

Neonatal Seizure Detection Using Blind Adaptive  
Fusion

NEONATAL SEIZURE DETECTION USING BLIND ADAPTIVE  
FUSION

BY  
HUAYING LI  
B.Eng.

A THESIS  
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING  
AND THE SCHOOL OF GRADUATE STUDIES  
OF MCMASTER UNIVERSITY  
IN PARTIAL FULFILMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF  
MASTER OF APPLIED SCIENCE

© Copyright by Huaying Li, September 2010

All Rights Reserved

Master of Applied Science (2010)  
(Electrical & Computer Engineering)

McMaster University  
Hamilton, Ontario, Canada

TITLE: Neonatal Seizure Detection Using Blind Adaptive Fusion

AUTHOR: Huaying Li  
B.Eng., (Electrical & Biomedical Engineering)  
McMaster University, Hamilton, Ontario, Canada

SUPERVISOR: Dr. Aleksandar Jeremic

NUMBER OF PAGES: x, 56

# Abstract

Seizure is the result of excessive electrical discharges of neurons, which usually develops synchronously and happens suddenly in the central nervous system. Clinically, it is difficult for physician to identify neonatal seizures visually, while EEG seizures can be recognized by the trained experts. Usually, in NICUs, EEG monitoring systems are used instead of the expensive on-site supervision. However, it is time-consuming to review an overnight recording, which motivates the researchers to develop automated seizure detection algorithms.

Although, there are few detection algorithms existed in the literature, it is difficult to evaluate these mathematical model based algorithms since their performances vary significantly on different data sets. By extending our previous results on multichannel information fusion, we propose a distributed detection system consisting of the existing detectors and a fusion center to detect the seizure activities in the newborn EEG. The advantage of our technique is that it does not require any prior knowledge of the hypotheses or the detector performances, which are often unknown in real applications. Therefore, this proposed technique has the potential to improve the performances of the existing neonatal seizure detectors.

In this thesis, we first review two newborn EEG models, one of which is used to generate neonatal EEG signals. The synthetic data is used later for testing purpose.

We also review three existing algorithms and implement them to work as the local detectors. Then, we introduce the fusion algorithms applied in the fusion center for two different scenarios: large sample size and small sample size. We finally provide some numerical results to show the applicability, effectiveness, and the adaptability of the blind algorithms in the seizure detection problem. We also provide the testing results obtained using the synthetic to show the improvement of the detection system.

# Acknowledgements

I wish to express my deep and sincere gratitude to my supervisor, Dr. Aleksandar Jeremic, for his encouragement and guidance from the initial to the final stage of my graduated study and for his support and detailed comments on the completion of this thesis.

I wish to express my thanks to my parents, for their continuous support on my study from the day I went aboard from home. Without their financial support, it is impossible to complete my undergraduate study at McMaster University. Without their support on taking care of my daughter, it is impossible to complete the research and write this thesis.

I also would like to thank my husband who gives me lots of understanding and my daughter who brings me lots of pleasure in my life.

Finally, I appreciate all of those who gave me supports during the completion of my graduate study.

Huaying Li

# Contents

<b>Abstract</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Outline . . . . .	4
<b>2 Neonatal EEG Synthesis</b>	<b>6</b>
2.1 A Nonstationary Model . . . . .	7
2.1.1 EEG Background Model . . . . .	7
2.1.2 EEG Seizure Model . . . . .	10
2.2 A Nonlinear Nonstationary Wiener Model . . . . .	14
2.2.1 EEG Background Model . . . . .	16
2.2.2 EEG seizure model . . . . .	18
<b>3 Neonatal EEG Seizure Detection Techniques</b>	<b>21</b>
3.1 Liu's Algorithm . . . . .	21
3.2 Golman's Algorithm . . . . .	25

3.2.1	Dominant Frequency $f_M$ . . . . .	26
3.2.2	Bandwidth $B$ . . . . .	27
3.2.3	Power Ratio $R_P$ . . . . .	27
3.2.4	Detection . . . . .	27
3.3	Celka's Algorithm . . . . .	28
3.3.1	Singular Spectrum Analysis (SSA) . . . . .	29
3.3.2	Minimum Description Length (MDL) . . . . .	30
3.3.3	Pre-processing . . . . .	30
<b>4</b>	<b>Distributed Detection System</b>	<b>32</b>
4.1	Overview of the System . . . . .	33
4.2	Local Detectors . . . . .	34
4.3	Fusion Center . . . . .	35
<b>5</b>	<b>Results and Discussions</b>	<b>42</b>
5.1	Simulations Results . . . . .	42
5.1.1	Large Data Set . . . . .	42
5.1.2	Adaptability to the Changes of Phenomenon . . . . .	45
5.1.3	Small Data Set . . . . .	46
5.2	System Performance with Surrogate Data . . . . .	47
5.3	Summary . . . . .	48
<b>6</b>	<b>Conclusions</b>	<b>51</b>

# List of Tables

2.1	Beta Distributions of $R_k$ and $V_n$ . . . . .	12
2.2	Parameters of $\hat{\beta}_{p_i}$ . . . . .	18
2.3	Parameters of PWLFM Signal $s(k)$ . . . . .	18
2.4	Parameters of $\hat{\beta}_{s_i}$ . . . . .	20
3.1	Scale for Scoring Autocorrelation Function . . . . .	24
3.2	Boundaries of Decision Regions . . . . .	28

# List of Figures

2.1	Newborn EEG Background and Power Spectrum . . . . .	8
2.2	Synthetic and Real Newborn EEG Background . . . . .	10
2.3	Synthetic and Real Newborn EEG Seizures . . . . .	14
2.4	Nonlinear Nonstationary Model of EEG Seizure . . . . .	15
2.5	Background Model and Inverse Background Model . . . . .	16
2.6	Seizure Model . . . . .	19
3.1	An EEG Epoch (30 seconds) and its Autocorrelation Function . . . . .	22
3.2	Moment Center of Autocorrelation Function . . . . .	23
3.3	Full Width at Half Maximum . . . . .	26
4.1	Parallel Distributed Detection System . . . . .	33
5.1	Averaged False Alarm Rate over 5000 Realizations . . . . .	43
5.2	Averaged Missed Detection Rate over 5000 Realizations . . . . .	44
5.3	Averaged Error Probabilities over 5000 Realizations . . . . .	44
5.4	Estimated $\hat{P}_1$ when the Statistical Property of the Signal Change . . . . .	45
5.5	Estimated $\hat{P}_1$ Using a Sliding Window . . . . .	46
5.6	Comparison . . . . .	47
5.7	False Alarm Rate of the Local Detectors . . . . .	48
5.8	Missed Detection Rate of the Local Detectors . . . . .	49

5.9	The Overall Error Probability of the Detection System . . . . .	49
-----	---	----

# Chapter 1

## Introduction

### 1.1 Motivation

A seizure is defined clinically as a paroxysmal alteration in neurologic function, i.e., behavioural, motor, or autonomic function. It is a result of excessive electrical discharges of neurons, which usually develop synchronously and happen suddenly in the central nervous system (CNS). It is critical to recognize seizures in newborns, since they are usually related to other significant illnesses. Seizures are also an initial sign of neurological disease and a potential cause of brain injury [18].

Clinical signs of seizures are different for adults and newborns. In adults, seizures are often accompanied by uncontrollable jerky movements of the body or the tonic flexion of the muscles. In newborns, many EEG seizures are not accompanied by any clinical signs. They may be accompanied with some nonobvious symptoms, such as sustained eye opening with ocular fixation, repetitive blinking of eyelids, or other slight facial expressions or body movements [14]. Since the clinical signs of seizures in newborns are subtle, it is difficult for physicians to identify the seizure occurrences

visually. Beyond clinical observations, seizures could be recognized in EEG recordings by some particular pattern, such as rhythmicity and repetitiveness [14]. Electroencephalographers are trained to identify EEG seizures from the recordings, but the number of these expertise usually is limited in the hospitals. Therefore, instead of the expensive on-site supervision, EEG recording systems are usually used in neonatal intensive care units (NICUs) to monitor the high risk babies [4].

The recorded EEG signals contain all the brain activities of the neonatal patient, normal or abnormal. Although on-site supervision is not required, reviewing an overnight recording is time-consuming because long-term monitoring generates huge amount of data, but only a small portion of which represents seizures. Nowadays, under the help of the powerful computational facilities, we have the potential to automatically detect seizure activities in newborns. Thus, automated seizure detections becomes a popular research topic. Several neonatal seizure detection algorithms exist in the literature. They use different biomedical signal processing techniques. The following three methods are the most popular ones, which are cited frequently.

1. Based on the rhythmicity of the neonatal EEG seizures, Liu *et al.* perform analysis on the autocorrelation function of the EEG signals in [8].
2. In [5], Gotman *et al.* divided neonatal seizures into three main groups: rhythmic discharges, multiple spikes, and very slow rhythmic discharges. They used spectral analysis in the frequency domain to detect the rhythmic discharges. They also provides two additional methods to detect the other two types of seizures.
3. The algorithm proposed by Celka and Colditz in [1] checked the complexity of neonatal EEG in order to identify newborn seizure activities.

Any of these well-known algorithms can be viewed as a single local detector. Usually, the performance of a single detector can be checked by its probability of false alarm and missed detection. In detection theory, the false alarm occurs when a non-target is detected as a target, while the missed detection occurs when a target is identified as a non-target. It is difficult to evaluate the performances of the existing seizure detectors since all of them are based on mathematical models whose performances vary significantly on different data sets and the statistical properties of the underlying EEG signals are time-dependent and patient-dependent.

In hospitals, a physician usually orders more laboratory tests when it is difficult to use the current test results to judge if a surgical operation is necessary or not. Similarly, in the seizure detection problem, multiple detectors can be used in order to accurately determine if there are seizure activities in the EEG or not. These multiple detectors observe the common phenomenon, the neonatal EEG, and make decisions on their own observations. The decisions are sent to a central processor, named as the fusion center. In the fusion center, the final decision is made by combining the received decisions in some way. The phenomenon, multiple local detectors, and the fusion center are the basic components of a distributed detection system. Usually, when the local decision rules are fixed, the fusion center requires the perfect knowledge on the prior information of the phenomenon and the performances of the detectors to optimally fuse the local decisions. However, such knowledge is not always available in real applications.

In our previous work, we proposed a blind algorithm for the distributed detection problem with  $M$  hypotheses. The advantage of this proposed fusion rule is that it does not require the prior knowledge of the hypotheses or the performances of the

local detectors.

In this work, we propose to combine the existing single seizure detectors to form a distributed detection system and apply our previously proposed blind algorithm on multichannel information fusion [9, 10]. First, we formulate the set of nonlinear equations consisting of the unknown a priori probabilities of the binary hypotheses and the unknown probabilities of false alarm and missed detection. Then, we estimate these unknowns using the corresponding multinomial distribution, maximum likelihood estimation and actual count of decisions made by different detectors. Finally, we present the analytical expression of overall error probability when the true values of the parameters are given and explore the effect of our blind algorithm to the overall seizure detection. To the evaluation purposes, we use a proposed neonatal EEG model [13] to generate neonatal EEG signals.

## 1.2 Outline

This thesis is organized as follow:

In Chapter 2, we introduce two neonatal EEG models proposed in [13] and [2]. We use the first model to generate neonatal EEG signals in order to evaluate the performance of our proposed detection system.

In Chapter 3, we review the three well-known algorithms on neonatal seizure detection.

In Chapter 4, we propose the distributed detection system by giving the structure of the system, describing the local decision rules, and introducing the fusion rule. When we introduce the blind algorithm used in the fusion center, we consider two scenarios: the number of the local decisions is large and the number of the local

decision is small.

In Chapter 5, a series of testings are conducted to evaluate the performance of the distributed detection system. The results provide the evidences that the proposed technique is applicable on the neonatal seizure detection problem.

In the last chapter, we give the summary of the completed work. We also list some the future research direction on this topic.

## Chapter 2

# Neonatal EEG Synthesis

To evaluate the performance of our proposed seizure detection system, it requires sufficient amount of EEG data. Due to the limited amount of real data, we decide to use simulated EEG. Simulation is a useful technique for engineers and researchers to examine the performance of a proposed algorithm [12]. If the proposed algorithm does not work well with the simulated data, it has little chance to perform well with real data. Another reason of using simulated data is that the artifacts in real neonatal EEG may interfere the evaluation of the performance since they sometimes mask the useful information in the signal [1, 4].

