An Investigation of the Outlier Processing Method for Single-Trial Event-Related Potentials

by

SAMINA TAJWAR

B.A.Sc., The University of Waterloo, 1991

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF APPLIED SCIENCE in

THE FACULTY OF GRADUATE STUDIES
DEPARTMENT OF ELECTRICAL ENGINEERING

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
December 1995
© Samina Tajwar, 1995
In presenting this thesis in partial fulfillment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of **Electrical Engineering**

The University of British Columbia
Vancouver, Canada

Date **Dec 16/95**
This thesis is an investigation of the Outlier Processing Method, which was developed to eliminate the dependence of conventional techniques on a priori information for the detection of event-related potentials (ERPs) from EEG signals. Instead of attempting to model the ERP, the OPM assumes the ERP to be an outlier signal which is imbedded in a background EEG signal. The background EEG is modelled as an auto-regressive signal and approximated using robust statistical parameter and signal estimation. The ERP estimate is obtained by subtracting the estimate of the background EEG signal from the observed EEG signal.

The basis of the robust statistical parameter estimation used in OPM was the generalized maximum likelihood (GM) estimate. It required the generation of initial parameter estimates to start the process. MEM, a least squares estimator used in the original OPM, was found to be a more consistent estimator when compared to median regression estimation. Also, a method of automatically generating the tuning constants used by the influence functions in the GM estimate was proposed.

The robust signal estimation was a robust variation of a Kalman filter which required a cleaner function to minimize the effect of outliers such as ERPs. Of the three different cleaners that were evaluated, the fixed-lag cleaner had the best overall performance. A method of detecting the location of the outlier was developed and was found to be effective for signals with low SNRs and a large amount of spectral overlap between the outlier and the background EEG signal.
The improvements to the OPM were incorporated into a new algorithm referred to as OPM2 and tested against the original OPM, referred to as OPM1, and MMSE, a previously published method used as a benchmark. The three algorithms were tested on real EEG signals of a ballistic movement of the index finger.

All three techniques successfully extracted information related to the finger movement. However, both OPM1 and OPM2 were successful without any a priori information, whereas MMSE uses a filter based on the average of the raw data. OPM1 and OPM2 were significantly better than MMSE in rejecting false positives.
Table of Contents

ABSTRACT .................................................. ii

List of Tables ............................................. vii

List of Figures ............................................ viii

Chapter 1) INTRODUCTION ............................... 1

1.1 Background ........................................... 1

1.2 Goal ..................................................... 3

Chapter 2) ROBUST PARAMETER ESTIMATION ............ 4

2.1 The EEG Model ......................................... 4

2.2 Initial Parameter Estimates ............................. 6

2.3 The Influence Function and Its Tuning Constants ....... 8

2.3.1 M-Estimate .......................................... 8

2.4 Automatic Generation of Tuning Constants ............. 12

2.4.1 Asymptotic Theory .................................. 12

2.4.2 Asymptotic Variance Function ...................... 13

2.4.3 Adaptation to the OPM ............................. 17

2.5 Summary of Issues ...................................... 23
Chapter 3) ROBUST SIGNAL ESTIMATION ............................................. 25

3.1 Filter-Cleaner ................................................................. 26

3.2 Fixed-Lag and Smoother Cleaners ................................. 29

3.2.1 Fixed-Lag Cleaner .................................................... 29

3.2.2 Smoother Cleaner .................................................... 30

3.3 Time-Varying Influence Function .................................... 31

3.3.1 SNR and Spectral Overlap .......................................... 32

3.4 Summary of Issues .......................................................... 34

Chapter 4) THE IMPROVED OPM ................................................. 35

4.1 Data Simulation ............................................................. 35

4.2 Initial Parameter Estimates ............................................. 37

4.3 Automatic Generation of Tuning Constants ..................... 41

4.4 Cleaners ........................................................................ 44

4.5 Time Varying Influence Function ..................................... 47

4.6 Summary ........................................................................ 52

Chapter 5) TESTS ON REAL EEG .............................................. 53

5.1 Analysis of EEG signals .................................................. 53

5.1.1 Ensemble Averages .................................................... 53

5.1.2 SNR ........................................................................ 56

5.1.3 Ranking Trials ............................................................ 57
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2 Test Procedure</td>
<td>59</td>
</tr>
<tr>
<td>5.2.1 MMSE</td>
<td>60</td>
</tr>
<tr>
<td>5.2.2 Performance Measures</td>
<td>61</td>
</tr>
<tr>
<td>5.3 Results</td>
<td>64</td>
</tr>
<tr>
<td>5.3.1 Ensemble Averages and Single-Trials</td>
<td>64</td>
</tr>
<tr>
<td>5.3.2 Performance Measurements</td>
<td>72</td>
</tr>
<tr>
<td>Chapter 6) CONCLUSIONS</td>
<td>76</td>
</tr>
<tr>
<td>6.1 Contributions</td>
<td>76</td>
</tr>
<tr>
<td>6.2 Future Work</td>
<td>78</td>
</tr>
<tr>
<td>References</td>
<td>80</td>
</tr>
</tbody>
</table>
List of Tables

Table 1  MSEs averaged over 100 trials ................. 45
Table 2  Difference in MSEs between cleaners. ........ 46
Table 3  The AR parameters for the different outlier sequences. . . . 47
Table 4  Average and maximum SNRs at Fz, C3 and Fz-C3 . . . . 57
Table 5  Results of DTW cost comparisons ............... 75
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>OPM block diagram</td>
<td>3</td>
</tr>
<tr>
<td>Figure 2</td>
<td>The Huber Psi Function</td>
<td>11</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Tukey's Biweight Psi Function</td>
<td>17</td>
</tr>
<tr>
<td>Figure 4</td>
<td>GM2</td>
<td>18</td>
</tr>
<tr>
<td>Figure 5</td>
<td>A block diagram of the IWLS procedure for GM.</td>
<td>20</td>
</tr>
<tr>
<td>Figure 6</td>
<td>A block diagram of GM-estimation including the automatic generation of tuning constants. The dashed boxes signify the proposed changes to the OPM.</td>
<td>24</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Hampel's Psi Function</td>
<td>27</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Comparison of Estimated Parameters between MR and MEM</td>
<td>38</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Percentage of best trials between MEM and MR</td>
<td>39</td>
</tr>
<tr>
<td>Figure 10</td>
<td>The sum of the squared errors for the varying contamination levels.</td>
<td>40</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Comparison of the parameter estimates generated by GM and GM_agtc.</td>
<td>42</td>
</tr>
<tr>
<td>Figure 12</td>
<td>Percentage of best trials between GM and GM_agtc.</td>
<td>43</td>
</tr>
<tr>
<td>Figure 13</td>
<td>Percentage of Best Trials by each Cleaner</td>
<td>44</td>
</tr>
<tr>
<td>Figure 14</td>
<td>The percentage of best trials, excluding the smoother-cleaner.</td>
<td>46</td>
</tr>
</tbody>
</table>
Figure 15  The spectral plots of the four different outlier sequences corresponding to the AR parameters listed in Table 3. The spectrum of the background sequence is the dotted line in each plot. ................................. 47

Figure 16  The MSEs of GM2 and MEM for the three different types of outliers. ........................................ 49

Figure 17  Examples of CD profiles generated from observed signals with −5dB SNR and outlier no.3. .......... 51

Figure 18  Ensemble averages of the active and idle trials at electrodes $F_z$ and $C_3$. ............................ 55

Figure 19  Test Procedure .................................................. 60

Figure 20  Classification with Dynamic Time Warping ........ 63

Figure 21  Ensemble averages of the active ERP estimates .... 65

Figure 22  Ensemble averages of the active ERP estimates .... 66

Figure 23  Examples of single trials of active ERP estimates .... 68

Figure 24  Examples of single trials of active ERP estimates .... 69

Figure 25  Examples of single trials of idle ERP estimates ...... 70

Figure 26  Examples of single trials of idle ERP estimates ...... 71

Figure 27  Standard template used for DTW for OPM1 estimates ... 73

Figure 28  Standard template used for DTW for OPM2 estimates ... 73

Figure 29  Standard template used for DTW for MMSE estimates ... 74
Chapter 1
INTRODUCTION

Advances in assistive technology have the potential to provide some levels of independence for people with severe movement-related disabilities. But, often, the effectiveness of current technology is limited by the inadequacies of the human-machine interface (HMI). The ideal universal HMI coupled with recent advances in technology for peripheral devices could provide a much improved life-style for people of this group. This research is an attempt to put us one step closer to developing such an interface.

1.1 Background

Electrical brain signals are one of the many human physiological signals that could be used for an ideal HMI. Movement information from signals elicited from the brain could be directly utilized to communicate with peripheral devices such as a robotic arm.

The electrical activity of the brain can be recorded in a non-invasive manner with an electroencephalograph (EEG). EEG potentials, measured at the surface of the scalp, represent the combined effect of potentials from the cerebral cortex (the cellular layer forming the top layer of the brain) and various points beneath. The continuous neural activity of a person will be referred to as background EEG. When an event occurs, such as a movement or a response to a stimulus, the brain generates a characteristic signal which is referred to as an event-related potential (ERP). ERPs are transient signals imbedded in the non-stationary background EEG. In addition, the SNR of an ERP lies below 0dB [1]. Given these conditions, researchers have been challenging the problem of ERP extraction
for decades with varying levels of success.

Initial studies modelled the ERP as a deterministic function for which the best estimate could be achieved by averaging over a number of responses, assuming the background EEG was random and zero mean. However, it was recognized by several independent researchers [2], [3], [4], that ERP components vary in amplitude and latency from trial to trial rendering the ensemble averaged model as an inaccurate representation of an ERP. Although, it does provide a general idea of the underlying ERP waveform, there is a significant loss of unique single-trial information.

