired Technology
CMOS	Complementary Metal-Oxide Semiconductor
DNA	DeoxyriboNucleic Acid
ECoG	ElectroCorticogram
EEG	ElectroEncephaloGram
FET	Field Effect Transistor
FN	False Negative
FP	False Positive
FPR/FDR	False Prediction/Detection Rate
GTC	General Tonic-Clonic
HEX	HEXadecimal
ICA	Independent Component Analysis
ISFET	Ion Sensitive FET
IT	Intervention Time
LE	Lypunov Exponent
SG	Savitsky-Golay
SOH	Seizure Occurrence Horizon
SPH	Seizure Prediction Horizon
SQL	Structure Query Language
TN	True Negative
TP	True Positive
WHO	World Health Organisation
YLD	Years of Life Lost due to Disability
YLL	Years of Life Lost due to Mortality

Chapter 1– Introduction

This project was initiated by one question: *what is normality?* In healthcare there are numerous parameters used to define your current and future health. Given that we, as humans, are so different, using the same health parameters across all patients would seem a simplistic way to assess your health. The question is then: *Is there a norm that is valid for everyone and where there is not can we provide a more realistic view of our own health?*

1.1 Motivation

With increased government regulations, e-business challenges, rising patient health expectations and demand for lower healthcare costs, the worldwide healthcare industry is facing immense pressure to utilize information technology to streamline healthcare processes providing cost effective healthcare and personalised services with improved efficiency.

Technology should ideally help companies, physicians and patients by providing custom healthcare software development services to manage these challenges. The challenge is to offer expert custom healthcare software applications development and application management services in the areas of healthcare like pharmaceutical, medical and life sciences.

In the Centre for Bio-Inspired Technology (CBIT), from which this work grew, a number of enabling technologies have been developed to extract biomedical parameters for diagnostics, early detection and therapy. Indeed this work first took this path in developing a sensor platform for Urea to Creatinine ratio calculation (to be discussed later) but we soon realized that there was a larger picture to consider in the way technology interacts with personalized healthcare.

Mainly, there is a need for healthcare solutions to come in the form of software implementations that aim to improve on existing healthcare systems for more presonalised healthcare analysis.

Furthermore the domain of organizational technologies in healthcare could thus be depicted as a rough circle around the nexus of four quadrants: policies, procedures, guidelines, and standard practices. Healthcare workers, according to this view, are necessarily knowledge workers, because even where health care is routine, individual differences among patients need to be accommodated. The appropriate administrative structure would be mixed, accommodating the need for clinicians to vary routines, solve problems, and be creative in meeting both clinical and management demands. To support this practice, health informatics would have to provide for communication to include ad hoc queries and responses, order entry, results reporting, vertical and horizontal messaging, and consultation. Decision support would include access to a variety of knowledge resources and databases, to algorithms and creative problem-solving tools, and to policies, procedures, guidelines, and standard practices.

The ability of the clinician to incorporate these issues into the clinical decision with greater or lesser certainty and by intuitive or logical processes defines whether, in that moment, health care is being practiced predominantly as craft, routine, engineering, or non-routine. The role of informatics is to reduce uncertainty in the human decision-making process by delivering state-of-the-art information, cues, and alerts and to offer well-defined decision algorithms where possible. It remains a human grace and a professional responsibility to make intellectual leaps when uncertainty is unavoidable and to be creative when algorithms are inadequate

The motivation is then to provide personalised health care to each individual. How can one combine information about me, myself and I to improve and manage our own health (Figure 1.1).

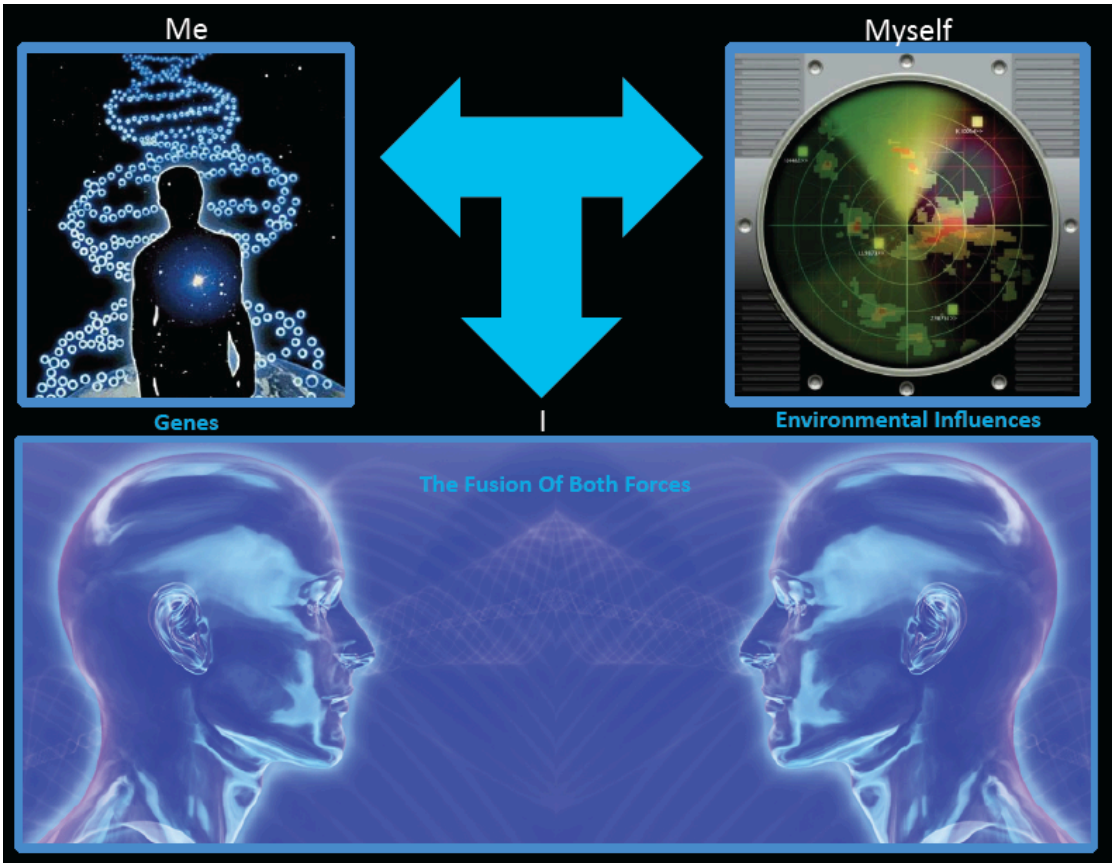


Figure 1.1 – Me, myself and I – what it means.

1.2 Aims

This work aims to target two particular solutions that covers a range of health care activities including:

- Electronic Medical Records
- Patient Normal Values Solutions.

We aim to develop customized databases and portals for healthcare organizations that provide them with a platform to share information about patients. Our solutions are based

on a content management architecture that facilitates and eases the task of publishing and managing content appearing on a website.

Further to this we aim to quantify what falls outside the norm, i.e. anomalies, associated with a pattern and customise this to each persons medical profile. The application chosen in this work is that of epilepsy and seizure prediction. This is an area that has seen multiple attempts to quantify prediction methods but still yet to find a reliable one.

1.3 Report Outline

The project is organized in three key chapters. Chapter 2 briefly focuses on the model of developing databases and solutions as previously described. It discusses further how the model is envisioned and can be implemented. In addition it describes a novel implementation of low power electronics applied to a healthcare application.

The third chapter focuses on the application of signal processing methodologies that have been developed for seizure prediction and detection. Here we explore both the clinical and algorithmic perspective of brain signals relating to seizures and potentially their prediction.

Chapter 4 derives the implementation of our novel pattern recognition signal processing method and database structure to support the analysis of brain signals in epileptic patients. Chapter 5 then applies this to two clinical data sets to show its value for seizure prediction and detection. Both sets of data are published data sets used for several benchmarking trials in the literature.

Finally in Chapter 6, we discuss the significance of this work, including future visions of extending the applications to more generalized personal healthcare, incorporating genetic and metabolic patient data.

Chapter 2- Healthcare and Technology

As we have stated in the introduction the aim of this work is towards databases and portals for personalizing healthcare. In this chapter we aim to expand this concept and how it has developed in the course of this work.

At the CBIT, as mentioned previously, there have been a number of projects involved with developing systems aimed towards personalized healthcare. One example work that this project initially started on is a sensor interface for monitoring kidney function [1] (see Appendix C - Conference and Journal Paper). This is a good example of how technology can be used to improve personalized healthcare.

2.1 Example: Kidney Function monitoring

The group has a strong history on developing micro chemical sensors called ion-sensitive field-effect transistors (ISFET) and the circuitry that interfaces with them. In an attempt to understand this technology and application we worked on a sensor interface for monitoring the ratio between urea and creatinine.

2.1.1 ISFET Basics

There are generally three categories of chemical sensors: *amperometric* (applying a DC bias and measuring current), *voltammetric* (sweeping DC bias and measuring current [2]) and *potentiometric* (passively measuring the potential- without forcing a bias potential or current). Being passive in nature, the *potentiometric* approach is therefore best suited for interfacing biology.

One such potentiometric sensor that is becoming increasingly popular due to the fact that they can be fabricated in unmodified CMOS technology is the Ion-sensitive field effect

transistors (ISFETs). ISFETs have predominantly been used for pH measurements [3]. These sensors use a membrane to electrostatically couple the charge to modulate the underlying active device. This can be achieved in CMOS by exploiting the silicon nitride passivation as the sensing membrane over a floating gate field-effect transistor (FET) [CmosISFET]. ISFETs can be made sensitive to specific ions (often referred to as ChemFETs) by the deposition of an appropriate polymer membrane (ionophores [4]), permeable to specific ions [3], on the device sensing area. The ChemFET response results from the potential that develops at the membrane-electrolyte interface due to the reversible binding of the analyte with the ion carrier. An intermediate layer is often applied, to establish a reversible electron transfer pair at the ion-selective membrane-ISFET interface.

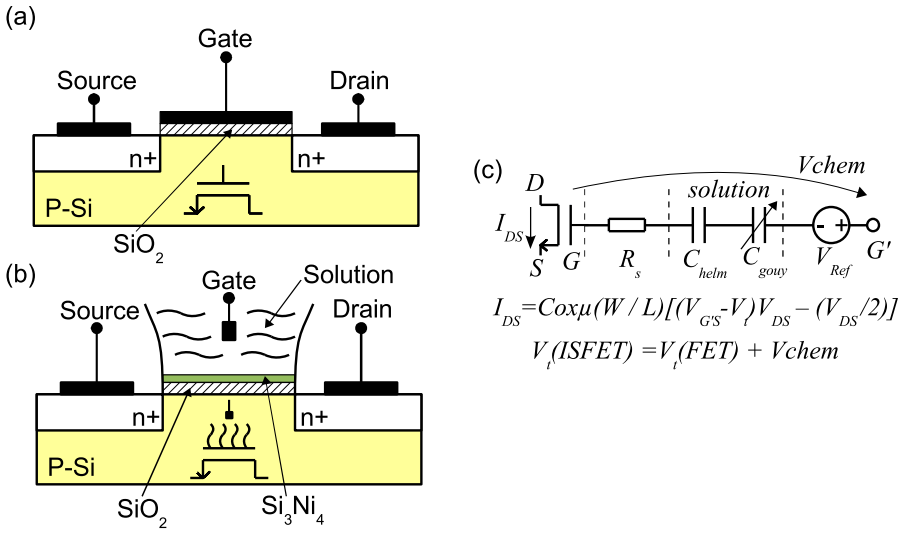


Figure 2.1 - Figure of a CMOS FET (a) with modification to ISFET (b) and its model (c).

2.1.2 Application to Urea:Creatinine

It has been stated that the ratio of urea to creatinine concentration is an important indicator of renal failure [5]. Urea is usually measured as the blood urea nitrogen (BUN) where BUN = urea/2.14. The BUN to Creatinine ratio has three main categories listed in Table 2.1.

Table 2.1 – Significance of particular BUN to creatinine ratios

BUN:Cr ratio	Location of Renal Disorder
>20:1	Pre-renal disease (before glomeruli)
10-20:1	Normal or post renal disease (within ureter)
<10:1	Intrarenal disease (within kidneys)

Here, the golmeruli is a capillary that performs the first step in filtering blood to form urine and the ureter is a muscular tube that transports urine from kidney to bladder. Hence, this ratio is a strong indicator of renal failure. To be able to measure this ratio we developed a sensor based on the ion-selective field effect transistor (ISFET). This is a modified transistor with gate removed and instead the gate charge is controlled by the pH of the solution as H+ ions, which are related to pH modify the transistors behaviour (by bonding to the Silicon Nitride passivation) [3, 6]. Adding a membrane onto the transistor allows sensing of other chemical concentrations. In this example urease and creatinase can be immobalised to sense urea and creatinine respectively [7].

The other feature that these systems use is the weak inversion state of a transistor. This allows low power operation and since there is a log relationship between the chemical signal and the pH then weak inversion, which utilizes a log relationship, linearises the chemical recording.

The proposed circuit, published in [1] is shown in Figure 2.2. Here Iurea and Icr are given by:

$$I_{urea} = I_0 e^{VGS/nUt} K_{urea} [urea]^{a_{urea}/n} \quad (1)$$

$$I_{Cr} = I_0 e^{VGS/nUt} K_{Cr} [Cr]^{a_{Cr}/n} \quad (2)$$

which is based on the weak inversion operation of a transistor [8, 9]. Using this principle of

a translinear loop (a method for exploiting exponential relationships) we can formulate a linear relationship.

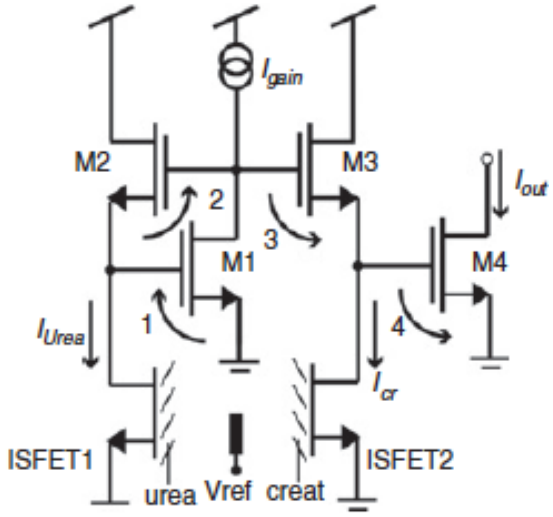


Figure 2.2 – Urea: Creatinine sensor using weak inversion transistor [1].

In this case VGS1+VGS2 = VGS3+VGS4 so substituting the weak inversion currents we eventually obtain the equation:

$$I_{out} = I_{gain} \frac{K_{urea}[urea]^{a_{urea}/n}}{K_{Cr}[Cr]^{a_{Cr}/n}} \tag{3}$$

Details of the derivation can be found in [1]. We simulated this circuitry in a CMOS design environment (Cadence) and showed that the ratio can be computed as per Table 2.1 but with monitoring of the current output of Figure 2.2.

2.2 Algorithm of biological system

The previous section describes a very simple example of how one can utilize technology in this form to derive novel solutions to existing medical problems. However, we soon came to realize that this type of technology is a small portion of the personalized healthcare space. In conventional medicine there is a large reliance on databases and Information Technology to produce more efficient healthcare solutions.

2.1.1 Initial concept

The initial concept is based on the need to gather a multitude of information available for one person to understand the person's health status as a whole. These would include genetic, blood work and even neurological data. Based on an inter-disciplinary study field that focuses on complex interactions in biological systems, this research will bring new perspective to diagnosis and health care by combining all this information into one framework.

Figure 2.3 aims to encompass this vision, which shows that a person has many inputs and outputs that can feedback to form a critical evaluation of a person's health status. This model is applicable to both personalized healthcare in the sense of self-evaluation to more in depth diagnostic and treatment solutions in the medical industry space. Technology such as the urea:creatinine sensor falls into this scope quite well. It is our aim to expand on this further and explore the need and model for such future generation of personalized healthcare systems.

Given the expertise available in CBIT, we chose to explore the neural signal processing space within this framework. The next chapter describes a novel algorithm developed with the aim of assessing neurological anomalies.

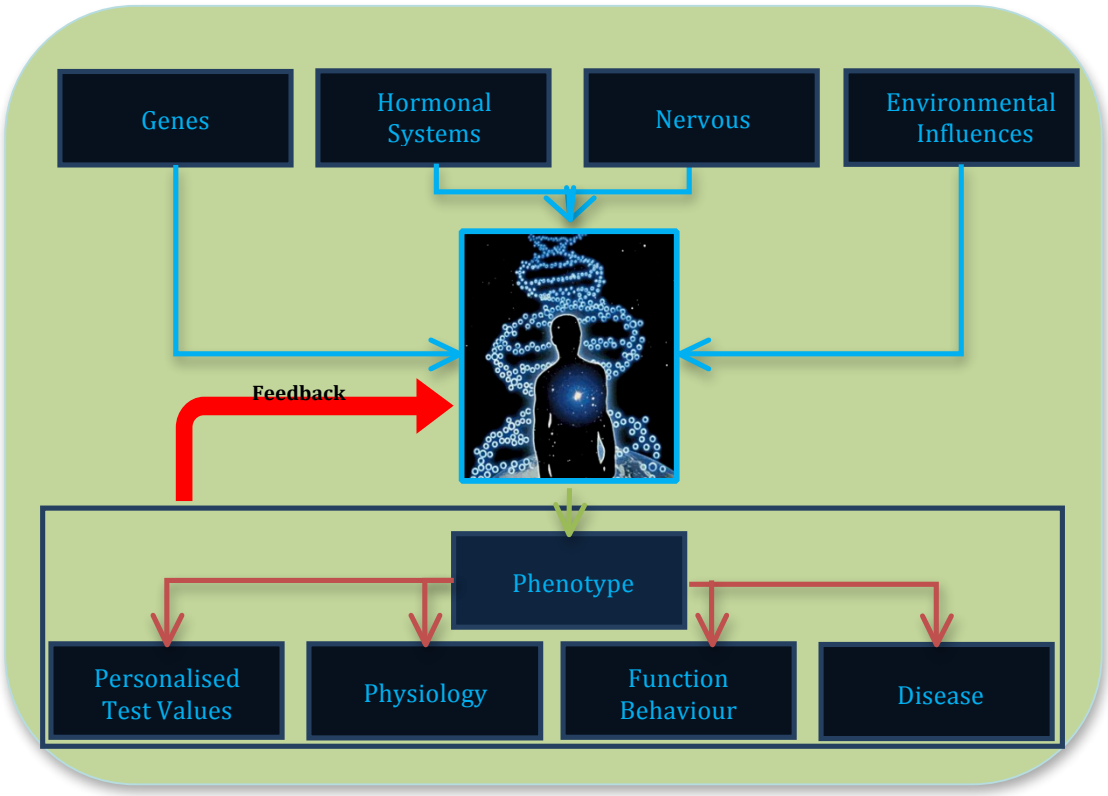


Figure 2.3 – Algorithm of life

2.3 References

[1] W. Juffali, P. Georgiou, and C. Toumazou, "ISFET based urea:creatinine translinear sensor " *IEEE Electronics Letters,* vol. 46, pp. 746-748, 2010.

[2] B. A. Patel, C. A. Anastassiou, and D. O' Hare, *Biosensor Design and Interfacing*: Springer London, 2006.

[3] P. Bergveld, "Thirty years of ISFETOLOGY - What happened in the past 30 years and what may happen in the next 30 years," *Sens and Act B,* vol. 88, pp. 1-20, 2003.

[4] S. Sharma, B. C. T. Field, and C. Toumazou et al, "An integrated silicon sensor with microfluidic chip for monitoring potassium and pH," *Micro and Nanofluidics,* vol. 10, pp. 1119-1125, 2011.

[5] D. Morgan, M. Carver, and R. Payne, "Plasma creatinine and urea:creatinine ratio in patients with plasma urea," *British Medical Journal,* vol. 2, p. 929, 1977.

[6] J. Baussells, J. Carrabina, A. Errachid, and A. Merlos, "Ion-Sensitive field-effect transistors fabricated in commercial CMOS technology," *Sensors and Actuators B,* vol. 57, pp. 56-62, 1999.

[7] B. Premanode and C. Toumazou, "A novel, low power biosensor for real time monitoring of creatinine and urea in peritoneal dialysis," *Sensors and Actuators B,* vol. 120, pp. 732-735, 2007.

[8] L. Shepherd and C. Toumazou, "A biochemical translinear principle with weak inversion ISFETs," *IEEE transactions of Circuits and Systems* vol. 52, pp. 2614-2619, 2005.

[9] L. Shepherd and C. Toumazou, "Weak Inversion ISFETs for ultra-low power biochemical sensing and real-time analysis," *Sensors and Actuators B,* vol. 107, pp. 468-473, 2005.

Chapter 3 - A State of the Art Review: *Neural Signal Analysis*

3.1 Introduction

Over the last several decades neuroscientists have interfaced with the nervous system through technology in a multitude of application spaces. These advanced neural interfaces have allowed us to capture single-neuron recordings and in depth spatial mappings of brain function. One of the key challenges with handling these recordings is extracting the spatio-temporal signatures related to activity and neurological function.

To overcome this challenge, signal processing techniques have acted as the bridge between recording and understanding, whether through statistical methods or complex nonlinear state-space mapping. Apart from opening a window of understanding to brain function these methods also aim to act as a bridge between function and treatment.

For example, in Parkinson's disease closed loop deep brain stimulators that have been designed, and have been clinically tested, show that medication and tremors can be reduced [1]. These systems record, process and stimulate deep brain structure of the brain.

In general there are several neurological disorders of interest that could benefit from the utilization of technology. One of the bottlenecks of the research in this area has been data consistency; having data sets well annotated, defined and databased. From this it would already seem that it fits into our model. To better clarify this, we will briefly look at Neurological Disorders in general and focus on one - Epilepsy. This will give us a perspective on what is missing that can help support work in this area.

3.2 Neurological Disorders

The brain is estimated to contain up to 100 billion neurons, and the nerves that make up the communication network from brain to periphery is a 45 mile long network that can transmit at speeds of up to 120 m/s (270 mph)[1].

Like many elements of the human body a lot of things can and do go wrong. This can be caused by injuries, diseases/viruses, cancer or other disorders. It is difficult to assess how a neurological disorder can affect individuals so in an attempt to do so the World Health Organisation (WHO) published an online book [2].

In it the top neurological disorders are stated (Figure 3.1). Additionally, an attempt was made to quantify the effect of these disorders on individuals. To achieve this they use two measures:

1. Years of Life Lost due to Mortality (YLL)
2. Years of Life Lost as a result of Disability (YLD)

Adding these two figure together results in a measure called the Disability Adjusted Life Years (DALY). In essence *"One DALY can be thought of as one lost year of healthy life and the burden of disease as a measure of the gap between current health status and an ideal situation."* [2].

Neurological disorders in neuropsychiatric category	Disorders/injuries with neurological sequelae in other categories
Epilepsy	Cerebrovascular disease
Alzheimer and other dementias	Poliomyelitis
Parkinson's disease	Tetanus
Multiple sclerosis	Meningitis
Migraine	Japanese encephalitis

Figure 3.1 – A list of the most common and affecting neurological disorders [2].

[1] www.hypertextbook.com/facts/2002/DavidParizh.shtml

With this measure a projection was made as to the estimated DALY for 2030. This is shown in Figure 3.2 that estimates that in the year 2030, approximately 103 million DALYs for the top disorders. This would mean that for a life span of 70 years (for Brazil) this equates to 10% of Brazil's population or in the UK the whole population of Northern Ireland.

Cause category	2005		2015		2030	
	No. of DALYs (000)	Percentage of total DALYs	No. of DALYs (000)	Percentage of total DALYs	No. of DALYs (000)	Percentage of total DALYs
Epilepsy	7308	0.50	7419	0.50	7442	0.49
Alzheimer and other dementias	11078	0.75	13540	0.91	18394	1.20
Parkinson's disease	1617	0.11	1762	0.12	2015	0.13
Multiple sclerosis	1510	0.10	1586	0.11	1648	0.11
Migraine	7660	0.52	7736	0.52	7596	0.50
Cerebrovascular disease	50785	3.46	53815	3.63	60864	3.99
Poliomyelitis	115	0.01	47	0.00	13	0.00
Tetanus	6423	0.44	4871	0.33	3174	0.21
Meningitis	5337	0.36	3528	0.24	2039	0.13
Japanese encephalitis	561	0.04	304	0.02	150	0.01
Total	**92392**	**6.29**	**94608**	**6.39**	**103335**	**6.77**

Figure 3.2– The estimated DALY for 2005, 2015 and 2030 [2].

There are many other considerations to be made including the social and economic implications. Socially, it is difficult to assess because figures are harder to gather in developing countries. However, up until the late 1800's disorders such as Epilepsy were treated as being caused by evil influences [3] and still are in many places.

Epilepsy is a useful example of a neurological disorder and the one that the remainder of this work focuses on although future work will extend to other areas.

3.3 Epilepsy

Epilepsy is a neurological disorder that affects approximately 1% of the world's population [3]. It manifests in most patients as seizures that can cause the patients to lose consciousness. There can be many causes but mostly it is described as idiopathic (no known cause).

In the UK alone 456,000 people are recorded as having epilepsy (1 in 131 people) of which only 70% could be controlled with proper anti-epileptic drugs (AED). Of the remaining, 3% are eligible for surgery, but there are up to 500 new applications per year and surgery is back-logged[2]. In the UK alone, it is estimated that the yearly cost will be £0.5 billion to treat the current population of children with epilepsy[3]. In the US, the cost is estimated at $15 billion per year from a population of 3 million Americans (1% of the population) with 200,000 new cases per year. Epilepsy in the US is the 3rd most prevalent Neurological Disorder[4].

Generally epilepsy is diagnosed based on clinical history and through electroencephalogram (EEG) analysis. EEG is a method whereby dozens of electrodes are placed on the scalp to measure the cumulative activity of several thousands of neurons. These electrodes can come in many forms, in some cases needing a gel or saline solution to improve conductivity or promote adhesion to the skin surface. There are many theories as to how the electrodes pick up the electrical activity and the distribution of this measurement. This however is not the focus of this work and as such we refer the reader to [4], which is a definitive publication on the topic.

There are many other ways of accessing brain activity (implanted electrodes, fMRI, MEG etc…) but given this work is tailored towards personalized healthcare systems and general use in hospitals EEG is still the optimum choice of analysis.

[2] http://www.epilepsysociety.org.uk/AboutEpilepsy/Whatisepilepsy/Epilepsy-didyouknow
[3] www.jointepilepsycouncil.org.uk
[4] www.epilepsyfoundation.org

3.4 Analysing EEGs

EEGs have been analysed through linear, nonlinear (state-space) or multivariate analysis techniques depending on the application [5, 6]. In the clinical field there are a number of features used to assess the activity. With Epilepsy there are a number of particular syndromes associated with it that include[5]: West syndrome, Lennox-Gastaut Sydrome, childhood absence epilepsy, benign partial epilepsies and juvenile myoclonic epilepsy. Although as mentioned a number of epilepsy conditions are idiopathic (no known cause). With each condition and in general EEG analysis, clinicians usually use frequency quantifiers (Table 3.1)[7].

Table 3.1 – Frequency Measures used in analyzing EEGs.

Type	Frequency/Hz
Delta	< 4
Theta	4-7
Alpha	8-12
Beta	12-30
Gamma	30+

Epilepsy studies have aimed to extract features from neurological recordings for prediction and/or detection of seizures. However, although excellent work has emerged [5, 6, 8, 9], predicting seizures is still an open problem with mixed results as to the optimum approach, which we shall discuss later.

This is being tackled by many groups including through the European Union FP7 grant[6]. The field is still open for new methods applied to neurological signals that can give more insight into the signal dynamics surrounding epileptic seizures. Although in the area

[5] http://www.epilepsy.org.uk/info/syndromes
[6] http://www.epilepsiae.eu/

of Neonates where seizure detection is sought there has been some success [10-12], but nothing translated to usable systems.

3.4.1 Seizure Prediction and Detection

Generally, seizures and EEG-based signals are treated in two different frames of thought: detection and prediction. Prediction has seen work ever since the 1950s applying linear, nonlinear (state-space) or multivariate analysis techniques to the EEG and derivatives of it. The ability to predict a seizure would allow for intervention strategies, such as brain cooling, electrical stimulation or drug delivery [13-15] to be administered for those patients where medication and surgery has had little or no effect. The state of prediction is really best summarised in [5, 8] . In the 1990's up until 2000 there was an optimistic drive towards prediction and the methods and data used varied but all showed optimistic results. Chaotic measures such the Lyapunov Exponent [16] and Correlation Dimension [17] were presented as possible markers .

From 2000 until present a more formal approach emerged to validate results and organise results/data, including more robust statistical validation methods (e.g. surrogate data analysis, sensitivity and specificity quantifies [18, 19]). This quickly showed that all current methods were not much better than chance and that linear methods (energy, entropy etc…) were just as powerful as the more nonlinar and nonstationary methods. Indeed over that last couple of years, the momentum appears to have slowed with still no real answer to whether seizures can be predicted. Currently there are a number of new emerging theories including microseizures, high frequency activity and a high pre-seizure state. There is also an emerging perspective that detection and prediction need to go hand-in-hand, which was discussed in the Freieburg Seizure Prediction Workshop of 2007.

However, not much significant work has emerged from this and prediction is still an open question.

Seizure detection in most cases has followed a different path. The majority of literature has focused on Neonatal EEG. Seizures in neonates can be indications of Neonatal Encephalopathy (NE) - the manifestation of abnormal neonatal brain function - and can affect from 0.5-4 neonates per 1000. The incidence of NE can have devastating long-term effects, including cerebral palsy, blindness, deafness, autism and in 10-15% of cases can be fatal. As a result litigation claims for the NHS since 1996 number approximately £3 Billion. One of the issues with neonatal care in the area of seizure detection is that there doesn't exist an automated system for aiding non-expert clinicians in spotting seizure behaviour. Although the initial methods saw value in this observation their results were poor, even when retested with optimisations more recently [20]. State-of-art developments of these approaches have extended to algorithms that include time-frequency methods (wavelet transform [21]), information theory (entropy [22, 23]) and chaotic measures (Lypanov exponent [16]). Our interpretation of the results shows three main insights, which are key driving points for our approach.

Firstly, the characteristics of rhythmicity, frequency and complexity of the raw EEG signal are the three of the most important factors in quantifying EEG activity. Second is that although frequency pre-processing steps can be used as a primary measure they can also be a valuable step to enhancing other quantification measures. For instance entropy measures are best suited under low-noise conditions; hence denoising methods such as simple filtering to wavelet decomposition have proven to improve these results [20]. Thirdly, complex methods such as chaotic measures do not seem to improve the analysis, at least for

seizure detection. These types of methods are computationally expensive and rely on any chaotic nature of the EEG to be related to the seizure activity, which has not been shown.

The rhythmicity/frequency is related to spiking activity which can be associated with seizures while the signal complexity suggests quantification of gestational age and expected states of the neonate brain for that age (sleep/wake cycles). Conventional clinical evaluation of seizures in EEG is based on frequency and amplitude changes of particular frequency bands (such as theta and delta) and spiking activity. Spike activity and frequency band information have led to high sensitivity of detection of seizures but with poor specificity, showing false prediction rates (FPR) of 0.66-1.5 per hour [24, 25]. The highest sensitivity was based on an optimisation of 40-60 parameters across a neural network (91% sensitivity, 1.17 FDR) [26]. This method was very powerful in optimising relevant features for seizure detection in data sets of different abnormal background activity [26], with one of its successes being the pre-removal of artefacts prior to analysis. Other studies have been positive but have been very specific to the seizure itself and have not considered that it may be of clinical interest to detect and classify any activity that may be seizure related (e.g. inter/post ictal).

Below we outline in more detail the algorithms and methods used in the art of seizure prediction. This review is adapted and extended from [27].

3.4.2 Clinical EEG

Epileptologists use the frequency characteristics and direct observations to discern pathologies. Table 2.3 shows the typical characteristics that are noted in an EEG analysis.