To the best of our knowledge, there are two models for simulating newborn EEG. In this chapter, we first review the newborn EEG model proposed by Rankine *et al.* in [13]. Then, we introduce another model proposed by Celka and Colditz in [2]. In this thesis, the synthetic neonatal EEG signals used for plotting and testing are referred to the ones generated from the former model unless otherwise stated. Using the synthetic signals, we can have the control of the testing data, such as the location, duration, and occurrence number of the seizure activities.

## 2.1 A Nonstationary Model

The newborn EEG model proposed in [13] consists of two sub-models: the newborn EEG background model and the newborn EEG seizure model.

### 2.1.1 EEG Background Model

The upper plot in Figure 2.1 is a typical neonatal EEG background epoch ( $T = 4s$ ) and its power spectrum is shown as the solid line in the lower plot. Base on the fact that the power spectrum can be approximated by a power law, shown by the dashed line in the same plot, Rankine *et al.* proposed a EEG background model in [13], given by

$$S_n(f) = \frac{c}{|f|^{\gamma_n}} \quad (2.1)$$

where  $S_n(f)$  is the power spectrum of the  $n$ th epoch,  $c$  is a constant, and  $\gamma_n$  is the power law exponent of the  $n$ th epoch. Since the EEG signal is nonstationary [1, 13], but it is assumed to be quasi-stationary for a certain epoch, the exponent  $\gamma_n$  varies for different epoches, but stays constant within an epoch.

The model in (2.1) is the frequency representation of the EEG background model. It can be rewritten as the product of two conjugate terms:  $X_n(f)$  and  $X_n^*(f)$ , given by

$$S_n(f) = \frac{c}{|f|^{\gamma_n}} = X_n(f)X_n^*(f) \quad (2.2)$$

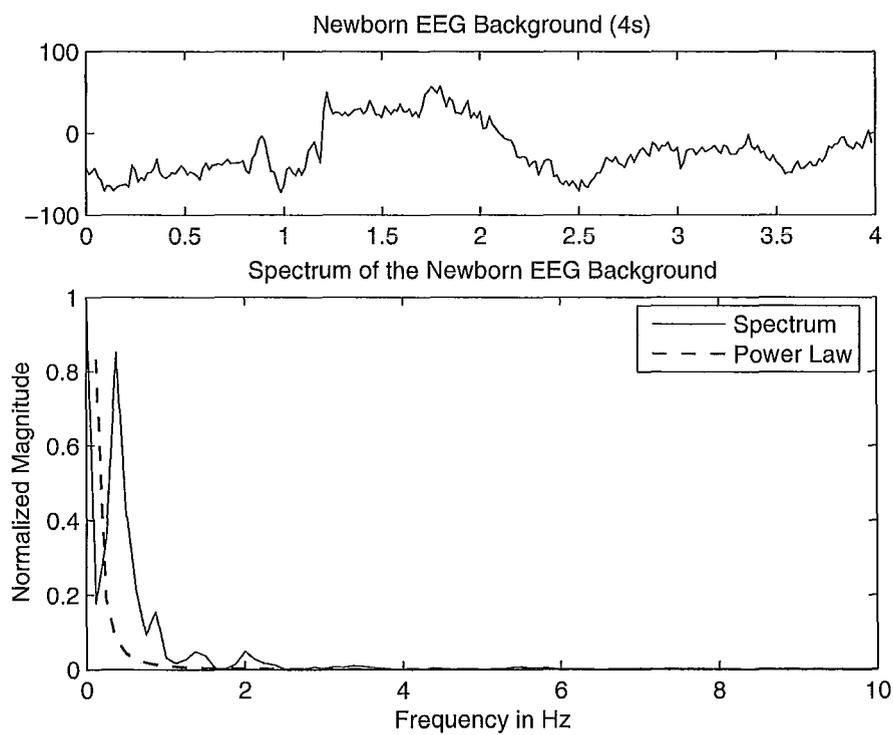


Figure 2.1: Newborn EEG Background and Power Spectrum

where

$$X_n(f) = \frac{\sqrt{c}}{|f|^{\gamma_n/2}} e^{j\theta_n(f)} \quad (2.3)$$

The term  $X_n(f)$  is considered to be the Fourier Transform of the  $n$ th epoch  $x_n(t)$ , with the phase spectrum  $\theta_n(f)$ . The synthetic  $n$ th epoch  $x_n(t)$  could be obtained from the inverse Fourier Transform of  $X_n(f)$  as

$$x_n(t) = \int_{-\infty}^{\infty} X_n(f) e^{j2\pi ft} df. \quad (2.4)$$

However it has a smooth power spectrum. To simulate the phenomenon that the spectrum of real EEG background fluctuates, several sub-epoches are generated by setting the power law exponents same and the phase spectra different. We denote the different phase spectra as  $\theta_{n,l}(f)$  for  $l = 1, \dots, L$ , where  $L$  is the number of the sub-epoches used. The sum of  $L$  sub-epoches

$$y_n(t) = \sum_{l=1}^L x_{n,l}(t) \quad (2.5)$$

represents the  $n$ th epoch of synthetic EEG background in the time domain.

Rankine *et al.* provided the estimation results for the background model in [13]. The estimated distribution of the power law exponents  $\gamma_n$  is obtained using a maximum likelihood estimator. It is a Beta distribution with parameters  $\alpha = 7.82$  and  $\beta = 7.44$ . The phase spectrum  $\theta_n(f)$  is assumed to be a realization of a random process, which is uniformly distributed in  $[0, 2\pi)$ . Some EEG signal generated using this model is shown in Figure 2.2, compared with some real EEG background signal.

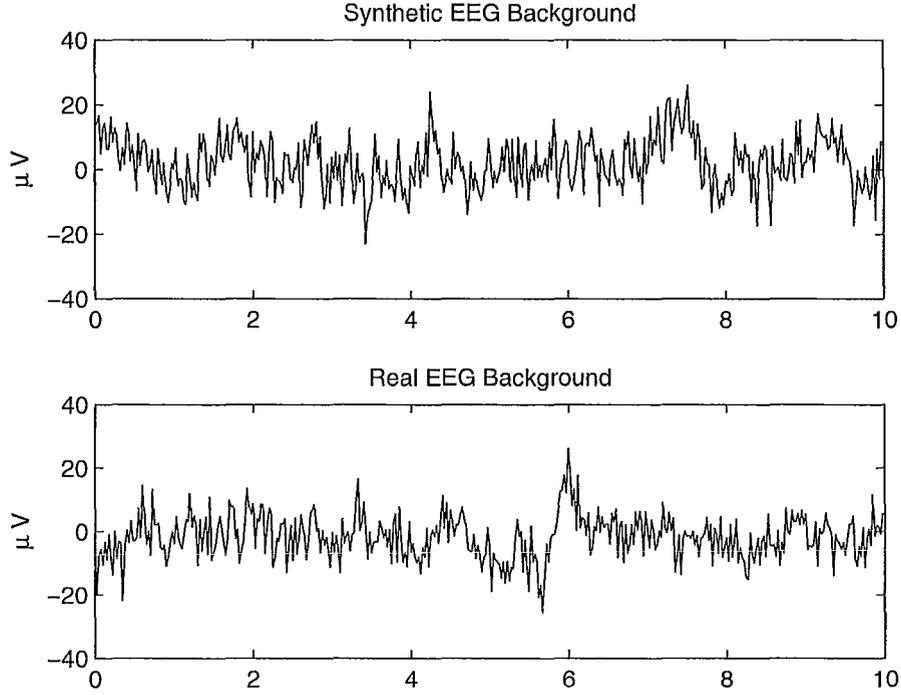


Figure 2.2: Synthetic and Real Newborn EEG Background

### 2.1.2 EEG Seizure Model

Based on the amplitude modulated, piecewise LFM, and multiple harmonic characteristic of EEG seizure, Rankine *et al.* proposed a neonatal EEG seizure model in [13]

$$s(t) = \sum_{k=1}^K a_k(t) \cos\left(2\pi \int_0^t f_k(\tau) d\tau + \theta_k\right) \quad (2.6)$$

which involves functions of the amplitude modulation  $a_k(t)$ , instantaneous frequency  $f_k(\tau)$ , and initial phase  $\theta_k$ .

### 2.1.2.1 Amplitude Modulation $a_k(t)$

The amplitude modulation of each harmonic  $a_k(t)$  is a function of the gain factor  $R_k$ , normalized variation  $V_n$ , and the number of turning points  $P$ ,

$$a_k(t) = \Phi(R_k, V_n, P; t) \quad (2.7)$$

The gain factor  $R_k$  is the ratio between the average amplitude of the  $k$ th harmonic and the fundamental. It implies  $R_1 = 1$ . The normalized variation  $V_n$  has the mean of 0.33. Suppose the  $P$  turning points occur randomly at

$$q = \frac{N(p + X)}{P} \quad (2.8)$$

where  $p = [0, \dots, P - 1]$  and  $X$  is a stationary random process, which is uniformly distributed in  $[0, 1]$ . The amplitudes associated with these  $P$  turning points are obtained from the gain factor  $R_k$  and the normalized variation  $V_n$  by

$$a_k(q) = R_k(0.67 + V_n) \quad (2.9)$$

The amplitude modulated function  $a_k(t)$  is then obtained from Equation (2.9) using a cubic spline interpolation.

The estimation results for the function  $a_k(t)$  are provided in [13]. The distribution of  $P$  is a discrete Beta distribution with parameters  $\alpha = 1.8$  and  $\beta = 3.0$ . The

Table 2.1: Beta Distributions of  $R_k$  and  $V_n$ 

	$R_2$	$R_3$	$R_4$	$R_5$	$V_n$
$\alpha$	1.7	1.5	1.9	1.4	3.9
$\beta$	3.2	4.1	3.6	1.2	8.0
min	0.2	0.2	0.2	0.2	0
max	1.2	1.0	0.6	0.4	1

minimum of  $P$  is 1 and the maximum is 8. The distributions of the gain factor  $R_k$  and the normalized variation  $V_n$  are all Beta distributions, whose parameters are shown in Table 2.1.

### 2.1.2.2 Time-varying Instantaneous Frequency $f_k(\tau)$

The fundamental instantaneous frequency  $f_1(t)$  is a piecewise linear frequency modulated (LFM) function with  $M$  pieces. Its mathematical representation is shown as follow:

$$f_1(t) = \sum_{m=1}^M F_m(\xi_m, C_m; t) \text{rect}\left(\frac{t - 0.5(B_{m+1} - B_m)}{B_{m+1} - B_m}\right) \quad (2.10)$$

where

$$F_m(\xi_m, C_m; t) = \xi_m t + C_m \quad (2.11)$$

and

$$C_m = \begin{cases} f_{st} & m = 1 \\ F_{m-1}(\xi_{m-1}, C_{m-1}; B_m) - \xi_m B_m & m \geq 2 \end{cases} \quad (2.12)$$

The starting frequency of LFM is given by  $f_{st}$  in Eq. (2.12). It is randomly

chosen from a Log-Normal distribution with parameters  $\mu = -0.17$  and  $\sigma^2 = 0.55$ . The minimum allowed value of  $f_{st}$  is 0.425. The gradients of LFM are given by  $\xi = [\xi_1, \xi_2, \dots, \xi_M]$ , which can be drawn from Beta distribution with parameters  $\alpha = 69.1, \beta = 69.8$ , falling in the range of  $[-0.06, 0.06]$ . The turning points  $B = [B_1 = 0, B_2, B_3, \dots, B_M, B_{M+1} = N]$ , where  $N$  is the discrete length of the seizure, is a stationary random process, which is uniformly distributed across the epoch. In order to guarantee the continuity of function  $f_1(t)$ , the alignment intercept  $C_m$  is computed using Eq. (2.12).

The instantaneous frequency of other harmonics  $f_k(t)$  can be derived from the fundamental by

$$f_k(t) = k f_1(t) \quad (2.13)$$

### 2.1.2.3 Simulation of Newborn EEG Seizure

The model (2.6) show us that the functions  $a_k(t)$ ,  $f_k(t)$ ,  $\theta_k$ , and  $K$  are all necessary to simulate the EEG seizure of newborns. We already describe the functions  $a_k(t)$  and  $f_k(t)$  in the previous sections. We now introduce the rest of the required functions and parameters.