Current research has been focused towards developing signal processing methods for ERP extraction on a single-trial basis. To date, researchers have based their single-trial techniques on using training sets of data to obtain averaged templates or statistical characteristics of ERPs. This included methods incorporating filtering [5], correlation matrices [6] and neural networks [7]. Although these methods have achieved some level of success, their reliance on a priori knowledge of the ERP would limit the scope of a universal human-machine interface.

Birch et. al [8] developed a method, called the Outlier Processing Method (OPM), that successfully extracted ERPs without any requirement for a priori information of the ERP. Instead of attempting to model the ERP, Birch modelled the background EEG signal. The algorithm was based on the assumption that ERPs were additive outliers which did not conform to the structure of the background EEG.
1.2 Goal

The OPM is basically comprised of two fundamental elements. The first generates a set of robust autoregressive parameters to capture the statistical structure of the background EEG. Based on these parameters, the second element, a signal estimator, generates an estimate of the background signal. In the final step, the difference is taken between the background estimate and the observed signal to reveal the outlier content which contains ERP information.

Figure 1: OPM block diagram

The goal of this thesis was to independently analyze and improve the robust parameter and signal estimation to lead to the development of a new OPM in terms of accuracy, efficiency and facility.
Chapter 2
ROBUST PARAMETER ESTIMATION

2.1 The EEG Model

The OPM was developed by Birch in 1988 with the deliberate intention of designing a single-trial ERP extraction technique that was not dependent on a priori information. The method and its development are described in detail in [9]. However, it would be beneficial to briefly review the EEG model for the purposes of this thesis.

Martin and Thompson [10] had suggested that the Additive Outlier model described by (2.1.1–2.1.2) would apply to patchy type outliers whose behaviour seemed to be uncorrelated with the behaviour of the rest of the data. ERP’s are transient waveforms that would be like patchy type outliers. Seeing its applicability to the EEG waveform, Birch modelled the observed EEG signal as the sum of an autoregressive (AR) process, \(x_i\) and an additive outlier, \(v_i\).

\[
y_i = x_i + v_i \tag{2.1.1}
\]

\[
x_i = \sum_{k=1}^{p} x_{i-k} \theta_k + u_i \tag{2.1.2}
\]

where

- \(y_i\): observed EEG signal
- \(x_i\): background EEG signal
- \(v_i\): ERP
- \(u_i\): residual error term
\{\theta_1, \ldots, \theta_p\}: \text{AR parameters}

The Additive Outlier model required the background EEG signal to be modelled by an AR process. Parametric models such as the AR model (2.1.2) are often used to describe a set of data. Classical statistical methods typically assume that the data will exactly follow the chosen parametric model. However, outliers in the data set, in this case ERPs, can greatly hinder the results of classical statistical procedures. Robust statistics takes into account the existence of outliers realizing that parametric models are only approximations to reality. Therefore, in order to produce a set of AR parameters that would best model the background EEG signal, it was necessary to use a robust parameter estimator.
2.2 Initial Parameter Estimates

The first step in robust parameter estimation is to generate a set of initial parameter estimates, based on classical statistical methods. These initial parameter estimates are then improved upon through an iterative process of estimation and removal of outlier data.

In the original OPM, MEM (Maximum Entropy Method), a least squares estimator had been selected for the initial parameter estimates. A least squares estimate of $\hat{\theta}$ is that $\hat{\theta}$ which minimizes the sum of the squared errors or residuals (2.2.1). This type of estimator is also referred to as an L-2 estimator.

$$\sum_{i=1}^{n} (y_i - x_i^T \hat{\theta})^2 \quad (2.2.1)$$

When the errors are normally distributed, the least squares estimator is optimal in the class of all unbiased linear estimators. Unfortunately, it is extremely sensitive to additive outliers and the efficiency dramatically decreases when the error distribution is non-Gaussian.

An alternative to least squares are L-1 estimators. An L-1 estimate minimizes the sum of the absolute errors (2.2.2).

$$\sum_{i=1}^{n} |y_i - x_i^T \hat{\theta}| \quad (2.2.2)$$

Birch [8] and Lind [9] had suggested investigating the L-1 estimate, inferring that it might prove to be a better starting point than least squares. L-1 estimators are known to be optimal for linear regressions where the errors have double exponential tails [10]. They also resist undue effects from a few, large outliers but the effects of patchy additive
outlier has not been well-documented. Median regression (MR) is an example of an L-1 estimator. MR is based on a modified simplex algorithm proposed by Barrodale and Roberts [2] with a slight change by Koenker [11]. MEM and MR were compared, in Section 4.1, to determine which one would produce the best initial parameter estimates with patchy type additive outliers.
2.3 The Influence Function and Its Tuning Constants

Another critical step in robust parameter estimation is to minimize the effect of outlier data. One approach involves the application of the influence function. The influence function can be interpreted as a heuristic tool which discriminates against the effect of outliers by weighting the residual error appropriately. The influence function originated with the development of the M-estimate.

M-Estimate

The M-estimate is a generalization of the well-known maximum likelihood estimate. Suppose that \((x_1, x_2, \ldots, x_n)\) are an independently and identically distributed (i.i.d.) set of observations with probability density function \(f(x - \theta)\) for an unknown true location parameter \(\theta\). The maximum likelihood estimate is that value of \(\theta\) that maximizes (2.3.1) or minimizes (2.3.2).

\[
\prod_{i=1}^{n} f(x_i - \theta) \quad \text{(2.3.1)}
\]

\[
\sum_{i=1}^{n} [-\ln f(x_i - \theta)] \quad \text{(2.3.2)}
\]

Huber [12] proposed to generalize (2.3.2) to

\[
\sum_{i=1}^{n} \rho(x_i - \theta) \quad \text{(2.3.3)}
\]
where $\rho$ is a differentiable continuous function. If the derivative of $\rho$ is defined as $\psi(u) = \left(\frac{\partial \rho(u)}{\partial \theta}\right)$ then the estimate $\theta$ satisfies

$$\sum_{i=1}^{n} \psi(x_i - \theta) = 0. \tag{2.3.4}$$

An M-estimate is identified by the choice of $\psi$ or influence function used in (2.3.4). The key to robust M-estimation is to find influence functions that will discriminate against outliers for better estimation results.

Taking it one step further, (2.3.5) is a scale invariant M-estimator where $\sigma$ is a normalization term representing a robust estimate of the standard deviation around $\theta$. $\psi$ is then independent of the scale of the process.

$$\sum_{i=1}^{n} \psi\left(\frac{x_i - \theta}{\sigma}\right) = 0 \tag{2.3.5}$$

When applied to a linear regression model, $y_i = x_i^T \theta + u_i$, the M-estimate is defined by

$$\sum_{i=1}^{n} \rho\left(\frac{y_i - x_i^T \theta}{\sigma}\right) \tag{2.3.6}$$

where

- $y_i$: $i$th observation
- $x_i^T$: $i$th row of design matrix $X \in \mathbb{R}^{n \times p}$
- $\theta$: is a $p \times 1$ vector of unknown parameters
- $u_i$: $i$th residual error with distribution $F(u/\sigma)$
- $\sigma$: robust estimate of the standard deviation of $u$
Chapter 2 — ROBUST PARAMETER ESTIMATION

Taking the partial derivatives with respect to \( \theta \) results in the linear regression M-estimator (2.3.7).

\[
\sum_{i=1}^{n} x_i^T \psi \left( \frac{y_i - x_i^T \theta}{\sigma} \right) = 0
\]  

(2.3.7)

By defining \( x_i^T = \{ y_{i-1}, y_{i-2}, \ldots, y_{i-p} \} \), (2.3.7) becomes the M-estimator for the autoregression model (2.1.2).

In (2.3.7), \( y_i - x_i^T \theta \) is called the residual error term. By scaling the residuals, we hope to reduce their variance to approximately one, but that is largely dependent on their distribution, \( F \). Influence functions are specifically designed to handle data with particular residual error distributions. If the scaled residuals are large, there is a high probability that the data contains outliers. Thereby, the role of the influence function is to set a threshold beyond which all residuals would be down-weighted to counteract the effect of outliers.

An influence function is characterized by its shape and its relative size which is defined by tuning constants. One of the simplest influence functions is the Huber Psi function (2.3.8) designed for an error distribution, \( F \), which is normally distributed in the center with double exponential tails. The tuning constant, \( k \), is the point of discrimination. When the magnitude of the residual is less than or equal to \( k \), it is treated as data with no outlier content and the output equals the input. When the residual magnitude increases above \( k \), the function compensates for outlier content by limiting the output to \( k \).

\[
\psi(u) = \begin{cases} 
-k & , u < -k \\
u & , |u| \leq k \\
k & , u > k 
\end{cases} 
\]

(2.3.8)
In practice, the amount of outlier content is unknown and so it is important for the accuracy of a parameter estimation technique to use an optimum value of the tuning constant. A sub-optimal setting of the tuning constant could lead to the downweighting of valid data or the inclusion of data with outliers, thereby adversely affecting the parameter estimates.

Trial and error methods are typically used for determining the tuning constants of an influence function. However, in the following section, a method of automatically generating the optimal tuning constants is developed for certain influence functions.
2.4 Automatic Generation of Tuning Constants

Asymptotic Theory

In order to understand the premise for calculating the optimal tuning constant for a particular influence function, we must first consider the properties of asymptotic theory \[13\].

Let \((y_1, ..., y_n)\) be a set of independent and identically distributed observations with empirical distribution \(F_n\). Let \(T_n(y_1, ..., y_n)\) or \(T_n(F_n)\) be the estimators of the set of observations, where \(\{T_n; n \geq 1\}\) can be viewed as the sequence of estimators, one for each possible sample size \(n\).

