ONLINE AUTOMATIC EPILEPTIC SEIZURE DETECTION FROM ELECTROENCEPHALOGRAM (EEG)

By

HUI LIU

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2005
ACKNOWLEDGMENTS

I thank my advisor Dr. José C. Príncipe for making every aspect of this research possible. I appreciate his constant enthusiasm for guiding, searching, facilitating and driving the research progress; his patience in editing revisions of the manuscript. I thank Drs. John G. Harris, J. Chris Sackllares, Mark C. K. Yang and Janise McNair for taking time from their busy schedules to serve in my committee. I thank Drs. J.B. Gao, Kenneth Hild II and Jeongho Cho for their support in the process of the system build up.

I thank Ganesan Ramachandran the Computational NeuroEngineering Laboratory (CNEL) brain dynamic group team member; and Linda Kahila, the graduate student advisor, for their encouragement, when I was confused, they kindly offered helping hands. I want to thank Pastor Joel Osteen for using simple stories to help me streamline my understanding of life during the past 5 years. Somehow I feel my appreciation can reach him although he dose not know me.

I thank Song Lin, Feng Nian Lin, my brothers and sister and all my friends back home for taking care of my parents while I was far away. Despite all of my dreams and ambition, the thought that “I cannot let them down” keep me going. I am forever grateful.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>ACKNOWLEDGMENTS</strong></td>
<td>iii</td>
</tr>
<tr>
<td></td>
<td><strong>LIST OF TABLES</strong></td>
<td>vii</td>
</tr>
<tr>
<td></td>
<td><strong>LIST OF FIGURES</strong></td>
<td>viii</td>
</tr>
<tr>
<td></td>
<td><strong>LIST OF OBJECTS</strong></td>
<td>xii</td>
</tr>
<tr>
<td></td>
<td><strong>ABSTRACT</strong></td>
<td>xiii</td>
</tr>
<tr>
<td>1</td>
<td><strong>INTRODUCTION</strong></td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td><strong>REVIEW</strong></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>- Seizure Classification and Artifact Categories</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>- Seizure Detection Algorithms</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>- From Adult EEG</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>- From Neonatal EEG</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>- Seizure Prediction Algorithms</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>- Seizure Prediction through Linear and Nonlinear Signal Processing</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>- Seizure Prediction through Nonlinear Dynamic Theory</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>- Dealing with Artifacts</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>- Summary</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td><strong>SEIZURE-DETECTION SCHEME</strong></td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>- Data</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>- Feature Extraction</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>- Artifact Sources</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>- Automatic Artifacts Correction by Template Matching</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>- System Schematic</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>- Contributions</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td><strong>FEATURE SELECTION</strong></td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>- Short Term Lyaponov Exponent Feature and Seizure Detection</td>
<td>45</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>2-1</td>
<td>Categorization of seizures.</td>
<td>9</td>
</tr>
<tr>
<td>3-1</td>
<td>General information of the five study sets.</td>
<td>40</td>
</tr>
<tr>
<td>3-2</td>
<td>Detail of P93 set.</td>
<td>40</td>
</tr>
<tr>
<td>3-3</td>
<td>Detail of P185 and P148 sets.</td>
<td>41</td>
</tr>
<tr>
<td>3-4</td>
<td>Detail of P92 and P256 sets.</td>
<td>41</td>
</tr>
<tr>
<td>4-1</td>
<td>Performance.</td>
<td>53</td>
</tr>
<tr>
<td>4-2</td>
<td>Parameters.</td>
<td>54</td>
</tr>
<tr>
<td>4-3</td>
<td>Good compromise point of detection vs. false alarm (T1 feature)</td>
<td>59</td>
</tr>
<tr>
<td>4-4</td>
<td>Good compromise point of detection vs. false alarm.(Teager feature)</td>
<td>63</td>
</tr>
<tr>
<td>5-1</td>
<td>Rotation matrix calculation</td>
<td>80</td>
</tr>
<tr>
<td>8-1</td>
<td>Scalp EEG study set information.</td>
<td>114</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>ECoG seizure samples</td>
<td>8</td>
</tr>
<tr>
<td>2-2</td>
<td>Artifacts samples</td>
<td>8</td>
</tr>
<tr>
<td>3-1</td>
<td>Basic electrode placement</td>
<td>39</td>
</tr>
<tr>
<td>3-2</td>
<td>Seizure detection system</td>
<td>44</td>
</tr>
<tr>
<td>4-1</td>
<td>STLmax feature</td>
<td>47</td>
</tr>
<tr>
<td>4-2</td>
<td>Background and in-seizure modeling</td>
<td>48</td>
</tr>
<tr>
<td>4-3</td>
<td>Power spectral density of background modeling error</td>
<td>49</td>
</tr>
<tr>
<td>4-4</td>
<td>Autocorrelation of background modeling error</td>
<td>50</td>
</tr>
<tr>
<td>4-5</td>
<td>Model and modeling error collection details</td>
<td>50</td>
</tr>
<tr>
<td>4-6</td>
<td>Seizure detection scheme</td>
<td>52</td>
</tr>
<tr>
<td>4-7</td>
<td>Test ROCs</td>
<td>54</td>
</tr>
<tr>
<td>4-8</td>
<td>T1 Illustration</td>
<td>56</td>
</tr>
<tr>
<td>4-9</td>
<td>T1 feature</td>
<td>57</td>
</tr>
<tr>
<td>4-10</td>
<td>Regional T1</td>
<td>58</td>
</tr>
<tr>
<td>4-11</td>
<td>Seizure detection ROC</td>
<td>59</td>
</tr>
<tr>
<td>4-12</td>
<td>Clinical seizure detection ROC</td>
<td>60</td>
</tr>
<tr>
<td>4-13</td>
<td>Teager energy feature</td>
<td>61</td>
</tr>
<tr>
<td>4-14</td>
<td>Seizure detection performance</td>
<td>62</td>
</tr>
<tr>
<td>4-15</td>
<td>Clinical seizure detection performance</td>
<td>62</td>
</tr>
<tr>
<td>4-16</td>
<td>T1 feature variation</td>
<td>66</td>
</tr>
<tr>
<td>Page</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>4-17</td>
<td>Sine wave and Teager energy.</td>
<td>68</td>
</tr>
<tr>
<td>4-18</td>
<td>T1 of different sine waves extracted under different radius.</td>
<td>68</td>
</tr>
<tr>
<td>4-19</td>
<td>Sine wave with different frequency.</td>
<td>69</td>
</tr>
<tr>
<td>4-20</td>
<td>Sine wave with different amplitude.</td>
<td>70</td>
</tr>
<tr>
<td>4-21</td>
<td>EEG samples.</td>
<td>71</td>
</tr>
<tr>
<td>4-22</td>
<td>Background trajectories.</td>
<td>72</td>
</tr>
<tr>
<td>4-23</td>
<td>In-seizure trajectories.</td>
<td>72</td>
</tr>
<tr>
<td>5-1</td>
<td>MRMI-SIG algorithm block diagram.</td>
<td>78</td>
</tr>
<tr>
<td>5-2</td>
<td>Artificial mixed signal separation example.</td>
<td>83</td>
</tr>
<tr>
<td>5-3</td>
<td>Electrode pop artifact.</td>
<td>83</td>
</tr>
<tr>
<td>5-4</td>
<td>Line artifact.</td>
<td>84</td>
</tr>
<tr>
<td>5-5</td>
<td>Muscle artifact.</td>
<td>84</td>
</tr>
<tr>
<td>6-1</td>
<td>Density distribution $Q_j$ and transition distribution $Y_{ij}$ of a trained template</td>
<td>88</td>
</tr>
<tr>
<td>6-2</td>
<td>The first two dimension of Lorenz (x) SOM template.</td>
<td>88</td>
</tr>
<tr>
<td>6-3</td>
<td>MG(27)-MG(30)-MG(27) mixture</td>
<td>92</td>
</tr>
<tr>
<td>6-4</td>
<td>One dimension vs, high dimension Kullback-Leibler divergence</td>
<td>92</td>
</tr>
<tr>
<td>6-5</td>
<td>Effects of PE number (MG30 is the template)</td>
<td>94</td>
</tr>
<tr>
<td>6-6</td>
<td>Noise effect (10 x 10 MG30 template)</td>
<td>95</td>
</tr>
<tr>
<td>6-7</td>
<td>Divergences between MG30 template and the artificial mixed time series</td>
<td>98</td>
</tr>
<tr>
<td>6-8</td>
<td>First order dynamic correlation coefficients between MG30 template and the artificial mixed time series</td>
<td>99</td>
</tr>
<tr>
<td>6-9</td>
<td>Seizure ECoG samples.</td>
<td>99</td>
</tr>
<tr>
<td>6-10</td>
<td>Comparison of two ECoG segments.</td>
<td>100</td>
</tr>
<tr>
<td>6-11</td>
<td>Relationship between KL and CC measures.</td>
<td>102</td>
</tr>
<tr>
<td>7-1</td>
<td>P93 feature.</td>
<td>103</td>
</tr>
<tr>
<td>7-2</td>
<td>P148 feature.</td>
<td>104</td>
</tr>
</tbody>
</table>
8-12 Components of Figure 8-11. ................................................................................ 121
8-13 EEGLAB drop down menu.................................................................................. 122
9-1 Samples of EEG and the corresponding frequency character.............................. 125
9-2 Wavelet decomposition scheme........................................................................... 126
9-3 Wavelet preprocessed seizure detection scheme. ................................................ 126
9-4 Performance of seizure detection with wavelet preprocessing. ......................... 127
9-5 KL divergences between the in-seizure template and the ECoG signal of LTD1....................................................................................................................... 128
9-6 First order dynamic correlation coefficients between the in-seizure template and the ECoG signal of LTD1. ............................................................. 128
9-7 Component topography of Figure 8-7.................................................................. 130
## LIST OF OBJECTS

<table>
<thead>
<tr>
<th>Object</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-1</td>
<td>Sine wave amplitude change and T1.</td>
</tr>
<tr>
<td>4-2</td>
<td>Sine wave frequency change and T1.</td>
</tr>
<tr>
<td>4-3</td>
<td>Background EEG and T1.</td>
</tr>
<tr>
<td>4-4</td>
<td>In-seizure EEG and T1.</td>
</tr>
</tbody>
</table>
Abstract of Dissertation Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

ONLINE AUTOMATIC EPILEPTIC SEIZURE DETECTION FROM ELECTROENCEPHALOGRAM (EEG)

By

Hui Liu

December 2005

Chair: José C. Príncipe
Major Department: Electrical and Computer Engineering

We examined epileptic seizure detection methods. Our aim was to detect seizures online and automatically, thus reducing the workload imposed on the neurologist who reads days of EEG recordings. We anticipate a good detection rate, with a tolerable false-alarm rate.

To detect seizures from EEG, a critical issue is feature extraction. Feature extraction decides the performance of the detection and also the complexity of the algorithm. Up-to-date, linear feature-extraction methods have been well studied. Nonlinear dynamic methods were mainly applied to seizure prediction and generated many positive results. Our intention was to test the applicability of nonlinear dynamic methods for seizure detection.

We tested nonlinear features: Short Term maximum Lyaponov exponent (STLmax), recurrence time statistics T1 and Teager energy. Based on the “online automatic” requirement, T1 was the chosen feature.
To reduce false alarm, we proposed blind source separation and a self-organizing map-based template-matching algorithm as a preprocessor before feature extraction and the combined performance were tested. The preprocessing step automatically corrected some artifacts presented in ElectroCorticoGram (ECoG) data, but online automatic implementation on scalp EEG was troublesome because of the complexity of artifacts presented in scalp EEG. Thus the scheme is not applicable to EEG in general.

To detect seizures from both scalp EEG and ECoG, we proposed and tested wavelet preprocessing combined with T1 feature extraction. On 4 sets of scalp EEG, (124 hours recording with 44 seizures), and 5 sets of ECoG (235 hours with 71 seizures), the average detection rate reached 93% when producing 0.25 false alarms per hour. Wavelet preprocessing combined with T1 feature extraction holds promise for seizure detection in real clinical applications.
CHAPTER 1
INTRODUCTION

Epilepsy is not a disease: it is a set of disorders of the nervous system. Epileptic patients tend to have unprovoked epileptic seizures, with no periodicity. Seizures usually disturb the normal electrical rhythms of the central nervous system and are concurrent with clinically convulsive attacks, oftentimes with clouding of consciousness. Seizures triggered by clear precipitants, termed “acute symptomatic seizures”, do not constitute epilepsy, even if repeated [1], [2].

It is generally accepted that the prevalence of active epilepsy is in the range of 5 to 10 cases per 1000 people, and the incidence is around 50 per 100,000 per year in developed countries. Around 70 to 80% of the diagnosed cases respond well to antiepileptic drugs, although the drug influence on the natural history of epilepsy is unknown. The remaining 20 to 30% of patients are classified as sufferers of chronic refractory epilepsy [2]-[4].

Data from the industrialized world indicate that people with epilepsy have a greater risk of death than those without epilepsy [5], [6]. The increased risk is most pronounced in people with chronic refractory epilepsy [3].

One option for the patient with chronic refractory epilepsy is surgery. The first step before deciding on surgical intervention is accurate diagnosis and identification of the epileptogenic brain regions (the foci) responsible for the seizure. Selected surgery candidates must undergo painstaking and laborious electrophysiological, structural (neuro-radiological), and neuro-physiological tests.
The most useful test in epilepsy diagnosis is EEG. Both scalp and invasive EEG (invasive EEG is usually called ECoG) may be used. Usually the EEG recording is concurrent with patient behavior video monitoring. When scalp EEG artifacts obscure the side of seizure onset, or when clinical seizure (behavior) onset precedes electrical seizure onset, invasive tests (ECoG recorded from subdural, epidural or depth electrode insertion) are usually needed [7].

Recorded EEG is not used to make major surgical decisions, but to lateralize a focus when the hemisphere of ictal onset is not known with certainty, or to localize a focus when the hemisphere of seizure onset is defined clearly but the location and/or limits of the focus are not clear [1], [7].

To lateralize and localize seizure focus, patients must be under EEG monitoring for days, as the recordings must capture multiple seizures to be useful (it is impossible to make any diagnosis based on one seizure); and more often than not, the seizure event is rare and unpredictable. This means long-term monitoring and large amounts of data collecting.

Depending on the type and patient, seizures usually last from seconds to minutes [8]. To find these few short in-seizure sections, neurologists spend large amounts of time scoring days of EEG recording. To make things worse 95 to 99% of the recording offers little information for evaluating the epileptic disorder [9]. Thus effective automatic seizure-detection algorithms have long been sought by research groups and individuals around the world.

In following eight chapters, the development of a seizure detection scheme is presented. Chapter 2 reviewed the literatures related with seizure detection that have been
published so far. The review was intended to develop a sense of where we are, and why we chose to study certain aspect of the signal. Chapter 3 summarized the final seizure-detection scheme and the contributions of our research. Chapter 4 detailed the feature selection process. Chapter 5 studied the artifacts sources decomposition algorithm. Chapter 6 presented an automatic artifacts correction method. Chapter 7 hooked up the building blocks to test the performance on ECoG. Chapter 8 tested the applicability on scalp EEG. Chapter 9 proposed and tested a practical scheme for general seizure detection and some thought on future research.
CHAPTER 2
REVIEW

The irreplaceable position of the EEG signal in the process of epilepsy diagnosis has been the driving force behind many EEG studies. Beginning 1961, patients with temporal lobe epilepsy of suspected deep focal origin were implanted bilaterally, with depth-recording electrodes for the purpose of localizing the site of seizure onset. But until the early 1970s, recorded seizures were inspected only when seizure occurrences were reported by the ward nurses or the patient; unreported seizures were simply discarded because it was not economical to find the few minutes of relevant data from many hours of recording [10], [11].

The attempt to automatically detect seizures started in the early 1970s and the trend has been growing since. On the one hand, EEG recording techniques open a tiny window that attracts ever more researchers to study the most interesting natural system-our brain. But the recording packed all the activity onto one signal. EEG recorded both seizure activities and large amount of complex background activities including varieties of artifacts. Further more each brain is a unique system: in-seizure signal character varies from brain to brain. Within the same brain, seizures also vary. On the other hand, theoretical (neural network, wavelet, nonlinear dynamic, digital signal processing, etc.) and technological (the processing ability of computer, digital electroencephalographic technology, advance of medical devices, etc.) development gave researchers more tools to extract information from EEG signals. New tools might help researchers better understand brain mechanism.
To build past development, we reviewed present seizure-detection techniques and related issues. First, we reviewed the classification of seizure based on clinical symptoms and the category of artifacts; then we reviewed the progress on seizure detection methodologies. Because a good seizure predictor would also be a good seizure detector, we reviewed prediction methods. As methods rely on EEG analysis have to deal with artifacts, we also reviewed artifact-rejection methodologies.

**Seizure Classification and Artifact Categories**

The International League against Epilepsy (ILAE) developed an international classification of epileptic seizures in 1981 [1], [8], [12], [13]. Seizures were divided into two major classes: partial-onset and generalized-onset. Partial-onset seizures begin in one focal area of the cerebral cortex. While generalized-onset seizures have an onset recorded simultaneously in both cerebral hemispheres. The seizures that are hard to fit in both classes were considered as unclassified.

Partial seizures were further classified into three subcategories: simple partial, complex partial, secondarily generalized tonic-clonic seizures. The key defining element of simple partial seizure is the occurrence of seizure with preservation of consciousness (aura of different types in Table 2-1). When consciousness is impaired, it is a complex partial seizure. The secondary generalized tonic-clonic seizures are those with tonic-clonic symptoms: they might start from simple partial, complex partial, simple to complex partial; then evolve into tonic-clonic.

Generalized-onset seizures (grand mal) represent a maximal epileptic response of the brain. The response is not exactly uniform and may vary somewhat from person to person, or from seizure to seizure. The attack may be generalized from the start, or maybe ushered in by focal (partial) seizure activity (like aura). It was classified into six subtypes.
• **Absence seizures**: brief episodes of impairment of consciousness, with no aura or post-ictal confusion; episodes typically last less than 20 seconds and are accompanied by few or no automatisms.

• **Myclonic seizures**: brief, arrhythmic, jerking, motor movements that last less than a second.

• **Clonic seizures**: rhythmic, motor, jerking movements, with impairment of consciousness.

• **Tonic seizures**: sudden-onset tonic extension or flexion of head, trunk, and/or extremities (usually last a few seconds).

• **Tonic clonic seizures**: several motor behaviors, including generalized tonic extension of the extremities lasting for few seconds, followed by clonic rhythmic movements and prolonged post-ictal confusion. Clinically, the only behavioral difference between these seizures and secondarily generalized tonic-clonic seizures is that these seizures lack an aura.

• **Atonic seizures**: significant neurological abnormalities and brief loss of postural tone, often resulting in falls and injuries. The ictal EEG correlate is similar to abnormalities observed in tonic seizures.

• **Unclassified seizures**: defy any of the above classifications; this group of seizures represents a particular challenge for the epileptologist.

Figure 2-1 shows sample seizures from the ECoG data sets we used in our study.

Concise seizure classification is given in Table 2-1.

Although EEG recording is intended to capture cerebral activity, it unavoidably picks up any electrical activities arising from sites other than the brain. Recorded activity not of cerebral origin is termed “artifact” and can be divided into physiologic (body) and extra-physiologic (environment, equipment) [14]. Basically artifacts are classified as

• Muscle (electromyogram) activity
• Glossokinetic artifacts
• Eye activity, ECG artifacts
• Pulse
• Respiration artifacts
• Skin artifacts
• Electrodes problem
• Line noise
• Movements in environment

Detailed descriptions of each type are given in any EEG artifacts atlas reference. Figure 2-2 shows some sample plots of artifacts.

**Seizure Detection Algorithms**

Depending on the needs of diagnosis, the recorded EEG may come from scalp, epidural, subdural, or depth. Scalp EEG is more susceptible to artifacts than the remaining three; thus seizure detection algorithm based on it might have more false alarms because of artifacts. Furthermore waveform character is different for neonatal patients than for adult. Thus our review differentiates these aspects. We also examined methodology, the rationale behind the method (if available) and the test performance.

**From Adult EEG**

Prior et al. [15] used the observation that a typical tonic clonic seizure produces a sudden high level of cerebral activity as the seizure discharge occurs and during the period of post-ictal extinction of the electrical activity of the brain, the level of cerebral activity recorded falls abruptly, to build an cerebral function monitor (CFM). Recorded EEG passes through a filter with pass band 2 to 15 Hz, and the pass band signal was amplified with gain slope 12 dB/decade. Then the filter output went through a logarithmic amplitude compression and rectified before being written out to chart recorder. There were plots in the paper to illustrate that seizure were captured automatically, but no test result in terms of number were presented.
Figure 2-1. ECoG seizure samples. (a) Secondary generalized complex partial seizure. (b) Another secondary generalized complex partial seizure. (c) Complex partial seizure. (d) Complex partial seizure from different patient. (e) Complex partial seizure from another different patient. (f) Simple partial seizure.