The discrete epoch length is  $N = 256$ . The number of harmonic  $K$  is suggested to be 5. M-piece LFM is chosen to have 3 pieces. Initial phase  $\theta_k$  is randomly chosen from a stationary random process, uniformly distributed in  $[-\pi, \pi)$ . Let us define a vector

$$\mathbf{c} = [R_k, V_n, P, f_{st}, \xi, B, K, M, \theta_k] \quad (2.14)$$

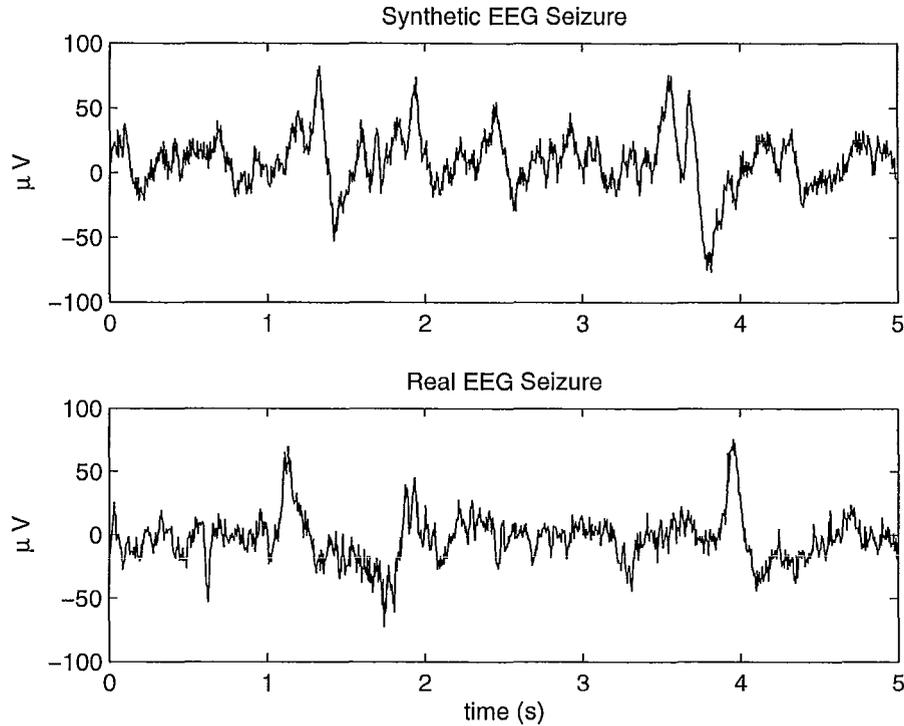


Figure 2.3: Synthetic and Real Newborn EEG Seizures

which contains all the parameters that are required to simulate the EEG seizure. We name it as the control vector of the EEG seizure model. To generate different neonatal EEG seizure signals, we can randomly choose the parameters in the control vector  $\mathbf{c}$ . A piece of synthetic EEG seizure signal is shown in Figure 2.3, compared with real EEG seizure.

## 2.2 A Nonlinear Nonstationary Wiener Model

Celka and Colditz proposed a nonlinear nonstationary model in [2] to simulate the neonatal EEG. The reason we introduce another neonatal EEG model is because

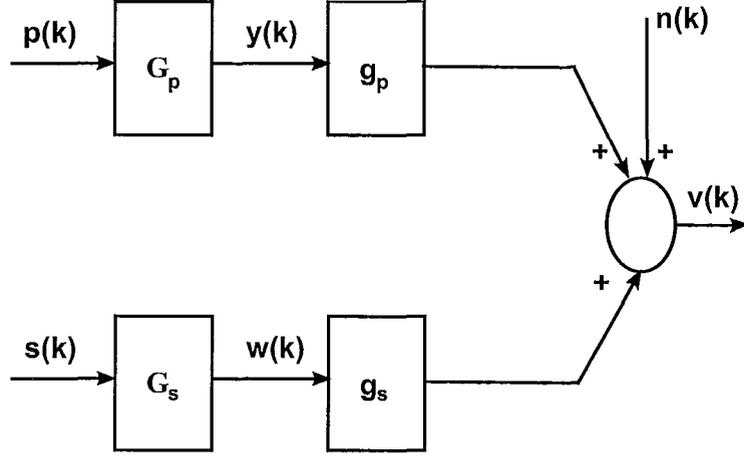


Figure 2.4: Nonlinear Nonstationary Model of EEG Seizure

this model is the basics of an existing detection algorithm which is reviewed later in this thesis. As shown in Figure 2.4, the upper branch is used to model the non-gaussianity of pure background brain activity and the lower branch is used to model the non-stationarity of pure seizure activity.

The mathematical representation of the model is given by

$$\begin{aligned}
 v(k) &= g_s[w(k)] + g_p[y(k)] + n(k) \\
 &= g_s\left[\sum_{i=0}^{\infty} q_s(i)s(k-i)\right] + g_p\left[\sum_{i=0}^{\infty} q_p(i)p(k-i)\right] + n(k) \quad (2.15)
 \end{aligned}$$

where  $q_p(i)$  has  $z$  transform  $G_p(z) = H_1(z)/H_2(z)$ ,  $q_s(i)$  has  $z$  transform  $G_s(z) = 1/H_2(z)$ , and  $n(k)$  is the measurement Gaussian white noise (GWN) with mean zero and variance  $\sigma_n^2$ .

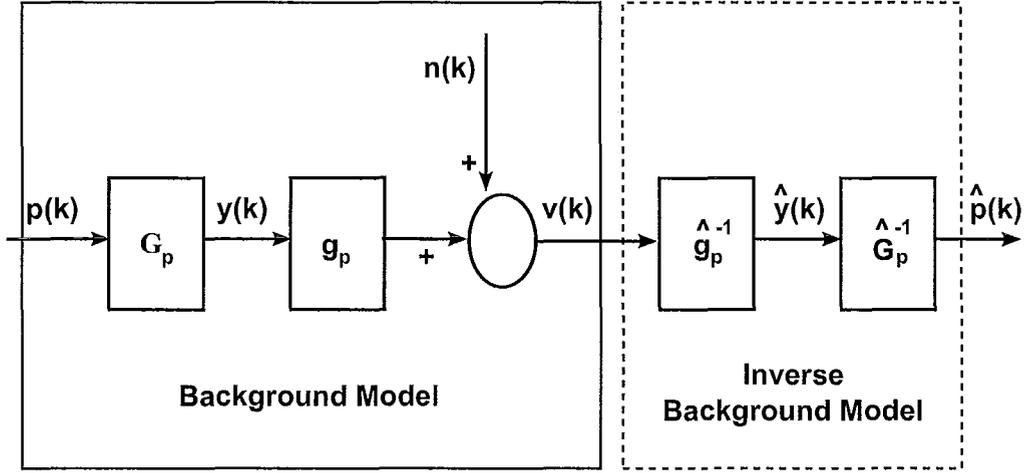


Figure 2.5: Background Model and Inverse Background Model

### 2.2.1 EEG Background Model

When there is no input signal in the lower branch of the model (Figure 2.4), i.e.,  $s(k) = 0$ , it turns to be the EEG background model, shown in the bold rectangle in Figure 2.5.

The input signal  $p(k)$  is supposed to be a Gaussian white noise. The first block of the background model  $G_p$  is an auto-regressive moving average (ARMA) filter. It is excited by the input  $p(k)$  and followed by a nonlinear shaping function  $g_p$ . The inverse background model, shown in the dashed rectangle in Figure 2.5, contains the inverse shaping function  $\hat{g}_p^{-1}$  and the inverse ARMA whitening filter  $\hat{G}_p^{-1}$ .

To estimate the background model, we start from the estimation of the inverse shaping function  $g_p^{-1}$ . The algorithm used to achieve this goal is called shaping algorithm. Suppose  $gauss(k)$  is a Gaussian variate time series. It has  $L$  samples, zero mean, and unit variance. Ranking  $gauss(k)$  in ascending order produces another series, named  $gauss_r(k)$ . In a similar way,  $v_r(k)$  is obtained from the measured EEG

signal  $v(k)$ .

Suppose  $\text{rank}(v(k)) = r$ , i.e.,  $v(k) = v_r(k)$ . The intermediate signal  $\hat{y}(k)$  can be obtained by

$$\hat{y}(k) = \text{gauss}(r) \quad \text{for } k = 0, \dots, L-1 \quad (2.16)$$

which has a Gaussian pdf. It is also the output of the inverse shaping function  $\hat{g}_p^{-1}$  when  $v(k)$  is treated as the input, i.e.,  $\hat{y}(k) = g_p^{-1}[v(k)]$ .

Let us define two monotonic increasing graphs:  $\Gamma_p^{-1} = \{v_r(k), \text{gauss}_r(k)\}$  and  $\Gamma_p = \{\text{gauss}_r(k), v_r(k)\}$ . The inverse shaping function  $g_p^{-1}$  and shaping function  $g_p$  can be estimated from graph  $\Gamma_p^{-1}$  and  $\Gamma_p$  respectively. However, the parametric model of  $g_p^{-1}$  is not useful because we can obtain the intermediate signal  $\hat{y}(k)$  by applying the shaping algorithm. Thus, only the estimation of the parametric model of  $g_p$  is necessary.

To estimate the shaping function  $g_p$ , the parametric model is defined as

$$g_p[x] = \beta_{p_0} + \beta_{p_1}x + \beta_{p_2}x^2 + \beta_{p_3}x^3 \quad (2.17)$$

where the parameters  $\beta_{p_i}$  (for  $i = 0, \dots, 3$ ) can be estimated using a least-square method on graph  $\Gamma$  and a Levenberg-Marquardt search algorithm. The estimates of  $\beta_{p_i}$  are given in Table 2.2.

We use the knowledge of  $\hat{y}(k)$  to estimate the inverse ARMA filter  $\hat{G}_p^{-1}$ . It can be done using MatLab routine *armax* with the model order  $D = 10$  and  $N = 10$ , which are the orders of the polynomials of  $H_1(z)$  and  $H_2(z)$  respectively. Once the inverse ARMA filter  $\hat{G}_p^{-1}$  is estimated, we can have the access to the estimated ARMA filter

Table 2.2: Parameters of  $\hat{\beta}_{p_i}$ 

Parameters	$\hat{\beta}_{p_0}$	$\hat{\beta}_{p_1}$	$\hat{\beta}_{p_2}$	$\hat{\beta}_{p_3}$
Averaged	0.014	0.76	-0.02	0.071
Maximum	0.016	0.87	-0.002	0.12
Minimum	0.013	0.60	-0.0039	0.039

Table 2.3: Parameters of PWLFM Signal  $s(k)$ 

Parameters	$\hat{\alpha}[s^{-2}]$	$\hat{\alpha}_1[s^{-2}]$	$\hat{\alpha}_2[s^{-2}]$	$\hat{f}_m[Hz]$
Maximum	0.092	0.039	0.087	1.55
Minimum	-0.083	-0.078	-0.048	0.67

$\hat{G}_p$  easily.

## 2.2.2 EEG seizure model

To simulate the seizure activities in the EEG signal, the input signal  $s(k)$  of the lower branch in Figure 2.4 is chosen to be a piecewise linear frequency modulated (PWLFM) sawtooth signal [2]. For  $t = k/F_r$ , the input signal  $s(k)$  is given as follow,

$$s(t) = z(t)e^{j\phi(t, \tilde{t})} \quad (2.18)$$

$$\phi(t, \tilde{t}) = 2\pi t f_{PWL}(\tilde{t}) \quad (2.19)$$

$$f_{PWL}(\tilde{t}) = \frac{\alpha_1 - \alpha}{2} |\tilde{t} - Bp_1| + \frac{\alpha - \alpha_2}{2} |\tilde{t} - Bp_2| + \frac{\alpha_1 + \alpha_2}{2} \tilde{t} - \frac{\alpha_1}{2} Bp_1 - \frac{\alpha_2}{2} Bp_2 + f_m \quad (2.20)$$

where  $z(t)$  is a periodic sawtooth signal and the estimates of parameters  $\alpha$ ,  $\alpha_1$ ,  $\alpha_2$  and  $f_m$  are shown in Table 2.3. The sum of  $Bp_1$  and  $Bp_2$  is set to 20 seconds.