In addition to these characteristics seizures are typically broken into different categories. Although there are currently several debates as to the best EEG and seizure classifications and an effort by the International League Against Epilepsy (ILAE) to re-

establish a classification norm [28]. Regardless, the most common classification is shown in Table 3.3. In addition we have partial seizures that may not result in loss of consciousness (this includes absence seizures) and complex partial, that are partial seizure that evolve to more generalized ones.

Table 3.2- Interpreting the EEG

Overall Amplitude Left/Right Symmetry Alpha/Beta/Theta Rhythm Generalised or Bilateral Delta Epileptiform Potentials Normal variants	Response to activation (photic and hyperventilation) Changes during sleep Correlates of notable behaviors Summary Comments

Table 3.3 - Typical seizure classifications

Generalised Seizures
Tonic-clonic
Absence
Typical/Atypical
Mycolonic
Tonic / Atonic
Clonic

Partial seizures are usually characterised by spike-wave burst patterns of 1-10 second durations. The period of these waves can help determine other types of syndromes, such as the Lennox-Gostaut syndrome which is typically characterized by slow spike-wave bursts. Seizures generally start with local EEG rhythms (e.g. alpha or theta waves), focal spiking or electrodecremental (low voltage and high frequency) activity. Tonic and clonic seizures contain muscle artifact with more high frequency and amplitude activity can be seen.

With waveform types, clinical symptoms, frequency and amplitude variations and spatial distribution we can discern relatively well the type of seizure it is likely to be. This is done with visual heuristics and can distinguish between non-epileptic seizures (such as

psychogenic seizures), artifacts (muscle, eye movement), partial/focal and generalised seizures.

Detection is done visually (by human eye), usually post-seizure event (back annotation on EEG records). There are some forms of seizure detection algorithms used for trial purposes but as yet there is no standard. The majority of predictive elements are for commercial and research purposes, none of which have been realized in clinics.

3.4.3 Seizure Prediction

Prediction of seizures using EEG-based signals has gone through several stages over the last several decades. Methods applied to EEG signals have evolved with science and technological developments. As methods were developed and computer infrastructures developed to allow these methods to be applied efficiently, nonlinear methods were introduced into the field. Since then a multitude of works in several disciplines has focused on specific measures, both linear and nonlinear in an attempt to distinguish three states surrounding an epileptic seizure.

These states are commonly referred to as pre-ictal, ictal and inter-ictal (ictal refers to seizure state). Excellent chronological reviews of the field of analysis on EEGs can be found in [5, 6, 9] among others. We also note, that some methods involve the use of multiple channels (multi-variate), looking at correlations and/or synchronicity in localized areas of the brain.

In the next section we review some of the basic methods of linear and nonlinear analyses that has been applied to EEG-based signals.

3.4.4 Linear Analysis

In the 1970 S. S. Viglione [29] carried out linear analysis on generalized (Grand-Mal) seizures. However, this was abandoned when the results were found to be inconclusive but more recently has been re-examined. We will discuss the results of these methods later in this section.

After the initial linear work on EEGs in the 1970s and 80s, nonlinear analysis was introduced. This then overshadowed any linear work and as such it is only recently that work has re-established the relevance of both types. Only by a comprehensive comparison between both linear and nonlinear methods can we establish the validity and statistical value of any algorithm. This is especially illustrated in the work of Mormann et al. [5] and McSharry et al. [30].

Part of the motivation of using linear methods is that in cases even though the general dynamics may be unpredictable, i.e. nonstationary and nonlinear, the feature/activity we are interested in may actually have linear characteristics or can be approximated as such. We note that the following methods are simply described, the equations and derivations can be found extensively described in [5, 27].

Moments

In a time series, such as the EEG, the mean is actually referred to as the arithmetic mean. This is also known as the *first moment* or expected value (denoted *E(X)* in most text books). The second moment (variance) is a measure of the deviation from this expected value. The third moment (skewness, χ) and fourth moment (kurtosis, κ) give other indications such as symmetry ($\chi = 0$ for symmetric waveforms) and peakedness/flatness of amplitudes [5, 31], [32].

These sorts of methods were applied based on the theory that seizures are as a result of hypersynchronous neuronal activity, which results in more periodic (stationary) and linear features that these moments can extract.

Hjorth Parameters: In the 1970's Hjorth [4] defined a set of parameters to be useful in clinical analysis of the EEG that are simple extensions of these classical moments we have just described. These measures are mobility, activity and complexity [5, 31]. Activity is proportional to the variance of a signal, complexity measures the variance of the slope changes in relation to an ideal sine wave, and mobility measures the variance of the slopes of the EEG normalized by the variance of the amplitude distributions. Mobility and complexity are estimated through the second and fourth statistical moment of the power spectrum.

Energy and Power

The energy of a signal is given by the sum of the square of its discreet values, which is essentially the variance. In [33] a measure was used which calculates the *accumulative energy*. Each window is summed in order to extract the evolution of energy around a seizure.

Power, a measure of energy as a function of time, can be quantified in several ways. There are two measures of power that have been proposed in the EEG literature. The first is to split the frequency components of the signal into its frequency bands (Table 3.1). Then calculate the *spectral band power* of these.

The second power measure we have is an extension to this, known as the *spectral edge frequency*, used by Stanski et al. in 1984 [34]. This is defined as the minimum frequency in which 50% of the power in the 40Hz bandwidth is contained [5, 31]. This frequency is

chosen since the majority of the frequency content of an EEG is believed to be within the 40Hz range.

Autocorrelation Function

The autocorrelation function [5] measures how correlated a signal is to itself at a later point in time. The theory is that the autocorrelation function will vary from A(0) (i.e. $\tau = 0$). Close to A(0), the autocorrelation function decays, while for large values of τ the function varies around 0. The slower the decay from A(0) the more linear correlations in the series. Hence a measure of the time of the first zero crossing of A(τ), can quantify the range of linear correlations in the signal.

Autoregressive modeling

The previous measures characterised statistical properties of a signal but another approach involves the use of modeling the signal. By assuming any point in the time series is a linear sum of previous values one can extrapolate a linear model for the system [35]. This is known as linear regression.

We can also include multiple channels of data [36]or a random process into the equation [5, 31]. These models are known as an autoregressive moving average (ARMA). Once modeled we can calculate the error between model and original signal to define the changes taking place, then use the z-transform of this error [35]. The evolution of the transfer poles, coefficients, prediction error and/or model order can then be analysed.

Linear cross-correlation

We mentioned that multiple channels can be used and measures between them extracted. The cross correlation is a standard metric used in many applications. Rosenblum [37] defined a measure that quantifies the similarity between two signals. The measure is the

maximum value of a normalized cross-correlation function [5]. Correlation will vary between 0 and 1, with 0 being no correlation between the two signals.

Coherence

Coherence is similar to the cross-correlation except we define the synchronization between the two signals, at a particular frequency rather than over the full signal.

Linear Summary

These methods are the basic linear measures applied to seizure prediction. Very few additions have been made in the last several years and those we have listed here have been extensively retested in [5, 32]. Undoubtedly, other linear measures, such as those illustrated in [38], which aim to work on both spatial (electrode relationships, neural sources etc..) and temporal (time-series of single channels) modeling can be used. What needs to happen, and is the reason for the retesting of so many of these prediction algorithms, is for a universal statistical testing method in which to test the results [5, 18]. This we will describe in later sections but for now we will describe the nonlinear methods applied to EEG signals.

3.4.5 Nonlinear Analysis

A nonlinear signal is any one that shows a higher order power relationship in the underlying equations that describe it. Many mechanisms can be classified nonlinear, some of which are illustrated in Figure 3.3.

	Number of Variables (n)				
	n=1	**n=2**	**n≥3**	**n>>1**	**Continuum**
Linear	*Growth, decay or equilibrium*	*Oscillations*		*Collective phenomena*	*Waves and patterns*
	Exponential growth	Linear oscillator	Civil engineering	Coupled harmonic oscillators	Elasticity
	RC circuit	Mass and spring	Electrical engineering	Solid-state physics	Wave equations
	Radioactive decay	RLC circuit		Molecular dynamics	Electromagnetism (Maxwell)
		2-body problem (Kepler, Newton)		Equilibrium statistical mechanics	Quantum mechanics (Schrödinger, Heisenberg, Dirac)
					Heat and diffusion
					Acoustics
					Viscous fluids
Nonlinear			*Chaos*		*Spatio-temporal complexity*
	Fixed points	Pendulum	Strange attractors (Lorenz)	Coupled nonlinear oscillators	Nonlinear waves (shocks, solitons)
	Bifurications	Anharmonic oscillators	3-body problem (Poincaré)	Lasers, nonlinear optics	Plasmas
	Overdamped systems, relaxational dynamics	Limit cycles	Chenical kinetics	Nonequilibrium statistical mechanics	Earthquakes
	Logistic equation for single species	Biological oscilaltors (neurons, heart cells)	Iterated maps (Feigenbaum)	Nonlinear solid-state physics (semiconductors)	General relativity (Einstein)
		Predator-prey cycles	Fractals (Mandelbrot)	Josephson arrays	Quantum field theory
		Nonlinear electronics (van der Pol, Josephson)	Forced nonlinear oscillators (Levinson, Smale)	Jheart cell synchronisation	Reaction-diffusion, biological and chemical waves
			Practical uses of chaos	Neural netwroks	Fibrillation
			Quantum chaos	Immune systems	Epilepsy
				Ecosystems	Turbulent fluids (Navier-Stokes)
				Economics	Life

Figure 3.3 - Table from Strogatz, [39] illustrating the classification of various natural mechanisms, including biological.

To extract features from a time-series we first translate it into a multi-dimensional space. From a single time-series or time-correlated multi-channel data we use either time- or spatial embedding. Time-embedding is a mapping of signal to a *d*-dimensional space, *D*, defined as $D(t) = x(t), x(t + \delta\tau),..., x(t + N\ \delta\tau)$ where $\delta\tau$ is some fraction of delayed time. Spatial embedding constructs this space as $D = x_1(t) + x_2(t) + ...+ x_N(t)$ where each of $x_i(t)$ represents the sample value of multi-channel data. While time-embedding extracts local dynamics (in relation to the source activity) spatial-embedding extracts more global brain dynamics [40].

Time-embedding assumes that the time-series portion is small enough to be considered stationary. Therefore τ and δ are typically determined through empirical methods or entropy based derivations leading to debate as to whether the true nature of the signal dynamics is being captured [22, 23].

With this understanding we can described the nonlinear algorithms used for prediction. The two foundation methods are the Lyapunov exponent (LE) and correlation dimension (CD). These and the foundations of their formation are the basis of several of the following algorithms. Sets of measures such as *local flow* and *loss of recurrence* are similar to these other measures but follow the flow of trajectories more than measures based on distinct points. The final sets of algorithms we shall consider look at spatial multi-channel patterns.

Lyapunov Exponent (LE)

This measure is described as an indication of how the signal flows in our constructed multi-dimensional space and how this flows has a sensitive dependence on the initial conditions of the system. This sensitivity produces flows on this space that exponentially diverge or converge from each other in a seemingly random way.

Any positive LE is an indication that chaotic mechanisms exist on the attractor (the name for the signal plot in the multi-dimenstional space). To measure it we have to quantify the exponential divergence of adjacent trajectories on the phase space. There are several methods in computing this from a time series for one channel [41]. The basic concept is that in each dimension of the attractor there will be a divergence that characterizes that principal axes (x, y, z, etc...).

Under particular conditions it has been shown that the largest LE gives accurate insight into the trajectory divergences. By calculating the largest LE, the average divergence is known.

A group led by Iasemidis et al. defines a bivariate (two channels) measure known as *dynamical entrainment*. Defining a vector Xi as the d-dimensional point in the phase space

they define the Largest Lyapunov exponent (STLmax). This measure is also extended to construct the second trajectory X(tj) from another time series for multi-channel data.

For the number of dimensions of the system there will be an equivalent number of exponents we can extract. The main drawback with calculating LEs is their sensitivity to noise and signal fluctuations [42, 43]. Hence, although LEs can be used as indication that chaotic behaviour exists in the system, under noisy conditions the results can be misleading.

Another interesting insight into progress of this measure, such as in [16] indicates that there is predictive power. However, the statistical validation carried out in [44, 45] considered the results insufficient. We will review these results later. The evolution of the LE measure from its first quantification is a good example of how a lot of the algorithms have progressed. Initially it was used with no knowledge of the system dynamics, but as time progressed it was apparent that some parameter optimisation was needed for specific EEG changes.

Correlation Dimension

Our second standard measure is the correlation dimension, which is an estimate of the dimension of attractor space (strange attractor in chaos). For strange attractors this dimension is not necessarily a rational number, such as the Lorenz system that has a correlation dimension of 2.06. The general name given to this dimension is *fractal* dimension.

The *correlation sum*, an essential foundation to the dimension calculation, first defined by [46] and corrected in [47] count the number of vectors in state space that are within a hypersphere (a sphere that goes through all dimensions) of radius *r*.

Using this measure we determine the dimension from the slope of the correlation sum. From the above it would appear that we would need an infinite set of data points, N. However Lehnertz and Elger [17] showed how an effective estimate of the correlation dimension can be found when an almost constant range of values is observed.

In relation to prediction there has been some speculation as to the ability of this computation to quantify any brain dynamical changes preceding a seizure. In fact the results on EEG analysis first cited by [17] have been found to be questionable in [48]. They use the same computational procedure but have variations in normalising the correlation sum.

The above two measures, correlation dimension and LE, are probably the most well known measures in the field of chaos and the ones described now are derived from these.

Similarity Index

A similarity index was developed by Le Van Quyen et al. in 1999 [49]. In it a time embedding of a predefined reference window is made and then is compared with time windows on the EEG through a cross-correlation sum.

Their first task was to minimise data usage as time-embedding and correlation sums for a large embedding dimension is computationally intensive. To do this they use the positive threshold crossings (such as in quantisation) of a signal as time markers and measure the time intervals between two successive crossings. This method is applied to test windows and compared to a reference. There have been doubts raised in these results from [50, 51] in response to some of these results. However, the doubts were defended as being due to different testing procedures, including filtering and computation of the algorithm.

An improvement in the result of this algorithm has been published that uses an initial wavelet denoising [52]. They use a slightly modified correlation sum that uses a Gaussian

function instead of the Heavyside one. This allows for a softer boundary and as such changes will be smoother in calculating the similarity index.

Dissimilarity

Another measure, closely related again to the correlation sum is the dissimilarity measure. As in the similarity index a reference window is chosen and a probability density function is defined for each window based on it a multi-dimensional phase space representation. Effectively a similar scheme is used on the test and reference window as in the similarity index but instead other measures are used, the $\chi 2$ test and L1 norm [53]. The $\chi 2$ test quantifies frequency discrepancies between the two signal while L1 distance is related to the total invariant measure on the attractor [53].

Marginal Predictability

Another measure that is a derivation from the correlation sum is the marginal predictability [54]. Intuitively, this measure is the conditional probability that given *d* points in the hypersphere of radius *r*, the *(d + 1)* point will also be in this sphere. According to [54] this measures how much additional predictive information is contained in the data point *(d+1)* compared with the prediction information of the previous *d* samples, reflecting in some way the Markovian nature of the signal.

Local Flow

Used by [64] and additionally described in [5] the local flow looks at how trajectories pass through overlapping hypercubes in the multi-dimensional space. The j^{th} hypercube will be passed n times by a given trajectory. Each of these vectors is summed and normalized by the number of passes.

This measure allows us to distinguish between a nonlinear deterministic flow and a linear stochastic one. The thinking is that in a deterministic system there will be trajectories in the state-space close to each other and moving in the same direction. While in a random system, such closeness of trajectories is not present.

Algorithmic Complexity

This is a measure, described by Mormann et al. [31], that is based on symbolic dynamics. The time series is transformed into a symbol sequence by partitioning the sampling values and denoting each by a symbol that denotes its interval (analogous to quantisation). Each value of the time series is then replaced by its corresponding symbol. The resulting symbol sequence is then investigated for its complexity by some manner (in [44] the Lempel- Ziv parsing was used [55]). Finally, the algorithmic complexity is defined as this value normalized by the expected asymptotic value of such a random sequence of symbols.

Recurrence Loss

This measures the degree of nonstationarity of the system which is not a straightforward task. Rieke et al. [56] describes this algorithm by considering the state-space, constructed through time-embedding. Vectors that are the nearest neighbours in this space are extracted and counted.

This is then compared with an expected time-lag if the system was assumed stationary and the difference is compared using a distribution function. So then the measure reflects the systems loss of recurrence of the signal to a state-space area over time. The *stationary* distribution function interpolated from the reference signal helps quantifies this.

Interdependence

For multichannel or bivariate data analyses we consider nonlinear correlation measures. First is the *interdependency* of two EEG signals. Consider two time-series signals xi and yi which are both time-embedded to generate their state space, *xn* and *yn* respectively.

We define the points which are considered nearest neighbours of *xn* and *yn* and denote the time indices of these *k* points as *rn,j* and *sn,j* respectively (j = 1, ..., k). Four correlations are quantified that measure the interdependence of *xn* to its nearest neighbours and to *yn* based on the distance between these points. The same is done for *yn* to itself and *xn*.

With this definition insight into the dependence of the two signals can be extracted. A good discussion of the various types of measures and adaptations in the way these are defined can be found in [57].

Phase Synchronization

The final measure we shall consider in detail is in relation to synchronisation. Synchronization in systems has been identified for many hundreds of years starting with Hyugens in 1673 who noticed synchronization of clocks on his wall [5, 6].

There is a lot of interest in this topic because of the insight it can give to the coupling between different signals. There is a lot of work to study synchronization as a general topic [6] but the work we shall consider here is EEG related.

In the application of EEG signals, the type of synchronization that has been of interest is *phase synchronisation*. The three described in [5] are the ones we will refer to here and then we will describe some other concepts of synchronisation from general sources. If two oscillatory systems are phase locked, then there is a constant phase difference between them. In [58] a measure called the *mean phase coherence* is used.

Under a phase locking condition coherence will reach the value 1 when unsynchronized and 0 when it is. Another measure used by Rosenblum et al. [37] is based on the Shannon entropy (the uncertainty of a random variable).

Nonlinear Summary

There are several ways to quantify the nonlinearity of single and multi-variate signals. Indeed there are a lot of techniques, both statistical and algorithmic, which have not been mentioned here, such as nonlinear forecasting and mutual dimension measures [6]. Most of the algorithms and measures we have described are relevant as they have been applied to EEG-based signals and more specifically seizure prediction.

There are several methods that have emerged over the last several years, including cross-frequency coupling [59], patient specific bivariate-synchony [60], rule-based system [61] to new studies showing other features in EEG-based signals, including high frequency oscillations [62] and studies of cortical excitability using transmagnetic stimulation [63] and more are emerging all the time.

Within the prediction field there have been a number of testing and retesting of the above algorithms on different data sets. Although not all results are illustrated here, sufficient summaries will be made based on the most significant findings of these works. To understand them we must first look at the methods that quantify their accuracy.

3.4.6 Statistical Measures

There are several measures used to quantify the ability of an algorithm to detect specific features of data. These are summarised in the following table:

Table 3.4 - Table describing the typical metrics used in prediction results.

		Condition		
		True	**False**	
Test Result	**Positive**	True Positive (TP)	False Positive (FP)	*= Positive Predictive value*
	Negative	False Negative (FN)	True Negative (TN)	*= Negative predictive value*
		= Sensitivity	*= Specificity*	

1. **Sensitivity** - This is a measure of $TP/TP+FN$ which highlights how many events (i.e. seizures) of the total that are considered important (true) were detected.

2. **Specificity** – This measures $TN/TN+FP$ which quantifies the algorithms ability to identify areas where no events (i.e. seizures) occurred.

3. **False prediction/detection rate** (FPR/FDR) measures the number of false predictions/detections made within a time interval, e.g. 1/h (false prediction / hour). This is sometimes used instead of or to complement specificity, as with specificity it is difficult to assess true negatives and specificity does not track linearly with data size.

Surrogate Data Analysis

In an attempt to further validate the above measures, a surrogate time-series is constructed (see [18] and references therein). Here, with EEG data, the seizures, or detectable events are left where they are and all other data is removed. This data is then replaced with data from the same patient but during the interictal (between seizure) period but with similar statistics (mean, variance etc…). The above measures are re-evaluated on this new series, and if performance is poor compared with the original data set then the methodology/algorithm under test is further validated.

Prediction Specific

In prediction, you not only want to determine if a seizure will occur but when one would expect it to. In an effort to aid prediction validation, in 2003 a framework was established to compare and assess prediction methods [50]. The key aspect of this framework is a defined *seizure prediction horizon* (SPH) and a *seizure occurrence period* (SOP). As illustrated in Figure A2, the alarm indicates that in a certain period of time (SPH), the patient will enter the SOP, a period in which a seizure may occur.

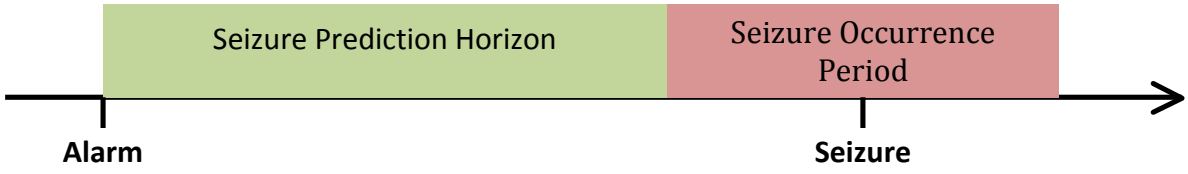


Figure 3.4 - Figure highlighting the seizure prediction and occurrence areas.

In addition to this recent work has derived measures of determining what a random predictor should be able to achieve given a specific SOP and SPH [18, 19]. Here a binomial probability defined for the number of independent features (r) and electrodes/channels (n) for prediction k of N seizures:

$$P_{binom}(k;N;P) = 1 - \left(\sum_{j<k} \begin{pmatrix} N \\ J \end{pmatrix} P^j (1-P)^{N-j} \right)^d \quad (1)$$

where, d is the number of independent features given as ranging from *1* to $d_{max,r}(n) = \binom{n}{r}$. The upper and lower sensitivity given by this range of d is determined as $S = max_k\left(P_{binom,d}(k;N;P) > \alpha\right) * 100\%$ where α is a significance level of 0.05. P is a probability defined as the *SOP*SPH*.

3.4.7 Results

Prior to and including 2006 there were a number of papers aiming to analyse the current state of the art methods (those we have previously described) and determine their predictive power. These include those of Mormann et al. in 2005 [31], Winterhalder et al. 2003 [50], Mormann et al. again in 2006 [5] among many smaller studies.

This was all prior to the observation that all methods need to be studies with statistical analyses to validate the strength of them versus chance. Mormann et al. of 2005 studies a variety of linear and nonlinear measures using surrogate data to validate them.

The data used in this work compromised 5 patients (of which 4 were used for surrogate data analysis) with 46 seizures over a 311 hours of recording time. The data were tested under 4 categories, to distinguish the measures' ability to determine a preictal state based on single channel measures for each seizure separately and universally.

Two methods were used to assess the data. First, a threshold that distinguishes preictal and interictal is defined. This threshold would then have to be consistent for all seizures analysed. Secondly, given the number of electrode sites and seizures different evaluations can be made based on the following four schemes: (1) All seizures are analysed, (2) Each channel was tested separately, (3) Each seizure is analysed separately on each

channel and (4) Is the same as in scheme 3 except the preictal period is only compared with the previous interictal state. This is to account for any baseline shift that may be present in the signals. Finally tests were carried out with different preictal periods (5/30/120/240 minutes) and using an ROC (sensitivity vs. specificity) curve preictal and ictal states.

The conclusions from these results are that linear and nonlinear measures perform just as well as each other. Also univariate measures are able to distinguish changes for a shorter preictal period (5 mins) compared with more significant changes measured by bivariate measures for larger periods (240 mins).

With the introduction of the seizure prediction horizon a more extensive assessment can now be made. Winterhalder et al who originally proposed it determined that both similarity index, accumulated energy and correlation dimension did not have significant predictive value. Mormann et al continued their assessment in 2006 indicating that most methods have no statistical validation so could not be differentiated from chance.

This has led to studies that counter previous claims. For example Harrison et al. in two papers related to correlation dimension [48] and accumulated energy [64]. The original work on correlation dimension claimed that a preictal state is detectable with a prediction time of up to 12 minutes. In Harrison et al. [48] the changes found in these signals that the previous studies revealed was a result of the time-frequency characteristics of the signal. Using surrogate data analysis they showed that methods based on the correlation sum, such as the dimension and similarity index, provide no real prediction for these signals.

Some of the more recent studies have employed the statistical methods we have previously described [65, 66]. In these they determine the significance of seizure prediction based on a binomial probability of seizures being predicted. Aarabi et al. [65] recently showed that predicting at 10 seconds prior to the seizure onset zone leads to a prediction of

90%. Although this is with a 50 minutes seizure occurrence period and it is not clear where the seizure occurs relative to this start point. In [66] a comparison between subclinical and clinical seizures is made showing sensitivities of 20-40% with not many several exceeding the lower and upper critical sensitivity values.

3.4.8 Seizure Prediction: Conclusions

In the previous sections we have looked at a number of methods applied to EEG-based signals with the aim of predicting epileptic seizures. We have looked at the EEG signal as a time-series of a single channel in which we apply linear and nonlinear measures.

With EEG signals there is no guarantee that this assumption holds true. If it were, then periodic solutions would exist and as such this whole work would be solved. But we don't have a periodic solution, nor can we assume that a full sinusoidal cycle will occur (which is needed to assume a signal contains a particular frequency in fourier analysis). Unless we are talking about instantaneous frequencies of course, then we can make such assumptions (we will discuss this in later chapters).

What we have presented here is by no means the full spectrum of work carried out on the area of brain dynamics. Instead, we chose here very specific pieces of work to illustrate the topic of seizure prediction. Of course the explanation of the epileptic process and the population statistics was important in justifying the motivation of this work.

The algorithms (linear and nonlinear) described have mostly been applied to EEG signals with the aim being to determine and distinguish states within the brain surrounding an epileptic event (seizure and nonseizure).

What is clear is that statistical methods need to be applied to all methods to truly quantify their predictive capabilities. A few recent progresses have helped this cause. First

is the work describing the generally accepted methods to apply [5, 18, 50]. Second is the understanding that a universal database, well annotated and described. Although not able to be utilised in this project a database has just been come available as part of an EU project (http://epilepsy-database.eu/). Similar work is being pursued with the Mayo Clinic and Brian Litt in the USA (http://mayoresearch.mayo.edu/).

3.5 Conclusions

This chapter as a whole has aimed to give the reader insight into how epileptic seizure forecasting is addressed both in clinical and academic environments. Although only a snapshot of the field it gives us insight into the complex ever growing world of seizure prediction.

Interestingly, in the large studies of seizure prediction, we have seen that results favour a patient-specific approach. The results in those cases outweighed any generalised trends. In retrospect we should probably not be surprised given the pure complexity of the human brain.

Whether seizures can be solved so that the millions of people suffering from it can be given an alternative treatment employing technology depends wholly on good research outcomes. As we have described in Chapter 2, this also depends on the ability to integrate computing, engineering and biology into a framework that support this type of research.

As a first step in this work, we next present a novel method that has been designed to analyse biopotential signals and extract dynamics that can then be used for seizure detection and prediction.

3.6 References

[1] P. Krack et al, "Five-Year Follow-up of Bilateral Stimulation of the Subthalamic Nucleus in Advanced Parkinson's Disease," *The New England Journal of Medicine,* vol. 349, pp. 1925-1934, 2003.

[2] W. H. Organisation. (2006). *Neurological Disorders: Public Health Challenges*. Available: http://www.who.int/mental_health/neurology/neurodiso/en/index.html

[3] E. S. Goldensohn, R. J. Porter, and P. A. Schwartzkroin, "Epilepsy History - 50th Anniversary of the American Epilepsy Society," *Epilepsia,* vol. 38, p. 150, 1997.

[4] J. Malmivuo and R. Plonsey, *Bioelectomagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*, 1995.

[5] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure Prediction: the long and winding road," *Brain Advance Access,* pp. 314-333, 2006.

[6] C. J. Stam, "Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field," *Clinical Neurophysiology,* vol. 116, pp. 2266-301, 2005.

[7] *The JOHN HOPKINS Atlas of Digital EEG*, 2006.

[8] R. Aschenbrenner-Scheibe, T. Maiwald, M. Winterhalder, H. U. Voss, J. Trimmer, and A. Schulze-Bonhage, "How well can epileptic seizures be predicted? An evaluation of a nonlinear method," *Brain,* vol. 126, p. 11, 2003.

[9] L. D. Iasemidis, "Epileptic seizure prediction and control," *IEEE Tran Bio Eng,* vol. 50, pp. 549-558, 2003.

[10] A. Aarabi, F. Wallois, and R. Grebe, "Automated neonatal seizure detection: A multistage classification system through feature selection based on relevance and redundancy analysis," *Clinical Neurophysiology,* vol. 117, pp. 328-40, 2006.

[11] P. Celka and P. Colditz, "A Computer-Aided Detection of EEG Seizures in Infants: A Singular-Spectrum Apparoach and Performance Comparison," *IEEE Transactions of Biomedical Engineering,* vol. 49, pp. 455-62, 2002.

[12] S. Faul, G. Boylan, S. Connolly, L. Marnane, and G. Lightbody, "An evaluation of automated seizure detection methods," *Clinical Neurophysiology,* vol. 116, pp. 1533-41, 2005.

[13] G. A. Stein, G. H. Eder, E. D. Blum, A. Drachev, and S. R. Fisher, "An automated drug delievery system for focal epilepsy," *Epilepsy Research,* vol. 39, pp. 103-14, 2000.

[14] W. M. Hill, M. Wong, A. Amarakone, and M. S. Rothman, "Rapid Cooling Aborts Seizure-Like Activity in Rodent Hippocampal-Entorhinal Slices," *Epilepsia,* vol. 41, pp. 1241-8, 2000.

[15] I. Osorio, G. M. Frei, and et al., "Automated Seizure Abatement in Humans Using Electrical Stimulation," *Annals of neurology,* vol. 57, pp. 258-68, 2000.

[16] W. Chaovalitwongse, L. D. Iasemidis, P. M. Pardalos, P. R. Carney, D. S. Shiau, and J. C. Sachellares, "Performance of a seizure warning algorithm based on the dynamics of intracranial EEG," *Epilepsy Research,* vol. 64, pp. 93-115, 2005.

[17] K. Lehnertz and C. E. Elger, "Can epileptic seizures be predicted? evidence from nonlinear time series analysis of brain activity," *Physical Review Letters,* vol. 80, pp. 5019-22, 1998.

[18] H. Feldwisch-Drentrup, A. Schulze-Bonhage, J. Timmer, and B. Schelter, "Statistical validation of event predictors: A comparative study based on the field of seizure prediction," *Physical Review E,* vol. 83, p. 066704, 2011.

[19] M. Winterhalder, B. Schelter, T. Maiwald, A. Brandt, A. Schad, A. Schulze-Bonhage, and J. Timmer, "Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction," *Chaos,* vol. 16, p. 013108, 2006.

[20] S. Janjarasjitt, M. S. Scher, and K. A. Loparo, "Nonlinear dynamical analysis of the neonatal EEG time series: The relationship between sleep state and complexity," *Clinical Neurophisiology,* vol. 119, pp. 1812-1823, 2008.

[21] S. Janjarasjitt, M. S. Scher, and K. A. Loparo, "Nonlinear dynamical analysis of the neonatal EEG time series: The relationship between sleep state and complexity," *Clinical Neurophisiology,* vol. 119, 2008.

[22] L. Pezard, J.-P. Lachaux, N. Thomasson, and J. Mertineri, "Why bother to spatially embed EEG? comments on Pritchard er al," *Psychophysiology,* vol. 36, pp. 527-531, 1999.

[23] W. S. Pritchard, "On the validity of spatial embedding: A reply to Pezard et al," *Psychophysiology,* vol. 36, pp. 532-5, 1999.