Figure 2-2. Artifacts samples.
Table 2-1. Categorization of seizures

<table>
<thead>
<tr>
<th>Partial seizures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple partial seizures</td>
</tr>
<tr>
<td>1. With motor signs</td>
</tr>
<tr>
<td>2. With somatosensory or special sensory symptoms</td>
</tr>
<tr>
<td>3. With autonomic symptoms or signs</td>
</tr>
<tr>
<td>4. With psychic symptoms</td>
</tr>
<tr>
<td>Complex partial seizures</td>
</tr>
<tr>
<td>5. Simple partial onset, followed by impairment of consciousness</td>
</tr>
<tr>
<td>6. With impairment of consciousness at onset</td>
</tr>
<tr>
<td>Partial seizure evolving to secondarily generalized seizures</td>
</tr>
<tr>
<td>7. Simple partial seizure evolving to generalized seizures</td>
</tr>
<tr>
<td>8. Complex partial seizures evolving to generalized seizures</td>
</tr>
<tr>
<td>9. Simple partial seizure evolving to complex partial seizures evolving to generalized seizures</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Generalized (convulsive or non-convulsive) seizures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absence seizures</td>
</tr>
<tr>
<td>Myoclonic seizures</td>
</tr>
<tr>
<td>Clonic seizures</td>
</tr>
<tr>
<td>Tonic seizures</td>
</tr>
<tr>
<td>Tonic-clonic seizures</td>
</tr>
<tr>
<td>Atonic seizures*(Combinations of the above may occur)*</td>
</tr>
</tbody>
</table>

| Unclassified epileptic seizures                       |

Ives et al. [16] built telemetry and computer analysis for long term petit mal absence seizures. First the telemetry system recorded and stored EEG on analog tape. Then the recording played back and digitized in 100 Hz, analyzed in 2.5 seconds periods. The amplitudes whose value was between 2 to 3 times of the average background activity were searched first in the observation period. If more than 10% of the points fell into this range then the previous 5 seconds would be output as seizure discharge, and these output discharges would be analyzed by technician. The paper presented on how much time the automatic system could save. Ives et al. [17] applied depth electrode to record the EEG signal, and the computer analysis, used two criteria to automatically select the recording section with the seizure discharge. If the output of a band pass filter exceeded a level (patient and seizure pattern dependent) for a certain length of time, and/or if the activity...
of the recording channels fell within an amplitude window for certain period of time, then computer would record the 2.5 seconds prior and continue the recording till the discharge stops. Test result showed that during 2 weeks, 13 seizures were captured, among them only 4 were noted by nurse and patients.

Babb et al. [11] built an electronic circuit to detect seizure from depth electrodes recording. The detector utilized the phenomenon that in-seizure EEG has higher amplitude concurrent with higher frequency; was composed of two main stages and three auxiliary stages. First stage converted pulse rate to voltage level, second stage checked the duration of certain level voltage from the 1st stage (intended to detect seizures that lasted 5 seconds or more) to trigger alarm. The auxiliary stages were multiple pulses preventing at the detector output, alarm latch light waiting for manual reset and pulse generator for signaling attention. The detector was tested on four patients, generated a total of 64 seizure alarms, 18 of them were false alarms (most due to equipment), 20 of the detected seizures were not noticed by nurse. The paper concluded that machine detector was superior to human in terms of detection rate and inferior in terms of false alarm.

Principe and Smith [18] designed a microcomputer based real-time petit mal seizure detection system. Their system captured the rhythmic spike and wave complex pattern that was often observed in petit mal seizure. In their system two band pass filters (0.8 to 6 Hz and 10 to 25 Hz) were used to extract the slow wave and spike characters presented in the signal, then the period of the individual components of the filtered data was measured. Unlike former methods that only used amplitude, for the slow wave they had set up a set of waveform description rules (peak to valley half period, maximum time
between zero crossing, peak to peak amplitude) based on a training set of petit mal seizures. The detected spike and slow waves combination were feed to a petit mal pattern recognizer to finally decide whether a petit mal seizure was captured or not. The system was first tested on one frontal channel 70 hours of scalp EEG from 6 patients then tested on a 6 hour recording taken one month later. They found that the system performs well in clinically significant epochs, and they had tabled and analyzed the seizure detection performance in detail.

Gotman [19] proposed a method to detect seizures from both scalp and intracranial EEG. The detection procedure was composed by 5 steps: (1) Sampling frequency 200 Hz, notch filter (3 tap) eliminating 60 Hz interference (preprocessing). (2) Decompose in half-waves according to method presented in [20], [21]. (3) Character of one epoch (2 seconds), the ratio of average amplitude of half waves vs. background, the average duration, the variation of duration, the coefficient of duration variation vs. the duration mean (measure the rhythm). (4) Background average half wave amplitude (each 16 seconds window which was 12 seconds before the present epoch). (5) Detection criteria, channel detect: relative amplitude equal or greater than 3, average duration 3 to 20 Hz, coefficient of variation less than 0.6; seizure detect: at least two channel detect or adjacent two epoch detect in one channel. The method had been tested on 16 patients with 24 scalp EEG average length 12.4 hours each, the seizure detect rate is 22%, paroxysmal burst 58%, artifacts 20%; and 4 patient with 44 intracranial EEG average length 18.7 each, detect rate 2.5 %, paroxysmal burst 79%, artifacts 18.5 %. Later on the method was improved by adding more waveform description parameters and tested on a much larger data set in [22], then independently tested in [23]. Further improvement was
proposed by Qu and Gotman [24] to reduce false detection. The improvement targeted those false detections that caused by patient specific artifacts. Their method took initial known artifacts features as template and created a distance measure between later extracted feature and the template; if similar the new alarm would be rejected as false alarm. And the method was tested again independently by [25]. In summary the Gotman method and its derivatives has been commercialized and widely used worldwide after its introduction.

Murro et al. [26] developed a method to detect complex partial seizures from depth EEG. They extracted three features, relative amplitude, dominant frequency and rhythmicity from two channels of training epochs (both seizure and background). Then in-seizure and in-background probability of a test epoch was determined by the density of nearest neighboring to the known seizure epochs and known background epochs, their discriminant function was the ratio between these two probabilities. A threshold was finally applied to decide whether the test epoch was in-seizure or in-background. The method was tested on eight sets of data they showed that the detection rate was in the range of 90 to100% with false positive in the range of 1.5 to 2.5 per hour.

Harding [25] detected seizures from intracranial EEG using an epoch wave pattern descriptor. The ECoG signal was digitized with 141 Hz frequency; took 5 seconds epoch of samples, built histogram for the sample-to-sample difference (MDV) and time difference (DT) between MDV. And these two extracted parameters combined with signal to background ratio and adaptive threshold were used to judge whether a seizure was ushered in or dying out. The method had been tested on 40 patients and achieved average clinical seizure detection rate 97% and sub-clinical seizure detection rate 93%.
Pietilä et al. [26] classified the scalp EEG wave into 370 types and expressed each type in the form of feature vector with 13 components. Then they used a function that considered average magnitude and differences between consecutive samples to segment EEG into variable length, extract feature vector from each segment, apply Euclidian distance to check the closeness between extracted feature and the class models. The seizure type was the spike-slow-wave burst. Based on a test set of 6 patients, twelve 30 minutes EEG segments test their sensitivity was 97%. And the reference also compared their method with Gotman’s method.

Webber et al. [29] built a three layer feed forward (node number 31-30-8) artificial neural network (ANN) to segment scalp EEG recordings into eight types: small seizure (<300uV), large seizure (>300uV), normal, theta, alpha, muscle, chewing and noise. The segmentation was achieved through following steps: training samples of the above eight types EEG were collected, then 31 features of each type was extracted for each 2 seconds epoch, these features are presented to the ANN. After the ANN was well trained, the test data features were then discriminated by ANN epoch by epoch. They have tested the method on 78 files (34 seizure file, 44 other files) randomly selected from 50 patients, the selectivity was 87% and the sensitivity 76%. Klatchko et al. [30] enhanced the performance of this system by using a clustering algorithm to reduce false detection. A so-called generic detector (where the detection algorithm operating on a single channel one epoch at a time) was attached it to the ANN detector. The generic detector is a clustering algorithm, while took the output from the ANN detector for each channel as its input. In stead of making the detection decision epoch by epoch, the clustering algorithm searching for channel relations along the direction where epochs proceeded. The final
detection was achieved by setting a threshold to the in-seizure probability, duration and the number of active channels of the interested clusters. It was claimed that the method reduced the false alarm up to 50%.

Gabor et al. [31] applied another form of neural network, the self organizing map, to detect seizures from scalp EEG. In the training stage, they extracted a scale index using 10-scale wavelet filters from a set of in-seizure training samples, the index and wavelet filter were the preprocessor for the testing data; then they applied FFT to extract 2D spectrograms (time, frequency) for every 4 seconds epoch (512 samples), {procedure: in each epoch use 50% overlapping window and FFT(every 250ms/32samples), the 2D spectrogram of the epoch was generated by arranging the small FFT window in time frequency dimension}, the 2D spectrogram was transformed into one dimensional feature vector. All the feature vectors from in-seizure training examples were feed to train a one dimensional 6 nodes SOM. In the testing phase, EEG was preprocessed by the processor extracted from the training data, then grouped the channels according to user’s wishes/intuitions, and the EEG amplitude was averaged in one group, each epoch of the average time series (4 seconds) went through the same 2D spectrogram feature extraction and the extracted feature was then compare with the trained SOM, the logarithm of the minimum Euclidian distance of each cluster was calculated, and finally a threshold was applied to the log-distance to decide whether the signal was in-seizure or in-background. They tested the method on 529 hours of data from 22 patients, achieved average detection rate 90% with false detection 0.71 per hour. In [32] the method was further evaluated on 4554 hours of data from 65 patients, the average detection rate was 92.8% with mean false detection 1.35 per hour. The detection performance was also compared with audio-
transformation (Medilog 9200) method and seizure monitor (Stellate system Ver.8, the Gotman method), and the SOM based neural network method was claimed to outperform both.

Shindler et al. [33] developed a method based on simulated leaky integrate and fire units (LIFU) to detect seizures recorded in foramen ovale and scalp EEG. The detection scheme had five stages. (1) multiple channels of contra-lateral time recordings were averaged. (2) each channel recording was then compared with the average; then the difference feed to next stage. (3) the absolute value of the derivative from last stage, was taken and compared with a threshold to yield 1 if exceed threshold otherwise 0 was recorded (convert continuous signal into discrete pulse train). (4) The pulses train from each channel was then feed into a parallel RC integration circuit (simulate the neuronal activity), the output of each integrators (ipsi-lateral side) were added together (to simulate soma), the summation was compared with a threshold, if exceed a unit spike was output to the next stage and a delayed feed back to integrators of each channel to reset the integrator. (5) The spiking rate ratio between right and left of the recording was calculated by Eq.2-1

\[
SRR = \frac{SR_R - SR_L}{[SR_R + SR_L]} \tag{2-1}
\]

\(SR_R\) and \(SR_L\) were the mean spiking rate of the right and left recording side. The \(SRs\) and \(SRR\) were used as a measure of seizure detection. They tested 22 short sections (around 10 minutes each) and 4 long sections (around 18 hours each) foramen ovale and scalp EEG from 10 drug resistant partial epilepsy patients with a total 36 of seizures in the recording, and found that the method correctly detected all seizure without false alarms.
Khan and Gotman [34] developed a wavelet assisted seizure detection algorithm to detect seizures from intracerebral EEG recording. Based on the observation that the intraseizure EEG rhythmic was concentrated in the 3 to 25 Hz (their data sampling rate was 200 Hz), their method first decomposed a 5 seconds epoch into 5 scales, then the following features were extracted for the 3 to 5th scale: (1) the relative energy between the average energy of each scale vs. the summation of all scale; (2) extract coefficient of variation for each scale using the wave segmentation method from [19], (3) the relative amplitude of present epoch vs. background. Then they set a threshold for the three features, the epochs that exceed the three thresholds were regarded as preliminary detections. After preliminary detection their method set up heuristic rules to eliminate false over-threshold epochs caused by paroxysmal rhythmic discharges and burst of alpha, theta activities. They had tested on 229 hours of data contained 66 seizures from 11 patients. The false alarm dropped from 2.4 per hour using method of [21] to 0.3 per hour with the same detection rate of 90%.

From Neonatal EEG

Liu et al. [33] applied autocorrelation technique to detect neonatal seizures. The test was based on the fact that if signal is periodic then the corresponding autocorrelation should also be periodic and the observation that neonatal seizures appeared periodic in different frequency. The method sampled the scalp EEG recordings with 40 Hz and sequenced each recording channel with length of 30 seconds. The 30 seconds were divided into 5 windows (6.4 seconds, i.e. 256 samples), and the autocorrelation of each window was calculated through FFT, then they exam the periodicity of the autocorrelation by calculating the ratio between peak moment center and the primary period. This ratio was rounded to the nearest integer, and the number, that the ratio
needed to be rounded to, was then scored by a nonlinear scale. This scale value was used to decide whether the epoch contained seizures. They compared the method with power spectral density (PSD) method developed by Aziz et al. [34] using Burg Fourier analysis, and showed that autocorrelation revealed the periodicity seizure pattern better that PSD.

Gotman et al. [37], [38] pointed out that the autocorrelation method tested only on selected spectrum of neonatal seizure type, to overcome the description deficiency a more comprehensive method was proposed to cover more seizure variety. Their method has sub-methods: (1) Spectral analysis to find the rhythmic discharges at various frequencies (FFT of 10.24 seconds 2048 samples then find dominant frequency, width of dominant spectral peaks, power ratio, stability of current epoch, 60 Hz indicator, patient disconnect indicator). (2) Spike detection to find groups of abnormal spikes that were not rhythmic. (3) Low-pass digital filtering to find slow discharge with less or equal to 0.5 Hz frequency. The method had been tested on 55 newborn data in [37] and 54 in [38] respectively. And the detection results were similar to those commonly used method for adult seizure monitoring.

In stead of extracting features from recorded EEG, Roessgen et al. [39] used EEG modeling technique to detect seizures. In [39] the model was described by 11 model parameters, based on these parameters the EEG signal could be decomposed into background spectrum and seizure spectrum. The corresponding parameters of each EEG epoch were estimated by Whittle’s approximation (a simple form of maximum likelihood estimator). And the final detection was by setting a threshold to the ratio between summation of estimated seizure spectrum and summation of background spectrum. The
method was tested on two channels of two baby data with total length of 137 minutes, and they claimed that the result was better than quadratic detection filters.

Celka et al. [40] proposed singular spectrum analysis and Rissanen minimum description length (SSA-MDL) scheme to detect seizure. The method assumed that EEG signal was composed of background generated by Gaussian noise and seizure rhythmic activity. They first preprocessed EEG to attenuate the background. After preprocessing the data were segmented into 10 seconds (overlapping 8.75 seconds) windows, then each window was embedded into a high dimensional matrix, SVD was applied to the matrix to extract eigenvalues for the window (SSA). They calculated the minimum description length (MDL) from the eigenvalues, the windows with MDL greater than 1 were flagged, and the final in-seizure flags were smoothed by a median filter to eliminate those isolated flags. They had compared the method with both Gotman method and Liu’s method. The comparison was detailed in tables.

Boashash et al. [41] suggested decomposing the time domain EEG into time-frequency and they applied a time-frequency matched detector to find seizures. The method was tested on model simulated EEG data only.

Altenburg et al. [43] applied synchronization likelihood method from [42] to detect seizures. The method took $M$ simultaneously recorded time series $x_{k,i}$ where $k$ denoted the channel number and $i$ discrete time index, $X_{k,i}$ was the embedded vector of the corresponding time series. Then took probability Eq.2-2 of the two vectors near to each other within $\varepsilon$ for the channel $k$, critical $\varepsilon_{k,i}$ is determined by $P_{k,i}^{\varepsilon_{k,i}} = P_{ref}$ and $P_{ref} << 1$.

$$P_{k,i}^{\varepsilon_{k,i}} \equiv \frac{1}{2(w_2 - w_1)} \sum_{w_1 < |i-j| < w_2} \theta(\varepsilon - |X_{k,i} - X_{k,j}|)$$  \hspace{1cm} (2-2)
Then defined another closeness variable among channels by Eq.2-3

\[ H_{i,j} = \sum_{k=1}^{M} \theta(e_{k,i} - |X_{k,i} - X_{k,j}|) \]  

(2 – 3)

The synchronization likelihood for the \( X_i \) and \( X_j \) pair is defined by Eq.2-4.

\[ S_{k,i,j} = \begin{cases} \frac{H_{i,j} - 1}{M - 1} & \text{if } |X_{k,i} - X_{k,j}| < e_{k,i} \\ 0 & \text{if } |X_{k,i} - X_{k,j}| \geq e_{k,i} \end{cases} \]  

(2 – 4)

And finally took the average over a window by Eq.2-5.

\[ S_{k,i} = \frac{1}{2(w_2 - w_1)} \sum_{j=1}^{N} S_{k,i,j} \]  

(2 – 5)

This measure described how strongly channel \( k \) at time \( i \) was synchronized to all the other \( M-1 \) channels. They had tested on 42 sections (length 21 seconds each, half seizure sections and half normal sections) and 2 complete EEG from 22 patients all with 12 bipolar recording channels. When applied certain threshold to the synchronization measure they could detect 85% of seizures.

**Seizure Prediction Algorithms**

The earliest attempt to predict seizure was in 1975. Viglione et al. [44] use EEG spectra features to predict seizures. Their system was abandoned later because of too many false predictions. Compared with seizure detection, relatively less research has been done using linear signal processing method on seizure prediction since then.

**Seizure Prediction through Linear and Nonlinear Signal Processing**

Rogowski et al. [45] applied autoregressive modeling technique to track the pattern of the filter pole before seizure occurs. They used an order 16 predictor on a 1.7 seconds window, and Wiener-Hopf algorithm to generate the filter coefficients, then calculated
the poles for the corresponding window. The window was moved 0.1 seconds forward, the process repeated, and the trace of poles vs. time was plotted. Test on 28 seizure sections from 12 patient showed that the trace of the most mobile poles might be used to predict seizure, as they found 25 seizures out of 10 patients had obvious change in terms of pole location. They explained the change might be due to underlying neuronal parameter change.

Katz et al. [46] tracked the spike changes before seizure using the automatic spike detector from depth EEG for 10 patients. They had monitored 0-5, 5-10 and 0-60 minutes epoch before 45 seizures, by comparing the repeated measures 1-way ANOVA, they found the total spikes in all recording channels did not differ among the three pre-ictal epochs.

Duckrow et al. [47] studied the neuronal synchronization pattern during seizure onset. They took FFT of 6.4 seconds (half overlapping) window of depth EEG, and averaged every 4 windows. Then the phase cross spectrum and coherence value was calculated for all possible pairs of contacts (for the frequency range 1.17 to 48.44 Hz, 0.39 Hz increment). They found that on the way to generate seizures there were transient interaction between brain regions.

Alarcon et al. [48] studied the power spectrum and intracranial EEG patterns of 78 complex partial seizures from 15 patients. They took FFT for each channel use window size 1.28 seconds; then extract (a) the amplitude summation of 12 different bands between 2 to 100 Hz (with different resolution), (b) activity, mobility and complexity in 2 to 100 and 2 to 40 Hz band respectively. Test result showed that the time track of these variables changes before seizures. They claimed that the method was effective for
detecting low-amplitude high-frequency activity, the kind of waveforms often time could be observed before seizure occur.

Qu and Gotman [49] developed a patient specific seizure warning system. The system took a group of seizure onset patterns and some background data to train a classifier. The classifier was composed by a cloud of 5 dimensional points. The 5 dimension were average wave amplitude, average wave duration, coefficient of variation of wave duration, dominant frequency and average power in a main energy zone respectively. In the testing session the same features were extracted for each 2.56 seconds window (512 samples) and the feature was compared with the classifier by weighted Euclidean distance. Those channels which had features close to seizure onset template point within a fixed defined range were considered predictive of an impending seizure, and the final warning announcement also considered the spatial in-seizure characteristic. The method was tested on 12 sets of data from 12 patients with total of 47 seizures, they achieved 100% onset detection rate with only 0.02 false alarms per hour. And the average of the pre-seizure warning is 9.35 seconds.

Osorio et al. [46] described a two-step filtering algorithm to detect seizure from ECoG. The first filter was a 22 orders FIR filter generated from 3-level Daubichies 4 wavelet, it intended to extract 5 to 45 Hz signal components from ECoG. The second was a median filter. They median filtered every 2 seconds (480 samples) energy signal epoch from the first filter output to obtain the foreground (FG) sequence, then built background (BG) sequence from the FG sequence through decimating FG, exponential decay and another median filtering. The feature they used to announce seizure was the ratio $r$ between FG and BG in each epoch, a threshold $T$ was applied to the summation of $r$ from
all recording channels. They had tested on 55 hours of 10 minutes concatenated data from 16 subjects (total 125 partial seizures), when $T=20$ the method achieve 100% detection with 0.36 false alarm per hour. And also 92% of the seizures could be predicted by mean of 15.5 seconds ahead. In [51] the method was further tested on consecutive recording data from 15 subjects (8 channels each). Meng et al. [52] modified the algorithm by applying Gaussian mixture model to the ECoG features. The revised algorithm use the same 3-level Daubechies 4 wavelet filter to decompose ECoG into 4 scales, then median filter the energy of each scale (2 seconds window). Twenty four dimensional features were extracted from the signals that came out of the median filter, these dimensions were: (a). the energy of each sub-band and the total energy of all sub-band (5 features), capture the energy change. (b). the gradients (25 seconds of data prior current point) of each sub-band and the total energy (5 features), capture large changes in short-term. (c). the FG to BG ratio of each sub-band and the total energy (5 features), capture change intensity at short timescales compared to long-term trend. (d). the normalized FG deviation from BG (5 features) to detect change volatility and reduce highly unstable background caused false alarms. (e). fraction of total energy found in each sub-band (4 features) to capture those seizures with rhythmicity change but no magnitude change. In the training phase, the 24 dimensional features extracted from seizure and background were used to train two Gaussian Mixture Model (GMM) using Expectation Maximization (EM) algorithm, while in the testing phase new features extracted from test data were compared with the models to check the likelihood of its in-seizure or in-background tendency. Test result showed that GMM method improved speed of detection over the previous filtering method with similar performance.
Petrosian et al. [53] developed a recurrent neural network based seizure prediction method. They used wavelets to decompose the EEG recording into high pass and low pass sub-band, the decomposed signal as input to train and test a recurrent neural network (extended Kalman filtering algorithm was used to train the Recurrent Neural Network (RNN)). Four pre-seizure sections and 3 normal sections were tested; they claimed that there was a pre-ictal state some minutes before seizure occur.