The estimation of the ARMA filter  $G_s$  can be obtained from the results in Section 2.2.1, since  $G_s(z) = 1/H_2(z)$  is only related to the denominator of  $G_p(z) =$

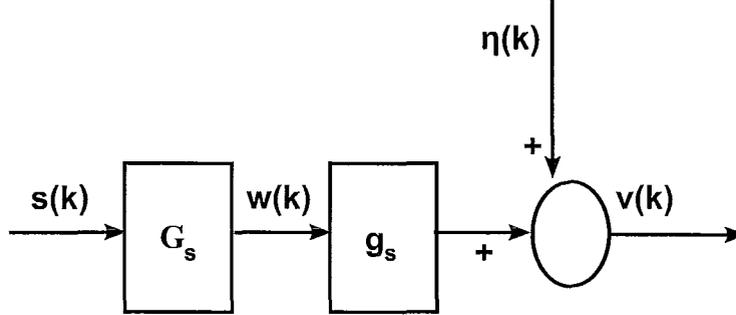


Figure 2.6: Seizure Model

$H_1(z)/H_2(z)$ . The mathematical model (2.15) shows that the EEG signal  $v(k)$  has three parts: the pure seizure activity, the pure background activity, and the noise. Consider the background activity and the noise together as the stochastic part of the model, named as  $\eta(k)$ . Therefore, the model (2.15) can be rewritten as

$$v(k) = g_s[w(k)] + \eta(k) \quad (2.21)$$

where  $\eta(k) = g_p[y(k)] + n(k)$ . For a typical EEG seizure signal, the contribution from  $\eta(k)$  is much less than that from the seizure activity ( $g_s[w(k)]$ ) [1, 2]. Therefore, we can reconstruct the model to be the one shown in Figure 2.6. The intermediate signal  $\hat{w}(k)$  is obtained from the estimated ARMA filter  $\hat{G}_s$  excited by the input signal  $s(k)$ , given by

$$\hat{w}(k) = \sum_{i=0}^{\infty} \hat{q}_s(i) s(k-i) \quad (2.22)$$

where  $\hat{q}_s(i)$  are obtained from  $\hat{G}_s = 1/\hat{H}_2(z)$ .

Using the shaping algorithm explained in the previous section, we create another graph  $\Gamma_s = \{\hat{w}_r(k), v_r(k)\}$  to estimate the shaping function  $g_s$ . Its parametric model

Table 2.4: Parameters of  $\hat{\beta}_{s_i}$ 

Parameters	$\hat{\beta}_{s_0}$	$\hat{\beta}_{s_1}$	$\hat{\beta}_{s_2}$	$\hat{\beta}_{s_3}$
Averaged	0.097	1.03	0.10	-0.02
Maximum	0.19	1.16	0.16	0.06
Minimum	-0.00056	0.83	0.08	-0.06

is defined as

$$g_s[x] = \beta_{s_0} + \beta_{s_1}x + \beta_{s_2}x^2 + \beta_{s_3}x^3 \quad (2.23)$$

where estimates of parameters  $\beta_{s_i}$  (for  $i = 0, \dots, 3$ ) are shown in Table 2.4.

# Chapter 3

## Neonatal EEG Seizure Detection Techniques

Several neonatal EEG seizure detection algorithms exist in the literature. To our best knowledge, the algorithms proposed by Liu *et al.* [8], Gotman *et al.* [5], and Celka and Colditz [1] are cited most frequently in this research field. In this chapter, we provide the reviews of the detection techniques used in these well known algorithms. They are also implemented to work as the local detectors in our distributed detection system.

### 3.1 Liu's Algorithm

Liu *et al.* focused on the rhythmic characteristic of neonatal EEG seizure and proposed a detection algorithm in [8] using autocorrelation analysis. Autocorrelation is the cross-correlation of a signal with the delayed version of itself. It is a useful mathematical tool in signal processing to detect repeated patterns in a signal. For

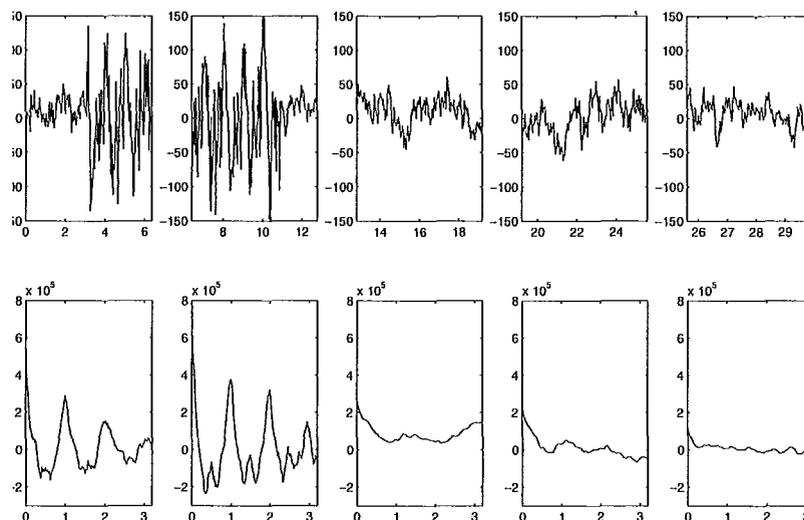


Figure 3.1: An EEG Epoch (30 seconds) and its Autocorrelation Function

example, it can be used in the detection of periodic signals buried under noises [7].

Neonatal EEG seizure is an ictal event that consists of repetitive waveforms lasting at least ten seconds [3]. Due to the periodicity of EEG seizure, its autocorrelation function has more peaks with similar periodicity of the original signal. In contrast, normal neonatal EEG does not have clear periodicity, so its autocorrelation usually has irregular peaks. This phenomenon is verified clearly by the plots in Figure 3.1. The upper plots are the segments of the EEG signal and the lower plots are their corresponding autocorrelation functions. The first two lower plots both have repeated peaks and the corresponding EEG segments contain seizure activities. The last three lower plots all have irregular peaks and no seizure activities appear in their corresponding EEG segments.

A scoring system is used by Liu *et al.* in [8] to determine the degree of periodicity of the EEG signal quantitatively in order to identify the existences of the seizure

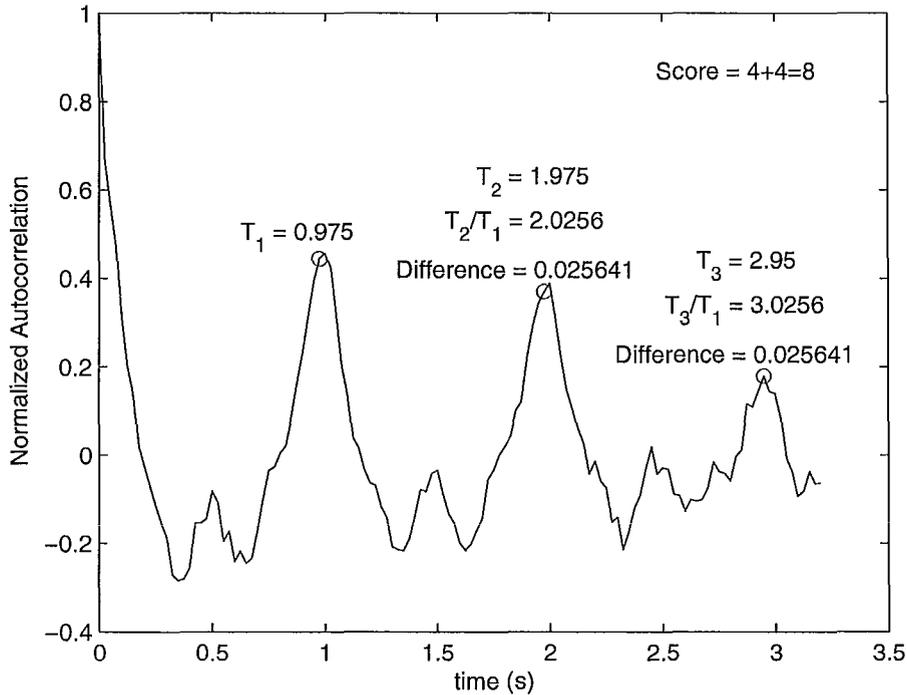


Figure 3.2: Moment Center of Autocorrelation Function

activities. For a single-channel EEG, 30-second data is considered as an epoch. As shown in Figure 3.1, this EEG epoch is divided into five consecutive windows. Each window is approximately 6-second long. The breaking points of the time interval  $[0, 30]$  are 6.4s, 12.8s, 19.2s, and 25.6s.

To check the periodicity, the moment centers are first calculated. The moment center of a certain peak is defined as the moment that halves the area between zero-crossings of that peak. For example, in Figure 3.2,  $T_1$  is the moment center of the first positive peak, so it is called the primary moment center.  $T_2$  and  $T_3$  are the second and third moment center respectively. (If more positive peaks exist, nomination is similar.) Then, we calculate the ratios of second and third moment center to the

Table 3.1: Scale for Scoring Autocorrelation Function

Difference	Score
0.000-0.025	+5
0.025-0.050	+4
0.050-0.075	+3
0.075-0.100	+2
0.100-0.125	+1
0.125-0.150	0
0.150-0.175	-1
0.175-0.200	-2
0.200-0.300	-3
0.300-0.400	-4
0.400-0.500	-5

primary moment center, i.e.,  $R_2 = T_2/T_1$  and  $R_3 = T_3/T_1$ . Finally, the difference between each ratio and its nearest integer is calculated to obtain scores using the scale shown in Table 3.1. In our example, according to Figure 3.2, ratio  $R_2$  is 2.0256. The difference between  $R_2$  and its nearest integer 2 is 0.0256. Ratio  $R_3$  is 3.0256, whose difference with its nearest integer is also 0.0256. According to Table 3.1, the second and third peak both obtain score of 4. Thus, the total score of this particular window is  $4+4 = 8$ .

Therefore, the EEG signal containing seizure activities should obtain high score because its autocorrelation has more repeated peaks. The normal EEG signal with irregular peaks in its autocorrelation function should obtain low score. The criteria proposed in [8] is used to determine the existences of seizure activities in a 12-channel EEG signal. Since we are only interested in single-channel EEG, we modify the criteria to be: if one of the following is satisfied, the epoch contains seizures.

1. 2 consecutive window scores within a channel were  $\geq 2$ , with a sum of the 2 window scores  $\geq 10$

2. 3 consecutive window scores within a channel were  $\geq 2$ , with a sum of the 2 window scores  $\geq 14$
3. a score from a single window was  $\geq 12$

## 3.2 Golman's Algorithm

Gotman *et al.* proposed three different seizure detection methods in [5] to detect three types of seizures: rhythmic discharges, multiple spikes, and very slow rhythmic discharges, respectively. In this thesis, we only focus on the rhythmic discharge detection since it could identify 90% of the seizures detected by all three detection algorithms [4].

Power spectrum analysis is an important tool in signal processing. It is frequently used in EEG classification and detection problem [6]. The algorithm reviewed in the previous section is based on the rhythmicity of the neonatal EEG seizure in the time domain. Gotman *et al.* proposed the detection method in [5] using the same characteristic of neonatal EEG seizure, but they performed the analysis in the frequency domain.

The rhythmicity of a signal can be represented in the frequency domain by a high and narrow peak at the frequency of that signal. Therefore, in the spectrum of the EEG segment containing seizure activities, a large distinct peak is expected to appear at the main frequency of EEG seizure. Smaller peaks may appear at one or two other frequencies. In the rest of the spectrum, only little power is expected.

Gotman *et al.* considered 10-second as an current epoch and 20-second as a background epoch. Between current and background epoch, there is a 60-second gap.

