SINGLE TRIAL EEG SIGNAL ANALYSIS USING OUTLIER INFORMATION

By

GARY EDWARD BIRCH

B.A.Sc., The University of British Columbia, 1983

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

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

June, 1988

In presenting this thesis in partial fulfilment 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
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date Sept. 22, 1988

ABSTRACT

The goal of this thesis work was to study the characteristics of the EEG signal and then, based on the insights gained from these studies, pursue an initial investigation into a processing method that would extract useful event related information from single trial EEG. The fundamental tool used to study the EEG signal characteristics was autoregressive modeling. Early investigations pointed to the need to employ robust techniques in both model parameter estimation and signal estimation applications. Pursuing robust techniques ultimately led to the development of a single trial processing method which was based on a simple neurological model that assumed an additive outlier nature of event related potentials to the ongoing EEG process. When event related potentials, such as motor related potentials, are generated by a unique additional process they are "added" into the ongoing process and hence, will appear as additive outlier content when considered from the point of view of the ongoing process. By modeling the EEG with AR models with robustly estimated (GM-estimates) parameters and by using those models in a robust signal estimator, a "cleaned" EEG signal is obtained. The outlier content, data that is extracted from the EEG during cleaning, is then processed to yield event related information.

The EEG from four subjects formed the basis of the initial investigation into the viability of this single trial processing scheme. The EEG was collected under two conditions: an active task in which subjects performed a

skilled thumb movement and an idle task in which subjects remained alert but did not carry out any motor activity. The outlier content was processed which provided single trial outlier waveforms. In the active case these waveforms possessed consistent features which were found to be related to events in the individual thumb movements. In the idle case the waveforms did not contain consistent features. Bayesian classification of active trials versus idle trials was carried out using a cost statistic resulting from the application of dynamic time warping to the outlier waveforms. Across the four subjects, when the decision boundary was set with the cost of misclassification equal, 93% of the active trials were classified correctly and 18% of the idle trials were incorrectly classified as active. When the cost of misclassifying an idle trial was set to be five times greater, 80% of the active trials were classified correctly and only 1.7% of the idle trials were incorrectly classified as active.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank everyone in the Psychophysiology Laboratory at UBC who have helped to facilitate this work. In particular, I would like to thank Adelle Forth and Timothy Harpur for their invaluable assistance in gathering the EEG data.

I would like to express my deep gratitude to Dr. J. C. Lind, who introduced me to robust statistics and who reminded me that in some cases outliers are like diamonds in a sea of semi-precious stones.

I am grateful to my committee members, Drs. M. P. Beddoes, G. Dumont, and R. D. Hare, for their support and expert advice. I would like to give special additional thanks to Dr. Hare for making his laboratory facilities available to me and for his strong support of my work since the beginning.

I extend my deepest appreciation to my supervisor Dr. P.D. Lawrence for his guidance, encouragement and dedicated support throughout my graduate career.

To my family, I am thankful for their encouragement throughout my education. I would like to especially thank Brenda for her patient and nurturing support.

CHAPTER 1

INTRODUCTION

1.1 Problem Statement

Most often the greatest failing of technical aids for severely disabled persons is the inadequacy of the man/machine interface. With a universal effective and efficient interface, current technology has the capability of providing substantial independence and hence, a greatly improved quality of life for even the most severely disabled persons. In pursuit of such an ideal interface, researchers have been studying the feasibility of utilizing electrical brain potentials to directly communicate to peripheral devices. Two important example applications would be the supervisory control of a robotic arm and the method of input to a personal computer system. Such an interfacing capability would also prove to be very useful in many man/machine interface problems in the able-bodied population. The ultimate goal of this researcher is to provide a direct communication system between man and external devices using electrical brain activity.

Brain potentials are comprised of continuous random electrical activity which when recorded are referred to as electroencephalographic (EEG) signals. Embedded within the EEG are event related potentials (ERP) and typically, there is a -6 db to a -9 db signal-to-noise ratio between the ERP and the EEG. This activity is usually recorded using surface electrodes on the scalp, since the current risk versus benefit situation does not justify electrode implantation. The EEG signals as measured from the scalp surface are in the order of 5-50 micro-volts and are easily contaminated by other

bio-electrical signals such as electroocular (EOG) potentials and muscle (EMG) potentials, especially those of the scalp and face. To date, the study of ERP's has mostly been confined to the averaging of EEG recorded during a specific event, such as a flash of light, over many trials. With averaging, the random signal not related to the event tends to average out to zero at a rate generally proportional to the square root of the number of trials averaged. The signal related to the event is assumed to remain constant and hence becomes more recognizable as the background random signal decreases. This approach has many drawbacks. In the context of this work, two of the greatest drawbacks are firstly, that it is not amenable to real-time processing of event related potentials for use in closed loop control applications and, secondly, there is a significant loss of unique single trial information. However, it is a useful tool for obtaining a general idea of the underlying waveform of an event related response.

One of the most significant obstacles that must be overcome in pursuing the ultimate goal is the establishment of a signal processing method that can extract event related information from single trial EEG. There have been some single trial processing schemes, proposed by various researchers (see for example [1,2,3]), that were designed to detect features in the EEG that are related to specific external events. The usefulness of their results has been generally limited because their schemes have often been partly dependent on fundamental assumptions about the statistical characteristics of the EEG which at the present time are not well understood and, more importantly, have been critically dependent on complex external visual stimuli.

The goal of this thesis work was to study the characteristics of the

EEG signal and then, based on the insights gained from these studies, pursue an initial investigation into a processing method that would extract useful event related information from single trial EEG. The fundamental tool used to study the EEG signal characteristics was autoregressive modeling. Early investigations pointed to the need to employ robust techniques in both model parameter estimation and signal estimation applications. Pursuing robust techniques ultimately lead to the development of a single trial processing method which was based on a simple neurological model that assumed an additive outlier nature of event related potentials to the ongoing EEG process. The EEG from four subjects formed the basis of the initial investigation into the viability of this single trial processing scheme. The EEG was collected under two conditions: an active task in which subjects performed a skilled thumb movement and an idle task in which subjects remained alert but did not carry out any motor activity.

1.2 Motor Potentials

Brain signals that are related to movement are a type of event related potential and were first reported by Kornhuber and Deecke [4]. These potentials are produced by the sensory-motor cortex prior to and during voluntary movements of the body. Figure 1.1 [5] shows a cross-section of the motor cortex which illustrates how various parts of the body are functionally mapped onto the motor cortex.

The use of ERP's related to movement have many advantages over other ERP's in the context of an interface system. Motor potentials are produced as a result of a self-initiated cognitive process. This is unlike the

Primary Motor and Somatosensory Areas of the Cerebral Cortex

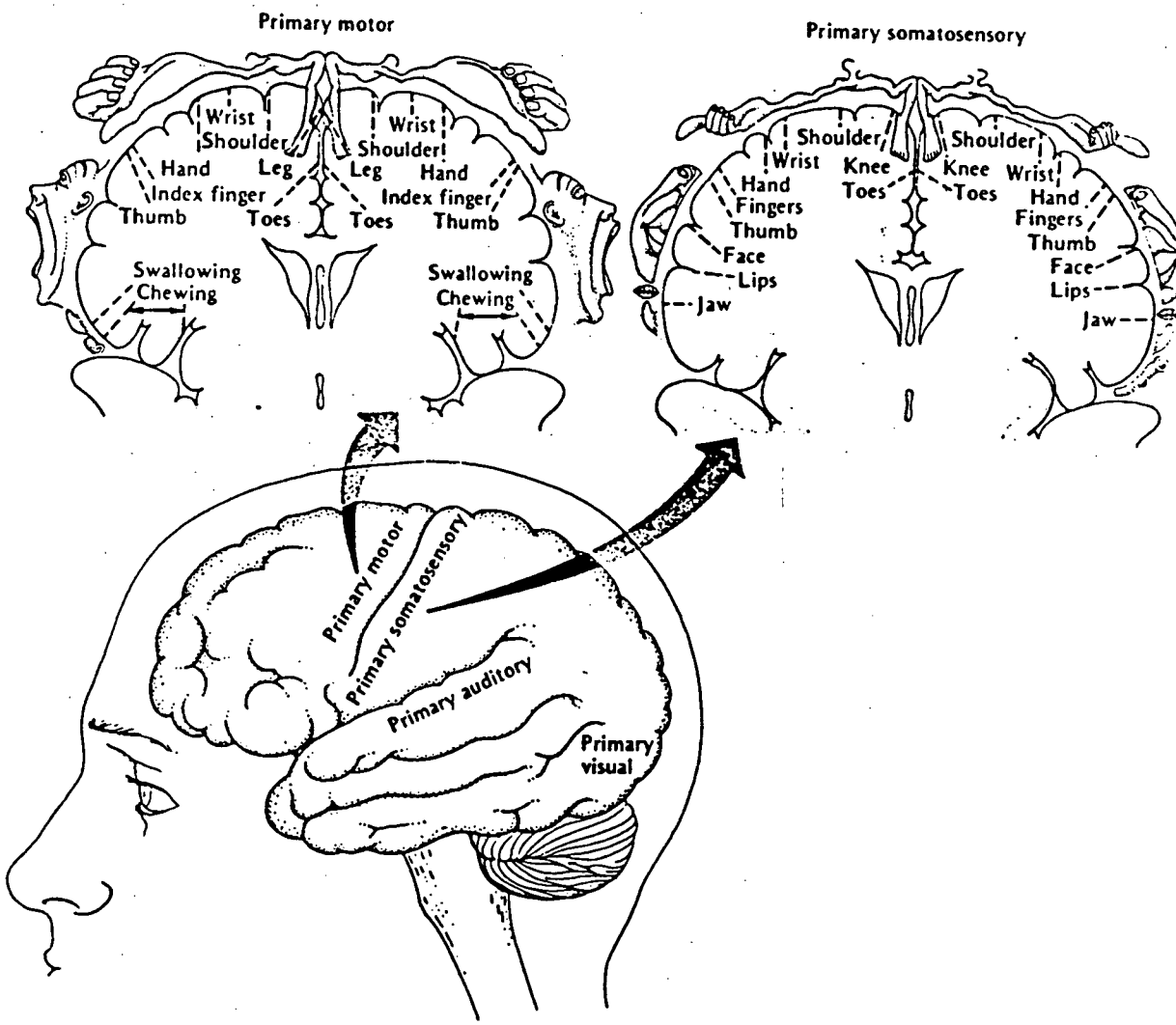


Figure 1.1

perhaps better known subset of ERP's, called evoked potentials, where an external stimulus is required to elicit a response. (for example: visual, somatosensory or auditory evoked potentials.) Motor potentials that are associated with parts of the body over which a disabled person does not have cognitive control, could be utilized to obtain unique control of a given peripheral device and thereby minimizing the likelihood of unintentional activation of that device. As suggested by Figure 1.1, there is in fact some capability to discern what part of the body is being moved by utilizing spatial information from across the motor cortex with the use of surface electrodes. Some of the recent work in this area has been carried out by Brunia and Van den Bosch [6] in which they exploit the ipsilateral and contralateral properties of the motor potentials to demonstrate an ability to discriminate between hand and foot movements as well as right versus left body movements. The extent to which this discrimination capability can be refined needs to be pursed further but at the very least it can provide some diversity in the control functions that could be derived from the motor potentials. It has also been demonstrated [7] that, in the averaged motor potentials, there is a sustained response (in the order of one second) throughout a prolonged task requiring substantial cognitive involvement. This property could be particularly exploited in an interface system that requires a continuous control function such as in the task of guiding a robotic arm.

A specific issue that will need to be addressed in future work, in terms of an interface systems for the disabled, is the role of peripheral afferent feedback in the generation of motor responses. Due to the early work of Vaughan [8] it was thought that the positive components after onset

of the movement were due to afferent feedback. However, Vaughan et al. [9] four years later unexpectedly found a similar positivity in deafferented monkeys while they where carrying out a self paced task. Papakostopoulos [10] postulated that as long as there are elements of skill required in the task, the positivity will be developed despite the absence of afferent feedback. The uncertainty surrounding this issue remains in relatively recent work (see Grunewald et al. [7]) and to resolve it completely will require the ability to analyze the motor potentials on a single trial basis because a disabled person can not provide a movement trigger for conventional averaging techniques. The single trial analysis method described in this thesis provides this required ability and therefore, will ultimately facilitate the study of motor potentials from persons who lack peripheral afferent feedback.

CHAPTER 2

MODELING THE EEG SIGNAL

2.1 Need for Statistical Understanding of EEG

A prerequisite to the mathematical modeling of a given signal is an adequate understanding of its fundamental characteristics. Previous investigations (see Persson [11] or McEwen and Anderson [12]) have noted that the random character of EEG makes the theory of random processes applicable to EEG signals. Therefore, if the approach to the signal analysis of EEG is to be based on random process signal theory then basic statistical characteristics of the EEG signal should be well understood.

Statistical properties, particularly assumptions about Gaussianity, are often key factors in the resulting performance of many of the signal processing methods that have been conventionally applied to EEG. For instance, applications of Wiener filtering, Kalman filtering and AR parameter estimation to EEG have met with mixed and inconsistent results (see McGillem et al. [13]). This could very well be explained if in fact the signal characteristics were at different times ranging approximations to the required Gaussian assumptions. On occasions when the signal processing was carried out while the EEG was relatively close to Gaussian, the result would have been close to optimal and relatively good performance would be expected. However, as will be demonstrated in Section 3.4, if estimation had been carried out on EEG that was relatively non-Gaussian the performance would have likely been very poor.

Hence, to utilize signal processing methods which make certain statis-

tical assumptions, requires both a statistical knowledge about the target signal as well as knowledge about the ramifications of using a given method when those assumptions are not to some extent met. In the type of statistical modeling employed in this study, details of which are provided in subsequent sections, a satisfactory understanding of these issues is very important. The following section begins this process of understanding by reviewing previous investigations into the statistical character of EEG.

2.2 Previous Stochastic Studies on EEG

There have been relatively few investigations into the statistical characteristics of spontaneous EEG activity. The resulting conclusions from these few investigations have been largely contradictory and indecisive. In general, most investigators were attempting to measure the degree of wide-sense stationarity and to estimate the amplitude probability distribution.

McEwen and Anderson [12] did some early extensive work in this area. To test for wide-sense stationarity they divided a given EEG segment into two halves and then carried out a two-sample Kolmogorov-Smirnov (K-S) test on both the sample amplitude and spectral distribution functions. This test required that both the amplitude and spectral distributions from each half could not be significantly different for the whole EEG segment to be considered as wide-sense stationary. They tested for the Gaussianity of a given EEG segment by using its amplitude distribution in a K-S goodness of fit test with unknown mean and variance using a 0.05 level of significance. They rejected their null hypothesis that EEG from awake resting subjects with eyes closed was Gaussian and wide-sense stationary approximately 15% of the time

over two second epochs and approximately 60% of the time over 8 second epochs. Persson [14] in commenting on their results, pointed out that the statistical tests that they used assume independent samples (observations) but the digitization rates used resulted in samples that were highly correlated. In fact, McEwen and Anderson noted that too high a sample rate would cause the efficacy of the statistical tests to be adversely affected and they consequently recommended sampling at a rate as little above the Nyquist rate as possible.

Persson [14] went on to argue that the maximum tolerable correlation coefficient between adjacent samples is about 0.5 and in previous work he showed, based on an estimated autocorrelation function from real EEG, that to meet this requirement the sample rates should not be much greater than 10 Hz. The obvious resulting conundrum is that if only approximately two second epochs can be considered stationary and a sample rate in the order of 10 Hz is used then the resulting number of samples would be so small that a reasonable inference cannot be made about the amplitude distribution.

Weiss [15] approached this problem by developing a correction factor, based on the second and fourth spectral moments, for the Kolmogorov-Smirnov goodness of fit test which is designed to compensate for correlation in the data two sample points back in time. He tested this method on simulated EEG data which was generated by a second order autoregressive processes. Although he reports good results on this simulated data, its usefulness is still generally limited by its ability to compensate for the correlation over only two sample points. In addition, its effectiveness, if applied to actual EEG, will be further limited by the accuracy of the estimated spectral moments. In Section 2.4 a different approach is discussed which would be

potentially more flexible and ultimately provide more information about the EEG signal characteristics.

2.3 Neural Basis for the Gaussian Nature of EEG

R. Elul was responsible for the initial work devoted to the stochastic aspects of EEG based on neuronal activity. He first suggested [16,17,18] that each individual neuron generator was independent of the summed contributions from all the neuron generators. Therefore, this resultant sum, the EEG, could be thought of as the sum of statistically independent or nearly independent neuronal contributions and since the contribution from each neuron is very small relative to the resulting EEG there must be a very large number of neurons contributing at any given time. Based on these arguments the application of the Central Limit Theorem (CLT) is justified; that is, the sum of neuronal activity will tend toward Gaussianity. However, Elul [19] later carried out an experiment in which he administered tetrodotoxin (TTX) into the brain of cats. The amount of TTX that was given to the cats should have caused about a 10% drop in neural activity. The resulting EEG activity was reduced way below the level that could be accounted for based on his concept of independent or nearly independent neural activity.

A. Siegel [20] followed up on Elul's work and he proposed the idea that a substantial proportion of the neurons belong to synchronized groups. These groups, however, would be necessarily restricted in size due to the existence of many competing inputs to a given neuron which would result in attenuation of the synchronizing effect as one moves along a chain of interacting neurons. He further postulated that because of this restricted size

there would still be a very large number of "internally synchronized but mutually unsynchronized groups of neurons". This explanation of neuronal activity was able to predict the dramatic reduction of EEG activity which occurred in the TTX experiment. It also allows for Elul's basic idea of the summed activity being independent of the activity of the individual generators. Elul [18] also suggested that different degrees of independence between neurons, as was alluded to earlier when the terms "independent" and "nearly independent" were used, would be the major influence on the degree of Gaussianity: as the dependence becomes greater the resulting distribution becomes less Gaussian.

Siegel [20] elaborates on this concept and in so doing suggests a mechanism that would produce this result. He utilizes Bernstein's [21] theory of applying the CLT to dependent variables. Roughly speaking, it states that as long as the dependence between variables decreases with separation then the CLT can be applied. Therefore, Siegel argues that as long as the dependency of two neuronal generators decreases quickly enough with increased separation, the application of the CLT can still be justified. It also follows that, at periods of time when dependency is less, the effective number of independent contributors increases and the CLT is more closely approximated. Elul [18] suggested the application of this concept to various levels of mental activation: performing an active mental task would require a greater degree of interneuronal coupling than would a mental idle state. Hence, the degree of Gaussianity would decrease during performance of mental tasks. He carried out some empirical work with EEG, which showed some support for this idea, but the statistical analysis suffered from the sample dependency problem that was described in the previous section as well as

other methodological problems.

As attractive as the above concept at first seems, it is however, contradictory to a common assumption about EEG. As Elul states: "low-voltage, fast activity implies 'desynchronized' (active) EEG, and high-voltage slow activity is indicative of 'synchronization'" (idle EEG). Siegel [20] resolves this apparent paradox with the following argument. In the idle state there are, as stated previously, groups of neurons which are internally synchronized but mutually independent, which results in summation of activity that is relatively high-voltage and appears to be in relative synchrony. During a mental task the relationship between the "within group" neurons becomes more complicated than simple "in-step" synchrony. As Siegel [20] states: "Essentially, this is because neurons must be related in configurations which correspond in complexity to that of the task itself." So the electrical neuronal activity, although more interdependent does not have the same appearance of synchronicity and hence there will be a greater amount of neuronal activity canceling each other out resulting in lower voltage EEG.

In some recent work by Anninos, Zenone and Elul [22], they studied neuronal activity and the resulting EEG based on a rigorous artificial neural net model. Their principle conclusion was that the main factor in causing the summed neural activity to deviate from Gaussianity was in fact the level of interneuronal connectivity: greater connectivity caused greater deviations from Gaussianity. This result occurred independent of the probability distribution of the membrane potential in the individual elements. In addition, they discovered that for a given level of connectivity in their model, when external input was applied, as would be the case when afferent signals were applied to the neural net, the resulting distribution became

less Gaussian. This finding may add some interesting insight to the possible differences between the motor related potentials in normal and disabled persons.

To date, it appears that the idea of relating various levels of mental activation to various levels of Gaussianity has not been carefully confirmed or rejected by empirical measurements. This is probably due to the fact, as noted in Section 2.2, that a satisfactory tool to measure the EEG statistics does not seem to be available.

2.4 Applying AR Modeling to EEG

A very general linear model for the modeling of stochastic discrete-time processes is the autoregressive moving average (ARMA) model. It is given by

$$x_i = \sum_{k=1}^{p} a_k x_{i-k} + \sum_{j=0}^{\ell} b_j e_{i-j} \qquad 2.1$$

where x_i is the discrete signal sequence of length n, i = 1,2...n, e_i is the residual error sequence and a_k, k=1,2...p and b_j, j=0,1,2....l are weighting parameters on past values of the signal and residuals respectively. The autoregressive (AR) model is the "all pole" version of the ARMA model and it has the following form

$$x_i = a_1 x_{i-1} + a_2 x_{i-2} + \ldots + a_p x_{i-p} + e_i \qquad 2.2$$

where again x_i is the signal sequence, a_i are weighting parameters, e_i is the white residual error sequence and p is the order of the AR model. The signal x_i at a given time i is assumed to be a linearly weighted sum of p

past values of x_i plus a random (white) error term e_i. Often this last term is referred to as the predictive error since it is the difference between the measured value and the predicted value.

The AR model has some significant practical advantages over the more general ARMA model. Firstly, a closed form solution to the minimization problem for the estimation of the ARMA model parameters does not exist and hence iterative numerical optimization approaches must be utilized. Whereas, in the case of the AR model, the closed form solution of the minimization problem does exist and computationally efficient methods have been developed to estimate the AR model parameters. Secondly, as noted by Kay and Marple [23], the Wold decomposition theorem demonstrates that any stationary ARMA process (in fact, any MA processes as well) of finite variance can be represented by a unique AR model which may be of infinite order. The implication is, even if it is argued that for a given signal an ARMA model is the most appropriate model, a reasonable approximation can still be achieved by utilizing an AR model with an appropriately chosen model order. Similarly, in a study by Beamish and Priestly [24] they note that the time series does not have to exactly conform to a finite AR model but rather assumes it can be modeled by an infinite AR model. Then by choosing an appropriate order which will provide in some sense the optimal fit with a finite model, the signal can be well represented. Selection of this appropriate model order is a very important issue and it is dealt with in detail in Section 4.4.

Previous work has indicated that AR modeling would prove to be a useful tool in the investigation of the EEG signal. Jansen et al. [25] note the statistical definition of regression is: "a functional relationship between two or more correlated variables used to predict values of one vari-

able when given values of the others." If these variables are time related, as would be the case with EEG, then an autoregressive model can be applied [26]. In general, it is not clear how to definitively assess whether an AR model at a given model order is adequately representing a segment of EEG. For purposes of this study, the appropriateness of AR modeling will be discussed in terms of its relative performance when it is applied to the spectral estimation of EEG (see Section 4.4).

The AR model as applied to EEG can be utilized in several ways. The estimated parameters could possibly be used as features in a signal detection/discrimination problem. As indicated above, the parameters can be used in spectrum estimation. Many researchers [23,25,27] have demonstrated that there are some distinct advantages of this approach over the conventional FFT methods of spectrum estimation. Another benefit of applying AR modeling to EEG is the fact that residuals, ideally, have the correlation of the process removed (whitened) and since the AR process is linear, the statistical characteristics of the original process are still contained in the residuals. Although Andrews [28] demonstrates that residuals from a non-Gaussian process will tend to mask the evidence of non-Gaussianity, Chambers and Heathcote [29] have developed a method of characterizing the Gaussianity of a process based on a scale factor which is determined by the characteristic function of the residual error distribution. The main benefit of this approach is that it overcomes the problems of correlated data samples, as was discussed in Section 2.2.