Litt et al. [54] simply tracked accumulated energy of intracranial EEG recording for 5 patients (each recording was in the range of 3 to 14 day). They found that hours before seizure the accumulated energy began to increase, and deducted that quantitative measure of pre-seizure electrical activity could possibly be used to predict seizure.

Schindler et al. [55] tested a parameter modified simulated leaky integrate and fire unit (details are described in [33]) for seizure prediction. They tested 9 long term (average 16 hours each) scalp and foramen ovale EEG from 7 patients, and the method could predict on average 83 minutes before seizure occurs and had 2 false predictions during the whole testing.

Niederhauser et al. [56] applied sign periodogram transform to capture the low amplitude rhythmic activity before seizure. They took the derivative of EEG time series \( x[n] = x_{EEG}[n] - x_{EEG}[n-1] \), then the sign of the derivative Eq.2-6.

\[
s[n] = \text{sign}(x[n]) = \begin{cases} 
1, & \text{if } x[n] \geq 0 \\
-1, & \text{if } x[n] < 0 
\end{cases} \tag{2 - 6}
\]

Finally N-point FFT of \( s[n] \) was taken by Eq.2-7.

\[
Jm[k] = \left| \frac{1}{N} \sum_{n=0}^{N-1} s[n + m - N]e^{-j\frac{2\pi kn}{N}} \right|^2 \tag{2 - 7}
\]
Where $m = 0, M, 2M, \ldots$, $M$ was the sliding window displacement and $N$ was the window length. The prediction announcement was based on the value of a median filter output $D[m]=\text{median}(J_{f_{lo}}^{f_{hi}}[m-i])_{i=0,M,2M \ldots (O-1)M}$, where $J_{f_{lo}}^{f_{hi}} = \max(J_{m}[k])_{k=f_{lo}, \ldots, f_{hi}}$ and $f_{lo}, f_{hi}$ defined the frequency range. They tested on 2 to 6 days depth EEG recording from 10 patients found that 5 patients had the low amplitude rhythmic activity (5 to 80 seconds) before seizure and the other 5 did not have the pattern.

Shoeb et al. [57] proposed a patient specific method to detect seizure onset from scalp EEG. Their method decomposed an epoch (2 seconds, their sampling frequency was 256 Hz) of recording into 4 scales using wavelets. The logarithm of the energy in each scale was regarded as a morphology descriptor, and these descriptors from all channels were arranged sequentially into a feature vector. First they chose a training set (both in-seizure and background) from each patient, used the feature vectors extracted above to train a support vector machine classifier. Then in the testing phase feature from new testing epoch was classified by the classifier. The method was tested on 36 pediatric patients 60 hours of data contained 139 seizures, detection rate 94% with 0.25 false alarms per hour, with predictability of on average 8 seconds before seizure onset.

**Seizure Prediction through Nonlinear Dynamic Theory**

The advent of physical-mathematical theory of nonlinear dynamics in the early 1980s made new techniques available for researcher to analyze the irregular behavior of EEG. Apply these techniques has produced large body of evidence that EEG is not only composed by interictal and ictal period, there is a preictal period exist, this means seizure prediction is feasible [58], although there are debates on whether the results justify the algorithmic complexity [59] [60].
In nonlinear dynamic theory an observed time series is regarded as generated by multiple unknown variables of the system, and the behavior of the variables (thus the system) can be revealed through embedding this time series. The following review tracks researches that have been done by groups and individuals around the world, emphasis on the methods, and test results. Most of seizure prediction results were based on intracranial EEG study, exceptions will be pointed out.

Iasemidis et al. [61] introduced the idea of using nonlinear dynamical parameters to characterize intracranial EEG. The largest Lyapunov exponent (L_{max}) [62] was extracted from window-wise ECoG (Wolf’s method with practical implementation revision). They found that as seizure approach the L_{max} drop. Through observing the evolution of L_{max} they found that besides inter-ictal and ictal states, a pre-ictal state also exists [63]. The spatio-temporal dynamics was described in [64] by T-index defined by Eq.2-8

\[
T_{i,j}(t) = \frac{E\{|L_{max,i}(t) - L_{max,j}(t)|\}}{\frac{\sigma_{i,j}}{\sqrt{N}}}
\]

(2-8)

Where N was the moving window length and \(\sigma_{i,j}\) was the standard deviation of L_{max} difference between channel i and channel j. Reference [65] summarized the expansion of this group’s work. They applied optimization theory to select the most entrained electrodes based on T-index. Tested on 5 patients (58 seizures contained in 266.2 hours of ECoG recording) data, they found over 90% of the seizures could be predicted 19.8 minutes to 42.9 minutes in advance.

Martinerie et al. [67] applied nonlinear measure to predict 19 seizures from 11 patients. Their phase space was constructed through the time series amplitude embedding. In the phase space, the correlation density \(D\) was extracted for each sliding
window, they used surrogate data technique to prove that there was deterministic nonlinear process under the observed signals. And further more they assume that at the cellular level there was a gradual increase in excitability which caused reduction in threshold for seizure onset force $R$, the force is induced in part by recruitment of distant neurons in an epileptic-prone neural tissue. The force itself was unobservable, but the seizure route plot generated by the change of correlation density from window to window provided a representation of $R$. They plot the $D$ transition in a two dimension plots ($D_i$ versus $D_{i+1}$ axis). When in background state, the $D$ gingering around in a smaller region, when seizure began to form, the $D$ escaped the region, if this leaving persisted for certain amount time they predicted that seizure was on the way. Test result showed that 17 of 19 were correctly predicted between 2 to 6 minutes beforehand. The authors later found that their method was sensitive to amplitude variance fluctuation which was an unwanted character. To overcome the amplitude influence Quyen et al. [68] proposed a similarity measure to quantify the dynamic closeness between pairs of windows. In stead of using amplitudes of the time series, they embedded the time interval between positive going threshold crossings to construct the phase space. They first build a 16 dimensional background reference trajectory, then applied Singular Value Decomposition (SVD) to find lower dimensional projection base $V$ (the eigenvectors corresponds to larger singular values) and the projected reference into $Y(\text{ref})$. In the test phase the phase trajectory generated using the same embedding technique for the test window was projected onto $V$ to get $X(\text{test})$. Correlation integral $C(Y(\text{ref}) X(\text{test}))$ was calculated and cross-correlation ratio between $C(Y(\text{ref}) X(\text{test}))$ and square root of $C(Y(\text{ref}) Y(\text{ref}))C(X(\text{test}) X(\text{test}))$ was the final decision making features. They had tested the method on 23 recording (40
minutes each) from 13 patients, all of the seizures were able to be predicted, but the prediction advance time varies even for the same patient. They concluded that seizure emergence was a complex non-repetitive process. This same method was applied in reference [69] to study the spatial extension of preictal transition compared with the epileptogenic zone on a homogeneous group of patients (with medial temporal lobe epilepsy), they found that the spatial distribution of preictal changes often involves an extended network projecting beyond the limits of the epileptogenic region; they also applied the method on scalp EEG in [70] and neocortical partial epilepsy patients in [71]. The group found that their method has considerable therapeutic implications on the possibility of anticipating the onset of seizures. Reference [72] summarizes the research of this group.

Lehnertz et al. [73] highlighted nonlinear analysis techniques developed in their group. To characterize the pre-seizure features they had developed two measures, one univariate and one bivariate. The univariate measure named neuronal complexity loss $L^*$ was introduced in [74]. To generate $L^*$ they extracted correlation integral from window-wise ECoG then calculated the correlation dimension $D_2$. The $L^*$ was defined as the summation of ratios between $D_2$ and a preset threshold $D_u$ over a span of observation time. The measure was tested on data (98 sections of interictal, 28 sections of pre and postictal, each with around 20 minutes in length) collected from 20 patients with unilateral temporal lobe epilepsy, they found that $L^*$ decrease (loss) in the epileptogenic area. The method was further tested in reference [75]-[79]. To investigate the spatial interaction between channels, bivariate measure was also investigated. In [80] mean
phase coherence defined by Eq.2-9 was employed to characterize the phase synchronization between channels.

\[
R \equiv \frac{1}{N} \sum_{j=0}^{N-1} e^{i\phi_{1,1}(j\Delta t)}
\]  

(2 – 9)

Where \(\phi_{1,1}(t)\) was the instantaneous phase difference between two signals, \(\Delta t\) was the sampling period, \(N\) was the number of samples over which the average was taken, the instantaneous phase was obtained through Hilbert transform, they found that \(R\) began to drop before seizure occur (explained as synchronization decrease). In [81] \(R\) was window-wise extracted as feature for 117 hours of data from 18 patients (contain 32 seizures), based on the depth of \(R\) drops and duration relative to baseline, the final decision was made on whether a seizure was approaching. They claimed that 26 out of 32 were properly predicted. In [82] maximum linear cross correlation \(C_{\text{max}} = \max \{C(s_a, s_b)(\tau)\}\) was added to characterize lagged synchronization, where \(C(s_a, s_b)(\tau)\) was defined by Eq.2-10 and 2-11.

\[
C(s_a, s_b)(\tau) = \frac{\text{corr}(s_a, s_b)(\tau)}{\sqrt{\text{corr}(s_a, s_a)(0) \ast \text{corr}(s_b, s_b)(0)}}
\]  

(2 – 10)

\[
\text{corr}(s_a, s_b)(\tau) = \int_{-\infty}^{+\infty} s_a(t) \ast s_b(t + \tau) dt
\]  

(2 – 11)

The \(R\) and \(C_{\text{max}}\) were combined to predict seizure automatically. They tested the method on data from 10 temporal lobe epilepsy patients (35 sections of interictal EEG average 25 minutes each, 14 preictal sections average 50 minutes each), had correctly predicted 12 out of 14 seizures.

Hively et al. [83] developed two measures to characterize changes in nonlinear time series. Their embedded high dimension phase space was divided into hypercube, the
ratio between number of points in each cube and the total number of points in a window was regarded as the discrete data distribution. They chose a base case as a reference and the test data was compared with the base by two measures namely $L_i$ distance and $\chi^2$ statistics.

$$L = \sum_i |Q_i - R_i| \quad \text{and} \quad \chi^2 = \sum_i \frac{(Q_i - R_i)^2}{Q_i + R_i} \quad \text{Eq. 2-12}$$

$Q$ and $R$ were discrete distributions of reference and test respectively. Similarly the transition distributions between cubes were also quantified using the same measures to capture the underlying dynamics. The method was applied on short sections (0.31 to 0.86 hours) of scalp EEG from 9 patients and longer sections (0.61 to 5.5 hours) of scalp EEG from 11 patients with different sampling rate [84], each section contained a seizure. They found that the measure were superior to traditional measures (mutual information, correlation dimension and Kolmogorov entropy) on detecting the condition change. In reference [85] they further analyzed 41 sections of scalp EEG data (length in the range between 1.5 hours and 8.25 hours) from 41 patients (each contain at least on seizure event), and used additional 20 sets of data without seizure as reference. It was found that the measures have potential to forewarn seizures. In [86] the seizure prediction criteria was fine tuned based on the extracted $L$ and $\chi^2$ feature to improve the forewarning consistency.

Li et al. [87] and Drury et al. [88] applied a nonlinear measure called marginal predictability $\delta_d$ to scalp EEG. The nonlinear measure was based on the findings of [89]-[91]. Its extraction started from correlation integral of a windows of EEG recording. The correlation integral Eq.2-13 of $d$ dimensional vectors can be interpreted as the probability of the norm between each pair vectors less than $\varepsilon$. 

$$\sum_{i} P(||x_i - x_j|| < \varepsilon)$$

$$P(x) = \frac{1}{2\pi \sigma^2} e^{-\frac{x^2}{2\sigma^2}}$$
Then they defined predictability Eq.2-14.

\[ \delta_d \equiv \frac{R_d - 1}{R_d} \]  
(2 – 15)

\[ R_d = \frac{S_d}{S_{d-1}} = \frac{C_{d+1}C_{d-1}}{C_d^2} \]  
(2 – 16)

Where \( \delta_d \) was interpreted as the marginal predictability, they claimed that their measure has both time and spatial seizure predictability. It was not so clear how many data they had tested.

D’Alessandro et al. [92] developed a multiple feature seizure prediction system. They divided the data into training and testing set (both contain portion of pre-ictal data and baseline data). The training set was used to train a classifier. The major training process was feature selection. The selection was composed by two parts: (1) Three level of feature extraction, for each window of EEG recording, the first level has (a) curve length (equivalent to fractal dimension); (b) average energy; (c) average nonlinear energy (Teager’s algorithm); (d) spectral entropy; (e) the sixth power; (f) energy of the wavelet packets; the second and third level features were derived from the class conditioned pdf for those visually observed seizure focus channels of the first level features, include minimum, maximum, median, mean, variance, standard deviation, skewness, kurtosis, slope, integral, derivative, sum. These features were feed to next step. (2) Genetic algorithm and probabilistic neural network classifier for feature selection. The test set
was checked after the prediction system was properly trained. The experiment was conducted on 46 preictal records and 160 hours of baseline (70% used on training and 30% for testing) from 4 patients they found that 62.5% of seizures were properly predicted with 0.27 false predictions per hour.

Moser et al. [93] applied Lyapunov theory to intracranial (invasive and semi-invasive) and scalp EEG data collected from three persons (two epilepsy patients and one normal). They extracted Lyapunov exponents (all the available dimensions using Eckmann’s method) from windowed EEG, then they observed the spectra of each exponents of certain interested window, the Kolmogorov entropy and dissipation vs. the time was plotted for all the observation windows. They found that the spectra density character was well suited to classify the EEG states (ictal, interictal, preictal).

**Dealing with Artifacts**

Any study based on EEG has to face artifacts. There are three approaches to deal with artifacts: (1) avoid, (2) eliminate, or (3) minimize [94]. Avoid can only be done before signal recording, and the constraint it imposed on the recording can hardly be applied on epileptic seizure research. To automatically eliminate artifacts, methods are needed to detect artifacts first. Rohalova et al. [95] had applied auto-regressive Kalman filtering and extended Kalman filtering radial bases feed forward neural network to detect artifact sections. Brunner et al. [96] extracted FFT feature of each 4 seconds window of EEG and use adaptive background and median filtering to detect muscle artifacts section. These methods might be applicable to research like sleep studies, as the signal is roughly classified into two types: artifacts section and normal section (eliminate or not); adjustment is needed for seizure detection to accommodate the additional in-seizure section which might present in EEG recording.
To minimize artifacts, most research has taken “divide and conquer” approach until recently. The common practice to line noise is notch filtering, and the most studied artifact is EOG, in the sense that fewer papers can be found that is dedicated to other types of artifacts [97], except the work by Barlow and Dubinsky who had subtracted the EKG artifacts [97] and Larsen et al. applied auto-regressive filtering to smooth out EKG (order 1) and muscle artifacts (order 7) [98] [99].

For EOG artifacts, the very early attempt was in 70s, when Girton and Kamiya [100] designed an analog comparator circuit to minimize EOG interference in EEG recording, but they discarded lots of data.

References [101], [102] and [103] began to use time domain regression technique. The regression function calculated $B$ the propagation coefficient [104], the proportion of one variable that is explained by another. The estimation of $B$ was as following: $X_i$ the EOG recording, $Y_i$ the EEG recording at time $i$, the $Y_i = B \cdot X_i + EEG_{true_i}$. In [101] and [102] $B$ is estimate by Eq.2-17.

$$\hat{B} = \frac{Y_{t_{\text{max}}} - Y_{t_{\text{began}}}}{X_{t_{\text{max}}} - X_{t_{\text{began}}}}$$ (2-17)

Where $t_{\text{max}}$ and $t_{\text{began}}$ were the time of EOG at maximum value and EOG started to occur), and [103] the estimation of $B$ by Eq.2-18

$$\hat{B} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = r_{xy} \cdot \frac{sd_y}{sd_x}$$ (2-18)

Where $r_{xy}$ was the cross correlation coefficient between EOG and EEG, the correction was done through Eq.2-19.

$$EEG_{true_i} = Y_i - (\hat{B} \cdot X_i) - (\bar{X_i} - \bar{Y_i} \cdot \hat{B})$$ (2-19)
If one more EOG reference recording $Z_i$ was available then they use Eq.2-20, 2-21.

$$
\tilde{B}_{yx,z} = \frac{(r_{yx} - r_{yz} \ast r_{zx})}{(1 - r_{yx}^2)} \ast \frac{sd_y}{sd_x}
$$

$$
\tilde{B}_{yz,x} = \frac{(r_{yz} - r_{yx} \ast r_{zx})}{(1 - r_{yz}^2)} \ast \frac{sd_y}{sd_z}
$$

And the true signal was estimated by Eq.2-22.

$$
\text{EEG}_{\text{true}i} = Y_i - (\tilde{B}_{yx,z} \ast X_i) - (\tilde{B}_{yz,x} \ast Z_i) - (\overline{X}_i - \overline{Y}_i \ast \tilde{B}_{yx,z}) - (\overline{Z}_i - \overline{Y}_i \ast \tilde{B}_{yz,x})
$$

The same concept applied if multiple reference EOG channels were available.

Whitton et al. [105] developed a frequency domain regression method to correct EOG artifacts. They FFT every 4 seconds of EEG and EOG recording, after filtering and scaling the spectral of the two windows, the EOG spectral was subtracted from EEG spectral, subtraction in frequency domain is equivalent of filtering in time domain, the filter property was determined by EOG. Woestenburg [106] used regression analysis in frequency domain also, in their method Eq.2-23 relation was assumed.

$$
V_{eog_i}(t) = \sum_{k=0}^{M} V_{EOG_i}(t - k) \ast p(k)
$$

Where $V_{eog}$ was the EOG reached EEG recording electrode, $V_{EOG}$ was the EOG electrode recording and $p$ the linear filter connecting the two signals, the EEG recording was composed by two components $i.e. \text{EEG}_i = V_{eog_i} + V_{eeg_i}$. By converting the time domain convolution into frequency domain multiplication, they derived Eq.2-24.

$$
Z_{\text{EEG}_i}(w) = P(w) \ast (Z_{\text{EOG}_i}(w) - Z_{\varepsilon_i}(w)) + Z_{\text{eeg}_i}(w)
$$

Where $\varepsilon_i$ accounted for the EEG leak into EOG recording (assume small). $Z_{\text{EEG}_i}$ was the FFT of the $i^{th}$ EEG recording series, $Z_{\text{EOG}_i}$ was the FFT of the $i^{th}$ EOG recording series, $Z_{\text{eeg}_i}$ was the FFT of the $i^{th}$ EEG recording series after EOG cleaning, $P(w)$ was the
linear transfer function. Apply least square method to these frequency domain equations, the estimated transfer function Eq.2-25.

\[
\hat{P}(w) = \frac{\sum Z_{EEG_i}(w)Z_{EOG_i}^*(w)}{N} - \frac{Z_{EEG_i}(w)Z_{EOG_i}^*(w)}{\sum |Z_{EEG_i}(w)|^2} - \frac{|Z_{EOG_i}(w)|^2}{\sum |Z_{EOG_i}(w)|^2}
\]  

(2-25)

And the ‘true’ EEG could be recovered accordingly.

In order to regress EOG artifacts out of recording either in time or in frequency domain, reference channels of EOG recordings are needed, there are many research study regression technique on different kind of eye activities. References [94], [107] had detailed discussions on all the research ramifications. To date EOG correction research has focused on problems relating to the estimation of the correction coefficients \( \hat{B} \) [108].

There were also researches dealing with multi-types of artifact at one time. Givens et al. [109] had used frequency domain method to reject body movement, muscle and eye movement. In [110] Koles applied PCA to extract abnormal component and/or artifacts from EEG. He built spatial covariance matrix \( R_n \) from multiple know normal EEG recording section (multiple channels), took in an abnormal section and calculate the spatial covariance \( R_a \), then SVD decompose the \( (R_a+R_n) \), the eigenvectors and the eigenvalues were used to map the abnormal section, the normal and abnormal waveforms were visually judged after the mapping.

