Design of a Computational Biomarker for Epileptogenesis: A Machine Learning Approach

by

Ankit Saxena

B.E., Netaji Subhas Institute of Technology, University of Delhi, 2010

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Master of Science

Department of Electrical, Computer & Energy Engineering

2016

This thesis entitled:
Design of a Computational Biomarker for Epileptogenesis: A Machine Learning Approach
written by Ankit Saxena
has been approved for the Department of Electrical, Computer & Energy Engineering

_____

Prof. François G. Meyer

_____

Prof. Michael C. Mozer

_____

Prof. Juan G. Restrepo

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Abstract

Saxena, Ankit (M.S., Electrical Engineering)

Design of a Computational Biomarker for Epileptogenesis: A Machine Learning Approach

Thesis directed by Prof. François G. Meyer

We describe here the recent results of a multidisciplinary effort to design a biomarker that can actively and continuously decode the progressive changes in neuronal organization leading to epilepsy, a process known as epileptogenesis. Using an animal model of acquired epilepsy, we chronically record hippocampal evoked potentials elicited by an auditory stimulus. Using a set of reduced coordinates, our algorithm can identify universal smooth low-dimensional configurations of the auditory evoked potentials that correspond to distinct stages of epileptogenesis. We use a continuous distribution hidden Markov model to learn the dynamics of the evoked potential, as it evolves along these smooth low-dimensional subsets. We provide experimental evidence that the biomarker is able to exploit subtle changes in the evoked potential to reliably decode the stage of epileptogenesis and predict whether an animal will eventually recover from the injury, or develop spontaneous seizures.

Dedication

This thesis is dedicated to the memory of Savitri Saxena.

Acknowledgements

First of all, I would like to express my sincere gratitude to Prof. François G. Meyer, my advisor, for his support, encouragement, and guidance. I would also like to thank Prof. Michael C. Mozer and Prof. Juan G. Restrepo for reviewing this thesis. Finally, I would sincerely like to thank my family for their encouragement and support.

Contents

Chapter

Tables

Table

Figures

Figure

Chapter 1

Introduction

Epilepsy is a neurological disorder that is characterized by the occurrence of several unprovoked seizures. Distinct etiologies (gene mutations, injuries, etc.) have been proposed for the different types of epilepsies. Despite various causes of epilepsy and varying degrees of disease severity in the human population, hippocampal sclerosis is the most consistent neuropathological feature of temporal lobe epilepsy [35].

At present, Epilepsy is the fourth most common[21] neurological disorder. When the incidence of epilepsy[1] is looked over a life time, one in 26[20] people will develop epilepsy sometime in their life. In the United States, the current estimate of the prevelance of epilepsy[2] is 2.2 million people or 7.1 for every 1000[21] people. Also, epilepsy accounts for eight to seventeen[44] percent of the total number of deaths. There is yet no cure for epilepsy and the medication only helps to manage the seizures.

Animal models have been developed to study the neuronal changes underlying the clinical manifestations of epilepsy (chronic-spontaneous seizures). One popular model relies on controlled administration of a convulsant drug (e.g., pilocarpine) to induce status-epilepticus, a life-threatening condition in humans characterized by loss of consciousness and generalized convulsive tonic-clonic seizures. This condition is followed by a latent seizure-free period of weeks to months, where progressive neuronal damage and network reorganization eventually leads to the development of spontaneous seizures.

[1] Looks at the number of new cases of epilepsy in a given year. Usually represented as a ratio of 'x' out of 1000.
[2] Looks at the total number of people suffering from epilepsy. Usually given in millions of people or a ratio

Most of our understanding of the progression of epilepsy, or **epileptogenesis**, is derived from such animal models. It is therefore critical to define a **biomarker** to monitor epileptogenesis and understand the mechanisms that lead to epilepsy. An accurate biomarker would be invaluable for the design of novel anti-epileptogenic drugs and could eventually be translated into a diagnostic tool for humans. Unfortunately, there are very few measurable biological variables that can be used to consistently monitor epileptogenesis and predict disease onset; none of these variables can actively probe the hippocampal circuit in living animals during epileptogenesis.

The presented work addresses this problem and proposes for the first time a "**computational biomarker**" that relies on actively probing the excitability of the hippocampus using an auditory stimulus. The biomarker is validated experimentally using a status epilepticus rat animal model of acquired temporal lobe epilepsy. We borrow the concepts from continuous distribution hidden Markov model, which have found extensive application in Automatic Speech Recognition (ASR)[36] methods, and apply them to the acquired temporal lobe epilepsy model. We aim to study the progression of the disorder without having any access to the physiology of the animal other than recording the response to the auditory stimulus. In doing so, we describe formal methods that can be used to track epileptogenesis.

1.1 Organisation

The remainder of the thesis is organised as follows. The next chapter introduces the methods used for measuring the excitability of the hippocampus using electrophysiological electrodes. The third chapter gives an overview of the approach being proposed. The fourth chapter describes the implementation in detail. The fifth chapter provides a detailed description of the experiments and their results. The last chapter is the conclusion.

Chapter 2

Active Evoked Responses: A Measurement of Epileptogenesis

2.1 Introduction

The chapter introduces two different approaches that are undertaken towards the development of predictive biomarkers: passive recording and active recording. We advocate the latter because the active approach is more reliable and can yield an estimate of the progression of the disorder. In comparison, the passive recording is a 'wait and watch' approach, where we monitor the electroencephalography recordings and look for the high frequency oscillations (HFO) or the interictal spikes (IIS)[41]. We argue that the passive approach does not provide sufficient information to observe early changes in the neuronal excitability. In the presented work, we are interested in actively monitoring the evoked response from the hippocampus to study progressive changes in the neuronal organization. This process of modification in neuronal organization is known as epileptogenesis and leads to the development temporal lobe epilepsy. The monitoring is performed by taking electrical measurements from the electrodes placed on the hippocampus. In this chapter, we discuss the concept of epiletogenesis by presenting the evidence obtained from the actively monitored evoked responses from the hippocampus in response to an auditory stimulus.

2.2 From Passive Recording to Active Sensing

Current efforts toward the development of a reliable and predictive biomarker of epileptogenesis (e.g., see [34] and references therein) fall in three main classes: molecular and cellular biomarkers [26], imaging biomarkers [40], and electrophysiological biomarkers [41]. The presented

work focuses on the electrophysiological class of biomarkers that rely on recordings of the electrical activity associated with neuronal firing. The development of electrophysiological biomarkers has focused on the analysis of both epileptiform spikes[22], and high frequency oscillations [14]. While recordings of spontaneous neuronal spiking can be indicative of neuronal excitability, and therefore correlate with the propensity for seizures, one could argue that the **passive** electrophysiological recordings may not provide enough information to observe early changes in neuronal excitability associated with epileptogenesis.

By contrast, we advocate an **active** approach whereby we probe the excitability of the hippocampus, a brain region known to be epileptogenic [18]. Because the hippocampus also receives several sensory inputs (through the entorhinal cortex), we propose to record in the hippocampus the evoked potential elicited by an auditory stimulus. Since epilepsy does not modify the primary auditory cortex, any alterations in the evoked potential should be indicative of neuronal changes in the hippocampus underlying epileptogenesis. To quantify the property of this biomarker, we use a pilocarpine animal model of temporal lobe epilepsy, and chronically record hippocampal auditory evoked potentials during epileptogenesis. We design a decoding algorithm to demonstrate that changes in the morphology of the hippocampal auditory evoked potential have universal predictive value and can be used to accurately quantify the progression of epilepsy.

We can see in Figure 2.1, that there are subtle changes in the evoked responses between the different states. In baseline (blue), we show the natural response to the auditory stimulus. However in silent (cyan), the response is almost flat. This is attributed to the fact that the rat is still recovering from the injection of pilocarpine. When the rat wakes up, the latent (green) period, we can see that there is a subtle change in the shape of the response. In the chronic (red) state, when the rat has developed spontaneous seizures , we see the response has slightly different characteristics as compared to the responses in the latent period.

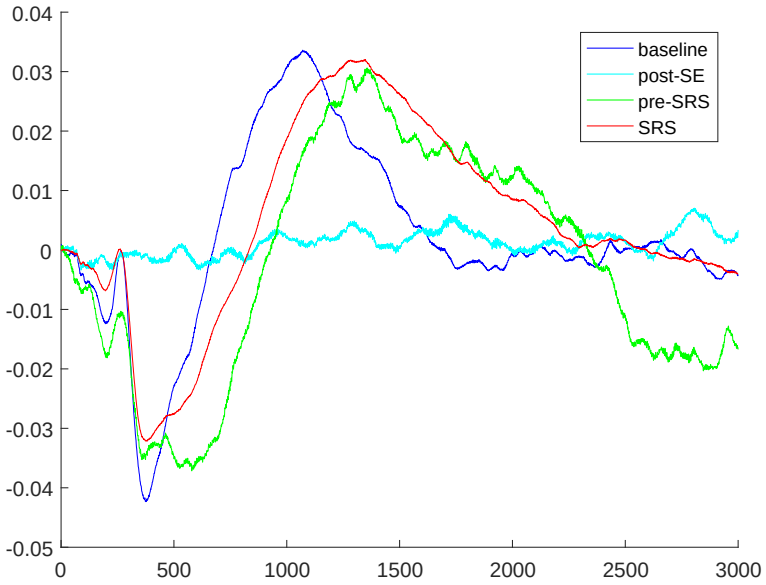

Figure 2.1: Active Recording: Averaged evoked potentials of all states for rat H41. Baseline (blue), Silent(cyan), Latent(green) and Chronic(red).

Thus, we can see from Figure 2.1 that there is evidence of the progression of epilepsy by just visually inspecting the shape of the evoked responses. In the next chapter we discuss our approach on extracting information about the shape of the evoked responses using stationary discrete wavelet coefficients. This forms the mathematical basis of our visual interpretation about the shape of the evoked responses in relation to the process of epileptogenesis.

2.3 Related Work

The authors are not aware of any work that uses machine learning methods to construct a biomarker of epileptogenesis. Counter to this situation, the study of the mechanism underlying the generation of seizures, known as **ictogenesis** [37] – a related and yet completely different problem – has been intensively explored using statistical methods. Specifically, time-delay embedding has been used to try to predict seizures using intracranial EEG recordings [28]. Time-delay embedding was also used in conjunction with machine learning classifiers to predict seizures in [27]. More

recently, reinforcement learning has been used in [6] to control seizure using electrical simulation. Another feature of our approach involves the geometric analysis of the low-dimensional sets formed by embedding time-delay coordinates. This question has been extensively explored using linear methods (e.g., [47] and references therein). Recently nonlinear methods, such as those used to learn the geometry of manifolds, have also been proposed [4, 43]. An important note is in order here: unlike earlier works on ictogenesis, we find it problematic to rigorously apply the deterministic Takens' embedding theorem, or some stochastic version thereof [47]. Rather, we use the time-delay wavelet coordinates to characterize the local dynamics of $\boldsymbol{w}^{(k)}$ at time k, the denoised representation that is extracted from the evoked potential (see sub-section 4.2.1 for a detailed discussion).

2.4 Discussion

Though the idea and approach is straightforward, there are some inherent challenges in dealing with these electrical recordings. Firstly, the recordings are highly susceptible to noise as the evoked potentials are only a few hundred microvolts and since the animal is allowed to move freely, further noise is introduced in the recordings. Secondly, rats are primarily nocturnal animals, thus most of the recordings during the night emit weak response to the auditory stimulus. Therefore, we have to filter almost half of the electrical measurements. We now move on to the next chapter where we describe the animal model and the decoding framework using these hippocampal auditory evoked potentials. We use the pilocarpine model of temporal lobe epilepsy[11] to describe our animal model. This model rests on the use pilocarpine to artificially induce status epilepticus (SE), which is the onset of epileptogenesis. Later on, the presence of hippocampal sclerosis leads to reorganization of neuronal networks and ultimately occurrence of sporadic seizures in the rat.

Chapter 3

Overview of the Approach

3.1 Introduction

With the electrical recording described in the previous chapter, we move on to establishing formal methods of processing this information. This chapter details the pilocarpine model of temporal lobe epilepsy[11] to give the reader a perspective of the setup of the laboratory experiment. We describe the decoding framework which briefly outlines our approach to develop a computational biomarker. We then describe our approach to handle the noise in the measurement by looking at very specific regions of the waveform to extract meaningful information.

3.2 The Animal Model: a Pilocarpine Model of Temporal Lobe Epilepsy

Figure 3.1 provides a detailed timeline, along with the nomenclature of the different periods associated with the progress and eventual onset of epilepsy. All procedures were performed in accordance with the University of Colorado Institutional Animal Care and Use Committee guidelines for the humane use of laboratory rats in biological research. The data of the evoked response was collected from the laboratory of Professor Daniel Barth. The data was generated for the purpose of another study being conducted.

Twenty-four male Sprague-Dawley rats (200-250 gm) were implanted with a hippocampal wire electrode, a ground screw, and a reference screw.

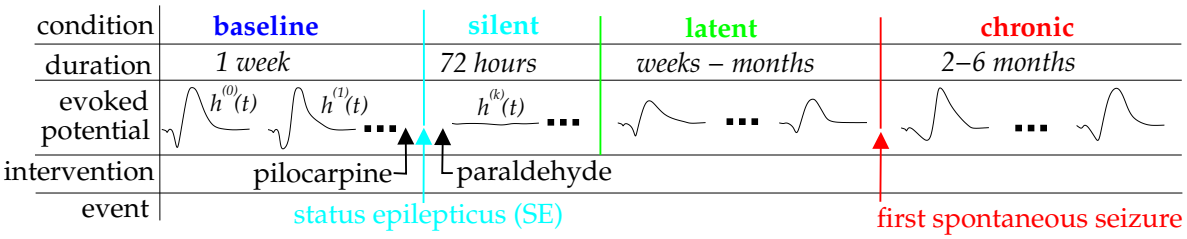

Figure 3.1: Timeline and nomenclature of the different conditions.

Animals were tethered to an electrode harness and slip-ring commutator permitting free movement for 24/7 video and EEG monitoring. The 24 rats were divided into two groups: a group of 17 rats that received lithium-pilocarpine, and a control group of 7 rats. The control group was composed of 2 rats that received all drug injections associated with the lithium-pilocarpine model except for pilocarpine, which was substituted with saline; and 5 rats that received no drugs. After full recovery from the electrode implantation (2 weeks) and at least one additional week of chronic recording of baseline video/EEG, 17 rats were given an injection of lithium chloride followed by an injection of pilocarpine hydrochloride 24 hours later (see Fig. 3.1). After one hour of status epilepticus, the animals were administered a dose of paraldehyde to terminate convulsions. Throughout the experiment at every $\Delta t = 30$ minutes, an auditory stimulus composed of a sequence of 120 square-wave clicks (0.1 ms duration, 2 sec ISI, 45dB SPL) was played in a top-mounted speaker. The 300 ms hippocampal responses to each click were filtered and sampled at 10 kHz, and the average of the 120 responses was computed. In the remainder of the paper, we denote by $h^{(k)}$ the average evoked potential, measured at time $k\Delta t$. To further simplify the exposition, $h^{(k)}$ is simply referred to as the **evoked potential measured at time** k.