The greatest benefit of applying the AR model to EEG in terms of this thesis work lies in the fact that it has a very convenient state-space representation, which allows for the straight forward use of state-space

techniques such as Kalman type filtering [30]. These techniques play a major role in the EEG single trial processing scheme that is discussed in detail in Chapter 4.

CHAPTER 3

THEORY OF AR MODELING

3.1 Conventional AR Parameter Estimates

EEG signal characteristics are changing over time and hence, a single time-invariant model can not be applied. This results in the need to estimate model parameters from the EEG signal in a manner that will attempt to account for time varying characteristics. The estimation of the AR model parameters can generally be carried out either by block mode estimation or by recursive estimation. Recursive methods sequentially update the parameters data point by data point. They have the potential advantage of being set up such that the estimation of parameters adapts to time varying characteristics of a signal [31]. This is essentially accomplished by assigning greater weight to newer information than to older information. It has been demonstrated for EEG signals, which are slowly time-varying over long epochs, that adaptive recursive estimation schemes can be effective [32,33]. However, Jansen [25,32] provides evidence which indicates that the adaptation process is not rapid enough for short segment analysis of EEG signals that are expected to be time-varying relatively quickly. Short segments in the order of 1 to 2 seconds are typical for the work carried out in this thesis. Hence, a block mode approach, where the AR model parameters are estimated on the basis of a short data segment, is employed throughout this work.

Various block mode methods have been used to estimate a set of AR model parameters from a sample signal segment. The following discussion describes the most common conventional methods for AR parameter estimation.

In the past, the standard method to estimate AR parameters was often based on the Yule-Walker equations. These equations provide a relationship between the autocorrelation function and the AR model parameters (see Ulrych and Bishop [34]. The derivation of these equations is reviewed by Kay and Marple [23] and the final result expressed in matrix form is

$$\begin{bmatrix} R_{xx}(0) & R_{xx}(-1) & \cdots & R_{xx}(1-p) \\ R_{xx}(1) & R_{xx}(0) & \cdots & R_{xx}(2-p) \\ \vdots & & & \vdots \\ R_{xx}(p-1) & R_{xx}(p-2) & \cdots & R_{xx}(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ ap \end{bmatrix} = \begin{bmatrix} R_{xx}(1) \\ R_{xx}(2) \\ \vdots \\ R_{xx}(p) \end{bmatrix} \qquad 3.1$$

where $R_{xx}(k)$ is the autocorrelation function for lag k, p is the AR model order, and a_k k=1,2...p are the AR model parameters. Therefore, by obtaining estimates of the autocorrelation function, estimates of the AR parameters can be obtained by solving the system of Equations 3.1. Kay and Marple recommend, to achieve low mean-squared error, estimation of the autocorrelation function at specific lags with the following expression

$$\hat{R}_{xx}(m) = \frac{1}{n} \sum_{i=0}^{n-m-1} x_{i+m} x_i \qquad 3.2$$

Note also that for a stationary process the conjugate symmetry property of the autocorrelation function $R_{xx}(m)=R_{xx}(-m)$ can also be utilized in solving 3.1. In addition, the Yule-Walker equations yield an expression that allows the variance of the residuals to be calculated [23]

$$R_{xx}(0) = \sum_{k=1}^{p} a_k R_{xx}(k) + \delta_e^2 \qquad 3.3$$

where δ_e^2 is the variance of the residual sequence. Equation 3.1 can be augmented with 3.3 to yield the following alternative form of the Yule-Walker equations [23]

$$\begin{bmatrix} R_{xx}(0) & R_{xx}(-1) & \cdots & R_{xx}(-p) \\ R_{xx}(1) & R_{xx}(0) & & R_{xx}(-(p-1)) \\ \vdots & \vdots & & \vdots \\ R_{xx}(p) & R_{xx}(p-1) & \cdots & R_{xx}(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \delta_e^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad 3.4$$

Least square (LSQ) estimation of the AR parameters is another very common method. In fact, as will be shown later, all the conventional methods discussed in this section can be shown to be based on least squares minimization criteria. The method that is most commonly referred to as the "LSQ" method can be derived in the following manner [23]. From Equation 2.2 the prediction error can be written as

$$e_i = x_i - \sum_{k=1}^{p} a_k x_{i-k} \qquad 3.5$$

The sum of squared prediction errors is then

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [x_i - \sum_{k=1}^{p} a_k x_{i-k}]^2 \qquad 3.6$$

To find the AR parameters that minimize 3.6 the partial derivatives with respect to each a_k are taken and set equal to zero. That is

$$\frac{\partial (SSE)}{\partial a_q} = 0 \qquad q=1,2...p \qquad 3.7$$

The result of applying 3.7 to 3.6 is

$$\sum_{k=1}^{p} a_k \sum_{i=1}^{n} x_{i-k} x_{i-q} = \sum_{i=1}^{n} x_i x_{i-q} \qquad q=1,2...p \qquad 3.8$$

Then by substituting 3.8 into 3.6, the minimum SSE can be shown to be [23]

$$SSE_{min} = \sum_{i=1}^{n} x_i^2 + \sum_{k=1}^{p} a_k \sum_{i=1}^{n} x_i x_{i-k} \qquad 3.9$$

Now by expanding 3.8 into matrix form for model order p results in

$$\begin{bmatrix} \sum x_{i-1}x_{i-1} & \sum x_{i-2}x_{i-1} & \cdots & \sum x_{i-p}x_{i-1} \\ \sum x_{i-1}x_{i-2} & \sum x_{i-2}x_{i-2} & & \vdots \\ \vdots & & & \vdots \\ \sum x_{i-1}x_{i-p} & \cdots & & \sum x_{i-p}\sum x_{i-p} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \sum x_i x_{i-1} \\ \sum x_i x_{i-2} \\ \vdots \\ \sum x_i x_{i-p} \end{bmatrix} \qquad 3.10$$

By calculating the summations in 3.10, from a data sequence of n points, this system of equations can be solved to determine the LSQ estimate of the AR parameters. Note that if the autocorrelation estimate given in 3.2 is used, except for the scaling factor, 1/n, which does not effect the solution for the AR parameters, the above equations reduce to the Yule-Walker method of parameter estimation. Therefore, the Yule-Walker estimates are equivalent to the LSQ estimates for sufficiently large n.

An implicit assumption, which is particularly evident in the development of the Yule-Walker solution, is that the autocorrelation function is assumed to be zero outside the data segment of interest. In practice, when relatively short segment lengths are used, this truncation of the autocorrelation function can result in relatively poor parameter estimates [25,23]. Burg [35] addressed this problem by using extrapolation of the autocorrelation function based on concepts of maximum entropy and he formulated a method of AR parameter estimation, known as the Burg method or the maximum entropy method (MEM). Many authors have noted the superior performance of this method over the Yule-Walker type methods when applied to relatively short data segments (for example see [23,25,36,37]). Therefore, in this thesis work, because typically relatively short epochs of EEG are utilized, MEM was selected as the conventional AR parameter estimation method.

The fundamental idea behind Burg's method is to provide a non-zero extrapolation of the p_{th} autocorrelation function beyond the known lags (up to and including the p lag: see Equation 3.1) as opposed to the implied zero extrapolation as in the Yule-Walker equations. Burg argued that the extrapolation should impose the fewest possible constraints on the extrapolated autocorrelation function without compromising any information about the known lags. To achieve this, he required that the hypothetical time series, which would be represented by the extrapolated autocorrelation function, should have maximum entropy. This requirement maximizes the randomness of that time series, given the constraints of the estimate of the function, and hence produces a minimum bias solution. From a spectral point view, the maximum entropy estimate is based on choosing a spectral estimate such that the entropy (E) per sample

$$E = \int_{-1/2f_s}^{1/2f_s} \ln \hat{F}_x(f)\, df \qquad 3.11$$

where $F_x(f)$ is the spectral estimate of the data segment and f_s is the sample frequency, is maximized subject to

$$\int_{-1/2f_s}^{1/2f_s} F_x(f) \exp-(j2\pi k\frac{f}{f_s})\, df = \hat{R}_{xx}(k) \qquad k=1,2, \ldots p \qquad 3.12$$

It can be shown [38] that the spectral estimate which maximizes entropy subject to the constraint that its first p Fourier coefficients correspond exactly to the sample autocorrelation function evaluated at the first p lags is the estimate of the spectral density function of an AR model of order p. In addition, it is shown that the estimates of the parameters and the power

in the residual sequence, $\hat{P}_p$, can be written as

$$\begin{bmatrix} \hat{R}_{xx}(0) & \cdots & \hat{R}_{xx}(p) \\ \vdots & & \vdots \\ \hat{R}_{xx}(p) & \cdots & \hat{R}_{xx}(0) \end{bmatrix} \begin{bmatrix} 1 \\ \hat{a}_1 \\ \vdots \\ \hat{a}_p \end{bmatrix} = \begin{bmatrix} \hat{P}_p \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad 3.13$$

Note that these equations are of the same form as the augmented Yule-Walker Equations 3.4 where the variance of the residuals for the zero mean residual sequence is equal to P_p.

The algorithm that was utilized in this thesis work to calculate MEM parameter estimates is based on a procedure outlined by Anderson [36]. The system of Equations 3.13 is solved in a sequential manner. Beginning with p=0, P_o, is estimated by

$$P_o = \frac{1}{n} \sum_{i=1}^{n} x_i^2 \qquad 3.14$$

Then the model for p=1 is determined as that which minimizes the power in the forward prediction error sequence averaged together with the power in the backward prediction error sequence. This average power is given by

$$\pi_1 = \frac{1}{2} \frac{1}{n-1} \sum_{i=1+1}^{n} [(x_i - a_1(1)x_{i-1})^2 + (x_{i-1} - a_1(1)x_i)^2] \qquad 3.15$$

For the general case of progressing from order p-1 to order p, it is shown that

$$\pi_p = \frac{1}{2} \frac{1}{n-p} \sum_{i=p+1}^{n} [(x_i - \sum_{k=1}^{p} a_k(p)x_{i-k})^2 + (x_{i-p} - \sum_{k=1}^{p} a_k(p)x_{i-p+k})^2] \qquad 3.16$$

and it can be minimized with respect to the single parameter $a_p(p)$. Note that the arguments on the parameters indicate the current maximum model order at a given stage in this sequential procedure. The dependence of the other model parameters at the pth stage are given by the Levinson recursion

$$a_k(p) = a_k(p-1) - a_p(p)a_{p-k}(p-1) \qquad 3.17$$

Applying Equation 3.17 to 3.16, taking $\frac{\partial \pi_p}{\partial a_p(p)} = 0$ and solving for $a_p(p)$ results in

$$a_p(p) = \frac{2 \sum_{i=p+i}^{n} b_{p-1}(i-1)e_{p-1}(i)}{\sum_{i=p+1}^{n} (|b_{p-1}(i-1)|^2+|e_{p-1}(i)|^2)} \qquad 3.18$$

where $b_p(i)= \sum_{i=0}^{p} a_p(i)x(i-p+k)$ and $e_p(i)= \sum_{i=0}^{p} a_p(i)x(i-k)$ are the backward and forward prediction errors. In the above notation it is assumed that $a_p(0)=1.0$. Hence, the value of $a_p(p)$ is calculated via Equation 3.18, with the remaining parameters being calculated via Equation 3.17 and then the backward and forward prediction error sequences are updated. The order is then increased to p+1 and the procedure is repeated until the desired model order is reached.

van den Bos [38] noted that MEM is equivalent to a least squares fitting of an AR model. More specifically, Kay and Marple [23] note that MEM can be viewed as a constrained least squares minimization problem because, as is indicated by Equation 3.15, the sum of both the forward and backward error energies (squared error terms), as given below

$$E = \sum_{i=p+1}^{n} |e_p(i)|^2 + \sum_{i=p+1}^{n} |b_p(i)|^2 \qquad 3.19$$

is the minimization criteria for MEM.

3.2 LSQ Parameter Estimation on Simulated Gaussian Data

LSQ methods, as is shown in Section 3.3, are optimal when applied to a Gaussian random process in terms of maximum likelihood estimation which provides the best linear unbiased estimates. Simulation studies using LSQ parameter estimation were carried out on computer generated Gaussian AR data. The simulated data were generated using a Gaussian sequence, e_i with variance 1.0 and mean 0.0, driving an 8th order AR model with the following parameters:

$a_1 = 0.838$ $a_2 = -0.471$ $a_3 = 0.638$ $a_4 = -0.429$

$a_5 = 0.518$ $a_6 = -0.304$ $a_7 = 0.182$ $a_8 = -0.243$

This set of parameters was calculated from EEG data and the set was selected as being typical for an 8th order AR model of actual EEG. The Gaussian sequence e_i was generated from a uniform white sequence via the procedure given by Box and Muller [39]. The uniform pseudo-random number generator was based on a "generalized feedback shift register algorithm" which is given by T.G. Lewis and W.H. Payne [40]. They demonstrate that this generator has excellent random properties and a sequence period of 2 raised to the jth power, where j is the integer word length which in this case was 31.

Fifty sets of estimations using different random data segments of lengths n=1000, n=500, n=100, n=64, and n=32 were carried out. The mean squared error (MSE) between the actual and estimated parameter values were

calculated for each set. The results are summarized in Figure 3.1. These results show that the MSE began to increase significantly when the segment length was decreased to n=100 and they became very large when the segment length was reduced to n=32. This points to, in terms of parameter estimation efficacy, a practical lower bound on segment size. Ideally, segment sizes in the order of n=500 would be preferable, but in the case of actual EEG, there is a need (see Section 4.5) to make the segment size as small as possible. The EEG data was sampled at 64 Hz (see Section 4.1), and therefore, a one second segment contains only 64 data points. Since the MSE and SE became so large at n=32 and since it was significantly less at n=64, it was decided that one second (n=64) should be the absolute lower bound on segment length for EEG data. Ultimately, a segment size of 1.5 seconds (n=96), moderately above this lower bound, was utilized in the single trial analysis method (see Section 4.4). Further discussion on segment length relating specifically to the EEG data used in this thesis work is provided in Section 4.5.

3.3 Deviations from Gaussianity

Several researchers, such as McGillem et al. [13], Jansen et al. [25] and Smith and Lager [41], employed AR parameter estimation in various applications involving EEG. In this work, the required parameter estimation was based on estimation methods involving least squares, which are generally optimal in Gaussian processes. However, it has been shown [42] that the performance of these methods can, under certain circumstances, significantly break down under even slight deviations from Gaussianity. Although it was discussed earlier in Section 2.2 and 2.3 that there was considerable evidence

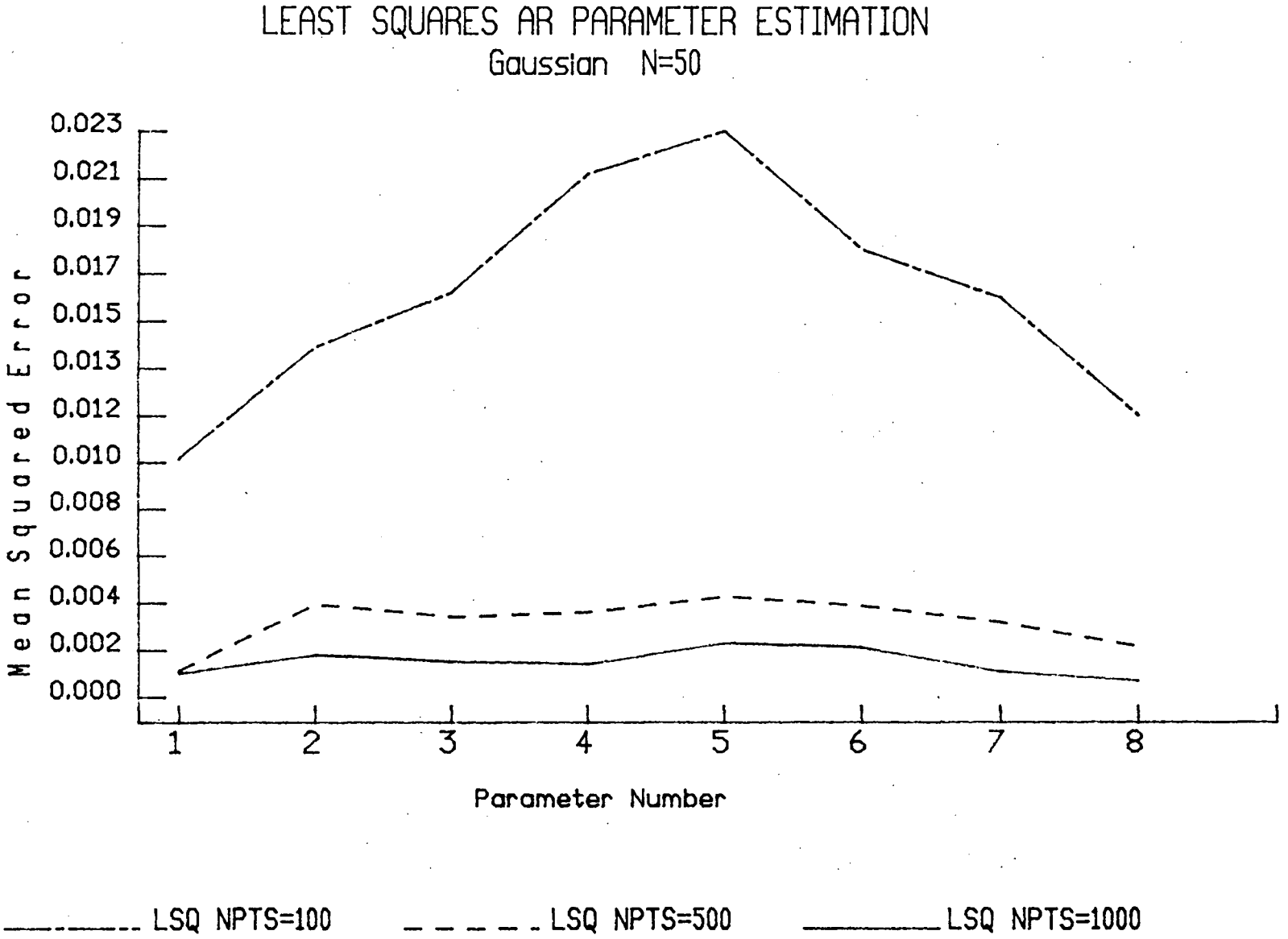

LEAST SQUARES AR PARAMETER ESTIMATION
Gaussian N=50
Mean Squared Error
0.023
0.021
0.019
0.017
0.015
0.013
0.012
0.010
0.008
0.006
0.004
0.002
0.000
1
2
3
4
5
6
7
8
Parameter Number
LSQ NPTS=100
LSQ NPTS=500
LSQ NPTS=1000

Figure 3.1a

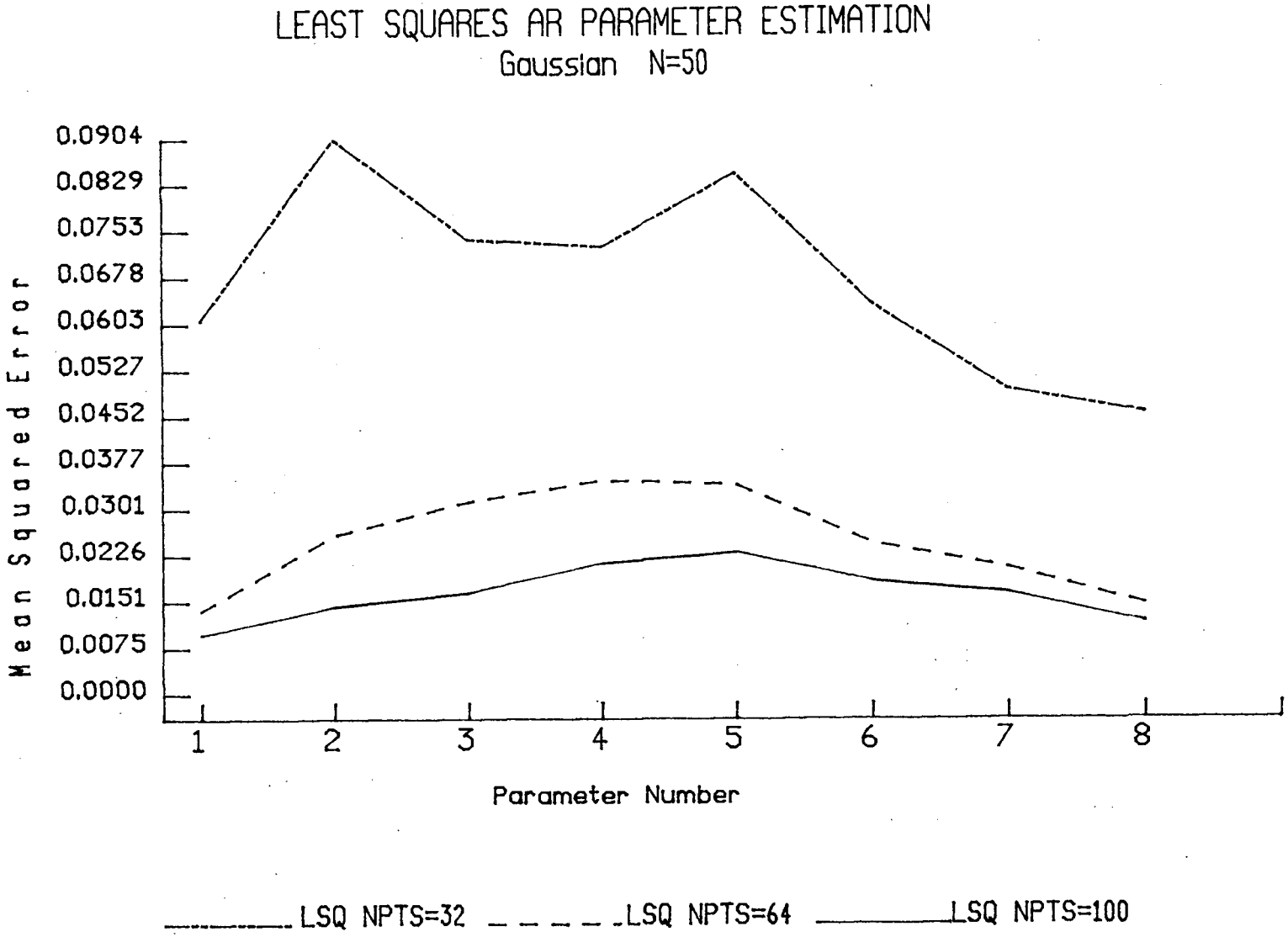

LEAST SQUARES AR PARAMETER ESTIMATION
Gaussian N=50
Mean Squared Error
0.0904
0.0829
0.0753
0.0678
0.0603
0.0527
0.0452
0.0377
0.0301
0.0226
0.0151
0.0075
0.0000
1
2
3
4
5
6
7
8
Parameter Number
LSQ NPTS=32
LSQ NPTS=64
LSQ NPTS=100

Figure 3.1b

to expect EEG to be generally Gaussian, there is certainly no guarantee that sample segments of EEG will always be strictly Gaussian. This prompts questions about both the Gaussian nature of sample segments of EEG, in particular relatively short segments, and the resulting performance of AR model parameter estimation techniques under conditions of varying degrees of deviation from Gaussianity.