[24] W. Deburchgraeve, P. J. Cherian, M. De Vos, R. M. Swarte, J. H. Blok, G. H. Visser, and P. Goavert, "Automated neonatal seizure detection mimicking a human observer reading EEG," *Clinical Neurophisiology,* vol. 119, pp. 2447-54, 2008.

[25] J. Rennie and et al., "Non-expert Use of the Cerebral Function Monitor for Neonatal Seizure Detection," *Arch. Dis. Child fetal neonatal Edition,* pp. F37-F40, 2004.

[26] S. Faul, G. Boylan, S. Connolly, L. Marnane, and G. Lightbody, "Chaos Theory: Analysis Of The Newborn EEG - Is It Worth The Wait," *Workshop on Intelligent Signal Processing (WISP), IEEE,* pp. 381-6, 2005.

[27] A. Eftekhar, "Hilbert-Hunag Transform: biosignal analysis and practical implemenation," PhD Thesis, Electrical and Electronic Engineering, Imperial College London, London, 2010.

[28] A. T. Berg et al, "Revised terminology and concepts for organization of seizures and epilepsies: Report of the ILAE Commission on Classification and Terminology, 2005–2009," *Epilepsia,* vol. 51, pp. 676-685, 2010.

[29] S. S. Viglione, "Applications of pattern recognition technology," in *Adaptive, learning, and pattern recognition systems theory and applications*, J. M. Mendel and K. S. Fu, Eds., ed New York: Academic Press, 1970, pp. 115-162.

[30] P. E. McSharry, L. A. Smith, and L. Tarassenko, "Prediction of epileptic seizures: are nonlinear methods relevant?," *Natural Medicine,* vol. 9, pp. 241-2, 2003.

[31] F. Mormann, T. Kreuz, C. Rieke, R. G. Andrzejack, A. Kraskov, P. Davi, C. E. Elger, and K. Lehnertz, "On the predictability of epileptic seizures," *Clin Neurophys,* vol. 116, pp. 569-587, 2005.

[32] R. G. Andrzejack, F. Mormann, G. Widman, T. Kreuz, C. E. Elger, and K. Lehnertz, "Improved spatial characterization of the epileptic brain by focusing on nonlinearity," *Epilepsy Research,* vol. 69, pp. 30-44, 2006.

[33] B. Litt, R. Estelle, J. Echauz, M. D'Alessandro, R. Shor, and T. E. A. Henry, "Epileptic seizures may begin hours in advance of clinical onset: a report of five patients,"" *Neuron,* vol. 30, pp. 51-64, 2001.

[34] D. R. Stanski, R. J. Hudson, T. D. Homer, L. G. Saidman, and E. Meathe, "Pharmacodynamic modeling of thiopental anesthesia," *Journal of Pharmacokinetics and Biopharmaceutics,* vol. 12, pp. 223-240, 1984.

[35] Z. Rogowski, I. Gath, and E. Bental, "On the prediction of epileptic seizures," *Biol. Cybern.,* vol. 42, pp. 223-240, 1981.

[36] Y. Salant and I. Gath, "Prediction of epileptic seizures from two-channel EEG," *Med Bio Eng Comput,* vol. 36, pp. 549-556, 1998.

[37] M. Rosenblum, A. Pkovsky, J. Kurths, C. Schafer, and P. A. Tass, *Phase Synchronizaion: from theory to data anlaysis*: Elsevier, 2001.

[38] L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda, "Recipes for the linear analysis of EEG," *NeuroImage,* vol. 28, pp. 326-341, 2005.

[39] S. H. Strogatz, *Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry, And Engineering*: Westview Press, 1994.

[40] J.-P. Lachaux, L. Pezard, L. Garnero, C. Pelte, B. Renault, F. J. Varela, and J. Martinerie, "Spatial extension of brain activity fools the single-channel reconstruction of EEG dynamics," *Human Brain Mapping,* vol. 5, pp. 26-47, 1997.

[41] M. T. Rosenstein, J. J. Collins, and C. J. De Luca, "A practical method for calulating the largest lyapunov exponents from small data sets," *Physica D,* vol. 65, pp. 117-134, 1993.

[42] M. Barahona and C.-S. Poon, "Detection of nonlinear dynamics in short, noisy time series," *Nature,* vol. 381, pp. 215-7, 1996.

[43] C.-S. Poon and M. Barahona, "Titration of chaos with added noise," *PNAS,* vol. 98, pp. 7102-7112, 2001.

[44] F. Mormann, C. E. Elger, and K. Lehnertz, "Letter to editor on: Performance of a seizure warning algorithm based on the dynamics of intracranial EEG," *Epilepsy Research,* vol. 71, pp. 241-242, 2006.

[45] M. Winterhalder, B. Schelter, A. Schulze-Bonhage, and J. Trimmer, "Comments on ”performance of a seizure warning algorithm based on the dynamics of intracranial EEG” by mormann et al," *Epilepsy Research,* vol. 71, pp. 80-81, 2006.

[46] P. Grassberger and I. Procaccia, "Characterization of strange attractors," *Physical Review Letters,* vol. 50, pp. 346-9, 1983.

[47] J. Theiler, "Spurious dimension from correlation algorithms applied to limited time-series data," *Physical Review A,* vol. 34, pp. 2427-2432, 1986.

[48] M. A. Harrison, I. Osorio, M. G. Frei, S. Asuri, and Y.-C. Lai, "Correlation dimension and integral do not predict epileptic seizures," *Chaos,* vol. 15, p. 33106, 2005.

[49] 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," *Computational Neuroscience,* vol. 10, pp. 2149-2155, 1999.

[50] M. Winterhalder, T. Maiwald, H. U. Voss, R. Aschenbrenner-Scheibe, J. Trimmer, and A. Schulze-Bonhage, "The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods," *Epilepsy Behaviour,* vol. 3, pp. 318-325, 2003.

[51] W. De Clercq, P. Lemmerling, S. Van Huffel, and S. Van Paescshen, "Anticipation of epileptic seizures from standard EEG recordings," *The Lancet,* vol. 361, p. 970, 2003.

[52] G. Ouyang, X. Li, Y. Li, and X. Guan, "Application of wavelet-based similarity analysis to epileptic seizure prediction," *Computers in Biology and Medicine,* vol. 37, pp. 430-7, 2007.

[53] L. M. Hively and V. A. G. Protopopescu, P. C., "Timely detection of dynamical change in scalp EEG signals," *Chaos,* vol. 10, pp. 864-75, 2000.

[54] D. Li, W. Zhou, I. Drury, and R. Savit, "Linear and nonlinear measures and seizure anticipation in temporal lobe epilepsy," *J. Computational Neuroscience,* vol. 15, pp. 335-45, 2003.

[55] A. Lempel and J. Ziv, "On the complexity of finite sequences," *IEEE Trans Information Theory,* vol. 22, pp. 75-81, 1976.

[56] C. Rieke, K. Sternickel, R. G. Andrezejak, F. Mormann, P. David, and K. Lehnertz, "Measuring nonstaionarity by analyzing the loss of recurrence in dynamical systems," *Physical Review Letters,* vol. 88, p. 244102, 2002.

[57] J. Arnhold, P. Grassberger, K. Lehnertz, and C. E. Elger, "A robust method for detecting interdependencies: application to intracranially recorded EEG," *Physica D,* vol. 134, pp. 00140-2, 1999.

[58] F. Mormann, K. Lehnertz, P. David, and C. E. Elger, "Mean phase coherence as a measure for phase synchronisation and its application to the EEG of epilepsy patients," *Physica D,* vol. 144, pp. 358-369, 2000.

[59] C. Alvarado-Rojas, M. Valderrama, A. Witon, V. Navarro, and M. L. Van Quyen, "Probing cortical excitability using cross-frequency coupling in intracranial EEG recordings: A new method for seizure prediction," *IEEE EMBC,* pp. 1632-5, 2011.

[60] L. Kuhlmann, D. Freestone, A. Lai, A. N. Burkitt, K. Fuller, B. Grayden, L. Seiderer, S. Vogrin, I. M. Y. Mareels, and M. J. Cook, "Patient-specific bivariate-synchrony-based seizure prediction for short prediction horizons," *Epilepsy Research,* vol. 91, pp. 214-231, 2010.

[61] A. Arabi and B. He, "A rule-based seizure prediction method for focal neocortical epilepsy," *Clinical Neurophysiology,* vol. In Press, pp. 1111-22, 2012.

[62] M. Zijlmans, P. Jiruska, R. Zelmann, F. S. S. Leijten, J. G. R. Jefferys, and J. Gotman, "High-frequency oscillations as a new biomarker in epilepsy," *Animals of Neurology,* vol. 71, pp. 169-178, 2012.

[63] R. A. B. Badawy, R. A. L. Macdonell, S. F. Berkovic, M. R. Newton, and G. D. Jackson, "Predicting seizure control: Cortical excitability and antiepileptic medication," *American Neurological Association,* vol. 67, pp. 64-73, 2010.

[64] M. A. Harrison, M. G. Frei, and I. Osorio, "Accumulated energy revisited," *Clinical Neurophysiology,* vol. 116, pp. 527-531, 2005.

[65] A. Aarabi and B. He, "A rule-based seizure prediction method for focal neocortical epilepsy," *Clinical Neurophysiology,* vol. 1, pp. 1111-1122, 2012.

[66] H. Feldwisch-Drentrup, M. Ihle, M. Le Van Quyen, C. Teizeira, A. Dourado, J. Timmer, F. Sales, V. Navarro, A. Schulze-Bonhage, and B. Schelter, "Anticipating the unobserved: Prediction of subclinical seizures," *Epilepsy & Behavior,* vol. 22, pp. S119-S126, 2011.

Chapter 4 - WiNAM: A novel tool for analyzing biosignals

In the previous chapters we discussed our objectives in defining a system that is (1) a patient specific analysis of patients and (2) a system that integrates an ICT infrastructure to support such an analysis.

Towards this, in this chapter, we describe a novel algorithm applied it to seizure prediction and detection. Here we shall first describe the original methods from which the algorithm was developed, followed by the extensions to it we have implemented to make it applicable to time-series data.

4.1 Pattern Recognition

Any system designed to classify data (such as seizure/non-seizure) has three main steps: (1) data acquisition, (2) data representation and (3) decision making [1].

One of the more popular methods of pattern recognition is a statistical approach where features of the data are extracted through methods such as Linear Discriminant Analysis or Independent Component Analysis (ICA) and then selected according to some criteria. Once extracted classification methods are used, such as template matching or nearest neighbour models. An excellent review of all these methods and the future perspective on this topic see [2] and references therein. The method that we originally took inspiration from is called an ngram.

4.2 Ngrams

Traditionally ngrams have been applied to language models. An ngram model extracts and counts the subsequences of a particular symbolic sequence. These subsequences or patterns can be directly translated to probabilities of words occurring given the previous *n*-

1 words [3, 4]. Hence ngrams can give direct insight into the Markovian nature of the sequence. For language learning models this is useful as a language can be modelled as the probability of a word occurring given a certain subset of previous ones.

Ngrams are derived from the original work of C. Shannon, who used the ngram approach of symbolic dynamics, entropy and complexity measures for the analysis of time-series, which is now widely used [5]. However, little attention has been given to an ngram approach for pattern recognition; applying probability and significance measures based on found patterns to quantify change in time-series dynamics.

Instead, a number of works have derived extensions of the Shannon entropy through symbolic dynamics and phase space reconstruction [6] in an effort to define complexity measures [7] and track signal dynamics. The most extensive being in the use of entropy and probability distributions in language series partitioning and segmentation [8]. The latter using ngram data mining to define sequence probabilities.

As we mentioned, typically, the target for ngram based approaches is language based, aiming to segment non-separated character sequences into meaningful sentences [8] regardless of language. In this work we consider an alternative method for using ngram-based analyses for pattern recognition in time-series analyses.

To understand then how we can apply this approach to time-series we need to first transform the data into a symbolic representation.

4.2.1 Symbolic Representation

Data can be compressed and represented in a symbolic fashion in one of many ways. We have mentioned some that look at phase space reconstruction but given the doubts of

embedding methods we chose one that compressed the signal in the one dimensional space. The most popular one being quantization. Quantisation here defined as:

$$Q(x_i) = \mathrm{sgn}(x_i) * \left\lfloor \frac{|x_i|}{\Delta} + \frac{1}{2} \right\rfloor \quad (1)$$

with $\Delta = {V_{ref}^{+} - V_{ref}^{-}}/{2^N}$ where V_{ref}^{+} and V_{ref}^{-} are the analog-to-digital reference voltages and N the bit level representation. Finer resolution can be achieved by having a non-binary divisor. This compressed signal can be represented in hexadecimal, hence in a symbolic way. Although we note, this is not necessary for the remaining methodology, it simply aids in showing the analogy to symbolic pattern recognition and the ngram approach. Also, biological data is generally acquired through some form of data acquisition using an analog-to-digital converter. The conversion, depending upon the application can result in 8-16bit data which when stored for software this information is represented at its lowest level in binary, but in a higher level of abstraction in hexadecimal (HEX). Hence, the data is already available in an alphanumeric format [9] (see Appendix C - Conference and Journal Paper).

With EEG signals one typically measures a signal of up to 100μV or 1-2mV [10] if measured using electrocorticogram (ECoG, surface of the brain). An example of quantization of an EEG signal is shown in Figure 4.1.

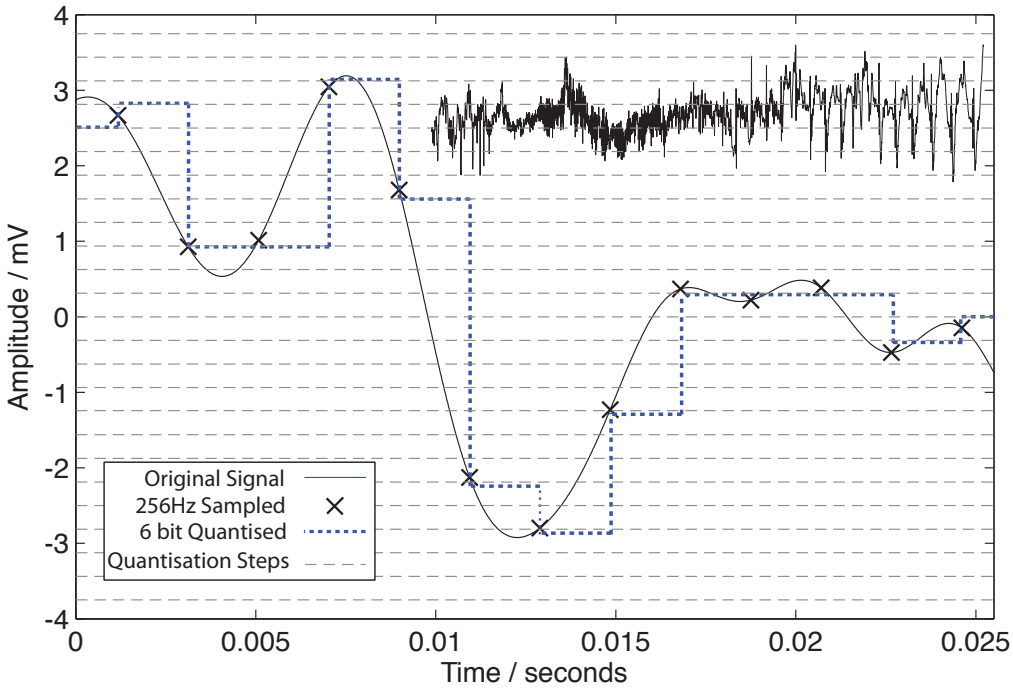


Figure 4.1 – Figure showing a sample EEG data set re-quantised from its original 16 bit to 6 bit using our outlined methodology.

The algorithmic methods is summarised in Figure 4.2 with the addition of a divide down factor (first block in pre-processing). Here we re-quantise the data at a lower bit level in order to minimize noise contributions. We will discuss the implications of this later. So an *N = 16bit* number could be divided down by *D = 8* bits resulting in an *8* bit number.

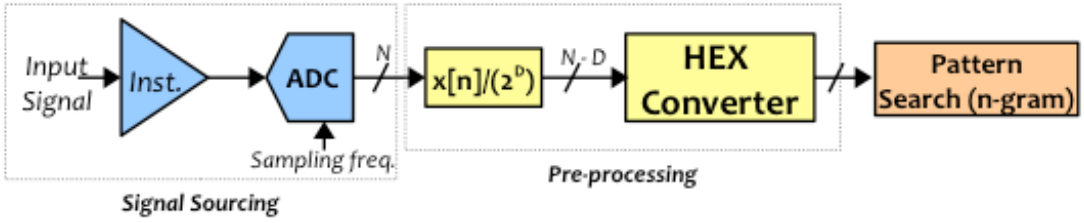


Figure 4.2 – Overview of how the Ngram technique is used with real-data

This will result in a list of HEX numbers each one of which represents a decimal value and can be used in the ngram pattern search. An example sequence is shown in Table 4.1.

Table 4.1 - Example of transition from sampled value to HEX value.

Sampled Value (+/- 10mV, 16bit)	16 bit value (signed)	8 bit value (signed)	HEX Value
-0.9726	-3187	-12	F4
-0.9726	-3187	-12	F4
-0.9067	-2971	-12	F4
-0.7761	-2543	-10	F6
-0.5942	-1947	-8	F8
-0.3815	-1250	-5	FB
-0.2863	-938	-4	FC

This table of values would result in the ngram of size 2 would count as:

Pattern	Count
F4 F4	2
F4 F6	1
F6 F8	1
etc…	…

With the time-series now in a symbolic representation we apply our ngram methodology to it, analogous to the time series being a language and then deriving sentences or word sequences from it.

4.2.2 Multiresolution Ngram

An ngram will produce a tree representing individual sequences and their counts. However, it by no means gives unique sequences but all possible combinations. This gives us some idea of the pattern spread but not on how the sequence should be separated into a meaningful one. This is analogous to segmenting words in a sentence into one defined as meaningful [8].

We periodically break a signal into segments (realizing a real-time analysis tool) and analyse each segments patterns and associated counts to determine the evolving significance of patterns in time segments of data. We construct a single one dimensional

ngram table, implemented using a combination of Matlab v7.12 and the ngram statistical package [77] for each data segment. A pattern is considered if it's count is beyond a particular threshold, *Th*. The total count for the ith time window of significant patterns are then calculated as:

$$CNT_i = (l(c(s_1^n), c(s_2^n), ..., c(s_k^n)) > Th)_{WIN_i} \quad (2)$$

where $c(s_1^n)$ is the count of symbol sequence s_1 of ngram length n, $l(...)$ is the length of the vectors contained within it and WIN_i is the i^{th} window.

In this simplified form we have no indication of the pattern sequence nor it's temporal evolution between the $(i-1)^{th}$ and i^{th} window. It will however, capture changes of periodicity within a signal over short time frames, details of which are dependant on Δ and ngram length, n.

This is shown in the same sample EEG signal shown in Fig. Here the seizure is characterised by the first 3 seconds of delta spikes (0-4Hz) then electrodecremental activity (low amplitude high frequency for 7 seconds) followed by larger delta spike burst for a further 10 seconds.

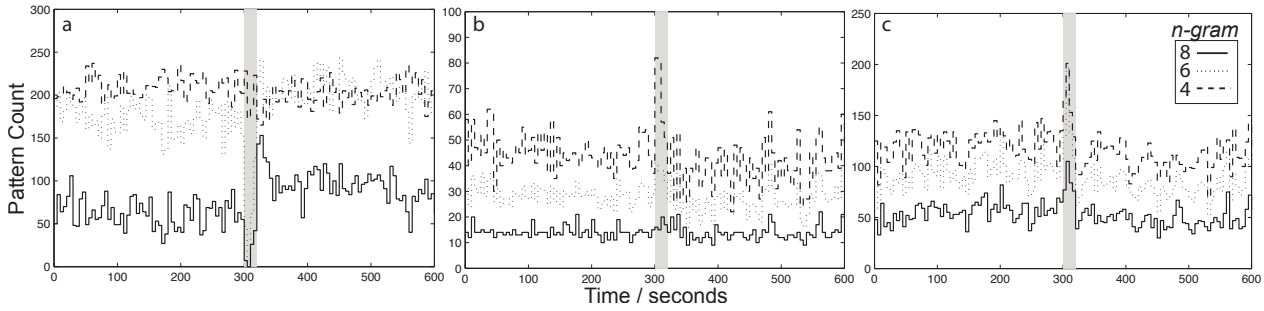


Figure 4.3 - Figure showing the pattern counts for three ngram sizes with three different quantisation levels: (a) 8, (b) 6 and (c) 4.

Thus we clearly see a few cases. A decrease in pattern size can be due to two reasons: (1) the quantisation to transform the signal to the symbolic domain fails to compress the

variabilities between similar symbol sequences such that they do not get counted and rather than see an increase, we see a decrease in pattern counts or (2) an increased periodicity results in fewer patterns. An increase in patterns may capture increases in the signal activity, such as intermittent activity. This can also be a result of larger patterns being broken up such that they appear as smaller pattern sizes.

What is clear is that on its own the ngram approach poses more questions than answers and needs to be improved to truly quantify the signal dynamics. This can be improved to have a refined pattern inference by searching for *unique* patterns. Quantisation which reduces signal states by a factor of 2 and inclusion of nonstationary and nonlinear processes that will misalign patterns, assigning them to different states, can result in similar patterns being considered separately. This can be countered by searching for *similar* patterns between those found.

To address ***uniqueness*** we propose to use a multi-resolution ngram that searches for the most significant non-overlapping patterns at a subset of pattern lengths. So we apply a methodology that during the ngram process derives the unique, non-overlapping patterns at all levels of pattern length from the ngram lengths of $Ngram = (N_1,...,N_n)$ with $N_n > N_1$. Hence for a time series in symbolic representation we search for the unique and most significant (highest count) pattern sizes in the *ngram* range. The process is described for a sequence x of length n and current symbol index i (note we start by searching for the largest ngram size):

1. Extract pattern, $P_k = \{x_i, x_{i+1}, ..., x_{i+N_n-1}\}$
2. Lookup in the table of patterns of length N_{n-j} where $j = 0,...,n-1$ is the index of Ngram length, initially $j=0$.

a. If this exists, $c(P_k) = c(P_k) + 1$, signal index $i = i + N_{n-j}$, is incremented. Determine patterns in Ngram tables with same first location and remove (see step 3 and note), return to step 1.

b. If doesn't exist and $N_{n-(j+1)}\} \geq N_1$ then $j = j + 1$ and $P_k = \{x_i, x_{i+1}, \ldots, x_{i+N_{n-j}}\}$ and return to step 2, else step 3.

3. Since no pattern pre-exists in the lookup tables each patter, P_k is stored in a table with last location address, i recorded and $i = i + N_1$ and returned to step 1.

Note: The largest pattern that repeats is considered dominant over any others with the same last location address regardless of whether in the overall scheme another pattern would have dominated. This significantly reduces the complexity of the process.

To explain further in the case that a no pre-found pattern is found in the pattern tables, then the pattern will be added to the pattern list for all pattern sizes with a specific unique flag. These flags allow us to determine, once all patterns are located, which is the dominant pattern and thus remove the others. Thus what remains is to some degree the *unique* set of patterns. An example on the same data set is shown in Figure 4.4.

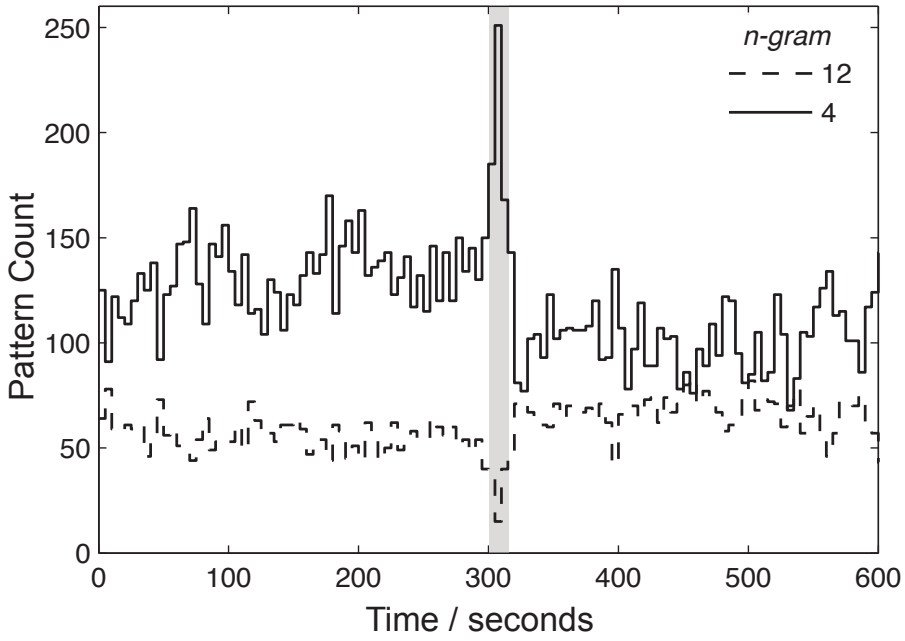


Figure 4.4 – Figure showing the results of the implementation of the unique multiresolution ngram approach.

What this shows us is that during the seizure onset, the number of unique large patterns reduces (due to periodicity), while the smaller pattern sizes increase somewhat. The latter is partially due to variations in patterns symbols that are in fact subsets of the larger pattern lengths; a member of the smaller pattern size varies in symbol classification it is missed as being a subset of the larger pattern. This can be partially resolved using similarity metrics.

This is the algorithmic framework within which we have implemented our system. Although preliminary code (Appendix A) was implemented in Matlab, we wanted to create an environment that is more efficient for uploading mass amounts of data with included metadata.

4.2.3 System Implementation

We have mentioned how one of the bottlenecks of analysis of epileptic seizure activity has been the ability to work with consistent data, well annotated and databased. Part of this

work is to establish such a framework for future work and to support signal processing methods, including our own. Any biomedical signal processing technique needs to be built into a framework that allows efficient processing and storage of results. We envision a system that follows the trends of biomedical healthcare in bringing a sensor front-end to processing and backend visualisation (or closed-loop functionality) into a modular system where different components can be placed in different modules (Figure 4.5).

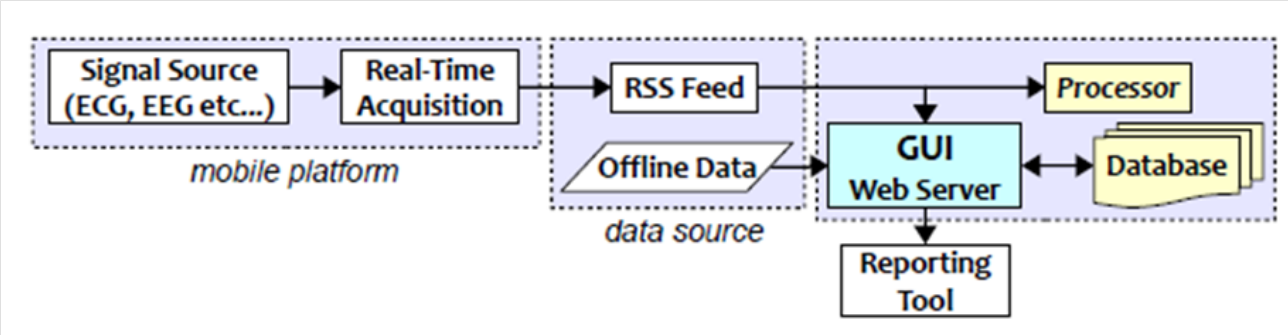


Figure 4.5 - General system overview envisioned for implementing a signal processing method for biological data.

In this figure we have a data acquisition, user interface and processing blocks. In theory each of these components could be placed in a different technological implementation, such as the acquisition being an implantable neural monitoring device, user interface on mobile phone or PC and processing units being a web-accessed cloud. The distribution of these elements will vary depending upon the signal processing requirements (computational complexity) and application space. The ngram method would ideally fall into the processor section of the web system.

We implemented this system on an online system using an SQL database structure (www.winam.net). The fundamental page where the processing parameters are stored is displayed in Figure 4.6. Here there are parameters that are associated with a case (i.e. patient) and specific information related to the data to be analysed for that patient.



Figure 4.6 – The main user interface for accessing the database and analysis.

So for each user (such as a doctor), they can have multiple cases (patients) of which they have multiple data sets (ECG, EEG etc…) at potentially different times. Each of these sets will then be subjected to analysis (runs) with different intervals/sections of the data analysed. The analysis being the multiple pattern sizes (A-D) searched for using the ngram analysis. This is all illustrated in Figure 4.7.

The case and run parameters used were based on some preliminary database parameters and common metadata used in analysis (patient age etc…). Once the database

structure and analysis were implemented we developed some visualization and printing tools to debug and analyse each stage of the system.

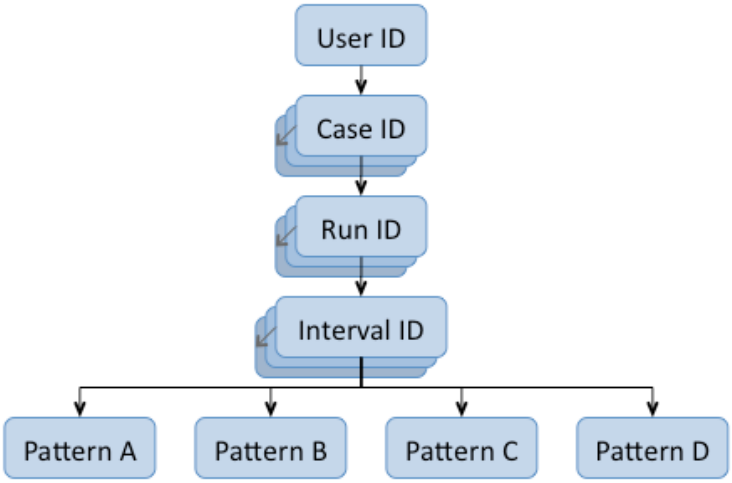


Figure 4.7 – Image of the hierarchy employed in the WiNAM database.

4.2.4 Data Requirements

The following set of information tables outlines some of the metadata to which we built the database. At present we have implemented most of these parameters and continue to build them as the database grows. Those with an ?? indicate that they are not at present relevant to our studies, but have been included for future work.

1. Patient

1a	Age Groups	Adult (>16)
1b	Sex	Mixed
1c	Ethnicity	??
1d	*Impairments (ICIDH-2)	??

* In reference to J. Engel Jr. "ILAE classification of epilepsy syndromes", Epilepsy Research 70S (2006): S5-S10.

2. Epilepsy

2a	Generalised Classification	Idiopathic, symptomatic, cryptogenic

2b	Syndrome Classifications	??
2c	Disease Classification	??
2d	Age of first onset	??
2e	AED response?	??

3. Seizure

3a	Onset Classifications	Idiopathic, focal, generalized
3b	Focal seizure types	Any
3c	Generalized seizure types	Any
3d	Status Epilepticus (continuous seizure types)	Yes
3e	Percipitating stimuli (reflex seizures)	??
3f	Brain onset areas (if focal)	Any
3g	Seizure frequency	≥1 per month
3h	Seizure duration	All durations
3i	Other Seizure types	Electrographic, clinical, benign

4. Data Gathering

4a	Sampling Frequency	≥256Hz
4b	Recording montage	10-10, 10-20, monopolar
4c	Type of electrode / invasiveness	Intracranial (ECoG/Deep Brain), EEG
4d	Data acquisition resolution	8-16bit
4e	Number of Patient data sets	100
4f	Number of non-epileptic controls	20
4g	Hours of data per patient	≥24
4h	Number of seizures per record	≥2 (ideally several)
4i	Number of significant epilepsy signatures	≥2
4j	Interictal duration (hours)	≥10

5. Other Specifications

5a	Artefacts (independent)	Sweating, heart, electrical, chewing, tongue, blinking
5b	Artefacts (with seizures)	*As per 5a*
5c	Number of artefacts in data	≥30

4.2.5 Data Input

The first stage is the uploading of the data (after case/run parameters have been inputted) – Figure 4.8. After this one can view the data graphically as shown in Figure 4.9.