We conclude this section with the description of the names that we use to describe the different stages of epileptogenesis (see Fig. 3.1). The period before the injection of pilocarpine is called **baseline**. Conversely, the period following the first spontaneous seizure is called **chronic**. We further define the **silent** period to be the 72 hour period of recovery immediately following the termination of status epilepticus, and the **latent** period to be the remaining period leading to the eventual onset of the first spontaneous seizure.

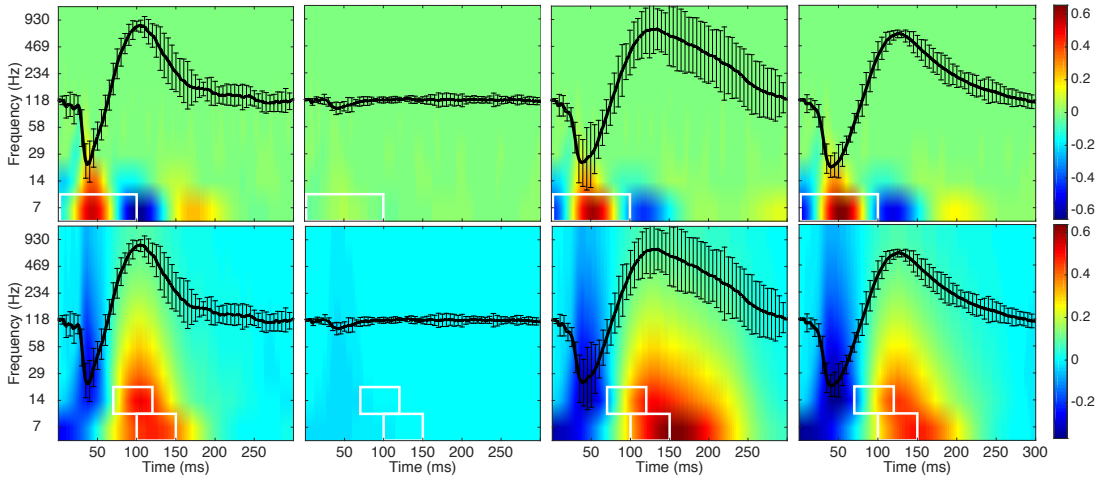

Figure 3.2: Top: average wavelet coefficients with the average evoked potential (across all animals). Bottom: average approximation coefficients. Only the scales $j = 3$ (top) to 10 (bottom) are displayed. White rectangles delineate the time-frequency blocks used to construct $\boldsymbol{w}^{(k)}$.

3.3 Denoising the Input Data

For each animal r, the evoked potentials $h_r^{(k)}, k = 0, 1, \ldots$ were normalized such that the average energy computed during the baseline condition, $\langle h_r^2 \rangle$, for that animal was one. In order to use the noisy evoked potentials to predict the state of epileptogenesis, we extract a denoised representation of $h^{(k)}$. We use a discrete stationary wavelet transform (CDF 9-7) to compute a redundant representation of $h^{(k)}$. Among the 10 scales and 3,000 time samples of the wavelet and approximations coefficients, we only retain the time intervals and the scales that most significantly separated (after controlling for false discovery rate[42]) the evoked potentials taken from the four conditions, across all the animals. Specifically, we form a vector $\boldsymbol{w}^{(k)}$ of 2,000 entries composed of 1,000 wavelet coefficients from the time-frequency region $[0, 100]$ ms $\times$ $[5, 10]$ Hz, and 1,000 approximation coefficients from the time-frequency regions $[70, 120]$ ms $\times$ $[10, 20]$ Hz and $[100, 150]$ ms $\times$ $[5, 10]$ Hz (see Fig. 3.2). This representation is consistent with reports of disruption in the θ rhythm (4-12 Hz) during the latent period preceding the onset of epilepsy [9].

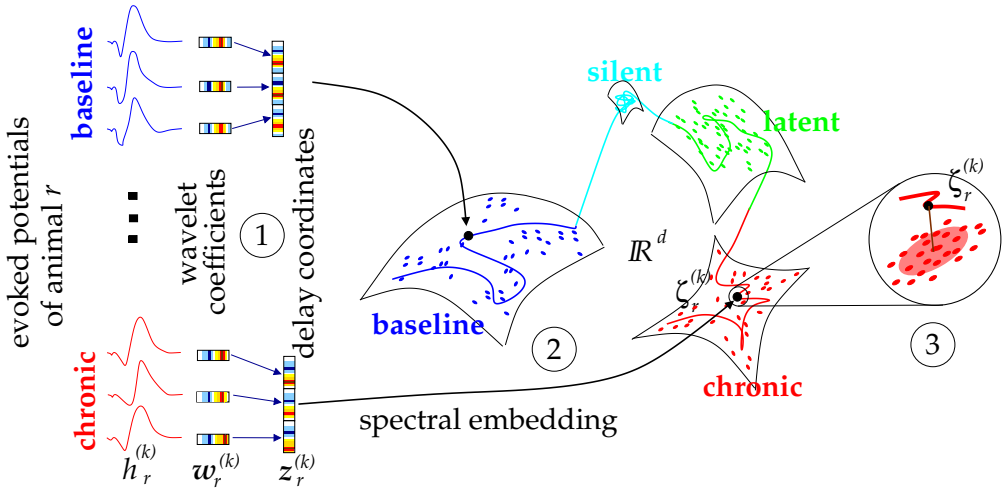

Figure 3.3: Overview of the decoding algorithm. ①: a vector of wavelet coefficients, $\boldsymbol{w}_r^{(k)}$, is computed from the evoked potential $h_r^{(k)}$. A vector of time-delay wavelets coordinates, $\boldsymbol{z}_r^{(k)}$, is formed by concatenating τ consecutive $\boldsymbol{w}_r^{(k)}$. ②: spectral embedding maps $\boldsymbol{w}_r^{(k)}$ to $\boldsymbol{\zeta}_r^{(k)}$. ③: the distance between $\boldsymbol{\zeta}_r^{(k)}$ and the low-dimensional structure formed by each condition is computed.

3.4 The Decoding Framework: Notations and Overview of the Approach

We present here a brief overview of the decoding approach. Given an animal $r = 1, \ldots, 24$, we consider the sequence of evoked potentials $h_r^{(0)}, h_r^{(1)}, \ldots$. The first stage involves the construction of a denoised representation of $h_r^{(k)}$ (see ① in Fig. 3.3). We decompose $h_r^{(k)}(t)$ using a discrete stationary wavelet transform and retain a vector of s (carefully chosen; shown in 3.2 and given in 3.3) wavelet and scaling coefficients, $\boldsymbol{w}_r^{(k)}$, (see ① in Fig. 3.3). The second stage involves characterizing the association between the condition of the disease (baseline, silent, latent, or chronic) and the vector of wavelet coefficients $\boldsymbol{w}_r^{(k)}$. We tackle this question by lifting $\boldsymbol{w}_r^{(k)}$ into $\mathbb{R}^{\tau \times s}$ using time-delay embedding: we concatenate the consecutive vectors $\boldsymbol{w}_r^{(k)}, \ldots, \boldsymbol{w}_r^{(k+\tau-1)}$ to form a $\tau \times s$ vector, $\boldsymbol{z}_r^{(k)}$, of **time-delay wavelet coordinates** (see ① in Fig. 3.3). Low-dimensional structures, which uniquely characterize the stage of epileptogenesis, emerge in the high-dimensional space. We use spectral embedding to parametrize these low-dimensional structures, and map $\boldsymbol{z}_r^{(k)}$ to $\boldsymbol{\zeta}_r^{(k)}$ (see ② in Fig. 3.3). The first decoding stage involves geometrically computing the likelihood that a given

vector $\boldsymbol{\zeta}_r^{(k)}$ corresponds to one of the four conditions. To this end, we quantify the distance of $\boldsymbol{\zeta}_r^{(k)}$ to the low-dimensional cluster formed by each condition (see ③ in Fig. 3.3). In the final decoding stage, we use a hidden Markov model to capture the intrinsic dynamics of epileptogenesis.

To alleviate the notation in the following discussion, and unless we explicitly compare or combine several animals, we dispense with the subscript r, denoting the dependency on rat r, when we discuss the analysis of the evoked potentials for a fixed animal.

3.5 Discussion

Figure 3.2 delineates the time frequency windows (white) for which the discrete stationary wavelet transform was computed. The wavelet transform coefficients are the denoised representation, $\boldsymbol{w}^{(k)}$ of the evoked potentials $h^{(k)}$. We can see in figure 3.2 that the region of $h^{(k)}$ that we are interested in is the trough that occurs in the time-frequency window of $[0, 100]$ ms $\times$ $[5, 10]$ Hz and the peak that occur in the time-frequency window of $[70, 120]$ ms $\times$ $[10, 20]$ Hz and $[100, 150]$ ms $\times$ $[5, 10]$ Hz. The idea is to extract the relevant information from the shape of these peaks and troughs that occur in the evoked responses. From the figure we can see that there are subtle differences in the shape of $h^{(k)}$ as the animal goes from baseline to chronic, with silent being the exception.

We provide a brief summary in Figure 3.3 of approach adopted to develop a biomarker for epileptogenesis. The framework establishes a step by step approach from computing the wavelet coefficients from the evoked responses to spectral embedding the time delayed coefficients. Spectral embedding effectively maps the evoked responses, $h_r^{(k)} \in \mathbb{R}^{3000}$, to the low-dimensional space $\boldsymbol{\zeta}^{(k)} \in \mathbb{R}^d$ where $d \ll 3000$. This is a huge deal as now we implement machine learning algorithms that take advantage of the low-dimensional space for classification of the different states of epiletogenesis. Since the notion of time is not lost when mapping from $h^{(k)} \mapsto \boldsymbol{\zeta}^{(k)}$, we can use a Markovian chain to define the intrinsic dynamics of epileptogenesis by using a hidden Markov model.

Chapter 4

Universal Configuration of Epileptogenesis

4.1 Introduction

This chapter walks the reader through the process of developing a computational biomarker. We first talk about time delay embedding which transforms the extracted denoised representation, $\boldsymbol{w}^{(k)} \in \mathbb{R}^{2000}$, to a higher dimensional space, $\boldsymbol{z}^{(k)} \in \mathbb{R}^{12000}$. In this high dimensional space we stipulate that the trajectory of the embedded vector actually lies in a low dimensional subspace. To extract this low dimensional space we use a graph Laplacian and come to the conclusion that the dimensions of the data set is actually five, i.e. $\boldsymbol{\zeta}^{(k)} \in \mathbb{R}^5$. We move on to address the underlying process of epileptogenesis which, as we have assumed, moves through four different stages and spends an indefinite amount of time in each state. This give a sense of time and evolution of dynamics which can be modelled by the Continuous Distribution Hidden Markov Model. The inherent properties of this model makes it a suitable choice to implement a computational biomarker. Lastly, we address a practical issue that arises due to the not having sufficient data populations of each state. We overcome this issue by developing a pseudo model using a Gaussian mixture model and then re-sampling it to obtain equal representation from each state. The intuition behind this approach arises from our assumption that epileptogenesis has a universal configurations which can be learned using statistical models.

4.2 Time Delay Embedding

We now describe the construction of universal (stable across all animals) low-dimensional smooth sets formed by all the hippocampal auditory evoked potentials collected during the same stage of epileptogenesis. These sets are created by lifting the wavelet coefficients $\boldsymbol{w}^{(k)}$ of each $h^{(k)}$ into high-dimension. This lifting effectively creates smooth low-dimensional coherent structures that can then be parametrized with a drastically smaller number of coordinates.

4.2.1 Lifting the Evoked Potential to a High-Dimensional Space

Given the time series $\left\{\boldsymbol{w}^{(k)}, k = 0, 1, \ldots\right\}$ for a given animal, we analyze the dynamics of this time series by considering the time-delay wavelet coordinates formed by concatenating τ consecutive vectors of wavelet coefficients,

$$\boldsymbol{z}^{(k)} = \left[\boldsymbol{w}^{(k)} \quad \boldsymbol{w}^{(k+1)} \quad \cdots \quad \boldsymbol{w}^{(k+\tau-1)}\right]. \tag{4.1}$$

We characterize the changes in the dynamics of the evoked potentials by studying the geometric structures formed by the trajectory of $\boldsymbol{z}^{(k)}$ in $\mathbb{R}^{\tau \times s}$, as k evolves. To help gain some understanding about the geometric structures formed by the $\boldsymbol{z}^{(k)}$, we make the following elementary observation. We note that we can compute a p-point ($p \leq \tau - 1$) forward finite difference of $\boldsymbol{w}^{(k)}$ from the time-delay wavelet coordinates, $\Delta^p \boldsymbol{w}^{(k)} = \sum_{j=0}^{p} a_j \boldsymbol{w}^{(k+j)}/(\Delta t)^p$. If this finite difference is small, then $\left|\sum_{j=0}^{p} a_j \boldsymbol{w}^{(k+j)}\right| \approx 0$. This last statement can be translated as follows: the distance of $\boldsymbol{z}^{(k)}$ to the hyperplane defined by the vector $[a_0 \ a_1 \ \cdots \ a_p \ 0 \ 0]^T$ is small. We conclude that if the vector of wavelet coefficients $\boldsymbol{w}^{(k)}$ varies slowly, then the trajectory of $\boldsymbol{z}^{(k)}$ lies near the intersection of several such hyperplanes. Furthermore, if the temporal derivatives of $\boldsymbol{w}^{(k)}$ are small for many orders $p = 1, 2, \ldots$, then the trajectory of $\boldsymbol{z}^{(k)}$ will be confined to a set of small dimension. We notice this phenomenon in our data; shortly after status epilepticus, the evoked potentials $h^{(k)}$ are very flat and exhibit very little variability across time k and across animals. Consequently, the evoked potentials measured during the silent period are clustered around a very low-dimensional region of the configuration space (see Fig. 4.1). In practice, the number of time-delay vectors, τ,

is determined using cross-validation. We found that $\tau = 12$, which corresponds to six hours, yields the optimal results.