A considerable amount of work has been carried out on robust estimation of location parameters and linear regression model parameters in the case of independent and identically distributed (i.i.d.) observations. The basic goal of a robust procedure, for the purposes of this thesis, is to provide good estimates when the data has a small number of outliers (in the order of 5 to 20 percent) causing the assumed, Gaussian in the case of EEG, distribution function to be contaminated. In addition, the robust procedure should provide estimation results which are not significantly different from the conventional LSQ methods when the data is not contaminated with outliers.

Location Case

There are a number of robust methods but the most satisfying to date appear to be those given by modifications to maximum likelihood. Hogg [43] provides a good background tutorial on robust methods and the following brief description of modified maximum likelihood methods is based on that tutorial. If $x_1, x_2, \ldots\ldots x_n$ are a random sample from a probability density function $f(x-\Theta)$ where Θ is a location parameter, then the logarithm of the likelihood function is given by

$$\ell nL(\Theta) = \sum_{i=1}^{n} \ell nf(x_i-\Theta) = - \sum_{i=1}^{n} \rho(x_i-\Theta) \qquad 3.20$$

The maximum likelihood method maximizes ln $L(\Theta)$ or in terms of the ρ function

minimizes

$$\sum_{i=1}^{n} \rho(x_i-\Theta) = K(\Theta) \quad . \tag{3.21}$$

Assume that the minimization can be achieved by differentiating and solving $K'(\Theta) = 0$. In other words, find the value of Θ that satisfies

$$\sum_{i=1}^{n} \psi(x_i-\Theta) = 0 \tag{3.22}$$

where

$$\psi(t) = \rho'(t) = -f'(t)/f(t) \tag{3.23}$$

The value of Θ that minimizes $K(\Theta)$ is termed the maximum likelihood estimate of Θ and is denoted as $\hat{\Theta}$. Robust M-estimates are generated via Equation 3.22 except that different psi functions are used than that described in Equation 3.23. Each different psi function describes a specific type of M-estimate. The basis of robust M-estimates is to find psi (ψ) functions that will protect against outlier data points that cause undue influence on the estimation result. An example of such a psi function is that due to Huber [44]. It is designed to deal with data that is distributed normally in the middle with double exponential tails ("heavy-tailed" distribution). The psi function is given by

$$\psi(t) = \begin{cases} -c & t < -c \\ t & |t| \leq c \\ c & t > c \end{cases} \tag{3.24}$$

where c is a tuning constant. A scale invariant version of the M-estimator is given as

$$\sum_{i=1}^{n} \psi\left(\frac{x_i-\Theta}{s}\right) = 0 \tag{3.25}$$

where s is a robust estimate of the process scale. In this case, $c = 1.5$

is often selected to allow data from a truly Gaussian distribution to be uninfluenced by this psi function while still providing the desired protection from outlying points. The solution to Eqn. 3.25 is typically found by a block mode iteration scheme, such as the iterated-weighted least squares (IWLS) procedure [45]. Other commonly used psi functions are Hampel's three part redescending and Tukey's Bi-weight.

It should be noted that, the calculation of the psi function for a Gaussian distribution demonstrates that the LSQ estimate is the maximum likelihood estimator. The Gaussian distribution is given by [46]

$$f(x-\Theta) = \frac{\exp\left[\frac{-(x-\Theta)^2}{2\sigma^2}\right]}{\sqrt{2\pi\sigma^2}} \qquad 3.26$$

and hence

$$\rho(x-\Theta) = \frac{1}{2}\,\ell n(2\pi\sigma^2) + \frac{(x-\Theta)^2}{2\sigma^2} \qquad 3.27$$

and

$$\psi(x-\Theta) = \frac{1}{\sigma^2}\,(x-\Theta) \qquad 3.28$$

Applying Eqn. 3.22 it follows that the maximum likelihood estimator is the value of Θ that satisfies

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\Theta) = 0 \qquad 3.29$$

yielding the well known result

$$\hat{\Theta} = \bar{x} \qquad 3.30$$

which is exactly the LSQ estimate of Θ [43].

Regression Case

The above methods can be extended to the block mode linear regression case (see Hogg [43]). Given the linear model

$$\underline{Y} = X\underline{a} + \underline{e} \qquad 3.31$$

where: X is n x p data matrix

$\underline{a}$ is a parameter vector of order p

$\underline{e}$ is a residual vector of order n

p is the order of the model

n is the number of data points

It then follows that the expression to be minimized is now

$$\sum_{i=1}^{n} \rho\left(\frac{y_i - \underline{x}_i^T \underline{a}}{s_e}\right) \qquad 3.32$$

where: y_i is the ith data point

$\underline{x}_i$ is the ith row of the matrix X

Considering the p first partial derivatives, the following set of p equations must be solved

$$\sum_{i=1}^{n} \underline{x}_i \, \psi\left(\frac{y_i - \underline{x}_i^T \underline{a}}{s_e}\right) = 0 \qquad 3.33$$

where s_e is a robust estimate of the standard deviation of the residuals. Again, this set of equations can be solved using a IWLS procedure and Hogg [43] recommends the following robust estimate

$$s_e = \frac{\text{median } |e_i - \text{median}(e_i)|}{0.6745} \qquad 3.34$$

Autoregression Case

The application of robust methods to time-series data has lagged behind the application to the i.i.d. case, probably, as suggested by Martin [47], because of the considerable diversity in qualitative features of time-series data sets as well as the possible dependency that may arise in the residuals due to data outliers that occur in patches (correlated). Martin

[47] applies robust estimation to time-series data and shows that many of the concepts from the i.i.d. observations case can be applied directly. The greatest difference seems to occur in the definitions of the robust qualities of the given methods due mostly to the added difficulties mentioned above. Martin [47] also points out that for data to qualify as a time-series outlier it only has to be "different" on the innovations (residual) scale not the process scale. Since the innovations scale is typically 10 to 10,000 times smaller than the process scale the outliers will often be impossible to visually detect in a plot of the raw data.

Martin [47] identifies three types of outliers that may occur in time-series data:

1) independent isolated gross-error outliers which may be caused by various recording (measurement) errors

2) patchy type outliers whose behavior seems to be uncorrelated with the behavior of the rest of the data - this may be caused by brief malfunctions in the data collection system, inherent behavior of the process or maybe other unaccountable effects

3) patchy outliers whose behavior does appear to be related to the rest of the data with the possible exception of an initial jump - this type of outlier may be caused by unusual events within the process

He also suggests that two types of outlier models can reasonably simulate the above types of outlier activity. The additive outlier model would apply for types 1 and 2 while the innovations outlier model would apply for type 3.

The Innovations Outlier (IO) model is described in [47] as

$$x_i = \sum_{k=1}^{p} a_k x_{i-k} + e_i \qquad 3.35$$

where the innovation sequence e_i is i.i.d. with a symmetric distribution G and the observations are given by

$$y_i = x_i \quad . \tag{3.36}$$

Innovation outliers occur when G is heavy-tailed. Martin and Thompson [45] have found that this type of deviation from Gaussianity does not cause, except perhaps in extreme cases, serious problems for the conventional estimation of parameters from a time-series.

In the case of the Additive Outlier (AO) model, the observations are given by

$$y_i = x_i + v_i \tag{3.37}$$

where x_i is defined as in 3.35 with G Gaussian, v_i is independent of x_i, and v_i has a symmetric distribution. A suitable distribution for v_i for the i.i.d. case is the contaminated-normal with degenerate central component which has the following form [45]

$$CND(\gamma,\sigma^2) = (1-\gamma)N(0,0) + \gamma N(0,\sigma^2) \tag{3.38}$$

where γ is the proportion of contamination and the notation $N(\mu,\sigma^2)$ represents a normal (Gaussian) distribution with mean μ and variance σ^2. With this distribution, the probability that $v_i = 0$ is the probability that $y_i = x_i$ which equals $1-\gamma$ and in typical applications $0.01 < \gamma < 0.25$. In contrast to the IO case, it has been shown [45,47,48] that conventional time-series parameter estimation is highly non-robust under this additive type of contamination. Although, v_i has been restricted above to the i.i.d. case it has been found [45] that schemes which deal well with this type of outlier also deal best with the patchy type of outliers. Furthermore, Martin and Thompson [45] point out that, in practice, the details of the outlier distribution for v_i are largely irrelevant because it would be a poor robust estimator that depended significantly on a given distribution for v_i.

A generalized robust M-estimate (GM-estimate) [47], a member of the

class of bounded-influence autoregression (BIFAR) estimates, was shown to have robust qualities for the AO case and is given by [42]

$$\sum_{i=p+1}^{n} \underline{x}_{i-p} \, W(\underline{x}_{i-p}) \, \psi \left(\frac{y_i - \underline{x}_{i-p}^T \underline{a}}{s_e} \right) = 0 \qquad 3.39$$

Note that except for the weighting factors $W(\underline{x}_i)$, these equations are of the same form as the equations given for the M-estimate. As described by Martin [47], the role of the additional weighting factors is to down-weight the summands of Equation 3.39 for which $\underline{x}_i^T \underline{a}$ is a poor predictor because one or more of the values in $\underline{x}_i$ are too large. He shows that an appropriate calculation of these weights can be achieved by letting

$$W(\underline{x}_i) = w(d_i) \qquad 3.40$$

where d_i is defined as

$$d^2(\underline{x}_i) = \frac{\underline{x}_i^T \, C^{-1} \, \underline{x}_i}{p} \qquad 3.41$$

and C is the pxp covariance matrix for the p^{th} order AR model of the process and w(.) is a non-negative decreasing weighting function, typically of the form

$$w(t) = c \, \psi(t/c)/t \qquad 3.42$$

where c is a tuning constant. Hence, there are at least two tuning constants required: one for the psi function and one for the w function. The value d^2 in 3.41 is proportional to $\underline{x}_i^T C^{-1} \underline{x}_i$; this expression, known as the Mahalanobis distance, provides, as noted by Martin [47], a natural metric by which the relative "largeness" of $\underline{x}_i$ can be determined.

The GM-estimates can be solved using an IWLS procedure as given below [45]

$$\sum_{i=p+1}^{n} \underline{x}_{i-p} W(\underline{x}_{i-p}) \; \beta_i^j \; (y_i - \underline{x}_{i-p}^T \; \hat{\underline{a}}^{j+1}) = 0 \qquad j = 1,2, \ldots \text{NIT} \qquad 3.43$$

where

$$\beta_i^j = \hat{s}_e^j \cdot \frac{\psi\left(\dfrac{y_i - \underline{x}_{i-p}^T \; \hat{\underline{a}}^j}{\hat{s}_e^j}\right)}{y_i - \underline{x}_{i-p}^T \; \hat{\underline{a}}^j} \qquad 3.44$$

The process to obtain GM estimates requires starting with model order p=1 and sequentially increasing p until the desired order is reached. At each model order in this sequence, a set of p equations resulting from Equation 3.43 are solved for each of the NIT iterations. MEM estimates for $\underline{a}$ and the corresponding robust estimate of s_e are used as starting values for the iterations of Equation 3.43. The reason that the model parameters must be estimated sequentially is due to the manner in which a robust estimate of C^{-1} is calculated. It has been found [47] that a successful approach to obtain a robust estimate of C^{-1} is based on the factorization

$$C^{-1} = A^T A \qquad 3.45$$

where A is upper triangular and is given by

$$A_{kl} = \begin{cases} \dfrac{-\underline{a}(p-k)_{l-k}}{s_e(p-k)} & l > k \\ \dfrac{1}{s_e(p-k)} & l = 1 \\ 0 & l < k \end{cases} \qquad 3.46$$

where $k = 1,2,\ldots,p-1$ and $\underline{a}(p-k)$, $s_e(p-k)$ are the parameter estimates and residual standard deviation for model order p-k and the required starting value of $s_e^2(0)$ is set equal to the variance of the original data sequence

y_i. Hence, the pxp matrix C^{-1} is represented in terms of the AR parameters and the corresponding residual standard deviations derived for model orders up to p-1. Therefore, by fitting AR models in succession, the prior GM parameter estimates at model order p-1 will enable the construction of A^p, the A matrix at the p^{th} iteration, which provides the robust estimate of C^{-1} via Equation 3.45 to be used for the current GM parameter estimate at order p.

3.4 GM Estimation on Simulated Contaminated Gaussian Data

Simulation studies were carried out to determine the relative performance of robust GM estimation methods as compared to the conventional LSQ (MEM) estimation method when applied to eighth order AR Gaussian data with 0%, 10% and 20% levels of AO contamination. Additive outliers are studied because conventional parameter estimation is not robust under this type of contamination (see Section 3.3) and also because, the single trial analysis method (see Section 4.2) is largely based on additive outlier concepts. The contamination for these studies was produced based on the AO model given in Equation 3.37 with the distribution G being Gaussian with variance 1.0 and mean 0.0. The distribution for v_i was of the form given in Equation 3.38 with σ^2=2.0 and γ=0, 0.1, and 0.2. The simulations were carried out on data segments with a length equal to 100 points because this reflects the segment length used on the actual EEG signals as described in Chapter 4.

Figures 3.2, 3.3, and 3.4, show the MSE performance of the LSQ, GM, GM1, and GM2 estimation methods on fifty (N=50) random simulated 8th order AR

Gaussian processes with 0%, 10%, and 20% levels of AO contamination respectively. The GM1 and GM2 methods are extensions to the GM estimate which will be described in Section 4.3. In these studies, as suggested by Martin and Thompson [45], the Huber psi function, as described in Equation 3.24, was used for the first two iterations (j=1,2) of Equation 3.43 and then Tukey's bi-weight

$$\psi(t) = \begin{cases} t(1 - (t/c)^2)^2 & |t| < c \\ 0 & |t| \geq c \end{cases} \qquad 3.47$$

was used for the last iteration (j=3=NIT).

Through trial and error, the various tuning constants were selected to provide the best performance in term of MSE and are summarized below

	c
Huber psi	1.0
Tukey psi	3.0
w based on Huber psi	1.3

These results show that the robust methods perform, in terms of MSE, almost as well as the LSQ method on the uncontaminated Gaussian data. With a 10% level of AO contamination the LSQ performance falls off dramatically. The robust methods perform much better with the GM2 method performing almost as well as in the uncontaminated case. At the 20% level of contamination the performance across all the methods has dropped significantly. However, the GM2 method is still clearly the best performer with a MSE generally less than the LSQ method in the 10% contamination case. The results support the expectations of robust estimation in that the robust methods perform nearly as well as the LSQ method on uncontaminated data but perform significantly better than the LSQ method on contaminated data.

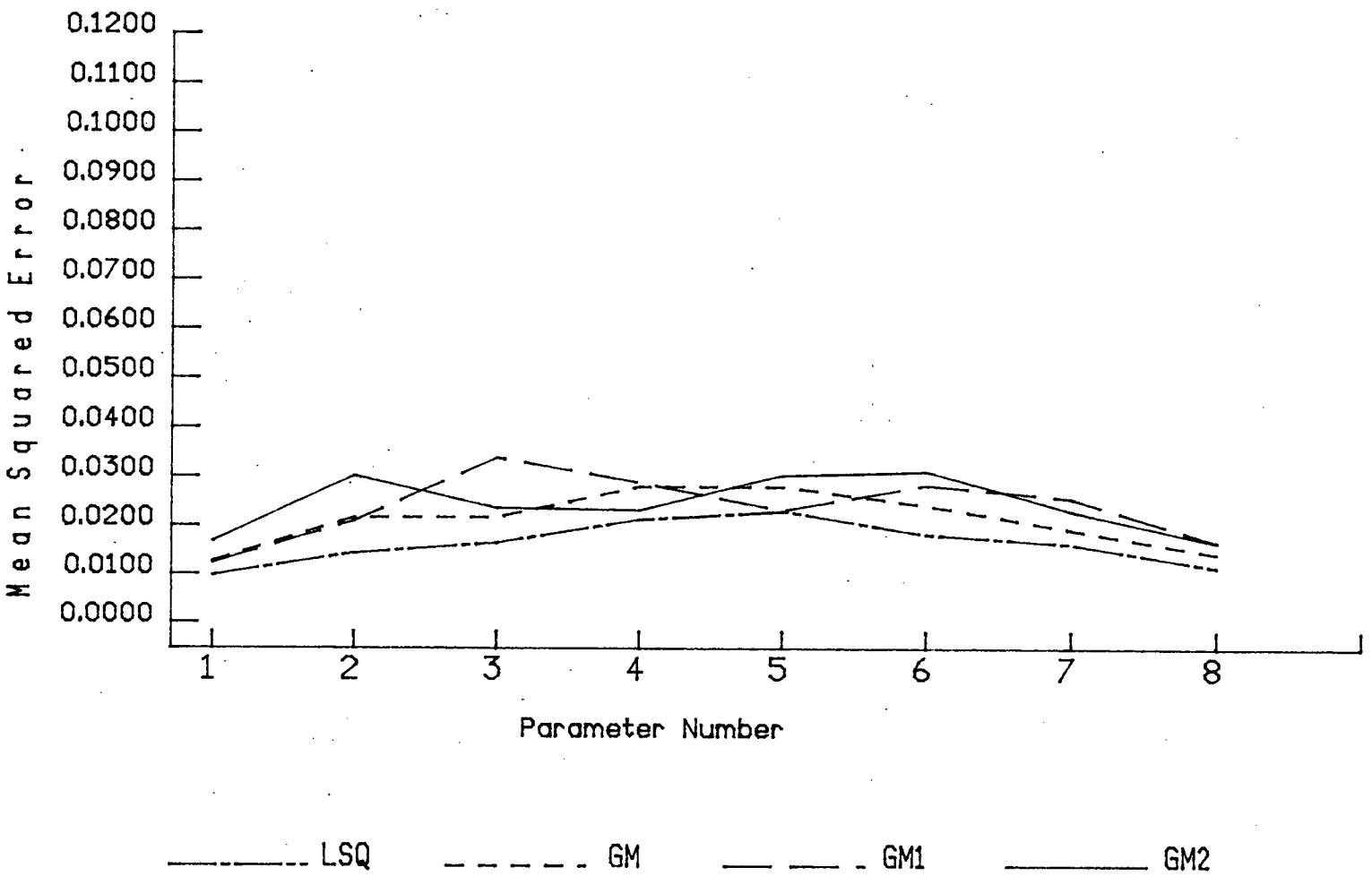

AR PARAMETER ESTIMATION
GAUSSIAN N=50
Mean Squared Error
0.1200
0.1100
0.1000
0.0900
0.0800
0.0700
0.0600
0.0500
0.0400
0.0300
0.0200
0.0100
0.0000
1
2
3
4
5
6
7
8
Parameter Number
LSQ
GM
GM1
GM2

Figure 3.2

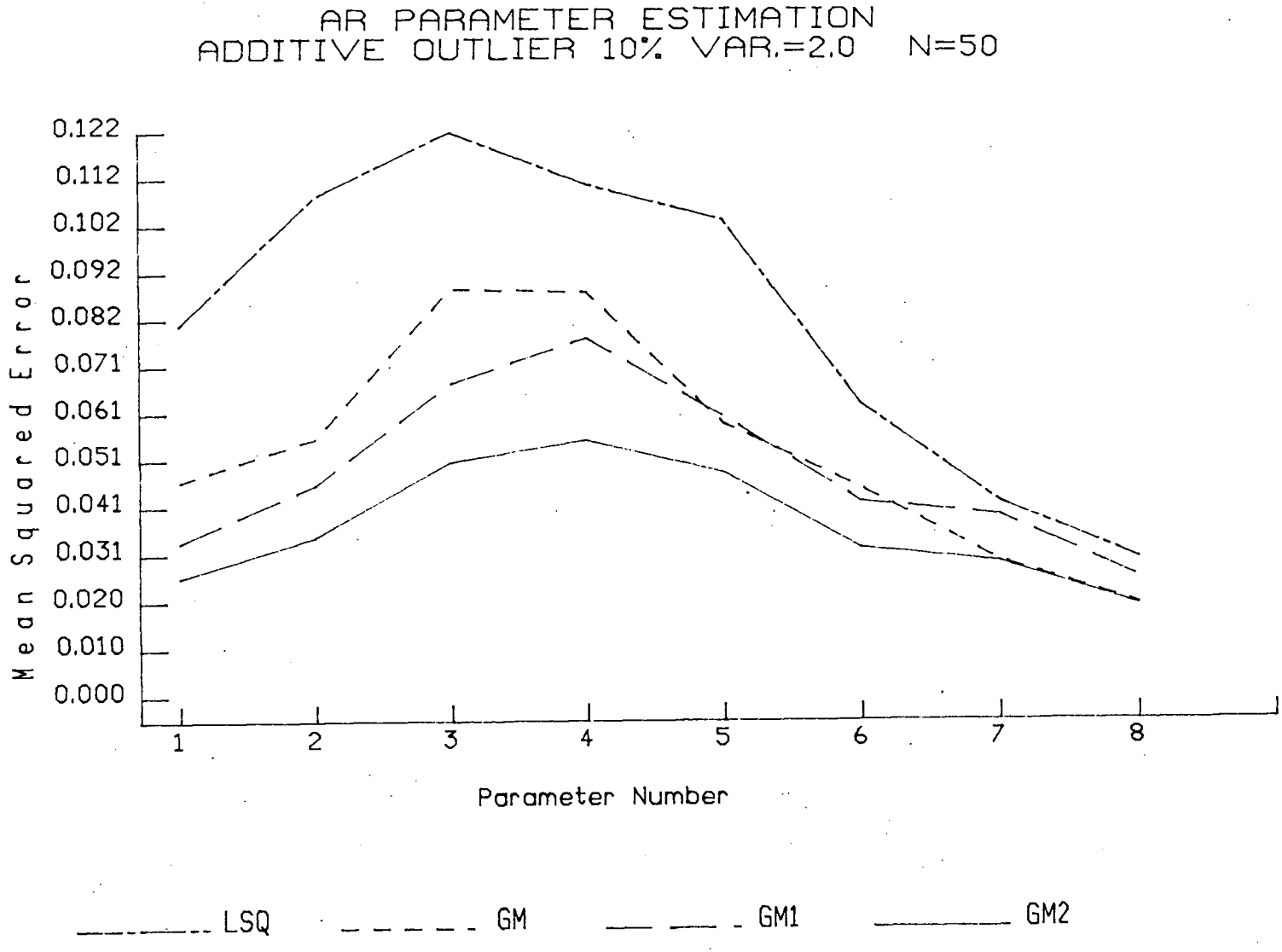

AR PARAMETER ESTIMATION
ADDITIVE OUTLIER 10% VAR.=2.0 N=50
Mean Squared Error
0.122
0.112
0.102
0.092
0.082
0.071
0.061
0.051
0.041
0.031
0.020
0.010
0.000
1
2
3
4
5
6
7
8
Parameter Number
LSQ
GM
GM1
GM2