If \(T_n(y_1, ..., y_n) \xrightarrow{n \to \infty} T(F)\), \(T(F)\) is called the asymptotic value of \(\{T_n; n \geq 1\}\) at \(F\), the true distribution. From this, the property of asymptotic normality is derived as

\[
\mathcal{L}_F(\sqrt{n}[T_n - T(F)]) \xrightarrow{n \to \infty} N(0, V(T, F))
\] (2.4.1)

\(\mathcal{L}_F\) is defined as the distribution of its argument under \(F\) and \(V(T, F)\) is the asymptotic variance of \(\{T_n; n \geq 1\}\) at \(F\).

For a parametric model, \(\{T_n; n \geq 1\}\) can be expressed as \(\{\tilde{\theta}_n; n \geq 1\}\) representing a sequence of estimators of \(\theta\), one for each possible sample size \(n\). A sequence of estimators \(\{\tilde{\theta}_n; n \geq 1\}\) is said to be consistent for a parameter \(\theta\) if \(\tilde{\theta}_n \xrightarrow{n \to \infty} \theta\). Thus, (2.4.1) can be written as

\[
\mathcal{L}_{F_\theta}(\sqrt{n}[\tilde{\theta}_n - \theta]) \xrightarrow{n \to \infty} N(0, V(F_\theta))
\] (2.4.2)
The asymptotic efficiency of an estimator is $e \propto \frac{1}{V(\theta)}$. Intuitively, the smaller the variance of the difference between the parameter estimate (based on n observations) and the true parameter, the better the performance of the estimator. Based on this relationship, the minimization of the asymptotic variance maximizes the asymptotic efficiency of the estimator. This, in turn, implies that the optimal value of the tuning constant for a particular influence function is that value which minimizes the asymptotic variance of the estimator.

These concepts can all be expanded to the multi-dimensional case with finite-dimensional vector parameters as described in [13].

**Asymptotic Variance Function**

John Lind [9] presented a method for choosing optimal tuning constants based on minimizing the asymptotic variance of some M-estimators. This method could not be directly applied to the OPM because the OPM uses a slightly different parameter estimation technique. The development of his proposed method will be discussed in this section and the adaptation of this method to the OPM will be discussed in Section 2.5.

The use of M-estimators in calculating robust estimates of regression parameters requires the practitioner to choose an appropriate influence function and specify the tuning constants. The choice of influence function is dependent on the underlying error distribution, F, of the data set. The tuning constants, which are typically chosen by trial and error, could be chosen through the minimization of the asymptotic variance function (AVF).
Huber designed the influence function described by (2.3.8) to specifically accommodate heavy-tailed distributions. Yohai [14] showed that the regression M-estimator with Huber's influence function had the properties of asymptotic normality and therefore, the AVF could be used to generate an optimal value for the tuning constant $k$. If $\tilde{k}$ is the estimate of $k$ and $\tilde{\theta}_k$ is the solution for the M-estimator, Yohai stated that $\sqrt{n}(\tilde{\theta}_k - \theta)$ converges to the normal distribution with mean 0 and multivariate covariance matrix

$$X_0^{-1}V(\psi_k, F)$$

where

$$X_0 = \lim_{n \to \infty} X'X/n$$

and

$$V(\psi_k, F) = \frac{\int_{-k}^{k} u^2 dF(u) + k^2 \left(1 - \int_{-k}^{k} dF(u)\right)}{\left(\int_{-k}^{k} dF(u)\right)^2}.$$

The sample estimate of the asymptotic variance function (2.4.5) is

$$\hat{V}(\psi_k, F) = \frac{\frac{1}{n} \left( \sum_{i=1}^{n} \psi_k^2 \left( \frac{y_i - x_i'\hat{\theta}}{\sigma} \right) \right)}{\left( \frac{1}{n} \left( \sum_{i=1}^{n} I_{[-k,k]} \left( \frac{y_i - x_i'\hat{\theta}}{\sigma} \right) \right) \right)^2}$$

1 where $I_{[-k,k]}$ is the indicator function defined as $I(t) = \begin{cases} 1 & \text{if } -k \leq t \leq k \\ 0 & \text{otherwise} \end{cases}$

Note that Yohai's original equation for (2.4.6) did not include the scaling factor in the argument for the $\psi$ function because he did not scale the residuals in the solution of (2.3.7).
Thus, for \( k \) belonging to a given interval, \( \tilde{k} \) is chosen as that value of \( k \) that minimizes the moment estimate of the asymptotic variance function (2.4.6). He also proved that this procedure had the same asymptotic efficiency for the unknown distribution of error.

Yohai's work was the basis of Hlynka et al.'s [15] derivations used to estimate the tuning constants, \( a_0 \) and \( u_0 \), for

\[
\psi(u) = \begin{cases} 
  u & |u| \leq u_0 \\
  u_0 \epsilon \left( \frac{a_0 - w - u}{a_0 - w - a_0} \right) \text{sgn}(u), & u_0 \leq |u| \leq u_0 - w \\
  0 & |u| \geq a_0 - w
\end{cases}
\]  

(2.4.7)

where \( \epsilon \) and \( w \) were known or predetermined. This influence function was referred to as the optimum solution to the M-estimator (2.3.7) given that the distribution of \( F \) was known to belong to a defined class of distributions. The advantages of their approach was that they chose a redescending influence function (returned to zero) that would totally eliminate gross outliers and allowed for a more general error distribution.

By establishing the asymptotic normality of \( \tilde{\theta}_{u_0,a_0} \), it was found that a reasonable estimate of the tuning constants would be that which minimized the asymptotic variance (2.4.8) for this M-estimator.

\[
V(\psi_{u_0,a_0}, F) = \frac{\int_{a_0-w}^{a_0+w} \psi^2_{u_0,a_0}(u) \partial F(u)}{\left( \int_{a_0-w}^{a_0+w} \psi'_{u_0,a_0}(u) \partial F(u) \right)^2}
\]  

(2.4.8)

The minimization of (2.4.8) becomes virtually impossible when the distribution of \( F \) is unknown, which is the case in most practical situations. Hlynka et al. tackled
this problem of choosing suitable values for the tuning constants of (2.4.7) where F was unknown. They stated that the tuning constants could be chosen by minimizing the moment estimate of the asymptotic variance function (2.4.9).

\[
\tilde{V}(\psi_{u_0, a_0}, F) = \frac{1}{n} \sum_{i=1}^{n} \psi_{u_0, a_0}^2 \left( \frac{y_i - x_i^T \hat{\beta}}{\sigma} \right) \left( \frac{1}{n} \sum_{i=1}^{n} \psi_{u_0, a_0}^t \left( \frac{y_i - x_i^T \hat{\beta}}{\sigma} \right) \right)^2
\]  

(2.4.9)

Due to the nature of the \( \psi \) function, the minimization of even the moment estimate (2.4.9) was still a complex problem. According to John Lind's [9] report, favorable results were obtained with a conjugate gradient minimization algorithm by Powell [16]. However, based on work subsequent to his technical report, he found that the AVF, as a function of two tuning constants was a fairly flat surface. As a result, in the finite sample case, Newton or quasi-Newton methods tended to hang up in small surface depressions. In practice, he suggested the implementation of a simpler redescending influence function such as Tukey’s Biweight (2.4.10) which is a function of only one parameter. He also recommended using a grid search to locate the minimum of the moment estimate of the AVF (2.4.11) as it would only require a one dimensional search.

\[
\psi(u) = \begin{cases} 
  u \left( 1 - \left( \frac{u}{c} \right)^2 \right)^2 & |u| < c \\
  0 & |u| \geq c 
\end{cases}
\]  

(2.4.10)
Adaptation to the OPM

**GM2-estimation**  As stated earlier, the OPM utilizes an estimation technique that is slightly different to the M-estimator. Denby and Martin [17] found that classical M-estimators lacked robustness when applied to the Additive Outlier model. They claimed that the generalized M-estimate (GM-estimate), for which the summands of the estimating equation are bounded and continuous functions of the data, is robust for the Additive Outlier case.

The OPM utilizes a GM2-estimate which is two iterations of GM, each followed by a cleaning. The purpose of the “cleaning” was not to extract the ERP, but to remove large gross outliers that may have been introduced into the signal from a number of sources. It was found that two iterations (GM2) was slightly better than one iteration.
(GM1) and that GM1 was significantly better than GM. A GM2–estimate is obtained through the following steps:

1. Calculate a GM-estimate of the parameters from the observed signal.
2. Remove large outliers in the observed signal with a filter-cleaner as described in Section 3.1.
3. Calculate a GM1–estimate of the parameters using the output of the filter-cleaner.
4. Repeat step 2.
5. Calculate a GM2–estimate of the parameters using the output of the filter-cleaner.

Figure 4: GM2

GM-estimation Since the basis of GM2 is the GM-estimate, the minimization of the asymptotic variance function was applied to the GM-estimate. It can be seen, by comparing the M-estimate (2.4.12) and the GM-estimate (2.4.13) that the essential difference between the two are the weighting factors, $W$. The purpose of the weighting factors is to downweight the summands in (2.4.13) when one or more components in $x_i$ is unreasonably large and produces a poor prediction.

$$\sum_{i=1}^{n} x_i^T \psi \left( \frac{y_i - x_i^T \theta}{\sigma} \right) = 0 \quad (2.4.12)$$

$$\sum_{i=p+1}^{n} x_i \psi \left( \frac{y_i - x_i^T \theta}{\sigma} \right) = 0 \quad (2.4.13)$$
Birch solved the GM-estimates using an IWLS (Iterated Weighted Least Squares) procedure as shown in Figure 5. The parameter estimates were sequentially obtained starting at order one to the desired order. At the beginning of each order, a least squares estimator, MEM produced the initial set of parameter estimates. Within each order, the parameter estimates were updated with every iteration of GM-estimation. Three iterations were performed, the first two with the Huber Psi function and the last iteration with Tukey’s Biweight. $W$ is dependent on the covariance matrix of the system which was updated with each order.
Chapter 2—ROBUST PARAMETER ESTIMATION

Figure 5: A block diagram of the IWLS procedure for GM.