To learn the geometry of the set formed by the different trajectories $\boldsymbol{z}_r^{(k)}, k = 0, 1 \ldots$, we consider the union – over all the animals – of the vectors $\boldsymbol{z}_r^{(k)}$, and define the set,

$$\mathbf{Z} = \bigcup_{r \in \text{ all animals}} \left\{ \boldsymbol{z}_r^{(k)}, k = 0, 1, \ldots \right\}. \tag{4.2}$$

4.3 Nonlinear Low-Dimensional Embedding of the Time-Delay Wavelet Coordinates

In order to identify the separate regions of $\mathbb{R}^{\tau \times s}$ that correspond to the four conditions (baseline, silent, latent, and chronic), we seek a smooth low-dimensional parametrization of $\mathbf{Z}$. Our results (not shown) indicate that the traditional linear approach (singular spectral analysis [47]) provides a very poor parametrization of the set $\mathbf{Z}$. A nonlinear approach, spectral embedding [17], yields a low-dimensional parametrization of $\mathbf{Z}$ that naturally clusters the different conditions into coherent disconnected smooth subsets. Briefly, we define a similarity matrix $\mathcal{K}$ that quantifies how any two evoked potentials $h^{(k)}$ and $h^{(l)}$ extracted from the same or from different conditions, and from the same or from a different animal, at the respective times k and l co-vary,

$$\mathcal{K}(k, l) = \exp\left(-\|\boldsymbol{z}^{(k)} - \boldsymbol{z}^{(l)}\|^2 / \sigma^2 \right). \tag{4.3}$$

The scaling constant σ was chosen to be a multiple of the median distance $\|\boldsymbol{z}^{(k)} - \boldsymbol{z}^{(l)}\|$, $k, l = 0, 1, \ldots$. We used the d eigenvectors of $\mathcal{K}$ that optimally separated the dataset $\mathbf{Z}$ into four clusters [25] related to the four conditions. The i^{th} eigenvector of $\mathcal{K}$ provided the i^{th} **reduced coordinate**, $\boldsymbol{\zeta}^{(k)}(i)$, of $\boldsymbol{z}^{(k)}$,

$$\boldsymbol{\zeta}^{(k)} = \begin{bmatrix} \boldsymbol{\zeta}^{(k)}(1) & \cdots & \boldsymbol{\zeta}^{(k)}(d) \end{bmatrix}^T. \tag{4.4}$$

In all our experiments, we found that $d = 5$ provided the optimal decoding performance.

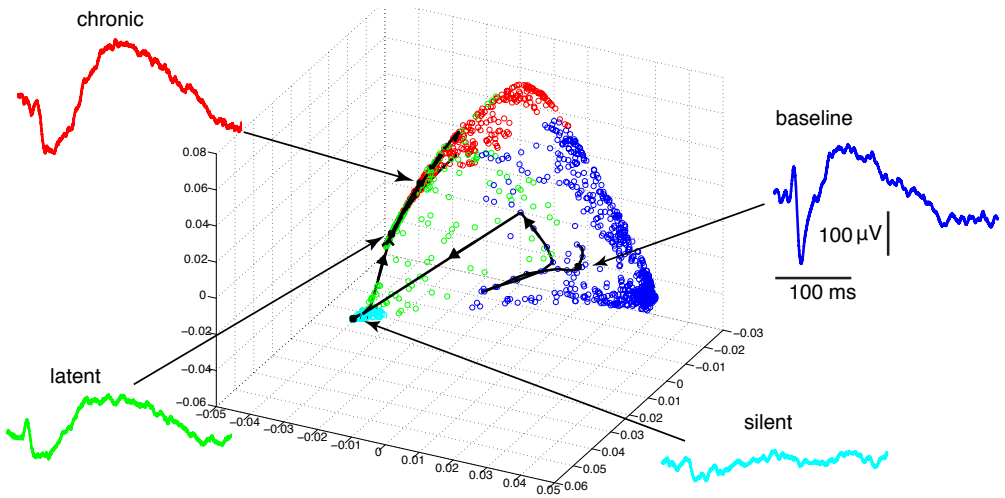

Figure 4.1: The training set of evoked potentials, $\mathbf{Z}$, displayed using the reduced coordinates $\boldsymbol{\zeta}^{(k)}$. Each condition, baseline (blue), silent (cyan), latent (green), and chronic (red), forms a coherent sub-cloud.

Figure 4.1 displays the nonlinear parametrization of $\mathbf{Z}$. Each dot represents an evoked potential $h_r^{(k)}$ measured at a time k from an animal r, and parametrized by the reduced coordinates $\boldsymbol{\zeta}_r^{(k)}$. The color indicates the condition during which $h_r^{(k)}$ was recorded. When collapsed across animals and time of recordings, four distinct clusters can be visually discerned, which can be interpreted in terms of the corresponding conditions. As expected, the evoked potentials in the silent condition are tightly grouped together around a low dimensional structure. Conversely, the latent condition displays the largest spread: some evoked potentials are morphologically close to silent evoked potentials while others are close to chronic evoked potentials. We also note that one can still separate the baseline evoked potentials from the chronic ones, despite their mutual proximity. To further help with the interpretation of this nonlinear representation of $\mathbf{Z}$, we trace the trajectory of $\boldsymbol{\zeta}_r^{(k)}$, $k = 0, 1, \ldots$ for a single rat (H37) over several weeks of recordings (see thick black line in Fig. 4.1). The animal first spends several days in the baseline cluster (blue), then briskly traverses the entire space to reach the silent cluster (cyan) after status epilepticus has been induced. Eventually, as the animal recovers, it joins the latent condition (green) and advances toward the chronic cluster (red). Four representative evoked potentials are shown along this trajectory, confirming changes in

the morphology of $h^{(k)}$ during epileptogenesis.

4.3.1 What Are the Reduced Coordinates Telling Us About Epileptogenesis?

We observe in Fig. 4.1 that the evoked potentials in the training set $\mathbf{Z}$ organize around smooth low dimensional subsets. The natural division of $\mathbf{Z}$ into coherent subsets, which correspond to well defined stages of epileptogenesis, suggests that a purely geometric algorithm could be used to quantify the development of epilepsy. Indeed, given an unclassified evoked potential, $\boldsymbol{\zeta}^{(k)}$, the distance from $\boldsymbol{\zeta}^{(k)}$ to each of the four clusters (baseline, silent, latent, and chronic) should provide an estimate of the likelihood of being in the corresponding condition at time k.

4.4 Estimating the Geometry and Dynamics of Epileptogeneis

Given the progression of epileptogenesis, the dynamics of the process seems to be dependent on the current state and the previous state. Concluding from experimental evidence we assume that these dynamics can be represented by a Markovian description. Therefore, we introduce a Continuous Distribution Hidden Markov model, Figure 4.2. The advantage of a continuous distribution model is that each state of epileptogenesis can be represented as a mixture of Gaussian. Our intuition is to model the geometry of each cluster, $\mathbf{Z}^c$, separately as we assume that the training $\boldsymbol{\zeta}^{(k)\prime}s$ can give us a good sense of the geometric structure of each cluster.

We denote by $\mathbf{Z}^c, c = 0, 1, 2, 3$, the four clusters formed by the evoked potentials in $\mathbf{Z}$ that were respectively collected during the baseline, silent, latent, and chronic conditions. Many techniques are now available to estimate the local geometry of point clouds (see e.g., [19, 10, 33, 7] and references therein). We found that using a mixture of Gaussian with a small number of components gives us a reasonable estimate of the geometric structure described by each $\mathbf{Z}^c$. Furthermore, unlike other models, a Gaussian mixture model does not require the stringent assumption that $\mathbf{Z}^c$ be a smooth sub-manifold. For the purpose of this thesis, a Gaussian mixture model (a mixture of Gaussian densities A) is being used to describe the different stages of epilepsy. Thus each epoch, *baseline*, *silent*, *latent* and *chronic*, are described by their own mixture model.

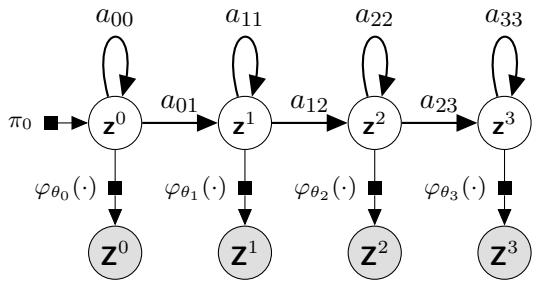

Figure 4.2: Continuous Distribution Hidden Markov Model: $A = a_{ij}$ is the probability state transition matrix, $P(q = j|q = i)$ and $\pi_0 = P(q = \mathbf{z}^0)$. φ_{θ_k} is the emission probability distribution, $p(o_t|q = k)$, given by a mixture of Gaussian (MOG). Each mixture model is defined per state, k, as $\varphi_{\theta_k}(o_t) = \sum_{i=1}^{M} w_i \mathcal{N}(o_t|\mu_{ki}, \Sigma_{ki})$, $\sum_{i=1}^{M} w_i = 1$ and $\mathcal{N}(o_t|\mu_{ki}, \Sigma_{ki})$ is a multivariate Gaussian distribution. $\mathbf{Z}^c, c = 0, 1, 2, 3$ are the clusters for baseline, silent, latent and chronic respectively. $\mathbf{z}^c, c = 0, 1, 2, 3$ are the baseline, silent, latent and chronic states of epileptogenesis respectively.

All the models are initialized with the *k-means++* algorithm[2][3] and the EM algorithm is used to maximize the likelihood. Please refer A for a detailed explanation of EM for a Gaussian mixture model and B for continuous distribution hidden Markov model.

4.4.1 Modified Approach to Continuous Distribution HMM

Even though the Continuous Distribution HMM provides us with a rich perspective to model the reduced coordinate system, $\boldsymbol{\zeta}^{(k)}$, we found that the results were not satisfactory. The problem was narrowed to insufficient evoked potential recordings. We found there was not enough data to estimate the mixture model parameters and that results from the decoded Viterbi path did not represent the actual hidden states. Given our data set of evoked potentials, we find that *chronic* state has the most number of recording as compared to the other three states.

We propose a new approach to implement the continuous distribution hidden Markov model and attempt to fix the issue of under representation from baseline, silent and latent clusters models. Firstly, we learn the geometry of each cluster by modelling it as a mixture of Gaussian densities. We define the number of mixture components, M, for each state, $Q \in [0, 3]$. The number of mixture

Algorithm 1 Oversampled Training of Continuous Density HMM

for $r \leftarrow 1, N \in Training\ Set$ **do** ▷ Concatenate the reduced coordinates of each cluster
  while $e \leftarrow 0, 3$ **do**
    $\mathbf{Z}^e \leftarrow \mathbf{Z}^e \cup \mathbf{Z}^e_r$
  end while
end for
while $e \leftarrow 0, 3$ **do**
  $GMM^e(\mu^e, \Sigma^e, \pi^e) \leftarrow$EM for $\mathbf{Z}^e$ ▷ Learn the parameters of GMM from each cluster
  $\mathbf{Z}^e_{new} \leftarrow$T samples from $GMM^e(\mu^e, \Sigma^e, \pi^e)$ ▷ Form new training set
end while
$\lambda^{t_{final}} \leftarrow \lambda^{t_0}(\mathbf{Z}_{new}, Q, M)$ ▷ Input Parameter to HMM for maximization using EM

components can be varied to get a better fit for the geometry of the respective cluster. Using the EM algorithm we learn the parameters of the mixture model. The learned parameters of the geometry are the means, μ^c, covariances, Σ^c and the posterior probability of each mixture component, π^c where $c = 0, 1, 2, 3$. With these learned geometric parameters, we generate new training samples, $\boldsymbol{\zeta}^{(k)}_{new}$ and group them in clusters, $\mathbf{Z}^c_{new}, c = 0, 1, 2, 3$. These new clusters have the equal number of representative members from each state. We use the new clusters, $\mathbf{Z}^c_{new}, c = 0, 1, 2, 3$, as our new training set and learn the parameters of the continuous distribution hidden Markov model using the Baum-WelchB (EM for HMM's) algorithm. This approach is also given in an algorithmic representation here 1.

4.5 Discussion

In this chapter, we have discussed in greater details our approach which was presented as an overview in the previous chapter. We talk about lifting the evoked potentials to the high dimensional space, $\boldsymbol{z}^{(k)}$, and using non-linear dimension reduction techniques we obtain a low dimensional representation, $\boldsymbol{\zeta}^{(k)}$, which is extremely helpful in visualizing the evoked potentials as clusters and implementing mixture models which define the geometry of each cluster. We also talk about modelling the intrinsic dynamics of epileptogenesis using continuous distribution hidden Markov models and propose a new method of dealing with under represented clusters. In the next chapter we provide experimental evidence of the methods described in this chapter.

Chapter 5

Experiments

5.1 Introduction

In this chapter we present the experimental evidence that was obtained by implementing the approaches described in the previous chapter. We present a sequence of the experiments and the goal of each experiment is to validate every aspect of the proposed approach. Due to challenges presented by the analysis of the real data, we first validate the algorithm using a series of synthetic datasets. These datasets become increasingly realistic and eventually we proceed to the evaluation of the real dataset using a leave one animal out cross validation scheme.

The computational biomarker was trained on ten epileptic rats and evaluated with one animal that was not part of the training data. For each animal in the set, the algorithm computes the stationary wavelet coefficients, time delayed embedded coefficients and the spectral embedding. Then the training animals are used to learn the geometry of various stages and the hidden Markov model is used to model the dynamics of epileptogenesis.

MATLAB® has been used to implement the machine learning algorithms. The Bayes Net Toolbox (BNT) for Matlab[30] has been used to implement the continuous density hidden Markov model.

Algorithm 2 Pure Simulation Model

procedure SIMULATION TEST($\mathbf{Z}$) ▷ $\mathbf{Z}$ is the Training set
  $T = 250, \tau = 25$
  for $r \leftarrow 1, N \in Training\ Set$ **do** ▷ Concatenate the reduced coordinates of each cluster
    while $e \leftarrow 0, 3$ **do**
      $\mathbf{Z}^e \leftarrow \mathbf{Z}^e \cup \mathbf{Z}_r^e$
    end while
  end for
  while $e \leftarrow 0, 3$ **do**
    $GMM^e(\mu^e, \Sigma^e, \pi^e) \leftarrow$EM for $\mathbf{Z}^e$ ▷ Learn the parameters of GMM from each cluster
    $\mathbf{Z}_{new}^e \leftarrow T$ samples from $GMM^e(\mu^e, \Sigma^e, \pi^e)$ ▷ Form new training set
  end while
  $\lambda^{t_{final}} \leftarrow \lambda^{t_0}(\mathbf{Z}_{new}, Q, M)$ ▷ Input Parameter to HMM for maximization using EM
  while $e \leftarrow 0, 3$ **do** ▷ Generating simulated test set
    $\mathcal{T}_{new}^e \leftarrow \tau$ samples from $GMM^e(\mu^e, \Sigma^e, \pi^e)$ ▷ Form new test set
  end while
  $[\delta]_{M\times\tau} = Viterbi(\lambda^{t_{final}}, \mathcal{T}_{new})$ ▷ Decoding most likely sequence of hidden states
end procedure

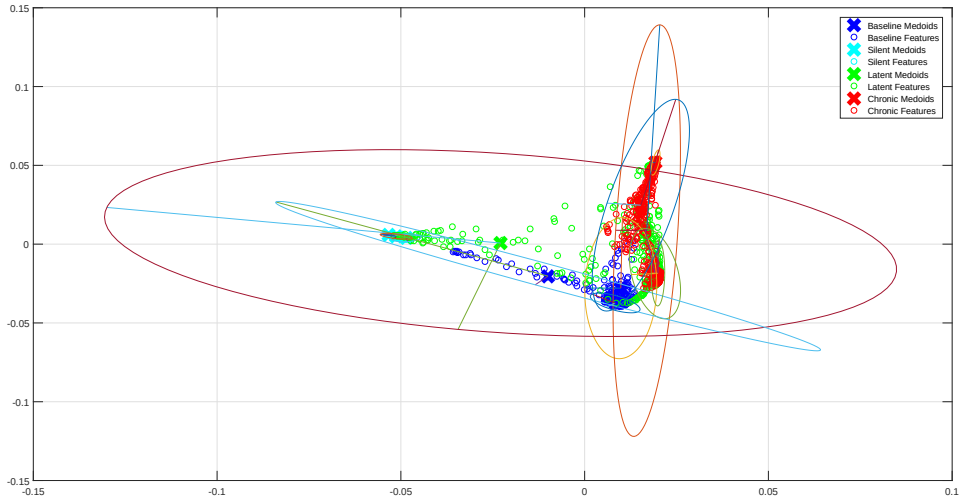

Figure 5.1: Epilepsy model fit using real data. M=3 components per state

5.2 The Simulation Model

We start by defining the model that is going to be used for this experiment. The model, as defined in Figure 4.2, has four states, $Q = 4$, three mixture component per state, $M = 3$. Thus, the total number of mixtures is $Q \times M = 12$. Therefore, the data is modelled by twelve Gaussian

densities each belonging to one of the four states. The dimensionality of the data is two, $d = 2$ and the length of the training set is chosen to be 900 and the testing set in chosen to be 100, thus $T = 1000$ which is the combined number of samples. The number of simulated training rats is 10. Thus, each rat is represented by 100 samples equally divided into baseline, silent, latent and chronic. Using leaving one rat out approach, we divide the dataset into nine simulated training rats and one simulated testing rat.