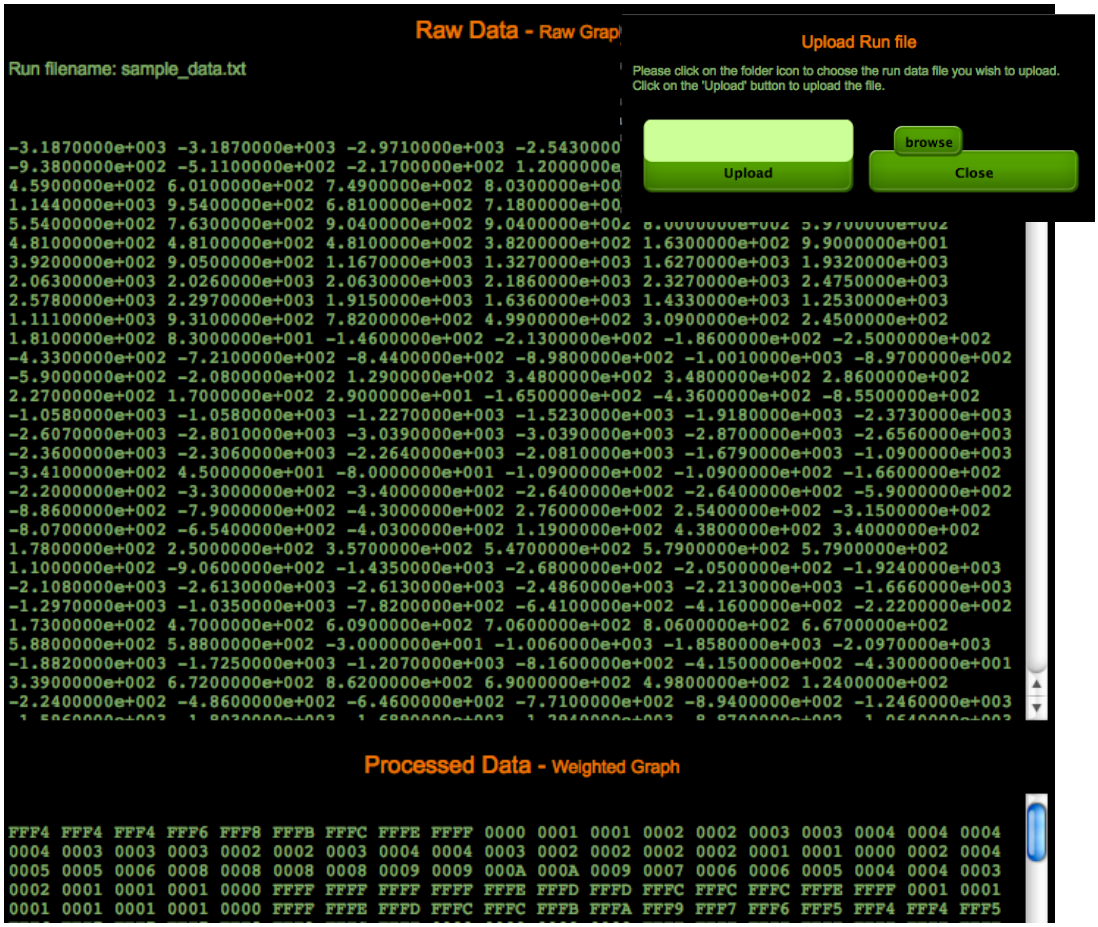


Figure 4.8 - The uploading file interface (top right) and the viewed data in its raw ascii form (engineering format) and processed (after weighting).

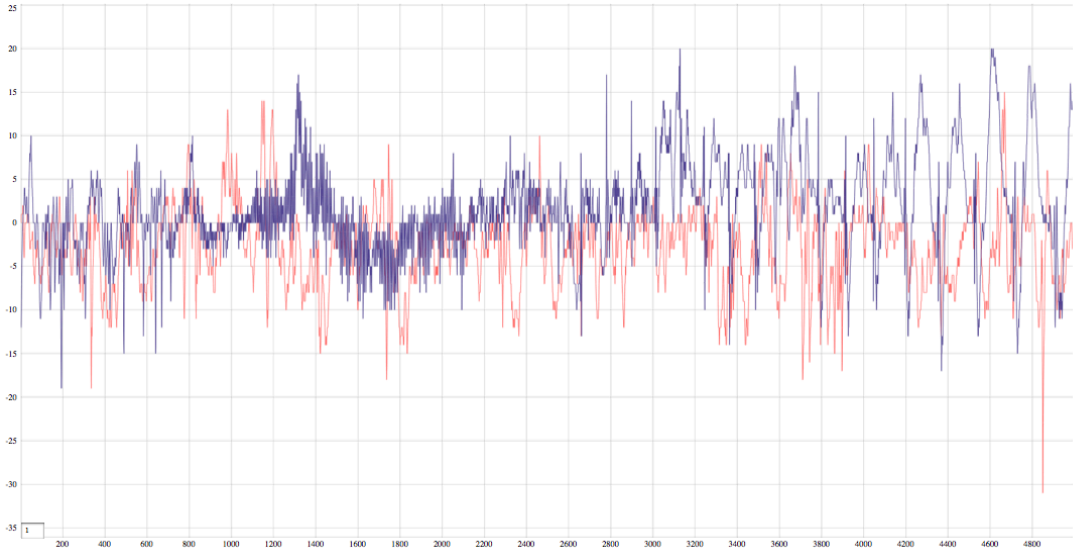


Figure 4.9 – Graphical representation of the seizure (blue) and nonseizure (red) area that can be viewed from the interface.

4.2.6 Data Analysis

After loading the data pattern counts can be chosen (A-D with A>B>C>D) and analysed as per the description of the ngram process. This can be viewed by clicking on the magnifier icon ( ) on the WiNAM NBP Counts part of (Figure 4.10). This also tells us the number of *anomalies* or non patterns.

4.2.7 WiNAM Parameters

The following parameters are those tuneable within WiNAM for optimisation tests.

Pattern length	0-24
Minimum occurrence	2-8
Standard deviation (SD)	0.5-4
Opt. Intervals	0-9
Pattern Weight	8-1024
Frame length	00:00:00- 99:99:99
Amplitude filter	
Frequency filter bandwidth	

The filtering tools have yet to be implemented at the time of writing this. Although some of our results (presented in the next chapter) carry out offline filtering in Matlab.

NBP Count Table

Case ID: 0001
Run ID: 0001
Run filename: sample_data.txt
File Size (bytes): 9,982
Total Patterns(bytes): 380
Total Anomalies (bytes): 9,602
Anomalies Percent: 96.19%

Sequence	Pattern	Count
0001	0003 0003 0003 0003 0004 0004 0003 0003 0002 0003	2
0002	0004 0004 0003 0003 0003 0003 0003 0003 0003 0003	2
0003	0005 0004 0004 0004 0004 0004 0004 0004 0004 0004	2
0004	0004 0004 0004 0004 0004 0004 0004 0004 0004 0003	2
0005	0001 0002 0002 0001 0001 0001 0001 0001 0001 0001	2
0006	0005 0005 0005 0005 0006 0006 0006 0005 0004 0004	2
0007	0004 0004 0004 0004 0003 0003 0003 0002 0002 0003	2
0008	0001 0000 0000 0000 0001 0001 0001 0001 0001 0001	2
0009	0004 0004 0004 0004 0004 0004 0004 0003 0003 0003	2
0010	0008 0008 0008 0008 0009 0009 0009 0009 0008 0008	2
0011	0010 0010 000F 000E 000D 000D 000D 000D 000E 000E	2
0012	0004 0004 0004 0004 0004 0004 0003 0003 0003 0002	2
0013	0003 0003 0003 0004 0004 0003 0003 0002 0003 0003	2

Figure 4.10 – Showing the ngram pattern count.

Finally, for all cases, one can look at the number of anomalies for all the runs of a given case (Figure 4.11). To elaborate from this figure, there are 380 significant patterns in the file meaning 9,602 non-patterns which means 96.19% are anomalies.

Case 0001 Ratio Table

Sequence	Run ID	File Name	Comment	Size - File	Size - Patterns	Size - Anomalies	Anomalies %
0001	0001	sample_data.txt	Seizure	9,982	380	9,602	96.19%
0002	0002	sample_data.txt	Normal	9,982	2,142	7,840	78.54%

Figure 4.11 – The summary of the runs ratio (expressed as anomalies percentage) for a case.

4.3 Summary/Conlusions

In the work presented here we aimed to quantify and database neurological data with the aim of assessing normal behaviour and quantifying anomalies within it. To do these, two elements had to be developed, the database and a quantifying measure.

The measure chosen was a novel one, based on a pattern recognition method called ngram. To develop this work further we first needed to develop the analysis system and database so we could analyse large amounts of data sets automatically, in real-time with multiple patterns. Ambulatory recordings and more data will help develop a more rounded analysis scheme for quantifying neurological behaviour.

In the next chapter we apply this methodology to the real data, quanitifying seizure prediction and detection capabilities of our system.

4.4 References

[1] S. B. Needleman and C. D. Wunsch, "A general method applicable to the search for similarity in the amino acid sequences of two proteins," *J. Mol. Biol.,* vol. 48, pp. 443-453, 1970.

[2] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical Pattern Recognition: A Review," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, pp. 4-37, 2000.

[3] S. Banerjee and T. Pederson, "The Deisgn, Implementation and Use of the Ngram Statistics Package," *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics,* pp. 370-81, 2003.

[4] Y. Bengio, R. Ducharme, P. Vincent, and c. Jauvin, "A Neural Probabilistic Language Model," *Journal of Machine Learning Research,* vol. 3, pp. 1137-55, 2003.

[5] M. Winterhalder, B. Schelter, T. Maiwald, A. Brandt, A. Schad, A. Schulze-Bonhage, and J. Timmer, "Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction," *Chaos,* vol. 16, p. 013108, 2006.

[6] C. Rao, A. Ray, S. Sarkar, and M. Yasar, "Review and comparative evaluation of symbolic dynamic filtering for detection of anomaly patterns," *SIViP,* vol. 3, pp. 101-114, 2009.

[7] N. E. a. H. Clapp, L. M., "Method and Apparatus for Extraction of Low-Frequency Artifacts from Brain Waves for Alertness Detection," US Patent, 1997.

[8] H. Feldwisch-Drentrup, M. Ihle, M. Le Van Quyen, C. Teizeira, A. Dourado, J. Timmer, F. Sales, V. Navarro, A. Schulze-Bonhage, and B. Schelter, "Anticipating the unobserved: Prediction of subclinical seizures," *Epilepsy & Behavior,* vol. 22, pp. S119-S126, 2011.

[9] W. Juffali, A. Eftekhar, and C. Toumazou et al, "The WiNAM project: Neural data analysis with applications to epilespy," *IEEE BioCAS,* pp. 45-48, 2010.

[10] J. Malmivuo and R. Plonsey, *Bioelectomagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*, 1995.

Chapter 5- Clinical Studies and Results

In the previous chapter we developed our novel pattern recognition method for analysis of various brain signals (including EEG and ECoG). In this chapter we apply this method to the area of seizure detection and prediction, which was discussed in Chapter 3. The aim being to quantify the ability of our method to predict seizure in a specified time prior to their onset and benchmark the results against current state of the art significance tests.

5.1 EPI Data

The data obtained from the EPI Epilepsy Centre in Zurich is a single case with 36 hours of data, sampled at 256Hz. At approx. 18 hours a seizure occurred (Figure 5.1). We pre-filtered the data at low pass frequencies (-3dB cut-off) of 5, 10, 15, 20, 30 and 70Hz using a 4th order Butterworth implementation.

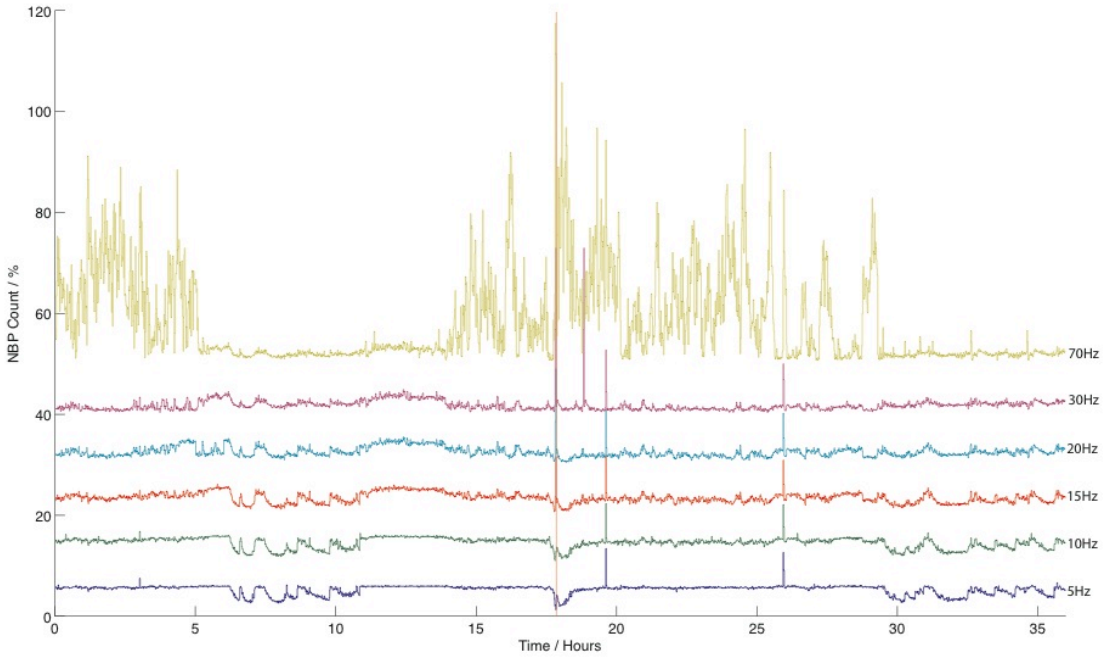


Figure 5.1 – Figure showing the results using WiNAM at different low-pass filtering frequencies

What is clear from this figure is that the seizure is detectable at all filtering frequencies but we see similar patterns associated with the sleep periods. Hence we analysed the 5Hz filtered data with different voltage reference levels to account for the reduced amplitude of the data (due to filtering). This is shown in Figure 5.2, where we clearly see that the seizure onset is much higher than the associated sleep patterns. The two large spikes at approx. 20 hours and just post 25 hours are signal artefacts (due to movement).

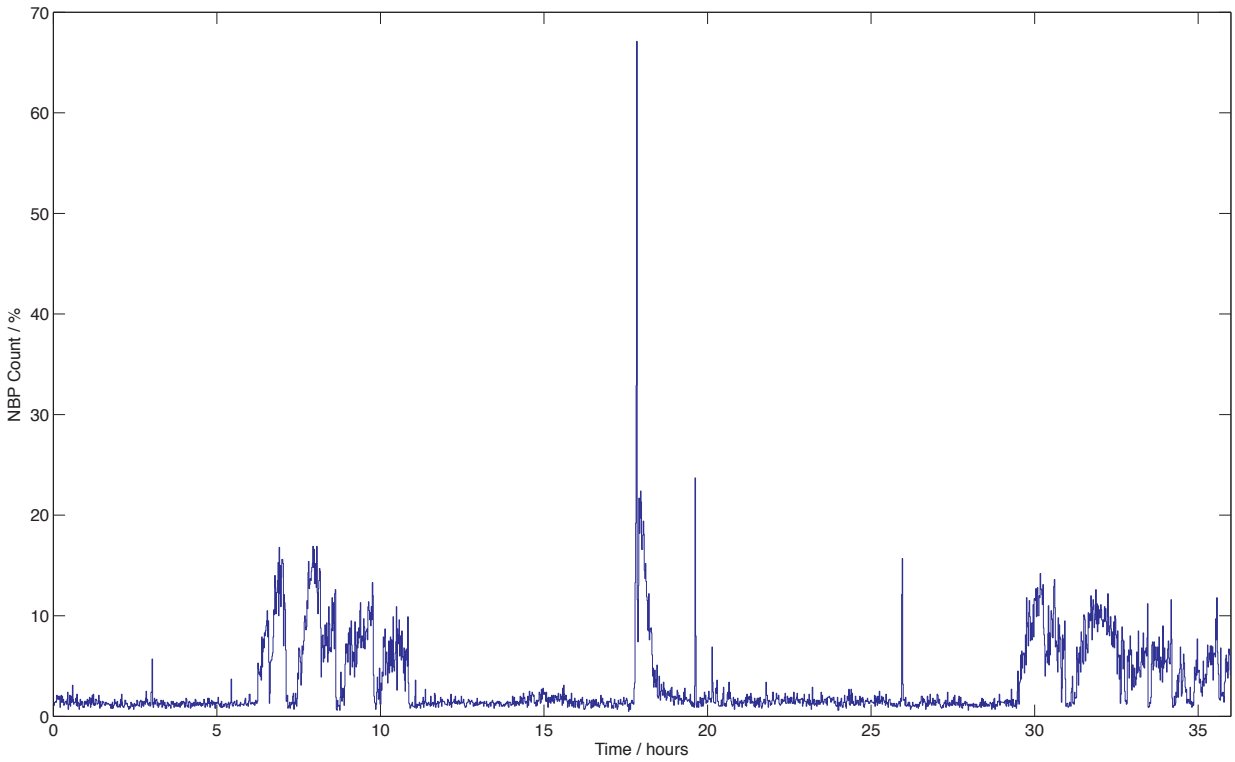


Figure 5.2 – Figure showing a 5Hz filtered version of the EPI data with voltage reference levels set to the maximum and minimum of the signal.

This is the extent to which we have analysed and worked with this data but it proved that initially our methodology extracts dynamics of the signal associated with significant and relevant features of the data set.

We are working with EPI to generate a well-annotated data set similar to this case (including sleep and artefact) to quantify and analyse the signal behaviour more closely.

Interestingly the seizure is predicted up to 10 minutes prior to onset although in this case, these changes are associated with typical ictal discharge phenomena.

5.2 Freiburg Data

Following this analysis we obtained a freely available data set from the University Hospital of Freiburg Epilepsy Centre, Germany. The data used were pre-sampled at 256Hz and quantised using a 128 channel 16-bit data acquisition [1-3]. This data is extracted from grid and depth electrode using 6 channels, 3 near the focus and 3 not.

We carried out two separate tests on this data set. The first a proof of concept that we published in 2010 [4], the second a study quantifying the detection further and also the prediction capability.

Patient	Sex	Age	Seizure type	H/NC	Origin	Electrodes	seizures analyzed	Interictal Duration	Info	Ictal	Size[Mb]	Interictal	Size[Mb]	Interictal duration	Interictal intervals
1	f	15	SP,CP	NC	Frontal	g,s	4	86400 s	info	001ictal	86,4	001interictal	242,2	24 h	1
2	m	38	SP,CP,GTC	H	Temporal	d	3	86400 s	info	002ictal	55,8	002interictal	220	24 h	2
3	m	14	SP,CP	NC	Frontal	g,s	5	86400 s	info	003ictal	82,5	003interictal	232	24 h	1
4	f	26	SP,CP,GTC	H	Temporal	d,g,s	5	86400 s	info	004ictal	113,9	004interictal	235,3	24 h	1
5	f	16	SP,CP,GTC	NC	Frontal	g,s	5	86400 s	info	005ictal	72,1	005interictal	175,7	24 h	3
6	f	31	CP,GTC	H	Temporo/Occipital	d,g,s	3	86400 s	info	006ictal	62,9	006interictal	221,1	24 h	1
7	f	42	SP,CP,GTC	H	Temporal	d	3	88597 s	info	007ictal	55,7	007interictal	235,2	25 h	1
8	f	32	SP,CP	NC	Frontal	g,s	2	86979 s	info	008ictal	35,1	008interictal	236,1	24 h	2
9	m	44	CP,GTC	NC	Temporo/Occipital	g,s	5	86163 s	info	009ictal	96	009interictal	447,3	24 h	2
10	m	47	SP,CP,GTC	H	Temporal	d	5	88047 s	info	010ictal	101,6	010interictal	226,8	24 h	1
11	f	10	SP,CP,GTC	NC	Parietal	g,s	4	86570 s	info	011ictal	76,3	011interictal	231,2	24 h	1
12	f	42	SP,CP,GTC	H	Temporal	d,g,s	4	89326 s	info	012ictal	74	012interictal	417	25 h	1
13	f	22	SP,CP,GTC	H	Temporo/Occipital	d,s	2	86400 s	info	013ictal	38	013interictal	229,2	24 h	1
14	f	41	CP,GTC	H, NC	Fronto/Temporal	d,s	4	85894 s	info	014ictal	66,5	014interictal	224,1	24 h	5
15	m	31	SP,CP,GTC	H, NC	Temporal	d,s	4	86400 s	info	015ictal	93,1	015interictal	222,4	24 h	1
16	f	50	SP,CP,GTC	H	Temporal	d,s	5	86400 s	info	016ictal	104,6	016interictal	211,2	24 h	2
17	m	28	SP,CP,GTC	NC	Tmporal	s	5	86634 s	info	017ictal	151,7	017interictal	241,5	24 h	1
18	f	25	SP, CP	NC	Frontal	s	5	89569 s	info	018ictal	132,8	018interictal	242,1	25 h	1
19	f	28	SP,CP,GTC	NC	Frontal	s	4	87780 s	info	019ictal	119,5	019interictal	227,8	24 h	3
20	m	33	SP,CP,GTC	NC	Tempo/Parietal	d,g,s	5	92219 s	info	020ictal	124,7	020interictal	239,6	26 h	1
21	m	13	SP,CP	NC	Temporal	g,s	5	86177 s	info	021ictal	110,5	021interictal	224,1	24 h	2

SP = simple partial CP = complex partial GTC = generalized tonic-clonic

Figure 5.3 – Figure showing the patient details breakdown, taken from https://epilepsy.uni-freiburg.de/freiburg-seizure-prediction-project/eeg-database

5.3 Test 1 - *Initial Results*

All 21 patients' first seizure was used (to initially avoid any inter-ictal phenomenon) from this data set. We use a preictal period of up to 1 hour in most cases (where such time was available) with seizure durations varying from 15-170 seconds. There were multiple types of seizures present including: simple partial, complex and general tonic-clonic.

The basic analysis we first used [4] analysed the difference between non-seizure (in the pre-ictal period) and seizure occurrence areas (Figure 5.4).

The non-seizure areas were gathered from 3 randomly selected areas (of the same length as the seizure areas) and pattern count averaged. Each set was reduced to 8bits and various ngram sizes ranging from 10 to 14 were used. The quantifier used, was the ratio between the pattern count of the two areas.

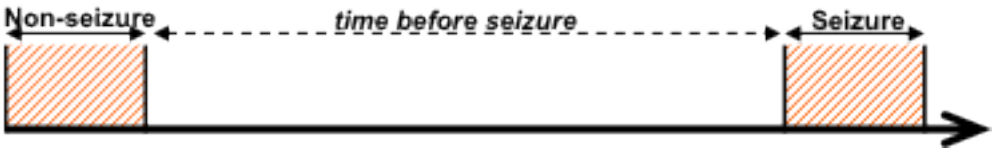


Figure 5.4 - Example of how seizure and non-seizure areas are selected and compared.

Table 5.1 taken from [4] shows an example of the types of ratios found. From this table we identified 18 out of the 21 patients seizures were detected on these criteria alone, whereby a change of ratio greater than 25% indicated seizure onset. Recently we have shown that the ratios by combining information on pattern sizes 10 and 12 can lead to a 19 out of 21 success (90% sensitivity). This proof-of-concept warranted further study into the use of this method to initially detect seizures.

Table 5.1 - Taken from [4] showing the initial results aiming to quantify the ratio of patterns between seizure and non-seizure.

			NGRAM = 12			NGRAM = 14		
Patient	Sdur/s	Pdur/hrs	S	P	Ratio	S	P	Ratio
1	19.49	6.10	514	778.9	0.66	269	393.3	0.684
2	147.35	0.89	6000	28727.5	0.209	2805	21811.3	0.129
3	108.63	1.67	11029	25517.8	0.432	8150	23742.1	0.343
4	77.30	1.03	743	324.6	2.289	577	83.2	6.935
5	14.98	0.99	1455	3246.8	0.448	1138	2991.8	0.38
6	107.94	1.80	2974	13409.7	0.222	1606	9053.5	0.177
7	79.55	1.70	938	16917.8	0.055	719	14890.9	0.048
8	178.87	1.10	27969	34800.6	0.804	19923	27931.6	0.713
9	38.56	1.04	990	6939.9	0.143	702	5465.6	0.128
10	87.07	1.18	14280	19114.6	0.747	10639	17416.9	0.611
11	41.55	1.05	2241	2216.3	1.011	981	861.8	1.14
12	61.33	1.00	2032	12806.2	0.159	1091	11541.4	0.095
13	76.13	1.18	4214	4809.5	0.876	1791	2451.3	0.731
14	70.77	1.34	14268	13165.1	1.084	12212	11027.6	1.11
15	31.51	0.94	5326	5523.1	0.964	4305	4452.6	0.97
16	170.61	1.36	29278	37282.8	0.785	21912	33935.4	0.646
17	58.87	1.81	1256	11120.3	0.113	988	9095.3	0.109
18	17.93	1.20	0	401.6	0	0	154.9	0
19	23.17	1.01	850	2688.7	0.316	469	1679.1	0.279
20	25.09	1.84	1323	4629	0.286	946	4098.1	0.231
21	89.91	0.93	17693	20965.5	0.844	15120	19932.2	0.759

5.3.1 Towards Real-time detection

Real-time detection requires analyzing segments of data and comparing the pattern counts to some predefined adaptive threshold. There are several ways to implement thresholds, most based on some statistics of the data, such as standard deviation.

Towards this goal of real-time some initial results were presented in this paper on single patterns, an example of which is shown in Figure 5.5. Here, we can clearly see a pattern count change at the onset (note: at the exact onset, not just high amplitude) of the seizure. This was done by analyzing the pattern count every 10 seconds.

To further study this, as simple thresholding was found to be unsuccessful we qualified the results by eye, to determine what types of changes are seen at the onset of seizure compared to *normal* activity. This is shown in Table 5.1 using ngram sizes of 10 and 14 with 10 and 5 second windows of analysis. With this the sensitivity reduced to 17 out of 21

(81%). However, some of the changes were difficult to quantify and this requires further automated studies.

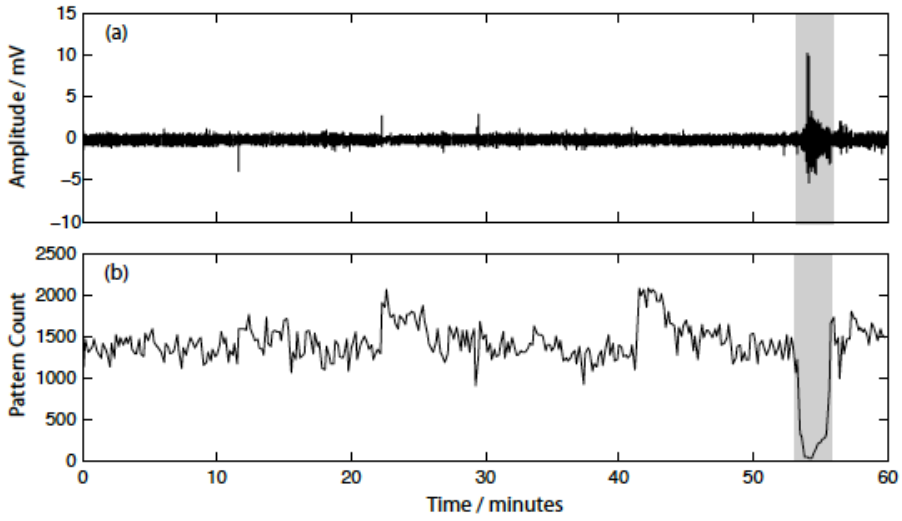


Figure 5.5 – Shows a typical ECoG signal (top) with seizure occurrence (shaded area) and the pattern count (bottom) for a 10 second window

Table 5.2 – Showing the qualified results of full analysis using single pattern counts over the full preictal period available.

	NGRAM = 10		NGRAM=14
Patient	10s	5s	5s
1	decrease	none	none
2	decrease	decrease	decrease
3	decrease	decrease	decrease
4	none	increase	increase
5	decrease	decrease	decrease
6	decrease	decrease/increase	increase
7	decrease	decrease	increase/decrease
8	decrease	decrease	none
9	decrease	decrease	increase
10	increase/decrease	increase/decrease	increase/decrease
11	none	none	none
12	decrease	decrease	increase/decrease
13	none	none	none
14	none	none	none
15	none	none	none
16	decrease/increase	decrease/increase	decrease/increase
17	decrease	decrease	increase
18	reduced variance	reduced variance	reduced variance
19	decrease/increase	decrease/increase	increase
20	decrease	decrease	decrease
21	none	decrease	decrease

There are some trends such as a sharp increase followed by a decrease that ones see quite often (Figure 5.6) in these data sets.

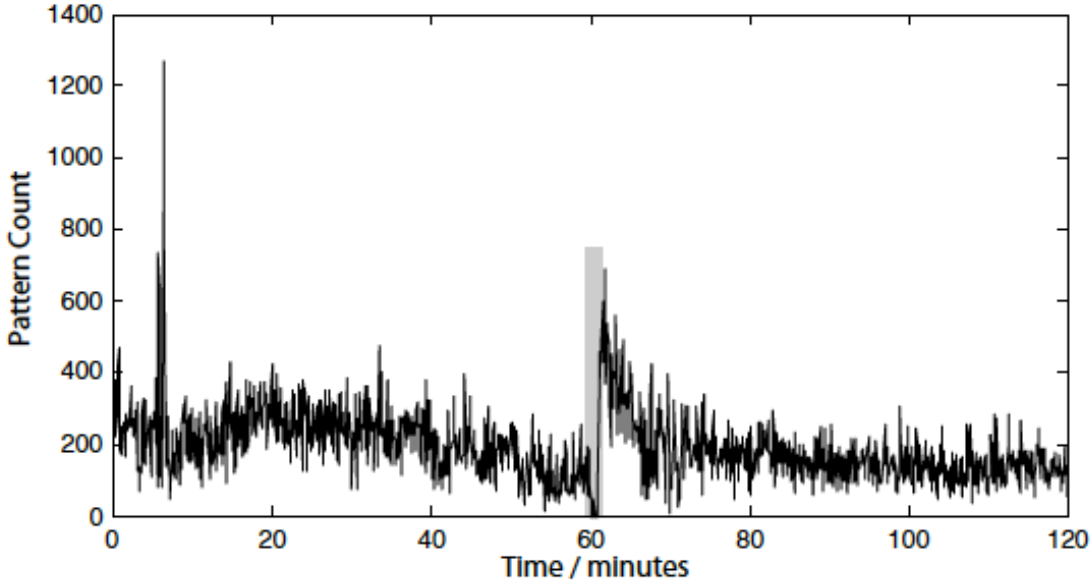


Figure 5.6 – Shows the quick increase and slow decrease seen in some sample data.

5.4 Test 2 - *Detection and Prediction*

Test 2 involves a more extensive analysis of the Freiburg data set. Based on the work defining a seizure prediction horizon we initially chose a 65-minute window prior to seizure onset to study predictive markers but also the detective ability of the algorithm. A summary of the data extracted from each patient is shown below where we identify the maximum data we could extract from what was available and by observation determined which channels were optimum and which may have contained GTC seizures.