Start

order = 1

MEM estimates

iteration = 1

\( \psi = \text{Tukey's Biweight Function} \)

\( \psi = \text{Huber's Psi Function} \)

update parameter estimates with GM-estimation

3 iterations ?

update covariance matrix

reached desired order?

Stop

increment iteration

increment order

No

Yes

< 2

No

Yes

20
The Asymptotic Variance Function for GM

The GM-estimate can be treated as an M-estimate where $\Phi$ is a function of $x$ and $u$ by letting $\Phi(x, u) = W(x)\psi(u)$. (2.4.13) can be rewritten as

$$\sum_{i=p+1}^{n} x_i \Phi(x_i, u_i). \quad (2.4.14)$$

Marrona and Yohai [18] showed that this type of estimator has the properties of asymptotic normality given certain assumptions. Based on those assumptions, the moment estimate of the asymptotic variance function was derived to be

$$\hat{V}(\Phi, F) = \frac{1}{n} \sum_{i=1}^{n} \phi^2 \left(\frac{x_i - \bar{x}\bar{y}}{\sigma}\right) \left(\frac{1}{n} \sum_{i=1}^{n} \psi^2 \left(\frac{y_i - \bar{x}\bar{y}}{\sigma}\right)\right)^2. \quad (2.4.15)$$

Assuming that $u_i$ is independent of $x_i$, (2.4.15) could be written in terms of $W$ and $\psi$ as

$$\hat{V}(\psi, F) = \frac{1}{n} \sum_{i=1}^{n} W^2(x_i) \psi^2 \left(\frac{y_i - \bar{x}\bar{y}}{\sigma}\right) \left(\frac{1}{n} \sum_{i=1}^{n} W(x_i) \psi^2 \left(\frac{y_i - \bar{x}\bar{y}}{\sigma}\right)\right)^2. \quad (2.4.16)$$

Therefore, the tuning constant that minimizes (2.4.16) would theoretically contribute to producing the best parameter estimates.

Before directly implementing the automatic generation of tuning constants we had to determine the effectiveness of the minimization with the influence functions. Birch used Tukey’s Biweight and the Huber Psi functions in the GM-estimation. For this reason, an initial study [19] was performed where both influence functions were applied to (2.4.16). The results of these simulations provided some insight into the characteristics of their respective AVFs.
The AVFs produced using the Huber Psi function were not very reliable. They were very "noisy" in the vicinity of the minima making it difficult to select the optimal tuning constant. Also, they were ineffective with small sample sizes, in that on many occasions there was no minimum. On the other hand, the AVFs for the Tukey's Biweight were fairly smooth curves where the minima were easily detectable with small sample sizes. The importance of processing the EEG in short data segments, and hence with small sample sizes, is so that the signal demonstrates a semblance of stationarity, as indicated in Birch [8].

Based on these results, it would not be feasible to implement the automatic generation of tuning constants using the Huber Psi function. Birch chose the Huber Psi to ensure that (2.4.13) produced a unique root, as discussed by Martin [20]. Martin stated that (2.4.13) has an essentially unique root with the Huber Psi function. The final iteration with Tukey's Biweight was to make use of the rejection character of the redescending function. Further iterations with Tukey's Biweight, according to Martin, could lead to extraneous roots.

This raised the question: would it be worthwhile to perform all three iterations with Tukey's Biweight to allow for the automatic generation of tuning constants and the complete elimination of trial and error with the trade-off of risking extraneous roots? This issue is addressed in Section 4.1.
2.5 Summary of Issues

The proposed issues with regards to the robust parameter estimation are summarized below and shown in Figure 6.

1. The automatic generation of tuning constants was compared to the trial and error methods to resolve whether the implementation of the OPM could be made easier.

2. MR was compared with MEM to determine which produced the best initial parameter estimates with a patchy additive outlier.
Figure 6: A block diagram of GM-estimation including the automatic generation of tuning constants. The dashed boxes signify the proposed changes to the OPM.
Chapter 3
ROBUST SIGNAL ESTIMATION

The technique used for signal estimation was also required to be "robust" to ensure that the background EEG was estimated without adverse effects due to the ERP. And as in the case of the parameter estimator, the influence function is the chosen approach, described in the next section. The autoregressive parameters generated through GM2-estimation are input, along with the observed signal, to a robust signal estimator which computes a robust estimate of the background EEG signal.
3.1 Filter-Cleaner

The original OPM algorithm used a filter-cleaner proposed by Martin and Thompson [20] for the signal estimation. It is basically a robust version of a Kalman filter as will be shown later.

The filter-cleaner relies on the Additive Outlier model, reviewed below.

\[ y_i = x_i + v_i \]  

(3.1.1)

The background EEG signal is represented by a pth order AR process defined as

\[ x_i = \sum_{k=1}^{p} x_{i-k} \theta_k + u_i \]  

(3.1.2)

where \( \{ \theta_1, \ldots, \theta_p \} \) are the AR parameters and \( u_i \) is a random error term. Written in the state variable form,

\[ \mathbf{X}_i = \mathbf{\Phi} \mathbf{X}_{i-1} + \mathbf{U}_i \]  

(3.1.3)

where

\[ \mathbf{X}_i^T = (x_i, x_{i-1}, \ldots, x_{i-p+1}) \]

\[ \mathbf{U}_i^T = (u_i, 0, \ldots, 0) \]

\[ \mathbf{\Phi} = \begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_{p-1} & \theta_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \]  

(3.1.4)

Robust estimates of \( \mathbf{X}_i \) are calculated according to the following recursion:

\[ \mathbf{\bar{X}}_i = \mathbf{\Phi} \mathbf{\bar{X}}_{i-1} + \mathbf{m}_i \mathbf{s}_i \psi \left( \frac{y_i - \tilde{y}_{i-1}}{s_i} \right) \]  

(3.1.5)
where \( m'_i = m_i / s_i^2 \) with \( m_i \) being the first column of the \( pxp \) matrix \( M_i \). \( M_i \) is recursively calculated as

\[
M_{i+1} = \Phi P_i \Phi^T + Q
\]

\[
P_i = M_i - w \left( \frac{y_i - \tilde{y}_i^{i-1}}{s_i} \right) \left( \frac{m'_i m'_i^T}{s_i^2} \right)
\]

(3.1.6)

The \( M_i \) and \( P_i \) matrixes are prediction and filtering error-covariance matrixes, respectively. \( Q \) is a \( pxp \) matrix with all zero elements except for \( Q_{11} = \sigma_u^2 \). \( s_i \) is a time varying estimate of the scale of the process and is defined by \( s_i^2 = m_{11} \). Birch used \( w(t) = \psi(t)/t \) with Hampel’s three-part redescending \( \psi \) (3.1.7).

\[
\psi(t) = \begin{cases} 
  t, & |t| \leq a \\
  a, & a < |t| \leq b \\
  \frac{a}{b-c}(c-t), & b \leq |t| < c \\
  0, & |t| > c
\end{cases}
\]

(3.1.7)

\( \tilde{y}_i^{i-1} \) denotes a robust one-step prediction of \( y_i \) and is defined by \( \tilde{y}_i^{i-1} = \left( \Phi \tilde{X}_{i-1} \right)_i \). Assuming that \( x_i \) is independent of \( v_i \), as in the AO model, a best predictor of \( y_i \) is also a best predictor of \( x_i \). The robust one-step predictor \( \tilde{x}_i^{i-1} \) of \( x_i \) satisfies \( \tilde{x}_i^{i-1} = \tilde{y}_i^{i-1} \).
The cleaned data at time $i$ is given by

$$\tilde{x}_i = \left(\tilde{X}_i\right)_1,$$

(3.1.8)

the first element of $\tilde{X}_i$.

Note that when $\psi(t) = t$, $w = 1$ and $s_i^2 = m_{13,i} + \sigma_v^2$ (where $\sigma_v^2$ is the variance of $v_i$), eq. (3.1.5) is effectively a Kalman filter. Unfortunately, the Kalman filter is not robust and a single additive outlier would throw off the predictions of $X_i$. 
3.2 Fixed-Lag and Smoother Cleaners

Fixed-Lag Cleaner

Suppose that we take a particular look at the prediction of $x_l$ where $l$ is a fixed point in time. The recursive equation (3.1.5) could be expressed at time $l$ as

\begin{equation}
\begin{bmatrix}
\tilde{X}_l & \\
\tilde{x}_l & \\
\tilde{x}_{l-1} & \\
\vdots & \\
\tilde{x}_{l-p} & \\
\end{bmatrix}
= \Phi
\begin{bmatrix}
\tilde{X}_{l-1} & \\
\tilde{x}_{l-1} & \\
\tilde{x}_{l-2} & \\
\vdots & \\
\tilde{x}_{l-p} & \\
\end{bmatrix}
+ m_l \psi \left( \frac{y_l - y_{l-1}}{s_l} \right)
\end{equation}

(3.2.1)

Now, according to Martin's filter-cleaner, the best prediction of $x_l$ is at time $l$ when $\tilde{x}_l$ is the first element of $\tilde{X}_l$. However, consider the next recursion when time is $l+1$.

\begin{equation}
\begin{bmatrix}
\tilde{X}_{l+1} & \\
\tilde{x}_{l+1} & \\
\tilde{x}_l & \\
\vdots & \\
\tilde{x}_{l-p} & \\
\end{bmatrix}
= \Phi
\begin{bmatrix}
\tilde{X}_l & \\
\tilde{x}_l & \\
\tilde{x}_{l-1} & \\
\vdots & \\
\tilde{x}_{l-p} & \\
\end{bmatrix}
+ m_{l+1} \psi \left( \frac{y_{l+1} - \tilde{y}_{l+1}}{s_{l+1}} \right)
\end{equation}

(3.2.2)

We still have a prediction of $x_l$ stored as the second element in $\tilde{X}_{l+1}$. That prediction is, theoretically, a better prediction because it is not only based on past and present observations ($y_l, y_{l-1}, ...$) but also on a future observation, $y_{l+1}$. $\tilde{x}_l$ will continue to be improved with future observations until it is the last element in the vector, $\tilde{X}_{l+p-1}$. Since
\( \bar{X}_i \) only holds \( p \) elements, it is only possible to improve the prediction of any data value with up to \( p-1 \) future observations. In general, a better prediction of \( x_i \) is

\[
\tilde{x}_i = (\bar{X}_{i+p-1})_p,
\]

the \( p \)th element of \( \bar{X}_{i+p-1} \).