We now train a mixture model using the training samples from the real data, $\mathbf{Z}^c, c = 0, 1, 2, 3$. The intuition is to learn the parameters from the real data and simulate the reduced coordinates, $\boldsymbol{\zeta}_{sim}^{(k)}$ and try to predict the progression of dynamics using the continuous distribution hidden Markov model. Figure 5.1 gives the mixture model fit of the real data with a mixture of Gaussian. The model is shown in 2D to emphasize the overlapping in the clusters. Table 5.1 lists the parameter obtained after training the mixture model with the real data. We next re-sample the mixture of Gaussian by the parameters defined in 5.1 to create a new set of training samples, 900 samples; 100 samples/rat for nine simulated training rats, $\mathbf{Z}_{new}$. We now create the test set by re-sampling as well, 100 samples for one rat, $\mathcal{T}_{new}$. We define the continuous distribution hidden Markov model with four states, $Q = 4$, and two mixture components per state, $M = 2$. A lesser number of mixture components is chosen to test the robustness of the model. Thus, we have a total of $Q \times M = 8$ mixtures for the model. We train the continuous distribution hidden Markov model using the new training set, $\mathbf{Z}_{new}$ and implement the Viterbi decoding algorithm using the new simulated test set, $\mathcal{T}_{new}$. Figure 5.2 shows the trained continuous distribution hidden Markov model with eight centres. Table 5.2 shows the new learned parameters of the mixture of Gaussian. Figure 5.3 shows the probability of the decoded Viterbi path of the most likely hidden states. The test sequence has 25 samples from each state as can be seen from the figure.

The results demonstrates the robustness of the continuous distribution hidden Markov model. As we can see from 5.3 that there is no error in the predicted states. Even when the training model was provided with fewer components the prediction was still accurate. An algorithmic representation of the procedure followed in this test is given in 1.

	M	π	μ	Σ
Baseline	1	0.3583	$\begin{pmatrix} 0.0098 \\ -0.0279 \end{pmatrix}$	$1 * 10^{-4} \begin{bmatrix} 0.0452 & 0.0088 \\ 0.0088 & 0.9441 \end{bmatrix}$
	2	0.4618	$\begin{pmatrix} 0.0083 \\ -0.0373 \end{pmatrix}$	$1 * 10^{-5} \begin{bmatrix} 0.2155 & -0.1302 \\ -0.1302 & 0.1849 \end{bmatrix}$
	3	0.1798	$\begin{pmatrix} -0.0099 \\ -0.0203 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.2592 & -0.1635 \\ -0.1635 & 0.1059 \end{bmatrix}$
Silent	1	0.1736	$\begin{pmatrix} -0.0485 \\ 0.0040 \end{pmatrix}$	$1 * 10^{-6} \begin{bmatrix} 0.7738 & 0.0957 \\ 0.0957 & 0.3012 \end{bmatrix}$
	2	0.1300	$\begin{pmatrix} -0.0533 \\ 0.0057 \end{pmatrix}$	$1 * 10^{-6} \begin{bmatrix} 0.1082 & -0.0190 \\ -0.0190 & 0.0077 \end{bmatrix}$
	3	0.6964	$\begin{pmatrix} -0.0513 \\ 0.0050 \end{pmatrix}$	$1 * 10^{-6} \begin{bmatrix} 0.8978 & -0.2270 \\ -0.2270 & 0.1612 \end{bmatrix}$
Latent	1	0.4678	$\begin{pmatrix} 0.0169 \\ 0.0086 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.0040 & 0.0218 \\ 0.0218 & 0.8044 \end{bmatrix}$
	2	0.2630	$\begin{pmatrix} -0.0229 \\ 0.0007 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.5476 & -0.0844 \\ -0.0844 & 0.1657 \end{bmatrix}$
	3	0.2692	$\begin{pmatrix} 0.0200 \\ -0.0184 \end{pmatrix}$	$1 * 10^{-4} \begin{bmatrix} 0.0016 & 0.0007 \\ 0.0007 & 0.2030 \end{bmatrix}$
Chronic	1	0.2760	$\begin{pmatrix} 0.0191 \\ -0.0187 \end{pmatrix}$	$1 * 10^{-4} \begin{bmatrix} 0.0229 & -0.0351 \\ -0.0351 & 0.3870 \end{bmatrix}$
	2	0.1973	$\begin{pmatrix} 0.0193 \\ 0.0518 \end{pmatrix}$	$1 * 10^{-5} \begin{bmatrix} 0.0068 & 0.0297 \\ 0.0297 & 0.3016 \end{bmatrix}$
	3	0.5267	$\begin{pmatrix} 0.0153 \\ 0.0247 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.0083 & 0.0296 \\ 0.0296 & 0.2128 \end{bmatrix}$

Table 5.1: Epilepsy model parameters before re-sampling

We find that the continuous distribution hidden Markov model provides a reliable method to estimate hidden states in a mixture of Gaussian with large overlap of the mixture components. We now try and test the limits of this design by reducing the number of simulated test samples for each state.

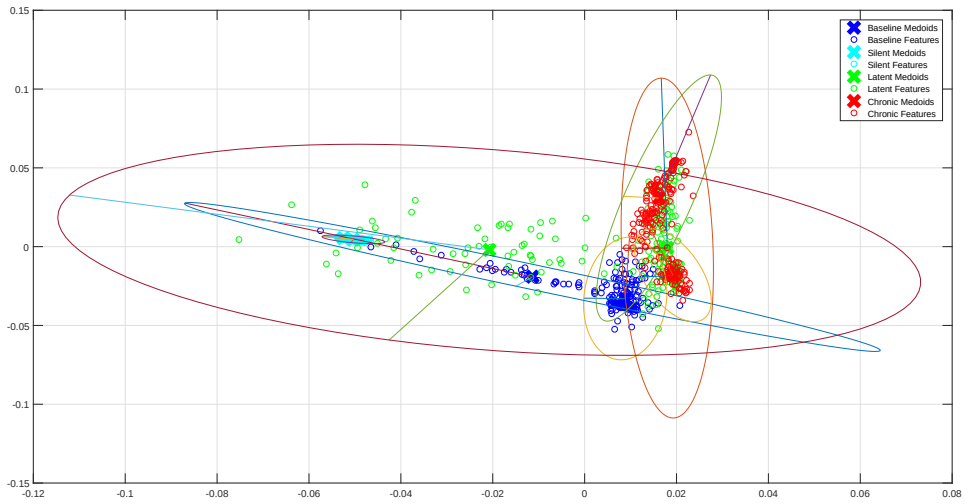

Figure 5.2: Epilepsy model fit after re-sampling. M=2 components per state

	M	π	μ	Σ
Baseline	1	0.8165	$\begin{pmatrix} 0.0088 \\ -0.0329 \end{pmatrix}$	$1 * 10^{-4} \begin{bmatrix} 0.0388 & 0.0162 \\ 0.0162 & 0.7140 \end{bmatrix}$
	2	0.1835	$\begin{pmatrix} -0.0114 \\ -0.0193 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.2705 & -0.1663 \\ -0.1663 & 0.1053 \end{bmatrix}$
Silent	1	0.2520	$\begin{pmatrix} -0.0525 \\ 0.0055 \end{pmatrix}$	$1 * 10^{-6} \begin{bmatrix} 0.9764 & -0.2259 \\ -0.2259 & 0.0601 \end{bmatrix}$
	2	0.7480	$\begin{pmatrix} -0.0503 \\ 0.0046 \end{pmatrix}$	$1 * 10^{-5} \begin{bmatrix} 0.2192 & -0.0538 \\ -0.0538 & 0.0353 \end{bmatrix}$
Latent	1	0.6991	$\begin{pmatrix} 0.0180 \\ -0.0009 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.0048 & -0.0066 \\ -0.0066 & 0.5487 \end{bmatrix}$
	2	0.3009	$\begin{pmatrix} -0.0208 \\ -0.0020 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.4160 & -0.0910 \\ -0.0910 & 0.2116 \end{bmatrix}$
Chronic	1	0.6511	$\begin{pmatrix} 0.0161 \\ 0.0309 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.0088 & 0.0411 \\ 0.0411 & 0.2879 \end{bmatrix}$
	2	0.3489	$\begin{pmatrix} 0.0192 \\ -0.0182 \end{pmatrix}$	$1 * 10^{-4} \begin{bmatrix} 0.0327 & -0.0632 \\ -0.0632 & 0.4093 \end{bmatrix}$

Table 5.2: Epilepsy model parameters after re-sampling

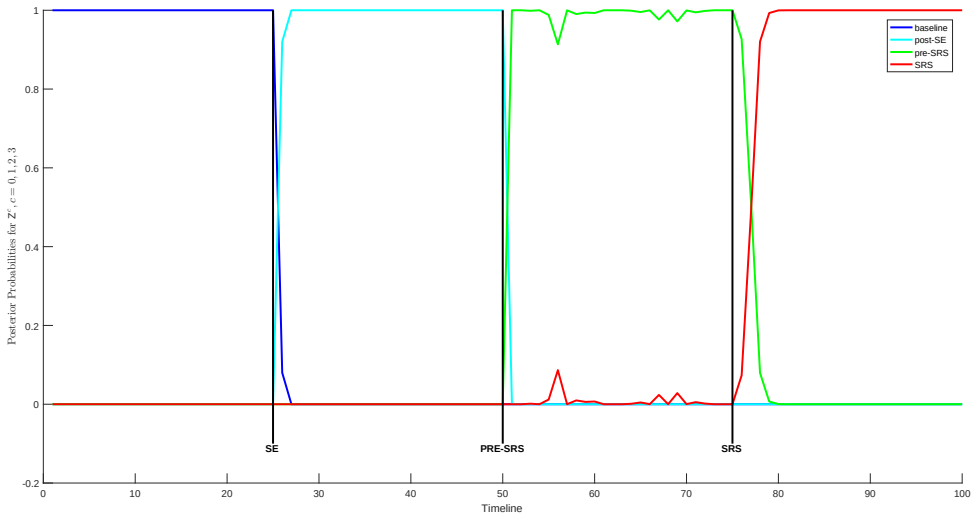

Figure 5.3: Estimated posterior probability of being in one of the states

Algorithm 3 Simulation Model: Oversampling State Space Test

procedure SIMULATION OVERSAMPLED TEST($\mathbf{Z}, \mathcal{T}$) ▷ $\mathbf{Z}, \mathcal{T}$ are the Training and Test sets
  $T = 900$
  for $r \leftarrow 1, N \in Training\ Set$ **do** ▷ Concatenate the reduced coordinates of each cluster
    while $e \leftarrow 0, 3$ **do**
      $\mathbf{Z}^e \leftarrow \mathbf{Z}^e \cup \mathbf{Z}^e_r$
    end while
  end for
  while $e \leftarrow 0, 3$ **do**
    $GMM^e(\mu^e, \Sigma^e, \pi^e) \leftarrow$EM for $\mathbf{Z}^e$ ▷ Learn the parameters of GMM from each cluster
    $\mathbf{Z}^e_{new} \leftarrow T$ samples from $GMM^e(\mu^e, \Sigma^e, \pi^e)$ ▷ Form new training set
  end while
  $\lambda^{t_{final}} \leftarrow \lambda^{t_0}(\mathbf{Z}_{new}, Q, M)$ ▷ Input Parameter to HMM for maximization using EM
  while $e \leftarrow 0, 3$ **do** ▷ Generating simulated test set
    $\tau \leftarrow sizeof(\mathcal{T}^e)$
    $\mathcal{T}^e_{new} \leftarrow \tau$ samples from $GMM^e(\mu^e, \Sigma^e, \pi^e)$ ▷ Form new test set
  end while
  $[\delta]_{M \times \tau} = Viterbi(\lambda^{t_{final}}, \mathcal{T}_{new})$ ▷ Decoding most likely sequence of hidden states
end procedure

5.3 Simulated Model: Oversampling State Space Test

From the previous experiment we saw that the continuous distribution hidden Markov model is extremely reliable for decoding hidden states from mixture of Gaussian. We want to have another

test which is closer to the real dataset. For the purpose of this experiment, the number of simulated test samples generated will be governed by the real dataset.

We start by providing the design of this test. The model, as defined in Figure 4.2, has four states, $Q = 4$, three mixture component per state, thus $M = 3$, so total number of mixture is $Q \times M = 12$. Therefore, the data is modelled by twelve Gaussian each belonging to one of the four states. The dimensionality of the data is two, $d = 3$ and the length of the training set is chosen to be 3600 and the length of the testing set will be determined from the real data. Thus $T = 3600 + \tau$ where τ is the number of simulated test samples to be determined from the testing set. The number of simulated training rats is 10. Thus, each of the simulated training rats are represented by 900 samples equally divided into baseline, silent, latent and chronic. But the simulated test rat samples are determined by real dataset. Thus, $\tau^c, c = 0, 1, 2, 3$ are the number of test samples from each state. Using leaving one rat out approach, we divide the dataset into nine simulated training rats and one simulated testing rat.