Figure 3.3

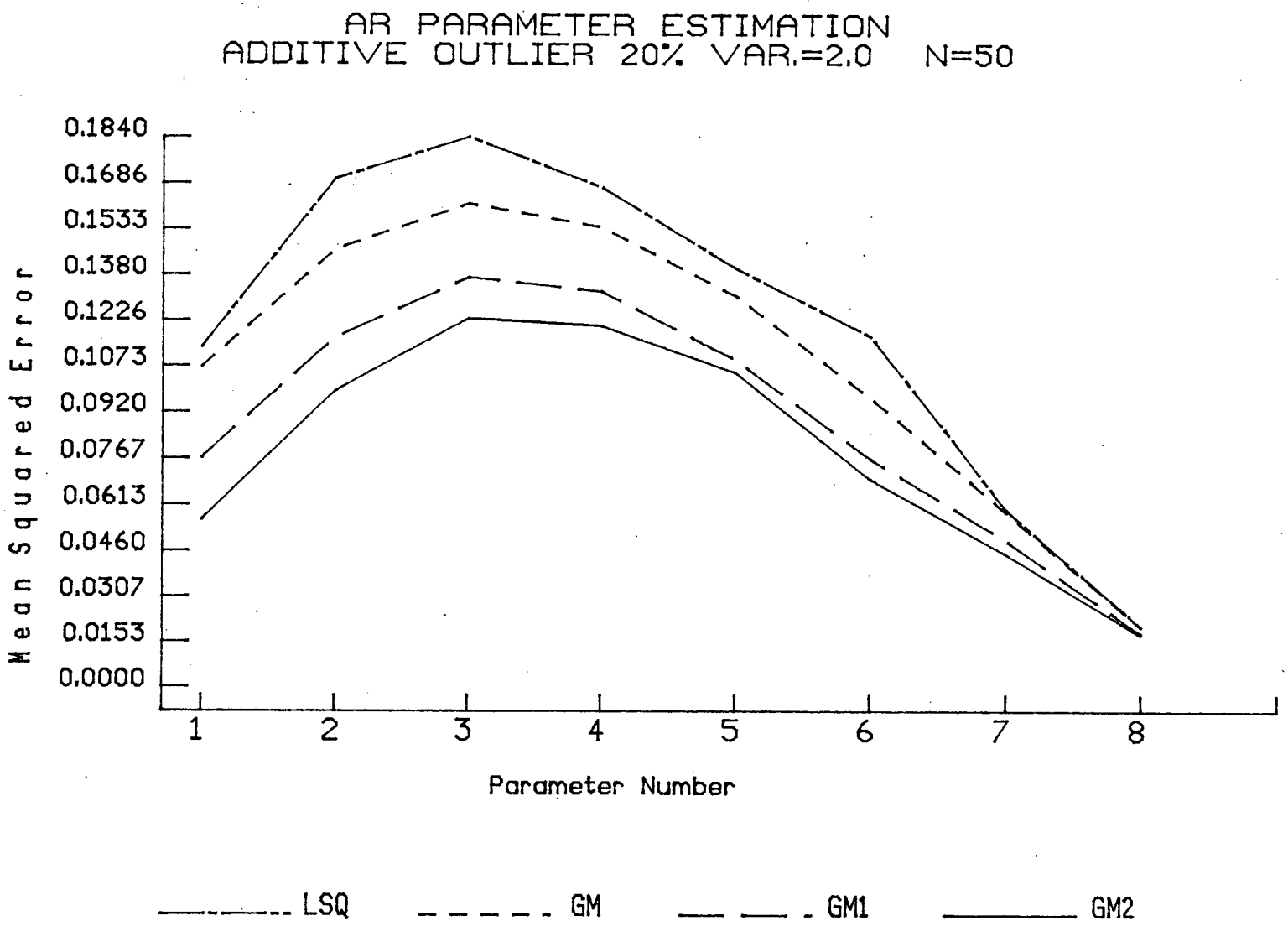

AR PARAMETER ESTIMATION
ADDITIVE OUTLIER 20% VAR.=2.0 N=50
Mean Squared Error
0.1840
0.1686
0.1533
0.1380
0.1226
0.1073
0.0920
0.0767
0.0613
0.0460
0.0307
0.0153
0.0000
1
2
3
4
5
6
7
8
Parameter Number
LSQ
GM
GM1
GM2

Figure 3.4

CHAPTER 4

OUTLIER PROCESSING OF SINGLE TRIAL EEG

4.1 Experimental Design and EEG Data Acquisition

The objective of the experimental data acquisition was to obtain EEG signals from subjects during a controlled voluntary skilled motor activity (active task) and during a controlled state in which the subjects were alert but not involved in any motor activity (idle task).

For the active task, subjects placed their right hand in an apparatus which oriented their hand in a standard position. The task required them to aim for a "target" position by performing a slow smooth (ramped) extension with their right thumb. During the movement the tip of their thumb pressed against a lever that provided a small opposing force to the thumb movement. In the starting position the lever rested on a support so that there was no initial load and the thumb was in a relaxed state. A potentiometer attached to the lever provided position information which was used to derive an encoded thumb movement signal. A sketch of the apparatus is given in Fig. 4.1a.

The duration of the ramped extension was approximately one second long. After subjects completed the extension and had returned their thumb to the starting position, visual feedback, via different colored lights, was presented to them indicating whether they had hit, overshot, or undershot the target position. During thumb movements subjects were asked to fixate on the visual feedback area in an attempt to minimize eye movements and to prevent subjects from looking at their thumb. After subjects were given some practice trials, they were asked to carry out fifty self-paced repetitions

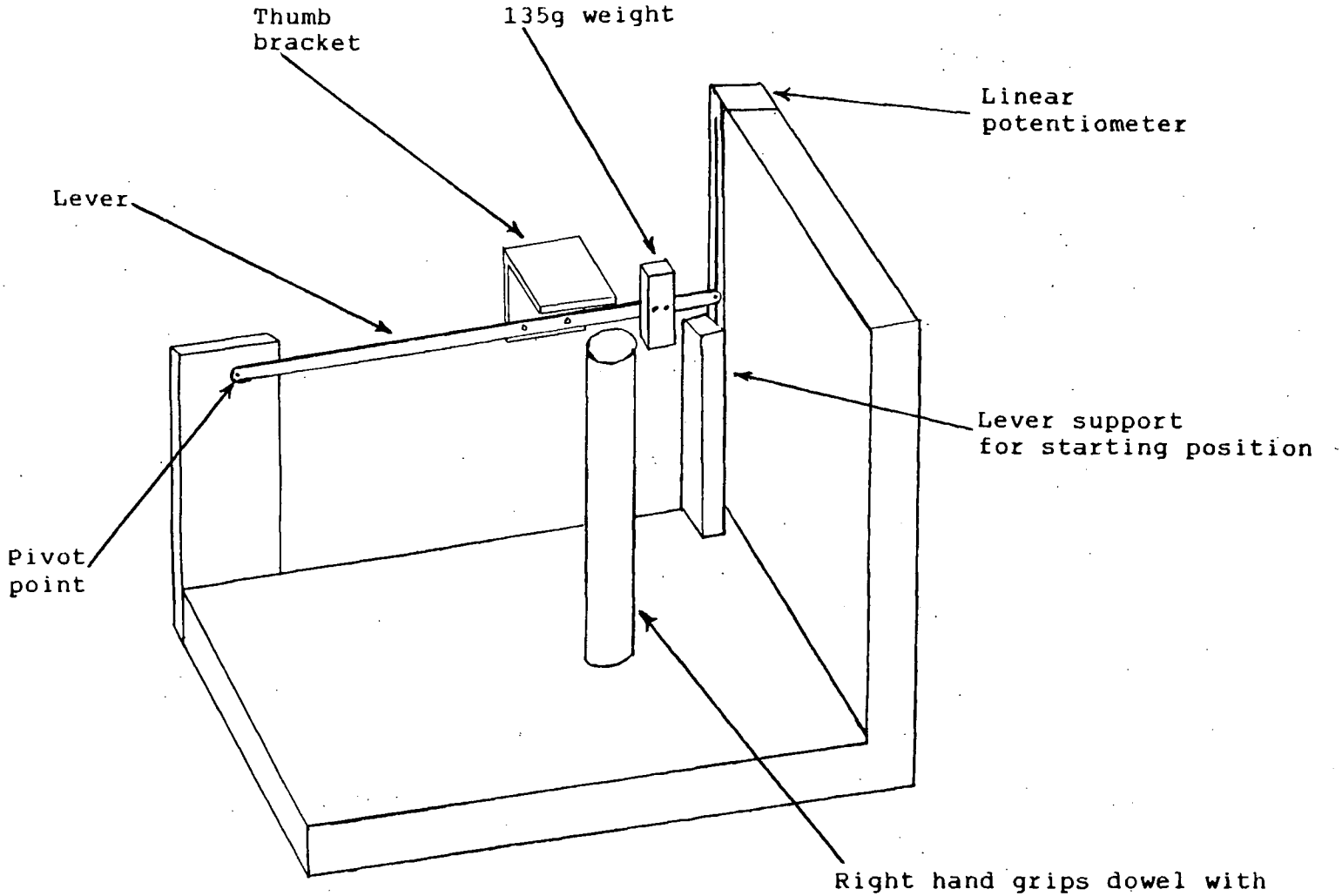
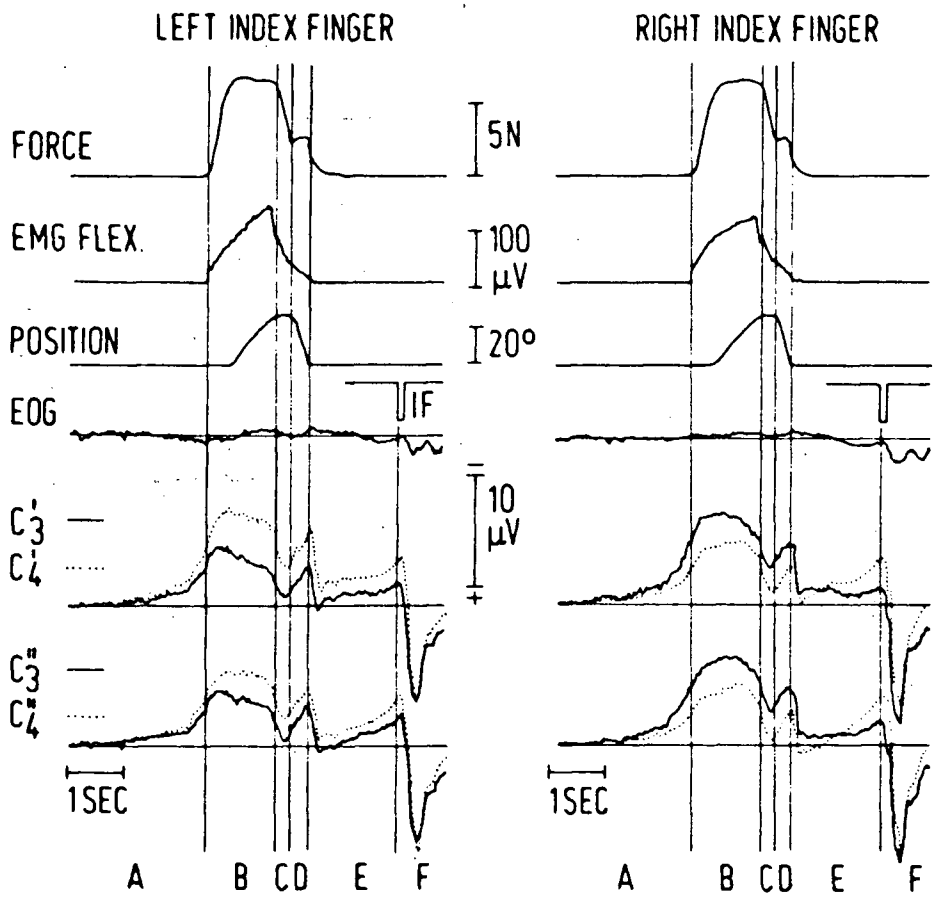

Thumb bracket
135g weight
Linear potentiometer
Lever
Lever support for starting position
Pivot point
Right hand grips dowel with thumb placed under the thumb bracket and perpendicular to the lever

Figure 4.1a

with their right thumb. On each trial, acquisition of the EEG started 1 second before thumb movement onset (determined by monitoring the encoded thumb movement signal) and then continued for 4.5 more seconds at which time the feedback was presented. The acquisition was then halted one second after presenting the feedback which resulted in a total epoch length of 6.5 seconds. The above active task was based on a previous study into motor potentials carried out by Grunewald and Grunewald-Zuberbier [7]. In their analysis they utilized conventional averaging techniques and their grand averages across seven subjects, 35 trials each, are given in Figure 4.1b.

For the idle task the subjects were kept in the same physical situation as in the active task but in this case they were not performing any thumb movements. Twenty epochs of EEG, 6.5 seconds in length, were collected from the subjects under this condition. After each epoch was collected, the feedback lights flashed to indicate to the subject that they had 10 seconds to relax before the onset of the next epoch. After taking a short pause the onus was on the subject to fixate their eyes on the feedback lights in preparation for the onset of the next epoch.

The EEG signals were recorded from the scalp using silver/silver chloride electrodes. This type of electrode possess the most appropriate characteristics for EEG recording in terms of low potential differences, long time constants and low resistance between the electrolyte and the metal surface of the electrode. The electrodes were "cupped" shaped with a small hole in the top and they were firmly attached to the scalp around the rim with the use of collodion. Electrolyte jelly was injected into the air space under the "cup" of the electrode which provided a good electrical connection between the scalp and the electrode. Typically the electrode impedance



Ramp positioning movements of the left and right index finger. Grand averages across seven right-handed subjects. Recordings of force, rectified EMG, position, vertical EOG, and slow potential shifts in left and right precentral (C3', C4') and postcentral (C3'', C4'') EEG (1cm anterior and 2 cm posterior to C3, C4 positions).

Figure 4.1b

between the scalp and the reference electrode was approximately 2500 ohms. The signal was initially recorded from three standard international 10/20 system electrode sites Cz, C3 and C4 (See Jasper [49]). The signal from each of these electrodes was referenced to linked ear lobes. A bi-polar EOG signal and the corresponding encoded thumb movements were also recorded. The EOG electrodes were placed on the supra orbital ridge and the external canthi of the right eye.

The EOG signal was utilized in a very conservative artifact rejection criterion which rejected any EEG epoch that had a corresponding EOG signal that at anytime during the epoch fluctuated above or below baseline by more than a given threshold value. This value was nominally set at 17 microvolts. As well, any EEG epoch that was not rejected by the above criterion but contained peak values that exceeded baseline by 43 microvolts, typically due to facial EMG, were also rejected as artifact-contaminated. In addition, for the active case, trials not containing a reasonable thumb movement were also rejected. Ultimately, based on the above selection criteria, 15 trials from both the active and idle cases were utilized from each of the four subjects. The EEG signal that was used for all the subsequent signal analysis was taken from the C3 electrode site which was contralateral to the thumb movement.

In all cases the EEG and EOG signals were initially amplified by a Beckman 711 polygraph using an analogue lowpass filter with a -3dB point at 100Hz (20dB per decade roll-off) and a highpass analogue filter with a time constant of 14.7 seconds. The signals for each epoch were digitized in real time at a rate of 1024 samples per second and were stored on a hard disk. Before any signal analysis schemes were applied, the EEG signals were preprocessed by a 201-point phaseless digital lowpass filter which had a cutoff frequency of 29Hz (-3dB point), a transition width of 3Hz (-24dB at

32Hz), and a minimum stopband attenuation of -27dB. It is generally agreed [50] that almost all the power in the normal EEG is between zero and thirty Hertz. Therefore, with the above digital filter the data was resampled at the relatively low rate of 64 Hz which is desirable because as the sample rate increases, there is a corresponding need to increase the AR model order since the time dependency is spread over a greater number of sample points. Generally, one wants to make the best possible trade-off between using a sufficiently high sample rate that will allow for the accurate representation of the highest frequency of interest and yet within that framework keep it as low as possible so that the required model order is minimized.

4.2 Neurological Premise

The concept that event related information is contained in EEG time series outliers is based on the following model of summation of electrical brain activity at a given point on the scalp. Under idle conditions the ongoing electrical activity that sums, spatially and temporally, at a given point on the scalp can be modeled as an overall ongoing process as "observed" from that point on the scalp during a particular time interval. When event related potentials, such as motor related potentials, are generated by a unique additional process they are "added" into the pre-existing ongoing process and would appear as additive outlier content when considered from the point of view of the ongoing pre-existing process. Therefore, if one could distinguish outlier points from pre-existing process points in the single trial active EEG time series then these outlier points could be used to provide information about event related potentials on a single trial basis.

The underlying p rinciple of the single trial processing scheme is to generate an AR model of the active EEG signal using a robust parameter estimation method that will represent the ongoing, underlying process by down-weighting unusual data points (see Section 3.3). This estimated model is then used in a robust signal estimator (see Section 4.3) which produces an estimated signal of the ongoing, underlying EEG process. The difference between the original measured signal and the estimated signal is considered to be additive outlier content (see Section 4.4.1). The outlier content is then processed (see Section 4.4.2) to produce waveform patterns that provide single trial event related information.

4.3 Signal Cleaning Process

An approach to detect outlier points in a time series was proposed by Martin and Thomson [45]. They used this system not to study the character, or information contained therein, of the outliers but rather to produce a "cleaned" time series which was used as part of a process that produced robust spectral estimates. This cleaning process is based on a "robustified" Kalman signal estimator. The objective of the cleaning process is to provide an estimate of the original signal without the AO content. It relies on a estimated pth order AR model of the process x_i as given in the AO model (3.37) which for convenience is repeated below

$$y_i = x_i + v_i \tag{4.1}$$

This AR process written in state variable form is given by

$$\underline{x}_i = \Phi \underline{x}_{i-1} + \underline{u}_i \tag{4.2}$$

where $\underline{x}_i^T=(x_i,x_{i-1},\ldots,x_{i-p+1})$, $\underline{u}_i^T=(e_i,0,0,\ldots,0)$, and

$$\Phi = \begin{bmatrix} a_1 & a_2 & \cdots & a_p \\ 1 & 0 & & 0 \\ 0 & 1 & & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots 1 & 0 \end{bmatrix} \qquad 4.3$$

Note that, given the above definition, the state $\underline{x}_i$ is equal to the current value of x_i and past values of x_i up to x_{i-p+1}.

"Robust" estimates of the state $\underline{x}_i$ are calculated recursively with the following expression [45]

$$\hat{\underline{x}}_i = \Phi \hat{\underline{x}}_{i-1} + \frac{\underline{m}_i}{s_i^2} s_i \psi \left(\frac{y_i - \hat{\underline{x}}^T_{i-1}\underline{a}}{s_i}\right) \qquad 4.4$$

where $\underline{m}_i$ is the first column of the pxp matrix M_i which is recursively calculated as follows

$$M_{i+1} = \Phi P_i \Phi^T + Q \qquad 4.5$$

and

$$P_i = M_i - w \left(\frac{y_i - \hat{\underline{x}}^T_{i-1}\underline{a}}{s_i}\right) \frac{\underline{m}_i \underline{m}_i^T}{s_i^2} \qquad 4.6$$

where Q is a pxp matrix with all zero entries except the first element which is equal to a robust estimate of the variance of the residual sequence, i.e. $Q(1,1)= S_e^2$ and s_i is a time varying scale defined by

$$s_i^2 = M_i(1,1) \qquad 4.7$$

The cleaned data at time i, will then be the first element of the estimated state $\underline{x}_i$, which is

$$\hat{y}_{ic} = \hat{x}_i \tag{4.8}$$

In other words, $\hat{y}_{ic}$ is an estimate of the process y_i without the influence of the additive outliers, v_i. Note that with the scaling given 4.7 and given that there is no influence from the psi function, as would be the desired result when there is no outlier content, $\hat{x}_i = y_i$.

Hampel's three part redescending psi function [43] was used in Equation 4.4 and it is given as follows:

$$\psi(t) = \text{sign}(t) \begin{cases} |t| & 0 \le |t| < a \\ a & a \le |t| < b \\ \frac{c-|t|}{c-b} & b \le |t| < c \\ 0 & c \le |t| \end{cases} \tag{4.9}$$

where a, b and c are tuning parameters. In a similar fashion as in the case of GM-estimation the w function is of the form

$$w(t) = \Psi(t)/t \tag{4.10}$$

The cleaning process described above was utilized as part of the procedure to obtain GM1 and GM2 AR parameter estimates. A GM1 estimate is based on the estimated cleaned time series and a GM2 estimate is a further iteration where the parameters from GM1 are used in the cleaner to provide a theoretically improved estimate of the cleaned time series. Figure 4.2 provides a block diagram of the procedure to obtain GM, GM1, and GM2 parameter estimates. Through the simulation studies described in Section 3.4 it was found that the best performance in terms of minimizing the MSE was obtained by setting the tuning parameters in Equation 4.9 as given below

ROBUST AR PARAMETER ESTIMATION

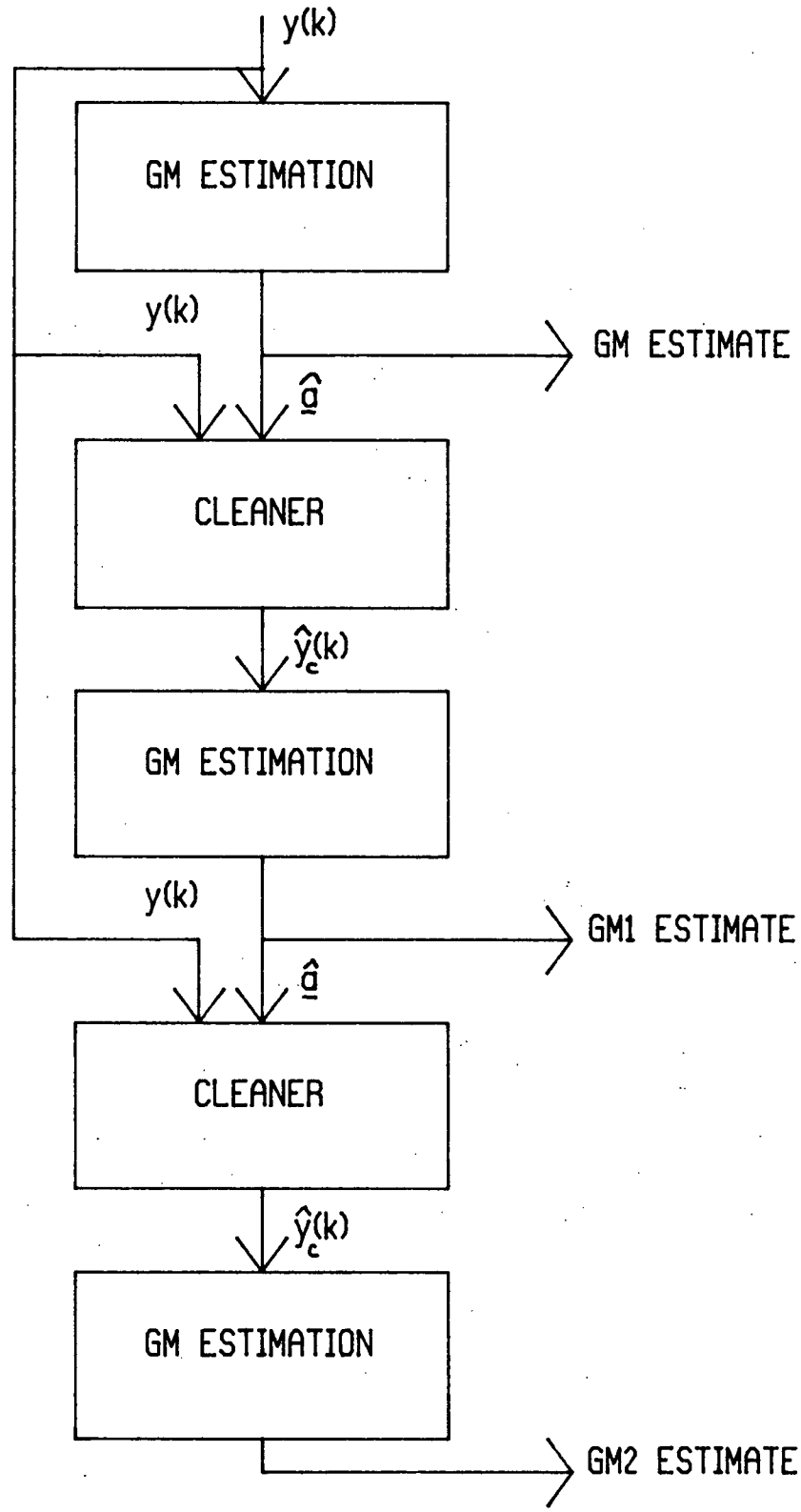


Figure 4.2

tuning parameter	
a	1.8
b	2.2
c	3.0

4.4 Extracting and Processing Outlier Information

4.4.1 Extracting Outlier Information

The outlier extraction process for the EEG data is accomplished by taking the epoch, 6.5 seconds long, and dividing it into 1.5 second segments with each segment overlapped by .75 seconds. Each segment is modeled at the order expected for idle task EEG (12-14) and hence, reducing the ability of the model to account for active task information in the EEG (see Section 4.5). The signal from each segment is then cleaned using the estimated model parameters in the cleaner described in Section 4.3. The outliers are then calculated by taking the difference between the original and cleaned signals (see Figure 4.3).