Martin [20] also noticed that there was room for improvement with the filter-cleaner. One of his suggestions was the fixed-lag cleaner, as described above, where the cleaned data is taken after a lag of \( p-1 \).

**Smooother Cleaner**

His second suggestion, a smoother cleaner, was to perform a backward recursion following the forward recursion using all the observations for a prediction of \( x^n_i \), where \( n \) denotes using the total number of samples. The backward recursion is defined as

\[
\tilde{X}_i^n = \tilde{X}_i + P_i \Phi^T M_{i+1}^{-1} (\tilde{X}_{i+1}^n - \tilde{x}_{i+1}^n)
\]

(3.2.4)

where \( i = n-1, n-2, ..., 1 \) with initial condition \( \tilde{X}_n^n = \tilde{X}_n \) and \( \tilde{x}_1^n = \left( \tilde{X}_1^n \right) \).

The background recursion is time consuming and requires the storage of the \( \Phi \), \( M \) and \( P \) matrixes. The smoother-cleaner would be, theoretically, the best choice as it uses all the available information for the prediction of each data value, but it does put demands on memory and processing time. There are also some fundamental problems with the smoother cleaner because the backward recursion depends on the inverse of the \( M \) matrix which is the prediction error-covariance matrix and has the potential of being singular at times.
3.3 Time-Varying Influence Function

The cleaner\(^2\) relies on the influence function to discriminate between outlier and non-outlier content in the data. This in turn implies that the tuning constants of the influence function play a critical role in the accuracy of the results as they directly affect the estimate of the ERP.

In Birch's implementation of the OPM, the tuning constants a, b and c for the Hampel $\psi$ function, Figure 7 were time invariant. A time invariant influence function would apply the same weighting during the entire EEG epoch. The ERP, however, is a transient signal that is only present for a portion of the EEG epoch. As a result, the influence function would be applying undue influence in the periods before and after the ERP and not enough influence during the ERP. Parts of the background EEG signal would be unduly weighted and considered as outlier whereas parts of the ERP would not be appropriately downweighted and treated as background EEG. This would lead to the extraction of noise and the degradation of the ERP estimate.

Noise suppression and ERP enhancement could be achieved with an adaptive influence function. The influence function would be required to be fairly insensitive (a large tuning constant) for the time periods before and after the ERP but would have to clamp down during the ERP. Mason [21] proposed a method to achieve this with an ERP location estimator. His modified OPM had required two passes through the cleaner. On the first pass, the cleaner estimated the location of the ERP using a time-invariant

---

\(^2\) The word "cleaner" would be applicable to any of the three cleaners as they are all just modifications of the basic filter-cleaner.
Chapter 3—ROBUST SIGNAL ESTIMATION

influence function. By keeping track of the statistical characteristics of the prediction error he was able to use a binary implementation of a time-varying influence function on the second pass. Mason achieved good results for varying combinations of simulated data at and above 0dB. The SNR of real EEG signals, however, tend to be less than 0dB, which suggested that Mason’s location estimator may not work as effectively with real data.

Mason had tested his location estimator over a range of SNRs and spectral overlaps between the ERP and the background EEG signals. His simulations identified that the algorithm performed better as the SNR was increased and amount of spectral overlap was decreased.

**SNR and Spectral Overlap**

Spectral overlap is a measure of how similar the ERP is to the background EEG signal in the frequency domain. If the frequency profiles are very similar, a predictive filter like the cleaner would have difficulty differentiating between the two signals. SNR is the ratio of the power in the ERP to the power in the background EEG. In this case, it is not so much how similar they are but how much smaller the power of the ERP is relative to the background signal.

To date, there is very little known about single-trial ERP. Through averaging and other techniques it is known that they exhibit low SNRs. But we have little information regarding their spectral characteristics.

Realizing that the sensitivity of the OPM to both of these factors is primarily a function of the signal estimator, we decided to analyze the effects of spectral overlap and SNR on the parameter estimator. Perhaps, the parameter estimates could provide
some insight into the spectral characteristics of the ERP. Further, if there is very little or no spectral overlap between the ERP and the background sequence, a location estimator could be developed for a time-varying influence function. The results of the simulations will be presented in Section 4.1.
3.4 Summary of Issues

In summary, two issues were explored with regards to the robust signal estimator:

1. The three types of cleaners,
   a. Filter-Cleaner
   b. Fixed-Lag Cleaner
   c. Smoother-Cleaner

   were tested on simulated data to determine if an improvement could be made to the
   filter-cleaner.

2. The effects of SNR and spectral overlap on the parameter estimator and how that
   information could be used to develop a time-varying influence function.
Chapter 4
THE IMPROVED OPM

Simulation studies were performed to test the issues discussed in the previous chapters and analyze their trade-offs. A new and improved OPM will be presented, based on the results of these simulations, and it will be tested on real EEG in the following chapter.

4.1 Data Simulation

The background EEG was simulated by applying a filter, based on a known set of autoregressive parameters, to a random Gaussian error distribution. Recall that an autoregressive sequence is described as

\[ x_i = \sum_{k=1}^{p} x_{i-k} \theta_k + u_i \]  (4.1.1)

where \( u_i \) represents an error term and \( \{\theta_1, ..., \theta_p\} \) are the AR parameters. For the purposes of these simulations the AR parameters were specified as

\[
\begin{align*}
\theta_1 &= 0.838 \quad \theta_2 = -0.471 \quad \theta_3 = 0.638 \quad \theta_4 = -0.429 \\
\theta_5 &= 0.518 \quad \theta_6 = -0.304 \quad \theta_7 = 0.182 \quad \theta_8 = -0.243.
\end{align*}
\]  (4.1.2)

The ERP was also simulated as an autoregressive sequence (4.1.3) which will be referred to as the outlier sequence. The reason that this sequence was chosen was because the spectral distribution of this sequence is somewhat overlapped with the spectral distribution of the background EEG signal. As will be discussed later in this chapter, spectral distribution can greatly affect the results of ERP extraction and for this reason...
a middle ground spectral distribution was chosen.

\[ \theta_1 = 0.6 \quad \theta_2 = -0.2 \quad \theta_3 = -0.2 \quad \theta_4 = 0.4 \]  

(4.1.3)

The length of the background sequence was set to 100 samples. The ratio of the lengths of the two sequences determined the level of contamination. For example, for 10% contamination, the length of the outlier sequence would be set to 10 samples. The outlier sequence was scaled to achieve a specified SNR and then added to the background sequence at a randomly chosen location, resulting in the simulated observation.
4.2 Initial Parameter Estimates

The first test dealt with the capabilities of MEM and MR to estimate the auto-regressive parameters characterizing the background EEG signals. The mean squared error between the parameter estimates and the true parameters (4.1.2) were calculated, where the mean was taken over the 100 trials for varying levels of contamination. Figure 8 depicts the mean squared error of each AR parameter as generated by the two estimation techniques. As expected, MEM outperformed MR in the case of no outlier content. The least squares estimator is optimal for that case, especially since the error distribution in the simulated data was Gaussian. However, the results were not so clear cut with contaminated data. In general, MR appeared to provide better estimates of the first few parameters and MEM appeared to produce the same or slightly better estimates of the last few parameters.
Because these results were not definitive, a second measure of comparison was investigated. Two new sets of background sequences based on the parameter estimates of MEM and MR were generated. The mean squared error was calculated between these signals and the original background sequences, which were based on the true parameters. The estimation technique that exhibited the lowest MSE was denoted as the “best” technique for that particular trial. The percentage of best trials for the different amounts of contamination is plotted in Figure 9. As expected, MEM had a far greater number of best trials than MR for the no contamination case. But MEM’s performance deteriorated and MR’s improved as the amount of contamination was increased. Following this trend
would indicate that in the case of a significant amount of outlier content (over 50%), MR might be the better choice.

Figure 9: Percentage of best trials between MEM and MR

However, Figure 10 indicated otherwise. It is a plot of the actual MSEs for each trial. Although better than MEM in some cases, MR produced extremely large errors for a number of trials. MR did not appear to be a very consistent estimator and even at 50% contamination, it outperformed MEM only 46% of the time. MEM appeared to be more successful in tracking the signal and maintaining a stable level of error. This could be due to the low SNR (0dB) or the patchy type of outlier. In any case, since both situations are expected in real EEG, MEM should be retained for the initial parameter estimates.
Figure 10: The sum of the squared errors for the varying contamination levels.