We now train a mixture model using the training samples from the real data, $\mathbf{Z}^c, c = 0, 1, 2, 3$. The intuition is to learn the parameters from the real data and simulate the reduced coordinates, $\boldsymbol{\zeta}_{sim}^{(k)}$ and try to predict the progression of dynamics using the continuous distribution hidden Markov model. Figure 5.4 gives the mixture model fit of the real data with a mixture of Gaussian. The model is shown in 3D to emphasize the overlapping in the clusters. Table 5.3 lists the parameter obtained after training the mixture model with the real data. We next re-sample the mixture of Gaussian by the parameters defined in 5.3 to create a new set of training samples, 3600 samples; 900 samples/rat for nine simulated training rats, $\mathbf{Z}_{new}$. We now create the test set by re-sampling as well. For this experiment $\tau^0 = 32, \tau^1 = 13, \tau^2 = 33$ *and* $\tau^3 = 43$, 121 samples for one rat, $\mathcal{T}_{new}$. We define the continuous distribution hidden Markov model with four states, $Q = 4$, and three mixture components per state, $M = 3$. Thus, we have a total of $Q \times M = 12$ mixtures for the model. We train the continuous distribution hidden Markov model using the new training set, $\mathbf{Z}_{new}$ and implement the Viterbi decoding algorithm using the new simulated test set, $\mathcal{T}_{new}$. Figure 5.5 shows the trained continuous distribution hidden Markov model with twelve centres. Table 5.4

shows the new learned parameters of the mixture of Gaussian. Figure 5.6 shows the probability of the decoded Viterbi path of the most likely hidden states. The test sequence has 121 samples in total and $\tau^0 = 32, \tau^1 = 13, \tau^2 = 33$ *and* $\tau^3 = 43$ is the number of samples from each state.

As we can see from 5.6 that there is no error in the predicted states. The results show that the oversampled fit can represent the geometry of the structure accurately. An algorithmic representation of the procedure followed in this test is given in 1.

The result from this experiment are encouraging. We find that the decoding works extremely well this approach. We can suggest that the number of samples is key to approximating the geometry of this model. This results give us confidence to test out the real test vectors based on the model over sampled training data.

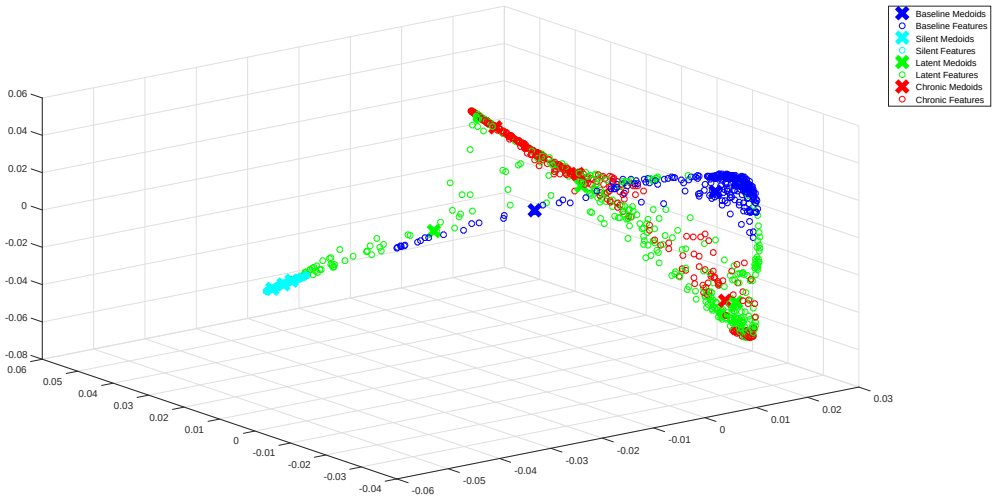

Figure 5.4: Observed epilepsy model using real data. M=3 components per state

	M	π	μ	Σ
Baseline	1	0.1334	$\begin{pmatrix} -0.0165 \\ -0.0160 \\ 0.0247 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.1779 & -0.1097 & 0.1548 \\ -0.1097 & 0.0703 & -0.0974 \\ 0.1548 & -0.0974 & 0.1366 \end{bmatrix}$
	2	0.4401	$\begin{pmatrix} 0.0083 \\ -0.0372 \\ 0.0403 \end{pmatrix}$	$1 * 10^{-4} \begin{bmatrix} 0.0159 & -0.0110 & -0.0380 \\ -0.0110 & 0.0203 & 0.0121 \\ -0.0380 & 0.0121 & 0.1151 \end{bmatrix}$
	3	0.4266	$\begin{pmatrix} 0.0097 \\ -0.0290 \\ 0.0286 \end{pmatrix}$	$1 * 10^{-4} \begin{bmatrix} 0.0533 & 0.0078 & -0.1866 \\ 0.0078 & 0.8670 & -0.4026 \\ -0.1866 & -0.4026 & 0.9033 \end{bmatrix}$
Silent	1	0.3020	$\begin{pmatrix} -0.0512 \\ 0.0048 \\ -0.0093 \end{pmatrix}$	$1 * 10^{-6} \begin{bmatrix} 0.3102 & -0.0937 & 0.2972 \\ -0.0937 & 0.0520 & -0.1066 \\ 0.2972 & -0.1066 & 0.2967 \end{bmatrix}$
	2	0.4135	$\begin{pmatrix} -0.0498 \\ 0.0045 \\ -0.0081 \end{pmatrix}$	$1 * 10^{-5} \begin{bmatrix} 0.1887 & -0.0526 & 0.1755 \\ -0.0526 & 0.0464 & -0.0727 \\ 0.1755 & -0.0727 & 0.1811 \end{bmatrix}$
	3	0.2846	$\begin{pmatrix} -0.0528 \\ 0.0056 \\ -0.0112 \end{pmatrix}$	$1 * 10^{-6} \begin{bmatrix} 0.3449 & -0.0880 & 0.3475 \\ -0.0880 & 0.0587 & -0.1185 \\ 0.3475 & -0.1185 & 0.3751 \end{bmatrix}$
Latent	1	0.3775	$\begin{pmatrix} 0.0188 \\ -0.0216 \\ -0.0401 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.0040 & 0.0104 & -0.0372 \\ 0.0104 & 0.0426 & -0.0901 \\ -0.0372 & -0.0901 & 0.3522 \end{bmatrix}$
	2	0.3733	$\begin{pmatrix} 0.0171 \\ 0.0196 \\ -0.0032 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.0046 & 0.0124 & -0.0125 \\ 0.0124 & 0.4583 & 0.2563 \\ -0.0125 & 0.2563 & 0.2546 \end{bmatrix}$
	3	0.2493	$\begin{pmatrix} -0.0249 \\ 0.0004 \\ 0.0078 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.4948 & -0.0994 & 0.3040 \\ -0.0994 & 0.1797 & -0.1487 \\ 0.3040 & -0.1487 & 0.2405 \end{bmatrix}$
Chronic	1	0.3973	$\begin{pmatrix} 0.0183 \\ 0.0454 \\ 0.0110 \end{pmatrix}$	$1 * 10^{-4} \begin{bmatrix} 0.0184 & 0.0992 & 0.0123 \\ 0.0992 & 0.6194 & 0.1155 \\ 0.0123 & 0.1155 & 0.0383 \end{bmatrix}$
	2	0.3194	$\begin{pmatrix} 0.0140 \\ 0.0161 \\ 0.0067 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.0086 & 0.0172 & -0.0149 \\ 0.0172 & 0.1139 & -0.0162 \\ -0.0149 & -0.0162 & 0.0393 \end{bmatrix}$
	3	0.2833	$\begin{pmatrix} 0.0191 \\ -0.0181 \\ -0.0420 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.0023 & -0.0046 & -0.0259 \\ -0.0046 & 0.0538 & 0.0708 \\ -0.0259 & 0.0708 & 0.3004 \end{bmatrix}$

Table 5.3: Epilepsy model parameters before re-sampling

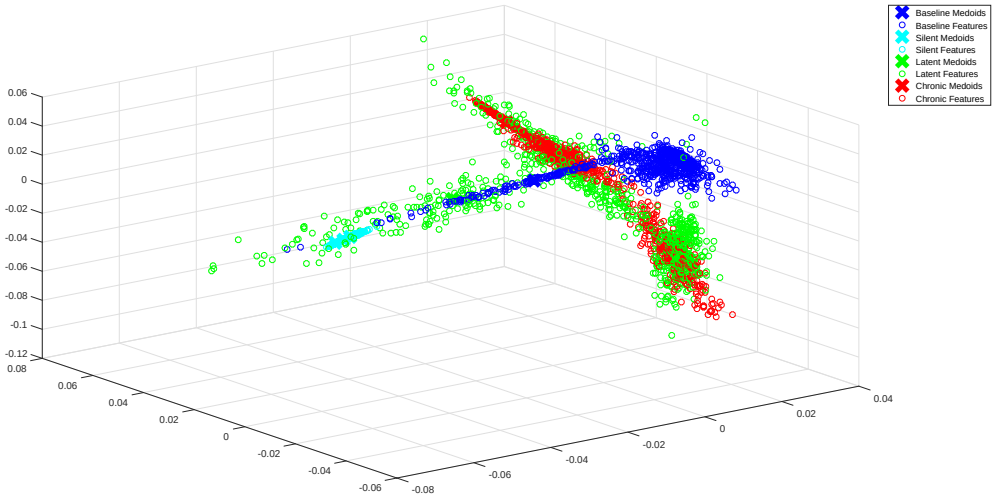

Figure 5.5: Trained epilepsy model fit after re-sampling. M=3 components per state

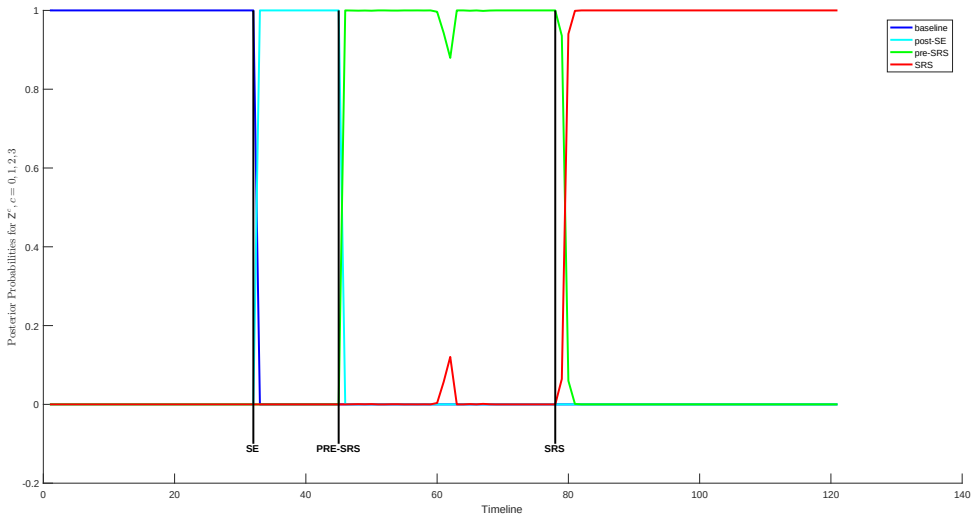

Figure 5.6: Estimated posterior probability of being in one of the states

	M	π	μ	Σ
Baseline	1	0.4405	$\begin{pmatrix} 0.0084 \\ -0.0373 \\ 0.0402 \end{pmatrix}$	$1 * 10^{-4} \begin{bmatrix} 0.0152 & -0.0102 & -0.0357 \\ -0.0102 & 0.0173 & 0.0127 \\ -0.0357 & 0.0127 & 0.1051 \end{bmatrix}$
	2	0.3911	$\begin{pmatrix} 0.0097 \\ -0.0290 \\ 0.0287 \end{pmatrix}$	$1 * 10^{-4} \begin{bmatrix} 0.0538 & 0.0098 & -0.1933 \\ 0.0098 & 0.9559 & -0.4553 \\ -0.1933 & -0.4553 & 0.9763 \end{bmatrix}$
	3	0.1683	$\begin{pmatrix} -0.0156 \\ -0.0164 \\ 0.0255 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.1614 & -0.0995 & 0.1398 \\ -0.0995 & 0.0642 & -0.0883 \\ 0.1398 & -0.0883 & 0.1233 \end{bmatrix}$
Silent	1	0.2963	$\begin{pmatrix} -0.0528 \\ 0.0056 \\ -0.0111 \end{pmatrix}$	$1 * 10^{-6} \begin{bmatrix} 0.3846 & -0.1033 & 0.3910 \\ -0.1033 & 0.0641 & -0.1351 \\ 0.3910 & -0.1351 & 0.4229 \end{bmatrix}$
	2	0.4072	$\begin{pmatrix} -0.0497 \\ 0.0045 \\ -0.0080 \end{pmatrix}$	$1 * 10^{-5} \begin{bmatrix} 0.1904 & -0.0519 & 0.1765 \\ -0.051 & 0.0462 & -0.0720 \\ 0.1765 & -0.0720 & 0.1815 \end{bmatrix}$
	3	0.2964	$\begin{pmatrix} -0.0512 \\ 0.0049 \\ -0.0094 \end{pmatrix}$	$1 * 10^{-6} \begin{bmatrix} 0.3423 & -0.0983 & 0.3245 \\ -0.0983 & 0.0543 & -0.1115 \\ 0.3245 & -0.1115 & 0.3204 \end{bmatrix}$
Latent	1	0.3712	$\begin{pmatrix} 0.0189 \\ -0.0214 \\ -0.0408 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.0038 & 0.0099 & -0.0352 \\ 0.0099 & 0.0426 & -0.0845 \\ -0.0352 & -0.0845 & 0.3341 \end{bmatrix}$
	2	0.2396	$\begin{pmatrix} -0.0249 \\ 0.0020 \\ 0.0067 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.4487 & -0.0629 & 0.2562 \\ -0.0629 & 0.1806 & -0.1318 \\ 0.2562 & -0.1318 & 0.2051 \end{bmatrix}$
	3	0.3892	$\begin{pmatrix} 0.0172 \\ 0.0189 \\ -0.0045 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.0041 & 0.0130 & -0.0098 \\ 0.0130 & 0.4810 & 0.2762 \\ -0.0098 & 0.2762 & 0.2605 \end{bmatrix}$
Chronic	1	0.2842	$\begin{pmatrix} 0.0192 \\ -0.0180 \\ -0.0434 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.0022 & -0.0048 & -0.0246 \\ -0.0048 & 0.0479 & 0.0714 \\ -0.0246 & 0.0714 & 0.2872 \end{bmatrix}$
	2	0.3306	$\begin{pmatrix} 0.0140 \\ 0.0161 \\ 0.0065 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.0075 & 0.0126 & -0.0135 \\ 0.0126 & 0.1085 & -0.0051 \\ -0.0135 & -0.0051 & 0.0415 \end{bmatrix}$
	3	0.3852	$\begin{pmatrix} 0.0184 \\ 0.0462 \\ 0.0111 \end{pmatrix}$	$1 * 10^{-4} \begin{bmatrix} 0.0187 & 0.0996 & 0.0118 \\ 0.0996 & 0.6213 & 0.1160 \\ 0.0118 & 0.1160 & 0.0396 \end{bmatrix}$

Table 5.4: Epilepsy model parameters after re-sampling

5.4 The Computational Biomarker

This experiment is similar in every way to the previous experiment except for the dimensionality of the data and the fact that we are using the real reduced coordinates for the test rat.

