NEONATAL SEIZURE DETECTION USING BLIND ADAPTIVE FUSION

Huaying Li and Aleksandar Jeremic
Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada

Keywords: Seizure detection, Distributed detection, Neonatal seizures, Biomedical signal processing.

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 in stead of the expensive on-site supervision. However, it is a waste of time 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 algorithm 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 this proposed technique is that it does not require any priori knowledge of the hypotheses and 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 paper, 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 on this topic and implement them to work as the local detectors of the system. 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.

1 INTRODUCTION

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 (Volpe, 2001).

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

Li H. and Jeremic A..
NEONATAL SEIZURE DETECTION USING BLIND ADAPTIVE FUSION.
DOI: 10.5220/0003294903650371
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing* (BIOSIGNALS-2011), pages 365-371
ISBN: 978-989-8425-35-5

single seizure detectors to form a distributed detection system and apply our previously proposed blind algorithm on multichannel information fusion. 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 multinominal 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 (Rankine et al., 2007) to generate neonatal EEG signals.

2 SIGNAL MODEL

2.1 Local Detectors

Several neonatal EEG seizure detection algorithms exist in the literature. In this paper we implemented the following three algorithms that have been proposed for the neonatal seizure detection:

Liu's Algorithm. In(Liu, 1992) the authors focused on the rhythmic characteristic of neonatal EEG seizure and proposed a detection algorithm using autocorrelation analysis. 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. A scoring system described in (Liu, 1992) can be used to determine the degree of periodicity of the EEG signal quantitatively in order to identify the existences of the seizure activities.

Gotmans's Algorithm. In (Gotman, 1997) the authors proposed three different seizure detection methods to detect three types of seizures: rhythmic discharges, multiple spikes, and very slow rhythmic discharges, respectively. In this paper, we only focus on the rhythmic discharge detection since it could identify 90% of the seizures detected by all three detection algorithms. 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.

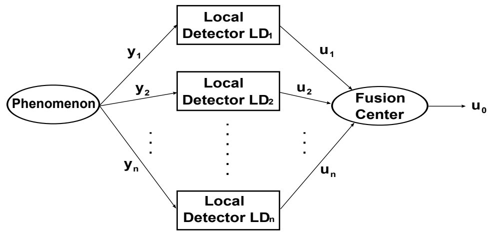

Figure 1: Parallel Distributed Detection System.

Celka's Algorithm. The algorithm reviewed in this section was proposed in (Celka and Colditz, 2002). 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: Preprocessing, singular spectrum analysis, and minimum description length.

2.2 Distributed Detection System

Each of the algorithms reviewed in the previous section can be considered as a single detector. Since the statistical properties of neonatal EEG can 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. Thus, it motivates us to combine the existing single detectors and utilize their strengths by extending previous results on blind multichannel information fusion (Liu et al., 2007). Figure 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.

2.3 Local Detectors

The local detectors LD_n have their own decision rules. We use the three algorithms reviewed in Section 2.1 to formulate the local decision rules.

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

H_0 : The EEG signal does not contain seizure

$$H_1: \quad \text{The EEG signal contains seizure}$$

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

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

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., $M = 3$.

2.4 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

$$P_e = P(H_0)P(u_0 = 1|H_0) + P(H_1)P(u_0 = 0|H_1) \tag{2}$$

The authors provided the optimality criterion for N local detectors in the sense of minimum error probability in (Varshney, 1986). We recall it here for the case of $N = 3$.

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

$$\text{where,} \quad w_0 = \log\left(\frac{P_1}{P_0}\right) \tag{4}$$

$$\text{and} \quad w_n = \begin{cases} \log((1-P_n^m)/P_n^f), & \text{if } u_n = 1 \\ \log(P_n^m/(1-P_n^f)), & \text{if } u_n = 0 \end{cases} \tag{5}$$

The probabilities of false alarm and missed detection of the nth 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 (Mirjalily, 2003; Liu et al., 2007). 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 $\{u_1 = i,\ u_2 = j$ and $u_3 = k\}$ is represented by $\ell = (ijk)_2$, where $i, j, k = 0$ or 1 (Mirjalily, 2003). 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 paper. The joint probability of decision $\{u_1 = i,$ $u_2 = j$ and $u_3 = k\}$ is also the occurrence probability of the ℓth decision combination, given by

$$\begin{aligned} P_\ell &= \Pr(u_1 = i, u_2 = j, u_3 = k) \\ &= P(u_1 = i|H_1)P(u_2 = j|H_1)P(u_3 = k|H_1)P_1 \\ &\quad + P(u_1 = i|H_0)P(u_2 = j|H_0)P(u_3 = k|H_0)(1-P_1) \end{aligned} \tag{6}$$

$$P(u_n = i|H_1) = \begin{cases} 1 - P_n^m, & \text{if } i = 1 \\ P_n^m, & \text{if } i = 0 \end{cases} \tag{7}$$

$$P(u_n = i|H_0) = \begin{cases} P_n^f, & \text{if } i = 1 \\ 1 - P_n^f, & \text{if } i = 0 \end{cases} \tag{8}$$

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_ℓ are known. Although P_ℓ is usually unavailable in practice, it could be replaced by empirical probability defined as

$$\begin{aligned} P_\ell &= \Pr(u_1 = i, u_2 = j, u_3 = k) \\ &\simeq \frac{\text{number of } u_1 = i, u_2 = j, u_3 = k}{\text{number of local decisions } N_t} \end{aligned} \tag{9}$$

where N_t is the number of decisions made by one of the local detectors. Eq. (9) is true usually when the number of decisions is large (Liu et al., 2007).