- 20% contamination
- 50% contamination
- 10% contamination
- No contamination
4.3 Automatic Generation of Tuning Constants

Since GM is the fundamental component of GM2, this test studied the advantages and disadvantages of using the automatic generation of tuning constants compared to using trial and error for GM-estimation.

If the tuning constants in GM are to be automatically customized to the data set, it was necessary to use only Tukey’s Biweight for the GM-estimation. This, however, introduced a risk of producing extraneous roots for the solution of GM.

To analyze the effects of this trade-off, the simulated data at 0dB, was processed by the original GM and the proposed GM\_agtc (automatic generation of tuning constants). The tuning constants selected for GM, determined through trial and error, were

\[ \psi_{\text{Huber}}: k = 1.0 \]
\[ \psi_{\text{Tukey}}: c = 3.0 \]

The mean squared error between the parameter estimates and the true parameters (4.1.2) were calculated, where the mean was taken over 100 trials for varying levels of contamination. The results are shown in Figure 11. For 50% and 0% contamination, both techniques were extremely close. For the cases of 20% and 10% contamination, GM produced slightly better estimates. Although, this was a positive result, it was intriguing that GM\_agtc did not do better than GM, considering that it’s tuning constants were supposedly optimal. This could be due to the replacement of the Huber Psi function with the Tukey’s Biweight for the first two iterations. Also, the AVF was derived based on
the assumption that \( n \to \infty \) and the small sample size of 100 may incur deviations from the true optimal tuning constant.

Figure 11: Comparison of the parameter estimates generated by GM and GM\(_{agtc}\).

Since these results were not definitive, two new sets of background sequences were generated based on the GM and GM\(_{agtc}\) parameter estimates. The mean squared error was calculated between these signals and the original background signals. The estimation
technique that had the lowest MSE was denoted as the "best" technique for that particular trial. Figure 12 shows the percentage of best trials for both techniques for the various levels of contamination. Once again, no one technique dominated. For all cases, the percentages hovered around 50%. These results seemed to indicate that we should expect to achieve at least equal performance from GM_agtc.

Figure 12: Percentage of best trials between GM and GM_agtc.
4.4 Cleaners

The filter-cleaner, fixed-lag cleaner and smoother cleaner were compared to each other by measuring how well each cleaner estimated the background sequences. The inputs to each cleaner were the true known AR parameters (4.1.2) and the simulated observation sequences. Each cleaner then output an estimate of the background sequence. The signal mean squared error between the estimated and the actual background sequences were calculated and used as a measure of performance. The cleaner that produced the lowest mean squared error was denoted as the "best" technique for a particular trial. The smoother cleaner, by far, outperformed the other two as shown in Figure 13. The fixed-lag cleaner ranked second and the filter-cleaner was the worst.

Figure 13: Percentage of Best Trials by each Cleaner
The MSEs between the cleaners, averaged over 100 trials, are listed in Table 1. The average MSEs for the fixed-lag cleaner are approximately 3/4 that of the original filter cleaner and, as predicted, the average MSEs of the smoother cleaner are even better. Although it would seem obvious that the smoother cleaner is the best method, there are a number of issues to consider with regards to this method. First, the smoother cleaner algorithm is susceptible to problems because it requires the calculation of the inverse of the M matrix which can be singular at times. Table 2 lists the differences between the MSE's and it is clear that the improvement with the fixed-lag cleaner over the filter cleaner is an order of magnitude greater than the improvement achieved with the smoother cleaner over the fixed-lag cleaner. In fact, when the smoother cleaner was removed from the comparison, the fixed-lag cleaner picked up every single trial for which the smoother cleaner had been ranked best, as shown in Figure 14.

<table>
<thead>
<tr>
<th>Percent Contamination</th>
<th>Filter-Cleaner</th>
<th>Fixed-Lag Cleaner</th>
<th>Smoother Cleaner</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>1.041</td>
<td>0.745</td>
<td>0.682</td>
</tr>
<tr>
<td>10%</td>
<td>1.074</td>
<td>0.760</td>
<td>0.689</td>
</tr>
<tr>
<td>20%</td>
<td>0.981</td>
<td>0.748</td>
<td>0.699</td>
</tr>
<tr>
<td>50%</td>
<td>1.243</td>
<td>1.12</td>
<td>1.10</td>
</tr>
</tbody>
</table>

Table 1: MSEs averaged over 100 trials
Table 2: Difference in MSEs between cleaners.

<table>
<thead>
<tr>
<th>Percent Contamination</th>
<th>MSE Difference Fixed-Lag and Filter (avg. over 100 trials)</th>
<th>MSE Difference Smoother and Fixed-Lag (avg. over 100 trials)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>0.296</td>
<td>0.063</td>
</tr>
<tr>
<td>10%</td>
<td>0.314</td>
<td>0.070</td>
</tr>
<tr>
<td>20%</td>
<td>0.232</td>
<td>0.049</td>
</tr>
<tr>
<td>50%</td>
<td>0.128</td>
<td>0.016</td>
</tr>
</tbody>
</table>

Figure 14: The percentage of best trials, excluding the smoother-cleaner.

The results of these simulations showed that the original filter-cleaner could definitely be improved upon. Considering that the smoother cleaner was only slightly more accurate than the fixed-lag cleaner, it would be more efficient to use the fixed-lag cleaner and save on time, memory and possible complications due to singular matrixes.
4.5 Time Varying Influence Function

In Section 3.3, it was suggested that some insight could be gained regarding the spectral characteristics of an ERP by analyzing the effect of spectral overlap on the parameter estimates. Outlier sequences based on three different sets of AR parameters were chosen to test varying amounts of spectral overlap. The spectral plots of the outlier sequences are shown in Figure 15 and their corresponding AR parameters are listed in Table 3. The first sequence exhibits the least amount of overlap whereas the third outlier sequence is heavily overlapped with the background signal. Note that outlier no.2 was used in the previous simulation tests. 100 trials were generated with each outlier sequence (1–3) at SNRs of 5dB, 0dB and −5dB.

Figure 15: The spectral plots of the four different outlier sequences corresponding to the AR parameters listed in Table 3. The spectrum of the background sequence is the dotted line in each plot.

Table 3: The AR parameters for the different outlier sequences.

<table>
<thead>
<tr>
<th>Outlier No.</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>$\theta_3$</th>
<th>$\theta_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>0.3</td>
<td>-0.1</td>
<td>-0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>2.</td>
<td>0.6</td>
<td>-0.2</td>
<td>-0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>3.</td>
<td>0.72</td>
<td>-0.31</td>
<td>0.15</td>
<td>-0.05</td>
</tr>
</tbody>
</table>
Chapter 4—THE IMPROVED OPM

The data was then processed by GM2 and MEM. GM2 was chosen as the best representative of a robust estimator and MEM was used as a reference. As seen in Figure 16, GM2's error level remained reasonably consistent across SNRs for each outlier sequence, reaffirming the superior performance of this robust estimator. The difference was in the MEM estimates. The error in the MEM estimates decreased as the SNR decreased and as the spectral overlap increased. In other words, as the outlier became less distinct and more similar to the background sequence, MEM's accuracy in estimating the background AR parameters improved. This quality of the MEM estimator was exploited to assist in revealing the spectral characteristics and the location of the outlier sequence.

The SNR of -5dB was of particular interest because this is the range at which real ERP information is expected. At -5dB the MEM and GM estimates were very similar. Earlier studies by Birch [8] on simulated data showed that when there was no outlier content, MEM was actually a better performer than GM2. Utilizing this information, for the case of SNR -5dB, the difference between the GM2 and MEM estimate would be significant when there was no outlier content and minimal in the vicinity of the outlier. By summing the square of the difference between the MEM and GM2 estimates for each parameter, a single quantity (Cumulative Difference) could be calculated to indicate the location of the outlier content.

\[
CD = \sum_{i=1}^{8} \left( \tilde{\theta}_{iGM2} - \tilde{\theta}_{iMEM} \right)^2 \tag{4.5.1}
\]

A set of 100 trials with each type of outlier was created to test the ability of this
Chapter 4 — THE IMPROVED OPM

Figure 16: The MSEs of GM2 and MEM for the three different types of outliers.
method to locate the outlier content. Each trial was generated with a background sequence of 1000 samples, a randomly placed outlier sequence of 200 samples with an SNR of −5dB. The background sequence length was increased to 1000 to accommodate a moving window of 100 samples to calculate the CD profile. Figure 17 consists of four examples of CD profiles with outlier no.3 where the true location of the outlier is within the dashed lines.

The results from the trials with outlier no.1 and no.2 showed that the outlier regions could not be clearly identified as they were in Figure 17. The reason for this might be that outlier no. 3 was the most heavily overlapped with the background EEG with respect to its spectral distribution and the MEM estimate would not be greatly affected by the outlier content. It would be expected that the MEM estimate would be consistent across the entire signal. However, the GM2 estimate would be a little off outside the outlier vicinity and improve in the outlier region, which is exactly why the difference is much larger outside the outlier in Figure 17. There was a slight lag in the CD profiles because a moving window of 100 samples in length was used.

Since, outliers no.1 and 2 begin to have spectral differentiation with the background EEG, the outlier content would begin to affect the MEM estimate and its performance would deteriorate with respect to GM2. Consequently, the two estimates would differ both outside and within the outlier region.
Figure 17: Examples of CD profiles generated from observed signals with -5dB SNR and outlier no. 3.
4.6 Summary

The results of these simulation studies are summarized below with the corresponding modifications to the OPM.

1. The automatic generation of tuning constants successfully equalled the performance of the original GM. By eliminating the trial and error methods, GM\_agtc is an easier tool to use. There is also an additional security in knowing that if the data, for some reason, required dramatically different tuning constants, this technique would adapt to the data. This method was incorporated into the new OPM.