We start by providing the design of this test. The model, as defined in Figure 4.2, has four

states, $Q = 4$, three mixture component per state, thus $M = 3$, so total number of mixture is $Q \times M = 12$. Therefore, the data is modelled by twelve Gaussian each belonging to one of the four states. The dimensionality of the data is two, $d = 4$ and the length of the training set is chosen to be 3600 and the length of the testing set will be same as the real data. Thus $T = 3600 + \tau$ where τ is the number of test samples. The number of training rats is 10. Thus, each of the training rats are represented by 900 simulated reduced coordinate samples equally divided into baseline, silent, latent and chronic. But the test rat samples are given by real dataset. Thus, $\tau^c, c = 0, 1, 2, 3$ are the number of test samples from each state. Using leaving one rat out approach, we divide the dataset into nine simulated training rats and one simulated testing rat.

We now train a mixture model using the training samples from the real data, $\mathbf{Z}^c, c = 0, 1, 2, 3$. The intuition is to learn the parameters from the real data and simulate the reduced coordinates, $\boldsymbol{\zeta}_{sim}^{(k)}$ and try to predict the progression of dynamics using the continuous distribution hidden Markov model. Table 5.5 lists the parameter obtained after training the mixture model with the real data. We next re-sample the mixture of Gaussian by the parameters defined in 5.5 to create a new set of simulated training reduced coordinates, 3600 samples; 900 samples/rat for nine training rats, $\mathbf{Z}_{new}$. For this experiment $\tau^0 = 32, \tau^1 = 13, \tau^2 = 33$ *and* $\tau^3 = 43$, 121 samples for one rat, $\mathcal{T}$. We define the continuous distribution hidden Markov model with four states, $Q = 4$, and three mixture components per state, $M = 3$. Thus, we have a total of $Q \times M = 12$ mixtures for the model. We train the continuous distribution hidden Markov model using the new training set, $\mathbf{Z}_{new}$ and implement the Viterbi decoding algorithm using the new simulated test set, $\mathcal{T}$. Table 5.6 shows the new learned parameters of the mixture of Gaussian. Figure 5.6 shows the probability of the decoded Viterbi path of the most likely hidden states. The test sequence has 121 samples in total and $\tau^0 = 32, \tau^1 = 13, \tau^2 = 33$ *and* $\tau^3 = 43$ is the number of samples from each state.

Figure 5.7 show the probability of being in either state. We can see that the computational biomarker is able to predict the onset of chronic state even before it occurs. An algorithmic representation of the procedure followed in this test is given in 1.

	M	π	μ	Σ
Baseline	1	0.2680	$\begin{pmatrix} 0.0095 \\ -0.0250 \\ 0.0284 \\ -0.0064 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.0043 & 0.0045 & -0.0151 & -0.0045 \\ 0.0045 & 0.0934 & -0.0599 & -0.1720 \\ -0.0151 & -0.0599 & 0.0808 & 0.1001 \\ -0.0045 & -0.1720 & 0.1001 & 0.3248 \end{bmatrix}$
	2	0.5951	$\begin{pmatrix} 0.0088 \\ -0.0369 \\ 0.0373 \\ 0.0199 \end{pmatrix}$	$1 * 10^{-4} \begin{bmatrix} 0.0357 & -0.0025 & -0.1361 & 0.0425 \\ -0.0025 & 0.0290 & -0.0360 & -0.0697 \\ -0.1361 & -0.0360 & 0.6228 & -0.0733 \\ 0.0425 & -0.0697 & -0.0733 & 0.2139 \end{bmatrix}$
	3	0.1369	$\begin{pmatrix} -0.0160 \\ -0.0164 \\ 0.0252 \\ -0.0030 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.1839 & -0.1152 & 0.1599 & 0.0075 \\ -0.1152 & 0.0751 & -0.1021 & -0.0116 \\ 0.1599 & -0.1021 & 0.1410 & 0.0104 \\ 0.0075 & -0.0116 & 0.0104 & 0.0177 \end{bmatrix}$
Silent	1	0.4680	$\begin{pmatrix} -0.0501 \\ 0.0046 \\ -0.0084 \\ 0.0031 \end{pmatrix}$	$1 * 10^{-5} \begin{bmatrix} 0.2510 & -0.0723 & 0.2366 & -0.2533 \\ -0.0723 & 0.0506 & -0.0904 & 0.0378 \\ 0.2366 & -0.0904 & 0.2397 & -0.2127 \\ -0.2533 & 0.0378 & -0.2127 & 0.3008 \end{bmatrix}$
	2	0.2918	$\begin{pmatrix} -0.0512 \\ 0.0048 \\ -0.0093 \\ 0.0042 \end{pmatrix}$	$1 * 10^{-6} \begin{bmatrix} 0.3297 & -0.0905 & 0.3098 & -0.3324 \\ -0.0905 & 0.0399 & -0.0957 & 0.0704 \\ 0.3098 & -0.0957 & 0.2988 & -0.2978 \\ -0.3324 & 0.0704 & -0.2978 & 0.3655 \end{bmatrix}$
	3	0.2403	$\begin{pmatrix} -0.0528 \\ 0.0056 \\ -0.0112 \\ 0.0057 \end{pmatrix}$	$1 * 10^{-6} \begin{bmatrix} 0.4617 & -0.1175 & 0.4661 & -0.5888 \\ -0.1175 & 0.0366 & -0.1245 & 0.1466 \\ 0.4661 & -0.1245 & 0.4759 & -0.5921 \\ -0.5888 & 0.1466 & -0.5921 & 0.7556 \end{bmatrix}$
Latent	1	0.3728	$\begin{pmatrix} 0.0189 \\ -0.0215 \\ -0.0406 \\ 0.0104 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.0036 & 0.0097 & -0.0340 & -0.0081 \\ 0.0097 & 0.0412 & -0.0836 & -0.0477 \\ -0.0340 & -0.0836 & 0.3239 & 0.0642 \\ -0.0081 & -0.0477 & 0.0642 & 0.0711 \end{bmatrix}$
	2	0.2444	$\begin{pmatrix} -0.0256 \\ 0.0011 \\ 0.0073 \\ -0.0131 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.4807 & -0.0734 & 0.2896 & -0.2637 \\ -0.0734 & 0.1538 & -0.1290 & -0.0328 \\ 0.2896 & -0.1290 & 0.2278 & -0.1091 \\ -0.2637 & -0.0328 & -0.1091 & 0.2496 \end{bmatrix}$
	3	0.3828	$\begin{pmatrix} 0.0169 \\ 0.0182 \\ -0.0026 \\ -0.0117 \end{pmatrix}$	$\begin{bmatrix} 0.0000 & 0.0000 & -0.0000 & 0.0000 \\ 0.0000 & 0.0005 & 0.0002 & 0.0005 \\ -0.0000 & 0.0002 & 0.0003 & 0.0001 \\ 0.0000 & 0.0005 & 0.0001 & 0.0012 \end{bmatrix}$
Chronic	1	0.3525	$\begin{pmatrix} 0.0140 \\ 0.0167 \\ 0.0069 \\ -0.0442 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.0072 & 0.0131 & -0.0132 & 0.0174 \\ 0.0131 & 0.1112 & -0.0031 & 0.0771 \\ -0.0132 & -0.0031 & 0.0417 & -0.0184 \\ 0.0174 & 0.0771 & -0.0184 & 0.1397 \end{bmatrix}$
	2	0.3712	$\begin{pmatrix} 0.0186 \\ 0.0467 \\ 0.0108 \\ 0.0271 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.0009 & 0.0052 & 0.0007 & 0.0207 \\ 0.0052 & 0.0430 & 0.0134 & 0.1469 \\ 0.0007 & 0.0134 & 0.0069 & 0.0365 \\ 0.0207 & 0.1469 & 0.0365 & 0.5354 \end{bmatrix}$
	3	0.2763	$\begin{pmatrix} 0.0191 \\ -0.0187 \\ -0.0428 \\ 0.0046 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.0023 & -0.0034 & -0.0249 & 0.0163 \\ -0.0034 & 0.0390 & 0.0522 & -0.0829 \\ -0.0249 & 0.0522 & 0.2822 & -0.2036 \\ 0.0163 & -0.0829 & -0.2036 & 0.2222 \end{bmatrix}$

Table 5.5: Epilepsy model parameters before re-sampling

	M	π	μ	Σ
Baseline	1	0.6013	$\begin{pmatrix} 0.0088 \\ -0.0368 \\ 0.0371 \\ 0.0197 \end{pmatrix}$	$1 * 10^{-4} \begin{bmatrix} 0.0358 & -0.0029 & -0.1369 & 0.0441 \\ -0.0029 & 0.0325 & -0.0375 & -0.0792 \\ -0.1369 & -0.0375 & 0.6293 & -0.0712 \\ 0.0441 & -0.0792 & -0.0712 & 0.2413 \end{bmatrix}$
	2	0.3987	$\begin{pmatrix} 0.0022 \\ -0.0227 \\ 0.0283 \\ -0.0056 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.1785 & -0.0713 & 0.0530 & -0.0185 \\ -0.0713 & 0.1054 & -0.0818 & -0.1253 \\ 0.0530 & -0.0818 & 0.1028 & 0.0788 \\ -0.0185 & -0.1253 & 0.0788 & 0.2467 \end{bmatrix}$
Silent	1	0.4890	$\begin{pmatrix} -0.0500 \\ 0.0046 \\ -0.0084 \\ 0.0030 \end{pmatrix}$	$1 * 10^{-5} \begin{bmatrix} 0.2526 & -0.0736 & 0.2382 & -0.2532 \\ -0.0736 & 0.0508 & -0.0914 & 0.0390 \\ 0.2382 & -0.0914 & 0.2412 & -0.2131 \\ -0.2532 & 0.0390 & -0.2131 & 0.2989 \end{bmatrix}$
	2	0.5110	$\begin{pmatrix} -0.0520 \\ 0.0052 \\ -0.0102 \\ 0.0049 \end{pmatrix}$	$1 * 10^{-5} \begin{bmatrix} 0.1045 & -0.0416 & 0.1130 & -0.1042 \\ -0.0416 & 0.0186 & -0.0465 & 0.0390 \\ 0.1130 & -0.0465 & 0.1233 & -0.1110 \\ -0.1042 & 0.0390 & -0.1110 & 0.1077 \end{bmatrix}$
Latent	1	0.6382	$\begin{pmatrix} 0.0008 \\ 0.0122 \\ 0.0007 \\ -0.0137 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.5850 & 0.1574 & -0.0191 & -0.0459 \\ 0.1574 & 0.3856 & 0.0130 & 0.2141 \\ -0.0191 & 0.0130 & 0.2510 & 0.0113 \\ -0.0459 & 0.2141 & 0.0113 & 0.7658 \end{bmatrix}$
	2	0.3618	$\begin{pmatrix} 0.0190 \\ -0.0217 \\ -0.0418 \\ 0.0107 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.0033 & 0.0087 & -0.0317 & -0.0070 \\ 0.0087 & 0.0374 & -0.0753 & -0.0427 \\ -0.0317 & -0.0753 & 0.3046 & 0.0551 \\ -0.0070 & -0.0427 & 0.0551 & 0.0645 \end{bmatrix}$
Chronic	1	0.6296	$\begin{pmatrix} 0.0164 \\ 0.0000 \\ -0.0165 \\ -0.0207 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.0116 & -0.0408 & -0.0861 & 0.0808 \\ -0.0408 & 0.4036 & 0.4756 & -0.4379 \\ -0.0861 & 0.4756 & 0.8136 & -0.7403 \\ 0.0808 & -0.4379 & -0.7403 & 0.7991 \end{bmatrix}$
	2	0.3704	$\begin{pmatrix} 0.0185 \\ 0.0463 \\ 0.0107 \\ 0.0256 \end{pmatrix}$	$1 * 10^{-3} \begin{bmatrix} 0.0009 & 0.0052 & 0.0009 & 0.0201 \\ 0.0052 & 0.0442 & 0.0144 & 0.1490 \\ 0.0009 & 0.0144 & 0.0071 & 0.0398 \\ 0.0201 & 0.1490 & 0.0398 & 0.5340 \end{bmatrix}$

Table 5.6: Epilepsy model parameters after re-sampling

Algorithm 4 Computational Biomarker

procedure COMPUTATIONAL BIOMARKER($\mathbf{Z}, \mathcal{T}$) ▷ $\mathbf{Z}, \mathcal{T}$ are the Training and Test sets
  $T = 900$
  for $r \leftarrow 1, N \in Training\ Set$ **do** ▷ Concatenate the reduced coordinates of each cluster
    while $e \leftarrow 0, 3$ **do**
      $\mathbf{Z}^e \leftarrow \mathbf{Z}^e \cup \mathbf{Z}^e_r$
    end while
  end for
  while $e \leftarrow 0, 3$ **do**
    $GMM^e(\mu^e, \Sigma^e, \pi^e) \leftarrow$EM for $\mathbf{Z}^e$ ▷ Learn the parameters of GMM from each cluster
    $\mathbf{Z}^e_{new} \leftarrow T$ samples from $GMM^e(\mu^e, \Sigma^e, \pi^e)$ ▷ Form new training set
  end while
  $\lambda^{t_{final}} \leftarrow \lambda^{t_0}(\mathbf{Z}_{new}, Q, M)$ ▷ Input Parameter to HMM for maximization using EM
  $[\delta]_{M \times \tau} = Viterbi(\lambda^{t_{final}}, \mathcal{T})$ ▷ Decoding most likely sequence of hidden states
end procedure

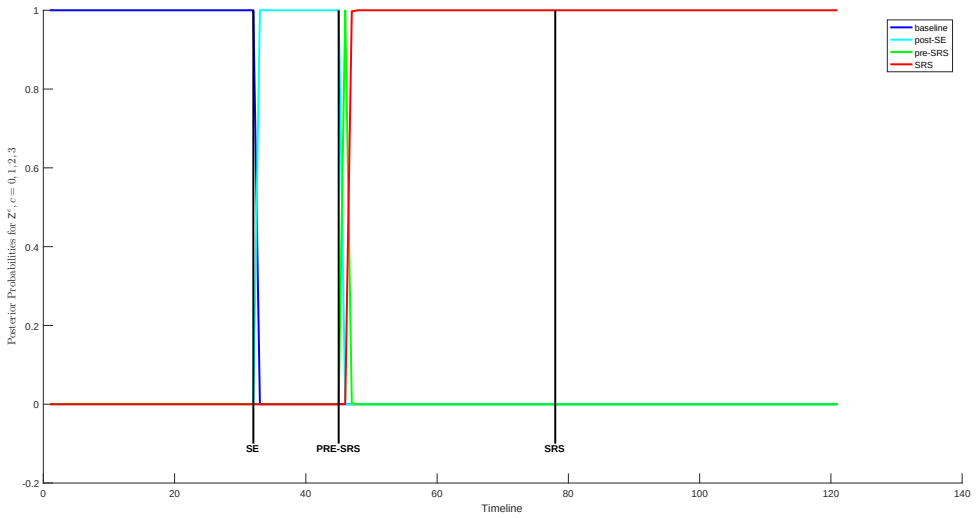

Figure 5.7: Estimated posterior probability of being in one of the states

5.5 Discussion

We can see from Figures 5.3 and 5.6 that the framework of continuous distribution hidden Markov models is successful in handling highly overlapped mixture models. We then are able to confirm in the last experiment that the computational biomarker can not only detect the onset of the chronic state but also predict before it happens 5.7. We have yet to test this biomarker on rats that do not develop epilepsy even though they were artificially induced with status epilepticus. But we are hopeful that this biomarker will give us promising results there as well.