The outlier extraction process was initially tested by applying it to Gaussian simulated 12th order AR data which contained 10% "patchy" (correlated) additive outlier contamination. This test confirmed that the extraction process had some distinct ability to recover the outlier content from the simulated signal. Patchy contamination was used since it was expected that in the case of real EEG the additive event related potentials would be correlated. As suggested by Martin and Zeh [51], the correlated values for v_i were generated by dividing the segment into equal halves.

PROCESS TO EXTRACT OUTLIER POINTS

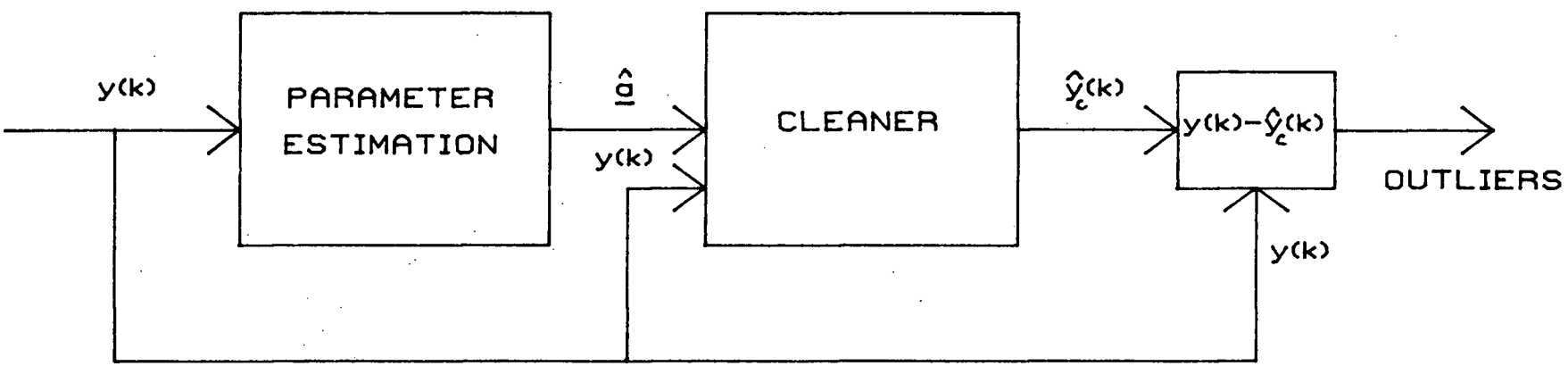


Figure 4.3

Immediately following the first non-zero v_i in each half the rest of the non-zero v_i's were grouped together. These grouped v_i's were used to produce correlated v_i^c's via the following expression

$$v_i^c = \Theta v_{i-1}^c + (1 - \Theta^2)^{1/2} v_i \qquad 4.11$$

where a value of $\Theta = 0.6$ was used in these simulation tests. This procedure results in a correlated outlier series which has roughly the same variance as in the independent case [51].

It was found in this application that the tuning parameters for the psi function given in Equation 4.9 needed to be set such that it provided a stronger influence than in the parameter estimation application discussed in Section 4.3. By trial and error the best performance of the extraction process was obtained by setting the tuning parameters as given below

tuning parameter	
a	1.0
b	1.2
c	1.8

Some example results from these tests, which qualitatively demonstrate the potential performance of the extraction process, are shown in Figure 4.4. Each example contains the same randomly generated v_i^c contamination which was used to contaminate different Gaussian AR sequences x_i. The broken line in each plot represents the actual v_i^c values and the solid line represents the outlier content that was extracted via the the outlier extraction process. Figures 4.4a through 4.4c show three different examples of using GM2 and LSQ parameters in the cleaning process. Figures 4.4d and 4.4e are the second and

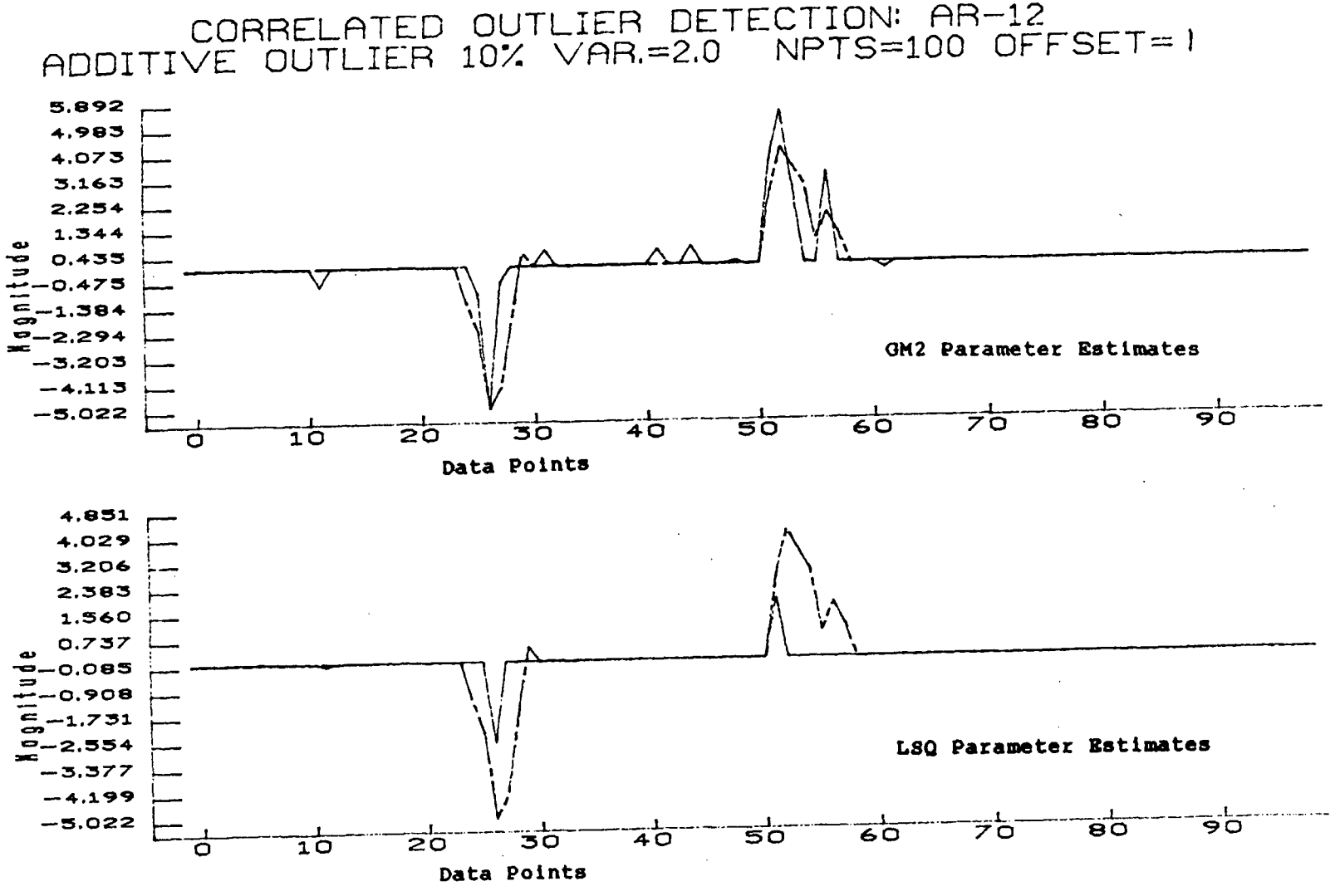

CORRELATED OUTLIER DETECTION: AR-12
ADDITIVE OUTLIER 10% VAR.=2.0 NPTS=100 OFFSET=1
Magnitude
5.892
4.983
4.073
3.163
2.254
1.344
0.435
-0.475
-1.384
-2.294
-3.203
-4.113
-5.022
0
10
20
30
40
50
60
70
80
90
Data Points
GM2 Parameter Estimates
Magnitude
4.851
4.029
3.206
2.383
1.560
0.737
-0.085
-0.908
-1.731
-2.554
-3.377
-4.199
-5.022
0
10
20
30
40
50
60
70
80
90
Data Points
LSQ Parameter Estimates
Extracted Outliers
Actual Outliers

Figure 4.4a

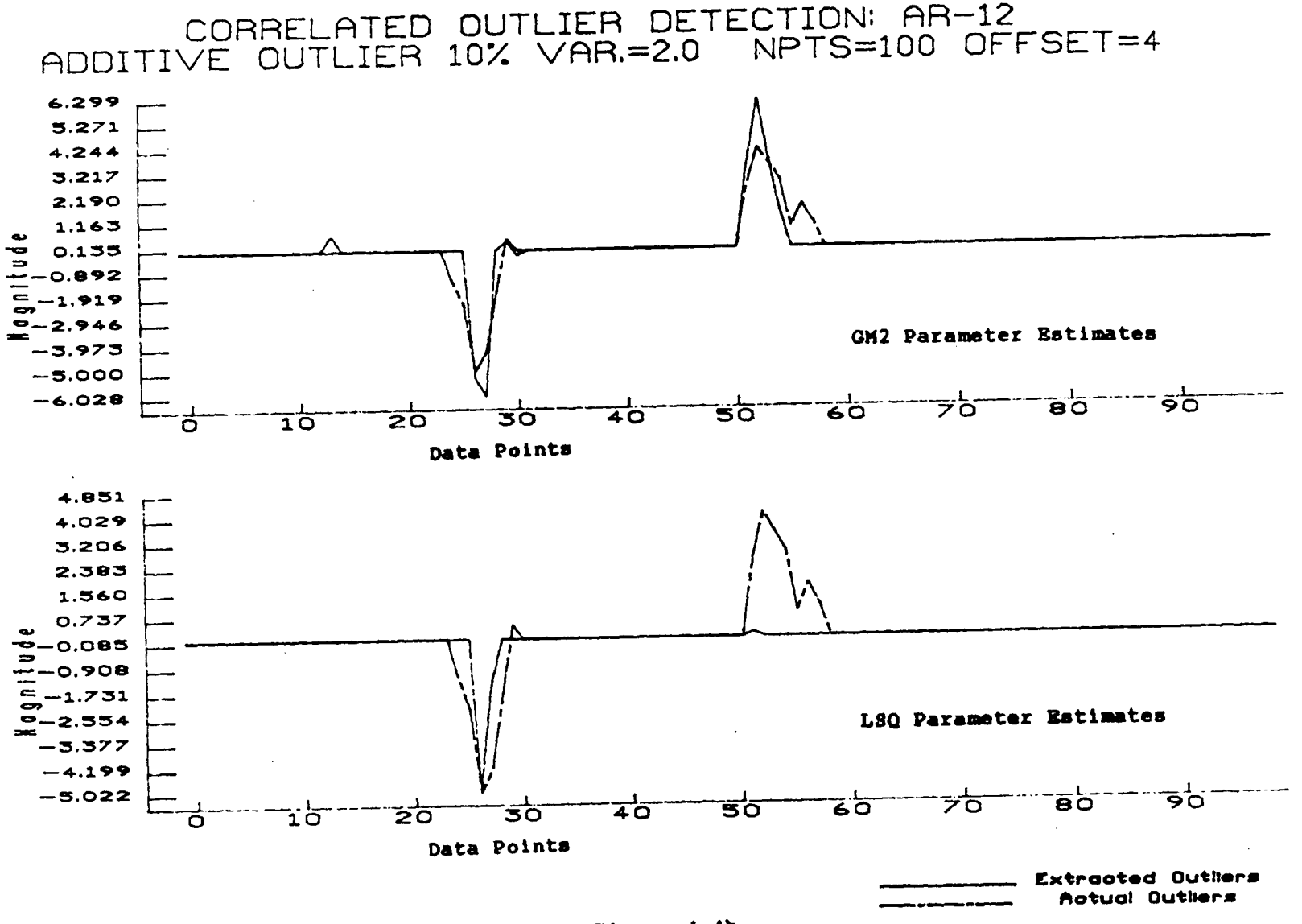

CORRELATED OUTLIER DETECTION: AR-12
ADDITIVE OUTLIER 10% VAR.=2.0 NPTS=100 OFFSET=4
Magnitude
6.299
5.271
4.244
3.217
2.190
1.163
0.135
-0.892
-1.919
-2.946
-3.973
-5.000
-6.028
0
10
20
30
40
50
60
70
80
90
Data Points
GM2 Parameter Estimates
Magnitude
4.851
4.029
3.206
2.383
1.560
0.737
-0.085
-0.908
-1.731
-2.554
-3.377
-4.199
-5.022
0
10
20
30
40
50
60
70
80
90
Data Points
L8Q Parameter Estimates
Extracted Outliers
Actual Outliers

Figure 4.4b

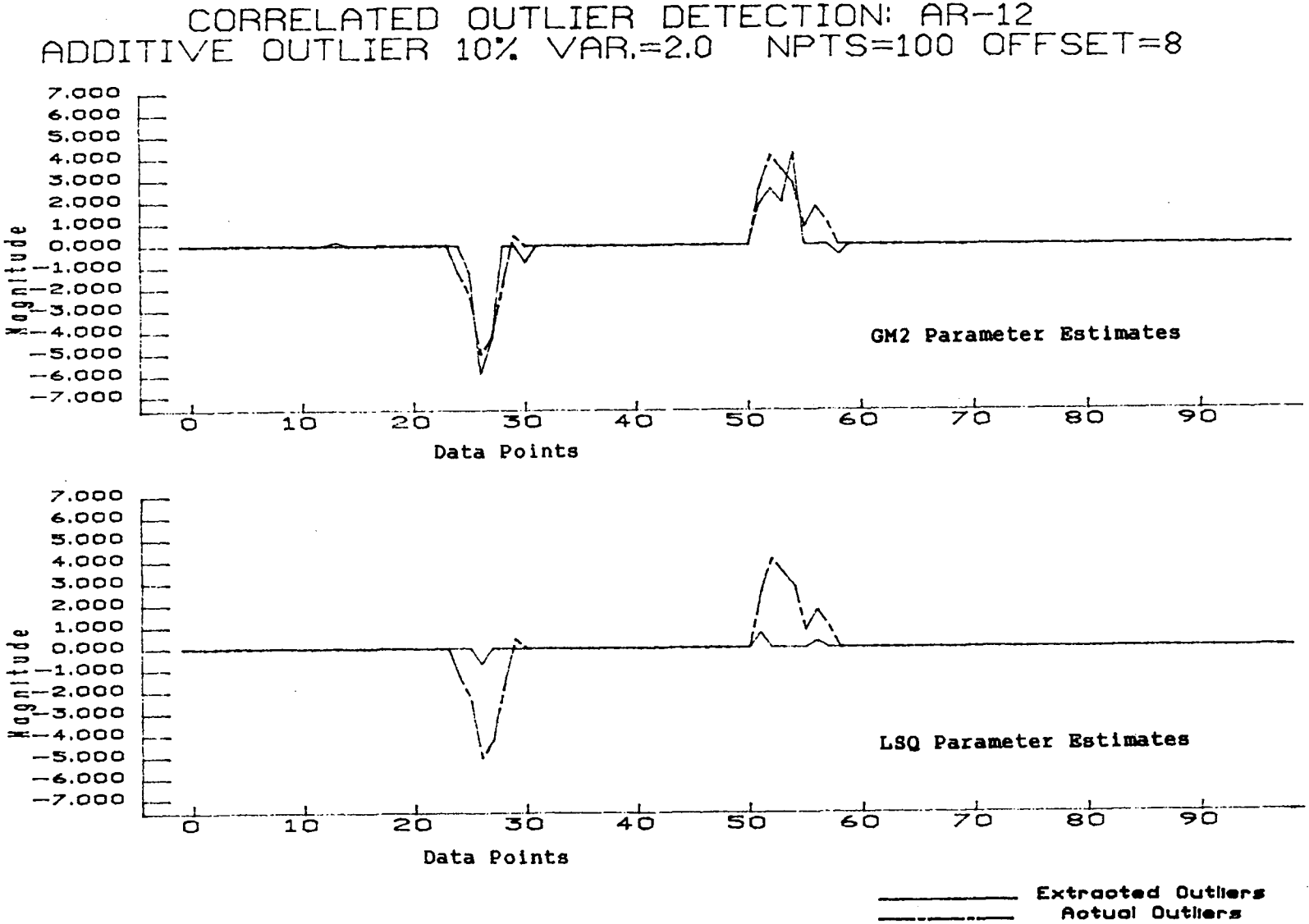

CORRELATED OUTLIER DETECTION: AR-12
ADDITIVE OUTLIER 10% VAR.=2.0 NPTS=100 OFFSET=8
Magnitude
7.000
6.000
5.000
4.000
3.000
2.000
1.000
0.000
-1.000
-2.000
-3.000
-4.000
-5.000
-6.000
-7.000
0
10
20
30
40
50
60
70
80
90
Data Points
GM2 Parameter Estimates
LSQ Parameter Estimates
Extracted Outliers
Actual Outliers

Figure 4.4c

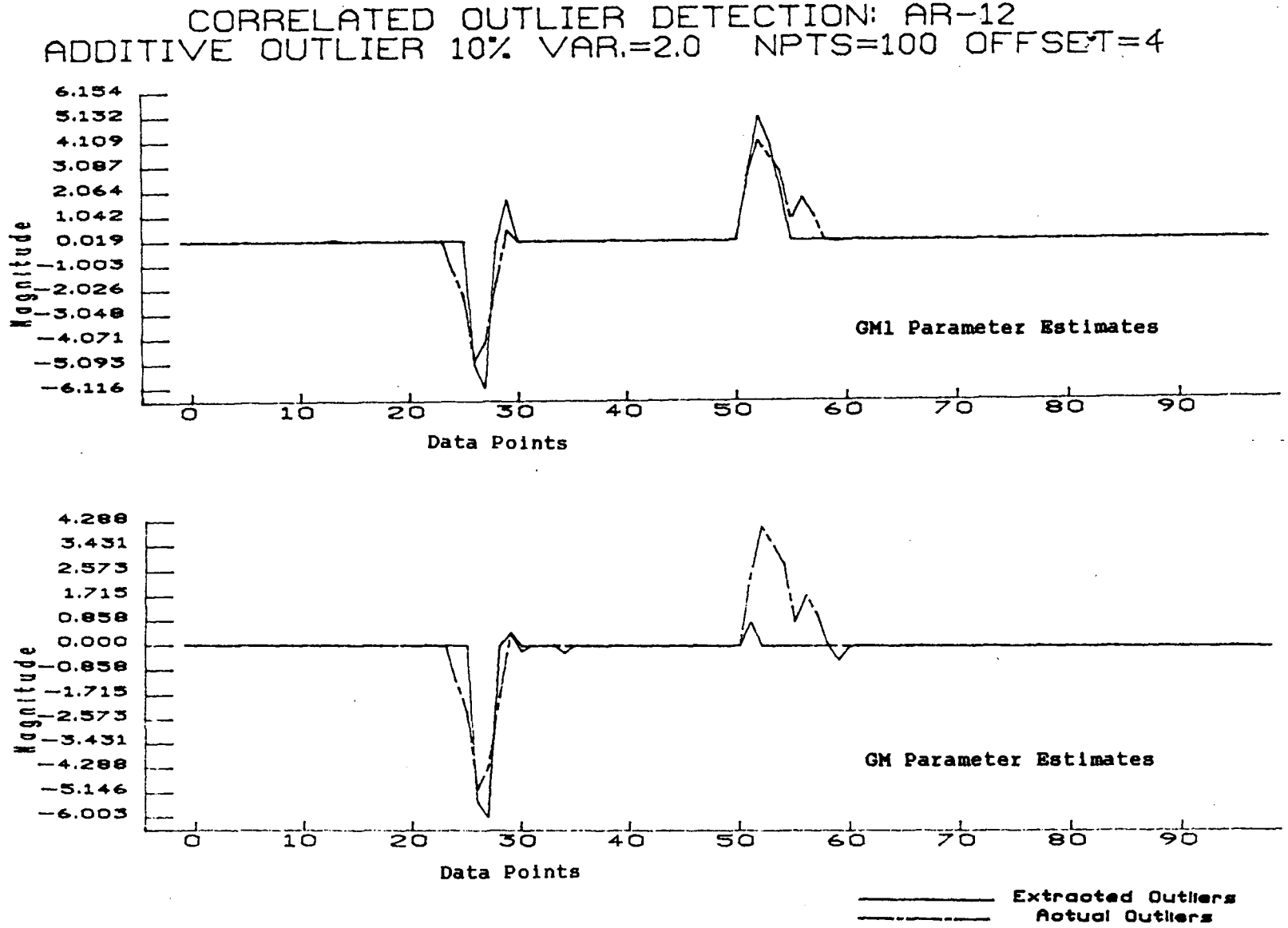

CORRELATED OUTLIER DETECTION: AR-12
ADDITIVE OUTLIER 10% VAR.=2.0 NPTS=100 OFFSET=4
6.154
5.132
4.109
3.087
2.064
1.042
0.019
-1.003
-2.026
-3.048
-4.071
-5.093
-6.116
Magnitude
0
10
20
30
40
50
60
70
80
90
Data Points
GM1 Parameter Estimates
4.288
3.431
2.573
1.715
0.858
0.000
-0.858
-1.715
-2.573
-3.431
-4.288
-5.146
-6.003
Magnitude
0
10
20
30
40
50
60
70
80
90
Data Points
GM Parameter Estimates
Extracted Outliers
Actual Outliers

Figure 4.4d

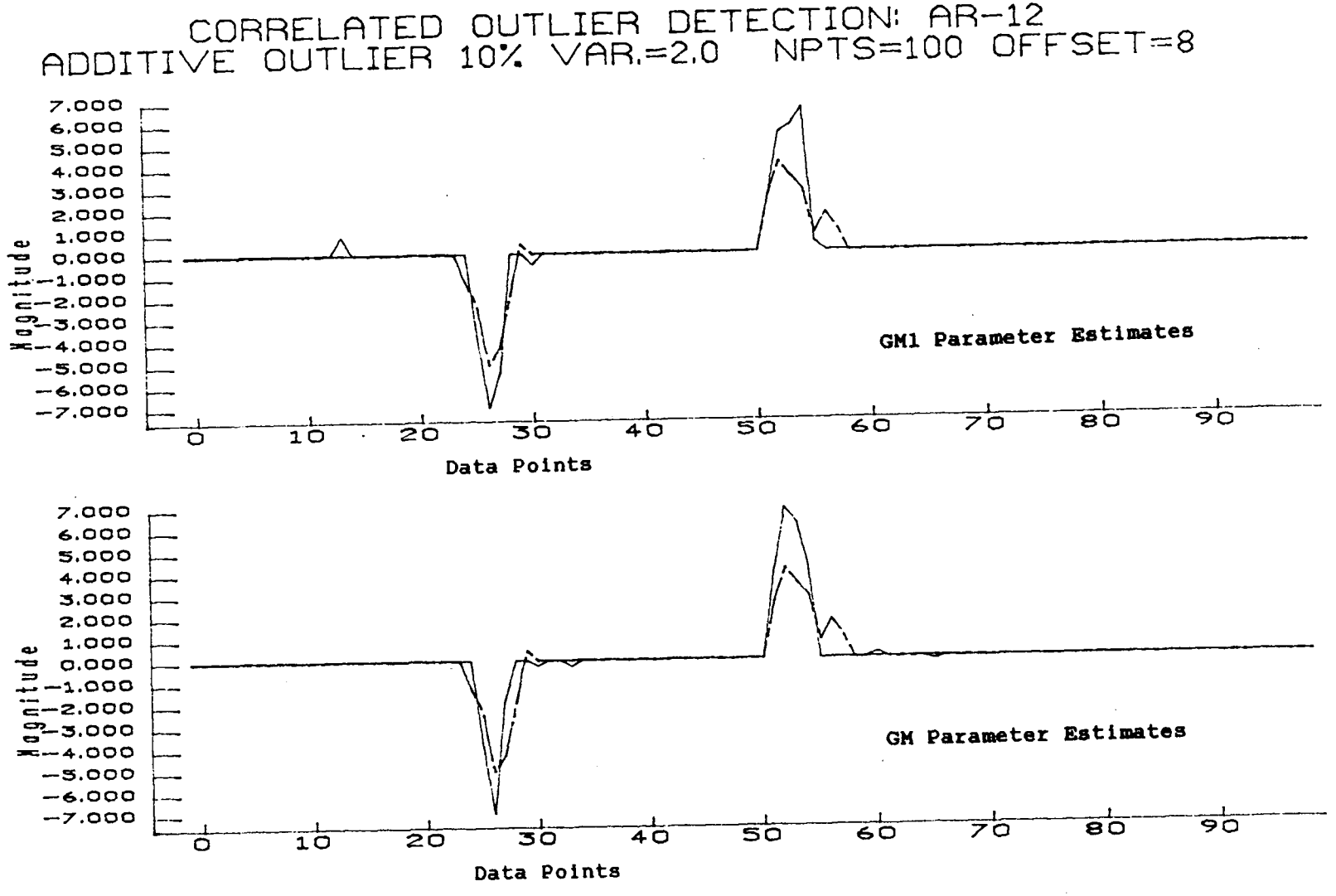

CORRELATED OUTLIER DETECTION: AR-12
ADDITIVE OUTLIER 10% VAR.=2.0 NPTS=100 OFFSET=8
Magnitude
7.000
6.000
5.000
4.000
3.000
2.000
1.000
0.000
-1.000
-2.000
-3.000
-4.000
-5.000
-6.000
-7.000
0
10
20
30
40
50
60
70
80
90
Data Points
GM1 Parameter Estimates
GM Parameter Estimates
Extracted Outliers
Actual Outliers

Figure 4.4e

third examples repeated using GM and GM1 parameter estimation. It is clear from these examples that the process performs better with GM2 estimates than with GM estimates and much better than with LSQ estimates. The performance of GM2 and GM1 are quite similar with perhaps some subtle improvements in the case of GM2. Since these tests revealed three clearly discernable jumps in performance in using LSQ, GM, and GM2 parameter estimates, it was decided that subsequent studies using outlier detection in this thesis work would be restricted to those three estimation methods.