Patient	No Seizures	Precital / hrs	Location	Electrode	Interictal / hrs
1	4	4.65	F	g,s	23.00
2	3	3.62	T	d	24.00
3	5	6.65	F	g,s	24.00
4	5	5.51	T	d,g,s	24.00
5	5	6.03	F	g,s	24.00
6	3	3.31	T/O	d,g,s	24.00
7	3	4.08	T	d	24.61
8	2	2.68	F	g,s	23.16
9	5	6.10	T/O	g,s	23.93

10	5	7.05	T	d	24.46
11	4	4.65	P	g,s	24.05
12	4	5.91	T	d,g,s	24.81
13	2	2.38	T/O	d,s	24.00
14	4	5.41	F/T	d,s	23.86
15	4	6.13	T	d,s	24.00
16	5	6.53	T	d,s	24.00
17	5	8.80	T	s	24.07
18	5	8.20	F	s	22.87
19	4	4.60	F	s	24.38
20	5	8.42	T/P	d,g,s	25.62
21	5	7.66	T	g,s	23.94

T: Temporal, F: Frontal, O: Occipital, P: Parietal (seizure area)
g: grid, s: strip, d: depth (electrode type)

5.4.1 Detection

The data was analysed with the methods described in the previous chapter. Initally, no optimization was carried and each data set was analysed as follows: each seizure was extracted with up to 65 minutes of data and analysed using parameters: *Ngram = [10 9 8 7]* and *Nbit = 7bit, Time Window = 1 minutes.* An example result is shown in Figure 5.7 where we see the spread of pattern counts for a specic case (Patient 1, seizure 1).

To automatically quantify detection we experimented with various thresholds based on the signals past standard deviation and mean. In general we found the mean to be more tolerant to fast changing signal components. Therefore, we calculated a moving average based on the previous *N* points. In this case *N=5*.

For a given patient and seizure an optimal value of the threshold is set for that seizure, then the same threshold is applied to the remaining seizures. This is done for each pattern length (4 in this case). For example, if we have two seizures and four pattern lengths, four thresholds are determined that detects the first seizure in the four pattern counts (minimizing false detections), and then this same threshold is applied to the second

seizure patterns. This is then reversed where the optimal threshold is determined for the second seizure and applied to the first. An example of thresholds is shown in Figure 5.8.

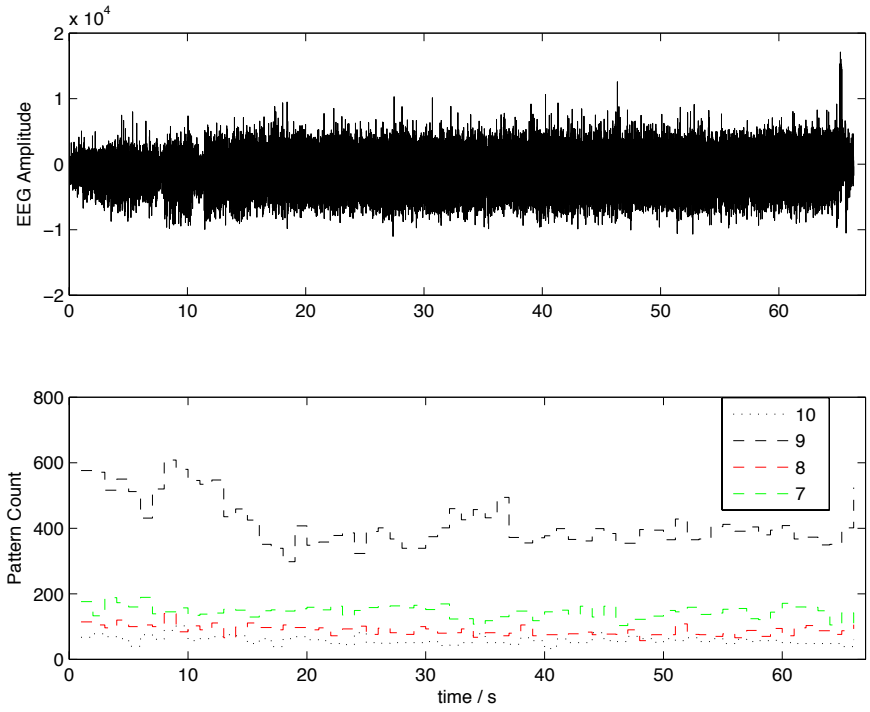


Figure 5.7 - Example of non-optimised ngram analysis of EEG signal (top) and ngram pattern counts (bottom). Seizure occurs at approx. t = 65mins.

Doing this without any other optimisations results in Figure 5.9. Although false detections are low, the sensitivity averages at 55% (0-100% range). Only in 7 out of the 21 cases (33%) do we see a sensitivity of 100%.

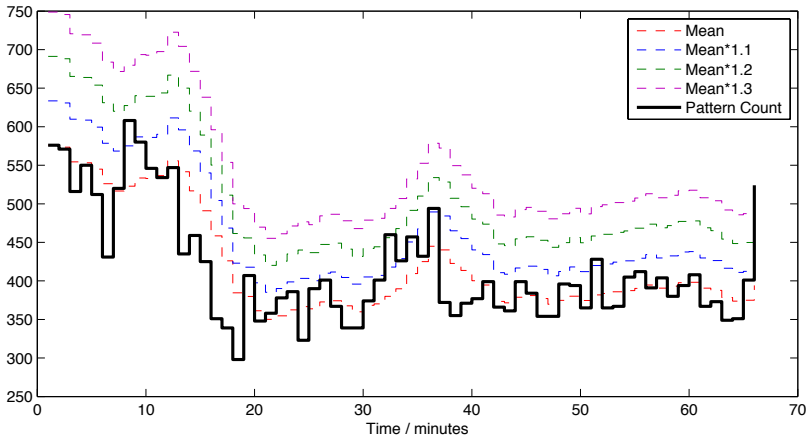


Figure 5.8 - Figure showing the thresholds as a function of signal moving mean.

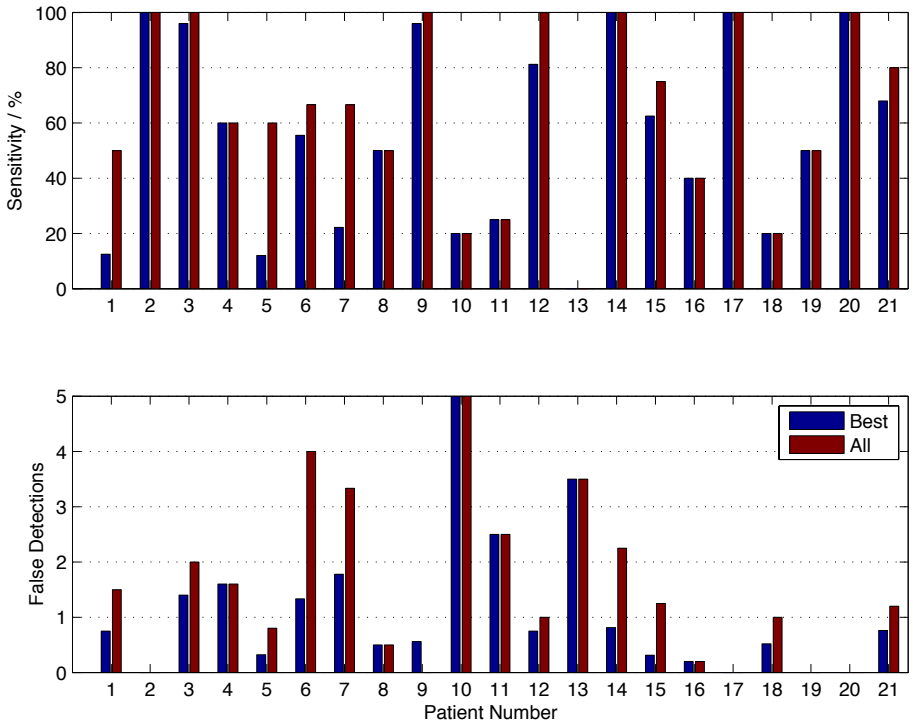


Figure 5.9 - Bar chart of the sensitivity (top) and false detections (bottom) of each Patient over all possible seizures used to determine the optimal threshold ("All") and for the best case scenario ("Best").

These results are poor, and on closer inspection with the data, this is partly due to artefacts and power line noise (examples of which are shown in Figure 5.10 and Figure 5.11).

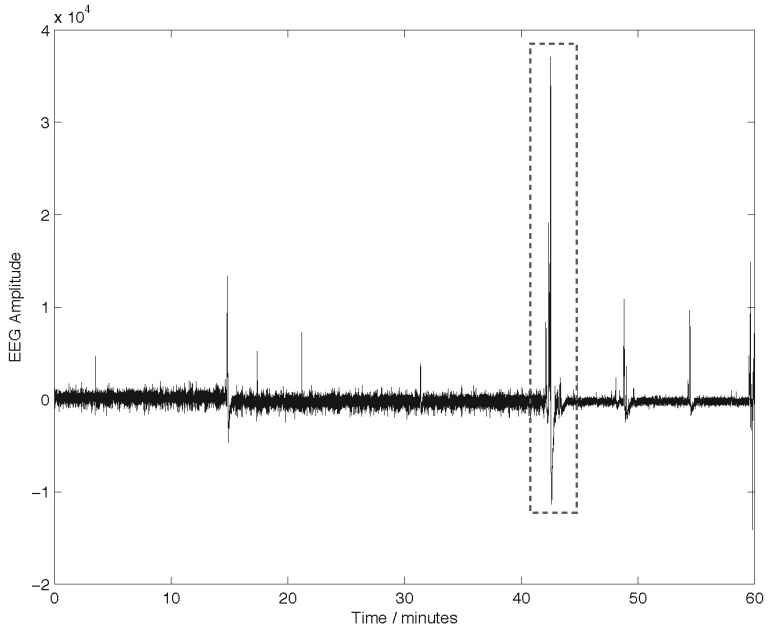


Figure 5.10 - Showing EEG plagued with artefacts (all large amplitude spikes), including at seizure onset (~60 minutes). Enclosed area indicates artifact shown in Figure 5.11.

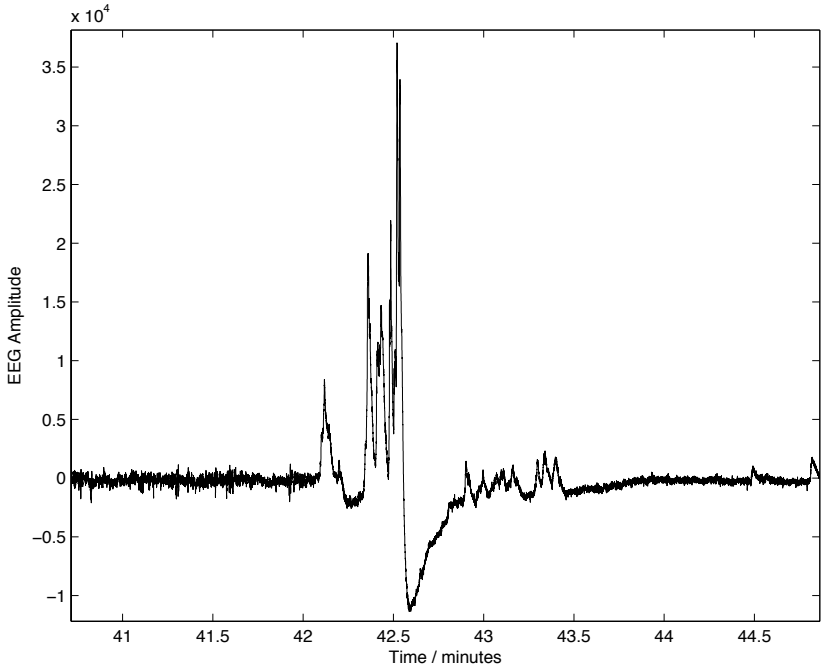


Figure 5.11 - A zoomed artifact of Figure 5.10.

As such we decided to apply interfere removal using two combined methods.

5.4.2 Interference Removal

The method we apply is to remove offsets, baseline drift and related electrode saturation. To do this we apply a similar method described in [5]. In [5] they describe a method not dissimilar to a Savitzy-Golay (SG) smoothing filter.

An SG filter takes the n points before and n points after signal point x_i and fits a polynomial to it. The polynomial fit at point x_i is effectively a filtered version (low frequency) component of the signal. Since both linear and quadratic functions may not be able to track the sharp changes associated with artefacts. We use a 3rd order polynomial on an example of which is shown in Figure 5.12.

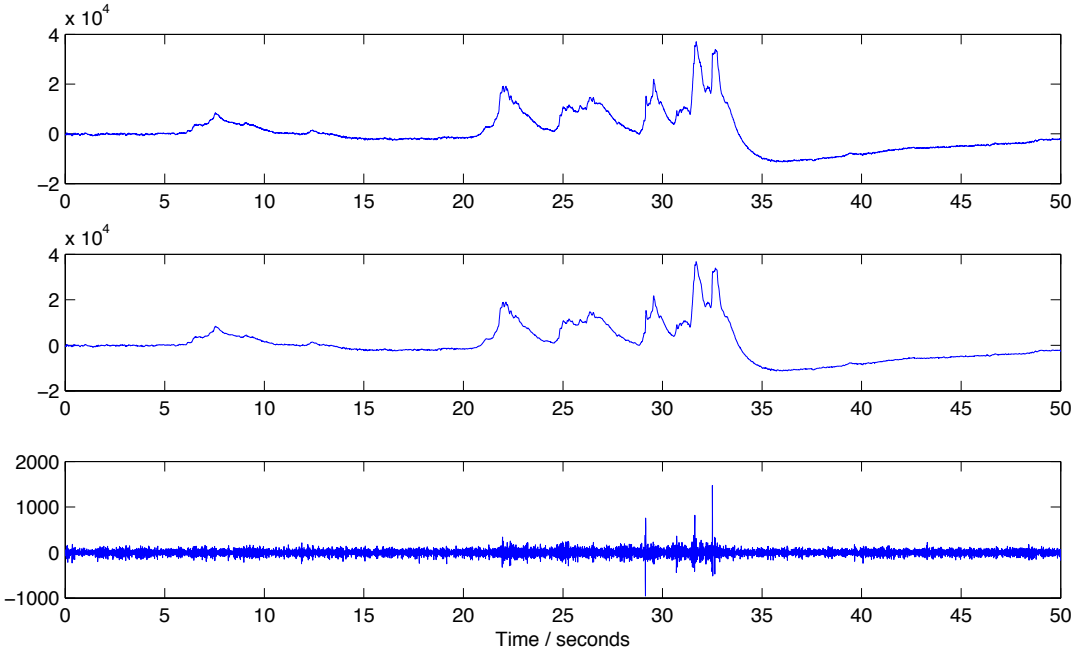


Figure 5.12 - Image showing (top) original EEG artifact signal, (middle) SG filter estimate of artifact and (bottom) EEG with artifact removed.

To this signal we then apply a notch filter (4^{th} order, -80dB attenuation and quality factor 10) to remove 50Hz noise.

With this applied to all data we apply the same analyses as before using *Ngram = [16 14 12 10 8 6 4]* and *7 bit* quantisation. We do this for 30 and 60 second time windows where a pattern is chosen as optimum for each seizure. The results are shown in Figure 5.13 and Figure 5.14 respectively. These produce sensitivities of 88.65% and 86.11% and average false detections of 6.42 and 3.73 over the entire analysed period.

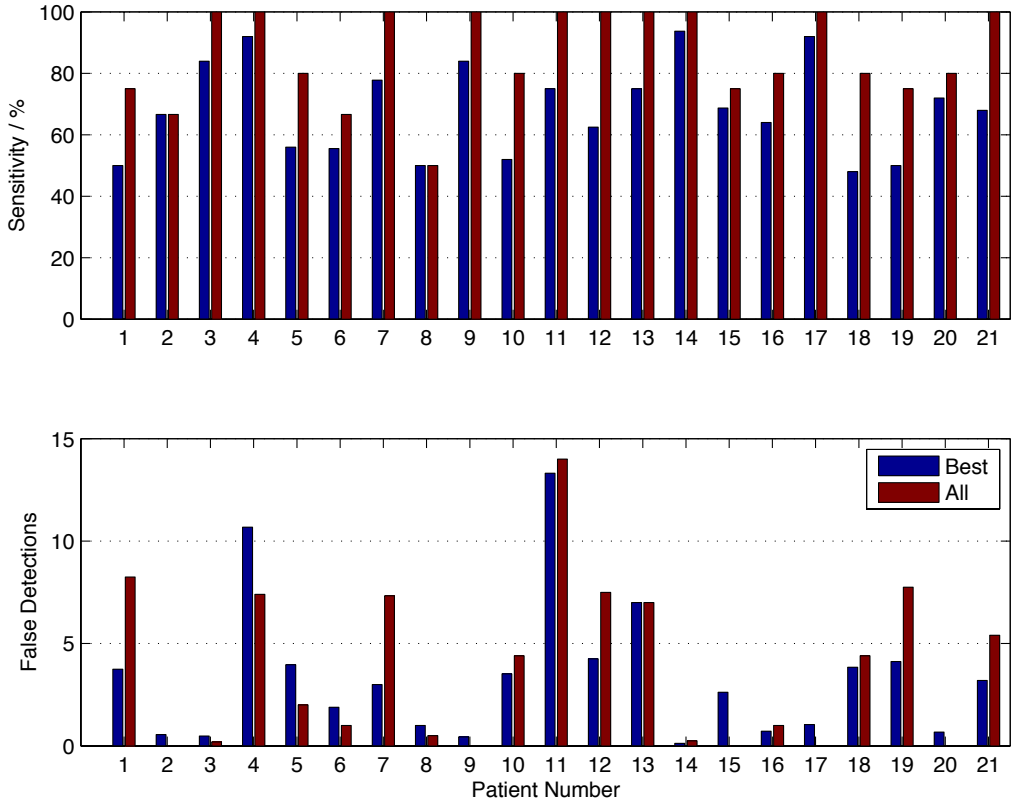


Figure 5.13 -Bar chart of the 60 second windows with artifact removed.

If we combine the best results of these two we have a sensitivity of 90.79% and mean false detection per hour of 1.54 (Figure 5.15). Even here the results have not been fully optimized for different pattern sizes, quantisation levels and only partly with window length. We believe that many cases could be further optimized.

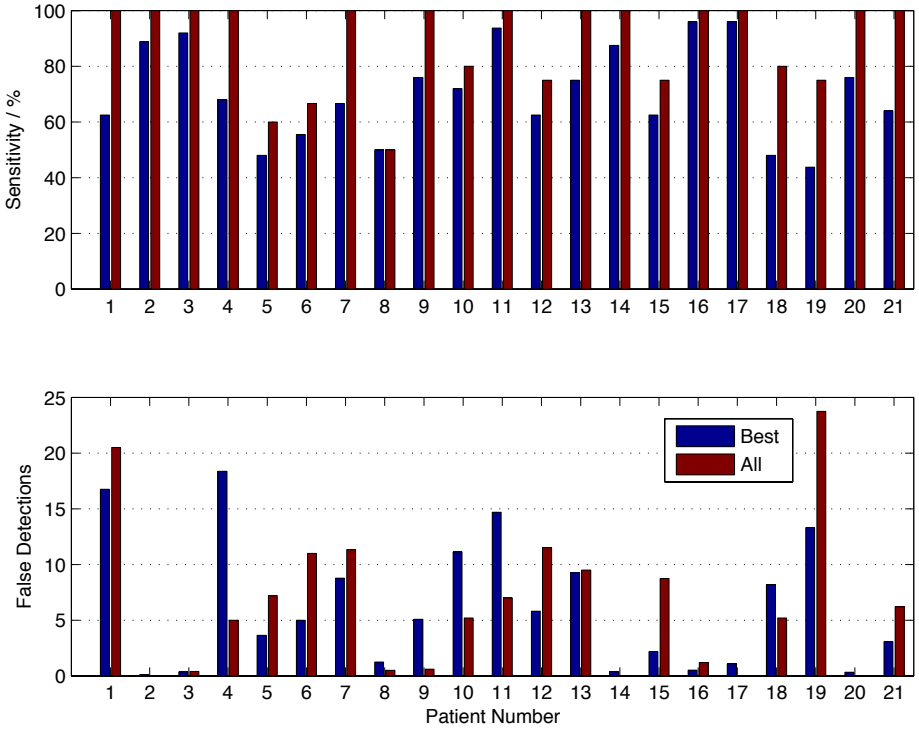


Figure 5.14 - Bar chart of the 30 second windows with artifact removed.

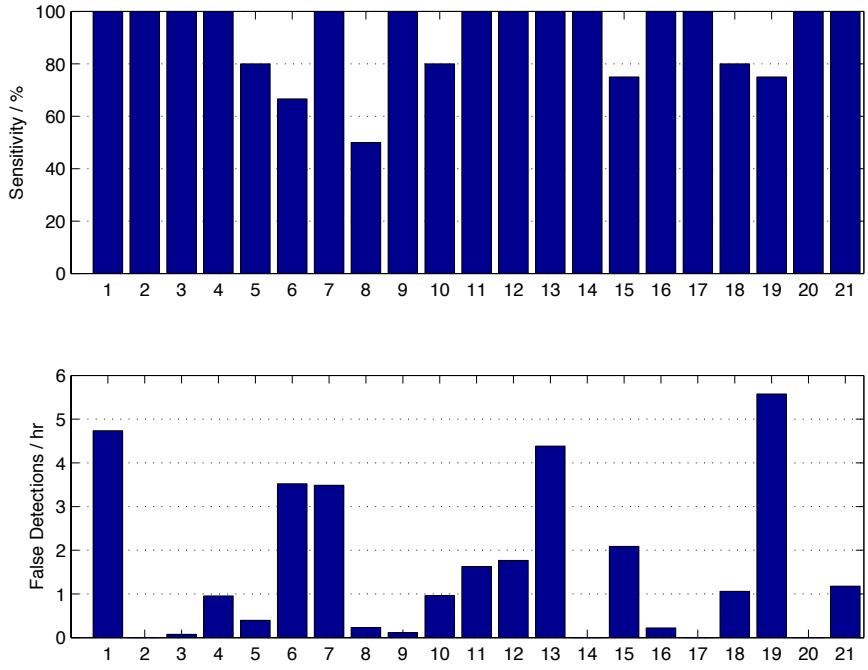


Figure 5.15 - Optimal results for combined 30 and 60 second windows.

Several cases show poorer results, specifically, case 6 and 8 show low sensitivity while 1, 6, 7, 13 an 19 show large false detections.

Closely analyzing these cases we find the following:

Sensitivity

Case 6 - Second case of this data has several inter- and pre-ictal spiking. As such these abnormal patterns make it difficult to isolate the seizure.

Case 8 - Both seizures are characterized by the large artefacts they create, meaning interference removal, which is imperfect in this case, results in reduction of results. Also, the optimal pattern length for both seizures is different.

False Detections

Case 1 - The high false detections are due to the first two seizures, the second of which is not detected. The high false detections are due to the slow changes associated with pattern changes (Figure 5.8). Possible removal of low frequency pattern changes would aid in this analysis.

Case 6 - As in the sensitivity, the second case has a lot of spiking, causing variations in patterns that subsequently cause false detections.

Case 7 - Here we could compromise sensitivity to reduce false detections. Reducing sensitivity to 67% (2 out of 3) reduces false detections to 0. This is because the last seizure appears to manifest just before the seizure index, thus biasing the threshold to be too low and have high false detections.

Case 13 - Both seizures are detectable, but given the different pattern sizes and mean and variances, it would appear that a more optimal threshold may exits.

Case 17 - We can reduce sensitivity to 50% and have lower false detections here. It does appear from the data that we could optimize better thresholds and pattern recognition parameters.

From this we have the following observations: (1) we can reduce sensitivity to minimize false detections considerably. This would completely depend on specific application requirements of detection. (2) There is a case for a more optimal threshold able to track the signal pattern changes better. We may not be able to improve sensitivity but would reduce false detections considerably. We must note that for this test, the optimal pattern was chosen for each seizure regardless of training seizure, showing the optimum results achievable. A more realistic situation aiming to match this is described next.

Our last methodology, then improves these results by optimizing for patterns. In the previous results a single pattern, for a single channel is chosen as the optimum for that patient. Here we use all channels that showed a detection of the training seizure. Detection is considered, whether false or seizure, only if a percentage of the patterns chosen show a detection. As we have 7 patterns, this meant that for different number of patterns chosen the following number of detections were required to be considered (Table 5.3).

Table 5.3 - Depiction of the number of detections need shared by multiple patterns to warrant a detection of seizure and/or false detection.

Number Patterns	No. of Detections
1	1
2	1
3	2
4	3
5	3
6	4
7	4

The results (Figure 5.16) show us a couple of observations: (1) Assuming we only use the first seizure, we would have a sensitivity of 71.3% with FDR of 0.7 per hour. This is primarily distorted by case 18, which when removed we have a sensitivity of 69.8% and FDR of 0.178, and (2) assuming we have multiple seizures to fine tune the process we can achieve a sensitivity of 92.2% with FDR of 1.6. This can be improved to 91.8% and 0.22 FDR with removal of case 1.

A final comment on these results is to iterate that any combination of FDR and sensitivity can be implemented. For example, case 1, a good result is using the second seizure for optimization, generates 50% detection and an FDR of 0.47.

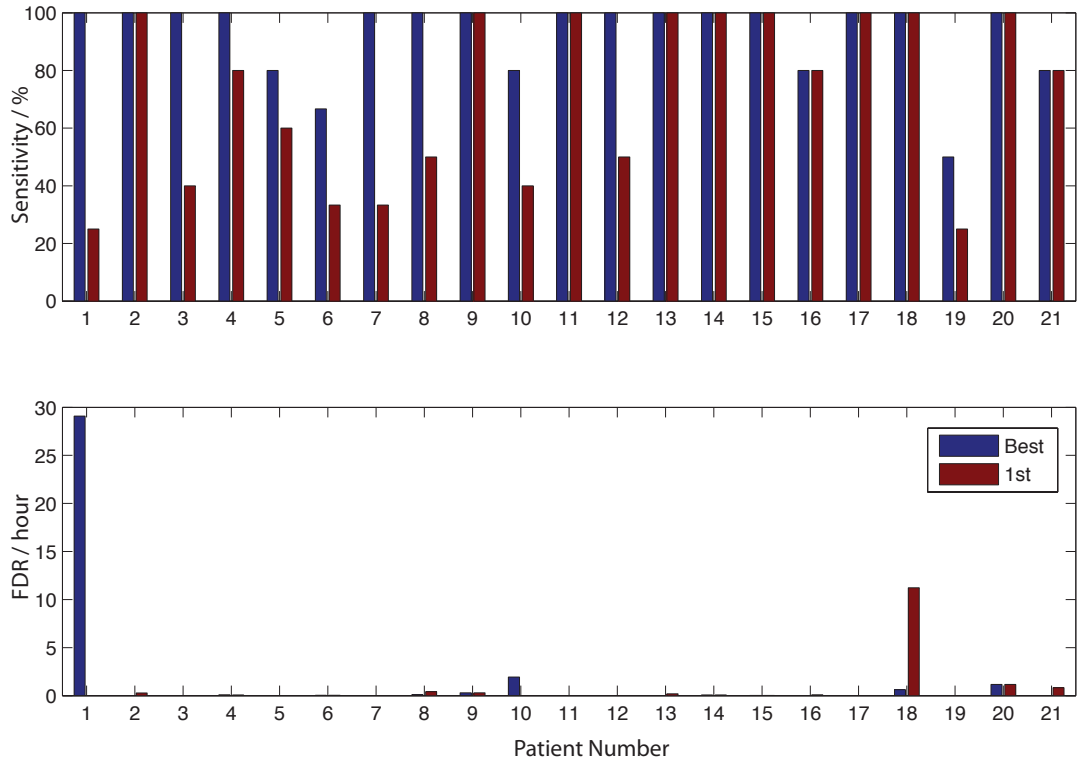


Figure 5.16 - Optimal results using the combination of multiple pattern detections to determine if a detection should be considered.

Our results overall show good detection results that has potential to be even further optimized to increase results. This requires a better understanding of seizure

characteristics that include the type of seizures (clinical/subclinical), duration and localization. This will require a well annotated data set.

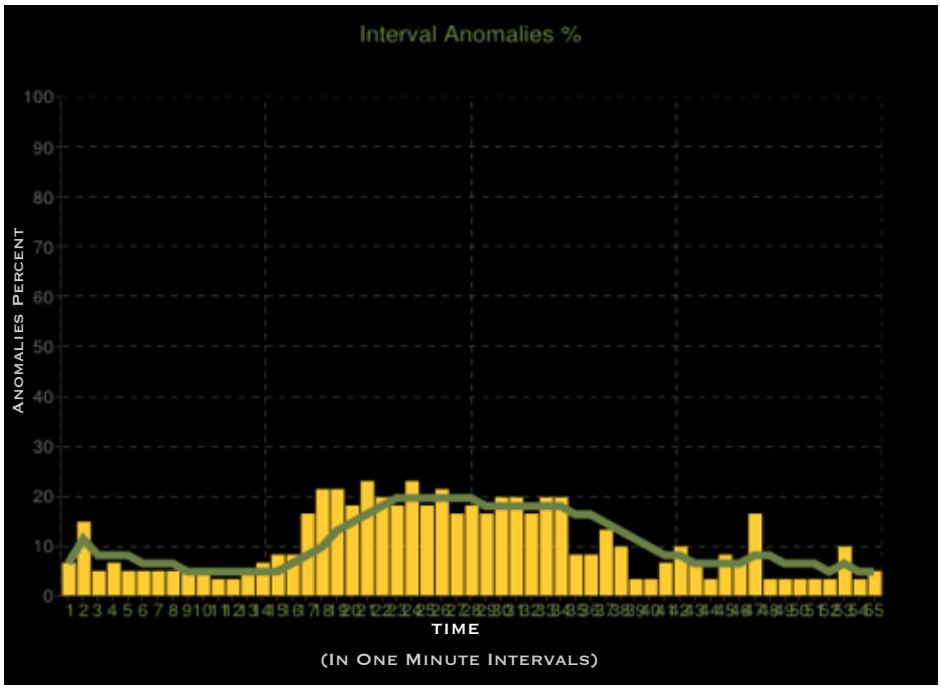
The second set of these results involves prediction. Here we decided to optimize pattern lengths and binary level by using interictal periods.

5.4.3 Prediction

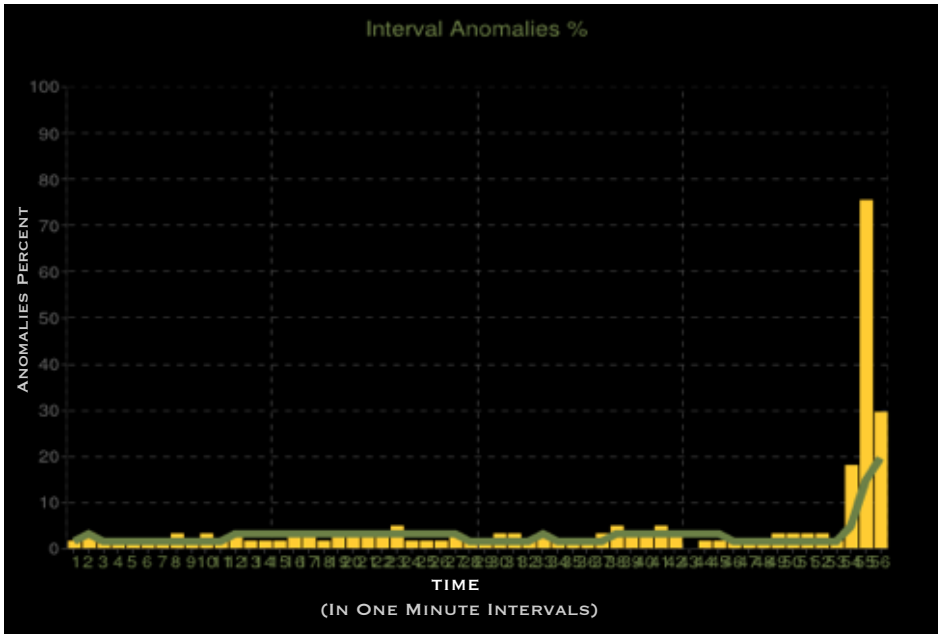
To quantify prediction we first carried out test comparing the interictal to preictal state of patients. The interictal period is used to define *Ngram* parameters that normalize the pattern count to a specific level. We initially applied this to the first seizure of each of the cases outlined in.

For 6 cases there was a clear variation of pattern counts prior to the seizure. An example of which is shown in Figure 5.17. Of course there is no clear indication on when this variance occurs in the data so we chose to investigate this further.