Chapter 6

Conclusion

6.1 Introduction

This work provides experimental evidence that it is possible to decode a quantifiable and reliable measure of the alterations in the neuronal circuitry triggered by an insult to the hippocampus. This novel computational biomarker was able to exploit subtle changes in the evoked potential elicited by an auditory stimulus, which was recorded from the hippocampus. Solely based on the shape of the evoked potential, the algorithm was able to predict whether an animal will eventually recover from the injury, or develop spontaneous seizures. The decoding combined three key concepts. First, the subtle variations in the morphology of the noisy evoked potentials were extracted using a stationary wavelet transform. Second, a time-delay embedding of the wavelet coefficients combined with a spectral embedding provided a reduced set of coordinates for each evoked potential. These coordinates revealed the presence of smooth low-dimensional structures, which were stable across all the animals, and corresponded to well defined stages of epileptogenesis. These "universal structures" could be used reliably to determine the stage of epileptogenesis. Combined with a hidden Markov model, the decoding algorithm estimated the posterior probability that a given animal be in one of the four stages of epileptogenesis, based on the evoked potential. In conclusion , the presented work leads the path to a new generation of computational biomarkers of epeleptogenesis.

6.2 Limitations

There are a few limitation with our current implementation. We identify a few of them as follows:

(1) We noticed that both the *latent* and the *chronic* conditions had the largest spread of the reduced coordinates. This suggests that these conditions need to be further subdivided. Indeed, there is evidence that the chronic stage is in fact a dynamic condition, wherein seizures reactivate mechanisms contributing to ongoing cumulative damage. A finer gradation of the latent and chronic conditions could provide a deeper understanding of the evolution of epileptogenesis

(2) The Markov Chain in an HMM is a model of the dynamics of the system. These dynamics solely model the internal evolution of the system and do not take into consideration the external effects. However, this is not true for out test case. We induce status epilepticus (SE) in the rat, which is not an internal process. Perhaps, a better model would be to start with *silent* and then model the evolution to *latent* and *chronic*. Though this model will be more suited for the premise defined by the Markov Chain, there are however some animals that do return to *baseline* after spending time in *chronic*. This would cause a problem as the model will not have enough training data to model the *baseline* state.

6.3 From Animal to Clinical Model

We would like to discuss the possibility for this research to transition into a clinical study. Currently, the electrophysiological electrodes are directly inserted on the hippocampus of the rats brain. This is a highly invasive procedure with inherent risks and might not be feasible for clinical trials. The question we ask here is that ***Is there another credible source of the such information?***. A lot of studies have shown the EEG recording have been helpful in predicting the onset of seizures but this study doesn't have any experimental results which can show its efficacy in implementing an efficient biomarker.

Bibliography

[1] M. Ghil, et al. Advanced spectral methods for climatic time series. Rev. Geophys., 40(1):3–1, 2002.

[2] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.

[3] Bahman Bahmani, Benjamin Moseley, Andrea Vattani, Ravi Kumar, and Sergei Vassilvitskii. Scalable k-means++. Proceedings of the VLDB Endowment, 5(7):622–633, 2012.

[4] Tyrus Berry, John Robert Cressman, Zrinka Gregurie-Ferencek, and Timothy Sauer. Time-scale separation from diffusion-mapped delay coordinates. SIAM J. Appl. Dyn. Syst., 12(2):618–649, 2013.

[5] Jeff A Bilmes et al. A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. International Computer Science Institute, 4(510):126, 1998.

[6] K. Bush and J. Pineau. Manifold embeddings for model-based reinforcement learning under partial observability. In NIPS 2009, pages 189–197, 2009.

[7] G. Canas, T. Poggio, and L. Rosasco. Learning manifolds with k-means and k-flats. In Advances in Neural Information Processing Systems, pages 2465–2473, 2012.

[8] O. Cappé, E. Moulines, and T. Rydén. Inference in Hidden Markov Models. Springer, 2009.

[9] Laetitia Chauvière, Nadia Rafrafi, Catherine Thinus-Blanc, Fabrice Bartolomei, Monique Esclapez, and Christophe Bernard. Early deficits in spatial memory and theta rhythm in experimental temporal lobe epilepsy. J. Neurosci., 29(17):5402–5410, 2009.

[10] G. Chen, A. Little, M. Maggioni, and L. Rosasco. Some recent advances in multiscale geometric analysis of point clouds. In Wavelets and Multiscale Analysis, pages 199–225. Springer, 2011.

[11] Giulia Curia, Daniela Longo, Giuseppe Biagini, Roland SG Jones, and Massimo Avoli. The pilocarpine model of temporal lobe epilepsy. Journal of neuroscience methods, 172(2):143–157, 2008.

[12] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. Journal of the royal statistical society. Series B (methodological), pages 1–38, 1977.

[13] M Dunham, K Murphy, et al. Pmtk3: Probabilistic modeling toolkit for matlab/octave, 2010.

[14] J. Engel and F.L. da Silva. High-frequency oscillations–where we are and where we need to go. Prog. Neurobiol., 98(3):316–318, 2012.

[15] G David Forney Jr. The viterbi algorithm. Proceedings of the IEEE, 61(3):268–278, 1973.

[16] Zoubin Ghahramani and Michael I Jordan. Learning from incomplete data. 1995.

[17] Michael Ghil, MR Allen, MD Dettinger, K Ide, D Kondrashov, ME Mann, Andrew W Robertson, A Saunders, Y Tian, F Varadi, et al. Spectral dimensionality reduction. Springer, 2006.

[18] E.M. Goldberg and D.A. Coulter. Mechanisms of epileptogenesis: a convergence on neural circuit dysfunction. Nat. Rev. Neurosci., 14(5):337–349, 2013.

[19] Aleksandr Nikolaevich Gorban', Balazs Kegl, Donald C Wunsch, and Andrey Zinovyev. Principal manifolds for data visualization and dimension reduction, volume 58. Springer, 2008.

[20] DC Hesdorffer, G Logroscino, EKT Benn, N Katri, G Cascino, and WA Hauser. Estimating risk for developing epilepsy a population-based study in rochester, minnesota. Neurology, 76(1):23–27, 2011.

[21] D Hirtz, DJ Thurman, K Gwinn-Hardy, M Mohamed, AR Chaudhuri, and R Zalutsky. How common are the common neurologic disorders? Neurology, 68(5):326–337, 2007.

[22] C. Huneau, P. Benquet, G. Dieuset, A. Biraben, B. Martin, and F. Wendling. Shape features of epileptic spikes are a marker of epileptogenesis in mice. Epilepsia, 54(12):2219–2227, 2013.

[23] Michael I Jordan and Robert A Jacobs. Hierarchical mixtures of experts and the em algorithm. Neural computation, 6(2):181–214, 1994.

[24] B-H Juang. Maximum-likelihood estimation for mixture multivariate stochastic observations of markov chains. AT&T technical journal, 64(6):1235–1249, 1985.

[25] J.R. Lee, , S.O. Gharan, and L. Trevisan. Multiway spectral partitioning and higher-order Cheeger inequalities. Journal of the ACM, 61(6):37, 2014.

[26] K. Lukasiuk and Albert J. Becker. Molecular biomarkers of epileptogenesis. Neurotherapeutics, 11(2):319–323, 2014.

[27] P. Mirowski, D. Madhavan, Y. LeCun, and R. Kuzniecky. Classification of patterns of EEG synchronization for seizure prediction. Clin. Neurophysiol., 120(11):1927–1940, 2009.

[28] Florian Mormann, Ralph G Andrzejak, Christian E Elger, and Klaus Lehnertz. Seizure prediction: the long and winding road. Brain, 130(2):314–333, 2007.

[29] Javier R Movellan. Tutorial on hidden markov models.

[30] Kevin Murphy. Hidden markov model (hmm) toolbox for matlab. online at http://www. ai. mit. edu/~ murphyk/Software/HMM/hmm. html, 1998.

[31] Kevin P Murphy. Machine learning: a probabilistic perspective. MIT press, 2012.

[32] Ian Nabney. NETLAB: algorithms for pattern recognition. Springer Science & Business Media, 2002.

[33] P. Niyogi, S. Smale, and S. Weinberger. A topological view of unsupervised learning from noisy data. SIAM Journal on Computing, 40(3):646–663, 2011.

[34] A. Pitkänen and R. Immonen. Epilepsy related to traumatic brain injury. Neurotherapeutics, 11(2):286–296, 2014.

[35] A. Pitkänen and K. Lukasiuk. Mechanisms of epileptogenesis and potential treatment targets. The Lancet Neurology, 10(2):173–186, 2011.

[36] Lawrence R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2):257–286, 1989.

[37] Sriram Ramgopal, Sigride Thome-Souza, Michele Jackson, Navah Ester Kadish, Iván Sánchez Fernández, Jacquelyn Klehm, William Bosl, Claus Reinsberger, Steven Schachter, and Tobias Loddenkemper. Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy. Epilepsy Behav., 37:291–307, 2014.

[38] Sriram Ramgopal, Sigride Thome-Souza, Michele Jackson, Navah Ester Kadish, Iván Sánchez Fernández, Jacquelyn Klehm, William Bosl, Claus Reinsberger, Steven Schachter, and Tobias Loddenkemper. Seizure detection, seizure prediction, and closed-loop warning systems in epilepsy. Epilepsy & Behavior, 37:291–307, 2014.

[39] Richard A Redner and Homer F Walker. Mixture densities, maximum likelihood and the em algorithm. SIAM review, 26(2):195–239, 1984.

[40] Sandy R Shultz, Terence J O'Brien, Maria Stefanidou, and Ruben I Kuzniecky. Neuroimaging the epileptogenic process. Neurotherapeutics, 11(2):347–357, 2014.

[41] R.J. Staba, M. Stead, and G.A. Worrell. Electrophysiological biomarkers of epilepsy. Neurotherapeutics, 11(2):334–346, 2014.

[42] John D Storey. False discovery rate. In International encyclopedia of statistical science, pages 504–508. Springer, 2011.

[43] K.M. Taylor, , M. Procopio, C. Young, and F.G. Meyer. Estimation of arrival times from seismic waves: a manifold-based approach. Geophys. J. Int., 185(1):435–452, 2011.

[44] Vera C Terra, Roberta Cysneiros, Esper A Cavalheiro, and Fulvio A Scorza. Sudden unexpected death in epilepsy: from the lab to the clinic setting. Epilepsy & Behavior, 26(3):415–420, 2013.

[45] M. Tipping and C. Bishop. Mixtures of probabilistic principal component analyzers. Neural computation, 11(2):443–482, 1999.

[46] Keiichi Tokuda, Takushi Masuko, Noboru Miyazaki, and Takuo Kobayashi. Hidden markov models based on multi-space probability distribution for pitch pattern modeling. In Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on, volume 1, pages 229–232. IEEE, 1999.

[47] T. Vialar. Complex and Chaotic Nonlinear Dynamics. Springer, 2009.

[48] Andrew J Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. Information Theory, IEEE Transactions on, 13(2):260–269, 1967.

[49] Fabrice Wendling, Pascal Benquet, Fabrice Bartolomei, and Viktor Jirsa. Computational models of epileptiform activity. Journal of neuroscience methods, 2015.

[50] CF Jeff Wu. On the convergence properties of the em algorithm. The Annals of statistics, pages 95–103, 1983.

Appendix A

Gaussian Mixture Model

A.1 Introduction

One of the most widely used models is the mixture of Gaussians (MOG), also called a Gaussian Mixture Model or **GMM**. The model has the form [31, chapter 11, p. 339]

$$p(x_i|\theta) = \sum_{k=1}^{K} w_k \mathcal{N}(x_i|\mu_k, \Sigma_k) \tag{A.1}$$

$$\sum_{k=1}^{K} w_k = 1 \tag{A.2}$$

$$\mathcal{N}(x_i|\mu_k, \Sigma_k) = \frac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} e^{\{-\frac{1}{2}(x_i-\mu_k)^T \Sigma_k^{-1}(x_i-\mu_k)\}} \tag{A.3}$$

The equation describes a distribution with K Gaussian components with π_k posterior probabilities. x_i is an d-dimensional vector and μ_k & Σ_k are the means and covariances respectivel,y of the k^{th} component. **GMM's** have found their application in various fields such as computer vision, speech recognition, speaker verification and many more in biometric systems. **GMM** parameters are estimated by using either Expectation Maximization (EM) algorithm[12][39][16][23][50][31, chapter 11, p. 350].

A.1.1 Expectation Maximisation (EM) Algorithm for GMMs

In this section we provide an outline for the EM algorithm for GMMs. The mixture has K components and with posterior probabilities π_k. [31, chapter 11, p. 349-351]

A.1.1.1 Log-likelihood

Let x_i be the observed variables and z_i are the hidden variables. The goal is to maximize the log likelihood of the observed variable.

$$l(\theta) = \sum_{i=1}^{N} \log p(x_i|\theta) = \sum_{i=1}^{N} \log p(x_i, z_i|\theta) \tag{A.4}$$

This is hard to optimize as the log cannot be pushed inside the sum. Thus we define the complete data log likelihood to be

$$l_c(\theta) \triangleq \sum_{i=1}^{N} \log p(x_i, z_i|\theta) \tag{A.5}$$

Since z_i is unknown we define an auxiliary function $\mathbf{Q}$ to give us the `expected sufficient statistics`. The expectation is taken with respect to the old parameters θ^{t-1} and the observed data $\mathcal{D}$.

$$Q(\theta, \theta^{(t-1)}) = \mathbb{E}\left[l_c(\theta)|\mathcal{D}, \theta^{t-1}\right] \tag{A.6}$$

We optimize the function Q in the M Step with respect to θ

$$\theta^t = \arg\max_{\theta} Q(\theta, \theta^{t-1}) \tag{A.7}$$

A.1.1.2 Auxiliary Function

$$\begin{aligned}
Q(\theta, \theta^{(t-1)}) &\triangleq \mathbb{E}\left[\sum_i \log p(x_i, z_i|\theta)\right] && \text{(A.8)}\\
&= \sum_i \mathbb{E}\left[\log\left[\prod_{k=1}^{K} (\pi_k p(x_i|\theta_k)^{\mathbb{I}(z_i=k)})\right]\right] && \text{(A.9)}\\
&= \sum_i \sum_k \mathbb{E}\left[\mathbb{I}(z_i = k) \log\left[\pi_k p(x_i|\theta_k)\right]\right] && \text{(A.10)}\\
&= \sum_i \sum_k p(z_i = k|x_i, \theta^{t-1}) \log\left[\pi_k p(x_i|\theta_k)\right] && \text{(A.11)}\\
&= \sum_i \sum_k r_{ik} \log w_k + \sum_i \sum_k r_{ik} \log p(x_i|\theta_k) && \text{(A.12)}
\end{aligned}$$

where $r_{ik} \triangleq p(z_i = k|x_i, \theta^{(t-1)})$ is the responsibility that cluster k takes for the data point i. This is computed in the E step.