2. It was recommended that MEM remain as the initial parameter estimator for GM. MR appeared to be an inconsistent estimator.

3. The smoother cleaner was the most accurate cleaner but the fixed-lag cleaner proved to be the most efficient cleaner. The fixed-lag cleaner was only slightly less accurate than the smoother cleaner and didn’t require as much memory or time. The fixed-lag cleaner replaced the filter-cleaner in the new OPM.

4. The cumulative sum of the difference between the GM2 and MEM parameter estimates would be effective at locating outlier content at low SNRs (−5dB) if the spectral characteristics of the outlier signal and the background EEG were heavily overlapped. Since this technique is specific to a particular type of signal, it was not implemented in the new OPM.

The new OPM will be referred to as OPM2. This new method was tested on real EEG signals and compared to two other algorithms in the next section.
5.1 Analysis of EEG signals

Ensemble Averages

Real EEG signals had been previously collected for a conference paper by Mason et. al [22]. A detailed description of the experiment and data collection can be found in that paper. 100 trials of active and 100 trials of idle EEG signals from a single subject were selected for this study. The active trials consisted of 2 seconds prior to, and 1.5 seconds after a ballistic flexion of the right index finger. The idle trials were of the same total duration (3.5 seconds) during which the subject had been asked to relax but remain alert.

The data had been sampled at 256Hz. There is a direct correlation between the sampling frequency and the AR model order. A higher sampling rate requires the use of a higher model order since the time dependency is spread over a greater number of sample points. It would be desirable to use the lowest model order possible without sacrificing information contained in the EEG signal. Since, the majority of the power in an EEG signal lies below 30Hz [8], that would be the highest frequency of interest. The EEG signals were lowpass filtered (cutoff 30Hz) and decimated by a factor of 4, reducing the sampling frequency to 64Hz.

Fifteen signals had been recorded for each trial. The first two signals were indicators of EOG\(^3\) activity and finger position. The remaining thirteen signals were EEG recorded from scalp electrodes. C\(_3\), C\(_4\), F\(_z\), C\(_z\), P\(_z\) and O\(_z\) are standard International 10/20 system

\(^3\) EOG is a measure of the variations in the corneal-retinal potential as affected by the position and movement of the eye.
electrode sites [22]. The additional electrodes had been added at equally spaced intervals between these electrodes. All of these signals had been referenced to the O\textsubscript{2} electrode.

It is generally believed [23], [24] that the supplementary motor area (SMA) and motor cortex are parts of the brain that play significant roles in the execution of voluntary movements. The SMA is located in the frontal lobe of the brain in the vicinity of the F\textsubscript{z} electrode. The contralateral region responsible for right index finger movement corresponds to C\textsubscript{3}, which is located on the motor cortex in the left hemisphere of the brain. Ensemble averages (sample averages over 100 trials) of both these electrode sites are depicted in Figure 18. Assuming that the background EEG had a zero mean, the process of averaging was used as an aid to reveal the underlying ERP that may be present but would not be apparent in the single-trial. The ensemble averages of the active trials revealed distinct waves of electrical activity coinciding with the finger movement. Note that the finger movement is marked at 2.0 seconds. In sharp contrast, the ensemble averages of the idle trials were random low amplitude signals with means close to zero.

The third signal of interest was F\textsubscript{z} - C\textsubscript{3}. The ensemble average of the difference signal was a very characteristic signal with a sharp transition from a negative peak to a positive peak.\textsuperscript{4} It would be easier for a method like OPM, which has demonstrated better performance in extracting sharp transitions, to pick out such a distinct signal. Once again, the ensemble average of the idle F\textsubscript{z} - C\textsubscript{3} signals, Figure 18f, was a random low amplitude signal.

\textsuperscript{4} All EEG signals will be portrayed with negative voltages above the zero line as is the convention.
Chapter 5— TESTS ON REAL EEG

Figure 18: Ensemble averages of the active and idle trials at electrodes Fz and C3.

a) Fz - active

b) Fz - idle

c) C3 - active

d) C3 - idle

e) Fz-C3 - active

f) Fz-C3 - idle
SNR

The calculation of SNR was not a straight-forward process. The SNR of any one active trial could not be calculated without attempting to extract the single-trial ERP. The extraction algorithm would then introduce an entire new set of factors that would have to be identified and accounted for. A simpler approximation of SNR is presented below.

The noise power was defined as the average power in the idle trials. This translated to calculating the power in each idle trial and then taking the average of the powers over the 100 trials.

\[
P_n = \frac{1}{100} \sum_{j=1}^{100} \left[ \frac{1}{T} \sum_{i=1}^{T} n_i^2 \right] 
\]  

The signal power was obtained from the ensemble average of the active trials' which was an approximation of a single-trial ERP. Two variations of signal power were calculated. The first was the average power in the ERP, where the ERP was defined by a window from a to b.

\[
P_{s-avg} = \frac{1}{(b-a) \cdot f_s + 1} \sum_{i=a \cdot f_s}^{b \cdot f_s} s_i^2, \quad f_s = 64 \text{Hz} 
\]  

The second was a calculation of the maximum signal power in the ERP. This was simply defined as the square of the maximum amplitude.

\[
P_{s-max} = \max(s_i^2), \quad i = (a, ..., b) \cdot f_s 
\]  

The average and maximum SNRs were calculated using the corresponding signal power with the noise power as described by (5.1.1). Because the signal powers were based on ensemble average ERPs, the SNRs listed in Table 4 are slightly under-estimated.
Single-trial ERPs are known to shift in time from trial to trial. By averaging across the 100 active trials, peaks that might exist in the time-shifted single-trial ERPs would tend to be smoothed over in the ensemble average. Consequently, these SNRs are conservative approximations and the actual SNRs should be slightly higher.

Table 4: Average and maximum SNRs at Fz, C3 and Fz-C3

<table>
<thead>
<tr>
<th>SNR</th>
<th>Fz</th>
<th>C3</th>
<th>Fz-C3</th>
</tr>
</thead>
<tbody>
<tr>
<td>average SNR</td>
<td>-8.9dB</td>
<td>-9.0dB</td>
<td>-9.3dB</td>
</tr>
<tr>
<td>maximum SNR</td>
<td>-5.3dB</td>
<td>-4.0dB</td>
<td>-4.0dB</td>
</tr>
<tr>
<td>a</td>
<td>1.85sec</td>
<td>1.0sec</td>
<td>1.0sec</td>
</tr>
<tr>
<td>b</td>
<td>2.15sec</td>
<td>3.5sec</td>
<td>3.5sec</td>
</tr>
</tbody>
</table>

**Ranking Trials**

Since not all of the trials were equally "good" trials, they were ranked according to the quality of the EOG and finger movement. An automatic system of rejecting trials with EOG artifact had been used during the data collection. The rejection criterion was defined by a pre-determined threshold. However, there were still trials with poor quality EOG that slipped through the system. Ideally, a good EOG signal would have a constant mean across the EEG epoch and a low variance. In other words, the flatter the EOG the better. The quality of the finger movement was judged by a signal that recorded the position of the finger relative to a 'white light' sensor. A weak finger movement may result in a poor ERP. Therefore, it was important to use only those trials with strong finger movements characterized by a sharp distinct change in the position of the finger as it was flexed. The active trials were initially given two ranks, EOG and finger movement.
A combination of the EOG ranking and finger movement ranking was used to provide an overall ranking for each active trial. The idle trials did not contain any finger movement and were therefore ranked by EOG exclusively.
5.2 Test Procedure

The improvements to the original OPM, as discussed in Chapter 4, were implemented and tested on the real EEG signals. F2–C3 was the signal chosen for processing because it displayed the most distinct features. It was found that 50 of the 100 trials would not be suitable for processing due to either poor EOG and/or finger movement. Therefore, only the top 50 ranked trials were tested. The procedure was comprised of three extraction techniques; 1) the original OPM (OPM1), 2) the improved OPM (OPM2) and 3) Minimum Mean Square Error (MMSE) filtering. MMSE was included to provide a comparison with a well-established and well-known extraction method. MMSE will be briefly overviewed in this section. The performance of each technique could not be evaluated with regards to how well it extracted the ERP because there was no true single-trial ERP to serve as a reference. The ensemble average would not be a good choice because it lacks unique single-trial information. The performance of each technique was measured by how well it could differentiate between idle and active trials. This was measured by a classifier which output the decision of whether the trial had contained an ERP or not. The classifier that was specifically used for these studies was Dynamic Time Warping.
MMSE

MMSE filters have been used for various types of signal extraction. In particular, MMSE was used by McGillem et al. [6] to extract visually evoked potentials, a type of ERP. These filters are based on the principles of minimizing the mean squared error. The EEG signal is represented by

\[ y_i = x_i + v_i \]  \hspace{1cm} (5.2.1)

where

- \( y_i \): observed signal
- \( x_i \): background EEG
- \( v_i \): ERP
An estimate of \( v_i \) is obtained by applying a filter to \( y_i \), where the filter coefficients are chosen such that they minimize the sum of the squared errors between \( \tilde{v}_i \) and \( v_i \). Thus, the filter coefficients are chosen to minimize

\[
\sum_i (\tilde{v}_i - v_i).
\]  

(5.2.2)

The filter coefficients are obtained from the following equation.

\[
h = R^{-1}g
\]  

(5.2.3)

where

\( R \): sample correlation matrix of \( y \)

\( g \): sample cross-correlation vector between \( y \) and \( v \)

\( h \): filter.

As the ERP is an unknown entity, the filter coefficients are based on using the ensemble average for \( v \). The ERP estimate, \( \tilde{v} \), by applying the filter to the observed signals, \( y \).