4.4.2 Processing Outlier Information

Processing the extracted outlier information is accomplished by taking the outlier content from each EEG segment and averaging it together with the outlier content from the corresponding overlapping segment. This results in an outlier pattern spanning the whole 6.5 second epoch. The outlier pattern is then smoothed by convolving it with a 16 point tapered smoothing window which is based on a minimum-bias spectral window suggested by Papoulis [52]. It is given by

$$W(k) = 16\pi^2 \frac{1 + \cos(2\pi\, k/16)}{[16(2\pi\, k/16)^2 - \pi^2]^2} \qquad 4.12$$

where $k = 0, +/-1, +/-2, \ldots, +/-16$

The resulting smoothed pattern constitutes the output waveform of the single trial processing method.

4.5 AR Spectral Analysis

Preliminary studies involving AR spectral analysis were useful in providing some measure of the ability of the AR model to represent the EEG signal. As well, these studies were instrumental in establishing appropriate EEG segment lengths and a procedure for the selection of the AR model order.

It can be shown [23] that the AR spectral estimate is

$$S(f) = \frac{S_e(f)}{\left|1 - \sum_{k=1}^{p} a_k \exp-\left(\frac{j2\pi kf}{f_s}\right)\right|^2} \qquad 4.13$$

where $S_e(f)$ is the power spectrum of the residual sequence e_i and f_s is the sample frequency. Since the term $S_e(f)$ applies to the residuals which are in theory white, the resulting power density function of the residuals should be flat and therefore $S_e(f)$ will be a constant independent of frequency. Ideally, the value of this constant (noting that the mean of the residuals is zero) will be proportional to the variance of the residuals [46]. Hence, the final expression for the conventional AR spectral estimate is obtained by replacing $S_e(f)$ in 4.13 with s_e^2/f_s, where s_e^2 is an estimate of the variance of the residuals and the $1/f_s$ term is included in the numerator so that the true power of the corresponding analogue signal will be represented [23].

The EEG signal characteristics from subjects, particularly during highly active mental states, are changing relatively quickly. Single trial AR spectral estimates from adjacent one second segments demonstrated that considerable change in signal characteristics could occur over this span of two seconds. An example of this is provided in Figure 4.5. It contains four consecutive AR spectral plots, each derived from a one second segment of

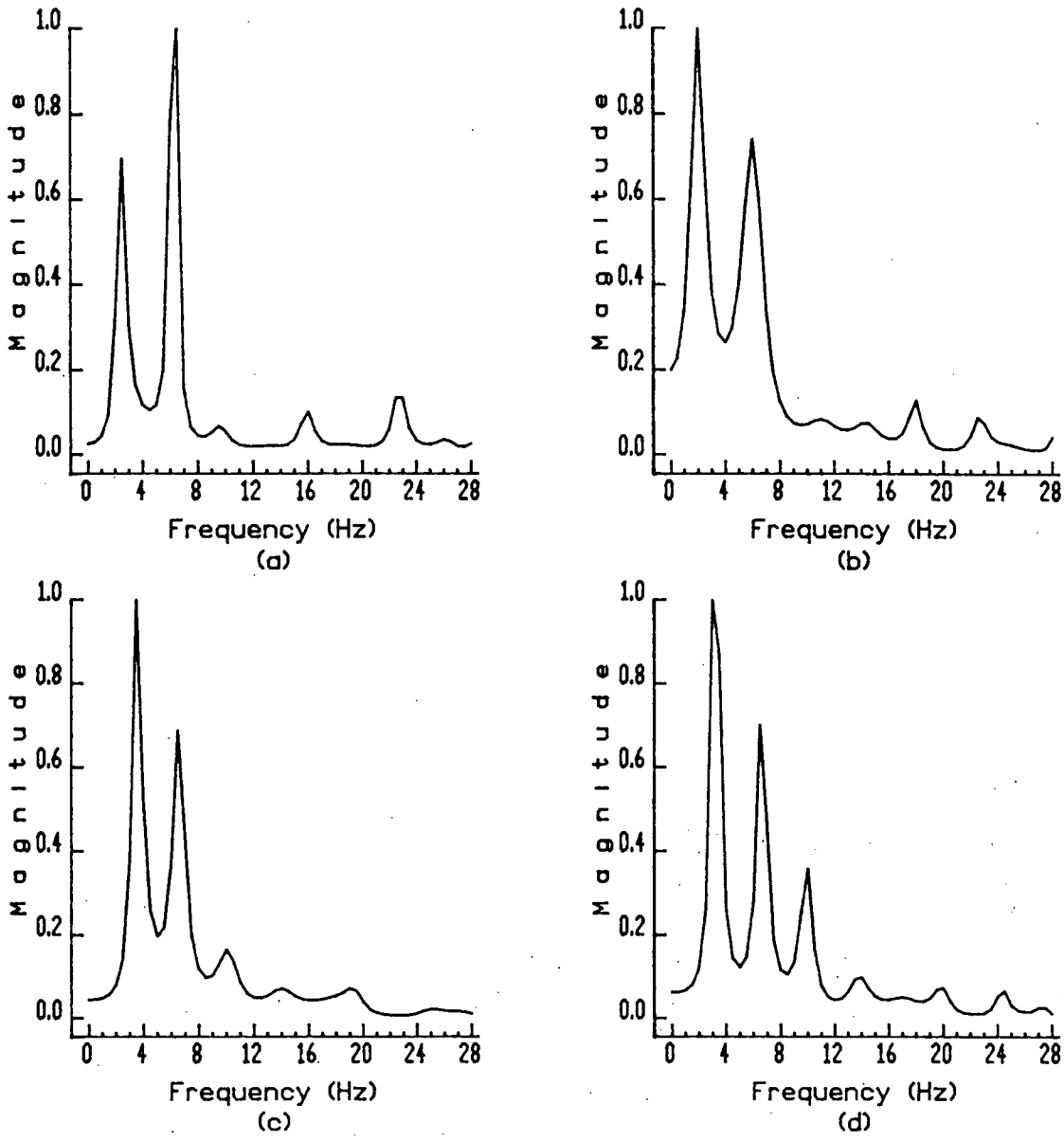


Figure 4.5 AR spectral estimates of one second segments of active trial EEG consecutively offset by a third of a second.

active EEG and each offset by 0.333 seconds (total span of two seconds). As was discussed in Section 3.2, it was found that a practical lower bound on segment length, from a parameter estimation point of view, was approximately one second. These findings ultimately lead to the utilization, as noted in Section 4.4.1, of a 1.5 second segment length with an offset of 0.75 seconds in the single trial processing method. This was an attempt to trade off the need for short segments because of the relatively rapid changing signal characteristics with the desire to raise the segment length above the lower bound for purposes of improving the parameter estimation efficacy.

It was found that selecting the model order via conventional methods such as Akaike's Information Criteria (AIC) does not work well with these short segments [53]. Conclusions were similar to Jansen [32] in that the selection of an appropriate model order requires some trial and error and, if possible, some a-priori knowledge of expected results. It was found useful to try a number of orders within a reasonable range (for a sample rate of 64Hz, somewhere between 8 to 25), following the trend of the estimate as the model order was increased. Features were identified that seemed reasonable based on both the a-priori knowledge of the condition under which the EEG was collected and a conventional FFT based estimate. The order was sequentially increased, expecting the features to become better defined, until spurious peaks began to occur. The appropriate model order was then selected to be two or three below that value. Typically, model orders were selected in the range of 12 to 14 from subjects during the idle task and in the range of 18 to 22 from subjects during the active task. Figures 4.6 and 4.7 provide example AR spectral plots to demonstrate this model order selection procedure for the idle and active task respectively.

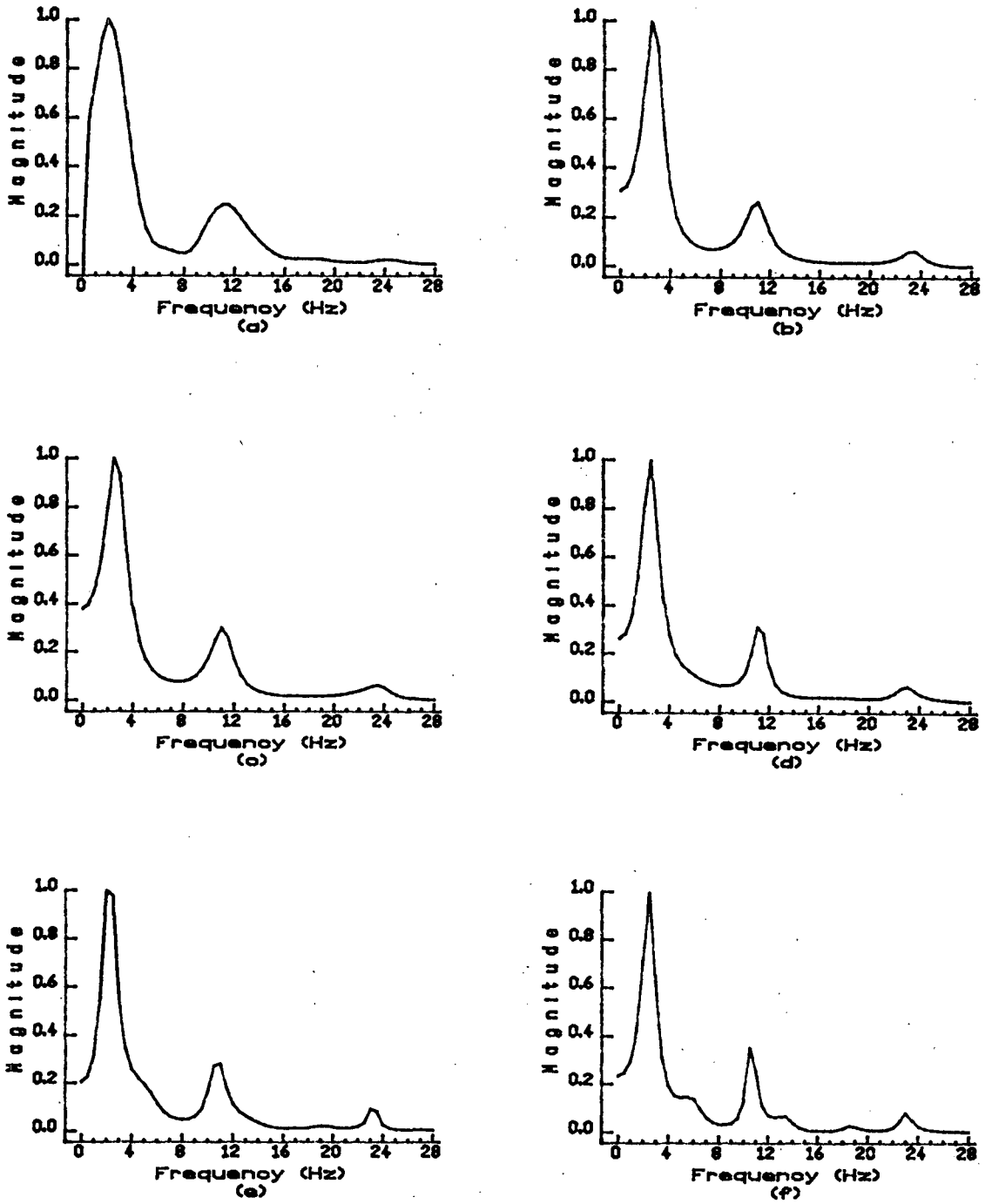


Figure 4.6 Progression of AR spectral estimates with increasing order using EEG data from an example idle trial.
a) conventional FFT b) model order 8 c) model order 10
d) model order 12 e) model order 14 f) model order 16

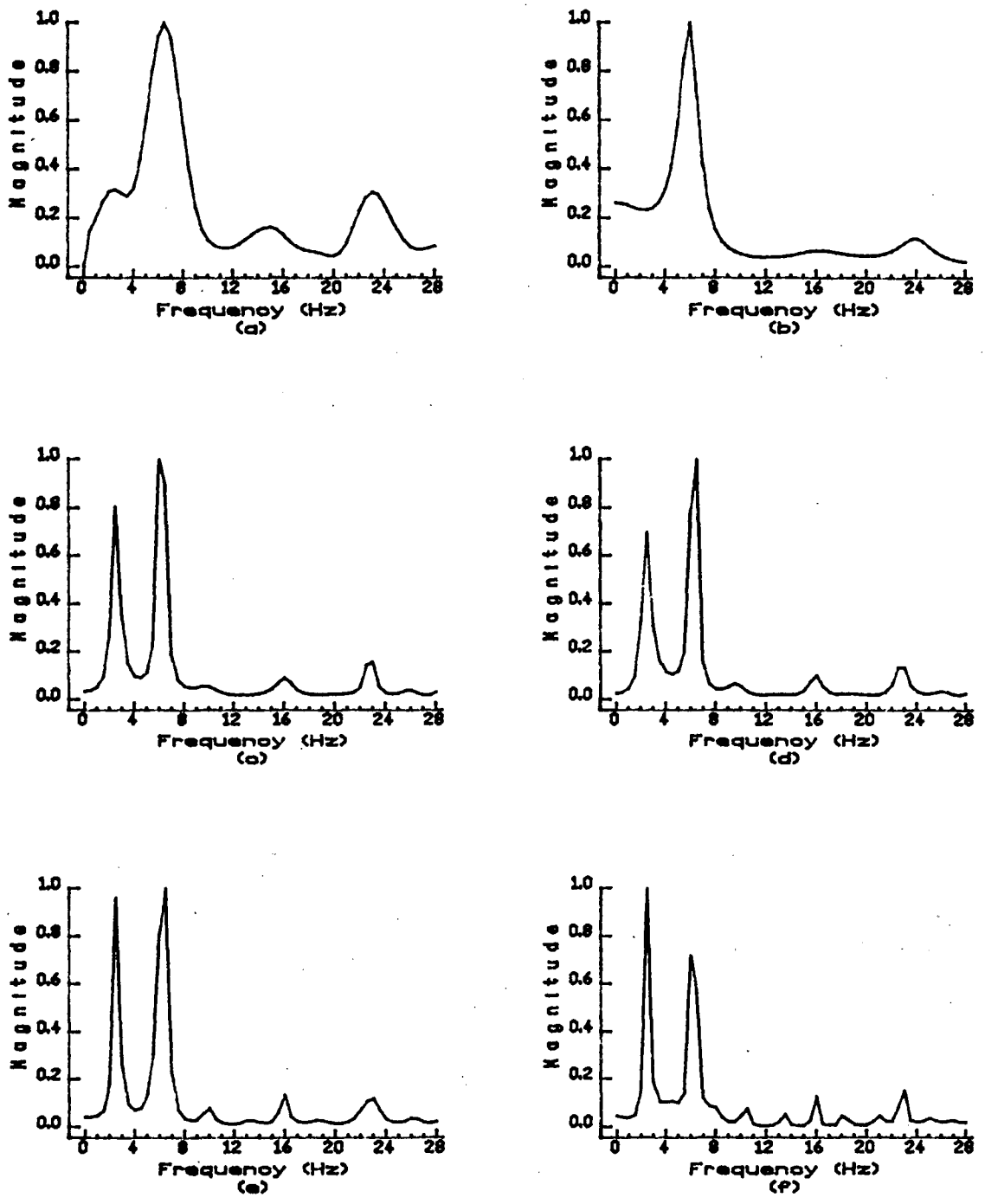


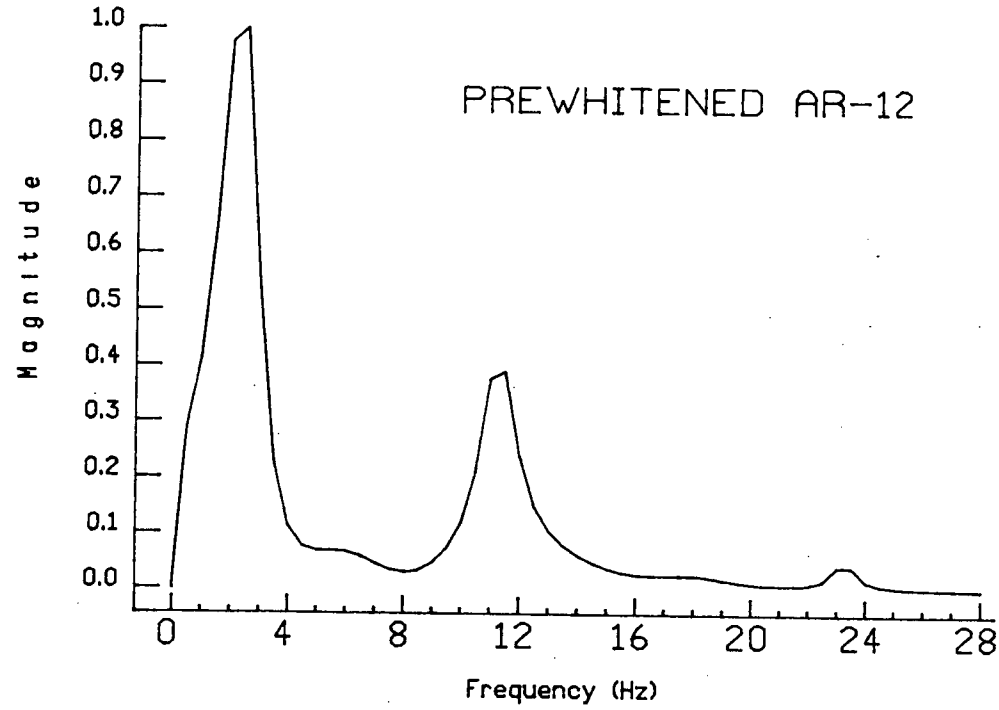
Figure 4.7 Progression of AR spectral estimates with increasing order using EEG data from an example active task.
a) conventional FFT b) model order 8 c) model order 10
d) model order 12 e) model order 14 f) model order 16

In the derivation of the autoregressive spectral estimate an alternative to assuming that the residuals will be perfectly white in the calculation of $S_e(f)$ would be to estimate that quantity with a conventional FFT based estimate. The residual signal can be thought of as a whitened signal because the information that can be represented by an AR model has been subtracted resulting in a signal with a much flatter spectrum. When the FFT is applied to this prewhitened signal the inherent drawback of leakage is greatly reduced. Application of conventional leakage control, such as Blackman windowing, serves to further reduce this problem. The prewhitened AR estimation method, therefore, combines the spectral information from both the AR model and the residual FFT spectral estimate. It is given by [45]

$$S(f) = \frac{S_{Ne}(f)}{\left| 1 - \sum_{k=1}^{p} a_k \exp - \left(\frac{j2\pi kf}{f_s}\right) \right|^2} \qquad 4.14$$

where $S_{Ne}(f)$ is a spectral estimate of the residual sequence e_i using a conventional FFT method. Some insight into the ability of the AR model to represent short segments of EEG was gained by pursuing studies using prewhitened AR spectral estimates. These studies demonstrated that when an appropriate model order was utilized the conventional AR spectral estimates were reasonably good compared to the prewhitened AR estimate which makes use of information retained in the residuals (see Birch et al. [53]). This indicates that the AR model, although not perfect, does represent much of the information contained in a short segment of EEG. An example of both a conventional and a prewhitened 12th order AR spectral estimate of idle task EEG is given in Figure 4.8.