As in the detection test, and based on the statistical prediction metrics described in Section 3.4.6, we chose prediction horizons of 5, 15, 30 and 60 minutes, where possible. This was measured against the interictal periods to look at general (all patient) and specific patient deviations). An example of a case is shown in Figure 5.18.



(a)



(b)

Figure 5.17 - Example of Case 1, between (a) Interictal period and (b) the ictal period with seizure occurring at interval 53 (where the histogram initially rises). The thick (green) line shows the standard deviation threshold previously defined.

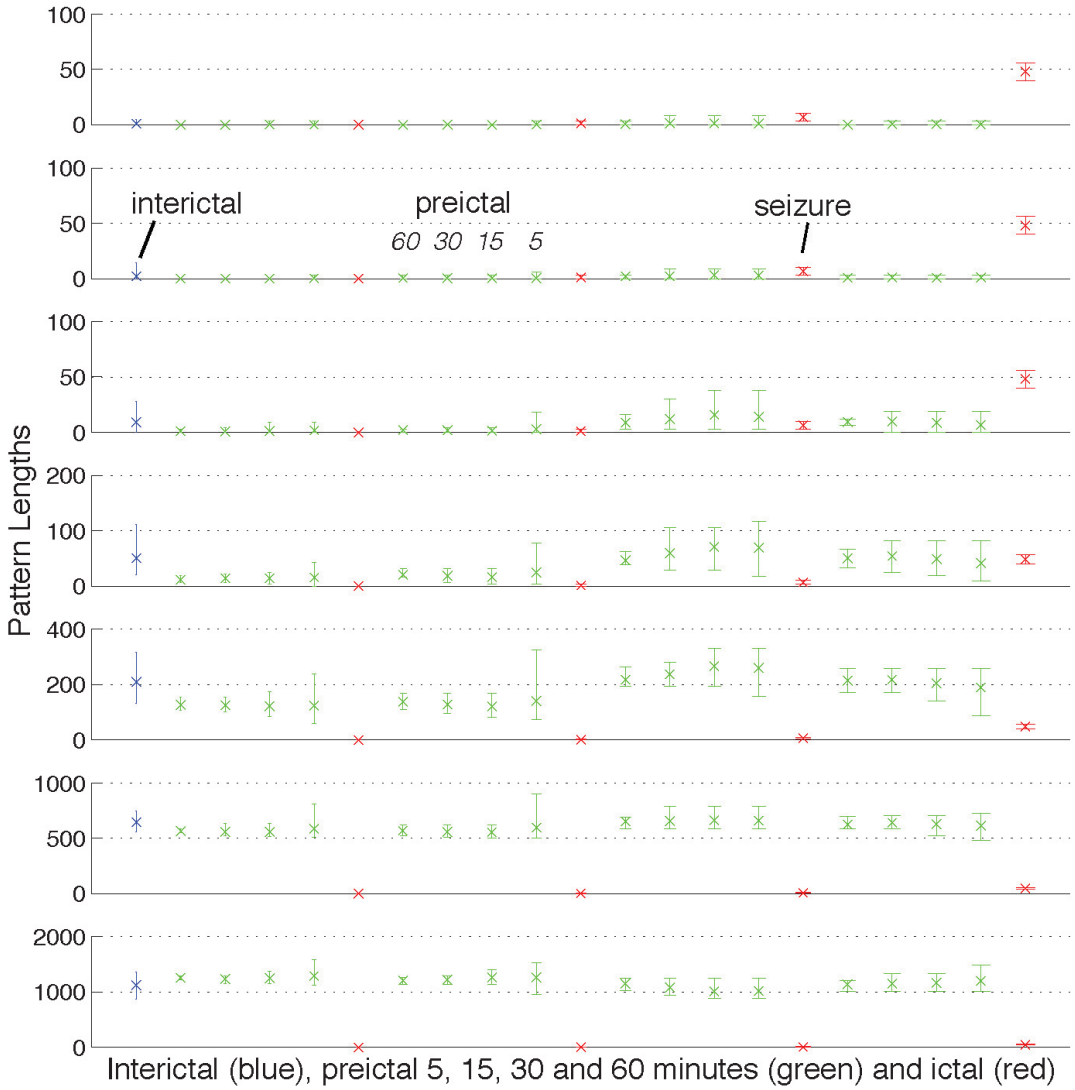


Figure 5.18 - Patient 1, all seizures with each of 4 seizures (red), the 4 preictal periods (in ascending order, green) and interictal (blue). The cross indicated the mean, with minimum and maximum indicated with the error bars. Each 7 plots are Ngram size 16 (top) to 4 (bottom) with intervals of 2.

We noticed the following observations:

1. For some specific pattern lengths, for example the 5th-7th subplots (ngram sizes 8, 6 and 4) we see that the ictal period is clearly differentiable from the interictal and preictal. Hence the seizure detection capability.
2. In the same subplots the max/min separation appears to decrease as we move towards the seizure (60 to 5 minutes).

3. There are some cases, where this evolution of max/min matches the interictal spread and moves outside the interictal period.

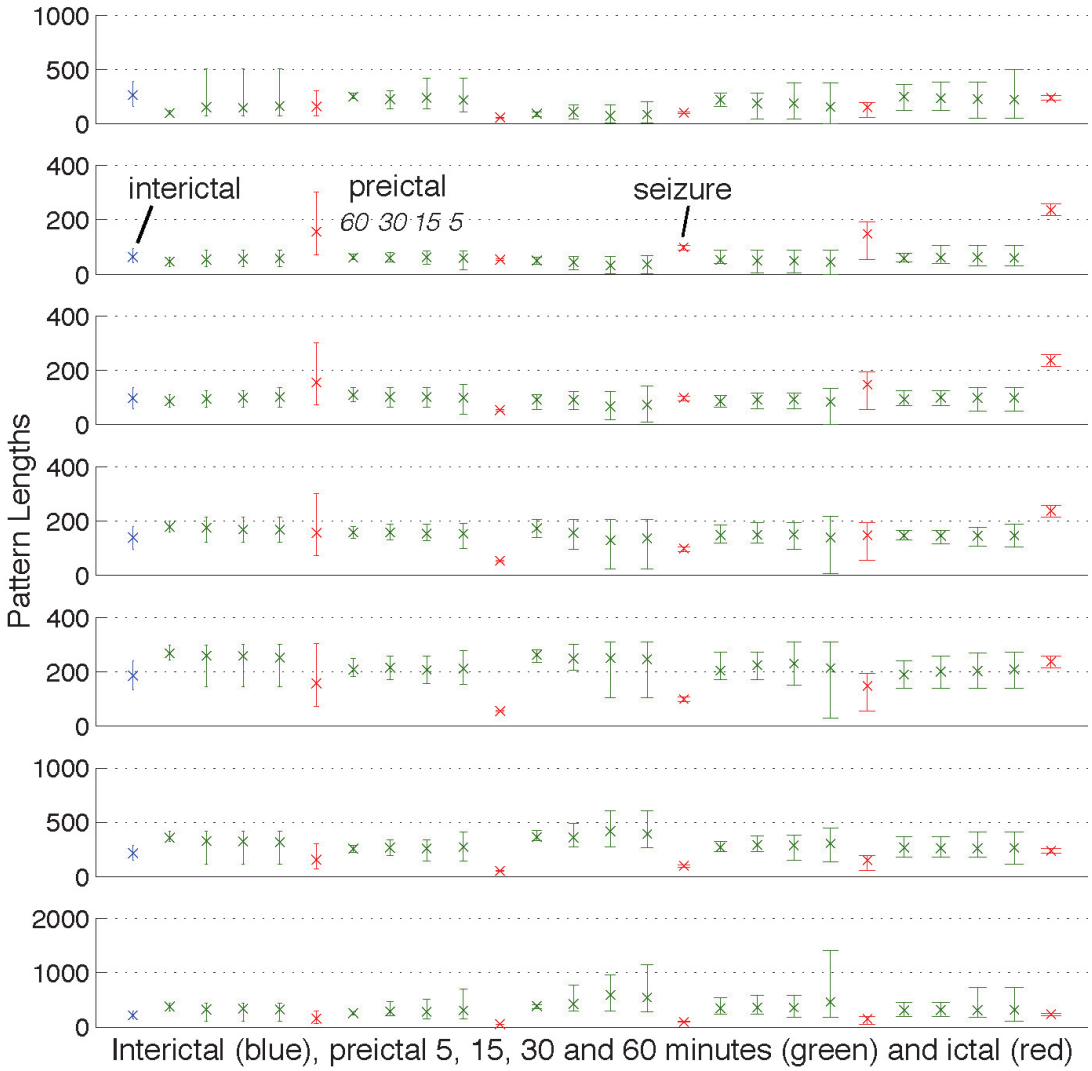


Figure 5.19 - Patient 10, all seizures with each of 4 seizures (red), the 4 preictal periods (in ascending order, green) and interictal (blue). The cross indicated the mean, with minimum and maximum indicated with the error bars. Each 7 plots are Ngram size 16 (top) to 4 (bottom) with intervals of 2.

Another example case for patient 10 is shown in Figure 5.19 where some of the subplots show similar pattern evolutions. These correlate to previous studies indicating that a level of synchrony between neurons occurs prior to the seizure. However, we have derived no

statistical proof for these case. Also the changes can be due to drifts in the signal leading up to the seizure. However, even a specific drift can be a preseizure warning.

First we look more closely at case 1, to divulge if what we are observing is a true indication of a prediction. This is shown in Figure 5.20 with indication of interictal mean, maxima and minima. It is clear that there are shifts to the seizure onset. Deriving whether these are statistically significant requires quantification and statistical validation.

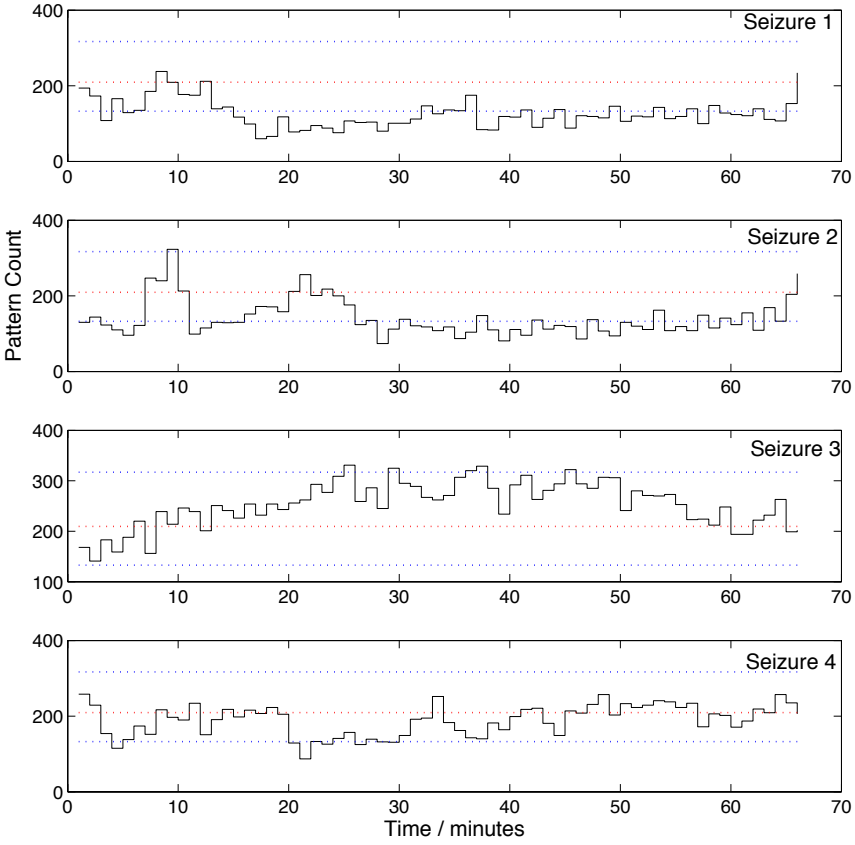


Figure 5.20 - Case 1, seizures 1 to 4, and indication of inteictal mean, maxima and minima.

To apply prediction in an automated fashion we first followed the same procedure as in our detection methods. We process each of the focal channels (three) separately. For a given patient and seizure an optimal value of the threshold is set for that seizure, then the same threshold is applied to the remaining seizures. The same multiple-detection per pattern is

followed. We repeated this using each seizure as the training seizure to have a statistical spread of the sensitivity and false predictions. The results are shown in Figure 5.21 and Figure 5.22 for a seizure prediction horizon of 40 and 20 minutes and occurrence period of 10 minutes (see Section 3.4.6 for details). We also computed a 60 minute and 10 minute IT.

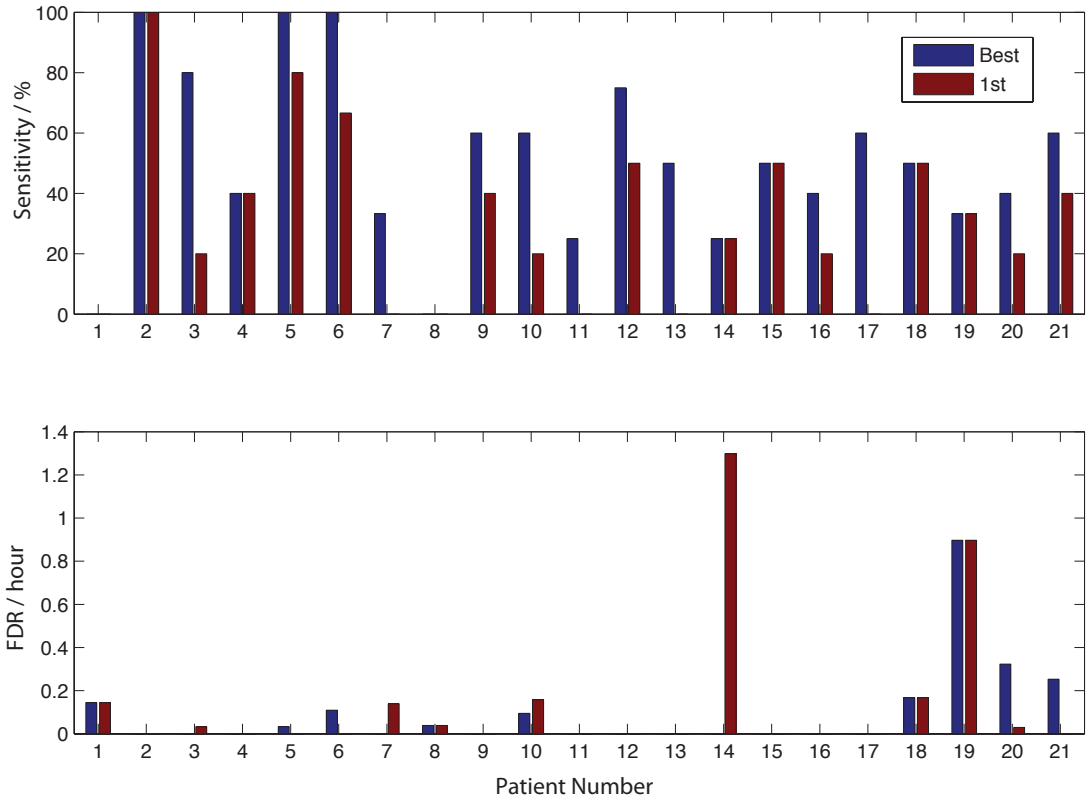


Figure 5.21 - Sensitivity and FDR for an IT of 40 minutes and SOP of 10.

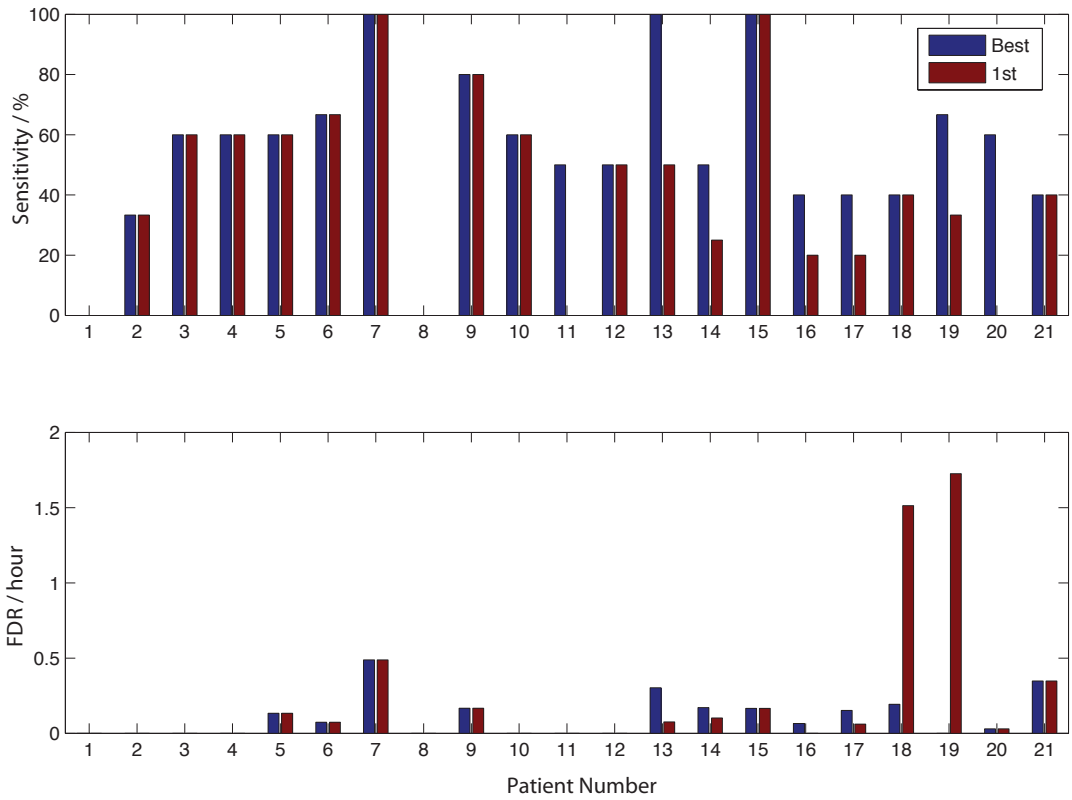


Figure 5.22 - Sensitivity and FDR for an IT of 20 minutes and SOP of 10.

For these IT periods we obtain sensitivities and FPR are:

IT (mins)	**60**		**40**		**20**		**10**	
	S	*FPR*	*S*	*FPR*	*S*	*FPR*	*S*	*FPR*
Best case	47.62%	0.3	36.43%	0.18	55.08%	0.11	55.95%	0.28
1st case	35.4%	0.39	22.46%	0.24	42.78%	0.23	36.35%	0.32

S: Sensitivity, **FPR**: False Prediction Rate (per hour)

Combining these results for the best of each patient, we obtain a sensitivity of 51.83% and FPR 0.08 when minimizing FPR and 65.32% and FPR of 0.29 when maximizing sensitivity.

5.4.4 Prediction Validation

With these prediction results, we can now apply the probability of predicting the same seizures with the same IT/SPH and SOP (Section 3.4.6).

As described in [6] for a FPR of 0.15 / hour (which is clinically relevant) and an SOP of 10 minutes, this yields values for each patient of their minimal and upper critical sensivity value as outlined in **Table 5.4**.

Table 5.4 - Lower and upper sensitivity (%) for two different SOP values. The significance of a seizure prediction result should be ideally place above the upper value.

	SOP = 10min	
Patient	**Lower**	**Upper**
1	25.00	25.00
2	33.33	33.33
3	20.00	20.00
4	20.00	20.00
5	20.00	20.00
6	33.33	33.33
7	33.33	33.33
8	0.00	50.00
9	20.00	20.00
10	20.00	20.00
11	25.00	25.00
12	25.00	25.00
13	0.00	50.00
14	25.00	25.00
15	25.00	25.00
16	20.00	20.00
17	20.00	20.00
18	20.00	20.00
19	25.00	25.00
20	20.00	20.00
21	20.00	20.00

Knowing this we can construct a plot to determine if our prediction results are above these critical sensitivities. Taking the best case IT for each patient, we obtain Figure 5.23.

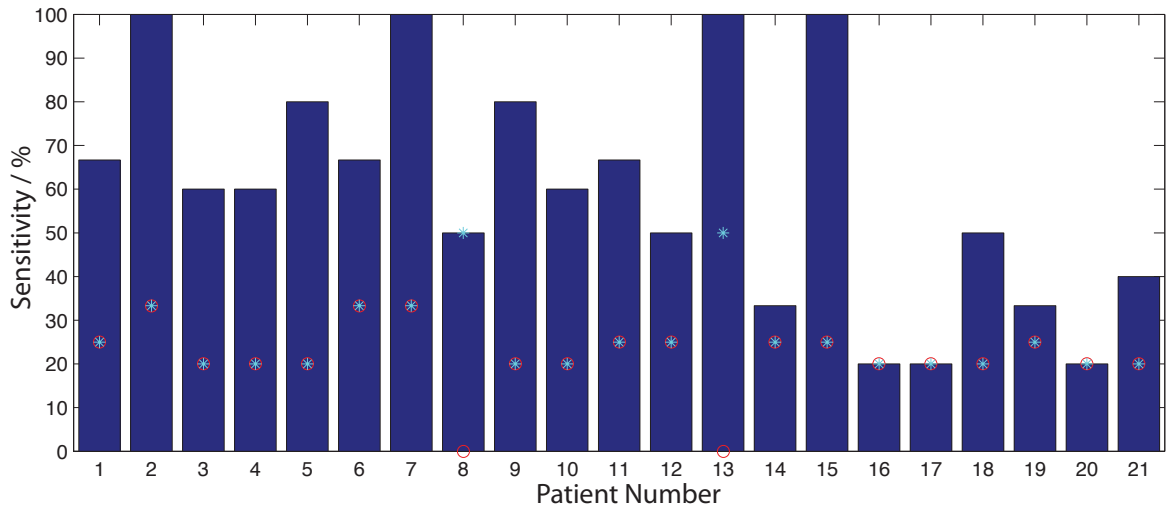


Figure 5.23 - Optimal sensitivity for each patient as a combination of all processed IT periods. Noted with an asterix (*) for the upper and circle (o) for the lower critical sensitivity values of Table 5.4, SOP of 10 minutes.

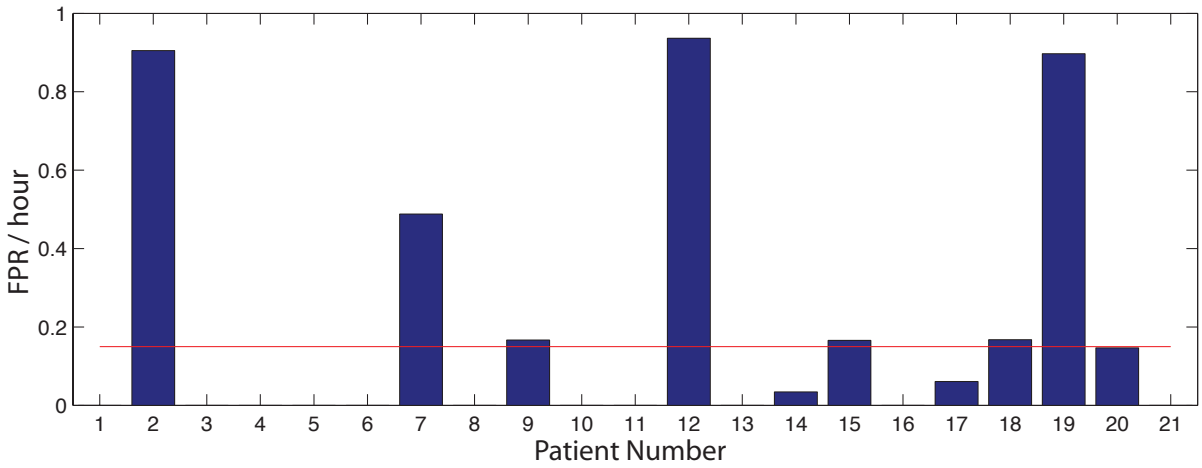


Figure 5.24 - FPR for the optimal sensitivity described in Figure 5.23 and the 0.15 FPR.

These results show that 17 out of 21 exceed the upper critical sensitivity of which only 11 also have an FPR below the 0.15 rate required. These results are comparable to those presented in [6] where overall a mean sensitivity of 35.2% (range 16.7-60%) for clinical seizures, 34.9% (18.8-83.3%) subclinical and 29.8% (18.2-58.3%) for both combined. Our results show larger numbers of patients exceeding the lower and upper critical value. Although they are not as high in sensitivity described in [7]. However the results of an IT of 5 and 10 minutes are not clearly described.

5.5 Discussions/Conclusions

In this chapter we have tested our methodologies to quantify seizure detection and prediction. We achieved this incrementally by analysing the data using various testing methods and criteria.

The last test, where we quantifiy statistically both detection and prediction is the most significant. We showed that in detection we can obtain significant results, similarly with prediction, where we showed statistical significance above chance level and with adequate FPR for just over 50% of cases.

Although these results are significant, they can still be improved by: optimising over more channels and features, improving thresholding for detection/prediction, adaptations of the algorithm to improve pattern uniqueness and improved data sets with better annotations. The latter will allow better understanding of the types and locations of seizure onset relating to our results.

On the other side of this we also must consider what the algorithm is doing and ways in which to improve it. Our pattern recognition technique involves extracting the counts of repeated sequence of electrical recordings. Hence, we are in effect measuring periodicity of features in a time window. Hence. It would make sense why we are able to detect most seizures, especially that those not detected lack clear periodic features.

Being periodic in this way we are partly resistant to single interictal spikes and artefacts of a similar sort. Also having removed baseline drift and some artefacts using the Savitzky-Golay smoothing function, we can be confident in the ability of our methods ability to be extracting significant events related to seizures. This, we have further proven with the statistical significant results.

5.6 References

[1] R. Aschenbrenner-Scheibe, T. Maiwald, M. Winterhalder, H. U. Voss, J. Trimmer, and A. Schulze-Bonhage, "How well can epileptic seizures be predicted? An evaluation of a nonlinear method," *Brain,* vol. 126, p. 11, 2003.

[2] T. Maiwald, M. Winterhalder, H. U. Voss, R. Aschenbrenner-Scheibe, J. Trimmer, and A. Schulze-Bonhage, "Comparison of three nonlinear seizure prediction methods by means of the seizure rpediction characteristic," *Physica D,* vol. 194, p. 368, 2004.

[3] M. Winterhalder, T. Maiwald, H. U. Voss, R. Aschenbrenner-Scheibe, J. Trimmer, and A. Schulze-Bonhage, "The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods," *Epilepsy Behaviour,* vol. 3, pp. 318-325, 2003.

[4] W. Juffali, A. Eftekhar, and C. Toumazou et al, "The WiNAM project: Neural data analysis with applications to epilespy," *IEEE BioCAS,* pp. 45-48, 2010.

[5] N. E. a. H. Clapp, L. M., "Method and Apparatus for Extraction of Low-Frequency Artifacts from Brain Waves for Alertness Detection," US Patent, 1997.

[6] H. Feldwisch-Drentrup, M. Ihle, M. Le Van Quyen, C. Teizeira, A. Dourado, J. Timmer, F. Sales, V. Navarro, A. Schulze-Bonhage, and B. Schelter, "Anticipating the unobserved: Prediction of subclinical seizures," *Epilepsy & Behavior,* vol. 22, pp. S119-S126, 2011.

[7] A. Aarabi and B. He, "A rule-based seizure prediction method for focal neocortical epilepsy," *Clinical Neurophysiology,* vol. 1, pp. 1111-1122, 2012.

Chapter 6- Conclusions and Future Work

In this work we have explored a methodology for analyzing biosignals that combines both an ICT infrastructure with a novel algorithmic implementation to address the future of healthcare innovation. We first presented some initial work on designing a system to measure the urea:creatinine ratio. Following from this we presented the vision of this work, namely how databasing and methods of analysis can be combined to form a personalized view of our future healthcare.

We applied this strategy with a novel algorithm and a database for assessing neural signals in cases of Epilepsy and although preliminary, the results showed the power of this methodology. In this final chapter we would like to summarise the various findings and contribution of each chapter and discuss the implications for future work that I will pursue.

6.1 Results - Discussions/Conclusions

Firstly we presented our work on a CMOS implementation of a urea to creatinine ratio sensor (Chapter 2). This work primarily showed that we could innovate novel solutions to healthcare problems in very simple ways. It also led to the foundation of this work, where we soon came to realize that this type of technology is a small portion of the personalized healthcare space. In conventional medicine there is a large reliance on databases and Information Technology to produce more efficient healthcare solutions. With this in mind we looked to expand our framework to this environment and from a business perspective what is missing.

Our concept is based on the need to gather a multitude of information available for one person to understand the person's health status as a whole. These would include

genetic, blood work and even neurological data. We encompassed this vision in an image (Figure 2.3) that shows how a person has many inputs and outputs that can feedback to form a critical evaluation of a person's healthcare status. This model is applicable to both personalized healthcare in the sense of self-evaluation to more in depth diagnostic and treatment solutions in the medical industry space. Technology such as the urea:creatinine sensor fall into this scope quite well. It is our aim to expand on this further and explore the need and model for such future generation of personalized healthcare systems.

With this vision in mind, we chose an application area to base our work on, namely neural signals (EEG and ECoG based) for the area of seizures in Epileptics. To do this we briefly covered the area (Chapter 3), from the clinical interpretation to the state of the art algorithmic implementations, to some of the current databasing systems emerging for the community to use. What was generally clear from this, was that that there are still unsolved elements and that a common database in which researchers and clinicians can work together is paramount to this area.

The implementation of our novel algorithm, based on an ngram approach, came in two forms. First was the software implementation in a Matlab environment, where we were able to show the initial workings of the algorithm, adapt it using some novel methods. We soon adapted this to a general SQL database online server to facilitate mass data upload, databasing and processing. This in itself is the foundation to our vision, allowing neural anomaly monitoring to be combined with efficient databasing.

Finally we applied this to real EEG-based data gathered from various sources (Chapter 5) aiming to quantify prediction and detection. Indeed with detection the key aspect of the method is that it is able to extract very good seizure detection with low false detections due to its resilience to artefacts.

Prediction is a difficult one to solve and indeed we have shown some predictive capability of this work. For future development to truly quantify the predictive value of our methods would rely on (1) well annotated data sets (including artifacts and feature descriptions) and (2) expert guidance and development from a neurologist or epileptologist. Our future work will follow this path to develop detection and prediction using our methods. Collaborations are being formed and developed for this purpose.

6.2 Thesis Contributions

This thesis has contributed in two key ways. It has:

1. Developed a novel pattern recogintion algorithm suitable of quantifying signal dynamics, and defining anomalous behaviour.
2. A novel system implementation that integrates both databasing and algorithmic implementation to facilitate the study and real-time analysis of patients biosignals.

These two contribution are both novel and succeeded in reaching the aims and objectives of this work. By providing both parts we now have a fully workable system to analyse biosignals in real-time for several applications.

6.3 Future work

Based on these contributions we aim to further develop this work in two parallel fronts. The first is in pursuit of our neural signal analysis for quantifying seizure detection and prediction. We shall pursue this with collaborations from leading Epilepsy clinics, including EPI Zurich and King's College London. With better annotated data and clinical expertise to guide the analysis we can create a system for all researchers and clinicians to use.

Another keen interest that has developed from this work, is the integration of personal biomarkers as a whole (including neural state) into diagnosis. This includes genetic traits, physiological parameters and metadata (family history) all integrated into a system that defines your personalized healthcare status. In the next section we briefly describe a project that works towards this aim that we will aim to pursue.

6.4 WiNORM

Personalised healthcare is rapidly moving beyond theoretical application into real-world medicine. Advances in understanding how genetic differences influence treatment and prevention of disease have leapt forward based on the understanding of the human genome. DNA sequencing and genetic profiles that used to take rather long to generate can now be defined in a short period of time. At this rate of discovery, sequencing a person's whole genome at a price lower than $10,000 is now achievable. However, despite these advances in understanding the genetic basis of disease, the post-genome era will need to focus on understanding how gene and protein networks interact with environmental stimuli to create complex human diseases. Here we emphasized the importance of predictive, preventive and personalized healthcare for patients. Greater understanding of the immediate application and future potential of these tools is necessary to affect the transformational benefit for patients receiving individualized genetically-based medical care.