Further development on EEG decomposition is the concept of blind sources separation (BSS)/independent component analysis (ICA). The BSS/ICA concept had been applied to artifacts sources decomposition by several researchers. Based on model \( X(t) = A*S(t) \) where the \( X(t) \) is multi-dimensional observation (the EEG recordings
from different channels) at time $t$, the $A$ is an unknown instantaneous mixing matrix, the $S(t)$ the unknown sources (to reject artifacts the special interesting sources are variety of artifacts and their mapping). Vigario [111] applied the FastICA algorithm to decompose ocular artifacts from EEG. Jung et al.[112] used extended InfoMax algorithm removed eye movement and muscle, eye blink and muscle, blink and blink related activities, line noise, line noise and harmonics with irregularly spaced frequencies. Potter et al. [113] tested extended InfoMax on artificially mixed white and pink noise with EEG and ECG. All the application had very positive results. Vorobyov and Cichocki [114] applied their ICA algorithm developed on different modeling assumption which suggested that $X(t) = AS(t) + \nu(t)$ where $S(t)$ contained a subset of useful sources (like ‘true’ EEG) with temporal structure and noise (like artifacts), and $\nu(t)$ an unknown additive white Gaussian/colored noise to represent the measurement and environment noise. They had simulated on a short section of scalp EEG laden with artifacts (they did not mention the kind of artifacts).

Summary

Literature review shows that seizure detection in the past is concentrated on using linear signal processing technique. Filters have been widely used to extract features, besides waveform descriptor has been used and tested on large amount of data by different researchers, the method was commercialized in 1980s and have been updated to version 8 up to date. And all, except one [58], linear signal processing technique on seizure prediction could predict seconds in advance of seizure occur. These methods are basically heuristic, which mimic the observed waveforms by mathematical description.
Neural network technique has been used on EEG signal segmentation and seizure detection. The segmentation, through nonlinear functions, map typical observed EEG segment into classes, including the in-seizure class. Neural network seizure detection, instead of multi-classifying EEG time series, converts EEG into two classes, the general background and in-seizure. The method extract in-seizure features use wavelets then compare test signal with the feature. The neural network technique take one step further from heuristic waveform description, modeling EEG signals through nonlinear functions.

The most active research in the past one more decade is the application of nonlinear dynamic theory on seizure prediction. By using the theory researchers found that there might be a pre-ictal state exists minutes before seizure occur, compare with tens of seconds prediction time which linear signal processing technique obtained, nonlinear dynamic theory do have an edge, especially Lyaponov exponent, the well derived mathematical parameter, used by Iasemidis et al. [61]. Based on STLmax, seizure could be predicted tens of minutes in advance, in addition the entrained electrode pairs and the best predictions channels have been quantified through T-index statistics and optimization method. But STLmax parameter, like many other nonlinear dynamic characterization parameters reviewed, has not been applied on seizure detection.

This motivated my research study the nonlinear dynamic theory on seizure detection. The starting point is to detect seizures based on STLmax parameter extracted from EEG.

For EEG analysis in general, any methods intend to extract information from the signal have to deal with artifacts one way or the other. Depends on the type extracted features/parameters, gain could come from dealing with artifacts before
features/parameters extraction or after. Most of methods developed before Jung et al. [112] divide artifacts type and conquer one by one. The very recent discovery is that blind source separation algorithms can decompose multiple types of artifact sources from artifacts carrying EEG recordings, based on one simple assumption that the artifact sources are statistically independent from ‘true’ neuronal signal, which looks like a reasonable assumption.

Armed with the knowledge of status quo of EEG research in general, an online automatic seizure detection scheme is proposed in section 3 and the building blocks are tested through section 4 to 9.
CHAPTER 3
SEIZURE-DETECTION SCHEME

From the signal processing point of view, the two factors that have decisive effects on the complexity of a seizure detection scheme are: the type of data where the algorithm builds upon and the feature/ features to be extracted. If data are collected from scalp, almost surely more artifacts are going to be picked up by recording electrodes than the data collected from depth. Therefore, to achieve similar performance, detection scheme for scalp data may have to be more complex in order to suppress the artifacts influence. The same is true for the extracted features. Up to now unique feature/features that could perfectly distinguish in-seizure EEG from background (all non-seizure activity) have not been found. Usually seizure detection relies on a group of features or a single feature to differentiate in-seizure EEG from background as much as possible, the less overlapping feature/ features between in-seizure and background the less complex the detection scheme. In addition, the complexity of a detection scheme is positively correlated with calculation time. Schemes with higher complexity that are not based on mathematical principles are not only less intuitive but also make online implementation harder. The preference for an intuitive scheme is that algorithms rely on parameters to link performance to the real signal features and it is easier to tune an intuitive scheme. With these thoughts in mind, the development of a seizure detection scheme is summarized in this chapter.
Data

At the beginning of this research some ECoG data is available. These data were collected from patients who were suffering pharmaco-resistant temporal lobe epilepsy and were admitted to Shands Teaching Hospital at University of Florida for surgery. The recordings were made during the invasive investigation of their epileptogenic foci. The number of electrodes and their placement were decided based on the previous non-invasive study. Figure 3-1 shows basic electrodes set up.

The ECoG was first recorded on VHS tapes with video monitoring. Then band pass filtered in 0.5 to 70 Hz range, digitized using 200 Hz sampling frequency and 10 bits precision.
In this study, the recording of each patient was first scored by neurologists. Then a study set of the corresponding patient is formed by concatenating his/her tapes with seizure recording in it, the concatenation follows the chronological recording order, all tapes were included as long as there are seizure recording in it. Table 3-1 shows general information of the constructed study sets, and Table 3-2 to 4 detail each set, the relative time is the seizure onset time relative to the starting point of the study set.

Table 3-1. General information of the five study sets.

<table>
<thead>
<tr>
<th>Data ID</th>
<th>Sex</th>
<th>Age</th>
<th>Length (h)</th>
<th>Number of channels</th>
</tr>
</thead>
<tbody>
<tr>
<td>P092</td>
<td>M</td>
<td>60</td>
<td>34</td>
<td>28</td>
</tr>
<tr>
<td>P093</td>
<td>F</td>
<td>41</td>
<td>64</td>
<td>30</td>
</tr>
<tr>
<td>P148</td>
<td>M</td>
<td>19</td>
<td>75</td>
<td>28</td>
</tr>
<tr>
<td>P185</td>
<td>M</td>
<td>29</td>
<td>47</td>
<td>28</td>
</tr>
<tr>
<td>P256</td>
<td>F</td>
<td>38</td>
<td>15</td>
<td>32</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>235</td>
<td></td>
</tr>
</tbody>
</table>

Table 3-2. Detail of P093 set.

<table>
<thead>
<tr>
<th>Seizure number</th>
<th>Type</th>
<th>Relative hour</th>
<th>Duration min: sec</th>
<th>EEG and clinical seizure feature briefing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PSG</td>
<td>1.54</td>
<td>1:36</td>
<td>Onset right hippocampus (RH) (asleep, arose, cry)</td>
</tr>
<tr>
<td>2</td>
<td>PSG</td>
<td>4.18</td>
<td>1:57</td>
<td>Onset from RH (asleep, arose, deep breath)</td>
</tr>
<tr>
<td>3</td>
<td>PSG</td>
<td>7.09</td>
<td>2:07</td>
<td>Onset from RH (asleep, breath, jerk head)</td>
</tr>
<tr>
<td>4</td>
<td>CP</td>
<td>8.87</td>
<td>1:38</td>
<td>Onset from RH (asleep, head sharp move)</td>
</tr>
<tr>
<td>5</td>
<td>CP</td>
<td>10.86</td>
<td>1:15</td>
<td>Onset from RH (asleep, eye flutter, head sharp move)</td>
</tr>
<tr>
<td>6</td>
<td>CP</td>
<td>13.29</td>
<td>1:44</td>
<td>Onset from left mesial temporal (awake, oral automatism)</td>
</tr>
<tr>
<td>7</td>
<td>CP</td>
<td>14.93</td>
<td>1:18</td>
<td>Onset from RH (reading in bed, grimace)</td>
</tr>
<tr>
<td>8</td>
<td>CP</td>
<td>23.39</td>
<td>1:23</td>
<td>Onset from RH (sitting in bed, chewing)</td>
</tr>
<tr>
<td>9</td>
<td>CP</td>
<td>25.39</td>
<td>1:09</td>
<td>Similar to right hippocampus (sitting in bed, exhale)</td>
</tr>
<tr>
<td>10</td>
<td>CP</td>
<td>27.17</td>
<td>0:59</td>
<td>Onset from RH (do cross puzzle in bed, laugh)</td>
</tr>
<tr>
<td>11</td>
<td>CP</td>
<td>31.67</td>
<td>0:22</td>
<td>Similar to 7-10 (awake quite in bed, eye, mouth open)</td>
</tr>
<tr>
<td>12</td>
<td>CP</td>
<td>35.78</td>
<td>1:16</td>
<td>Onset from RH (asleep, eye flutter, exhale forcefully)</td>
</tr>
<tr>
<td>13</td>
<td>CP</td>
<td>37.97</td>
<td>1:08</td>
<td>Onset from RH (asleep, blinking rose)</td>
</tr>
<tr>
<td>14</td>
<td>CP</td>
<td>41.29</td>
<td>1:14</td>
<td>Onset from RH (asleep, oral automatism)</td>
</tr>
<tr>
<td>15</td>
<td>CP</td>
<td>43.68</td>
<td>1:08</td>
<td>Onset from RH (eye close, blinking, deviate)</td>
</tr>
<tr>
<td>16</td>
<td>CP</td>
<td>47.12</td>
<td>1:34</td>
<td>Onset from RH (watching TV, stare, laugh)</td>
</tr>
<tr>
<td>17</td>
<td>CP</td>
<td>49.48</td>
<td>0:46</td>
<td>Onset from RH (lying, motionless open eye)</td>
</tr>
<tr>
<td>18</td>
<td>CP</td>
<td>50.61</td>
<td>0:33</td>
<td>Onset from RH (asleep, blink, eye open wide)</td>
</tr>
<tr>
<td>19</td>
<td>CP</td>
<td>50.94</td>
<td>0:50</td>
<td>Onset from RH (asleep, eye flutter, exhale)</td>
</tr>
<tr>
<td>20</td>
<td>CP</td>
<td>57.30</td>
<td>0:57</td>
<td>Onset from RH (sitting, eye surprise, mouth open)</td>
</tr>
<tr>
<td>21</td>
<td>CP</td>
<td>60.31</td>
<td>1:21</td>
<td>Onset from RH (mouth open, head tILT, exhale forcefully)</td>
</tr>
<tr>
<td>22</td>
<td>SC</td>
<td>62.28</td>
<td>0:55</td>
<td>Onset from RH (no behavior activity)</td>
</tr>
<tr>
<td>23</td>
<td>CP</td>
<td>63.05</td>
<td>1:05</td>
<td>Onset from RH (toothpick, eye open widely, exhale)</td>
</tr>
</tbody>
</table>

1 sub-clinical, 22 clinical seizures in 63.69 hours
Table 3-3. Detail of P185 and P148 sets.

<table>
<thead>
<tr>
<th>Seizure number</th>
<th>Type</th>
<th>Relative hour</th>
<th>Duration min: sec</th>
<th>Seizure number</th>
<th>Type</th>
<th>Relative hour</th>
<th>Duration min: sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SC</td>
<td>2.75</td>
<td>0:39</td>
<td>1</td>
<td>CP</td>
<td>3.13</td>
<td>1:27</td>
</tr>
<tr>
<td>2</td>
<td>SC</td>
<td>3.73</td>
<td>0:49</td>
<td>2</td>
<td>CP</td>
<td>7.33</td>
<td>0:22</td>
</tr>
<tr>
<td>3</td>
<td>SC</td>
<td>6.02</td>
<td>0:23</td>
<td>3</td>
<td>SC</td>
<td>17.56</td>
<td>2:20</td>
</tr>
<tr>
<td>4</td>
<td>CP</td>
<td>6.67</td>
<td>1:14</td>
<td>4</td>
<td>SC</td>
<td>25.44</td>
<td>0:47</td>
</tr>
<tr>
<td>5</td>
<td>CP</td>
<td>8.66</td>
<td>1:00</td>
<td>5</td>
<td>PSG</td>
<td>26.32</td>
<td>1:37</td>
</tr>
<tr>
<td>6</td>
<td>SC</td>
<td>10.84</td>
<td>1:30</td>
<td>6</td>
<td>PSG</td>
<td>31.01</td>
<td>1:29</td>
</tr>
<tr>
<td>7</td>
<td>SC</td>
<td>11.45</td>
<td>1:46</td>
<td>7</td>
<td>PSG</td>
<td>35.39</td>
<td>1:43</td>
</tr>
<tr>
<td>8</td>
<td>CP</td>
<td>12.65</td>
<td>1:25</td>
<td>8</td>
<td>SC</td>
<td>38.80</td>
<td>1:22</td>
</tr>
<tr>
<td>9</td>
<td>CP</td>
<td>16.02</td>
<td>1:56</td>
<td>9</td>
<td>SC</td>
<td>39.33</td>
<td>1:22</td>
</tr>
<tr>
<td>10</td>
<td>SC</td>
<td>17.85</td>
<td>2:34</td>
<td>10</td>
<td>CP</td>
<td>39.39</td>
<td>2:07</td>
</tr>
<tr>
<td>11</td>
<td>CP</td>
<td>19.18</td>
<td>1:00</td>
<td>11</td>
<td>CP</td>
<td>48.81</td>
<td>2:37</td>
</tr>
<tr>
<td>12</td>
<td>SC</td>
<td>24.09</td>
<td>0:10</td>
<td>12</td>
<td>SC</td>
<td>51.71</td>
<td>0:16</td>
</tr>
<tr>
<td>13</td>
<td>SC</td>
<td>24.41</td>
<td>0:11</td>
<td>13</td>
<td>SC</td>
<td>53.24</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>SC</td>
<td>25.32</td>
<td>0:20</td>
<td>14</td>
<td>PSG</td>
<td>54.84</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>SC</td>
<td>28.23</td>
<td>3:18</td>
<td>15</td>
<td>PSG</td>
<td>59.96</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>CP</td>
<td>32.74</td>
<td>0:22</td>
<td>16</td>
<td>SC</td>
<td>68.05</td>
<td>0:48</td>
</tr>
<tr>
<td>17</td>
<td>CP</td>
<td>35.29</td>
<td>0:19</td>
<td>17</td>
<td>PSG</td>
<td>74.83</td>
<td>1:16</td>
</tr>
<tr>
<td>18</td>
<td>CP</td>
<td>40.75</td>
<td>1:32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>SC</td>
<td>46.60</td>
<td>2:00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

11 sub-clinical, 8 clinical seizures in 46.62 hours

Clinical equipment used for these data to record had pre-set, un-adjustable maximal amplitude of $\pm 0.6$ mV. When amplitudes of the signal exceed this limit, they were clipped. Clipping situation was observed during some seizures and inter-ictal spikes.

**Feature Extraction**

Detection of seizures by features extracted from morphological observation combined with linear signal processing technique or neural network has been well studied during the past two more decades, the performance varies as can be seen from the
review. Although it is impossible to compare between methods because a standard is lacking [29], whenever new theory or technique appears, it always opens the possibility that they might help improve the status quo.

Nonlinear dynamic theory has been used to search for pre-ictal features and has generated some positive results. But the theory has not been applied to seizure detection. If these positive results can be regarded as an indication that the assumption behind the theory might be legitimate to the generation mechanism of EEG signal, the theory may also be applicable to seizure detection.

The STLmax [64] is considered first because it was available to predict epileptic seizures hence there will be no need to compute other features. It was observed that around the seizure the pattern of STLmax differs from background, i.e. the STLmax value gradually drops below the average background value then bounces back to overshoot the background. As seizures die out, the STLmax feature returns to the background level. Based on this observation a seizure detection scheme was developed and tested. The issue of STLmax feature based seizure detection scheme is that it is impossible to implement the method online.

An alternative nonlinear recurrence time statistics measure proposed by Gao [115] was then tested. It turns out that in the majority of cases the in-seizure feature produced by this measure has very good contrast to non-seizure state. Besides, the calculation expenses is very low (for 10.24 seconds data, the method needs 3.555 seconds to extracted the corresponding feature when implemented on Pentium IV PC with MATLAB 6.5) which is very encouraging. So the measure is studied in detail in Chapter 4, and it is compared with both the STLmax and Teager energy [116] for seizure
detection. Comparison with Teager energy is undertaken because it is also a nonlinear feature and its estimation is as simple as T1.

**Artifact Sources**

Review of the past artifacts minimization methods shows that BSS/ICA is preferred, because the same algorithm could decompose several types of artifacts as long as the artifacts generating sources are statistically independent of the ‘true’ brain signal. If the BSS/ICA algorithm has decent performance, the estimated artifacts sources signal should be free of any brain signal. The artifacts in the decomposed domain will then be easier to recognize because they are not contaminated by other activity.

The issue of applying BSS/ICA algorithm is that the estimated artifacts sources have gain and order uncertainty, so the artifacts sources have been picked by visual observation [112], BSS/ICA algorithm can tell that the estimated artifacts sources project the same observation through the corresponding de-mixing vectors as the ‘true’ artifacts sources through the ‘true’ mixing vectors.

For artifacts correction, the gain uncertainty is not a problem, as we only care about artifacts contribution to the recording (the final multiplication result), if the artifacts sources are known, their contribution to the recording can be simply subtracted. A data efficient minimize Renyi’s mutual information through stochastic information gradient (MRMI-SIG) ICA algorithm has been tested in Chapter 5 to decompose several noise clusters which the feature extraction method is sensitive to. The algorithm is able to converge in 6.005 seconds and decompose the artifacts components out of 30 channels of 10.24 seconds (2048 samples) ECoG data.

The only issue is that the order uncertainty (which component will be the artifact source) hampers BSS/ICA algorithms being used on large amount of data [112]. So to
correct artifacts automatically, methods to recognize artifacts are needed and have been developed as explained below.

**Automatic Artifacts Correction by Template Matching**

Artifacts correction by template matching is difficult because artifacts, like EEG, have patterns as they are listed on artifact atlas. Our methodology is as follows: Build offline templates for the most frequent artifacts which have similar features to in-seizure state and have caused the majority amount of false alarms; then present these templates after ICA decomposition, and compare with every artifacts template and the ‘matched’ components to reject these segments from the original recording. The recurrence time statistic feature can be extracted after cleaning the raw ECoG. In this way detection performance will be improved in the sense that false alarms will be reduced. The template matching methods is detailed in Chapter 6.

**System Schematic**

Summarizing all the pieces together, the automatic seizure detection system block diagram is depicted in Figure 3-2.

![Seizure detection system diagram](image)

Figure 3-2. Seizure detection system.

The system performance is tested in Chapter 7. During the progress of this research, some scalp EEG become available, the possibility of applying this detection
scheme on scalp data is tested in Chapter 8. Practical application of the methodologies and the extension of future research are discussed in Chapter 9.

**Contributions**

a) Tested the applicability of nonlinear feature STLmax for seizure detection, and demonstrates that a seizure detection scheme based on STLmax is feasible, but to obtain decent performance the scheme tends to be patient dependent and beside it is difficult to implement the scheme online and automatic.

b) Extracted and evaluated T1, the recurrence time statistic feature, demonstrated the applicability of T1 for seizure detection, and the advantage of T1 over STLmax and Teager energy.

c) Evaluated the applicability of MRMI-SIG blind source algorithm on EEG artifact sources decomposition.

d) Developed a self-organizing mapping based template matching method to correct artifacts sources after blind source separation.

e) Tested the performance of blind source separation-artifact correction-T1 feature extraction on ECoG seizure detection.

f) Demonstrated the poor performance of the previous scheme on scalp EEG artifacts correction.

g) Proposed and tested a seizure detection scheme use wavelet preprocessing and T1 feature extraction for EEG in general.
CHAPTER 4
FEATURE SELECTION

By applying nonlinear dynamic theory to EEG signal, the Brain Dynamics Bioengineering Research Partnership found the exciting phenomenon that the Short Term maximum Lyaponov (STLmax) exponent began to drop gradually. Based on this feature, seizures can be predicted tens of minutes in advance [66], the prediction time scale is much better than linear waveform based methodologies developed so far. Unlike linear methods, nonlinear dynamic theory embeds one dimensional EEG time series into a high dimensional phase space, and the STLmax is a feature to characterize the reconstructed phase space. It is expected that extracting features from phase space is better than amplitude dependent measures, as it is well know that the amplitude varies largely in EEG and the variance make the directly amplitude dependent algorithm susceptible to false decisions.

As nonlinear dynamic theory has not been applied to seizure detection before, the intention here is to explore its applicability to seizure detection.

Short Term Lyaponov Exponent Feature and Seizure Detection

Short Term Lyaponov Exponent Feature

The STLmax (for the convenience of reading, the algorithm is briefly listed in APPENDIX A) feature is available for the 5 study sets listed in the data section of chapter 3. Its gradually dropping character before seizure has been associated with the brain state and used on seizure prediction in [66]. Figure 4-1 is a short section to show this feature around one typical seizure.
Seizure Detection Based on STLmax

Based on this observation a seizure detection scheme was developed. The idea is that assuming feature extraction from phase space could avoid the EEG amplitude fluctuations (changes caused by different brain state other than seizure, artifacts modulation etc.), the STLmax feature extracted from EEG can be regarded as a piecewise stationary process, therefore a generalized likelihood ratio test (GLRT) [120] can be used to segment the feature into: the general in-background state and the in-seizure state.