A.1.1.3 E Step

The E Step is straight forward

$$r_{ik} = \frac{w_k p(x_i|\theta_k^{t-1})}{\sum_{k'} \pi_{k'} p(x_i|\theta_{k'}^{t-1})} \tag{A.13}$$

A.1.1.4 M Step

We optimize Q with respect to π and θ_k. For π we have

$$w_k = \frac{1}{N} \sum_i r_{ik} = \frac{r_k}{N} \tag{A.14}$$

where $r_{ik} \triangleq \sum_i r_i k$ is the weighted number of points assigned to cluster k. To derive the M Step for the μ_k and Σ_k terms.

$$\mu_k = \frac{\sum_i r_{ik} x_i}{r_k} \tag{A.15}$$

$$\Sigma_k = \frac{\sum_i rik x_i x_i^T}{r_k} - \mu_k \mu_k^T \tag{A.16}$$

After computing the new estimates we set $\theta^t = (w_k, \mu_k, \Sigma_k) \forall k \in [1, K]$ and go on to compute the next E Step.

A.2 Discussion

Above we have defined the training of the mixture of Gaussian model using the EM algorithm. This mixture model is used as the continuous observation model for hidden Markov Model.

Appendix B

Continuous Density Hidden Markov Models

B.1 Introduction

First order Markov chains are modelled with the assumption that the immediate past, q_{t-1}, capture all relevant information about the history of model. With complex dynamics like the epilepsy model, we assume that the underlying dynamics can be modelled by first order Markov chains and the data is a noisy observation of the process. This result is known as a **Hidden Markov Model or HMM**. The HMM has hidden dynamics are modelled by a hidden variable, q_t, at time t. The transition model can be defined as a conditional probability distribution of $p(q_t|q_{t-1})$. The observation model is given by the conditional probability distribution of $p(o_t|q_t)$.

Depending on the type of process being modelled the observation model may vary. The observation could be discrete, continuous or may even be modelled by a mixture model. We are interested modelling the observations as mixture model. These are also known as Continuous Density HMMs[46]. The observation model is

$$p(o_t|q_t = j, \lambda) = \sum_{k=1}^{M} g_{jk} \mathcal{N}(o_t|\mu_{jk}, \Sigma_{jk})$$

where λ are the HMM parameters and the continuous distribution is a Gaussian Mixture Model.

At each time step, the Continuous Density HMM generates a hidden state, q_t, as per the state to state transition model. Next, the model chooses a component, m_t, as per the state to component emission model. Once the component has been determines, the observation vector, x_t, is probabilistically chosen according to the Gaussian Probability distribution. For simplicity each

state is assumed to have the same number of components. The variable, v_{jk}, is used to represent the component k given the current state as j. q_t and m_t are random variables used to represent the state and the component respectively.

$$P_\lambda(o_t|v_{jk}) = P_\lambda(o_t|q_t = j, m_t = k) = \mathcal{N}(o_t|\mu_{jk}, \Sigma_{jk}) \quad \text{(B.1)}$$

where μ_{jk} and Σ_{jk} is the mean and covariance respectively of the k^{th} cluster belonging to the j_{th} state.

Given a state j, the system randomly chooses one its possible components according to the state to component emission probability, $g_{jk} = P(m_t = k|q_t = j)$. It is important to note that this probability is independent of the time, t, and thus is represented without the time index. Using equation B.1, the emission probability is now

$$b_j(o_t) = P_\lambda(o_t|q_t = j) = \sum_{i=1}^{M} P(m_t = i|q_t = j)P(o_t|q_t = j, m_t = i) \quad \text{(B.2)}$$

$$b_j(o_t) = \sum_{i=1}^{M} g_{ji}\mathcal{N}(o_t|\mu_{ji}, \Sigma_{ji}) \quad \text{(B.3)}$$

For each state, g_{ji}, should satisfy the stochastic constraint

$$\sum_{i=1}^{M} g_{ji} = 1 \quad \text{(B.4)}$$

$$g_{ji} \geq 0, \forall j \in [1, N], i \in [1, M] \quad \text{(B.5)}$$

so that $b_j(\cdot)$ are normalized, i.e.,

$$\int_{\mathbb{R}^K} b(o_t)do_t = 1 \quad \text{(B.6)}$$

B.1.1 Continuous Density HMM Training

We describe the complete set of parameters for a Continuous Density HMM as

$$\lambda = (A, B, \pi) \quad \text{(B.7)}$$

where A is the transition probability matrix, B is the emission probability matrix and π is the initial probability of each state. There are three basic approaches that we are focused on[5]

(1) Forward-Backward Algorithm: Find $p(O|\lambda)$ for some $O = (o_1, o_2, ..., o_T)$. We use this because it is more efficient that directly evaluating the Continuous Density HMM.

(2) Baum-Welch Algorithm: Also know as the EM algorithm for Continuous Density HMM. This algorithm provides a method to estimate $\lambda^\star = \arg\max_\lambda p(O|\lambda)$

(3) Viterbi Algorithm: Given the model parameters O and λ, find the best state sequence $q = (q_1, q_2, ..., q_T)$ that increases the likelihood of O.

B.1.1.1 Forward-Backward Algorithm

The probability of seeing a partial sequence and ending up in state i given the model parameters is

$$\alpha_i(t) = p(O_1 = o_1, \cdots, O_t = o_t, Q_i = i|\lambda) \tag{B.8}$$

This is know as the forward procedure. We can efficiently calculate $\alpha_i(t)$ recursively

$$\alpha_i(1) = \pi_i b_i(o_1) \tag{B.9}$$

$$\alpha_j(t+1) = \left[\sum_{i=1}^{N} \alpha_i(t) a_{ij}\right] b_j(o_{t+1}) \tag{B.10}$$

$$p(O|\lambda) = \sum_{i=1}^{N} \alpha_i(T) \tag{B.11}$$

Equation B.9 is the initialization step, B.10 is the recursive step and B.11 is the terminating step. The backward procedure is similar to the forward procedure

$$\beta_i(T) = 1 \tag{B.12}$$

$$\beta_i(t) = \sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \beta_j(t+1) \tag{B.13}$$

$$p(O|\lambda) = \sum_{i=1}^{N} \beta_i(1) \pi_i b_i(o_1) \tag{B.14}$$

Equation B.12 is the initialization step, B.13 is the recursive step and B.14 is the terminating step.

B.1.1.2 Expectation Maximisation (EM) Algorithm

Also know as the Baum-Welch algorithm for Hidden Markov Models, is used to train the Continuous Density HMM. The EM algorithm is used to estimate the parameters for the Continuous Density HMM. The algorithm is as follows[5]

We consider an observation sequence $O = (o_1, \cdots, o_T)$ and its underlying state sequence to be $q = (q_1, \cdots, q_T)$, which is hidden. The incomplete-data likelihood function is give by $P(O|\lambda)$ and the complete-data likelihood is $P(O, q|\lambda)$. Thus, the auxiliary function, Q, is defined as

$$Q(\lambda^t, \lambda^{t-1}) = \sum_{q \in \mathcal{Q}} \log P(O, q|\lambda^t) P(O, q|\lambda^{t-1}) \tag{B.15}$$

where λ^{t-1} is the current estimate and $\mathcal{Q}$ is the space of all the hidden states.

Given a particular state sequence $P(O, q|\lambda^{t-1})$ is given by

$$P(O, q|\lambda^{t-1}) = \pi_{q_0} \prod_{t=1}^{T} a_{q_{t-1} q_t} b_{q_t}(o_t) \tag{B.16}$$

The Q function becomes:

$$\begin{aligned} Q(\lambda^t, \lambda^{t-1}) = & \sum_{q \in \mathcal{Q}} \log \pi_{q_0} P(O, q|\lambda^{t-1}) \\ + & \sum_{q \in \mathcal{Q}} \left(\sum_{t=1}^{T} \log a_{q_{t-1} q_t} \right) p(O, q|\lambda^{t-1}) \\ + & \sum_{q \in \mathcal{Q}} \left(\sum_{t=1}^{T} \log b_{q_t}(o_t) \right) P(O, q|\lambda^{t-1}) \end{aligned} \tag{B.17}$$

The parameter set λ which maximizes the equation is subject to stochastic constraints

$$\sum_{i=1}^{N} \pi_i = 1 \tag{B.18}$$

$$\sum_{i=1}^{N} a_{ij} = 1, \ 1 \leq i \leq N \tag{B.19}$$

$$\sum_{m=1}^{M} g_{jm} = 1, \ 1 \leq j \leq N \tag{B.20}$$

Now optimizing each term of the B.17 individually. Thus, the first term of B.17 is

$$\sum_{q \in \mathcal{Q}} \log \pi_{q_0} P(O, q|\lambda^{t-1}) = \sum_{i=1}^{N} \log \pi_i p(O, q_0 = i|\lambda_{t-1}) \tag{B.21}$$

Since by selecting all $q \in \mathcal{Q}$ we are simply repeatedly selecting values of q_0, so the right hand side is just a marginal expression for time $t = 0$. Adding the Lagrange multiplier[24], γ, and using the constraint B.18 and setting the derivative to zero

$$\frac{\partial}{\partial \pi_i}\left(\sum_{i=1}^{N} \log \pi_i p(O, q_0 = i|\lambda^{t-1}) + \gamma\left(\sum_{i=1}^{N} \pi_i - 1\right)\right) = 0 \tag{B.22}$$

Taking the derivative, summing it over i to get γ, and solving for π_i we get:

$$\pi_i = \frac{P(O, q_0 = i|\lambda^{t-1})}{P(O|\lambda^{t-1})} \tag{B.23}$$

The second term of B.17 is

$$\sum_{q\in\mathcal{Q}}\left(\sum_{t=1}^{T} \log a_{q_{t-1}q_t}\right) p(O, q|\lambda^{t-1}) = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} \log a_{ij} P(O, q_{t-1} = i, q_t = j|\lambda^{t-1}) \tag{B.24}$$

because for this term, for each t, we are looking over all transitions from i to and weighting that by the corresponding probability − the right hand side is just sum of the joint-marginal for time $t-1$ and t. In a similar way, we can use a Lagrange multiplier[24] with the constraint B.19 to get:

$$a_{ij} = \frac{\sum_{t=1}^{T} P(O, q_{t-1} = i, q_t = j|\lambda^{t-1})}{\sum_{t=1}^{T} P(O, q_{t-1} = i|\lambda^{t-1})} \tag{B.25}$$

The third term of B.17 becomes

$$\sum_{q\in\mathcal{Q}}\left(\sum_{t=1}^{T} \log b_{q_t}(o_t)\right) P(O, q|\lambda^{t-1}) = \sum_{i=1}^{N}\sum_{t=1}^{T} \log b_i(o_t) p(O, q_t = i|\lambda^{t-1}) \tag{B.26}$$

because for this term, for each time t, we are looking for emissions for all the states and weighting each possible emission by the corresponding probability − the right side is just the sum of marginal for time t. For Gaussian mixtures, the for of the auxiliary function is different, i.e., the hidden variables must include not only the hidden state sequence, but also a variable indicating the mixture component for each state at each time. Therefore, we can write Q as:

$$Q(\lambda^t, \lambda^{t-1}) = \sum_{q\in\mathcal{Q}}\sum_{m\in\mathcal{M}} \log P(O, q, m|\lambda^t) P(O, q, m|\lambda^{t-1}) \tag{B.27}$$

where m is a vector $m = \{m_{q_1 1}, \cdots, m_{q_T T}\}$ that indicates the mixture component for each state at each time. If we expand this as B.17, the first and second terms remain unchanged as they are

not dependent on the mixture components. Thus, the third term becomes:

$$\sum_{q\in\mathcal{Q}}\sum_{m\in\mathcal{M}}\left(\sum_{t=1}^{T}\log b_{q_t}(o_t,m_{q_tt})\right)P(O,q,m|\lambda^{t-1}) = \sum_{i=1}^{N}\sum_{l=1}^{M}\sum_{t=1}^{T}\log(g_{il}b_{il}(o_t))p(O,q_t=i,m_{q_tt}=l|\lambda^{t-1}) \tag{B.28}$$

To find the expression for g_{il}, we introduce the Lagrange multiplier[24] with the constrain B.20 to get:

$$g_{il} = \frac{\sum_{t=1}^{T}P(q_t=i,m_{q_tt}=l|O,\lambda^{t-1})}{\sum_{t=1}^{T}\sum_{l=1}^{M}P(q_t=i,m_{q_tt}=l|O,\lambda^{t-1})} \tag{B.29}$$

$$\mu_{il} = \frac{\sum_{t=1}^{T}o_tP(q_t=i,m_{q_tt}=l|O,\lambda^{t-1})}{\sum_{t=1}^{T}P(q_t=i,m_{q_tt}=l|O,\lambda^{t-1})} \tag{B.30}$$

$$\Sigma il = \frac{\sum_{t=1}^{T}(o_t-\mu_{il})(o_t-\mu_{il})^TP(q_t=i,m_{q_tt}=l|O,\lambda^{t-1})}{\sum_{t=1}^{T}P(q_t=i,m_{q_tt}=l|O,\lambda^{t-1})} \tag{B.31}$$

B.23,B.25,B.29,B.30 and B.31 are implemented recursively till we obtain the most likely fit or the maximum number of iterations elapse.

B.1.1.3 Viterbi Algorithm[36]

Given a the model for the Continuous Density HMM, $\lambda = (A, B, \pi)$, it can be useful for decoding the most likely hidden states sequence, $\hat{q} = (q_1, \cdots, q_t)$, give then observation vector sequence, $O = (o_1, \cdots, o_T)$.

The most likely state sequence, $\hat{q}$, is calculates in a way similar to that for the forward algorithm, which is also refferred to as the Viterbi Algorithm[48][15]. Let $\delta_t(i)$ be the likelihood of the most likely state sequence ending in the i_th state at time t.

$$\delta_i(t) = \max_{q_1,\cdots,q_{t-1}} p(q_1,\cdots,q_{t-1},q_t=i,o_1,\cdots,o_t|\lambda) \tag{B.32}$$

and $\psi_t(i)$ be the array to keep track. The procedure to calculate the most likely state sequence is

(1) Initialise

$$\delta_1(t) = \pi_i b_i(o_1) \tag{B.33}$$

$$\psi_1(i) = 0 \tag{B.34}$$

(2) Recursion

$$\delta_t(j) = \max_i \left[\delta_{t-1}(i)a_{ij}\right] b_j(o_t) \tag{B.35}$$

$$\psi_t(i) = \arg\max_i [\delta_{t-1}(i)a_{ij}] \tag{B.36}$$

(3) Termination

$$\hat{P} = \max_i [\delta_T(i)] \tag{B.37}$$

$$\hat{q} = \arg\max_i [\delta_T(i)] \tag{B.38}$$

(4) Back-Tracking

$$\hat{q} = \psi_{t+1}(\hat{q}_{t+1}) \tag{B.39}$$

The difference between the forward algorithmB.9-B.11 and Viterbi algorithm is the maximisation stepB.35.

B.2 Discussion

Above we have defined the training of the Continuous Density hidden Markov model using the EM algorithm or Baum-Welch algorithm. This is model is used to learn the dynamics of epileptogenesis.