Our aim is to create a database that generates the "normal values" for each individual. Human illness is not solely determined by genetic inheritance. Interaction with environment and behaviour using sensor data is needed. Integrating results from various data modalities and knowledge retrieval engines are still challenging for clinically meaningful use. This database will take in account every specificity of each patient (Blood

test, genetic test, disease markers, ethnicity and lifestyle), combine intelligibly and provide this patient with new unique "normal value" applicable for diagnosis or simply health monitoring and management.

There will be two major application of this database;

1 Business application

- NHS government / clinicians or GPs
- Better handling of patients results and ease the analysis. Cost effective

2 Patient benefit

Considering the patient benefits, here we propose to consider the uniqueness of the individual patient and use the full range of the sciences and knowledge to provide the patient with personalised "normal values". This knowledge will provide the patient with an appropriate and unique solution to manage their own health. The health management will consist of data consultation among peers and the giving of advice based on expertise, rather than the normal model as used nowadays. Decision support would consist of access to a wide variety of knowledge resources in the databases, as well as to tools for creative problem solving and research.

A few key numbers:

The databases we are designing have great potential to help Medical professionals to do the best diagnosis. Indeed the NHS published a few results about the lack of good diagnosis within different departments.

- **50%** of medicaton errors are a result of insufficient information.
- **1 in 16** hospital admissions are a result of a mistaken diagnosis.
- **10.8%** of patients experience an adverse event on medical wards.
- **3rd** most common patient safety incident: medication errors.

Genetic disorders

Both environmental and genetic factors have roles in the development of any disease. A genetic disorder is a disease caused by abnormalities in an individual's genetic material (genome). The four different types of genetic disorders are (1) single-gene, (2) multifactorial, (3) chromosomal, and (4) mitochondrial.

(1) Single-gene (also called Mendelian or monogenic) - This type is caused by changes or mutations that occur in the DNA sequence of one gene. Genes code for proteins, the molecules that carry out most of the work, perform most life functions, and even make up the majority of cellular structures. When a gene is mutated so that its protein product can no longer carry out its normal function, a disorder can result. There are more than 6,000 known single-gene disorders, which occur in about 1 out of every 200 births. Some examples are cystic fibrosis, sickle cell anemia, Marfan syndrome, Huntington's disease, and hereditary hemochromatosis.

Single-gene disorders are inherited in recognizable patterns: autosomal dominant, autosomal recessive, and X-linked. More information on the different modes of inheritance is available from the following Web sites:

- Inheritance of Single Gene Defects - From the *The Merck Manual of Diagnosis and Therapy*.
- Inheritance Patterns of Monogenic Disorders - From the Genetic Interest Group in the U.K.
- Genetics - From the Medical Encyclopedia at MEDLINEplus.

(2) Multifactorial (also called complex or polygenic) - This type is caused by a combination of environmental factors and mutations in multiple genes. For example,

different genes that influence breast cancer susceptibility have been found on chromosomes 6, 11, 13, 14, 15, 17, and 22. Its more complicated nature makes it much more difficult to analyze than single-gene or chromosomal disorders. Some of the most common chronic disorders are multifactorial. Examples include heart disease, high blood pressure, Alzheimer's disease, arthritis, diabetes, cancer, and obesity. Multifactorial inheritance also is associated with heritable traits such as fingerprint patterns, height, eye color, and skin color.

(3) Chromosomal - Chromosomes, distinct structures made up of DNA and protein, are located in the nucleus of each cell. Because chromosomes are carriers of genetic material, such abnormalities in chromosome structure as missing or extra copies or gross breaks and rejoinings (translocations) can result in disease. Some types of major chromosomal abnormalities can be detected by microscopic examination. Down syndrome or trisomy 21 is a common disorder that occurs when a person has three copies of chromosome 21.

(4) Mitochondrial - This relatively rare type of genetic disorder is caused by mutations in the nonchromosomal DNA of mitochondria. Mitochondria are small round or rod-like organelles involved in cellular respiration and found in the cytoplasm of plant and animal cells. Each mitochondrion may contain 5 to 10 circular pieces of DNA.

An increasing number of gene tests are becoming available commercially, although the scientific community continues to debate the best way to deliver them to the public and to medical communities that are often unaware of their scientific and social implications. While some of these tests have greatly improved and even saved lives, scientists remain unsure of how to interpret many of the test results. Also, patients taking the tests may face significant risks, including jeopardizing their employment or insurance status. And because genetic information is shared, these risks can extend to family members as well. Developing

Winorm we propose to give a full understanding of these genetics tests by combining them with other markers and translating them in real normal value for the one specific patient.

Disease Markers

To make an accurate diagnosis a great deal of effort has been made in search of noninvasive tests for early identification and classification of the type of disease, to predict the severity and outcome as well as the patients' response to treatment. A major goal has been to identify biomarkers (proteins and molecules) that can be measured inexpensively for early diagnosis of disease in patients. So far a list of biomarkers for each disease is available and will be one of WiNorm input.

Bench mark Blood test

- Blood Lipids
    - Total Cholesterol in Vegans in the United States
    - Total Cholesterol in Western Vegans
    - Total Cholesterol: Do Vegan Men and Women Differ?
    - Why Are Vegans' Cholesterol Levels Lower?
    - LDL & HDL Cholesterol of Western Vegans
    - Triglycerides
- Blood Pressure
- Body Mass Index
- Body Fat

Databases already available

Database efforts have kept pace with the rapid rate at which this sequence data is being generated, providing investigators access to all public data in a practically instantaneous fashion. While most biologists are familiar with the databases comprising the International Nucleotide Sequence Database Collaboration (DDBJ, EMBL and GenBank), numerous other specialised databases have emerged. These specialized databases often arise out of a particular need, whether it is to address a particular biological question of interest or to better serve a particular segment of the biological community. So far no database is able to give an easy readout to the patient. Although we saw companies like 23andme[7] giving an idea about gene susceptibility to disease our database will go much further providing specific and customized information directly to the patients or medical professionals.

System

To create this database we are using the same strategy of system biology. The aim is to understand the body as a whole.

- Translation

  We will use experimental therapeutics results to implement data base and device design.

- Application and implementation

  Application and implementation – health care delivery

- Dissemination

  Provide personalized medicine at the community and individual user level

[7] https://www.23andme.com/

Applied Methods

We designed one model to organize a comprehensive collection of data based on 3 different levels; blood test results, diseases markers and genes analysis (Figure 6.1). In order to obtain a readable output we will analyse the input, by applying logic and linking the contents using an algorithm (Figure 6.2). The following steps comprise a proof of concept. We will do a case control study with one disease and validate the results.

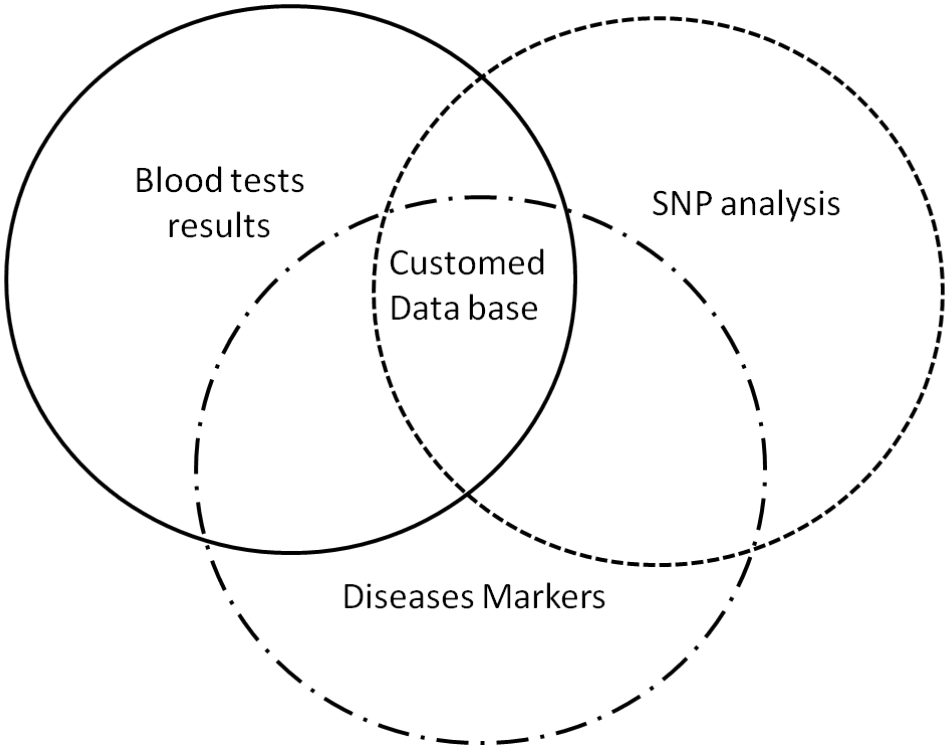


Figure 6.1 - WiNorm Database concept

Database design

The Database Design Process is a strategic decision making and business planning require accurate and manageable data. In order to build this database we needed to follow the steps presented here (Figure 6.3 and Figure 6.3).

1) Formulate an idea for a solution

- Identify an existing problem that could be improved with a database.
- Have a clear purpose in mind.
- Identify the problem:
- Describe existing difficulties:
- Propose a solution: Describe what can be done to make this data easier to utilize. Should anything else be a focus here?

2) Gather information and clarify needs.

- Talk to people currently involved with the data and anyone else who may have the potential to be a user of the application.
- Ask questions such as what information is available? Is it adequate to meet needs?
- How is the available information being used?
- Who maintains current data?
- Sketch out formatting of possible forms and/or reports (mailing labels, rosters, course offerings list, etc.) to make sure they are useful to users.

3) Develop a list of all useful fields.

- List all of the possible attributes that will be available within the database
- Break down information into the smallest meaningful units possible
- Eliminate redundancy
- Determine if all proposed fields will be needed

4) Organize the fields into logical tables.

- Divide fields into categories or subjects. Each of these categories will represent a table.

- Assign key fields. A key field or **primary key** is a unique identifier for a record/entity
- Consider normalization rules.
- Organize fields into tables**:** Create table names and list fields for each table.

5) Determine and define table relationships.

- Identify fields that match among tables
- Decide how data is related based on these matching fields
- Define relationships between the primary key of one table and the related or foreign key of another. These matching fields can have the exact same name or slightly different names, yet hold the same data.
- Add new fields and/or tables as needed to clarify relationships.
- Consider how relationships may be used to make queries and reports more efficient.
- Focus on the three types of table relationships.

Three Types of Table Relationships: There are three types of table relationships. Each has its own unique purpose in helping to organize the data within the database. These relationships can be used to determine joins in queries as well.

1) One-to-One Relationships: one record in a table is related to one

record in a related table; creates equally dependent tables

2) One-to-Many Relationships: one record in a primary table is

related to many records in a related table; however, a record in the

related table has only one related record in the primary table

3) Many-to-Many Relationships: several records in the primary table

are related to several records in a related table

Define relationships among tables: Draw lines to connect primary and foreign

key fields on related tables. This diagramming is referred to as ***Entity Relationship*** diagramming because it connects entities based on similarities.

6) Test, refine and improve.

- Analyze the overall database design to check for errors as well as to see that needs are being met.
- Review table structure to make sure all needed fields are present and existing relationships are valid and useful.
- Test table design by adding records and observing functionality.
- List points to consider when testing, refining, and improving design:
- Describe possible issues and resolutions.

7) Develop queries, forms, and reports.

- Look back at information gathered to get ideas for what type of queries, forms, and reports are needed.
- Consider table relationships and how they can be utilized.
- Try drafts of queries, forms, and reports to see if they show the expected data results.
- Modify as needed to meet needs.
- Consider this development process as ongoing throughout the use of the database.

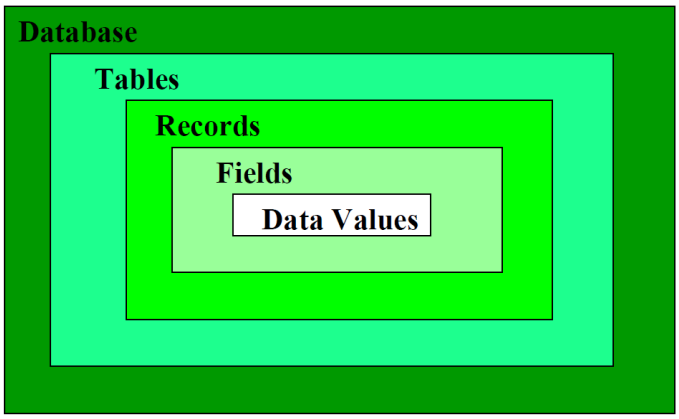


Figure 6.2 - Database design concept

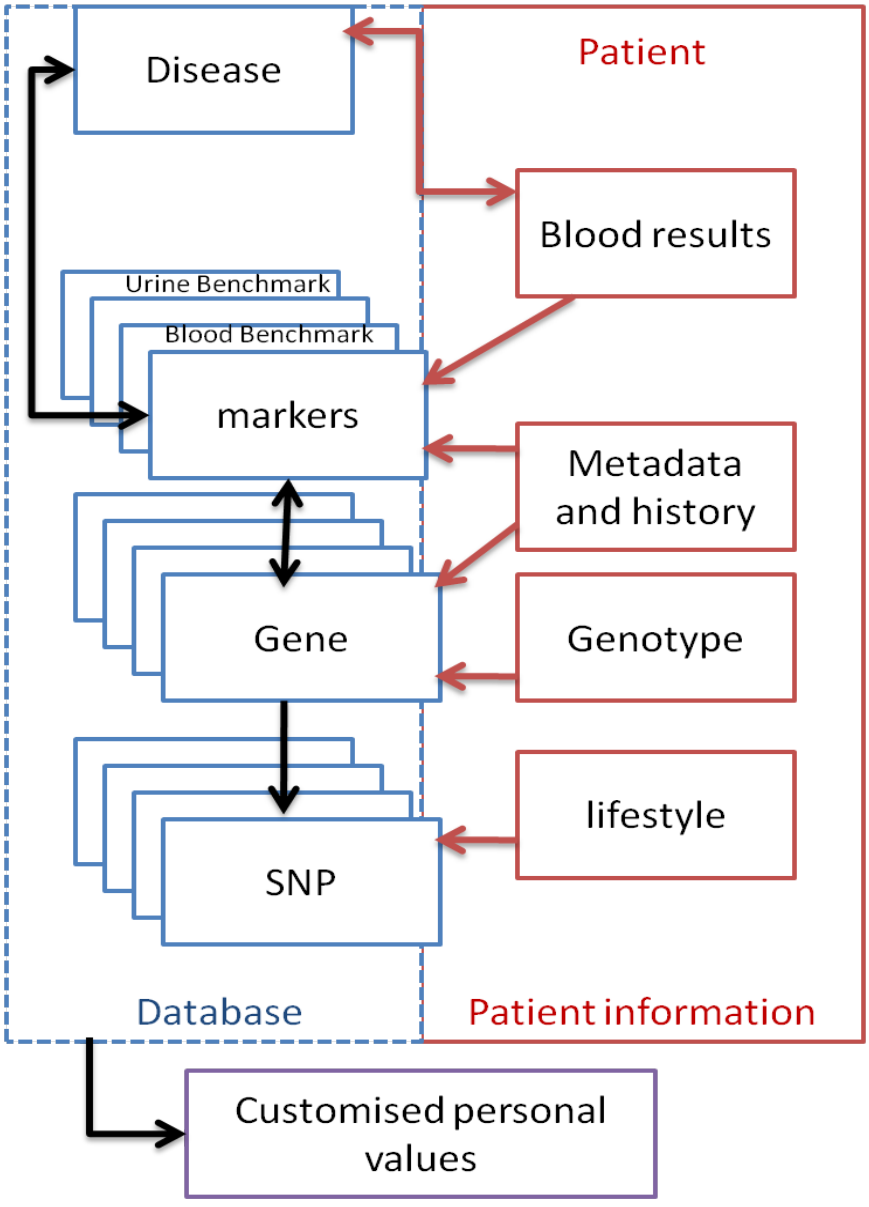


Figure 6.3 - Diagram of WiNorm database

6.5 Final Comments

What we have been able to show in this work is that by combining novel methods with efficient databasing we can successfully apply this to real data, for high impact applications. This work has a lot of future potential that we aim to leverage on and follow through with. Overall, I am confident that with the strong inter-disciplinary group that I have worked with on this thesis we will produce life-changing healthcare solutions for the future of medicine.

Appendix A – Matlab Code

This code was developed in collaboration Dr. Amir Eftekhar at the Centre for Bio-Inspired Technology.

DATA ADC

```
function [x_ADC] = dataADC(x,Vmax,Vmin,Nbit,plot_chk)

if nargin < 4
    Nbit = 8;
    plot_chk = 0;
end

x = x + abs(Vmin);

x_ADC = ADC(x,Nbit,0,Vmax-Vmin);

x_ADC = x_ADC - (2^Nbit)/2;

x_ADC = round(x_ADC);

%plot(x_ADC);
if plot_chk == 1
    figure;
    subplot(211)
    plot(x);

    subplot(212)
    plot(x_ADC,'--r')
end



function y = ADC(x,N,Vlow,Vhigh)

ADC_high = 2^N;

value = ((x - Vlow)./(Vhigh - Vlow)).*(ADC_high-1);

y = value;
```

HEX CONVERSION

```
function [DATA_HEX] = HEX_CONVERT(DATA,DATA_MAX,DATA_MIN,BITS)

%CONVERT DATA to HEX as defined from BIT_SIZE

%Variable s contains the HEX arrays for each
```

```
% DATA_MAX = 32767;
% DATA_MIN = -32768;


DATA_HEX = data2HEX(DATA,DATA_MAX,DATA_MIN,BITS,0);


function [s] = data2HEX(x,Vmax,Vmin,Nbit,plot_chk)

if nargin < 4
    Nbit = 8;
    plot_chk = 0;
end

x = x + abs(Vmin);

x_ADC = ADC(x,Nbit,0,Vmax-Vmin);

x_ADC = x_ADC - (2^Nbit)/2;

x_ADC = round(x_ADC);

%plot(x_ADC);
if plot_chk == 1
    figure;
    plot(x);
    hold on
    plot(x_ADC,'--r')
end

s = sdec2hex(x_ADC,Nbit);

function y = ADC(x,N,Vlow,Vhigh)

ADC_high = 2^N;


value = ((x - Vlow)./(Vhigh - Vlow)).*(ADC_high-1);

y = value;

function [y] = sdec2hex(x,N)

max_N = (2^N) - 1;

ind = find(x<0);
x(ind) = max_N + x(ind) + 1;

M=ceil(N/4);

y = repmat(char(M),1,length(x)*M);
if any(round(x) < 0)
```

```
    print('Ahhhhh');
end

s = char(dec2hex(round(x),M));

indx = 1:length(x);
for i=1:M
    indy = (indx.*M) - (M - i);
    y(indy) = s(:,i)';
end
```

SIGNED DECIMAL to HEX

```
function [y] = sdec2hex(x,N)

max_N = (2^N) - 1;

ind = find(x<0);
x(ind) = max_N + x(ind) + 1;

M=ceil(N/4);

y = repmat(char(M),1,M*length(x));

if any(round(x) < 0)
    print('Ahhhhh');
end

s = char(dec2hex(round(x),M));

indx = 1:length(x);

for i=1:M
    indy = (indx.*M) - (M - i);
    y(indy) = s(:,i)';
end
```

NW ALGORITHM

```
function [Score,Mx] = nw_alg(seq1,seq2)

Sij = 2;
nSij = 0;
w = -1;

ls1 = length(seq1);
ls2 = length(seq2);

Mx = zeros(ls1+1,ls2+1);
```

```
for i = 2:ls1+1
    for j = 2:ls2+1

        if strmatch(seq1(i-1),seq2(j-1))
            S = Sij;
        else
            S = nSij;
        end

        Mx(i,j) = max([Mx(i-1,j-1)+S Mx(i,j-1)+w Mx(i-1,j)+w]);

    end
end

Score = Mx(ls1+1,ls2+1)/sum([ls1 ls2]);
```

BASIC SINGLE PATTERN NGRAM

```
function [NGRAM_array,NGRAM_cnt] = NGRAM(HEXdata,Nhex,Ngram_size)
%% Function NGRAM
%
% Inputs:
%   HEXdata - characters of hex in array Nx1
%   Nhex - The data HEX size to use
%   Ngram_size - Token length to use
%
% Outputs:
%   NGRAM_array - Array of HEX patterns found (Ngram_size*Nhex)xM
%   Ngram_cnt - number of times each of M patterns repeats
%
% Designed by Amir Eftekhar on 8th February 2010

LEN_data = size(HEXdata,1);

Nforloop = floor(LEN_data/Nhex);

iarray = 1;
N = Ngram_size;
NGRAM_array = repmat(char(0),LEN_data,N*Nhex);
NGRAM_cnt = ones(1,LEN_data);

k=1;
%NGRAM_cnt = zeros(1,length(HEXdata));

for i = 1:(Nforloop- N+1)
    PAT_array = HEXdata(k:(k+(Ngram_size*Nhex)-1),:)';

    if i == 1
        NGRAM_array(iarray,:) = PAT_array;
        NGRAM_cnt(1) = 1;
```

```
        iarray = iarray+1;
    else

        indx = strmatch(PAT_array,NGRAM_array(1:iarray-
1,:),'exact');

        if ~isempty(indx)
            NGRAM_cnt(indx) = NGRAM_cnt(indx)+1;
        else
            NGRAM_array(iarray,:) = PAT_array;
            NGRAM_cnt(iarray) = 1;
            iarray = iarray+1;
        end
    end

    k = k + 2;
end

NGRAM_array = NGRAM_array(1:iarray-1,:);
NGRAM_cnt = NGRAM_cnt(1:iarray-1);
```

BASIC MULTIRESOLUTION NGRAM IMPLEMENTATION

```
function [NGRAM_array,NGRAM_cnt] = NGRAMv3(DATA,Ngram_size,Nbit)
%% Function NGRAM
%
% Inputs:
%   HEXdata - characters of hex in array Nx1
%   Nhex - The data HEX size to use
%   Ngram_size - Token length to use
%
% Outputs:
%   NGRAM_array - Array of HEX patterns found (Ngram_size*Nhex)xM
%   Ngram_cnt - number of times each of M patterns repeats
%
HEXdata = HEX_CONVERT(DATA,10000,-10000,Nbit)';

LEN_data = length(HEXdata);
Nhex = LEN_data/length(DATA);

Nforloop = floor(LEN_data/Nhex);

iarray = ones(1,length(Ngram_size));
N = min(Ngram_size);
NGRAM_array = cell(1,length(Ngram_size)); %
repmat(char(0),LEN_data,N);
NGRAM_SIZE = 0;
NGRAM_cnt = cell(1,length(Ngram_size)); % ones(1,LEN_data);
NGRAM_cellarray = cell(Nforloop- N+1,length(Ngram_size));
```

```
min_i =  Nforloop - Ngram_size +1;

k=1;
%NGRAM_cnt = zeros(1,length(HEXdata));

for i = 1:(Nforloop- N+1)

    stop_flag = 0;
    for j = 1:length(Ngram_size)

        if (stop_flag == 0) && (k <= Nhex*min_i(j))

            ind = k:(k+(Ngram_size(j)*Nhex)-1);
            PAT_array = HEXdata(ind,:)';

            if i == 1
                NGRAM_array{j}(iarray(j),:) = PAT_array;
                NGRAM_cellarray{iarray(j),j} = PAT_array;
                NGRAM_cnt{j}(1) = 1;
                iarray(j) = iarray(j)+1;
            else

                %indx = strmatch(PAT_array,NGRAM_array(1:iarray-
1,j),'exact');
                indx = strcmp(PAT_array,NGRAM_cellarray(:,j));
                indx = find(indx ~= 0);

                if indx ~= 0
                    NGRAM_cnt{j}(indx) = NGRAM_cnt{j}(indx)+1;
                    NGRAM_SIZE = Ngram_size(j);
                    stop_flag = 1;
                else
                    NGRAM_array{j}(iarray(j),:) = PAT_array;
                    NGRAM_cellarray{iarray(j),j} = PAT_array;
                    NGRAM_cnt{j}(iarray(j)) = 1;
                    iarray(j) = iarray(j)+1;
                end
            end

        end

    end
    %k = k + Ngram_size*Nhex;
    if stop_flag == 1
        k = k + NGRAM_SIZE*Nhex;
    else
        k = k + Nhex;
    end
end
%NGRAM_array = NGRAM_array(1:iarray-1,:);
%NGRAM_cnt = NGRAM_cnt(1:iarray-1);
```

PROCESS NGRAM WITH WINDOWED ANALYSIS

```
function [RAT1,RAT2,T,NGRAM_array,NGRAM_cnt] =
NGRAM_realtime(x,TH)

WIN_LENGTH = 256*60;
x_LENGTH = length(x);

Nloop = floor(x_LENGTH/WIN_LENGTH);

NGRAMs = [10:-1:7];
BINARY = [7];
MIN_PAT_CNT = 2;
NORMALISE = 0;

if nargin < 2
   TH = 0.95*ones(1,length(NGRAMs));
   %TH = round(NGRAMs .*(2/3))./NGRAMs;
else
   TH = round(NGRAMs .*(TH))./NGRAMs;
end

%RAT1 = zeros(Nloop,length(NGRAMs));
%RAT2 = zeros(Nloop,length(NGRAMs));
%disp(Nloop);
n = round(Nloop.*([10 20 30 40 50 60 70 80 90]/100));
N = length(NGRAMs);

RAT2 = zeros(length(NGRAMs),Nloop);
RAT1 = zeros(length(NGRAMs),Nloop);

parfor i = 1:Nloop
    disp(i)
%     if any(i == n)
%         disp(round((i/Nloop)*100));
%     end
%
    ind = ((i-1)*WIN_LENGTH + 1) : (i*WIN_LENGTH );

    [NGRAM_array{i},NGRAM_cnt{i}] =
NGRAMv3(x(ind),NGRAMs,BINARY);

%     for j = 1:length(NGRAMs)
%         ind2 = find(NGRAM_cnt{j} > MIN_PAT_CNT);
%         CNT2 = CNT2 + sum(NGRAM_cnt{j}(ind2));
%         CNT1 = CNT1 + length(ind2);
%
%     end
```

```
%
%      RAT1(i) = CNT1;
%      RAT2(i) = CNT2;
%      for j = 1:length(NGRAMs)
%         ind2 = find(NGRAM_cnt{i}{j} > MIN_PAT_CNT);
%         RAT2(j,i) = sum(NGRAM_cnt{i}{j}(ind2));
%         RAT1(j,i) = length(ind2);
%      end
     for j = 1:N
        ind2 = find(NGRAM_cnt{i}{j} > MIN_PAT_CNT);
        RAT2(j,i) = sum(NGRAM_cnt{i}{j}(ind2));
        RAT1(j,i) = length(ind2);
     end

end

%T = 1:WIN_LENGTH:length(x);
```