The analytical solution to the above nonlinear equations is given in (Mirjalily, 2003). However, the usage of Eq. (9) is limited when the number of decisions is not large enough. In our particular case the number of seizures occurring can be rather small and thus can yield inaccurate estimation results. To estimate those unknown probabilities in this situation, let us first define the random variable X_ℓ to represent the number of occurrences of the ℓth decision combination. Recall P_ℓ is the corresponding occurrence probability, defined earlier in Eq. (6). Let $\mathbf{X} = (X_1, X_2, \ldots, X_L)$ denote the occurrence numbers of all eight decision combinations, which are multinomially distributed with probability mass function (Liu et al., 2007)

$$P(X_1 = x_1, \ldots, X_L = x_L|N_t) = \frac{N_t!}{x_1! \ldots x_L!} P_1^{x_1} \ldots P_L^{x_L} \tag{10}$$

and $\mathrm{var}(X_\ell) = N_t P_\ell (1 - P_\ell)$, $\mathrm{cov}(X_s X_\ell) = -N_t P_s P_\ell$ for $s = 1, \ldots, L$ 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., $\mathbf{p} = [P(H_1)\ P_1^f\ P_2^f\ P_3^f\ P_1^m\ P_2^m\ P_3^m]$. Suppose z_ℓ is the estimate of the ℓth occurrence probability and

$$z_\ell = f_\ell(\mathbf{p}) + e_\ell, \quad \ell = 1, \dots, L \tag{11}$$

where e_ℓ 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 of probability equations can be rewritten in the matrix format as

$$\mathbf{z} = \mathbf{f}(\mathbf{p}) + \mathbf{e} \tag{12}$$

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. (10), 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 (10).

3 NUMERICAL RESULTS

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 models described in Section 2.1 to show the improvement of the detection system.

3.1 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 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 (Mirjalily, 2003), the global decision u_0 is then obtained by calculating the nonlinear 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

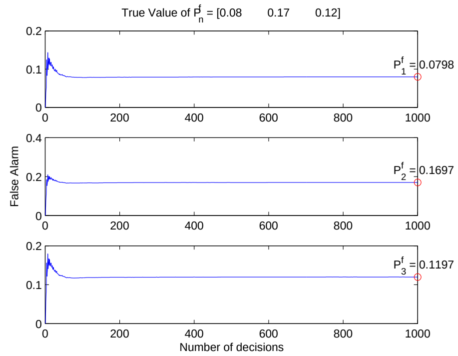

Figure 2: Averaged False Alarm Rate over 5000 Realizations.

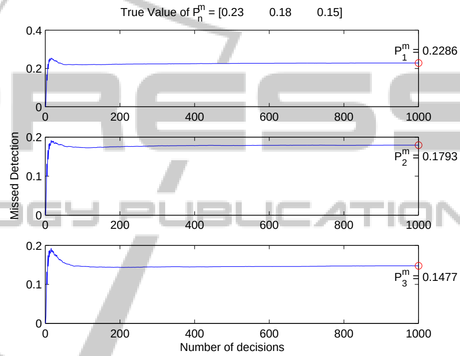

Figure 3: Averaged Missed Detection Rate over 5000 Realizations.

over 5000 realizations, and these are shown in Figure 2 and 3, respectively. As expected these plots show that the estimated values of the probabilities converge to their true values.

In Figure 4, 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.

3.2 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, it can be seen that the

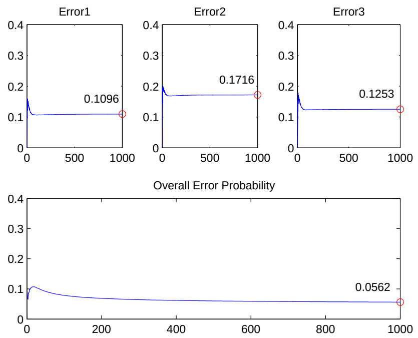

Figure 4: Averaged Error Probabilities over 5000 Realizations.

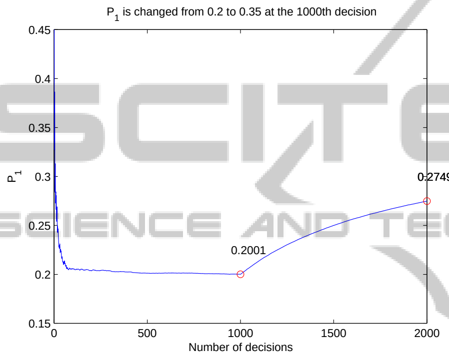

Figure 5: Estimated $\hat{P}_1$ when the Statistical Property of the Signal Change.