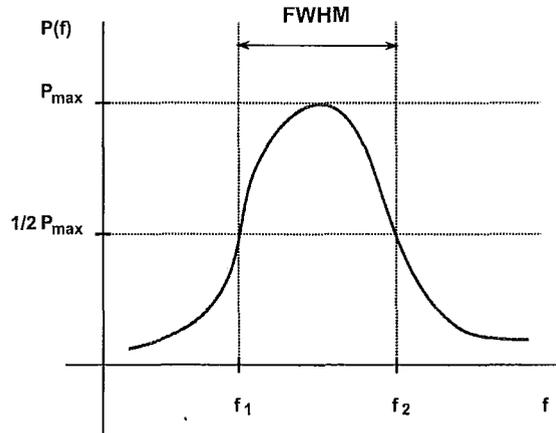


Figure 3.3: Full Width at Half Maximum

As the current epoch moving, the background epoch is updated constantly. Therefore, the detection results do not depend on the choice of background. Also, a 2.5-second sliding window is used in the detection algorithm. It provides an overlap of 75% of the current epoch, which guarantees that a 10-second seizure could be identified at least once.

Features of the current epoch, such as the frequency and the bandwidth of the dominant peak, are examined and compared with those of the background epoch. In the following sections, we introduce the features that used to detect seizure activities.

### 3.2.1 Dominant Frequency $f_M$

Let us first have a quick review on the term of full width at half maximum (FWHM). As shown in Figure 3.3, function  $P(f)$  has maximum value  $P_{max}$ . The values of function  $P(f)$  at  $f_1$  and  $f_2$  are both equal to the half of the maximum value  $P_{max}$ . The difference between  $f_1$  and  $f_2$  is defined as the full width at half maximum.

The spectrum of an EEG seizure epoch usually has several peaks. The dominant

peak  $M$  is the one with the largest average power in its full width half maximum band. The corresponding frequency of the dominant peak is called the dominant frequency, named as  $f_M$ . It represents the the most prominent rhythmic component of the epoch.

### 3.2.2 Bandwidth $B$

The bandwidth of the dominant spectral peak,  $B$ , is defined as the full width at half maximum of the dominant peak. For example, suppose the peak in Figure 3.3 is the dominant peak. The amplitude of  $f_1$  equals half of the maximum value in the rising slope of the peak. The amplitude of  $f_2$  has the same amplitude but appears in the falling slope. By definition, the difference between  $f_1$  and  $f_2$  is the width of the dominant spectral peak, which is the bandwidth  $B$ . The smaller this width is, the more rhythmic the signal is.

### 3.2.3 Power Ratio $R_P$

Suppose  $P_c$  is the power in the full width half maximum band of dominant peak in the current epoch and  $P_b$  is the power in the same frequency band but in the background epoch. The power ratio  $R_P$  is defined as the ratio of  $P_c$  and  $P_b$ . It is computed to determine if the rhythmic activity in the current epoch is new compared to the background.

### 3.2.4 Detection

Other features, such as stability  $I_s$ , 60-Hz indicator  $I_{60}$ , and disconnect indicator  $I_d$ , are also checked in this detection method to get rid of artifact or non-useful information.

Table 3.2: Boundaries of Decision Regions

	Dominant Frequency $f_M$	Bandwidth $B$	Power Ratio $R_P$
First Choice	0.5-1.5	$\leq 0.6$	3-4
Second Choice	1.5-10	$\leq 0.6$	2-4
third Choice	1.5-10	$\leq 1$	40-80

The current epoch is divided into four sub-epochs, each of which is 2.5 seconds. The average amplitude of each sub-epoch is computed. The stability of current epoch  $I_s$  is checked by the ratio of the highest value and the lowest. The 60-Hz indicator  $I_{60}$  is the ratio of the peak at 20 Hz to the total energy in the spectrum. Patient disconnection is noticed when the total energy in the spectrum is extremely low.

Define a 3-dimensional space  $\Phi$  using the three parameters described in Section 3.2.1 - 3.2.3,  $f_M$ ,  $B$ , and  $R_P$ . The boundaries of  $\Phi$  is used to determine where the detections belong, which is shown in Table 3.2. An epoch is consider to be seizure, if it is accepted according to Table 3.2 and  $I_s < 3$ ,  $I_{60} < 0.8\%$ , and  $I_d > 0$ .

### 3.3 Celka's Algorithm

The algorithm reviewed in this section was proposed by Celka and Colditz in [1]. The authors used a new detection technology compared with the ones used by Liu *et al.* and Gotman *et al.*. They performed the singular spectrum analysis and the information theoretic-based signal subspace selection to examine the complexity of the EEG signal. This detection algorithm has three main steps: Pre-processing, singular spectrum analysis, and minimum description length.

### 3.3.1 Singular Spectrum Analysis (SSA)

Singular Spectrum Analysis (SSA) is widely used in time-series analysis. It is a robust method of extracting information from a noisy environment. The measured EEG signal,  $\mathbf{x} = \{x(k)\}_{k=1}^N$ , is zero-meaned and normalized to have unit variance. Define  $\mathbf{x}_k = [x(k), x(k+1), \dots, x(k+(n_s-1))]^T$  as a state vector in  $\mathbb{R}^{n_s}$ .  $n_s$  is determined by  $n_s < \min\{F_r/\Delta B, (N/3+1)\}$ , where  $F_r$  is the sampling frequency and  $\Delta B$  is the bandwidth of the information bearing signal. In this thesis, the sampling frequency is  $F_r = 40$  Hz. For EEG seizure,  $\Delta B$  is in the interval of  $[1, 5]$  Hz, which leads  $8 < n_s < 40$ .

The trajectory matrix  $Z$  is defined as  $Z^T = [\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_{N_T}]$ , shown in (3.1), and its size is  $N_T \times n_s$ , where  $N_T = N - (n_s - 1)$ .

$$Z^T = [\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_{N_T}] = \begin{bmatrix} x(1) & x(2) & \dots & x(n_s) \\ x(2) & x(3) & \dots & x(n_s+1) \\ \dots & \dots & \dots & \dots \\ x(N_T) & x(N_T+1) & \dots & x(N) \end{bmatrix}^T \quad (3.1)$$

The covariance matrix  $A_{zz}$  has  $n_s$  eigenvalues,  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{n_0} \geq \dots \geq \lambda_{n_s}$ . The deterministic portion of the signal is related to the  $n_0$  largest eigenvalues, while the noise part is related to the rest  $(n_s - n_0)$  eigenvalues.

Compared with computing the covariance matrix  $A_{zz}$ , singular value decomposition (SVD) of  $Z$  is more robust against noise and more efficient to evaluate the eigenvalues of  $A_{zz}$  for a short time series. The relationship between the singular values  $\sigma_i$  and the eigenvalues  $\lambda_i$  is  $\sigma_i^2 = \lambda_i$  for  $i = 1, \dots, n_s$ .

### 3.3.2 Minimum Description Length (MDL)

The optimal dimension estimation is performed to determine the minimal dimension ( $n_0$ ) of the space embedded in the state space ( $n_s$ -dimensional). Rissanen's MDL criterion was used since it is suit for selecting subspace in a noisy environment.

$$\begin{aligned} \text{MDL}(l, \alpha_i) = & -l\ln \left[ \frac{\prod_{i=l+1}^{n_s} \alpha_i^{1/(n_s-1)}}{\frac{1}{n_s-l} \sum_{i=l+1}^{n_s} \alpha_i} \right]^{N_T(n_s-l)} \\ & + n_f(l) \left( \frac{1}{2} + l\ln[\gamma] \right) - \frac{n_f(l)}{l} \sum_{i=1}^l \ln \left[ \alpha_i \sqrt{\frac{2}{N_T}} \right] \end{aligned} \quad (3.2)$$

where  $\gamma = 32$  is for a floating point representation and  $n_f(l) = n_s l - (l^2/2) + l/2 + 1$  is the number of freely adjustable parameters. By using  $L_1$  norm, we have  $\alpha_i = \lambda_i$ . The minimum description length is determined by

$$n_0 = \arg \min_{l \in \{1, \dots, n_s\}} \text{MDL}(l, \alpha_i). \quad (3.3)$$

Usually, a signal consists of two parts: deterministic part containing information and stochastic part containing noise. The purpose of the SSA step is to separate the noise of EEG signal from the deterministic part of it. The role of  $n_0$  is to examine the complexity of the deterministic part of the signal. A pure white noise could be considered as a signal that only has stochastic part and  $n_0 = 1$ . For other signals containing both parts,  $n_0 > 1$ .

### 3.3.3 Pre-processing

The pre-processing is based on a model proposed by the same author in another paper [2]. We already introduced this model in Chapter 2, which was shown in Figure 2.4.

Also, we already have the access to the estimated inverse nonlinear function  $\hat{g}_p^{-1}$  and the inverse ARMA filter  $\hat{G}_p^{-1}$ . To Gaussianize the measured EEG, we apply  $\hat{g}_p^{-1}$  to  $x(k)$  and then apply  $\hat{G}_p^{-1}$  to  $x(k)$  in order to whitening the background. This procedure is described mathematically as follow,

$$x_n(k) = \hat{G}_p^{-1}(z)\hat{g}_p^{-1}[x(k)] = \underbrace{\hat{G}_p^{-1}(z)(\hat{g}_p^{-1} \circ g_s)[G_s(z)s(k)]}_{\text{deterministic part}} + \underbrace{\tilde{n}(k)}_{\text{stochastic part}} \quad (3.4)$$

which shows that  $x(k)$  consists of a deterministic part and a stochastic part. For EEG background, there is no seizure activity ( $s(k) = 0$ ). Therefore, the signal  $x(k)$  becomes whitened after pre-processing and does not contain any deterministic part. The MDL result should be  $n_0 = 1$ . Otherwise, for EEG with seizure activities, the result should be  $n_0 > 1$ . Based on the complexity of EEG seizure,  $n_0 \approx 3$  for EEG seizure.

# Chapter 4

## Distributed Detection System

Diagnosis of neonatal seizure is a typical detection problem. Each of the algorithms reviewed in the previous section can be considered as a single detector. Since the statistical properties of neonatal EEG are time-dependent and vary significantly from patient to patient, it is difficult to evaluate the performance of existing single detectors since they are all based on mathematical models whose performances change on different data sets. For example, a single detector may have different false alarm rates at different time of the day. The missed detection rate of a detector may change when examining patients with different ages. Thus, it motivates us to combine the existing single detectors and utilize their strengths by extending our previous results on blind multichannel information fusion [9].

In this chapter, we present a parallel distributed detection system, by first discussing the structure of the system, then briefly describing the local detectors and their decision rules, and finally introducing the optimal fusion rule applied in the fusion center. The advantage of this proposed technique is that it does not require any knowledge on a priori probability of seizure occurrence or the performances of

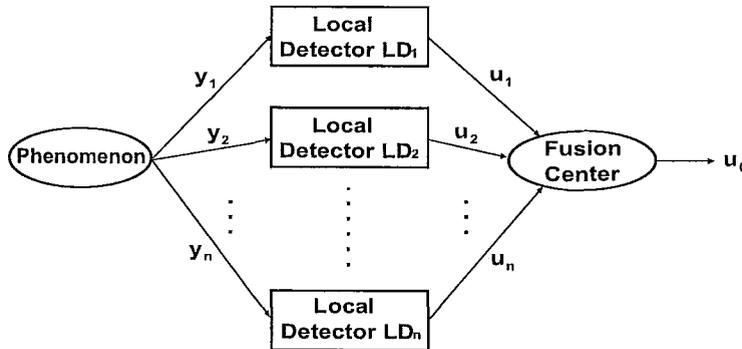


Figure 4.1: Parallel Distributed Detection System

the local detectors.

## 4.1 Overview of the System

Distributed detection is originally used in military surveillance, but now it has more applications in other fields, such as communication, biomedical engineering, and etc [9, 17]. There are four major topologies of distributed detection: parallel with or without fusion center, series, and tree [16]. We choose the topology of our detection system as parallel since the local detectors work on their own observations and do not communicate with others. Although, it is possible to enable the connections between local detectors, this approach increases the computational complexity of the system at the same time. Finally, a fusion center is required in this system in order to make a final decision based on the decision of the local detectors. The fusion center does not have access to the statistical parameters of the phenomenon.