AR SPECTRAL ESTIMATION OF IDLE TASK EEG



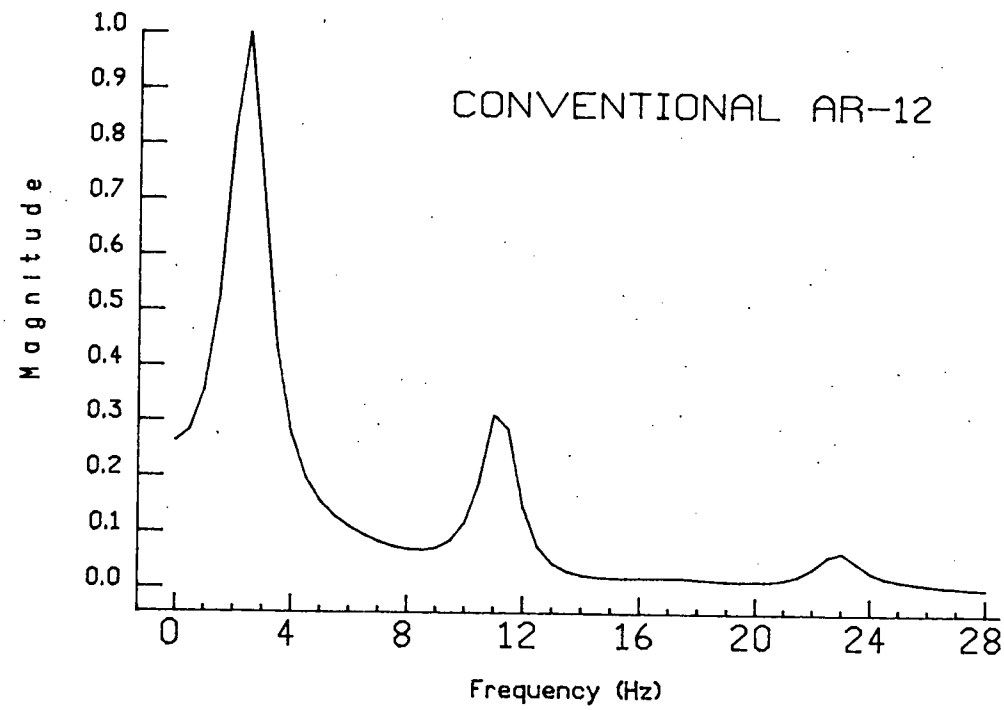


Figure 4.8

4.6 Applying Outlier Processing to Single Trial EEG

4.6.1 Comparison of Segmented Cleaned Active, Original Active and Original Idle Signals

It would be predicted, given the above neurological premise, that the original idle signal and the cleaned active signal should have little or no evidence of motor potential activity whereas the original active signal should contain evidence of motor potential activity. Figure 4.9 provides two sets of plots: one with N=6 trials the other with N=15 trials. Each set contains plots of conventionally averaged cleaned active, original active, and original idle signals. Motor potential activity in the active case should occur, approximately, during the first three seconds of the epoch, noting that the actual thumb movement began one second into the epoch. These plots demonstrate that the conventional averaging technique reveals some distinct motor activity in the original active case (raised level of positivity in the averaged signal during the first three seconds with a peak at about two seconds). However, in the cleaned active and original idle cases the averaging does not reveal any distinct motor activity. Hence, the above prediction is substantially borne out. The strong negative peak in the N=6 active task plot at about 6 seconds is the visual evoked response to the feedback light and is not due to motor potential activity.

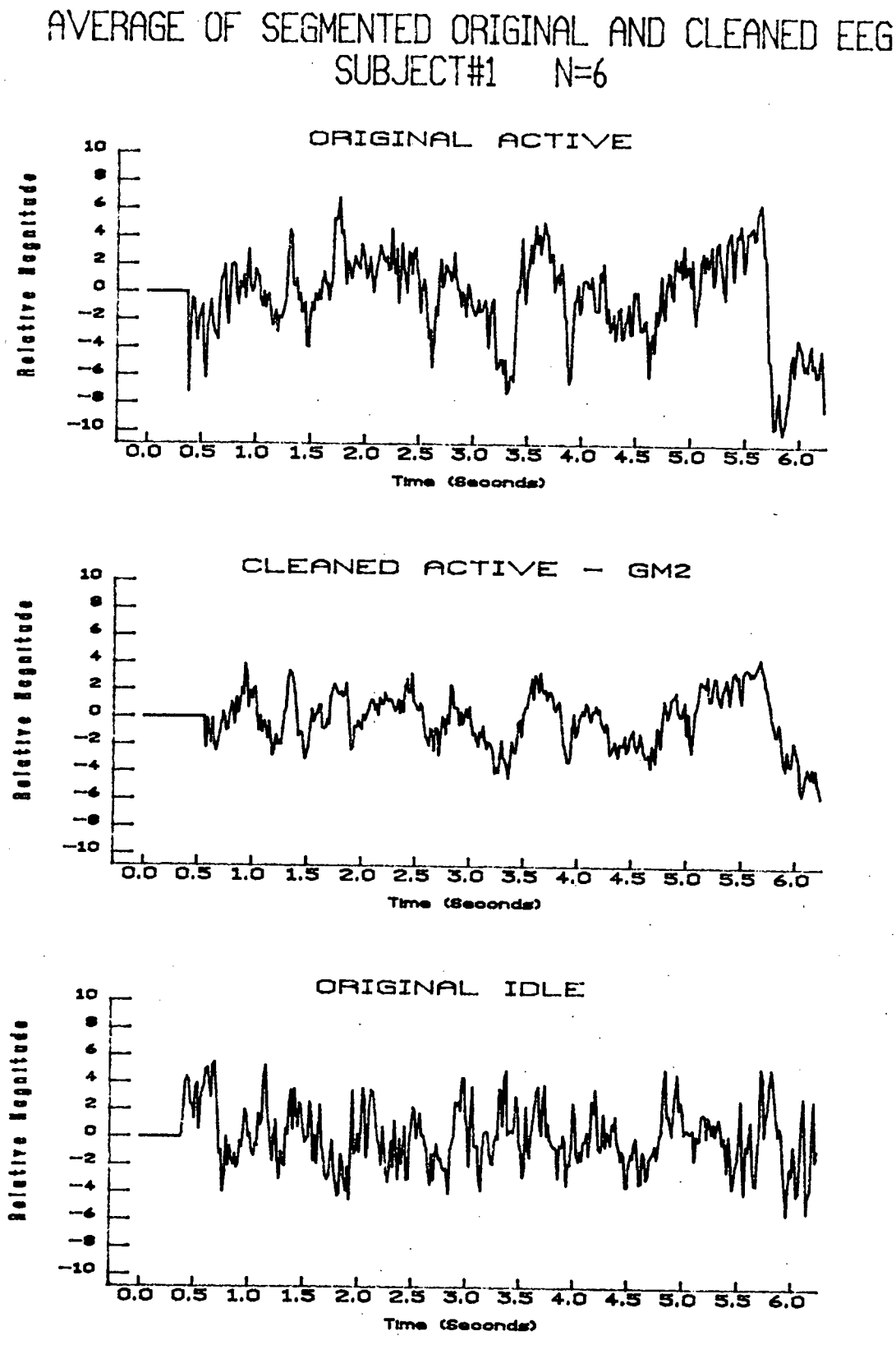

AVERAGE OF SEGMENTED ORIGINAL AND CLEANED EEG
SUBJECT#1 N=6
ORIGINAL ACTIVE
CLEANED ACTIVE - GM2
ORIGINAL IDLE
Relative Magnitude
10
8
6
4
2
0
-2
-4
-6
-8
-10
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0
Time (Seconds)

Figure 4.9a

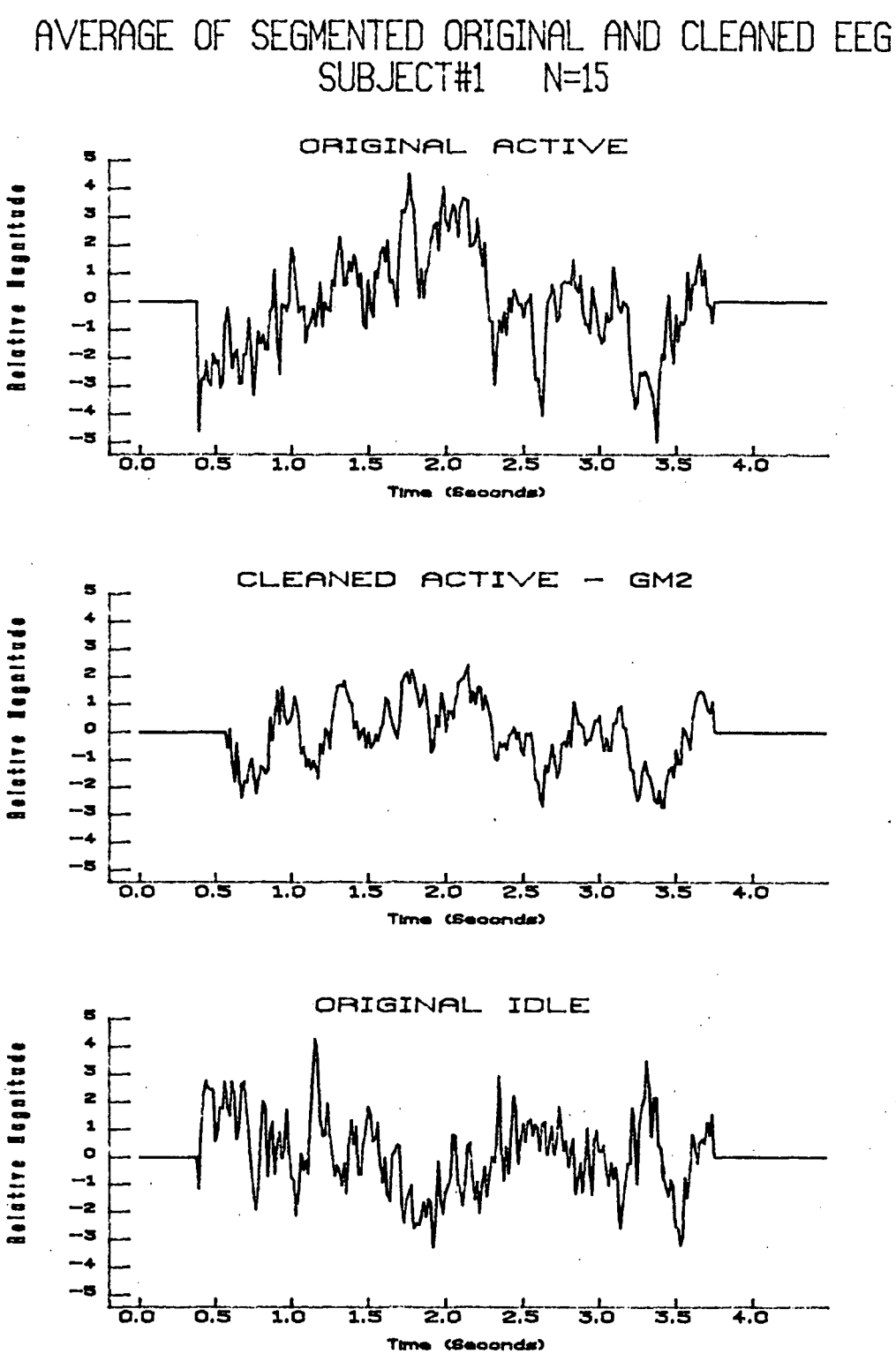

AVERAGE OF SEGMENTED ORIGINAL AND CLEANED EEG
SUBJECT#1 N=15
ORIGINAL ACTIVE
CLEANED ACTIVE - GM2
ORIGINAL IDLE
Relative Magnitude
5
4
3
2
1
0
-1
-2
-3
-4
-5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Time (Seconds)

Figure 4.9b

4.6.2 Examples of Single Trial Outlier Patterns and Single Trial Raw EEG

To qualitatively demonstrate the results of the single trial processing method, four example plots of the single trial outlier patterns using GM2 model parameters paired with the corresponding raw EEG are provided in Figure 4.10. Note the significant amount of information that is in the outlier patterns which can not be easily seen in the raw EEG signals. Results provided in the following sections demonstrate that the information in these outlier patterns is related to the thumb movements. However, at this point it is interesting to note the many similarities of these single trial patterns with the grand average waveforms from the Grunewald study cited in Section 4.1.

4.6.3 Comparison of Averaged Active Outlier Patterns, Averaged Idle Outlier Patterns and the Conventional Average of Active EEG

To demonstrate that there is some strong consistency in the active case outlier patterns and very little consistency in the idle case outlier patterns, the plots in Figure 4.11 have been provided. These plots contain the averaged outlier patterns for N=6 and N=15 using GM2 parameter estimates. As well, for comparison purposes, a plot of the conventional average for the active case is also included in this figure.

The fact that the average active case patterns maintain a general shape similar to the single trial patterns, strongly indicates that there is information related to the thumb movement that is consistent from trial to trial. On the other hand, the fact that the average reduces in magnitude and

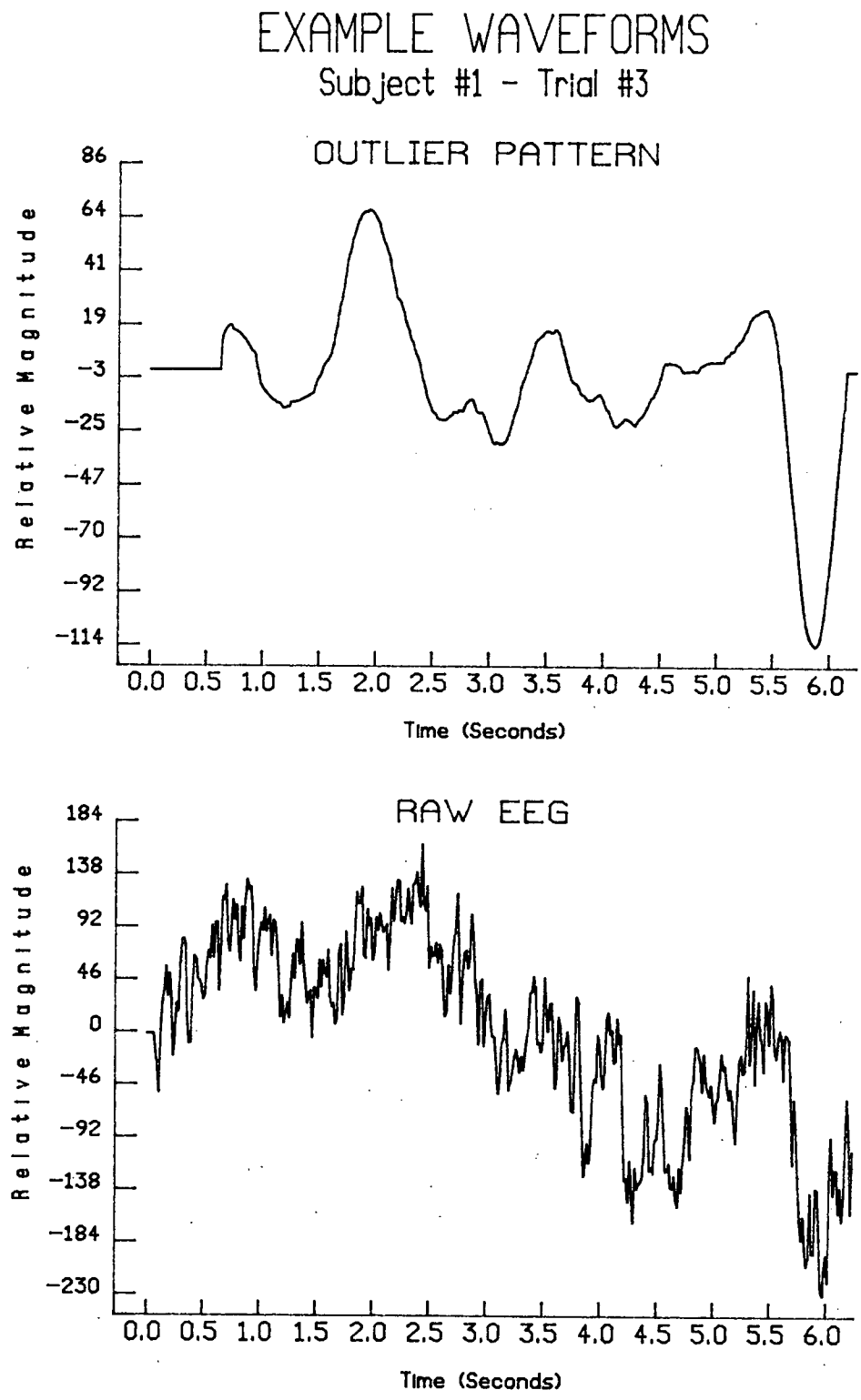

EXAMPLE WAVEFORMS
Subject #1 - Trial #3
OUTLIER PATTERN
Relative Magnitude
86
64
41
19
-3
-25
-47
-70
-92
-114
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0
Time (Seconds)
RAW EEG
Relative Magnitude
184
138
92
46
0
-46
-92
-138
-184
-230
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0
Time (Seconds)

Figure 4.10a

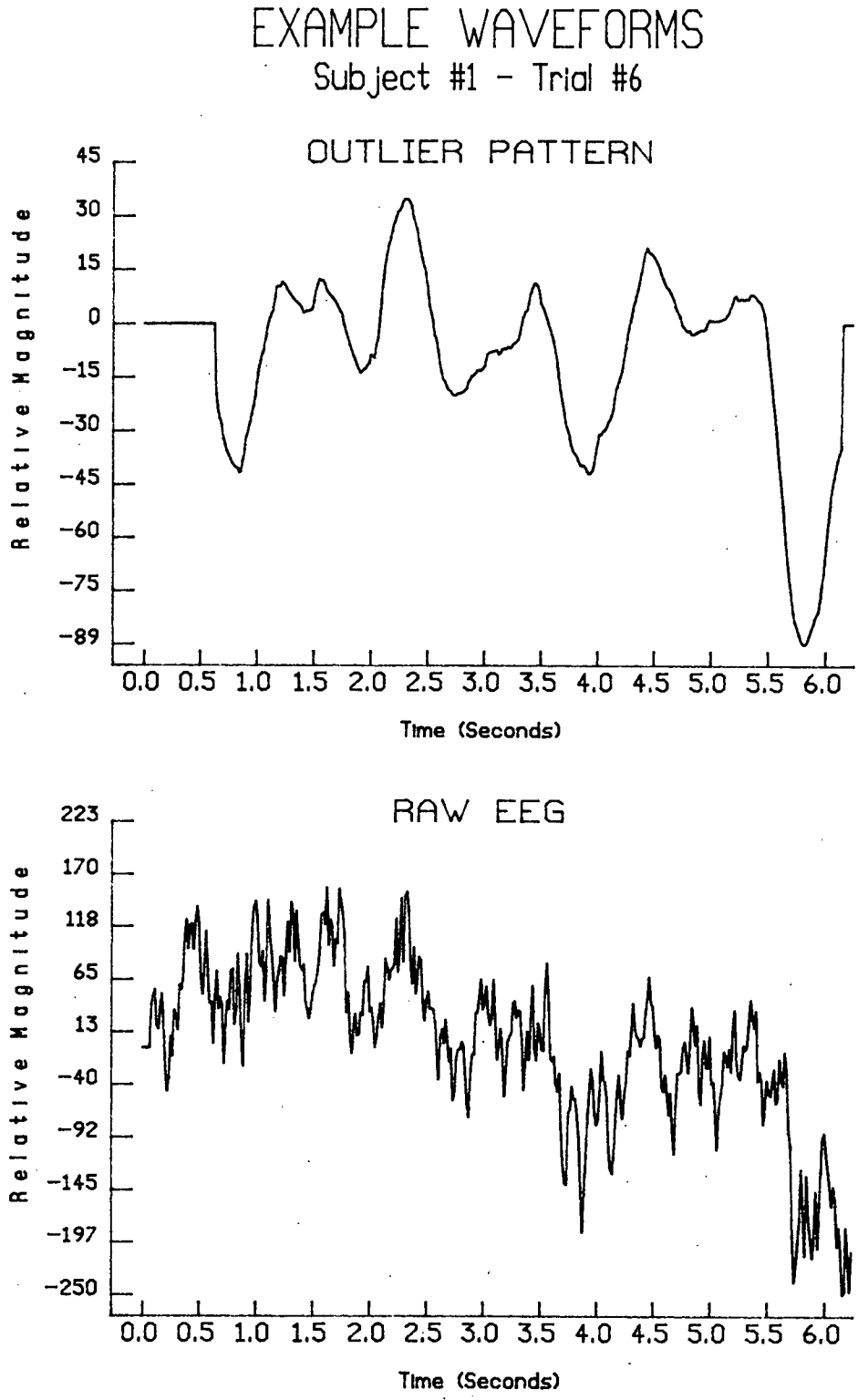

EXAMPLE WAVEFORMS
Subject #1 - Trial #6
OUTLIER PATTERN
Relative Magnitude
45
30
15
0
-15
-30
-45
-60
-75
-89
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0
Time (Seconds)
RAW EEG
Relative Magnitude
223
170
118
65
13
-40
-92
-145
-197
-250
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0
Time (Seconds)

Figure 4.10b

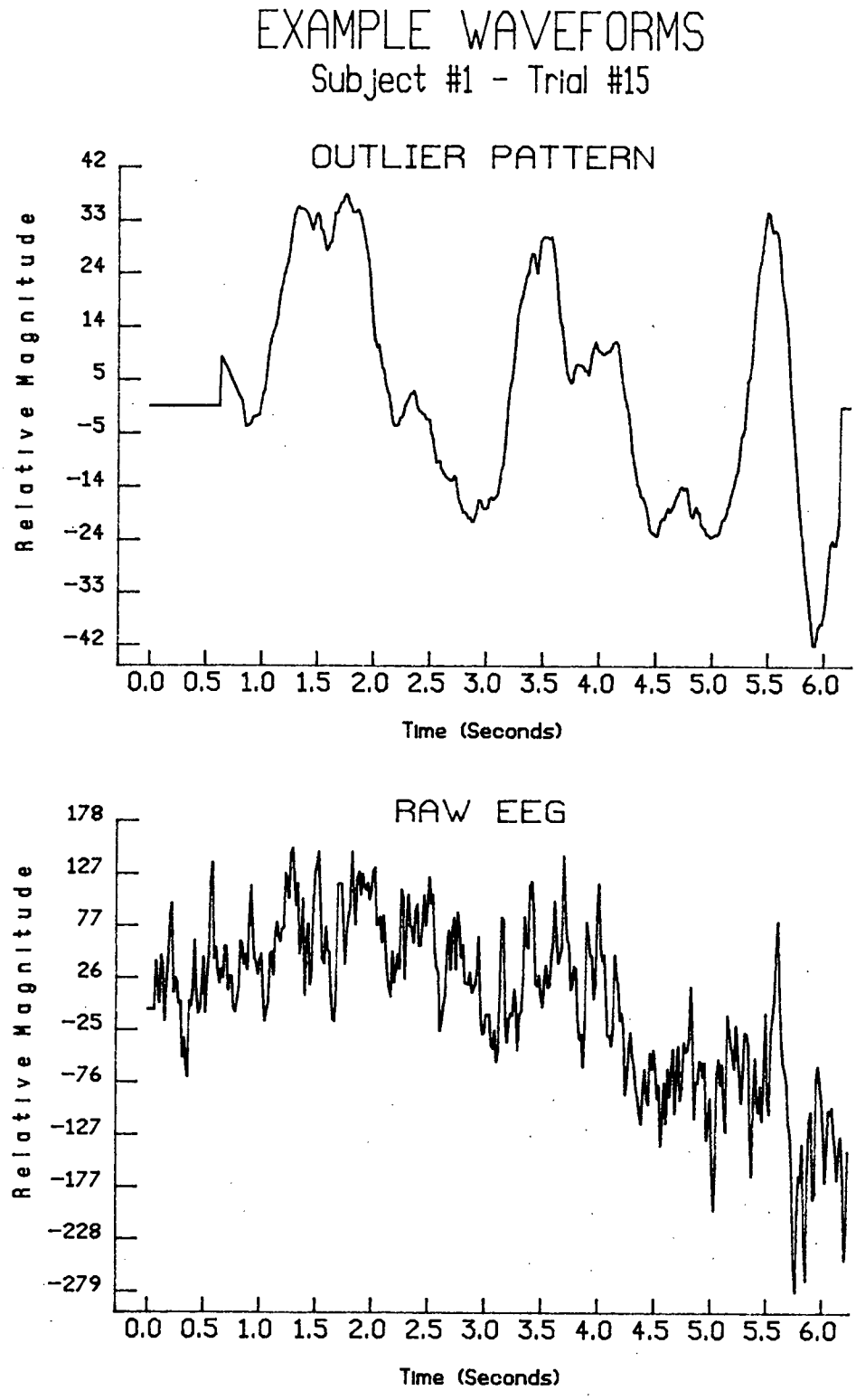

EXAMPLE WAVEFORMS
Subject #1 - Trial #15
OUTLIER PATTERN
Relative Magnitude
42
33
24
14
5
-5
-14
-24
-33
-42
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0
Time (Seconds)
RAW EEG
Relative Magnitude
178
127
77
26
-25
-76
-127
-177
-228
-279
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0
Time (Seconds)