Rough sketch of GLRT (refer to [120] [141] for thorough treat): assuming that there is a sequence of independent random variable $y(k)$, with probability $p_\theta(y)$ depend upon the parameter $\theta$, was observed, before an unknown time $t_0$ the parameter is $\theta_0$ and after $t_0$ the parameter becomes $\theta_1$. The interest is to find the $t_0$ where the change occurred based
on the observation. The logarithm of the likelihood ratio is defined by Eq.4-1 [120]. By 
checking the change of L(y) the changing time can be traced back.

\[ L(y) = \log \left( \frac{p_{\theta_1}(y)}{p_{\theta_0}(y)} \right) \]  

(4–1)

In our case, due to the observed in-seizure feature structure, GLRT is not directly 
applicable to the feature because the process is not independent identical distributed 
(i.i.d), but fortunately an inverse model could be used to convert the process into 
innovations which will be i.i.d., and segment the innovations is equivalent to segment the 
original series [141].

As there is no desired signal for modeling, a one step predictor is used. Typical 
seizure sections and some background sections are selected to train two predictors 
(background predictor and seizure predictor) to learn the model using Wiener-Hopf 
algorithm [117] \( R_0 w = p \), where \( R \) is the input M by M correlation matrix, \( p \) is M by 1 
cross correlation vector between tap input and desired, \( w_0 \) is the optimum tap weight, M 
is the order of the filter. Figure 4-2 is the model block diagram.

![Figure 4-2. Background and in-seizure modeling.](image)

The model is optimally trained by Wiener Hopf equations, and the model order is 
tuned until the error power spectral density (PSD) is approximately white, or equivalently 
the autocorrelation of non zero lags goes to zero. Figure 4-3&4 is an illustration of
background modeling error PSD and autocorrelation. The same procedure is used to tune the in-seizure model.

The cost function for automatic (unsupervised) segmentation is the mean square error after modeling. It was found by [141] that direct segmentation on this cost function generated smeared switching, and adding memory, as Figure 4-5 shows, to the cost function will reduce the smearing.

![Figure 4-3. Power spectral density of background modeling error (P063 ID was later re-numbered as P93).](image)

The error is collected through Eq.4-2, where $\alpha$ is a tunable parameter and $\alpha^{-1}$ represents the memory depth [141].

$$\varepsilon(k) = \alpha \cdot e(k)^2 + (1 - \alpha) \cdot \varepsilon(k-1), \quad \text{where} \quad 0 < \alpha < 1 \quad (4-2)$$

It is straightforward to see that $\varepsilon(k)$, called the innovation sequence, is an unbiased estimator of mean square error as $E[\varepsilon] = E[e^2]$ [141] [118] [119]. After these preprocessing steps, the equivalent innovation is ready to be segmented. In our case the process needs to be further refined since we have many EEG channels, so individual
channel assignments must be combined into a single event (seizure or background).

Figure 4-6 shows the block diagram.

Figure 4-4. Autocorrelation of background modeling error.

Figure 4-5. Model and modeling error collection details.

The accumulated error energy $\varepsilon(k)$ after error collection is feed into a cumulative sum (CUSUM) algorithm. Eq.4-3 is the algorithm. The reason to choose this algorithm is that typical behavior of the log-likelihood ratio shows negative drift before model change, followed by a positive drift after change [120]. The sequential ratio test algorithm restart the tracking as long as the drift is negative, start accumulate only when
the drift become positive, and the start point of the accumulation is regarded as the time where change occur.

\[
g^{(0)} = 0
\]

\[
g^{(k)} = \begin{cases} 
  g^{(k-1)} + \log \left( \frac{p_1(\epsilon^{(k)})}{p_0(\epsilon^{(k)})} \right) & g^{(k)} > 0 \\
  0 & g^{(k)} \leq 0
\end{cases}
\]

(4 - 3)

Here \(\epsilon^{(k)}\) is the collected modeling error energy at time \(k\). \(p_1\) is the pdf of \(\epsilon^{(k)}\) for seizure model which is assumed zero mean Gaussian and \(p_0\) is the pdf of \(\epsilon^{(k)}\) for background which is also assumed as zero mean Gaussian (their variances are estimated during model training) and \(g^{(k)}\) is the output of the algorithm at time instance \(k\).

This algorithm avoids negative accumulation, it dose not have to keep the entire memory contained in the past observation like the adaptive threshold CUSUM does [120], and the change can be picked up immediately whenever it occurs.

In real implementation, in stead of treating the immediate positive drift as a state switching point, a threshold is applied to the output \(g^{(k)}\) of the CUSUM algorithm, so that the accumulated \(g^{(k)}\) below the threshold is regarded as below the noise floor. The threshold setting is a value decided through simulation. Those \(g^{(k)}\)s that exceed a threshold are used in the final electrode voting to generate seizure alarms.

As there is no prior knowledge on which electrode is more ‘important’, all electrodes have equal weight. An added difficulty is that the time between seizures varies from seizure to seizure and the seizure duration also varies. Therefore, a specific voting scheme is set up to accommodate the data for the 5 sets of data, which we consider applicable to a wider group of patients.
It was observed that the minimum space between two consecutive seizures is around 30 minutes, and seizures length is in the range of 10 seconds to 3 minutes. So voting is conducted to every 15 minutes non-overlapping time bins. In 15 minutes, the algorithm counts the number of channels where $g(k)$ is over the threshold. If there are more than two channels exceed the threshold, an alarm will be announced, and if a seizure happened in this time bin, then the alarm is counted as machine detection, otherwise it is labeled as a false alarm.

In general, false alarms are expected to increase if the time bin is shorter, while longer time bins will generate lesser false alarms, because no matter how many switches happened inside each time bin, it only makes one announcement. However, this has a detrimental effect on the number of detections since several seizures may be combined together.
Performance Receiver Operating Characteristic curve

The corresponding ROC curve is generated by varying the threshold. Detection rate and false alarm rate are defined by Eq. 4-4 and 4-5.

\[
\text{Detection Rate} = \frac{\text{Detected seizures}}{\text{Total humanscored seizures of the test period}} \quad (4 - 4)
\]

\[
\text{False Alarm Rate} = \frac{\text{False Alarms}}{\text{Testing hours}} \quad (4 - 5)
\]

Figure 4-7 is the ROCs for the 5 sets of data and Table 4-1 is a “good” trade off snapshot taken from it. In detection, the trade off is between the detection rate and false alarm rate, so “good” here means on the ROC curve, within the tolerable false alarm rate range, the best detection rate the algorithm is able to achieve.

Parameters

The above performance is obtained by carefully tuning the parameters listed in Table 4-2, especially the ‘system delay’.

The reason for implementing a delay parameter is due to several reasons: the seizure starting point marked by different neurologist usually varies; the intrinsic delay of the whole detection scheme build upon STLmax (each STLmax value represents 10.24 seconds of EEG signal); the modeling of the STLmax has one step delay; the error collection has memory depth $\alpha^{-1}$ and CUSUM accumulation also has a delay depending upon the threshold.

Table 4-1. Performance.

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Detection rate</th>
<th>False alarm per hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>P092</td>
<td>80%</td>
<td>0.10</td>
</tr>
<tr>
<td>P093</td>
<td>80%</td>
<td>0.10</td>
</tr>
<tr>
<td>P148</td>
<td>80%</td>
<td>0.20</td>
</tr>
<tr>
<td>P185</td>
<td>90%</td>
<td>0.10</td>
</tr>
<tr>
<td>P256</td>
<td>80%</td>
<td>1.00</td>
</tr>
<tr>
<td>Average</td>
<td>~80%</td>
<td>~0.30</td>
</tr>
</tbody>
</table>
Table 4-2. Parameters.

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Number of seizure used on model training</th>
<th>Number of test seizures</th>
<th>System delay in samples</th>
<th>Background $\alpha_0$ / Seizure $\alpha_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>P092</td>
<td>2</td>
<td>5</td>
<td>30</td>
<td>0.1 / 0.5</td>
</tr>
<tr>
<td>P093</td>
<td>3</td>
<td>20</td>
<td>25</td>
<td>0.1 / 0.5</td>
</tr>
<tr>
<td>P148</td>
<td>3</td>
<td>14</td>
<td>15</td>
<td>0.1 / 0.5</td>
</tr>
<tr>
<td>P185</td>
<td>3</td>
<td>16</td>
<td>40</td>
<td>0.1 / 0.9</td>
</tr>
<tr>
<td>P256</td>
<td>3</td>
<td>5</td>
<td>20</td>
<td>0.1 / 0.9</td>
</tr>
</tbody>
</table>

Background model order $M=30$ Seizure model order $M=5$

Although final performance is acceptable, because of the complexity of tuning this seizure detection scheme and the delay caused by the complexity, the scheme is not able to be used online, and it is patient dependent. So the question becomes: is there any other nonlinear dynamic feature extraction method that could bypass these limitations?

Figure 4-7. Test ROCs.
Recurrence Time Statistic Feature and Seizure Detection

Recurrence time statistics is a method rooted in chaos theory. It was developed for stationary test and is constructed as follows. Assume that we have a scalar time series \( \{x(i), i=1, 2, \ldots, M\} \), where \( i \) is the time index. According to Takens’ embedding theory [121], the corresponding \( m \) dimensional phase space can be built by constructing vectors from the time series, \( \mathbf{X}_k = [x(k), x(k+L), x(k+2L) \ldots x(k+(m-1)L)] \), where \( L \) is the time delay. The vector sequence \( \{\mathbf{X}_k, k=1,2\ldots,N\} \) constitutes a trajectory in the phase space with \( N=M-(m-1)L \).

Then choose an arbitrary reference point \( \mathbf{X}_0 \) in this constructed phase space, and consider the recurrence to its neighborhood of radius \( r \): \( B_r(\mathbf{X}_0) = \{|| \mathbf{X}_j - \mathbf{X}_0 || \leq r, j \in [1 N] j \neq 0\} \). Denote the subset of the trajectory belongs to \( B_r(\mathbf{X}_0) \) by \( \mathbf{S}_1 = \{\mathbf{X}_t_1, \mathbf{X}_t_2, \ldots, \mathbf{X}_t_i, t_i \in [1 N] t_i \neq 0\} \). These points are called Poincaré recurrence points. And the Poincaré recurrence times are simply defined as \( \{T_1(i)=t_{i+1}-t_i, i=1,2,\ldots\} \). The T1 index of this reference point \( \mathbf{X}_0 \) is the mean of the above generated T1 set. The overall T1 of the whole phase space is the average of the T1 indices of all the reference points. Figure 4-8 illustrates the T1 generation of one reference point.

Implementation and Parameters

It is know that EEG electrodes record neuronal activity, and epileptic seizures are abnormal temporary manifestations of dramatically increased neuronal synchrony, either occurring focally or bilaterally [122]. The EEG itself, as a form of neuronal ongoing information processing, is non-stationary.

To implement this method, the raw EEG was partitioned into non overlapping 10.24 sec windows to keep the compatibility to the previous methods (i.e. 2048 samples...
at a sampling frequency 200 Hz). The phase space of each window is constructed according to Takens theorem, and two parameters need to be decided: the embedding dimension $m$ and time delay $L$.

Figure 4-8. T1 Illustration

According to Taken’s embedding theory, if the attractor’s dimension is $D$ (may be non-integer), then a constructed phase space, with $m > 2D+1$ ($m$ should be an integer) embedding dimension, is able to reveal the underlying dynamics.

For an unknown dynamical system like the brain, there is no established method to define this $D$ parameter at present. And some authors claimed that the seizure state can be described by low-dimensional dynamical system [84] [123]. So we decide the value $m$ by simulation. Start from a lowest number $D=1.5$ which is larger than limited cycle correlation dimension [64], $m=2*1.5+1=4$, and leave theoretic $D$ parameter deduction for future study.

$T_1$ subset of reference point $X_0$ is:

$S_1 = \{X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16}, X_{17}, X_{18}, X_{19} \}$

(total 18 recurrence points inside $r$)

$T_1(X_0) = (1+1+1+4+1+1+3+1+1+1+4+1+1+1)/18$
Delay \( L \) need to be small enough to capture the shortest change present in the data and large enough to generate the maximum possible independence between components of the phase space vectors [64]. We adopt the autocorrelation method introduced by [124] to decide \( L \), the first zero of the autocorrelation of in-seizure time domain ECoG is between 4 and 5, therefore we use \( L=4 \).

**Feature**

Figure 4-9 is a T1 plot (delay=4, dimension=4, radius=0.0039, window=2048) generated by approximately 5.5 hours of ECoG, where the two peaks correspond to the two seizures occurred during this recording. As can be observed, there is a distinct difference of the T1 value from the seizures to the background, which indicates a possible good feature for seizure detection.

![Figure 4-9. T1 feature.](image-url)
Figure 4-10 is the T1 plot over 64 hours of the 5 electrodes placed on the left temporal depth (LTD) region. The dotted lines indicate seizure starting times marked by a neurologist. In this data set there are 23 seizures, the first three are secondarily generalized complex partial seizures, the 22nd is a sub-clinical seizure, and the rest are complex partial seizures. We can observe that for a large majority of seizures there is a significant increase of the T1 index, which is a good prognostic for a reliable seizure detector.

**Performance ROCs**

The distinguishing T1 peaks make seizure detection easier. There is no need for further processing after feature extraction. The electrode voting procedure can be applied directly to the feature.

![P33 LTD Region](image)

Figure 4-10. Regional T1 (Top to bottom electrode: LTD1, LTD3, LTD5, LTD7, LTD9)
Figure 4-11 is the ROCs of the 5 sets of data. They are generated by the same bin electrode voting procedure as in the STLmax based seizure detection. In any non-overlapping time bin, if there are more than 1 electrode that has features exceed the threshold, an alarm will be announced. The definition of detected and false alarm is the same as Eq.4-3 and 4-4. The curves are obtained by varying the threshold equally over all the electrodes. Figure 4-12 is the ROC of the clinical seizure detection only.

Table 4-3. Good compromise point of detection vs. false alarm (T1 feature).

<table>
<thead>
<tr>
<th>Data ID</th>
<th>At false alarm rate (alarm/hour)</th>
<th>(Detected clinical seizures) / (Total clinical seizures)</th>
<th>(Detected clinical + Sub-clinical seizures) / (Total clinical + Sub-clinical seizures)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P92</td>
<td>0.20</td>
<td>100%</td>
<td>85%</td>
</tr>
<tr>
<td>P93</td>
<td>0.23</td>
<td>95.5%</td>
<td>95%</td>
</tr>
<tr>
<td>P148</td>
<td>0.00</td>
<td>90%</td>
<td>52%</td>
</tr>
<tr>
<td>P185</td>
<td>0.30</td>
<td>87.5%</td>
<td>92%</td>
</tr>
<tr>
<td>P256</td>
<td>0.35</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Average</td>
<td>~0.22</td>
<td>~95%</td>
<td>~85%</td>
</tr>
</tbody>
</table>

Figure 4-11. Seizure detection ROC.
Figure 4-12. Clinical seizure detection ROC.

Table 4-3 summarizes good compromise point from the curves. It can be seen that performance based on STLmax seizure detection is very close to the performance based on T1 feature, but the later is much simpler, can be implement online and automatic.

The T1 feature is better than STLmax based seizure detection from our criteria point of view, but there is one possibility that the performance might be out of pure luck, as it is known that EEG tends to have higher energy and higher frequency during seizure when compared with background. And for the T1 feature itself, is there any room for improvement?

**Teager Energy and Seizure Detection**

The T1 in-seizure peak reminds us of the Teager energy operator [54] [125] [126]. Teager energy is a nonlinear operator and it is defined by Eq.4-5, 4-6, 4-7 where the $\Psi_c$, $\Psi_d$ are continuous and discrete operator respectively. Combining the operator definitions with in-seizure EEG waveform character, it is expected that in-seizure Teager energy
shall be higher than in-background, just like the T1 feature. But the feature is simpler than T1 as it operates directly in the time domain, and the search for the embedding parameters can be avoided.

\[
\begin{align*}
\Psi_d[x(n)] & \equiv x(n)^2 - x(n+1)x(n-1) \\
\Psi_c[x(t)] & \equiv \left(\frac{dx(t)}{dt}\right)^2 - x(t)\left(\frac{d^2x(t)}{dt^2}\right) \\
\Psi_c[A\cos(\omega t + \theta)] & = A^2\omega^2
\end{align*}
\] (4 – 6)  
(4 – 7)  
(4 – 8)

Figure 4-13 is a section of Teager feature from the same data that generates the T1 feature of Figure 4-7 and the STLmax feature of Figure 4-1. As expected, the Teager energy has similar peaking behavior when a seizure occurs. So the same seizure detection procedure as for T1 can be applied directly to Teager feature as well. Figure 4-14, 15 and Table 4-4 are the performance ROCs.

Figure 4-13. Teager energy feature.
Figure 4-14. Seizure detection performance.

Figure 4-15. Clinical seizure detection performance.
Table 4-4. Good compromise point of detection vs. false alarm (Teager feature)

<table>
<thead>
<tr>
<th>Data ID</th>
<th>At false alarms rate (alarm/hour)</th>
<th>(Detected clinical seizures) / (Total clinical seizures)</th>
<th>(Detected clinical + Sub-clinical seizures) / (Total clinical + Sub-clinical seizures)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P92</td>
<td>0.18</td>
<td>100%</td>
<td>85%</td>
</tr>
<tr>
<td>P93</td>
<td>0.04</td>
<td>79%</td>
<td>79%</td>
</tr>
<tr>
<td>P148</td>
<td>0.50</td>
<td>80%</td>
<td>50%</td>
</tr>
<tr>
<td>P185</td>
<td>0.50</td>
<td>25%</td>
<td>30%</td>
</tr>
<tr>
<td>P256</td>
<td>0.50</td>
<td>40%</td>
<td>40%</td>
</tr>
<tr>
<td>Average</td>
<td>~0.34</td>
<td>~64%</td>
<td>~57%</td>
</tr>
</tbody>
</table>

Simulation result shows that the lucky comes from two set of data the P093 and P092, but on average it is worth the trouble to go through embedding use T1 as a seizure detection feature.

**Practical Implementation and Improvement**

One important aspect that requires further study is the parameter selection to estimate the T1 index, and its sensitivity to different ECoG signals. Assuming the available 5 sets of ECoG can be regarded as a representative sample set of general EEG, we study the parameter setting for best performance. Besides, the detection results in Figures 4-11 and 4-12 show that for some data sets the detection rate and false alarm is not that satisfactory. Closer look at the feature may help spot some aspects for potential improvement.

**Practical Parameter Selection**

The recurrence time statistics T1 needs four parameters (radius $r$, window size, embedding dimension $m$, embedding delay $L$) before implementation. Depends on where the method will be used, the parameters can practically be chosen in different way.

Window length can be long or short, overlapping (higher resolution) or non-overlapping (lower resolution). The radius $r$ can be larger (less sensitive to time series variation) or smaller (higher sensitivity). For the selection of embedding parameters, except the
method mentioned in [124], there are theoretical analysis [127] and plenty of other practical [128] [129] implementation references.

In this study the statistics T1 is directly used as a feature for seizure detection, so the magnitude contrast between T1 in seizure and background decide how susceptible the detection scheme is to false alarms. The higher this contrast is the lesser chances of generating false alarms.

Figure 4-16 is a systematic simulation on the influence of each parameter variation to the contrast. Fifteen clinical seizures each with length of twenty seconds and same amount of background are randomly picked from the data used, and the corresponding T1 is extracted using different parameter sets. Each subplot is fifteen randomly picked in-seizure T1 and background T1 against the variation of one parameter.

Subplot (a) is the effect of neighborhood radius $r$ variation (other parameters are fixed at $m = 4, L = 4$, window size $= 2048$). The simulation shows that when $r$ is greater than $0.0039$ the background T1 feature has almost no change, the in-seizure T1 feature drops with the increase of radius. Better contrast is in the smaller $r$ range (but $> 0.0039$).

Subplot (b) is the widow size variation (other parameters are fixed at $m = 4, L = 4, r = 0.0039$), we can deduct from the plot that, on average the widow size has little influence on the contrast (except one seizure in the samples). Intuitively the widow size shorter than the shortest seizure length should be a good choice.

Subplot (c) is the variation of embedding dimension (other parameters are fixed at window size $2048, L = 4, r = 0.0039$), the embedding dimension between 3 and 6 is good, but from the simulation it seems that 6 is the best for this set of data.
Subplot (d) is the variation of delay (other parameters are fixed at window size 2048 \( m = 4 \), \( r = 0.0039 \)). The deduction from (d) is that delay parameter has a wide working range between 3 to 12, and 8 is the ‘best’ delay in this simulation. Besides the applicable delay range seems to show periodical pattern, this reflect the fact that periodic time series autocorrelation has multiple maximum and minimum.

This simulation gives a rough sketch on the range of parameters that can practically be used on the ECoG data. Finding practical parameters for other data could follow a similar procedure. When choosing the parameters, one should remember that the embedding dimension and delay are not independent of each other.

**Conclusion on Feature Selection**

Three nonlinear features have been tested. For the feature extraction method itself, STLmax and T1 are based on nonlinear dynamic theory and the features are extracted from phase space constructed from EEG time series. Teager energy is a nonlinear feature that is directly waveform based. The reason to compare with Teager energy is that, first it is waveform based and has similar peaking behavior during seizure as the T1; second the complexity of Teager energy extraction is similar to T1 feature extraction.