PROCESS DETECTION CRITERIA OF RESULTS

```
load('/Volumes/Neural Data/Freiburg/Tests 12th July - All Preictal
Available/ascii_file_info_and_prepost_seiz.mat')
cd('/Volumes/Neural Data/Freiburg/Tests 12th July - All Preictal
Available/NGRAM_processed_v1_256smooth_60s');
iiDataDir = ('/Volumes/Neural Data/Freiburg/Tests 12th July - All Preictal
Available/interIctal_Processed_v1_256smooth_60s');


detected_final = cell(1,21);
fd_final = cell(1,21);
k = 1;



for i = 1:21


    m = k;

    cdir = pwd;

    cd(iiDataDir)
    d = dir(['Case' int2str(i) '_*.mat']);

    ii_ind2use = 1;

    disp('Loading Interictal Data...');
    xii = cell(1,length(d));

    for g = 1:size(d,1)
         xii{g} = load(d(g).name);
     end

    disp('Done');
```

```
    cd(cdir);


    for j = 1:seiz_num(i)
        disp(['Patient' int2str(i) '_Seiz' int2str(j)])

        for p = 1:3

            disp(['channel' int2str(p)]);

            load(['Case' int2str(i) '_seiz' int2str(j) '_ch' int2str(p)
'_NGRAM_processed.mat']);
            range = ((preseizure_available(k) + (0:(seiz_ind_end(k)-
seiz_ind_start(k)))) / (256*60));
            range = [range(1)-1 range range(end)+1];


            ind_RAT2use = 1:min([ceil(range(end))+1 length(RAT2)]);

            detected = zeros(1,size(RAT2,1));
            fd = ones(1,size(RAT2,1));
            fdii = ones(1,size(RAT2,1));

            n = zeros(1,size(RAT2,1));
            TYPE = cell(1,size(RAT2,1));

            for v = 1:size(RAT2,1)

                th = smooth_amir(RAT2(v,ind_RAT2use),5);
                thii = smooth_amir(xii{ii_ind2use}.RAT2{p}(v,:),5);

                % USE CURRENT SEIZURE AS TRAINING
                STOP = 1;

                fd1 = 1;
                fd2 = 1;
                fdii1 = 1;
                fdii2 = 1;

                detected1 = 0;
                detected2 = 0;

                iter = 1;
                n1 = 0;
                n2 = 0;



                while ~((detected(v) == 1) && (fd(v) == 0))  && STOP

                    thtn = th/(1+n1);
                    thtp = th*(1+n2);

                    thtnii = thii/(1+n1);
                    thtpii = thii*(1+n2);

                    detected_prev1 = detected1;
```

```
                    fdp1 = fd1;

                    detected_prev2 = detected2;
                    fdp2 = fd2;

                    [detected1,fd1] =
det_seiz(RAT2(v,ind_RAT2use),thtn,range,'Neg');
                    [detected2,fd2] =
det_seiz(RAT2(v,ind_RAT2use),thtp,range,'Pos');


                    [~,fpii1] =
det_seiz(xii{ii_ind2use}.RAT2{p}(v,:),thtnii,range,'Neg');
                    [~,fpii2] =
det_seiz(xii{ii_ind2use}.RAT2{p}(v,:),thtpii,range,'Pos');

                    if ~((detected1 == 1) && (fd1 == 0) && (fdii1 == 0))
                        n1 = n1 + 0.01;
                    else
                        detected(v) = detected1;
                        fd(v) = fd1;
                        fdii(v) = fdii1;
                        TYPE{v} = 'Neg';
                        n(v) = n1;
                    end

                    if ~((detected2 == 1) && (fd2 == 0) && (fdii2 == 0))
                        n2 = n2 + 0.01;
                    else
                        detected(v) = detected2;
                        fd(v) = fd2;
                        fdii(v) = fdii2;
                        TYPE{v} = 'Pos';
                        n(v) = n2;
                    end

                    if (detected_prev1 == 1) && (detected1 == 0) && (detected2
== 0)
                        detected(v) = detected_prev1;
                        fd(v) = fdp1;
                        fdii(v) = fdii1;
                        n(v) = n1 - 0.02;
                        STOP = 0;
                        TYPE{v} = 'Neg';
                    elseif (detected_prev2 == 1) && (detected2 == 0) &&
(detected1 == 0)
                        detected(v) = detected_prev2;
                        fd(v) = fdp1;
                        fdii(v) = fdii2;
                        n(v) = n2 - 0.02;
                        STOP = 0;
                        TYPE{v} = 'Pos';
                    end

                    iter = iter + 1;
                    if iter > 10000
                        STOP = 0;
                        disp('Iterations too high');
                    end

                end
```

```
            end

            ind_fd = find((fd< 5) & (detected == 1));

            if isempty(ind_fd)
                [M,ind_fd] = min(fd);
            end


            fdii_temp = zeros(length(ind_fd),size(d,1));
            ind_fdii = [];
           int_fdii = zeros(1,size(d,1));
           cdir = pwd;
           cd(iiDataDir)
           disp('Processing Interictal Data');
           for g = 1:size(d,1)


               for u = 1:length(ind_fd)

                   if strmatch(TYPE{ind_fd(u)},'Neg')
                       thii = smooth_amir(xii{g}.RAT2{p}(ind_fd(u),:),5) /
(1+n(ind_fd(u)));
                   else
                       thii = smooth_amir(xii{g}.RAT2{p}(ind_fd(u),:),5) *
(1+n(ind_fd(u)));
                   end


                   [~,fdii_temp(u,g),temp] =
det_seiz(xii{g}.RAT2{p}(ind_fd(u),:),thii,range,TYPE{ind_fd(u)});

                   ind_fdii = [ind_fdii temp];

               end


               un = unique(ind_fdii);

               for q = 1:length(un)
                   temp = sum(ind_fdii == un(q));

                   if temp >= round(length(ind_fd)/1.6)
                        int_fdii(g) = int_fdii(g) + 1;
                   end
               end

end

           cd(cdir);


            %PROCESS THE REMAINING
            for l = 1:seiz_num(i)


                load(['Case' int2str(i) '_seiz' int2str(l) '_ch' int2str(p)
```

```
'_NGRAM_processed.mat']);
                range = ((preseizure_available(m+l-1) + (0:(seiz_ind_end(m+l-
1)-seiz_ind_start(m+l-1)))) / (256*60));
                range = [range(1)-1 range range(end)+1];

                ind_RAT2use = 1:min([ceil(range(end)) length(RAT2)]);

                detected_temp = zeros(1,length(ind_fd));
                fd_temp = zeros(1,length(ind_fd));
                ind_fdi = cell(1,length(ind_fd));

                for u = 1:length(ind_fd)

                    if strmatch(TYPE{ind_fd(u)},'Neg')
                        th = smooth_amir(RAT2(ind_fd(u),ind_RAT2use),5) /
(1+n(ind_fd(u)));
                    else
                        th = smooth_amir(RAT2(ind_fd(u),ind_RAT2use),5) *
(1+n(ind_fd(u)));
                    end

                    [detected_temp(u),fd_temp(u),ind_fdi{u}] =
det_seiz(RAT2(ind_fd(u),ind_RAT2use),th,range,TYPE{ind_fd(u)});

                end


                for q = 1:length(ind_fdi)-1

                    if q == 1
                        temp = intersect(ind_fdi{1},ind_fdi{2});
                    else
                        temp = intersect(temp,ind_fdi{q+1});
                    end
                end

               if isempty(temp)
                    int_fdi = 0;
               else
                    int_fdi = length(temp);
               end




               if sum(detected_temp) >= round(length(detected_temp)/1.6)
                    detected_final{i}{p}(j,l) = 1;
               else
                     detected_final{i}{p}(j,l) = 0;
               end


                fd_temp_total = sum([int_fdi ; int_fdii']);

%                 [M,I] = min(fd_temp_total(detected_temp == 1));
%
%
%                 if isempty(I)
%                     [M,ind_fd] = min(fd_temp_total);
```

```
%                   else
%                       ind_fd = find(fd_temp_total == M,1,'last');
%                   end



                fd_final{i}{p}(j,l) = fd_temp_total;

            end

        end

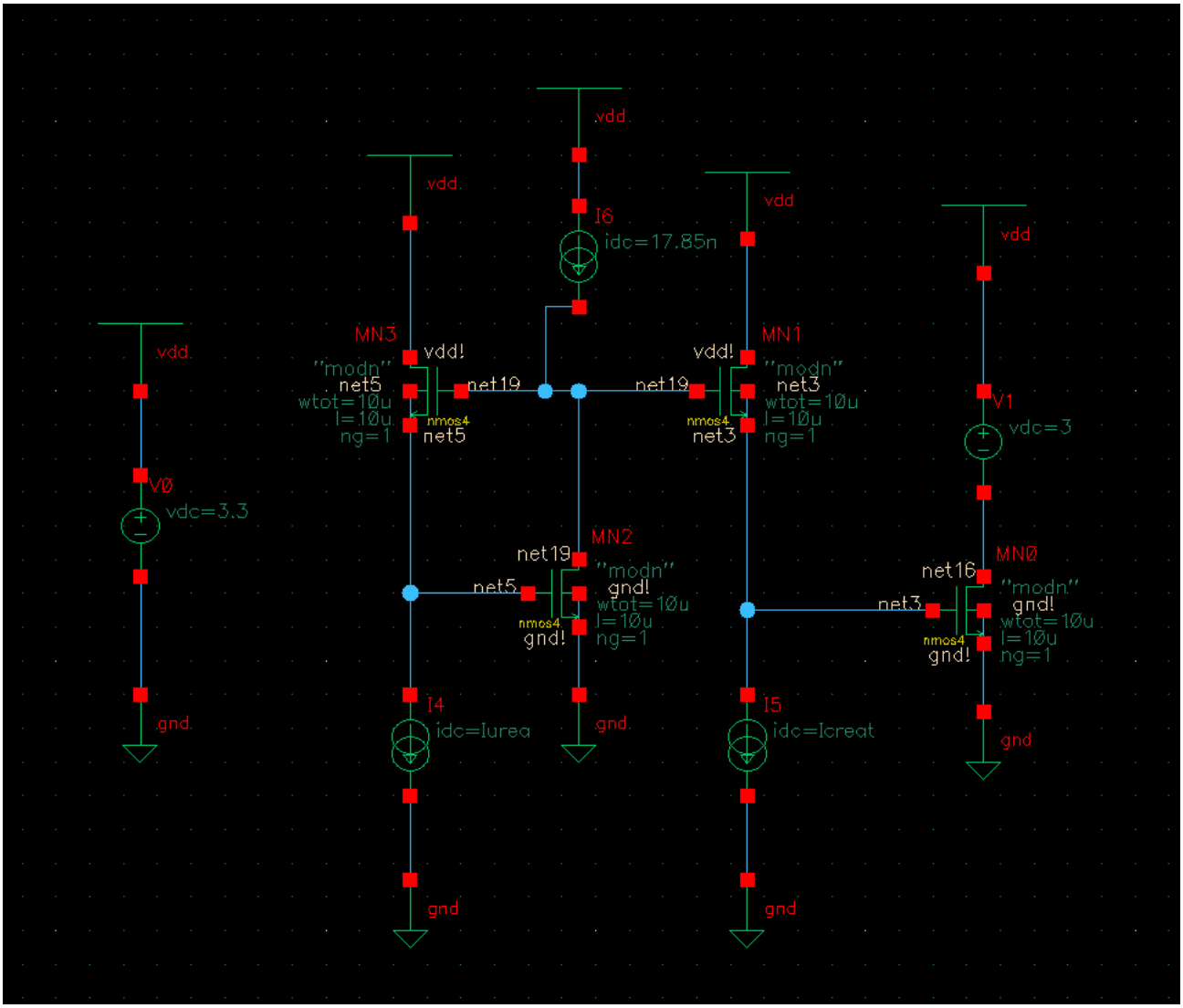

        k = k +1;
    end

end

save results_23rd_July_detection_AND_v2
```

Appendix B - Urea:Creatinine Schematics

Below is a schematic screenshot of our Cadence (version IC5) schematic, described in chapter 2.



Appendix C - Conference and Journal Paper

Biomedical Circuits and Systems Conference (BioCAS) 2010, pp 45-48.

The WiNAM Project: Neural Data Analysis with Applications to Epilespy

Walid Juffali[†], Jamil El-Imad[†], Amir Eftekhar[*†], Christofer Toumazou[*†]
[*]Department of Electrical and Electronic Engineering, Imperial College London, SW7 2BT, UK
[†]Centre for Bio-Inspired Technology, Institute of Biomedical Engineering, Imperial College London, SW7 2AZ, UK
Email: {w.juffali09,j.el-imad,a.eftekhar,c.toumazou}@imperial.ac.uk

***Abstract*—This work presents a novel algorithmic method based on an ngram approach and applies it ECoG and deep brain neural data for analysis of epileptic seizures. This is part of a project (WiNAM) to design an analysis framework suitable for analysing brain dynamical changes. By first exploring the ngram model and its traditional use we describe how to apply it to biological data for pattern recognition. We then use this methodology and apply it to neurological data to show good sensitivity to seizure onset. Through these tests we explore the parameters used when computing the ngram to evaluate what is needed to maintain this sensitivity. Additionally, we present the analysis framework (an online system) that is being developed to carry out the ngram analysis and database the data and results.**

I. Introduction

Over the last several decades neuroscientists have interfaced with the nervous system through technology for a multitude of application spaces. These advanced neural interfaces have allowed us to capture single-neuron recordings and in depth spatial mappings of brain function. One of the key challenges with handling these recordings is extracting the spatio-temporal signatures related to activity and neurological function. To vercome this challenge, signal processing techniques have acted as the bridge between recording and understanding, whether through statistical methods or complex nonlinear state-space mapping.

One area that has seen a multitude of algorithmic strategies applied to it is epilepsy. Epilepsy is a neurological disorder that affects approximately 1% of the world's population [1], has been analysed through linear, nonlinear (state-space) or multivariate analysis techniques. In the brain, it manifests as hypersynchronous neuronal firing that can result in loss of consciousness and in some cases, death. As well as hoping to open a window of understanding into brain function epilepsy studies have aimed to extract features from neurological recordings for prediction and/or detection of seizures. However, although excellent work has emerged, predicting seizures is still an open problem with mixed results as to the optimum approach [2]–[5]. One of the bottlenecks of the research in this area has been data consistency; having data sets well annotated, defined and databased. This is being tackled by many groups including through European Union FP7 grant[1].

Regardless, the field is still open for new methods applied to neurological signals that can give insight into the signal dynamics surrounding epileptic seizures. In this work we use a pattern recognition technique based on an ngram model and apply it to EEG-based signals (electrocortiograms (ECoG) and deep brain recordings). In Section II we will briefly describe the system framework we aim to develop and implement this algorithm into. In Section III we will overview the workings of the ngram approach and the way in which it can be used for biological data. After establishing this, we apply this method to real electrocortiogram (ECoG) data to identify seizure to non-seizure transitions (Section IV). We will continue to show how these changes can be captured in real-time to map pattern changes associated with seizure onset. Finally, we will describbe the system framework we have developed through an online interface for further studies of this method and for neurological data.

[1]http://www.epilepsiae.eu/

II. General System

We have mentioned how one of the bottlenecks of analysis of epileptic seizure activity has been the ability to work with consistent data, well annotated and databased. Part of this work is to establish such a framework for future work and to support signal processing methods. Any biomedical signal processing technique needs to be built into a framework that allows effcient processing and storage (databasing) of results. We envision a system which follows the trends of biomedical healthcare in bringing sensor front-end to processing and back-end visualisation (or closed-loop functionality) into a modular system where different components can be placed in different components (Fig. 1).

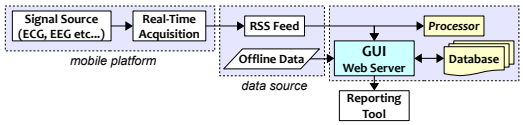


Fig. 1: General system overview envisioned for implementing a signal processing method for biological data.

Here we have data acquisition, user interface and processing blocks. In theory each of these components could be placed in a different technological implementation, such as the acquisition being an implantable neural monitoring device, user interface on mobile phone or PC and processing units being

a web-accessed cloud (such as the Amazon Elastic Compute Cloud). The distribution of these elements will vary depending upon the signal processing requeirements (computational complexity) and application space. Before we discuss the system analysis framework we have developed we will outline the pattern recognition method we employ in this work.

III. Applied Methods

Pattern recognition algorithms comes in many forms but almost always is made up of 3 items: (1) data acquisition, (2) data representation and (3) decision making [6]. One of the more popular methods of pattern recognition is the statistical approach where features of the data are extracted through methods such as Principal Component Analysis (PCA) or Linear Discriminant Analysis and then selected according to some criteria. Once extracted generally classification methods are used, such as template matching or k-nearest neighbour methods. An excellent review of all these methods and the future perspective on this topic see [6] and references therein.

A. Ngram

The ngram model extracts and counts the n subsequences from a particular sequence. Traditionally, it is applied in textual and speech recognition programs. Our method of use of the ngram model comes from the work described in [7] on a *Ngram Statistical Package* in Perl script. The descriptions that follow are sourced from this work. In order to identify patterns from text one first has to break the text into alphanumeric strings delimited usually by a space. Each individual block (between these spaces) is referred to as a *token*. Once these tokens are established then the ngram generates sequences of tokens (patterns) and a count of their occurences.

These patterns can be directly translated to probabilities of words occurring given the previous *n-1* words [8]. For language learning models this is useful as a language can be modelled as the probability of a word occurring givn all the previous ones (which reflects the markovian nature of the signal).

B. Biological Analysis

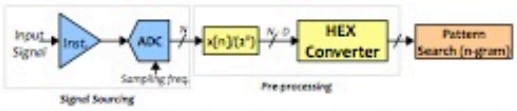


Fig. 2: Figure showing how the input signal is translated to hexadecimal in our system.

Given the ngram method outlined above, here we will describe how we apply this to biological data. Our motivation for using this methodology stems from the fact that biological data is generally acquired through some form of data acquisition. An analog-to-digital converter is used with typical sampling frequencies of 128-512Hz for EEG and ECoG to 10-30KHz for single neuron and local field potential (LFP) signals. The conversion, depending upon the application can result in 8-16bit data. When stored for software (and microcontroller hardware) this information is represented at its lowest level in binary, but in a higher level of abstraction in hexadecimal (HEX). Hence, the data is already available in an alphanumeric format.

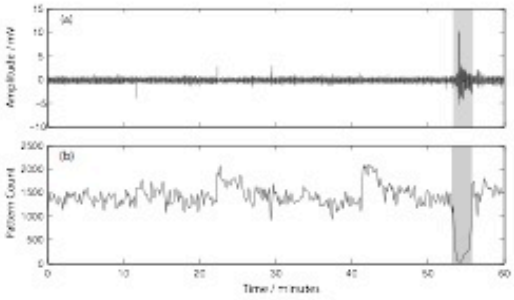


Fig. 3: Figure showing example of pattern count during preictal and ictal period.

Additionally, we can adjust the level of lossiness of the data representation by dividing the data (an N bit number) by an number 2^D where D is an integer, to result in a reduced data format (Fig. 2) - i.e. reducing a 16bit number to an 8bit one by dividing down with $D = 8$. After obtaining the HEX values we use the ngram process previously described to extract any patterns in the signals. Once patterns are extracted we counted the number of significant patterns (those that occured more than 2 times[2]) to determine if any changes occur at seizure onset. We then quantified these pattern counts as a ratio between current window and of analysis and a previous window during a interictal[3] state. All processing was carried out in Matlab version 7.9 with the perl script from [7] accessed through Matlab.

IV. Results

A. Data

Seizure data were obtained from the University Hospital of Freiburg Epilepsy Centre, Germany. The data used were pre-sampled at 256Hz and quantised using a 128 channel 16-bit data acquisition. All 21 patients' first seizure was used from this data set. We use a preictal period of up to 1 hour in most cases with seizure durations varying from 15-170 seconds. There were multiple types of seizures present including: simple partial, complex and general tonic-clonic.

B. *Analysis:* Seizure vs. Non-seizure

Using this data we carried out two sets of tests. The first aimed to quantify whether there were pattern differences

[2]We quantified significant pattern counts in 2 ways: (1) was to count out of the number of significant patterns the total number of occurrences of all these patterns and (2) out of the patterns found what percentage were significant. The former is shown in our results, the latter method quantified similar results so is not shown here.

[3]Ictal refers to the state during a seizure.

TABLE I: Table showing the initial results in aiming to distinguish preictal from ictal periods.

			NGRAM = 12			NGRAM = 14		
Patient	Sdur/s	Pdur/hrs	S	P	Ratio	S	P	Ratio
1	19.49	6.10	514	778.9	0.66	269	393.3	0.684
2	147.35	0.89	6000	28727.5	0.209	2805	21811.3	0.129
3	108.63	1.67	11029	25517.8	0.432	8150	23742.1	0.343
4	77.30	1.03	743	324.6	2.289	577	83.2	6.935
5	14.98	0.99	1455	3246.8	0.448	1138	2991.8	0.38
6	107.94	1.80	2974	13409.7	0.222	1606	9053.5	0.177
7	79.55	1.70	938	16917.8	0.055	719	14890.9	0.048
8	178.87	1.10	27969	34800.6	0.804	19923	27931.6	0.713
9	38.56	1.04	990	6939.9	0.143	702	5465.6	0.128
10	87.07	1.18	14280	19114.6	0.747	10639	17416.9	0.611
11	41.55	1.05	2241	2216.3	1.011	981	861.8	1.14
12	61.33	1.00	2032	12806.2	0.159	1091	11541.4	0.095
13	76.13	1.18	4214	4809.5	0.876	1791	2451.3	0.731
14	70.77	1.34	14268	13165.1	1.084	12212	11027.6	1.11
15	31.51	0.94	5326	5523.1	0.964	4305	4452.6	0.97
16	170.61	1.36	29278	37282.8	0.785	21912	33935.4	0.646
17	58.87	1.81	1256	11120.3	0.113	988	9095.3	0.109
18	17.93	1.20	0	401.6	0	0	154.9	0
19	23.17	1.01	850	2688.7	0.316	469	1679.1	0.279
20	25.09	1.84	1323	4629	0.286	946	4098.1	0.231
21	89.91	0.93	17693	20965.5	0.844	15120	19932.2	0.759

between seizure/ictal areas when compared against inter and preictal periods. To do this, sample preictal periods equal in size to the seizure period were extracted. The average pattern count between each of 10 sections of preictal data is computed in order to compare against the seizure pattern count. For this analysis we used $D = 8$ and analysed the data for ngram sizes of 12 and 14, where one token was one electrical reading in 1byte or 2HEX characters. These results (shown in Table I) show that in most cases the pattern count (P) compared to the seizure count (S) was considerably different; 18 out of 21 cases showed a ratio that indicated a greater than 25% change in pattern count.

C. *Analysis:* Real-time

The second set of analysis used the full preictal period, segmented into 5 and 10 second windows. We analysed the data using $D = 8$ and with ngram sizes 10 and 14. A typical pattern count found in these data sets is shown in Fig. 3. The results for all 21 patients are shown in Table II. We chose to quantify in this way to determine whether characteristics existed that warranted further study.

It is clear patterns do exist and interestingly some patterns change or exist for certain ngram sizes but not for others. An interesting pattern was an increase followed by a slow decrease over several minutes (Fig. 4).

These initial findings conclude that out of the 21 patients, 18 can be detected using the features outlined in Table II which show patterns at different ngram sizes. To investigate further we needed to identify patterns using different ngram sizes while making sure that patterns in an ngram of 12 are not replicated in an ngram size of 10, i.e. make sure patterns are unique and not a subset of a larger pattern in the data. Also, one needs to identify data parameters (e.g. types of seizure) to identify if certain patterns correlate with patient-specific

TABLE II: Table showing heuristic results (related to pattern count changes) when analysing the full data sets. The descriptions refer to the changes that occur that are visible changes compared to the preictal period.

	NGRAM = 10		NGRAM=14
Patient	10s	5s	5s
1	decrease	none	none
2	decrease	decrease	decrease
3	decrease	decrease	decrease
4	none	increase	increase
5	decrease	decrease	decrease
6	decrease	decrease/increase	increase
7	decrease	decrease	increase/decrease
8	decrease	decrease	none
9	decrease	decrease	increase
10	increase/decrease	increase/decrease	increase/decrease
11	none	none	none
12	decrease	decrease	increase/decrease
13	none	none	none
14	none	none	none
15	none	none	none
16	decrease/increase	decrease/increase	decrease/increase
17	decrease	decrease	increase
18	reduced variance	reduced variance	reduced variance
19	decrease/increase	decrease/increase	increase
20	decrease	decrease	decrease
21	none	decrease	decrease

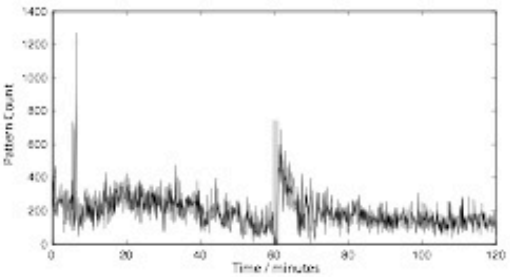


Fig. 4: Figure showing example of pattern count with pattern of increase and slow decrease.

information. To achieve this further study we first needed to design a modular analysis framework as described in Section II. The next section briefly outlines our work to this end.

V. Building the System

We constructed a system as an open online analysis tool (www.winam.net) with SQL database that we can used to easily examine multiple cases and runs on data sets (Fig. 5). As in Fig. 1 the structure is such that the data sourcing can be through an RSS feed, or offline data source. The processing (ngram) are implemented through a separate processing cluster allowing multiple parallel processing efficiently. This is an open system that allows users to freely access and carry out the algorithmic techniques described in this paper. The database structure itself is designed to allow users to input multiple patient cases, and for each case run particular data sets (EEG, ECoG, ECG etc...) or parts of these data sets. We



Fig. 5: Figure showing the graphical interface designed to analyse further data sets- from www.winam.net.

are implementing other importing functionality and reporting documentation to further visualise and analyse the results.

VI. Conclusion

In this paper we have shown a unique methodology for applying ngram models to biological data by taking advantage of the traditional acquisition and storage of the data. Through development of this analysis scheme we have shown initial results for seizure detection capabilities. Although the results are preliminary they show pattern count changes that further justify the study of this methodology. Further study will also be needed to understand how what these patterns reflect on in the data itself and what physiological interpretation can be made from them. To this end we have presented a recent online databasing and analysis tool for use with such data. This tool works through a framework that allows modular design.

There are many avenues to explore with such work, but first is to validate the ngram methodology on multiple data sets and quantify the optimum parameters. We have not quantified any preictal activity in this work but in several cases particular pattern changes were clear which we will investigate further. Overall, this paper has presented the initial steps to this project, named the WiNAM project, for neurological data analysis.

Acknowledgements

This work was supported through a collaborative effort with W Investments (UK) Limited.

References

[1] "http://www.wrongdiagnosis.com/e/epilepsy/stats-country.htm."
[2] F. Mormann, R. G. Andrzejak, C. E. Elger, and K. Lehnertz, "Seizure prediction: the long and winding road," *Brain Advance Access*, 2006.
[3] R. Aschenbrenner-Scheibe, T. Maiwald, M. Winterhalder, H. U. Voss, J. Trimmer, and A. Schulze-Bonhage, "How well can epileptic seizures be predicted? an evaluation of a nonlinear method," *Brain*, vol. 126, pp. 1–11, 2003.
[4] C. J. Stam, "Nonlinear dynamical analysis of eeg and meg: Review of an emerging field," *Clinical Neurophysilogy*, 2005.
[5] L. D. Iasemidis, "Epileptic seizure prediction and control," *IEEE Transactions on Biomedical Engineering*, vol. 50, 2003.
[6] A. Jain, R. Duin, and J. Mao, "Statistical Pattern Recognition: A Review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, 2000.
[7] S. Banerjee and T. Pederson, "The deisgn, implementation and use of the ngram statistics package," *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics*, 2003.
[8] Y. Bengio, R. Ducharme, P. Vincent, and c. Jauvin, "A Neural Probabilistic Language Model," *Journal of Machine Learning Research*, vol. 3, 2003.

IET Electronic Letters vol 46 No. 11

ISFET based Urea:Creatinine translinear sensor

W. Juffali, P. Georgiou and C. Toumazou

Presented is an ion-sensitive field-effect transistor (ISFET) based circuit capable of continuous analogue computation of the urea:creatinine ratio, a marker important for the prediction of pre-renal failure. The designed circuit utilises the weak-inversion of operation to derive sensor currents directly related to the urea and creatinine concentrations, as well as to implement linear division at extremely low power by using a translinear cell.

Introduction: Ion-sensitive field-effect transistor (ISFET) based sensors are becoming increasingly popular for use in integrated biochemical processing platforms used for the early detection and diagnosis of medical conditions. With their implementation in an unmodified CMOS technology [1], this has been made possible owing to the advantages of integration, scalability and low cost due to batch fabrication. In addition, ISFETs operating in the weak-inversion region of operation show great promise for processing biochemical signals owing to the inherent linearisation capability of the device [2]. This is mainly due to exponentially driven drain currents, governed by the diffusion of electrons across the channel of the device in weak-inversion, which linearise biochemical processes governed mostly by Boltzmann diffusion of ions. In this Letter we present a transilinear cell comprising two ISFETs, one for sensing the urea and one for the creatinine, capable of computing the Urea:Creatinine ratio, a marker important for the prediction of pre-renal failure.

Urea-Creatinine ratio: Measurements of the urea and the creatinine are vital markers for assessing kidney function. It is recognised that although a raised urea concentration primarily indicates renal failure, this may not always be the case. Therefore, by combining the measurement with the creatinine and looking at the ratio allows renal failure to be distinguished and also predicted [3]. The normal physiological ranges of these markers should be the urea between 2.5–8.3 mM and the creatinine 44–106 μM. Any deviation from these values is assessed medically according to the ratios of Table 1, indicating the type of renal disease. The urea is usually measured as blood urea nitrogen and the relation is urea = BUN*2.14.

Table 1: BUN:Creatinine ranges and significance

BUN:Cr ratio	Location of renal disorder
>20:1	Pre-renal disease (before glomeruli)
10–20:1	Normal or post-renal disease (within ureter)
<10:1	Intrarenal disease (within kidneys)

ISFETs for urea and creatinine in weak-inversion: When the ISFET is biased with a stable reference electrode (typically Ag/AgCl), ions in solution bind to the passivation causing an accumulation of charge which in turn modulates the threshold voltage of the device [4]. This results in a dependence of the ISFET threshold voltage on pH which can be observed as a shift in the threshold voltage of the device. Using this mechanism, the ISFET, however, can be modified to sense various other analytes and more specifically the urea and the creatinine by immobilisation of urease and creatinase on the surface, respectively [5].

Biasing the ISFET in weak-inversion has the advantage of low power chemical sensing, linearisation of naturally log compressed chemical signals and implementation of biochemical translinear circuits [6]. The generalised drain current of the device biased in weak-inversion is given by (1) whereby V_{GS} is the gate source voltage of the device, I_0 is the intrinsic current, n is the weak-inversion slope coefficient, U_t is the thermal voltage, K_{chem} is a grouping of constant chemical potentials, α is the sensitivity parameter and $[ionX]$ is the concentration of ions in solution. Equations (2) and (3) show the specific equations for the urea and creatinine ISFETs:

$$I_{D(WI)} = I_0 e^{V_{GS}/nU_t} K_{chem}[ionX]^{\frac{a_x}{n}} \tag{1}$$

$$I_{urea} = I_0 e^{V_{GS}/nU_t} K_{urea}[urea]^{\frac{a_{urea}}{n}} \tag{2}$$

$$I_{Cr} = I_0 e^{V_{GS}/nU_t} K_{Cr}[Cr]^{\frac{a_{Cr}}{n}} \tag{3}$$

Translinear cell: Circuits which exploit the exponential relationship between current and voltage in bipolar and weak-inversion MOS transistors are known as translinear circuits. These circuits exploit the linear relation between transconductance and drain current and can be used to perform multiplication, division and power law functions on current signals.

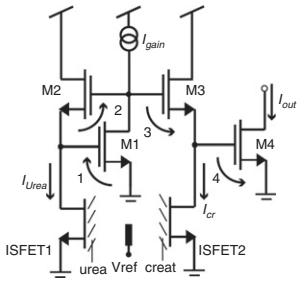


Fig. 1 *ISFET based translinear divider*

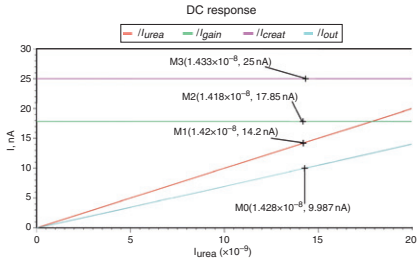


Fig. 2 *Simulation of translinear sensor*

Fig. 1 shows an ISFET based translinear cell capable of computing the division between the drain currents given by the urea and creatinine ISFETs, respectively. By performing a KVL on the loop shown we may write:

$$V_{GS1} + V_{GS2} = V_{GS3} + V_{GS4} \tag{4}$$

and substituting for the weak-inversion drain-currents yields:

$$nU_t \ln\left(\frac{I_{gain}}{I_0}\right) + nU_t \ln\left(\frac{I_{urea}}{I_0}\right) = nU_t \ln\left(\frac{I_{Cr}}{I_0}\right) + nU_t \ln\left(\frac{I_{out}}{I_0}\right) \tag{5}$$

and using the basic relation of addition of natural logs:

$$\ln(a) + \ln(b) = \ln(a\,b) \tag{6}$$

we can compute the output current as follows:

$$I_{out} = I_{gain}\frac{I_{urea}}{I_{Cr}} \tag{7}$$

whereby I_{gain} is a gain term on the ratio. Substituting (2) and (3) into (7) the output of our Urea:Creatinine sensor is now defined as follows:

$$I_{out} = I_{gain}\frac{K_{urea}[urea]^{\frac{a_{urea}}{n}}}{K_{Cr}[Cr]^{\frac{a_{Cr}}{n}}}$$

Simulated results: The circuit was designed in a typical 0.35 μm technology. Device size of W/L = 10 μm/10 μm was used for all transistors. Data from weak-inversion ISFETs for urea and creatinine in [5] were used for simulating the translinear cell. It was found that for 10 mg/dl of BUN, i.e. urea = 21.4 mg/dl, a 14 nA current was given

and likewise for 1 mg/dl of creatinine gave 25 nA of drain current. Considering this should give a ratio of 10:1, by setting I_{gain} to 17.85 nA can give an output indicative of this range as in Table 2. Fig. 2 shows the simulated results of the cell for $I_{Cr} = 25$ nA and a urea current ranging from 0–20 nA. What we can see is that for $I_{urea} = 14$ nA, $I_{out} = 10$ nA, meeting the specifications of Table 2 (considering urea = BUN*2.14). In addition, the cell shows good linearity during the division.

Table 2: BUN:Creatinine output of translinear cell

BUN:Cr ratio	I_{out}
>20:1	>20 nA
10–20:1	10 nA–20 nA
<10:1	<10 nA

Conclusion: We present a novel circuit capable of continuous low-power computation of the BUN:Creatinine ratio for the diagnosis of kidney function. The proposed circuit is capable of mapping the ratio directly as a division of the currents from urea and creatinine based ISFETs. This may be used in a system continually monitoring these markers for the early detection of renal degeneration.

© The Institution of Engineering and Technology 2010
17 February 2010
doi: 10.1049/el.2010.0460
One or more of the Figures in this Letter are available in colour online.

W. Juffali, P. Georgiou and C. Toumazou (*Institute of Biomedical Engineering, Imperial College London, South Kensington, London SW7 2AZ, United Kingdom*)

E-mail: pantelis@imperial.ac.uk

References

1 Bausells, J., Carrabina, J., Errachid, A., and Merlos, A.: 'Ion-sensitive field-effect transistors fabricated in a commercial CMOS technology', *Sens. Actuators B, Chem.*, 1999, **57**, (1–3), pp. 56–62
2 Shepherd, L., and Toumazou, C.: 'Weak inversion ISFETs for ultra-low power biochemical sensing and real-time analysis', *Sens. Actuators B, Chem.*, 2005, **107**, (1), pp. 468–473
3 Morgan, D., Carver, M., and Payne, R.: 'Plasma creatinine and urea: creatinine ratio in patients with raised plasma urea', *Br. Med. J.*, 1977, **2**, (6092), p. 929
4 Bergveld, P.: 'Thirty years of isfetology – what happened in the past 30 years and what may happen in the next 30 years', *Sens. Actuators B, Chem.*, 2003, **88**, (1), pp. 1–20
5 Premanode, B., and Toumazou, C.: 'A novel, low power biosensor for real time monitoring of creatinine and urea in peritoneal dialysis', *Sens. Actuators B, Chem.*, 2007, **120**, (2), pp. 732–735
6 Shepherd, L.M., and Toumazou, C.: 'A biochemical translinear principle with weak inversion ISFETs', *IEEE Trans. Circuits Syst. I, Reg. Pprs*, (see also IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.) 2005, **52**, (12), pp. 2614–2619