Figure 4.1 shows the structure of a typical parallel distributed detection system with  $N$  detectors. The role of the local detectors  $LD_n$  is to make local decision  $u_n$  based on their own observations  $y_n$ . All the local decisions are then sent to the

fusion center, where the global decision  $u_0$  is made based on a fusion rule in order to minimize the overall probability of error. In this work, we only focus on the case of three local detectors, i.e,  $N = 3$ , unless otherwise stated. Additional detectors can be added into the system whenever more information is required to make final decision. Although increasing the number of detectors has the potential to reduce the detection error probability, it also increases the computational cost.

## 4.2 Local Detectors

The local detectors  $LD_n$  have their own decision rules. We use the three algorithms reviewed in Chapter 3 to formulate the local decision rules.

We perform hypothesis testings (local decisions) with two hypotheses:

$$\begin{aligned} H_0 : & \quad \text{The EEG signal does not contain seizure} \\ H_1 : & \quad \text{The EEG signal contains seizure} \end{aligned}$$

for the local detector  $LD_n$ . The local decisions  $u_n$ ,  $n = 1, 2, 3$ , are made by

$$u_n = \begin{cases} 0, & \text{the } n\text{th detector favors } H_0 \\ 1, & \text{the } n\text{th detector favors } H_1 \end{cases} \quad (4.1)$$

We use  $P(H_0)$  and  $P(H_1)$  to denote the a priori probability of the hypothesis  $H_0$  and  $H_1$ , respectively.

A common assumption used here is the local observations  $y_n$  are conditionally independent, given the unknown hypothesis  $H_i$ , i.e.,  $P(y_j, y_k | H_i) = P(y_j | H_i)P(y_k | H_i)$

for all  $j \neq k$  and all  $i$ .

In a more general problem, the binary hypothesis testings could be replaced by the hypothesis testings with more hypotheses, i.e.,  $M = 3$ .

### 4.3 Fusion Center

After receiving the local decisions, the fusion center makes the global decision by applying an optimal fusion rule in order to minimize the final error probability. For a binary hypothesis testing problem, the error probability  $P_e$  is given by

$$P_e = P(H_0)P(u_0 = 1|H_0) + P(H_1)P(u_0 = 0|H_1) \quad (4.2)$$

and the two hypotheses are still

$H_0$  : The EEG signal does not contain seizure

$H_1$  : The EEG signal contains seizure

Varshney provided the optimality criterion for  $N$  local detectors in the sense of minimum error probability in [15]. We recall it here for the case of  $N = 3$ .

$$u_0 = \begin{cases} 1, & \text{if } w_0 + \sum_{n=1}^3 w_n > 0 \\ 0, & \text{otherwise} \end{cases} \quad (4.3)$$

$$\text{where, } w_0 = \log\left(\frac{P_1}{P_0}\right) \quad (4.4)$$

$$\text{and } w_n = \begin{cases} \log((1 - P_n^m)/P_n^f), & \text{if } u_n = 1 \\ \log(P_n^m/(1 - P_n^f)), & \text{if } u_n = 0 \end{cases} \quad (4.5)$$

As the short form of  $P(H_1)$ ,  $P_1$  is a priori probability of hypotheses  $H_1$ . Similarly,  $P_0$  is the short form of  $P(H_0)$ . The sum of  $P_0$  and  $P_1$  is 1. The probabilities of false alarm and missed detection of the  $n$ th local detector are denoted as  $P_n^f$  and  $P_n^m$ , respectively. The optimal fusion rule tells us that the global decision  $u_0$  is determined by the a priori probability and the detector performances, i.e.,  $P_1$ ,  $P_n^f$  and  $P_n^m$ . However, they are all unknown in our seizure detection problem, which is usually the case in many other real applications [9, 11]. In order to make the final decision, we need to utilize the information available to us: the local binary decisions  $u_n$ .

Suppose the decision combination  $\{u_1 = i, u_2 = j \text{ and } u_3 = k\}$  is represented by  $\ell = (ijk)_2$ , where  $i, j, k = 0$  or  $1$  [11]. In our system, the number of all the possible local decision combinations is  $2^3$  and will be denoted as  $L$  in the remainder of this thesis. The joint probability of decision  $\{u_1 = i, u_2 = j \text{ and } u_3 = k\}$  is also the

occurrence probability of the  $\ell$ th decision combination, given by

$$\begin{aligned}
P_\ell &= \Pr(u_1 = i, u_2 = j, u_3 = k) \\
&= P(u_1 = i|H_1)P(u_2 = j|H_1)P(u_3 = k|H_1)P_1 \\
&\quad + P(u_1 = i|H_0)P(u_2 = j|H_0)P(u_3 = k|H_0)(1 - P_1)
\end{aligned} \tag{4.6}$$

$$P(u_n = i|H_1) = \begin{cases} 1 - P_n^m, & \text{if } i = 1 \\ P_n^m, & \text{if } i = 0 \end{cases} \tag{4.7}$$

$$P(u_n = i|H_0) = \begin{cases} P_n^f, & \text{if } i = 1 \\ 1 - P_n^f, & \text{if } i = 0 \end{cases} \tag{4.8}$$

By listing all the possible decision combinations and substituting them into Eq. (4.6)

- Eq. (4.8), it generates a nonlinear equation set:

$$\left\{ \begin{aligned}
P_{000} &= P_1^m P_2^m P_3^m P_1 + (1 - P_1^f)(1 - P_2^f)(1 - P_3^f)(1 - P_1) \\
P_{001} &= P_1^m P_2^m (1 - P_3^m) P_1 + (1 - P_1^f)(1 - P_2^f) P_3^f (1 - P_1) \\
P_{010} &= P_1^m (1 - P_2^m) P_3^m P_1 + (1 - P_1^f) P_2^f (1 - P_3^f)(1 - P_1) \\
P_{011} &= P_1^m (1 - P_2^m)(1 - P_3^m) P_1 + (1 - P_1^f) P_2^f P_3^f (1 - P_1) \\
P_{100} &= (1 - P_1^m) P_2^m P_3^m P_1 + P_1^f (1 - P_2^f)(1 - P_3^f)(1 - P_1) \\
P_{101} &= (1 - P_1^m) P_2^m (1 - P_3^m) P_1 + P_1^f (1 - P_2^f) P_3^f (1 - P_1) \\
P_{110} &= (1 - P_1^m)(1 - P_2^m) P_3^m P_1 + P_1^f P_2^f (1 - P_3^f)(1 - P_1) \\
P_{111} &= (1 - P_1^m)(1 - P_2^m)(1 - P_3^m) P_1 + P_1^f P_2^f P_3^f (1 - P_1)
\end{aligned} \right. \tag{4.9}$$

In this nonlinear system, only seven out of eight equations are independent since  $\sum P_\ell = 1$  and there are seven unknowns  $P_1$ ,  $P_n^f$  and  $P_n^m$ , for  $n = 1, 2, 3$ . Thus, it can be solved theoretically when  $P_\ell$  are known. Although  $P_\ell$  is usually unavailable in practice, it could be replaced by empirical probability defined as

$$\begin{aligned} P_\ell &= \Pr(u_1 = i, u_2 = j, u_3 = k) \\ &\simeq \frac{\text{number of } u_1 = i, u_2 = j, u_3 = k}{\text{number of local decisions } N_t} \end{aligned} \quad (4.10)$$

where  $N_t$  is the number of decisions made by one of the local detectors. Eq. (4.10) is true usually when the number of decisions is large [9].

Mirjalily *et al.* provided the analytical solution to this nonlinear equations in [11], given by

$$\begin{cases} P_1 = 0.5 - \frac{X}{2\sqrt{X^2+4}} \\ P_n^f = \gamma_n - a_n \sqrt{\frac{P_1}{1-P_1}} \\ P_n^m = 1 - \gamma_n - a_n \sqrt{\frac{1-P_1}{P_1}} \end{cases} \quad (4.11)$$

where,

$$X = \frac{\gamma_* - \gamma_1\gamma_2\gamma_3 - (\gamma_1|\delta_{23} - \gamma_2\gamma_3| + \gamma_2|\delta_{13} - \gamma_1\gamma_3| + \gamma_3|\delta_{12} - \gamma_1\gamma_2|)}{\sqrt{(\delta_{12} - \gamma_1\gamma_2)(\delta_{13} - \gamma_1\gamma_3)(\delta_{23} - \gamma_2\gamma_3)}} \quad (4.12)$$

$$\begin{cases} \gamma_1 = \sum_{j,k} P_{1jk} \\ \gamma_2 = \sum_{i,k} P_{i1k} \\ \gamma_3 = \sum_{i,j} P_{ij1} \\ \gamma_* = P_{111} \end{cases} \quad (4.13)$$

$$\begin{cases} \delta_{12} = P_{110} + P_{111} \\ \delta_{13} = P_{101} + P_{111} \\ \delta_{23} = P_{011} + P_{111} \end{cases} \quad (4.14)$$

$$\begin{cases} a_1 = \sqrt{\frac{(\delta_{12} - \gamma_1 \gamma_2)(\delta_{13} - \gamma_1 \gamma_3)}{\delta_{23} - \gamma_2 \gamma_3}} \\ a_2 = \sqrt{\frac{(\delta_{12} - \gamma_1 \gamma_2)(\delta_{23} - \gamma_2 \gamma_3)}{\delta_{13} - \gamma_1 \gamma_3}} \\ a_3 = \sqrt{\frac{(\delta_{13} - \gamma_1 \gamma_3)(\delta_{23} - \gamma_2 \gamma_3)}{\delta_{12} - \gamma_1 \gamma_2}} \end{cases} \quad (4.15)$$

By time-averaging, the unknown a priori probability and the probabilities of false alarm and missed detection can be estimated from the local decisions  $u_n$ .

However, the usage of Eq. (4.10) is limited when the number of decisions is not large enough. To estimate those unknown probabilities in this situation, let us first define the random variable  $X_\ell$  to represent the number of occurrences of the  $\ell$ th decision combination. Recall  $P_\ell$  is the corresponding occurrence probability, defined earlier in Eq. (4.6). Let  $\mathbf{X} = (X_1, X_2, \dots, X_L)$  denote the occurrence numbers of all eight decision combinations, which are multinomially distributed with probability

mass function [9]

$$P(X_1 = x_1, \dots, X_L = x_L | N_t) = \frac{N_t!}{x_1! \dots x_L!} P_1^{x_1} \dots P_L^{x_L} \quad (4.16)$$

and  $\text{var}(X_\ell) = N_t P_\ell (1 - P_\ell)$ ,  $\text{cov}(X_s X_\ell) = -N_t P_s P_\ell$  for  $s = 1, \dots, L$  and  $s \neq \ell$ .

We also define a 7-dimensional vector  $\mathbf{p}$ , named performance vector, contain the true values of the a priori probability and the false alarm and missed detection probabilities, i.e.,  $\mathbf{p} = [P(H_1) P_1^f P_2^f P_3^f P_1^m P_2^m P_3^m]$ . Thus, Eq. (4.6) can be rewritten in a short form as

$$P_\ell = f_\ell(\mathbf{p}) \quad (4.17)$$

which gives all the occurrence probabilities  $P_\ell$ , for  $\ell = 1, \dots, L$ .

Suppose  $z_\ell$  is the estimate of the  $\ell$ th occurrence probability and

$$z_\ell = f_\ell(\mathbf{p}) + e_\ell, \quad \ell = 1, \dots, L \quad (4.18)$$

where  $e_\ell$  is the estimation error. Now we define a vector  $\mathbf{z} = [z_1 z_2 \dots z_L]^T$ ,  $\mathbf{f}(\mathbf{p}) = [f_1(\mathbf{p}) f_2(\mathbf{p}) \dots f_L(\mathbf{p})]^T$ , and  $\mathbf{e} = [e_1 e_2 \dots e_L]^T$ . Thus, the nonlinear system (4.9) can be rewritten in the matrix format as

$$\mathbf{z} = \mathbf{f}(\mathbf{p}) + \mathbf{e} \quad (4.19)$$

where  $\mathbf{z}$ ,  $\mathbf{f}(\mathbf{p})$  and  $\mathbf{e}$  are the matrices of the estimates of the occurrence probabilities, their true values, and the estimation error, respectively.