For the two nonlinear dynamic features, the STLmax contrast is not as obvious as T1 index, although STLmax has more parameters than T1 providing eventually more freedom in the implementation. However, the less obvious contrast forces the seizure detection methodology to take further steps, and these steps complicate the entire scheme, making automatic online detection very difficult. In this aspect, the T1 index has an advantage.
A final thought on STLmax. The EEG activity is a mixture of delta, theta, alpha, beta waves during normal background and becomes more regular and oscillatory during seizures as a combination of higher frequency activities in the beta and alpha ranges. From Eq.a-1 in APPENDIX-A, it can be seen that when the time series structure becomes “regular” the average STLmax for that window is bound to drop, as on average the divergence \( \delta X_{i,j}(\Delta t) \) in the numerator is going to be smaller. The “regularity” is usually being observed before seizure occurs, and this phenomenon has been used by reference [49] to extract feature for seizure prediction. During seizure EEG time series oscillate,
this is the reason that STLmaxs drops. As Lmax characterize the local regularity of the trajectory, it could capture those pre-seizure changes, it is proper for seizure prediction. On the contrary T1 feature is not able to capture the near seizure character, based on the present observation it is not applicable for seizure prediction.

From the feature in Figure 4-1, 4-9, 4-13, T1 and Teager is much better for seizure detection. And based on detection performance T1 index is preferred. Why Teager operator based detection generates more false alarms? After all, the peak feature in Figures 4-9 and 4-13 are very similar.

Unlike the T1 index, Teager operator is based on a linear signal model, analytical equation is more intuitive to understand what the method dose, so it is productive to observe how T1 and Teager behave when different sine waves are presented to the two algorithms. Maybe it might reveal where the difference comes from.

Figure 4-17 is the Teager energy for sine waves of different frequency and amplitude, it can be seen that the Teager responds to these changes by energy laddering the difference between waves. Figure 4-18 is the corresponding T1 index (the sine waves are embedded with dimension 2, delay 4, and window size 100) extracted under different radius. The T1 index is only able to differentiate the magnitude, not the frequency. What happened to T1 can be explained by Figure 4-19 and 4-20.

Figure 4-19 is 4 sine waves with different frequency, the blue trajectory is the lowest frequency, and the pink trajectory is the highest frequency (each of them has the same amount of time series samples). On each trajectory the small doted red and black circles indicate the trajectory starting point and reaching point after 30 steps, the r is the
neighborhood size used to calculate T1 feature for a reference point. The solid black circle indicates the traveling distance between steps.

Figure 4-17. Sine wave and Teager energy. (Left panel sine waves with different amplitude and frequency combination, right panel the corresponding Teager energy for the waves).

Figure 4-18. T1 of different sine waves extracted under different radius.
The recurrence time statistics $T_1$ for certain reference point is defined as the ratio between how long the trajectory left the neighborhood $r$ of the reference point forever (for convenience of following description it is named traveling time) and how many times the trajectory stay in the neighborhood during this period (named stay time).

Figure 4-19. Sine wave with different frequency.

It can be seen that frequency influences how fast the traveling speed is, because in the higher frequency case, the trajectory takes fewer steps to return to the neighborhood of the reference point, but it also needs fewer steps to leave the neighborhood (observe the black dash line on the trajectory). So, for a fixed reference point, the ratio between traveling time and stay time is almost constant no matter what the frequency is.

Figure 4-20 is 4 sine waves with different amplitudes (as above, the trajectory length are all the same). In this case, it can be seen that with lower amplitudes, the stay
time is longer, thus T1 is smaller. For higher amplitudes, the stay time is shorter, so the corresponding T1 is larger.

Figure 4-20. Sine wave with different amplitude.

Object 4-1 and 4-2 are two wave files to dynamically demonstrate the T1 concept and the frequency and amplitude influence using embedded sine waves. The sine waves are perturbed by random noise for the convenience of observation.

Object 4-1. Sine wave amplitude change and T1. (421MB, testA.avi, 1.5 Min.)
Object 4-2. Sine wave frequency change and T1. (428MB, testf.avi, 1.5 Min)

When dealing with normal background ECoG, the signal is broadband and the sensitivity of the Teager operator to frequency produces false peaks in high frequency activity. Since the T1 feature is not as sensitive to frequency as Teager operator, it is a draw back to differentiate pure tones, but in EEG signal processing it actually makes the seizure detection task into a zero one problem by responding only to amplitude. That is
the reason that Teager energy based detection has created more false alarms than T1 based.

Figure 4-21 are two section of real ECoG signal, the background section with a spike is chosen to give the trajectory some stretch.

Figure 4-21. EEG samples.

Figure 4-22 and 4-23 is the corresponding phase space trajectory. The solid line start from blue, at step n=30 changes into green and at step n=60 changes into cyan are the trajectory. The dotted red circle draws the neighborhood for reference point n=1, and the dotted black circles around n=30 and n=60 are the neighborhood of those two reference points.

From Figure 4-22, it can be observed that in background if without spike, the trajectory will crawl in a small area. For certain neighborhood size, the traveling time and stay time will be similar like the start reference point n=1. With spike the trajectory is stretched, and some reference point will have larger T1, like the reference n=30 and n=60. As T1 index is an average for the entire window, these few larger T1 will
eventually be averaged out by the majority of smaller T1, so on average the background T1 value is usually small, meaning the trajectory in general stays in the neighborhood.

Figure 4-22. Background trajectories. (The first 180 samples of the background ECoG)

Figure 4-23. In-seizure trajectories. (The first 100 samples of the seizure ECoG)
But for the seizure section, the trajectory is different, as can be observed in Figure 4-23. The almost rhythmic time series waveform in the top panel of Figure 4-21, make the trajectory stretch in phase space, so the majority reference points on the trajectory will have a large T1 if the neighborhood size is proper (meaning not too large so that the entire phase space is covered by the neighborhood), and on average T1 is larger for the seizure window.

Objects 4-3 is the WAV file of Figure 4-22 and Object 4-4 is the WAV file of Figure 4-23. The wave files are intended to dynamically show “traveling time”, “stay in” time, the trajectory stretch, the “regularity” of the trajectory, and their relation to the extracted T1 feature.

Object 4-3. Background EEG and T1. (1.72MB, testRb.avi, 10 Sec)
Object 4-4. In-seizure EEG and T1. (5.95MB, testRs.avi, 10 Sec)

**Improvements**

In this chapter the nonlinear recurrence time statistics has been selected for seizure detection. When only clinical seizures detections are considered, 95% of detection rate is achieved with 0.2 false alarms per hour. Considering the simplicity of the method and compared with all the published methods reviewed in chapter 2, this method is very effective for clinical seizure detection, although this kind of comparison is riddled with problem as the reported performance does not come from standard data sets.

Except performance, the method is also relatively insensitive to parameters and the parameters themselves are intuitive and easy to set. There is no need for model training before detection and further processing after feature extraction. Only one set of parameters, namely the window length, embedding dimension, delay and radius of neighborhood, need to be tuned. The complexity of implementing this method is
extremely low. It is easy to be implemented online and automatic, as the calculation expenses is low compare with STLmax based method and in general it can be patient independent.

Visualization of T1 index can be used by clinicians one feature plot summarizes tens of hours of EEG data. The interesting events (seizures) can be easily pin-pointed and traced back. This kind of plots may save neurologists huge amounts of time on searching for epileptogenic focus.

But for automatic seizure detection there are points that need to be improved, especially within the tolerable false alarm range:

1). Clinical seizure detection rate could not reach 100%, which raises the question of why some seizures are missed.

2). When including sub-clinical seizures, the overall detection rate is not that satisfactory. It seems that this method is not appropriate if sub-clinical seizure detection is the interest.

3) It was observed that T1 peaking behavior is not only generated by in-seizure ECoG, but also some burst of artifacts-carrying ECoG sections. Like the T1 peak cluster between seizure 3 and seizure 4 in Figure 4-10. These T1 peak clusters contribute to the false alarm. A pre-processing step before feature extraction might help to reduce the present false alarm rate.

For problem (1) and (2), cross check between feature and original signal shows that whenever in-seizure ECoG has lower amplitude, even if it oscillates in higher frequency, the T1 feature is not as obvious as those seizures with high amplitude, this is consistent with the method analysis. And most of the sub-clinical seizures and few missed clinical
seizures oscillate in lower amplitude, there are feature peaks when this kind of seizure occur, it is just that to catch them, the general detection has to tolerate high false alarm rate. The trade off depends on user, from the method itself there are no room to improve on these points.

The present work was based on limited 5 sets data, more data sets need to be tested to verify the applicability when new data sets are available in the future. Besides it is well known that depth EEG contains much less artifacts compare with scalp EEG. It is also very enticing to test the method on scalp EEG, and observe how it performs, if scalp data is available.
CHAPTER 5
ARTIFACT ELIMINATION BY SOURCE SEPARATION

To reduce the false alarms in the detection techniques, as explained in Chapter 3, caused by artifacts, the new techniques of blind source separation (BSS) methods are considered. An important assumption of BSS is that the observations are generated by statistically independent unknown sources [130]. This assumption is sufficiently met as long as the noise and/or artifacts can be represented as spatially distinct “sources” and if the set of these “sources” is independent of the group of ECoG generating sources. Another requirement needed for many BSS algorithms is that the number of sensors (ECoG source) is larger than or equal to the number of sources (in this case the number includes the number of noises and/or artifacts). While it is reasonable to assume that the first of these two conditions is met, it is not currently known whether the second condition is met since the effective number of spatially distinct sources in the ECoG is an open issue.

Assuming these two requirements are met (several other requirements for the BSS model, which are likely to be met for this data, can be found in [123]), the outputs of a properly trained BSS processor can be divided into a set of outputs which contain noise and/or artifact and another set of outputs that contain a mixture of the brain signals. Once this is done, the artifacts and/or noise components have to be identified. As the de-mixing matrix can be learned through the algorithm, it is trivial to remove the effects of the noise and/or artifacts from the original observations through subtraction.
Further assumption of any BSS algorithm is the mixing model of sources. Considering the time constants of ECoG recording electrodes relative to those unknown sources, a simple instantaneous mixing is assumed to simplify the question. As although other assumption had been assumed [114] and corresponding algorithm have been tested, there is no description of whether the complex model has advantage in terms of artifacts sources decomposition over instantaneous model in the reference. But whether this instantaneous mixing assumption is valid or not is another open issue.

**The Blind Sources Separation Algorithm**

With the above assumptions, a data efficient MRMI-SIG [131], was applied. This algorithm finds independent components in two steps: the observed mixtures are first whitened to reduce the number of tuning parameters; then rotation angles are determined through minimization of Renyi’s mutual information among the output components via a stochastic information gradient tuning. The algorithm estimates entropy directly using Parzen windowing estimator via. Gaussian kernel, the merit of this algorithm compare with other BSS/ICA algorithm can be found in reference [131].

Figure 5-1 is the block diagram of MRMI-SIG algorithm incorporating the assumptions above. Notations used in the block diagram are consistent with the relevant equation that follows. Further implementation details can be found in references [131] [132].

Where \( s_i(n), H, z_i(n) \) are unknown independent sources (without lose of generality and for convenience, sources will be assumed zero mean), unknown mixing matrix, known observed; \( W, R, y_i(n) \) are whitening matrix, de-mixing matrix, decomposed independent components; \( i=\{1,2\ldots N\} \) is the number of sources (in space); \( n=\{1,2\ldots L\} \) is the observed time sequence (time index). Define column vectors
\( \mathbf{S}_n = [s_1(n), s_2(n) \ldots s_N(n)]^T, \mathbf{Z}_n = [z_1(n), z_2(n) \ldots z_N(n)]^T, \mathbf{X}_n = [x_1(n), x_2(n) \ldots x_N(n)]^T, \mathbf{Y}_n = [y_1(n), y_2(n) \ldots y_N(n)]^T, \) and matrix \( \mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \ldots \mathbf{Z}_n], \mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n], \mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \ldots \mathbf{Y}_n]. \) Then vector \( \mathbf{Z}_n = \mathbf{H}_n \mathbf{S}_n, \mathbf{Z}_n = (\mathbf{W} \mathbf{R})^T \mathbf{Y}_n, \mathbf{Y}_n = \mathbf{R} \mathbf{X}_n \) if \( \mathbf{W} \) and \( \mathbf{R} \) can be found to make \( y_i(n) \) independent in certain sense, then the unknown sources could be traced back.

Figure 5-1. MRMI-SIG algorithm block diagram.

- \( \mathbf{W} = \Phi \Lambda^{1/2} \) where \( \Phi, \Lambda \) are the corresponding eigenvector and eigenvalue matrix of the autocorrelation matrix of \( \mathbf{Z} \), \( i.e. \ E[XX^T] = I_N \) (\( N \) x \( N \) identity matrix). It is known that, for instantaneous mixtures, the BSS problem can be decomposed into whitening by \( \mathbf{W} \) then followed by an orthogonal matrix rotation [133]. So the de-mixing matrix \( \mathbf{R} \) is constrained to be pure rotation. The algorithm needs to search the rotation angle space looking for the angles that make output \( y_i(n) \) maximum independent based on certain criteria.

The criteria of MRMI-SIG is minimizing Renyi’s mutual information between output components \( y_i(n) \) via stochastic information gradient. Renyi’s mutual information is defined by Eq.5-1.

\[
I_{R\alpha}(y) = \frac{1}{1 - \alpha} \log \left\{ \frac{\prod_{i=1}^{N} f_i(y_i)^{\alpha}}{\prod_{i=1}^{N} f_i(y_i)^{\alpha-1}} \right\} \tag{5-1}
\]
Eq. 5-2 is the difference between marginal entropy of output components and joint entropy.

\[
H_R(\alpha)(y) = \frac{1}{1-\alpha} \log \int_{-\infty}^{\infty} \prod_{i=1}^{N} f(y_i)^{\alpha} dy_i
\]

It has been analyzed in [132] that the minimization of Renyi’s mutual information is equivalent to minimize the sum of marginal entropy when sources are statistically independent and at most one component has Gaussian distribution. So the algorithm defines a cost function \( J \) (Eq. 5-3) and uses the quadratic form of Renyi’s entropy \( \alpha = 2 \) (Eq. 5-4 and 5-6). For the pdf estimation, Parzen windowing with a Gaussian kernel (Eq. 5-5) is used.

\[
J = \sum_{i=1}^{N} H_R(\alpha)(y_i)
\]

\[
H_R(\alpha)(y_i) = -\log \int_{-\infty}^{\infty} f(y_i)^{\alpha} dy
\]

\[
f(y_i) \approx \frac{1}{L} \sum_{j=1}^{L} G((y_i - y(j)), \sigma^2)
\]

\[
H_R(\alpha)(y_i) = -\log\left(\frac{1}{L^2} \sum_{k=1}^{L} \sum_{j=1}^{L} G((y_i(k) - y_i(j)), 2\sigma^2)\right)
\]

\[
\frac{d}{d \theta_{ij}} \sum_{k=1}^{N} H_R(\alpha)(y_k)
\]

\[
= -\sum_{k=1}^{N} \sum_{m=1}^{L} \sum_{n=1}^{L} G((Y_k(m) - Y_k(n)), 2\sigma^2) * (Y_k(m) - Y_k(n)) * (\nabla R_{ij})_k^T * (X(m) - X(n))
\]

\[
= -\sum_{k=1}^{N} \sum_{m=1}^{L} \sum_{n=1}^{L} G((Y_k(m) - Y_k(n)), 2\sigma^2) / \sum_{m=1}^{L} \sum_{n=1}^{L} G((Y_k(m) - Y_k(n)), 2\sigma^2)
\]
To minimize the cost function through iterations using a gradient approach, the updating equation becomes Eq.5-7, and Eq.5-8 is a simplified instantaneous version of Eq.5-7, which reduce the algorithm complexity from O(L^2) to O(L) and it is called the stochastic information gradient [132].

\[
\Delta L \theta_{ij} = \\
\frac{\sum_{k=1}^{L} G((Y_k(n) - Y_k(n-1))/2\sigma^2) * (Y_k(n) - Y_k(n-1)) * (\nabla R_{ij})_k^T * (X(n) - X(n-1))}{\sum_{k=1}^{L} G((X_k(n) - X_k(n))/2\sigma^2)}
\]  

(5-8)

The construction of the optimal matrix \( R \) is achieved from \( N(N-1)/2 \) Given’s rotations matrices \( R_{ij} \), which is equal to identity matrix \( I_N \) with replacement

\[
I_N(i,i) = \cos(\theta_{ij}), \quad I_N(j,j) = \cos(\theta_{ij}), \quad I_N(i,j) = \sin(\theta_{ij}), \quad I_N(j,i) = \sin(\theta_{ij}), \quad i = \{1,2,...,N-1\}, \quad j = \{i+1, i+2,...,N\}. \]

An example of \( N=4 \) case, the \( R \) calculation is illustrated by Table 5-1.

<table>
<thead>
<tr>
<th>Table 5-1. Rotation matrix calculation.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th></th>
<th>( i = 1 )</th>
<th>( i = 2 )</th>
<th>( i = 3 )</th>
<th>( i = 4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( i = j = 2 )</td>
<td>( i = j = 3 )</td>
<td>( i = j = 4 )</td>
<td></td>
</tr>
</tbody>
</table>
| \( R \)          | \[
\begin{bmatrix}
\cos(\theta_{12}) & \sin(\theta_{12}) & 0 & 0 \\
\sin(\theta_{12}) & \cos(\theta_{12}) & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}\] | \[
\begin{bmatrix}
\cos(\theta_{13}) & 0 & \sin(\theta_{13}) & 0 \\
0 & 1 & 0 & 0 \\
\sin(\theta_{13}) & 0 & \cos(\theta_{13}) & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}\] | \[
\begin{bmatrix}
\cos(\theta_{14}) & 0 & \sin(\theta_{14}) & 0 \\
0 & 1 & 0 & 0 \\
\sin(\theta_{14}) & 0 & \cos(\theta_{14}) & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}\] |
|                  | \( i = 2 \) | \( i = 3 \) | \( i = 4 \) |
| \( R \)          | \[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & \cos(\theta_{23}) & \sin(\theta_{23}) & 0 \\
0 & \sin(\theta_{23}) & \cos(\theta_{23}) & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}\] | \[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & \cos(\theta_{24}) & \sin(\theta_{24}) & 0 \\
0 & \sin(\theta_{24}) & \cos(\theta_{24}) & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}\] | \[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & \cos(\theta_{34}) & \sin(\theta_{34}) & 0 \\
0 & \sin(\theta_{34}) & \cos(\theta_{34}) & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}\] |
|                  | \( i = 3 \) | \( i = 4 \) |
| \( R \)          | \[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & \cos(\theta_{33}) & \sin(\theta_{33}) & 0 \\
0 & \sin(\theta_{33}) & \cos(\theta_{33}) & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}\] | \[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & \cos(\theta_{34}) & \sin(\theta_{34}) & 0 \\
0 & \sin(\theta_{34}) & \cos(\theta_{34}) & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}\] | \[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & \cos(\theta_{44}) & \sin(\theta_{44}) & 0 \\
0 & \sin(\theta_{44}) & \cos(\theta_{44}) & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}\] |

The derivative of \( R \) with respect of \( \theta_{ij} \) denoted as \( \nabla R_{ij} \) equals the corresponding \( ij \) identity matrix in \( R \) replaced with zero matrix \( O_N \) with elements \( O_N(i,i) = -\sin(\theta_{ij}), \quad O_N(j,j) = \cos(\theta_{ij}), \quad O_N(i,j) = -\sin(\theta_{ij}), \quad O_N(j,i) = \cos(\theta_{ij}) \). And \( (\nabla R_{ij})_k \) is the \( k \)th column of \( \nabla R_{ij} \).
Simulation

To check the performance of the MRMI-SIG algorithm on the mixture of noise and “pure” EEG signal, the algorithm is applied to a synthetic signal first. A section of noise component decomposed by ICA and Laplacian noise are mixed, by a randomly generated mixing matrix, with a section of a background evolving into seizure respectively. Then the mixture is processed by MRMI-SIG.

The original signals, the two mixtures, the decomposed sources and the difference between true signal and its estimation are plotted in Figure 5-2. Sub-panel g) shows the difference, which is taken between the signal component and the original, after the signal component is scaled down to the original signal range. The difference is in the range of 10% of the original signal, and it is regarded as acceptable performance. Some additional simulations produced similar results, so we proceed to test the algorithm on real signal.