**Performance Measures**

A critical problem encountered in this type of research is the inefficiency of available performance measures for the evaluation of ERP extraction methods. This problem is not limited to the scope of ERP signals alone and is applicable to other methods that extract transient signals.

Performance measures are needed to quantify the results obtained from extraction methods because it is not possible to visually inspect the large quantities of data processed.
during our studies. In addition, visual inspection is not useful to quantify the performance of any given method. A performance measure for a signal extraction technique is a quantity describing how closely the extracted signal matches the true signal.

**Dynamic Time Warping**

Due to the variable latencies of ERP components, Dynamic Time Warping as described by Roberts [25], was chosen as the performance measure for evaluating the performance of each technique. Given two waveforms, it finds the optimal time adjustment through shifting, stretching and contracting the time axis to minimize a specified cost. In this case, DTW produced a cost of warping an ERP estimate along the time axis to best match the standard wave template.

The cost of warping is defined by Roberts et al. [2] in terms of a correlation, \( Q \), and a penalty function, \( P \), as

\[
C = \frac{1}{Q(a, b, w) - \alpha P(w)}. \tag{5.2.4}
\]

The higher the cost, the worse the match. A randomly selected subset of 25 active ERP estimates were chosen as the active test set and the remaining 25 active trials were used to generate a standard wave template for DTW. To compare the same number of active and idle trials, an idle test set of 25 ERP estimates was also randomly selected. Both test sets were processed by DTW and a cost for warping each trial to the standard wave was generated.
The costs for the active test set were compared to the costs for the idle test set to determine the best threshold for discrimination. Each technique was evaluated on how well it discriminated between the active and idle trials.
5.3 Results

Ensemble Averages and Single-Trials

The ensemble averages of the top 50 active and idle ERP estimates were computed and plotted for each technique, Figures 21 and 22. The averages of the active ERP estimates indicate that the ERP related information was preserved by all three extraction methods. The finger movement is characterized by a negative peak at 2.0 sec. and a positive peak between 2.1 and 2.2 sec., which is consistent with the ensemble average of the original data, Figure 18e). In the case of MMSE, the waveform shape is very similar to Figure 18e), which is not surprising, since the MMSE filter that was applied to each signal was derived from Figure 18e). The ensemble averages of the ERP estimates resulting from OPM1 and OPM2 distinctly show the two peaks characterizing the ERP and exhibit the same negative and positive biases, on a smaller scale, preceding and following the two peaks as in Figure 18e). It is important to highlight at this point that the OPM algorithms were successful in preserving the ERP related information without any a priori information about the shape of the ERP waveform.

As expected, the idle ERP estimates in Figure 22 do not exhibit any characteristic peaks. The ensemble average of the raw idle data, Figure 18f), represents an ensemble average of the background EEG signal which appears to have a positive bias. This positive bias is also seen in the ensemble average of the idle signals processed by MMSE in Figure 22. However, the idle signals processed by OPM1 and OPM2 are reasonably flat signals with zero DC bias. This is probably due to the fact that the OPM1 and OPM2
techniques attempt to remove the background EEG signal which should ideally produce a completely flat signal.

Figure 21: Ensemble averages of the active ERP estimates
Figures 23 and 24 show the processed-ERP estimates for four different active trials. For the majority of the trials, OPM1, OPM2 and MMSE successfully extracted the peak at 2.0 sec. and the second peak between 2.1 and 2.2 sec. Given the extremely low SNR conditions, the successful extraction of the ERP-related information on a single-trial basis without a priori information, for the OPM1 and OPM2 algorithms, was a very encouraging result. This reaffirmed Birch's results [8] which used the OPM1 method.
Chapter 5 — TESTS ON REAL EEG

The results of processing four different idle signals through the three different extraction techniques are shown in Figures 25 and 26. As they should be, the OPM1 and OPM2 single trial idle signals were fairly flat and centered around zero volts, as there should not be any ERP related information or background EEG in these signals. The MMSE idle estimates were not so clean because the MMSE filter is based on the ensemble average of the active trials and does not attempt to remove the background EEG signal.
Figure 23: Examples of single trials of active ERP estimates

Trial 81

Trial 85
Chapter 5 — TESTS ON REAL EEG

Figure 25: Examples of single trials of idle ERP estimates

Trial 62

Trial 13
Figure 26: Examples of single trials of silent ERP estimates

Trial 58

Trial 91

Chapter 5 — TESTS ON REAL EEG
Performance Measurements

As described earlier in this section, Dynamic Time Warping was used to measure the performance of each extraction technique. The effectiveness of each technique was measured by determining how well each technique was able to detect an active signal and reject an idle signal. This was accomplished by determining a cost threshold, based on DTW, to differentiate between active and idle trials. All trials with costs above the threshold would be rejected whereas all trials with costs below the threshold would be accepted as active trials. The optimum cost threshold was determined for each extraction method by determining the cost that would maximize the number of active trials that were detected and minimize the number of idle trials that were mistaken for active trials, false positives.

A standard template of an active ERP estimate was generated for each technique using the method described by Roberts [25]. These standard templates are shown in Figures 27, 28 and 29.
Chapter 5 — TESTS ON REAL EEG

Figure 27: Standard template used for DTW for OPM1 estimates

Figure 28: Standard template used for DTW for OPM2 estimates
A separate template was generated for each extraction technique to ensure that the optimum template was used to determine the cost of warping the ERP estimates. Each active and idle processed signal was warped to its corresponding template and a cost was calculated for each signal. The idle and active costs were then analyzed for each extraction technique and a cost threshold was determined which would maximize the number of active trials that were detected and minimize the number of idle trials that were mistaken for signals containing ERP information. Table 5 lists the results of the DTW cost comparisons.
Table 5: Results of DTW cost comparisons

<table>
<thead>
<tr>
<th>Extraction method</th>
<th>Active trials detected</th>
<th>False Positives detected</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPM1</td>
<td>15/25</td>
<td>2/25</td>
</tr>
<tr>
<td>OPM2</td>
<td>13/25</td>
<td>0/25</td>
</tr>
<tr>
<td>MMSE</td>
<td>16/25</td>
<td>10/25</td>
</tr>
</tbody>
</table>

The ideal extraction method would have detected 25/25 of the active trials and detected 0/25 false positives, where a false positive is an idle trial that is mistaken for an active trial. However, since all of these methods fall short of ideal, it is important to note that in the real world situation of dealing with severely disabled individuals, a false positive is a far more serious error than rejecting an active trial. OPM1 and OPM2 proved to be better methods for rejecting idle data.
6.1 Contributions

Three different ERP extraction methods, MMSE, OPM1 and OPM2, were implemented and tested on simulated and real EEG signals. MMSE is the conventional approach to the problem and served as a benchmark for these tests. OPM1 was the first method that successfully extracted ERP related information from single trial EEGs without any a priori information. OPM2 was a method developed to improve upon the original OPM1.

Four components of the OPM1 algorithm were studied and tested with simulations to develop the OPM2 method.

MEM was retained as the initial parameter estimator for the robust parameter estimation in OPM2. Median Regression, which was an alternative method for parameter estimation, was shown to be an inconsistent estimator.

The automatic generation of tuning constants, used in OPM2, successfully equalled the performance of the original OPM1, which determined the tuning constants using trial and error methods. By eliminating the trial and error, OPM2 was an easier tool to use.

The smoother cleaner was the most accurate cleaner but the fixed-lag cleaner proved to be the most efficient cleaner. The fixed-lag cleaner was only slightly less accurate than the smoother cleaner and did not require as much memory or time. In the OPM2, the filter-cleaner was replaced by the fixed-lag cleaner.
The cumulative sum of the difference between the GM2 and MEM parameter estimates was shown to be effective at locating outlier content at low SNRs (−5dB) if the spectral characteristics of the outlier signal and the background EEG were heavily overlapped. Since this technique was specific to a particular type of outlier, it was not implemented in OPM2.

The results of the tests on real EEG showed that all three algorithms detected approximately the same number of active ERP trials. However, MMSE detected the largest number of false positives (10/25), OPM1 detected only 2/25 false positives, OPM2 was the best performer with 0/25 false positives.

This study showed that the OPM1 and OPM2 algorithms successfully extracted ERP related information on a single-trial basis from real EEG signals without any a priori information.
6.2 Future Work

The most difficult part of this research was acquiring valid usable data. 50% of the original collected data had to be discarded due to EOG artifact and/or poor finger movement. This statistic could be improved by collecting data based on a more pronounced movement and by using an automated and more sophisticated system for EOG artifact detection.

The tests on real EEG were performed on signals containing information related to a ballistic movement of the index finger. The SNR of these signals was quite low, approximately $-9\text{dB}$ on average and only $-5\text{dB}$ at the maximum. The low SNR could be attributed to the reflex nature of the finger movement. A more controlled and determined movement would probably result in a higher SNR and could significantly improve the performance of these extraction techniques.

Although these studies were limited to a single EEG signal, FZ-C3, each electrode location has a unique characteristic waveshape when an event occurs. Several theories have suggested that the ERP related information travels across the scalp and that spatial information could provide key characteristics of the ERP for better location detection.

The OPM algorithm is based on modelling the background EEG signal. The current model of a stationary autoregressive signal could be improved by using adaptive modelling techniques that would take into account the non-stationary characteristics of the background signal.

Dynamic time warping was used as the performance measure for these tests. This
performance measure was very susceptible to large peaks because it is based on a direct
correlation of the two waveforms being compared. Also, the penalty function is chosen
by trial and error. An automated method of choosing the optimal penalty function for a
set of data and an automated normalization incorporated into the algorithm would greatly
improve the efficiency of this method.
References


381, 1990.