Since the distribution of the occurrences of the decision combinations is given

by Eq. (4.16), we could apply maximum likelihood estimator to find the unknown parameters which make the observed outcome most likely to happen. It means that as long as the occurrence numbers are known, the ML estimator gives the value of  $\mathbf{p}$  that maximize Equation (4.16). The ML estimator can be written as

$$\hat{\mathbf{p}} = \arg \max_{\mathbf{p}} P(X_1 = x_1, \dots, X_L = x_L | N_t, \mathbf{p}) \quad (4.20)$$

# Chapter 5

## Results and Discussions

In this chapter, we present numerical results in order to show the applicability of the blind adaptive algorithms to the neonatal seizure detection problem. Due to a possibly non-stationary nature of the EEG signals, time-dependent approach may be needed in order to correctly estimate the time-varying parameters. In this case, the number of decisions available for estimation may be limited. To this purpose, we evaluate the algorithm for two scenarios: small sample size and large sample size. We also perform tests using the surrogate data generated by the model introduced in Chapter 2 to show the improvement of the detection system.

### 5.1 Simulations Results

#### 5.1.1 Large Data Set

Recall that we defined the performance vector as  $\mathbf{p} = [P(H_1) P_1^f P_2^f P_3^f P_1^m P_2^m P_3^m]$  in the previous chapter. In this example, we generate the binary local decisions  $u_n$

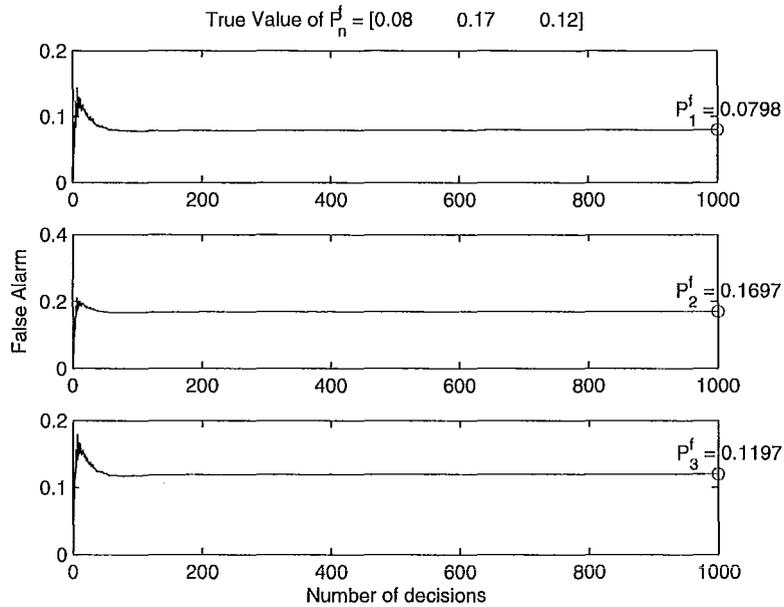


Figure 5.1: Averaged False Alarm Rate over 5000 Realizations

by using an arbitrarily chosen  $\mathbf{p} = [0.2 \ 0.08 \ 0.17 \ 0.12 \ 0.23 \ 0.18 \ 0.15]$ . Applying the blind adaptive algorithm [11], the global decision  $u_0$  is then obtained by calculating the non-linear set of equations. The decision number  $N_t$  is set to be 1000 and the simulation is performed 5000 times. The estimated unknown probabilities of false alarm  $P_n^f$  and missed detection  $P_n^m$  are then averaged over 5000 realizations, and these are shown in Figure 5.1 and 5.2, respectively. As expected these plots show that the estimated values of the probabilities converge to their true values.

In Figure 5.3, the upper plots show the averaged error probabilities of the local detectors and the lower plot shows the averaged overall error probability of the system. It is clear that by fusing the detection probabilities the overall performance of the detection system is much better than any of the local detectors in terms of low error probability.

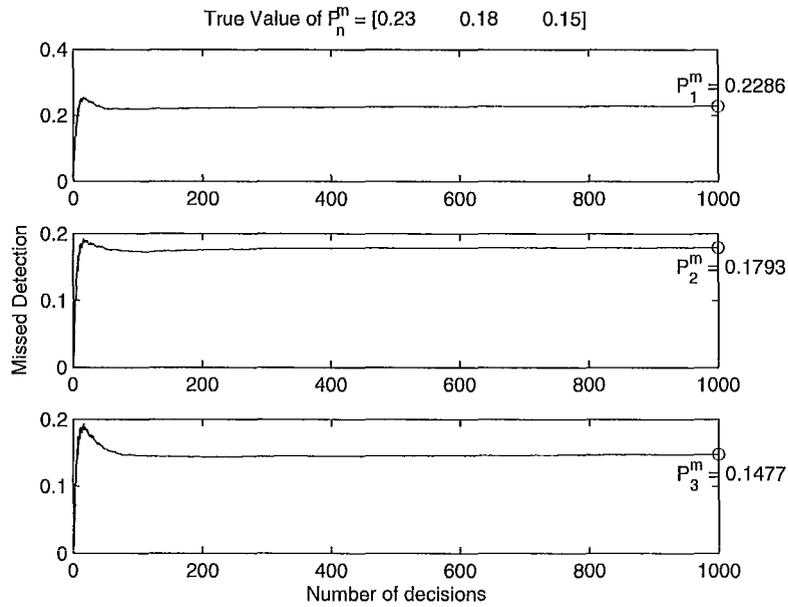


Figure 5.2: Averaged Missed Detection Rate over 5000 Realizations

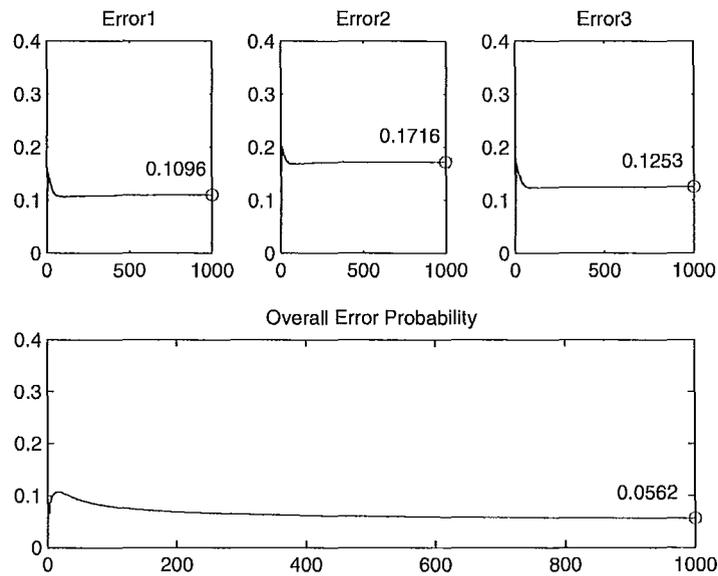


Figure 5.3: Averaged Error Probabilities over 5000 Realizations

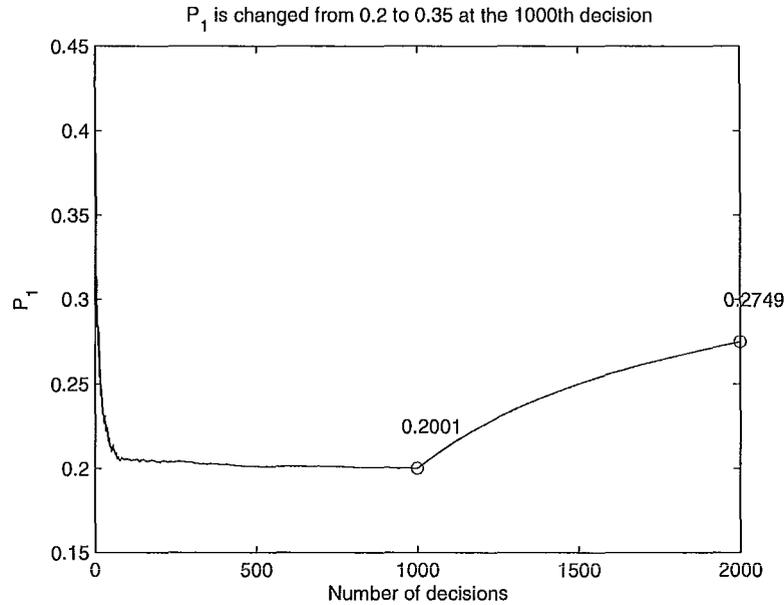


Figure 5.4: Estimated  $\hat{P}_1$  when the Statistical Property of the Signal Change

### 5.1.2 Adaptability to the Changes of Phenomenon

As we discussed before, the statistical properties of the neonatal EEG signals are time-dependent. We present a numerical example to show the adaptability of the blind algorithm. We use the same performance vector  $\mathbf{p}$  to generate the local decision  $u_n$ , except that the a priori probability  $P_1$  is changed from 0.2 to 0.35 at the 1000th decision. The total number of decisions is 2000 and the simulation is repeated 1000 times. From Figure 5.4, it can be seen that the algorithm can adapt to the changes of the unknown a priori probability.

However, since the algorithm is based on the time-averaging, it adapts to the changes quite slowly. To increase the speed of the convergence, we introduce the windowed approach. Suppose the length of the sliding window is  $N_s$ . At the  $N_i$ th decision,  $N_i > N_s$ , we use the previous  $N_s$  decisions including the  $N_i$ th decision for

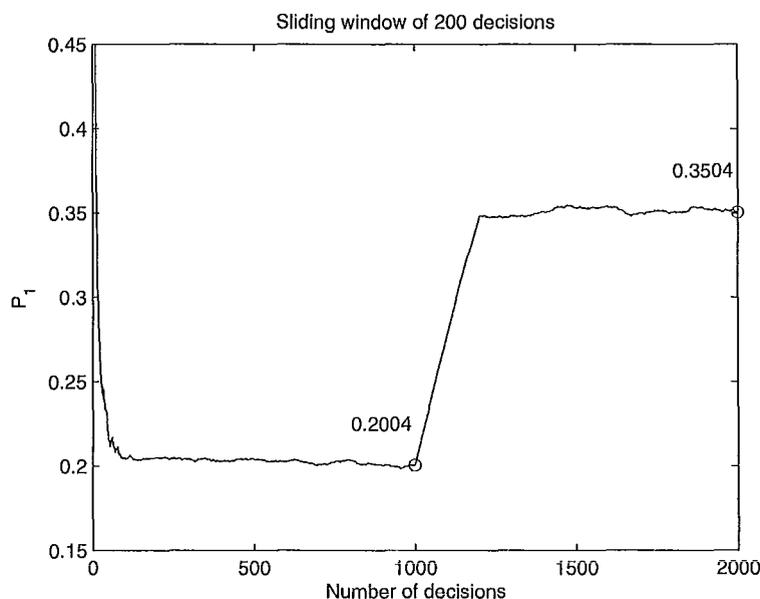


Figure 5.5: Estimated  $\hat{P}_1$  Using a Sliding Window

estimation instead of using all  $N_i$  decisions. Figure 5.5 provides the numerical result of using a sliding window of length 200. The plot shows the averaged value of  $\hat{P}_1$  over 1000 realizations. The rate of the convergence depends on the size of the sliding window. As a consequence, an effort is needed to determine an adequate window size for a particular dynamic of the system.

### 5.1.3 Small Data Set

In Section 4.3, we propose to estimate the unknown probabilities using the maximum likelihood estimator [9] when the size of the data set is small. Now, we present numerical comparison in order to show the effectiveness of the ML estimator. In Figure 5.6, the estimates obtained from the ML estimator converge to its true value much faster. Note that it is expected since the ML estimation accounts for statistical

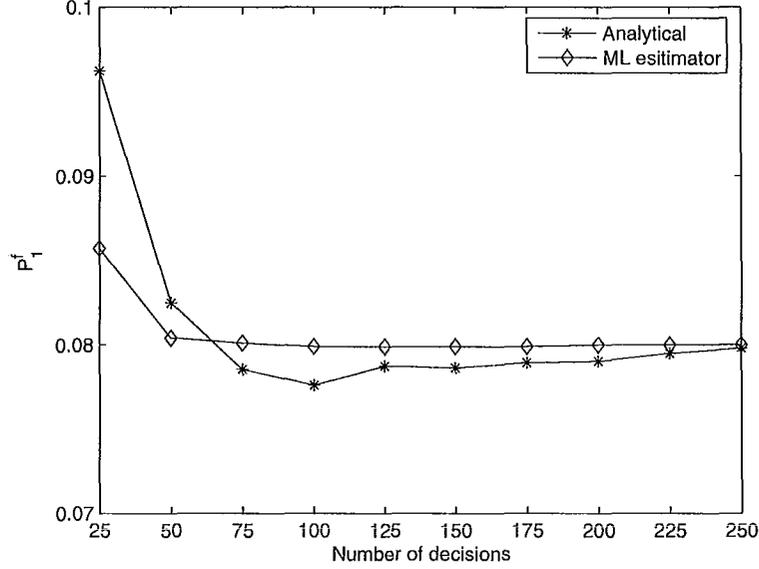


Figure 5.6: Comparison

properties of the error, while the trade off is computationally much more complex. The term analytical in Figure 5.6 means that the unknown probabilities are estimated using the analytical solution of the non-linear equation set under the assumption that the occurrence probabilities could be replaced by the empirical probabilities.