The study of the T1 index in Chapter 4 found that some high amplitude clusters generate many false alarms and the corresponding raw EEG was recognized as artifacts. Three such sections (10.24 sec.) each generating large T1 index, but from different types of artifacts, are picked from the marked ECoG. Figures 5-3, 5-4 and 5-5 depict these noisy ECoG sections, their decomposition and the “cleaned” ECoG after artifacts.
correction. The middle column of each plot are the decomposed components by the MRMI algorithm, the artifact components are visually picked (e.g. in Figure 5-3 the 4th component is regarded as the artifact component etc.), and their contribution are subtracted from the original recording using the obtained de-mixing matrix $\mathbf{WR}$.

$$\text{cleaned signal} = \begin{bmatrix} z_1(1) & z_1(2) & \cdots & z_1(L) \\ z_2(1) & z_2(2) & \cdots & z_2(L) \\ \vdots & \vdots & \ddots & \vdots \\ z_N(1) & z_N(2) & \cdots & z_N(L) \end{bmatrix} - \begin{bmatrix} \mathbf{WR}^{-1}(1,k) \\ \mathbf{WR}^{-1}(2,k) \\ \vdots \\ \mathbf{WR}^{-1}(N,k) \end{bmatrix} \ast \begin{bmatrix} y_k(1) \\ y_k(2) \\ \vdots \\ y_k(L) \end{bmatrix}$$

Notation here is in consistent with Figure 5-1. Assuming at the output of the algorithm, the $k^{th}$ component is the artifacts, we have $N$ observations at any time instance. $\mathbf{WR}^{-1}(\ast, k)$ is the $k^{th}$ column of the inverse of de-mixing matrix (whitening * rotation). $L$ is the observation length in time.

In order to reject the artifact automatically on large amount of data, the artifact component has to be differentiated from others by some additional algorithm that is able to recognize the artifacts automatically.

**Conclusion of Blind Sources Separation**

In this chapter the MRMI-SIG algorithm has been tested on synthetic data and on some short sections of real EEG data. The performance is acceptable for the purpose of reducing the false alarms in our detection scheme. The acceptance is in the context of T1 feature extraction point of view, for all the tested artifacts section simulated in this chapter, after knock out the artifacts component, T1 feature has been extracted for the cleaned signal, the scale of these T1 feature are all similar to clean background, so this suit our purpose to reduce false alarms. The issue is how to apply MRMI-SIG automatically on large amount of data. A method is proposed in the next chapter.
Figure 5-2. Artificial mixed signal separation example. (a) noise sources, (b) signal sources (ECoG), (c) random mixture 1, (d) random mixture 2, (e) decomposed component 1, (f) decomposed component 2, (g) difference between signal and ICA estimation.

Figure 5-3. Electrode pop artifact. Columns from left to right are ECoG, de-mixed components, after artifact correction. The 4th component in the middle column is subtracted.
Figure 5-4. Line artifact. Columns from left to right are ECoG, de-mixed components, after artifact correction. The 6th component in the middle column is subtracted.

Figure 5-5. Muscle artifact. Columns from left to right are ECoG, de-mixed components, after artifact correction. The 6th component in the middle column is subtracted.
CHAPTER 6
AUTOMATIC ARTIFACTS CORRECTION BY TEMPLATE MATCHING

Introduction

The idea of finding artifact components in the EEG by template matching is reasonable because artifacts are normally catalogued by EEGers/neurologists as muscle, glossokinetic, eye movement, ECG, pulse etc. Each corresponds to a different waveform but their number is limited in particular because the T1 index is sensitive to a few at least in this five set of data (for example T1 is not sensitive to EKG).

The reason to choose artifacts template matching instead of background template matching is that the EEG has many nuances not well understood and that affect the clinical information. Therefore, it is much safer to apply the template matching to the artifacts.

Conventional template matching uses AR or ARMA in conjunction with generalized likelihood ratio test, or matched filtering for short events [142]. These methods are not suitable to accommodate the complexity of artifacts and their performance deteriorates with the length increase of the segment due to the averaging inherent in their linear nature.

The centerpiece proposed for matching a BSS/ICA decomposed component with the artifact template is the utilization of the self-organizing map (SOM) [135]. As an infrastructure, the SOM is able to model the trajectories in phase space with the added advantage of discretization and neighborhood preservation. The SOM output space is a discrete projection of the high dimensional phase space which preserves neighborhoods
relation in the training samples. Therefore different trajectories in reconstruction space will be mapped into a different set of winning processing elements (PEs). The idea is to represent each trajectory by the histogram of the winning PEs, which will provide a signature in discrete and small dimensional space with which other histograms created from other trajectories can be compared against in a metric sense. The Kullback-Leibler divergence is proposed to estimate the distance between the two histograms, as it always exists in the low dimensional output space of the SOM and the estimation can use discrete probabilities due to the discretization operated by the SOM. Alternatively, one can compare the transition matrices of the winning PEs for each trajectory using a metric resembling the correlation coefficient since the SOM preserves neighborhood relations. These operations will be explained next.

**Template Building**

The self-organizing map is a neighborhood preserving vector quantizer that has been widely used in neural computing. The idea is to represent high dimensional clouds of points by a priori fixed number of processing elements (PEs) organized in 1D or 2D neighborhoods. There is a large volume of literature dedicated to the SOM and reference [134] is a good place to have a general view and [135]-[137] are good places to get started, especially [137] a textbook with a short chapter discuss the algorithm.

The SOM training algorithm is taken from [137]. There are three training steps: (1) the competition step selects the representing PEs, (2) in the cooperation step the winner PEs decides their neighborhood, (3) and finally in weight adjustment step the winner PE and its neighborhood are fine tuned. The algorithm is mathematically abbreviated in APPENDIX B, and details can be found in the cited references.
The stopping criterion for training is a preset value of the gradient of mean square error (MSE) between training samples and winning PEs weights from epoch to epoch. When the average gradient reached 0.001 in the past 100 epochs, the training stopped.

In order to train a SOM to represent a given scalar time series \(x(i)\), for \(i = 1, 2, \ldots, M\), where \(i\) is the time index, assume the time series is generated by a D dimensional dynamical system. From Takens’ embedding theorem, the corresponding \(m\)-dimensional phase space \((m \geq 2D + 1)\) can be constructed from the time series \(X_k = [x(k), x(k+L), x(k+2L), \ldots, x(k+(m-1)L)]^T\), where \(L\) is the time delay. The delay vector sequence \(X_k \{k = 1, 2, \ldots, K, \text{ and } K = M-(m-1)L\}\) constitutes a trajectory in phase space. This phase space is mapped onto a two dimensional SOM represented by \(N (N<<K)\) PEs each with weight \(W_j = [w_{j1}, w_{j2}, \ldots, w_{jm}]^T, j = 1, 2, \ldots, N\).

Given a trained SOM, the template corresponding to the segment of the time series is generated in the following two way: (1) calculate the distribution of each winner PE, \(Q\), over the SOM as the ratio between the numbers of times the \(j^{th}\) PE is fired divided by the number of training samples; (2) calculate the transitions distribution between \(i^{th}\) and \(j^{th}\) PEs, \(Y_{ij}\) as the ratio of the sample transitions from \(i^{th}\) PE to \(j^{th}\) PE divided by the total number of transitions (K-1).

Either the SOM histogram \(Q\) or the transition matrix \(Y\) extracts a template for the known segment of the time series. Figure 6-1 diagrammatically shows the procedure to create these quantities. Figure 6-2 is an example of a 2D SOM trained with 2000 training samples of the Lorenz system (the x variable, embedding dimension 3, delay 2). The template has 100 PEs. Figure 6-2 (a) plots in 2D space the winning PEs’ for the first 500 training samples, while (b) shows the histogram of the PE winners also in 2D space. This
histogram represents in the 2D space a projection of the density of points in the trajectory in the original phase space. Notice that the SOM discretized the phase space, which simplifies the distance calculation latter, but this discretization also introduces quantization errors.

Figure 6-1. Density distribution $Q_j$ and transition distribution $Y_{ij}$ of a trained template.

Figure 6-2. The first two dimension of Lorenz (x) SOM template. (a) Voronoi cell center $W_{js}$ and training samples, (b) Sample distribution $Q_{js}$ in each corresponding Voronoi cell.
Similarity Measures

After the template for a time segment is build, the goal is to find similar segments in the remaining portion of the time series. Two similarity measures are considered: (1) the Kullback-Leibler divergence measuring the delay vector distribution similarity; and (2) the Correlation Coefficient measuring a first order dynamic transition similarity.

The phase space of a window of the test data is created using the same embedding parameters of the training samples, and a histogram P of the winning PEs over the SOM is created.

The similarity between the template and the test window histogram is computed by Eq.6-1.

$$KL(P, Q) = \sum_{i=1}^{N} p_i \log \left( \frac{p_i}{q_i} \right)$$

The distance between the transition distributions is defined as the sum of the element wise product of the transition matrix entries, divided by the product of the sum of the square of each matrix elements:

$$CC(Z, Y) = \frac{1}{N} \times \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} Z_{ij} * Y_{ij}}{\sqrt{\left( \sum_{i=1}^{N} \sum_{j=1}^{N} Z_{ij} * Z_{ij} \right) * \left( \sum_{i=1}^{N} \sum_{j=1}^{N} Y_{ij} * Y_{ij} \right)}}$$

Where N is the number of PEs in the SOM template, Y is the transition matrix of the template, and Z is the transition matrix of the test window, built in the same way as for the template. Effectively this distance is the correlation coefficient between the norms of the two transition matrices. It is based on the Cauchy-Schwartz distance presented in [138] but without the log operation. Notice that if the first order transitions are exactly the
same between the template and the test segment transition matrix, then CC=1; conversely, if the two have no similarity at all then CC=0, so it is intuitive to name this quantity as ‘correlation coefficient’. Note however that CC(Z,Y) is always positive. This expression but can be more precisely interpreted as an inner product distance between two matrices instead of the more conventional Frobenious norm.

**Comparison**

For comparison, we implemented: the box method [83] which is a discrete method to compare two high dimensional phase space data distribution through box counting; and the Diks test [139], which is a continuous similarity measure between two distributions. The box method partition each dimension of the $m$ dimensional phase space into $S$ symbols via equation 6-3 [83], the $m$ dimensional phase space is then partitioned into $S^m$ cubes, the discrete distribution density is calculated by counting the number of points inside each cube, $S$ is finely tuned to gain better performance.

$$0 \leq s_i = \text{INT} \left[ \frac{S(x_i - x_{\min})}{x_{\max} - x_{\min}} \right] \leq S - 1$$  \hspace{1cm} (6 – 3)

Where $x$ is the value of the state at certain dimension, and INT is the integer operator.

While the Diks test use Gaussian kernel estimate the distribution density for given phase space represented by those discrete data points, and the distance between the test and the template is calculated through analysis equations. The kernel size (named bandwidth in the original reference) and the dependence among delay vectors should be tuned before calculation. The parameters used in our comparison will be clarified in the comparison and summary of model data section.
Simulation

The proposed methodology is applied to synthetic time series data first where Box, SOM, and Diks template matching were compared. To conclude the test short sections of real ECoG data were also compared between the Diks test and the SOM based method.

Simulation on Mathematical Model Generated Data

Time series were created from the Mackey Glass and Lorenz models using the Runge-Kutta Method with integration step 0.01, sampling rate 6. A 60,000 samples of Mackey Glass (MG) series with $\tau=30$, initial values 0.9 was generated. The embedding parameters were decided using the method introduced in [124], which yields an embedding dimension of $m=6$, delay $L=2$. The first 5,000 samples of this time series were used to train a MG30 SOM template (this training section was used for all the simulation results reported in this section).

Time series were created concatenating another 55,000 sample segment of the MG30 time series with other time series. We used Mackey Glass series with different $\tau=17, 25, 27, 28$, as well as a Lorenz series $x$ variable with $\sigma=10$ ($x$ variable), $r=28$ ($y$ variable), $b=8/3$ ($z$ variable), initial conditions $x(0)=1, y(0)=0, z(0)=0$. All time series are normalized to the range [-0.5 0.5] before template training and mixing. The results reported here are: (1) Lorenz($x$)-MG30-Lorenz($x$), (2) MG17-MG30-MG17, (3) MG25-MG30-MG25, (4) MG27-MG30-MG27, (5) MG28-MG30-MG28. Figure 6-3 is a short section of the time series MG27-MG30-MG27 for our visual evaluation.

Effect of Embedding Dimension in the Test Accuracy

The first test was the comparison of the effectiveness of the SOM based distance measure presented above with the application of the same KL divergence to the histogram created from the scalar time series samples, which can be thought as an
observation of the multidimensional deterministic dynamic system collapsed onto a single dimension. 5000 samples from MG30 (embedded in a 6\textsuperscript{th} dimensional space as explained above) are used to train a SOM template with 400 PEs, and the same segment generates the 1-D histogram with 400 bins.

Figure 6-3. MG(27)-MG(30)-MG(27) mixture.

Figure 6-4 panel (a) shows the evolution over the test segment of the KL divergence computed with the proposed SOM based method and (b) the one dimension KL computed every 1,000 samples from the histogram.

Figure 6-4. One dimension vs. high dimension Kullback-Leibler divergence.

Observe that the difference between MG17 and MG30 is clearly revealed in high dimensions, but hardly noticeable in one dimension. Therefore we conclude that an
increase to the embedding dimension is crucial to apply this similarity method to time series.

**Effect of Processing Element Number on the Estimated Similarity**

Intuitively, the number of PEs will affect the accuracy of the estimated similarity because the fidelity between the SOM output trajectory and the high dimensional input trajectory is a function of the size of the voronoi cells. For the present application the quantization error affects the ability to distinguish between similar templates, which impose a limit on the accuracy of the method. This consideration may suggest that the higher the number of PEs the better. However, other factors must be taken into consideration because SOM is an adaptive algorithm. In fact, higher number of PEs requires much higher number of samples for proper training, therefore there is an optimal number of PEs to well train the SOM for a given time series window. These aspects were experimentally addressed in this study by creating very similar Mackey-Glass time series generated with similar delay parameters and by adding white Gaussian noise to the time series.

Figure 6-5 simulates the influence of PE numbers on the proposed measures. Subplot (a) and (b) are the KL divergence of two different concatenations; Subplot (c) & (d) are the CC of the corresponding concatenations. We can observe from (a) that when the number of PEs increase, the differentiability of KL measure increases.

Notice that the difference between the matched regions is smaller than in the segments that came from the mismatched time series. There seems to be a saturation of the KL similarity above 14 x 14 SOMs. Panel (c) shows the effect of SOM size on the CC measure, and the trends are the same. When the time series are generated by two distinct dynamical systems (panels (b) and (d)) the difference in KL divergence is higher
and is not so dependent upon the size of the SOM, in particular for the CC measure that exploits the transitions between PEs.

Figure 6-5. Effects of PE number (MG30 is the template). (a) KL divergence of MG25-MG30-MG25 with different number of template PE. (b) KL divergence of Lorenz(x)-MG30-Lorenz(x). (c) CC of MG25-MG30-MG25 with different number of template PE. (d) CC of Lorenz(x)-MG30-Lorenz(x).

**Effects of Noise on the Similarity Measures**

To test how the measure behaves when noise is present, Gaussian noise with different powers was added to the test data. Figure 6-6 depicts the simulation result. The template is a MG30 10x10 SOM. As can be observed, when the signal to noise ratio (SNR) decreases, the differentiability of both measures deteriorates. When SNR reaches –2dB, the CC measure can not differentiate MG30 from MG25 any more. The additive
noise affects more the CC measure than the KL divergence. However, we should note that even when the SNR is 4 dB the separability is noticeable, which is an unexpected result and shows the robustness of the method to noise.

![Graphs showing noise effect on KL divergence and CC](image)

Figure 6-6. Noise effect (10 x 10 MG30 template). (a) KL divergence of MG25-MG30-MG25 with different SNR. (b) KL divergence of Lorenz(x)-MG30-Lorenz(x). (c) CC of MG25-MG30-MG25 with different SNR. (d) CC of Lorenz(x)-MG30-Lorenz(x).

Comparison and Summary of Model Data

The SOM based method is compared with both the box method and Diks’ test.

Figure 6-7 summarizes the three methods on model data. The similarity between MG30 and the other Mackey-Glass family member plus Lorenz (x) are compared. The MG30 template has 20 x 20 PEs for SOM based method, while the box method [83] has 9 bins in each dimension (this number has been tuned to obtain the best result). All the
embedding parameters are the same. And for Diks’ test the training samples used to train
SOM template is the reference in the distance calculation, the kernel size (bandwidth) is
0.06 and the dependence among delay vectors is 18. As there is no counterpart of CC
measure in Diks’ test, the CC measure is illustrated only for SOM based method in
Figure 6-8.

The conclusion we can draw from the simulation is: the SOM based phase space
discretization method performs closer to its continuous counterpart, the Diks’ test. The
SOM based method sensitivity is sufficient to differentiate the difference between most
of the templates of other Mackey-Glass family members, and Lorenz(x), in a similar
pattern as Diks’ test. As expected, the MG30 has larger difference both statistically and
dynamically from MG17 than from MG25, which can also be corroborated by visual
observation. As expected, the SOM quantization affects precision. The SOM based
method can not differentiate between MG29&MG30 and MG28&MG30, but Diks’ test
can.

The box method is able to differentiate between the MG29&MG30 and MG28 and
MG30, but it fails to provide resolution at the other end of the scale, collapsing the
distances of very different time series, unlike the other two methods. The method may
behave with high specificity, but lack of sensitivity.

Simulation on Real Data

To further verify the applicability on EEG, three methods are also tested on some
short sections of real data. Data from patient P93 has been used. The left temporal depth
(LTD) electrode 1 is used for the testing. An in-seizure template was created using 5,000
samples of the seizure recorded by the same electrode but in a different tape. Before
processing, the ECoG data was linearly scaled to [-0.5 0.5] range, the embedding
parameters are $L = 4$, $m = 4$ as has been used in chapter 4, and the SOM template has 400 PEs. Figure 6-9 is a short section of training samples and one of the in-seizure test samples. Notice how different the two events are in the fine detail, but they both can be recognized as seizure segment. Figure 6-10 presents the similarity test result of the three methods in the same segments.

These real data test results corroborate the model test conclusion, that is, the SOM based method is more close to continuous counterpart method than the box method because for the box method the difference between the drop in the seizure section and normal background is much smaller than the other two methods. The box method can even be misleading as shown in the second column of Figure 6-10. In ECoG, the performance of Diks test and of the SOM test are very similar and only a more thorough investigation would show which is better.

**Conclusion**

In this chapter a method to quantify similarity between know time series patterns and time series under testing is proposed. In essence the method is to measure similarity among trajectories of the time series. But instead of doing this in the original reconstruction space (like K-means clustering) or on the time series, SOM is used to project and discretize the trajectory to a 2 D space. Two distance measures are proposed and compared: the KL divergence which is a static measure of the density of points, and the correlation coefficient measure that is also sensitive to the time evolution of the trajectory. The performance of this approach is compare with Box method and Diks’ distance in both synthetic time series and real brain data. The overall conclusion is that the proposed methods perform on the same level as the Diks’ test, which encourages more systematic research.
Figure 6-7. Divergences between MG30 template and the artificial mixed time series. (a) Kullback-Leibler criteria via SOM template method. (b) Diks distance. (c) Distribution distance via box discretization method.
Figure 6-8. First order dynamic correlation coefficients between MG30 template and the artificial mixed time series

Figure 6-9. Seizure ECoG samples
Figure 6-10. Comparison of two ECoG segments.

One of the reasons to develop this method is to alleviate the computation burden to estimate Diks’ test for on-line processing. Since the SOM is trained off-line, the only computation to be performed online is the calculation of the histogram on the SOM and the calculation of the distances, but since everything is discritized by the SOM this step is reduced to summations instead of continuous variable calculations. This is a great savings, and may justify the slight degradation in performance for real world problems.

Many interesting things were observed during this research. First, it was surprising that the KL divergence between the histogram of the time series and the 2D SOM was so markedly different. It is known that similarity on the time series histogram is not very discriminate, but it is surprising to see the improvement on the histogram reconstructed in 2D space after the SOM based projection (that preserves local topology). This is great
news since distances in 2-D spaces are easy to compute with few samples. It is also interesting to analyze further the effect of dimensionality of trajectories in the accuracy of Diks test and the SOM based method. Here moderately small reconstruction spaces were used and so Diks test works well, in fact better than the SOM similarity distance due to the intrinsic quantization noise of the discretization. However, Dicks’ method breaks down in high dimensional spaces, so it is important to verify if the distances estimated in the projected SOM 2D spaces are more reliable.

To measure the static distribution similarity, the Kullback-Leibler divergence (any other measure that is based on probability density distribution will also be applicable) is used while the first order dynamic similarity is measured by the state transition correlation coefficient. The reason of considering two similarity measures is that static similarity may be realized by different dynamics. Dynamic similarity measures similarity in the transitions, therefore it is more specific. To illustrate this point, a two dimensional SOM template for uniform random variable was trained, then a Lorenz(x)-Random-Lorenz(x) test signal is compared with the template. Figure 6-11 shows that according to KL measure the test random signal is very close to the template, but according to CC measure only 18% of the transition route matches the template. However, noise will affect more the transitions than the PE winners, so static similarity is more robust to noise. It is not clear at this point which provides better similarity measure for the ECoG.

The SOM based similarity distance shines in speed of calculation. The most time consuming operation is the SOM training and it is done offline. Once the templates are built the computation is limited to discrete KL computations or CC between transition matrices of winners, which are basically counting algorithms. The calculation time of the
distance \( L(x) - MG(30) \) in figure 6-7 was tracked on a Pentium 4 PC with Matlab 6.5. The MG(30) template training spend 2758 seconds, generation of the KL distance by template matching needs 27 seconds, while Diks test need \( 1.7475 \times 10^5 \) seconds. Therefore, the proposed method can be easily implemented in real time on small computers. The SOM based similarity measures sacrifice some differentiability with respect to the Diks test by discretization, but they simplify enormously the calculations and are very appropriate for practical time series work.