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_ith decision, $N_i > N_s$, we use the previous N_s decisions including the N_ith decision for estimation instead of using all N_i decisions. Figure 6 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.

3.3 Small Data Set

In Section 2.4, we propose to estimate the unknown probabilities using the maximum likelihood estimator (Liu et al., 2007) 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 7, the estimates obtained from the ML es-

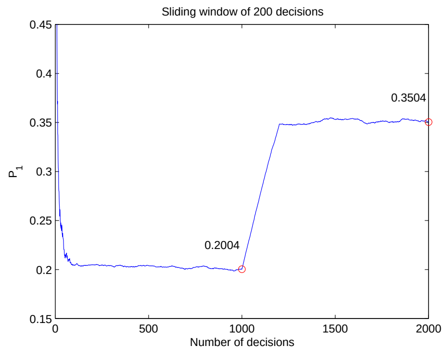

Figure 6: Estimated $\hat{P}_1$ Using a Sliding Window.

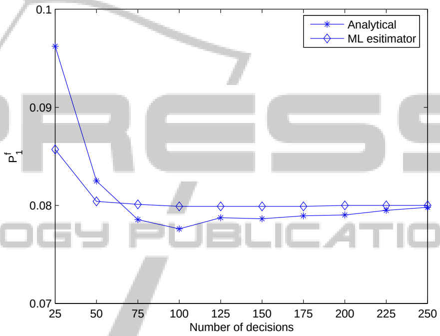

Figure 7: Comparison.

timator converge to its true value much faster. Note that it is expected since the ML estimation accounts for statistical properties of the error, while the trade off is computationally much more complex. The term analytical in Figure 7 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.

4 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 8 and 9, respectively. It 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

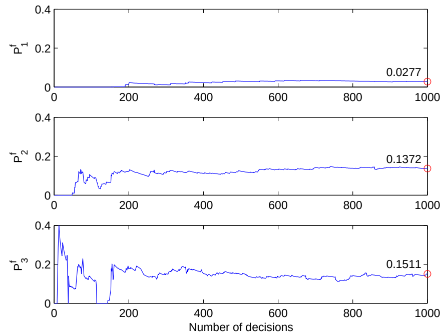

Figure 8: False Alarm Rate of the Local Detectors.

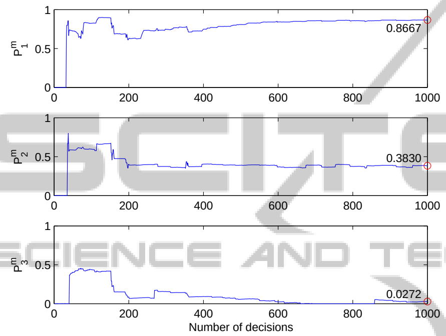

Figure 9: Missed Detection Rate of the Local Detectors.

of low missed detection rate. It is a good numerical example to show the effectiveness of the detection system.

In Figure 10, 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.

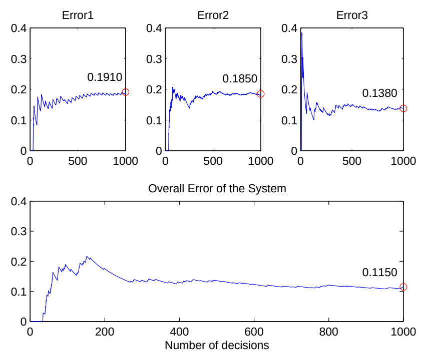

Figure 10: The Overall Error Probability of the Detection System.

5 CONCLUSIONS

In this paper, 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 (Mirjalily, 2003) 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 (Liu et al., 2007), 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 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.

REFERENCES

Celka, P. and Colditz, P. (2002). A computer-aided detection of EEG seizures in infants: asingular-spectrum approach and performance comparison. *IEEE Trans. on biomedical engineering*, 49(5):455–462.

Gotman, J. e. a. (1997). Automatic seizure detection in the newborn: methods and initial evaluation. *Electroencephalography and clinical neurophysiology*, 103(3):356–362.

Liu, A. e. a. (1992). Detection of neonatal seizures through computerized EEG analysis. *Electroencephalography and clinical neurophysiology*, 82(1):30.

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.

Mirjalily, G. e. (2003). Blind adaptive decision fusion for distributed detection. *IEEE Transactions on Aerospace and Electronic Systems*, 39(1):34–52.

Rankine, L., Stevenson, N., Mesbah, M., and Boashash, B. (2007). A nonstationary model of newborn EEG. *IEEE Trans. on Biomed. Eng.*, 54(1):19–28.

Varshney, P. (1986). Optimal data fusion in multiple sensor detection systems. *IEEE Trans. on Aerospace and Electronic Systems*, pages 98–101.

Volpe, J. (2001). *Neurology of the newborn*. WB Saunders Co.