## 5.2 System Performance with Surrogate Data

In this section, we first use the model reviewed in Section 2.1 to simulate neonatal EEG signal. The occurrence rate of EEG seizure is close to 0.2 and the length of the testing signal is about 1 hour. The generated neonatal EEG signal is fed into the local detectors  $LD_n$ , where the binary local decisions  $u_n$  are made. The global decision  $u_0$  is made by applying the blind adaptive algorithm. The unknown probabilities of false alarm  $P_n^f$  and missed detection  $P_n^m$  are shown in Figure 5.7 and 5.8, respectively. It

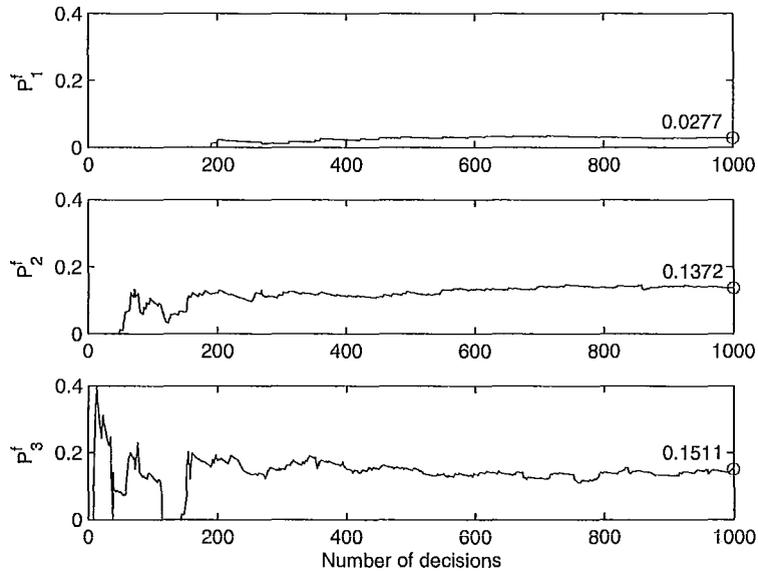


Figure 5.7: False Alarm Rate of the Local Detectors

is clear that the local detector  $LD_1$  is good in the sense of low false alarm rate and the local detector  $LD_3$  is good in the sense of low missed detection rate. It is a good numerical example to show the effectiveness of the detection system.

In Figure 5.9, the upper plots shows the error probabilities of the local detectors and the lower plot shows the overall error probability of the system. The distributed system has been improved by 2.3% (difference between 13.8% and 11.5%) compared with the local detector who performs the best individually.

### 5.3 Summary

The proposed technique allows us to evaluate the performance of the local detectors for any given patient. Namely the performance of the local detectors may vary significantly from patient to patient. Therefore, the probabilities of false alarm and

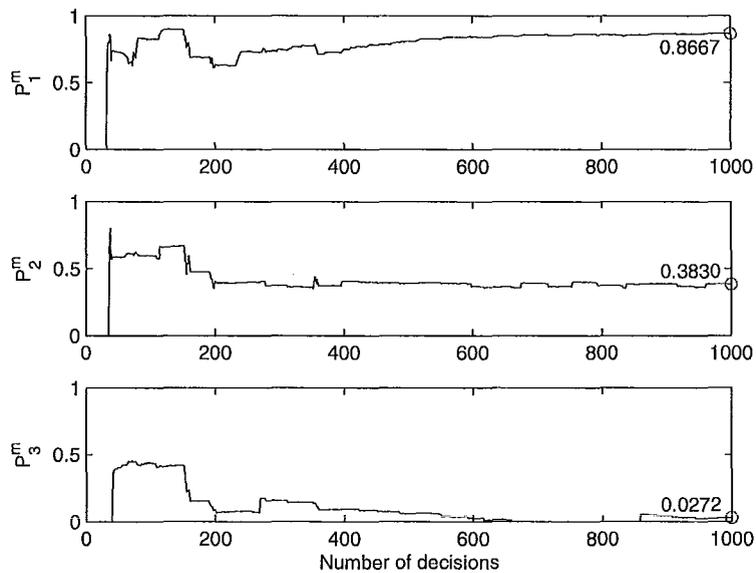


Figure 5.8: Missed Detection Rate of the Local Detectors

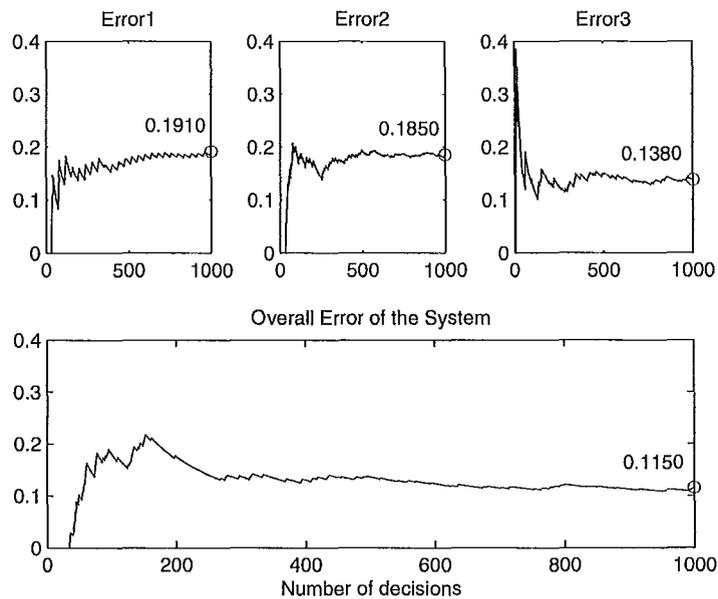


Figure 5.9: The Overall Error Probability of the Detection System

missed detection listed in tables and figures are random variables. Therefore, this uncertainties can significantly deteriorate the performance. Our algorithm has the potential of accounting for this randomness in an optimal way.

# Chapter 6

## Conclusions

In this thesis, we proposed a parallel distributed detection system for neonatal seizure detection problem using the blind adaptive fusion algorithms. The advantage of our technique is that it does not require any a priori probabilities of the hypotheses or the performance of the local detectors, which are usually unavailable in practice, especially the biomedical applications.

We first discussed two EEG models for simulating neonatal EEG signals. The first model was used to generate synthetic data in order to evaluate the proposed technique. The second model was the basics of one of the existing detection algorithms. We also presented three well-known neonatal seizure detection methods, each of which can be considered as a single seizure detector.

We then described the parallel structure of the system which enables us to combine heterogeneous detectors into one system, followed by introducing its components: the local detectors and the fusion centre. In practice, since the size of EEG data from the patients may be limited, we consider two cases: large data set and small data set. In the first case, we applied the blind algorithm proposed in [11] in the

fusion center, which solves the non-linear equation set formulated by the unknown probabilities. In the second case, we applied the blind algorithm proposed in our previous work [9], which uses maximum likelihood estimator to estimate the unknown probabilities. Note that since the EEG signal is non-stationary, it may require the windowed approach. Thus, the small data set may be the only option.

Further, we provided the numerical examples to show the effectiveness and applicability of the blind algorithms in the seizure detection application. We performed tests for both aforementioned cases of small and large sample sizes. We also demonstrated the decrease of the overall probability of error of the existing seizure detection algorithms by efficiently fusing their decisions.

The future research will include the following topics:

1. We are currently implementing the proposed local detectors and fusion algorithm on real neonatal EEG data sets.
2. An effort should be made to investigate the possibility of developing improved seizure detectors.
3. We can extend the hypothesis testing to allow for more hypotheses. For example, when  $M = 3$ , the possible hypotheses could be

$H_0$  : No seizure in the signal

$H_1$  : Seizures in the signal

$H_2$  : Not sure if there is seizure in the signal

4. An effort should be made to derive a statistically optimal detector to detect the

changes in phenomenon or the changes in the local detectors. For example, the adaptability shown in Figure 5.4 can occur faster using such a detector.

5. By developing an algorithm for automatic counting of the number of seizures, we can correlate the frequency of seizures with the brain development in neonates with cerebral pathologies.

# Bibliography

- [1] Celka, P. and Colditz, P. (2002a). A computer-aided detection of EEG seizures in infants: a singular-spectrum approach and performance comparison. *IEEE transactions on biomedical engineering*, **49**(5), 455–462.
- [2] Celka, P. and Colditz, P. (2002b). Nonlinear nonstationary Wiener model of infant EEG seizures. *IEEE Transactions on Biomedical Engineering*, **49**(6), 556–564.
- [3] Clancy, R. (1989). Interictal sharp EEG transients in neonatal seizures. *Journal of Child Neurology*, **4**(1), 30.
- [4] Faul, S., Boylan, G., Connolly, S., Marnane, L., and Lightbody, G. (2005). An evaluation of automated neonatal seizure detection methods. *Clinical Neurophysiology*, **116**(7), 1533–1541.
- [5] Gotman, J., Flanagan, D., Zhang, J., and Rosenblatt, B. (1997). Automatic seizure detection in the newborn: methods and initial evaluation. *Electroencephalography and clinical neurophysiology*, **103**(3), 356–362.
- [6] Jerger, K., Netoff, T., Francis, J., Sauer, T., Pecora, L., Weinstein, S., and Schiff, S. (2001). Early seizure detection. *Journal of Clinical Neurophysiology*, **18**(3), 259.

- [7] Lee, Y., Cheatham Jr, T., and Wiesner, J. (1950). Application of correlation analysis to the detection of periodic signals in noise. *Proceedings of the IRE*, **38**(10), 1165–1171.
- [8] Liu, A., Hahn, J., Heldt, G., and Coen, R. (1992). Detection of neonatal seizures through computerized EEG analysis. *Electroencephalography and clinical neurophysiology*, **82**(1), 30.
- [9] Liu, B., Jeremic, A., and Wong, K. (2007). Blind adaptive algorithm for M-ary distributed detection. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007*, volume 2.
- [10] Liu, B., Jeremic, A., and Wong, K. (2010). Optimal distributed detection of multiple hypotheses using blind algorithm. *to appear in IEEE Transactions on Aerospace and Electronic Systems*.
- [11] Mirjalily, G., Luo, Z., Davidson, T., and Bosse, E. (2003). Blind adaptive decision fusion for distributed detection. *IEEE Transactions on Aerospace and Electronic Systems*, **39**(1), 34–52.
- [12] Pritsker, A. (1989). Why simulation works. In *Proceedings of the 21st conference on Winter simulation*, page 9. ACM.
- [13] Rankine, L., Stevenson, N., Mesbah, M., and Boashash, B. (2007). A nonstationary model of newborn EEG. *IEEE Transactions on Biomedical Engineering*, **54**(1), 19–28.
- [14] Sanei, S. and Chambers, J. (2007). *EEG signal processing*. Wiley-Interscience.

- 
- [15] Varshney, P. (1986). Optimal data fusion in multiple sensor detection systems. *IEEE Transactions on Aerospace and Electronic Systems*, pages 98–101.
- [16] Varshney, P. and Burrus, C. (1997). *Distributed detection and data fusion*. Springer Verlag.
- [17] Viswanathan, R. and Varshney, P. (1997). Distributed detection with multiple sensors I. Fundamentals. *Proceedings of the IEEE*, **85**(1), 54–63.
- [18] Volpe, J. (2001). *Neurology of the newborn*. WB Saunders Co.