Figure 6-11. Relationship between KL and CC measures
CHAPTER 7
PERFORMANCE TEST

Templates and Test

The proposed preprocessing method is tested on three sets (P93, P92, and P256) of data where it is observed that artifacts create false T1 high amplitude clusters that will be misinterpreted as in-seizure by our detection algorithm. Figure 7-1 to 5 is the full feature spaces (without preprocessing) of the 5 data set.

Figure 7-1. P93 feature. (Top to bottom LTD (5 channel), RTD(5), LST(4), RST(4), LOF (4), ROF(4), RL(4))
Figure 7-2. P148 feature. (Top to bottom LTD (6) RTD(6) LOF(4) ROF(4) LST(4) RST(4))

Figure 7-3. P92 feature. (Top to bottom LTD (6) RTD(6) LST(4) RST(4) LOF(4) ROF(4))
Figure 7-4. P185 feature. (Top to bottom LTD(6) RTD(6) LST(4) RST(4) LOF(4) ROF(4))

Figure 7-5. P256 feature. (Top to bottom LOF(4) ROF(4) LAT(4) RAT(4) LTD(6) RTD(6) RPT(4))
Figure 7-6 is the detail block diagram of our proposed seizure detection methodology. Three artifact templates are built. The artifacts come from figure 7-1, 7-3 and 7-5 where the corresponding tapes are zoomed to have a closer view. And the equivalent artifacts sources are listed in figure 7-7.

Figure 7-6. Preprocess block diagram.

The multiple channels of T1 high amplitude clusters in figure 7-2 and 7-4 are not all caused by artifacts but by frequent spike and waves that generate the high amplitude clusters in these two plots as depicted in the zoomed Figure 7-8.

The SOM templates, trained by the data depicted in Figure7-7 artifacts sources, are applied to the output of the BSS algorithm. Since the BSS algorithm has a gain uncertainty and the SOM is a magnitude sensitive algorithm, the decomposed component of each window is scaled down to the range of the trained templates. The similarity threshold is set based on the simulations.
Figure 7-7. Artifacts sources from BSS decomposition

Figure 7-8. Frequent spikes that generate some of the high amplitude T1 index

Figure 7-9 is a plot to show the situation where artifacts switch on and off. Top panel is a short section of the T1 index from p92 set, without preprocessing. It can be observed that an artifact switches on and off, as reflected by the T1 peak profile. Second panel shows two Kullback-Leibler divergences between components and templates, it can
be seen that when the T1 peak appears one of the KL divergence drops, meaning that a matching has been found, and the artifact will be corrected.

Figure 7-9. Artifact burst and BSS-SOM preprocessing.

Figure 7-10 is a plot to show the T1 index of one channel from one tape of p92 set. There are three burst of artifact which creates high amplitude T1 clusters. The dotted line indicate features without preprocessing, the solid line are features after preprocessing. After preprocessing two T1 clusters caused by artifact return to background level, but one of them still remains. Apparently the preprocessing failed to perform in that section. Cross checking the original signal and BSS decomposition show that the original ECoG recording has both baseline wandering artifacts and muscle artifacts, and the BSS algorithm group the two artifacts together into one component. Because the SOM based template matching method is a magnitude sensitive method, the baseline wondering causes the artifacts component to deviate from template i.e. KL divergence greater than
threshold, so the artifact is not properly corrected. Baseline correction before BSS will be helpful to solve this problem. This situation is discussed in the conclusion section.

Figure 7-10. Comparison of T1 feature with and without artifacts rejection

Figure 7-11, 7-12 and 7-13 show the ROC of the detection performance comparison for the three data sets. It can be seen that in all cases the false alarms are reduced, but not reduced to zero. In fact the ROC plots raise to large detection values earlier (i.e. causing fewer false alarms), but the gain is not uniform across patients, perhaps indicating that not all the recordings contain these artifacts. As importantly, we see that there is no effect in the detection performance across all the patients tested.

Figure 7-11. P92 ROC
Conclusion

The proposed preprocessing method has been tested in this chapter and the result shows that the general performance depends on whether the artifacts sources have been properly separated. As long as the artifacts sources can be found by the BSS algorithm, the artifacts switch on and off and BSS algorithm channel switching seems to be handled by the SOM based template matching. Like any algorithm, BSS is build upon assumptions. For long stretch of EEG recording, the odds are very high that the conditions for the algorithm to perform well might be violated.

For example in Figure 7-12, the first burst of high amplitude T1 cluster is caused by baseline wandering. Figure 7-14 shows a window of 10 seconds original raw ECoG from this section, the artifacts source could not be properly separated by directly apply BSS to the signal. Baseline wonder is easily filtered by linear methods. If the baseline is removed before BSS algorithm, the decomposition is plotted Figure 7-15 and corrected signal in Figure 7-16. Observing the 28th channel of Figure 7-16, it can be seen that artifacts have been corrected properly. The conclusion is that the proposed methodology provides a facility to eliminate artifacts influence on seizure detection performance, but in general it is unrealistic to expect a zero false alarms in real applications due to the unconstrained signal processing environment.

As indicated at the beginning of this chapter, there are false alarms that are generated by frequent spikes and waves. Technically they are brain signal not artifacts. In order to decompose them properly, the assumptions of BSS algorithm as well as the physical explanation corresponding to each decomposed component have to be reconsidered carefully. For the present seizure detection task, unless a more complex decision statistics after feature extraction is considered, it is not possible to differentiate
seizure caused by T1 index or caused by frequent spike waves. Preprocessing is not helpful in this situation.

Figure 7-12. P93 ROC

Figure 7-13. P256 ROC

Figure 7-14. ECoG baseline wandering. (10 seconds window)
Figure 7-15. ICA Components after baseline removal.

Figure 7-16. Compare of original ECoG (red) and baseline corrected artifacts corrected (blue)
CHAPTER 8
APPLICABILITY TO SCALP EEG

Scalp EEG is usually taken in the first stage of seizure evaluation, and it is a non-invasive technique. Most of the analysis for epilepsy diagnostic is done in the EEG, not in the ECoG, therefore the clinical applicability of the proposed method will be a tremendous asset for automated EEG analysis if seizure detection performance remains at the same level. So it is very appealing to test the applicability of the proposed method on scalp data. The recurrence time statistic feature extraction, BSS artifacts decomposition and SOM based template matching have been tested on four sets of data.

Data

The detailed information of the four study sets is listed in table 8-1. The EEG was collected by the 10-20 standard set up, each with 22 recording electrodes. The process of digitization and study set formation is the same as the organization of the ECoG studies.

Detect Seizure Directly from Feature

The features are extracted directly from the raw EEG, Figure 8-1, 8-2, 8-3 and 8-4 are the feature space. The parameters used to extract these features were tuned to get better contrast between in-seizure and background. S148 was obtained by the same parameter as all the ECoG data (radius 0.0039, delay 4, dimension 4, window 2048), and the remaining three set were obtained by this parameter set (0.002, 7, 5, 2048). Direct threshold is applied to the feature and Figure 8-5 is the detection performance.

On average about 85% of the seizure can be detected when 1 false alarm per hour are tolerated. It can be seen that feature direct detection achieves decent performance on
three sets of data (decent in the sense of detection rate vs. false alarm relative to all the results in the published papers reviewed in chapter 2). The S203 set contributes the most to the average false alarm rate. From feature plot of S203 we can see that the contrast between in-seizure feature and background is not as high as other three data sets, so to detect more seizures threshold has to be lower, thus higher false alarms.

<table>
<thead>
<tr>
<th>ID</th>
<th>Sex</th>
<th>Age</th>
<th>Recording length (hours)</th>
<th>Number of seizures (sub-clinical)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S039</td>
<td>F</td>
<td>21</td>
<td>29</td>
<td>11</td>
</tr>
<tr>
<td>S148</td>
<td>M</td>
<td>18</td>
<td>37</td>
<td>11</td>
</tr>
<tr>
<td>S203</td>
<td>M</td>
<td>40</td>
<td>22</td>
<td>11</td>
</tr>
<tr>
<td>S240</td>
<td>F</td>
<td>32</td>
<td>36</td>
<td>11(1)</td>
</tr>
<tr>
<td>Summary</td>
<td></td>
<td>124</td>
<td></td>
<td>44</td>
</tr>
</tbody>
</table>

In set S240 the majority false alarms are caused by artifacts feature peak clusters, so the raw EEG are cross checked in the next section to see whether BSS can be used as an automatic preprocessor.

**Performance of BSS**

Sections of raw EEG which correspond to artifacts feature clusters in Figure 8-2 are tested. Figure 8-6 is a 10 seconds scalp EEG corresponding to artifacts causing a false alarm and containing several types of artifacts. Figure 8-7 is the decomposition. It seems that by visual observation the eye movement is isolated primarily to components 1, 2, and 10, 14, 15 and 16 for the muscle.

Figure 8-8 is the artifacts corrected EEG, from the figure it seems that still there are artifacts presented in certain channels, the correction is not fully accomplished.
Figure 8-1. S148 feature space.

Figure 8-2. S240 feature space.
Figure 8-3. S203 feature space.

Figure 8-4. S039 Feature space.
Figure 8-5. Detection performance direct from feature.

Figure 8-6. Window of scalp EEG.
Figure 8-7. ICA components.

Figure 8-8. “Cleaned” EEG
Many other sections of artifacts ridden EEG have been tested, the results are similar. It is not clear whether the artifacts have been decomposed properly. Without artifacts sources, the training of template can not be accomplished.

So the conclusion is: in scalp EEG the artifacts is much more complex than in ECoG, the method developed for ECoG cleaning is applicable to scalp EEG.

For scalp EEG, after seizures are detected, a very important function of EEG is to lateralize seizure focus, but often time when seizure ushers in muscle artifacts concurrent to obscure the signal, and make it impossible to deduct any useful information, although the seizure can be detected it is not helpful for further analysis.

Physically muscle artifact is independent of true EEG signal at the beginning of seizure and during seizure, so the basic condition for blind source separation algorithm still hold. If any BSS algorithm could decompose these around seizure muscle artifact, the application has tremendous significance.

Figure 8-9, 8-10, 8-11 and 8-12 are the test result on signals where the corresponding seizure has been detected, but is obscured by muscle artifacts.

Figure 8-9 is the muscle ridden EEG 30 seconds before seizure, Figure 8-10 is the components after apply BSS algorithm to the signal. It is not clear that which component is the muscle artifact.

Figure 8-11 is the muscle ridden EEG 10 seconds into seizure, Figure 8-12 is the components after apply BSS algorithm to the signal. It is not clear either that which component is the muscle artifacts source either.
Figure 8-9. Thirty seconds before seizure EEG.

Figure 8-10. Components of Figure 8-9.
Figure 8-11. EEG beginning 10 seconds into seizure.

Figure 8-12. Components of Figure 8-11.
Conclusion

The proposed SOM-BSS preprocessing method for ECoG needs template to function, but visually from the result of BSS decomposition, it is not clear which component is the artifact source. The BSS-SOM method closely depends upon the proper separation of artifacts from the EEG sources, and when this step fails the method breaks down. BSS decomposition around the seizure section generates similar result, although technically the assumption for BSS algorithm to function still holds. The very intriguing question is why muscle artifact can not be properly decomposed in the scalp data.

Figure 8-13. EEGLAB drop down menu

To deal with artifacts in scalp EEG is a much complex issue, as the variety of artifacts is much more than in depth EEG. Consider the issue from signal processing point of view, artifacts are signals generated from those mechanism that is not the interest of seizure detection, each mechanism is a unique system has its own characters which are reflected in the artifact it generated. So it comes as no surprise that in their EEGLAB analysis tool box, reference [140] developed a set of criteria to help the user reject artifact
component after BSS decomposition, as can be seen in Figure 8-13 the drop down manual.

For the purpose of reducing false alarm for seizure detection from EEG in general methods better BSS need to be developed, and this will be addressed in the next chapter.
CHAPTER 9
SCALP SEIZURE DETECTION AND FUTURE RESEARCH

**Practical Seizure Detection**

Practical seizure detection method is expected to be simple avoid training and the detection rate high with few false alarms. The T1 index meets almost all of these requirements except that it is sensitive to some artifacts. Since the BSS-SOM based preprocessing is the bottleneck of the full signal processing chain, when it does not perform well as in the scalp EEG, false alarm rate increase. To reduce the false alarms a wavelet filter can be implemented to preprocess the data and attenuate artifacts. Conceptually through a bank of filters, the EEG signal can be decomposed into detail sub-bands, to keep the majority band that constitute a seizure, and attenuate the other bands. The advantage of wavelet is that it is model free. Of course it is a compromise because the seizure activity will overlap the artifact activity in some of the sub bands. However, it may be more practical than BSS method, which tends to work well or not at all. Following this reasoning we propose wavelet preprocess to scalp EEG.

Based on the frequency character of the seizure Figure 9-1, it can be seen that the major seizure energy lies below 50 Hz. So a five level wavelet decomposition scheme is illustrated in Figure 9-2. For 200 Hz sampling frequency, the decomposed approximation coefficients correspond to (a) [0 to 3.125] Hz detail coefficients (b) (3.125 to 6.25] Hz, (c) (6.25 to 12.5] Hz, (d) (12.5 to 25] Hz, (e) (25 to 50] Hz, (f) (50 to 100] Hz respectively. The (a) and (f) sub-band coefficients are forced to zero to get rid of baseline wandering and high frequency contents. The remaining bands then go through the reverse
route of Figure 9-2 to reconstruct the signal. This signal is the input to the feature extraction in the general detection block diagram Figure 9-3.

Figure 9-1. Samples of EEG and the corresponding frequency character.

Regarding the filter selection, Daubechie (DAUB) 4 has been suggested by other researchers [34] [50] for reason of matching the frequency character of in-seizure EEG. We have tested other filters in preliminary simulation, but found little difference, so DAUB4 filter is used.

Figure 9-4 is the detection performance with this wavelet preprocessor followed by the T1 index for both scalp EEG and ECoG data tested in the previous chapters. Panel (a) and (c) are the detection performance with wavelet preprocessing of ECoG and scalp EEG, and panel (b) and (d) are the T1 feature direct detection performance for easy comparison. On average, about 93% of the seizures can be detected with about 0.25 false alarms per hour. Without preprocessing in chapter 4 the average detection rate is 85% with 0.22 false alarms per hour for ECoG, and in chapter 8 the average detection rate is
85% with 1 false alarm per hour. It can be seen that both detection rate and false alarm rate has significant improvement

**Application of SOM Template Matching**

SOM template matching, although it is not applicable to artifacts rejection on scalp EEG for lack of template, can be applied to any signal where specific features are known to exist. The immediate interest and available data in epilepsy are the subject seizures. Hence, seizure segments can be used as templates. Figure 9-5 is the KL divergence between the test channels from data set P93 to a SOM template trained with an in-seizure ECoG section of the same patient. And Figure 9-6 is the correlation coefficient of this
same test data. There are 23 seizures in this data set, and 18 of them are similar to the template in the KL divergence sense, while the CC is also 80% similar to the template.

Figure 9-4. Performance of seizure detection with wavelet preprocessing. (a) and (c) are the wavelet preprocessed detection performance of ECoG and scalp EEG respectively. (b) and (d) are the T1 feature direct detection performance.

It can be deducted that to apply SOM template method for seizure detection, the detection is patient dependent, even channel dependent. To average out the seizure signature as proposed by reference [31], the detection is going to loose specificity (meaning the dip when a match is found will not be as obvious as Figure 9-5 shows), and additional methods would have to be incorporated in the SOM template matching method.
Figure 9-5. KL divergences between the in-seizure template and the ECoG signal of LTD1.

Figure 9-6. First order dynamic correlation coefficients between the in-seizure template and the ECoG signal of LTD1.
As a conclusion, for practical seizure detection, the wavelet filtering combined with recurrence time statistics feature extraction is the simplest combination, and its performance surpasses the SOM-BSS-T1 method at the present stage of development.

**Future Research**

The wavelet combined with T1 method, when confront with frequent spikes and spike and waves as in Figure 7-2 is useless because these kind of signal cause false alarms. Perhaps a more sophisticated statistical decision method at the decision stage may help. In case of frequent spikes, if the BSS algorithm could decompose the signal and a proper explanation can be associated with the decomposed components, then the BSS-SOM have an edge over wavelet preprocessing.

For BSS–SOM preprocessing, if perform decently, the component out of BSS algorithm is very intuitive. The component topography like Figure 9-7 could help find the location of the active sources. A challenging question is in scalp EEG, when multiple artifacts active at the same time (which is usually the case), it is not clear whether the BSS algorithm can do a decent decomposition, if it doesn’t what is the reason. Another very useful application for BSS is that when an ushered-in seizure concurrent with muscle artifacts, present simulation shows that BSS algorithm dose not perform what is the reason? Is the assumption that muscle artifacts source independent of EEG signal still legitimate in this situation? If so, what can be done on the algorithm side to decompose this kind of signal?

The T1 feature extraction has been demonstrated as an effective method to detect clinical seizures, but for sub-clinical seizures the method has no room to improve its performance. Sub-clinical seizure detection is as significant as clinical seizure detection in the sense of understanding the physiological reason of seizure, it is important to
seizure prediction and the ultimate goal of seizure control too. Sub-clinical seizure detection is another challenging direction to extend the present research forward.

Figure 9-7. Component topography of Figure 8-7.
APPENDIX A
SHORT TERM MAXIMUM LYAPUNOV EXPONENT (STL\(_{\text{MAX}}\))

Short Term Maximum Lyapunov Exponent Nutshell:

Given a continuous time series \(x(t)\) of duration \(T\), sampled by sampling period \(D_t\), generate \(N\) discrete samples, \(T=(N-1)*D_t\);

Embedding the discrete time series to reconstruct the phase space with dimension \(\rho\) and time leg \(\tau\), the phase space vectors are: \(X(t_i)=(x(t_i), x(t_i+\tau), \ldots x(t_i+(\rho-1)*\tau)^T\)

where \(t_i \in [1, N-(\rho-1)*\tau]\)

Then the STL\(_{\text{MAX}}\) of the phase space attractor is calculated by:

\[
L = \frac{1}{N_a*\Delta t} \sum_{i=1}^{N_a} \log_2 \left( \frac{\delta X_{i,j}(\Delta t)}{\delta X_{i,j}(0)} \right) 
\]  
\(a-1\)

\[
\delta X_{i,j}(0) = X(t_i) - X(t_j) 
\]  
\(a-2\)

\[
\delta X_{i,j}(\Delta t) = X(t_i + \Delta t) - X(t_j + \Delta t) 
\]  
\(a-3\)

Where \(X(t_i)\) is a vector on the trajectory of the phase space;

\(X(t_j)\) is another properly chosen vector which is adjacent to \(X(t_i)\),

\(\delta X_{i,j}(0)\) is the displacement vector at \(t_i\), regard as the perturbation of trajectory at \(t_i\),

\(\delta X_{i,j}(\Delta t)\) is the evolution of the perturbation at time \(t_i+\Delta t\),

\(N_a\) is the number of local Lmax that will be estimated within duration of \(T\)

\(\Delta t\) is the evolution time for \(\delta X_{i,j}\), the time one allow \(\delta X_{i,j}\) to evolve in phase space.

The careful choice of \(\rho\), \(\tau\), \(\Delta t\), \(\delta X_{i,j}\),and \(T\) are discussed in the reference.
APPENDIX B
SELF ORGANIZING MAPPING (SOM) ALGORITHM

Self organizing map algorithm:

Random initialize $N$ ($m$ dimensional) weight vectors $W_j$

for $n=1$:(number of training epoch)

(a). Competitive process: select representative PEs

$$i(X_k) = \arg \min_j \|X_k - W_j\| \quad (b-1)$$

$j=1,2,…,N$. where $i$ is the best matching winner PE index

(b). Cooperative process

Winner takes care of its neighborhood by setting itself at the center max point, and define its neighborhood symmetrically around itself, decay neighborhood relation monotonically from center max corresponding to lateral Euclidian distance

$$h_{j,i(X)} = \exp \left( -\frac{d_{j,i(X)}^2}{2\sigma(n)^2} \right) \quad (b-2)$$

$$\sigma(n) = \sigma_0 \exp \left( -\frac{n}{\tau_1} \right) \quad (b-3)$$

where $h_{j,i(X)}$ is the topological neighborhood function, $d_{j,i(X)}$ is lateral distance. $\tau_1$ is the neighborhood decay time constant, $\sigma(n)$ is the neighborhood size, it decreases with the increase of training epoch $n$.

(c). Weight adaptation
\[ W_j(n + 1) = W_j(n) + \eta(n) h_{j,i}(X_j(n))(X - W_j(n)) \]  \hspace{1cm} (b - 4)

\[ \eta(n) = \eta_0 \exp \left( -\frac{n}{\tau_2} \right) \]  \hspace{1cm} (b - 5)

where \( \eta(n) \) is the learning rate, it decays with the increase of training epoch,

and \( \tau_2 \) is the learning rate decay constant

**end**
LIST OF REFERENCES


BIOGRAPHICAL SKETCH

Hui Liu received her B.S degree in industry and automation from China, and her M.S degree from the University of Florida in electrical and computer engineering. During her Ph.D. program in the Computer NeuroEngineering Laboratory, she developed a system for detection of epileptic seizures from electroencephalogram.