Figure 4.10c

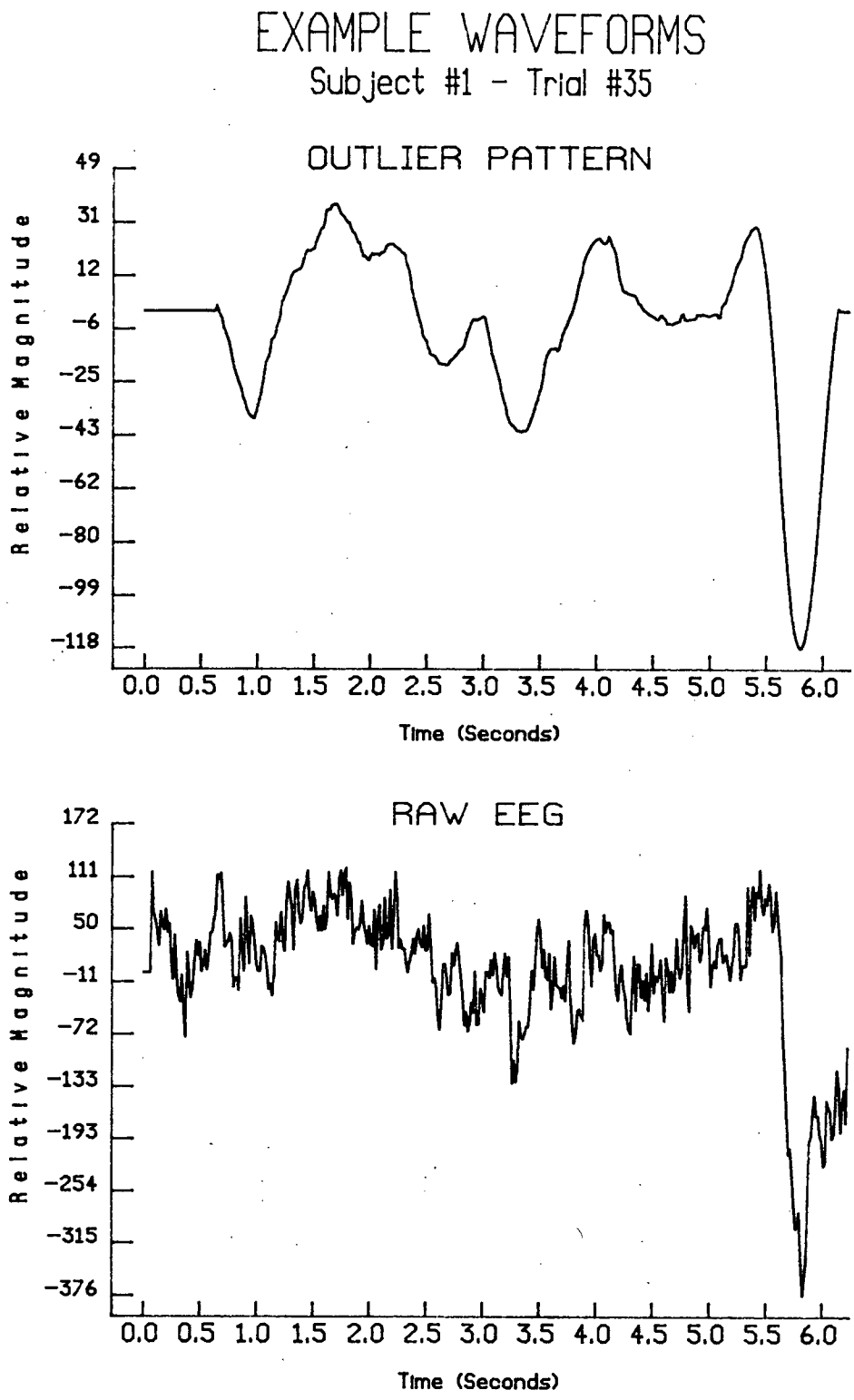

EXAMPLE WAVEFORMS
Subject #1 - Trial #35
OUTLIER PATTERN
Relative Magnitude
49
31
12
-6
-25
-43
-62
-80
-99
-118
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0
Time (Seconds)
RAW EEG
Relative Magnitude
172
111
50
-11
-72
-133
-193
-254
-315
-376
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0
Time (Seconds)

Figure 4.10d

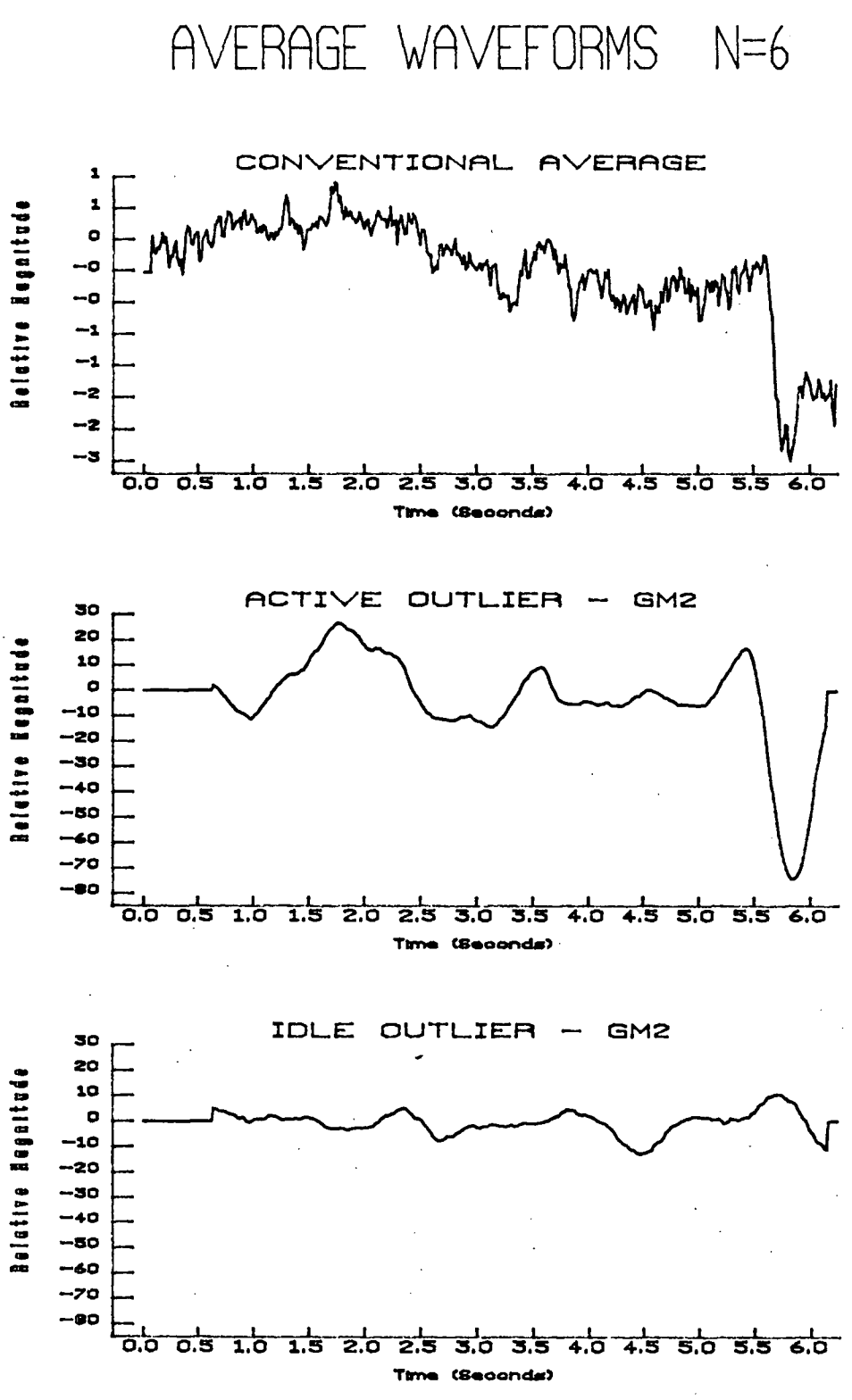

AVERAGE WAVEFORMS N=6
CONVENTIONAL AVERAGE
Relative Magnitude
1
1
0
-0
-0
-1
-1
-2
-2
-3
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0
Time (Seconds)
ACTIVE OUTLIER - GM2
Relative Magnitude
30
20
10
0
-10
-20
-30
-40
-50
-60
-70
-80
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0
Time (Seconds)
IDLE OUTLIER - GM2
Relative Magnitude
30
20
10
0
-10
-20
-30
-40
-50
-60
-70
-80
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0
Time (Seconds)

Figure 4.11a

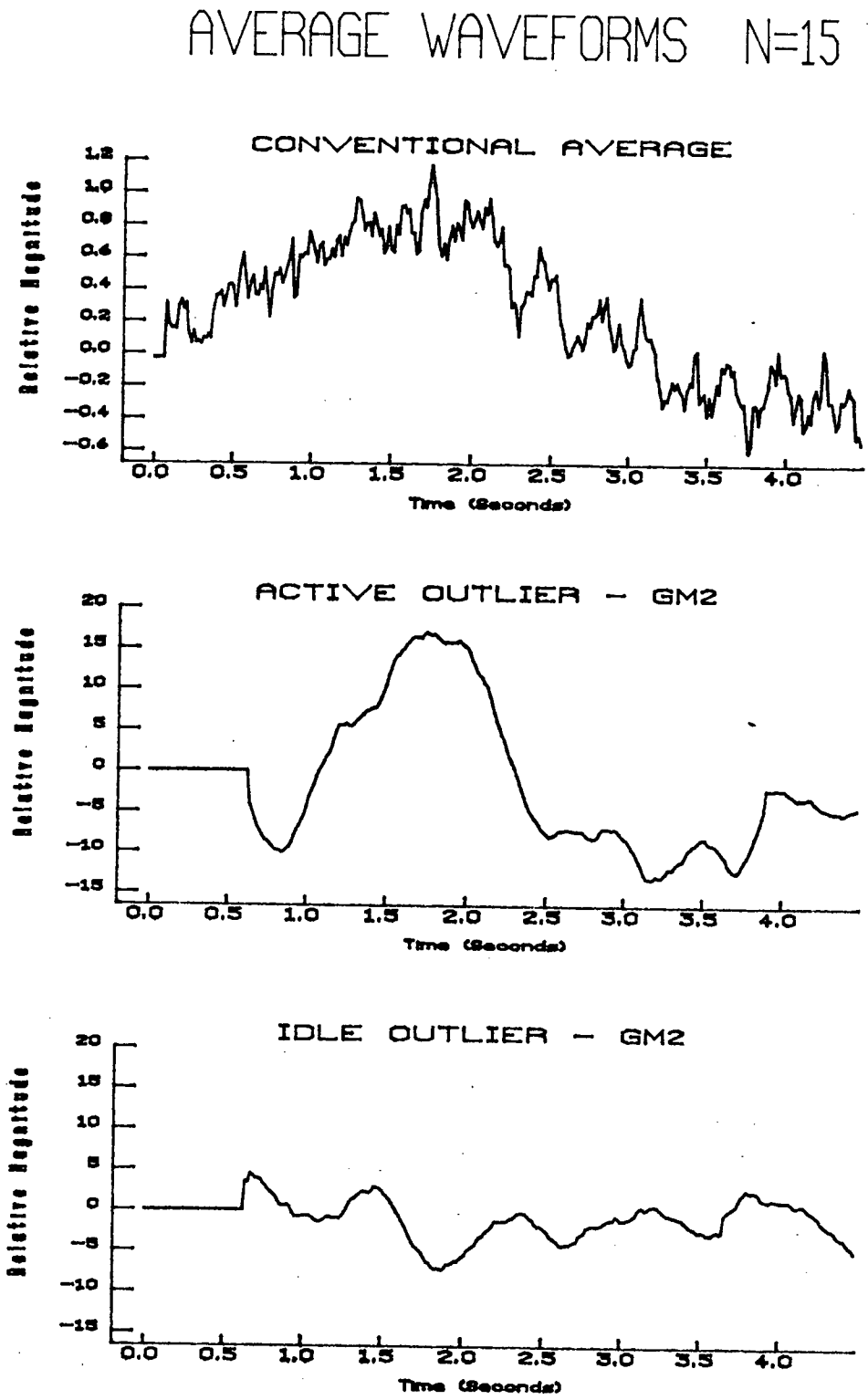

AVERAGE WAVEFORMS N=15
CONVENTIONAL AVERAGE
Relative Magnitude
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0
Time (Seconds)
ACTIVE OUTLIER - GM2
Relative Magnitude
20
15
10
5
0
-5
-10
-15
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0
Time (Seconds)
IDLE OUTLIER - GM2
Relative Magnitude
20
15
10
5
0
-5
-10
-15
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0
Time (Seconds)

Figure 4.11b

subtle features become less pronounced as N is increased, indicates that there is a significant amount of uniqueness in individual trials that is lost as many trials are averaged together. This uniqueness is certainly in part due to the variance in thumb movements from trial to trial which can be seen clearly in Section 4.7. It is also expected that additional trial by trial uniqueness is due to cognitive factors such as the mental intensity with which the subject carried out the task. The average of the idle task outlier patterns clearly demonstrates that there is no significant information that is being reinforced across idle task trials. The conventional average of active trials shows that with N's of 6 and 15 the motor potential information is quite limited and the "smearing" effect of event related information that is discussed above for the active case outlier patterns would also be occurring in these conventional averages. Hence, with the conventional averaging method, even with much greater N's as in the case of the Grunewald study (see Section 4.1), the information obtained will be limited to that which has remained relatively consistent across the trials.

4.6.4 LSQ Active Outlier Patterns Degrading with Higher Model Orders

It would also be expected, based on the neurological premise, that the single trial processing method would perform best when the AR model order was selected to best fit the idle case. As the model order is increased the AR model would be expected to gain some improved ability to represent the motor related activity in the active task EEG. Hence, the performance of the single trial method should begin to degrade since the cleaning process, which utilizes the higher order AR model, would lose some of its effectiveness in

detecting motor related outliers. A pair of averaged active outlier pattern plots using LSQ parameters for model order 12 (generally appropriate for the idle case) and model order 22 (generally appropriate for the active case) are shown in Figure 4.12. These plots demonstrate that the performance does degrade, in terms of both the amplitude and the detail of features in the averaged outlier pattern, when the model order is better matched to the active case.

4.7 Statistical Analysis of Features in the Outlier Patterns

The set of 15 active trial outlier patterns superimposed with the corresponding encoded thumb movements from Subject #1 are given in Figure 4.13. The patterns, although unique from trial to trial, do seem to posses a generally consistent waveform which contains features that appear related to events in the thumb movements. Statistical analysis was carried out to determine if features in the individual thumb movements are related to features in the corresponding outlier patterns. Two features in the thumb movement and three features in the outlier pattern were utilized in the statistical analysis. The features are described below and are shown in Figure 4.14.

Feature 1: Time from epoch onset to the point when the thumb movement first reaches the "on target" position.

Feature 2: Time from epoch onset to the point when the thumb movement first leaves the "on target" position.

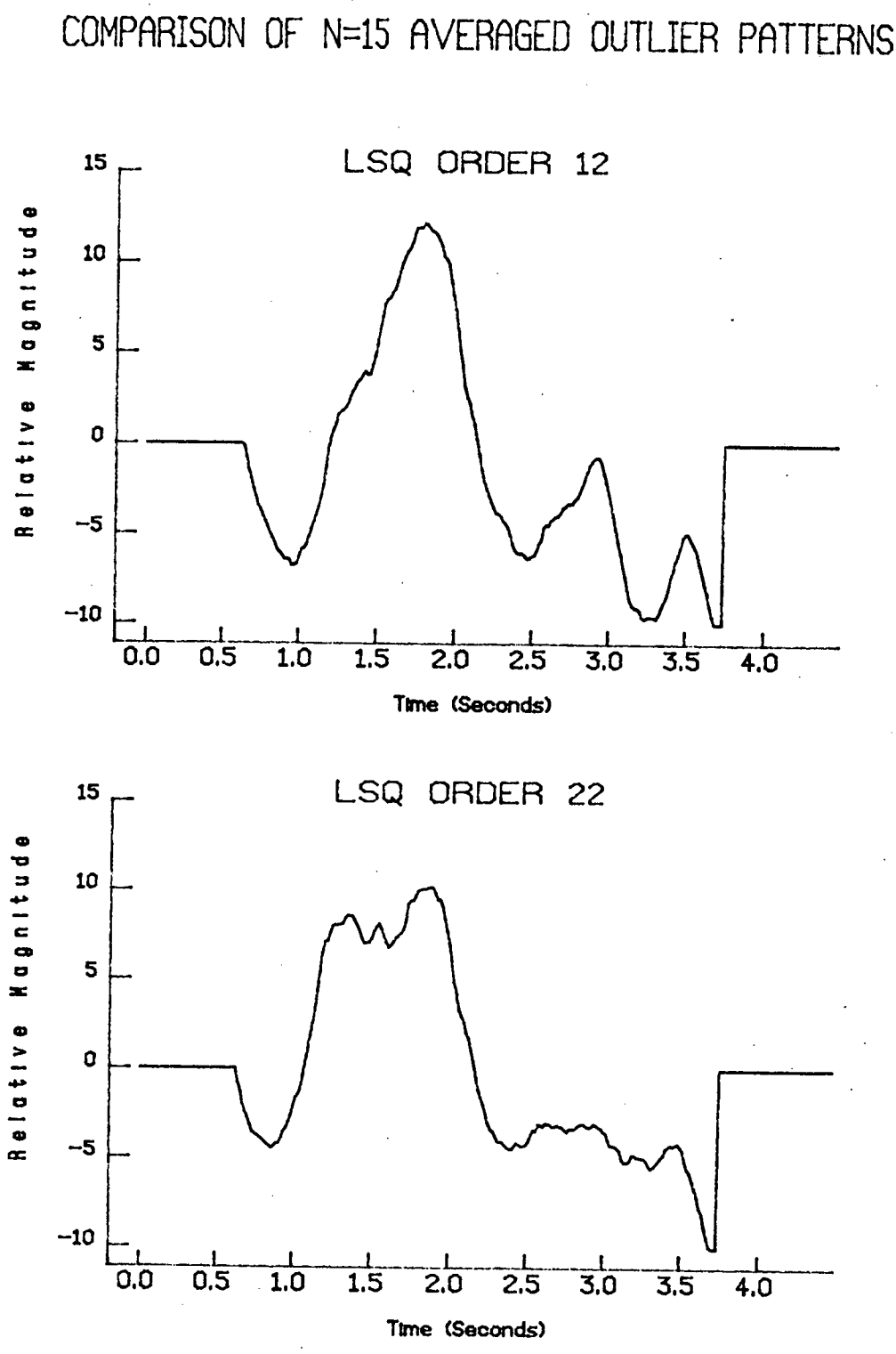

COMPARISON OF N=15 AVERAGED OUTLIER PATTERNS
LSQ ORDER 12
Relative Magnitude
15
10
5
0
-5
-10
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Time (Seconds)
LSQ ORDER 22
Relative Magnitude
15
10
5
0
-5
-10
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Time (Seconds)

Figure 4.12

SINGLE TRIAL GM2 OUTLIER PATTERNS
WITH CORRESPONDING THUMB MOVEMENT

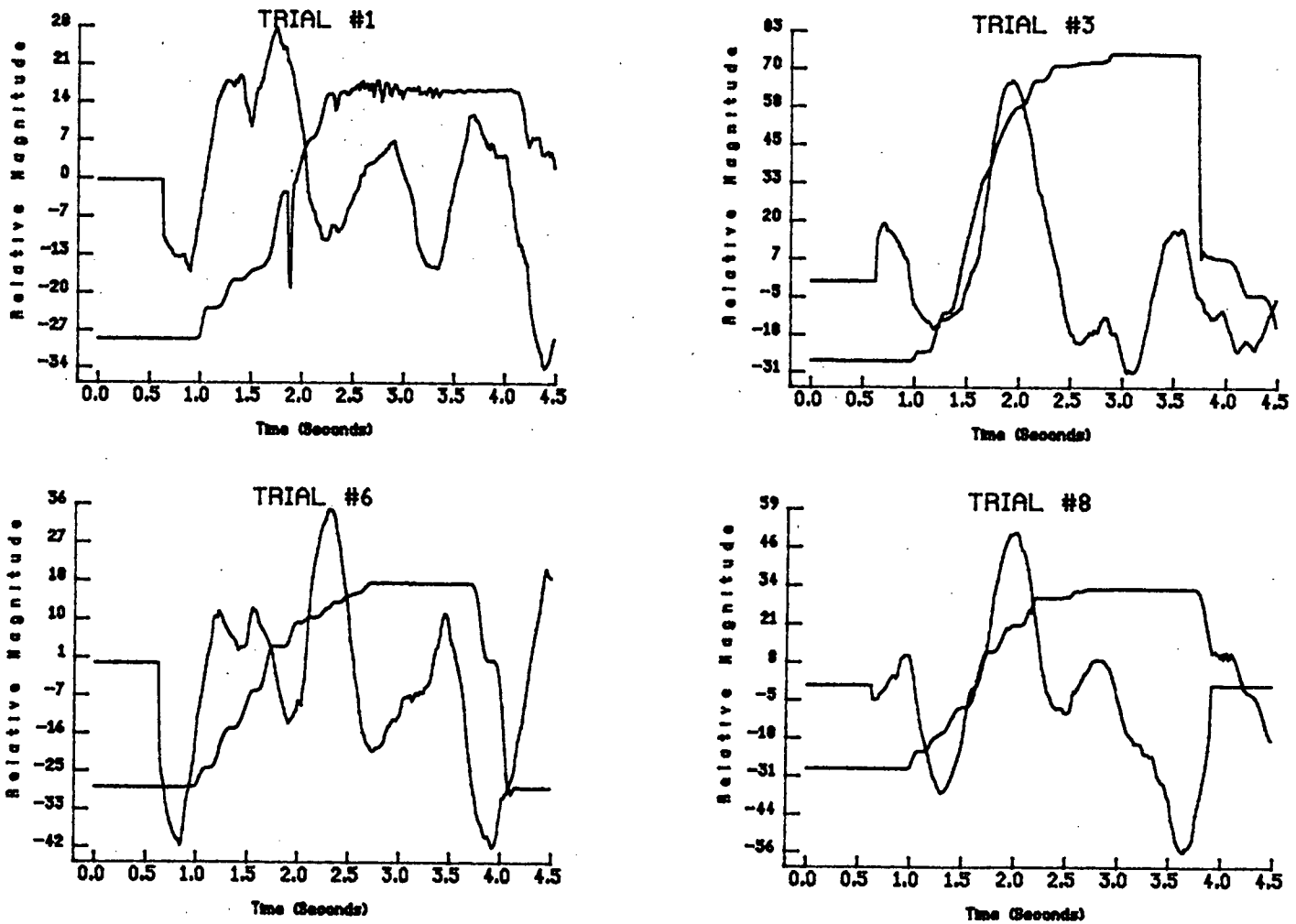


Figure 4.13a

SINGLE TRIAL GM2 OUTLIER PATTERNS
WITH CORRESPONDING THUMB MOVEMENT

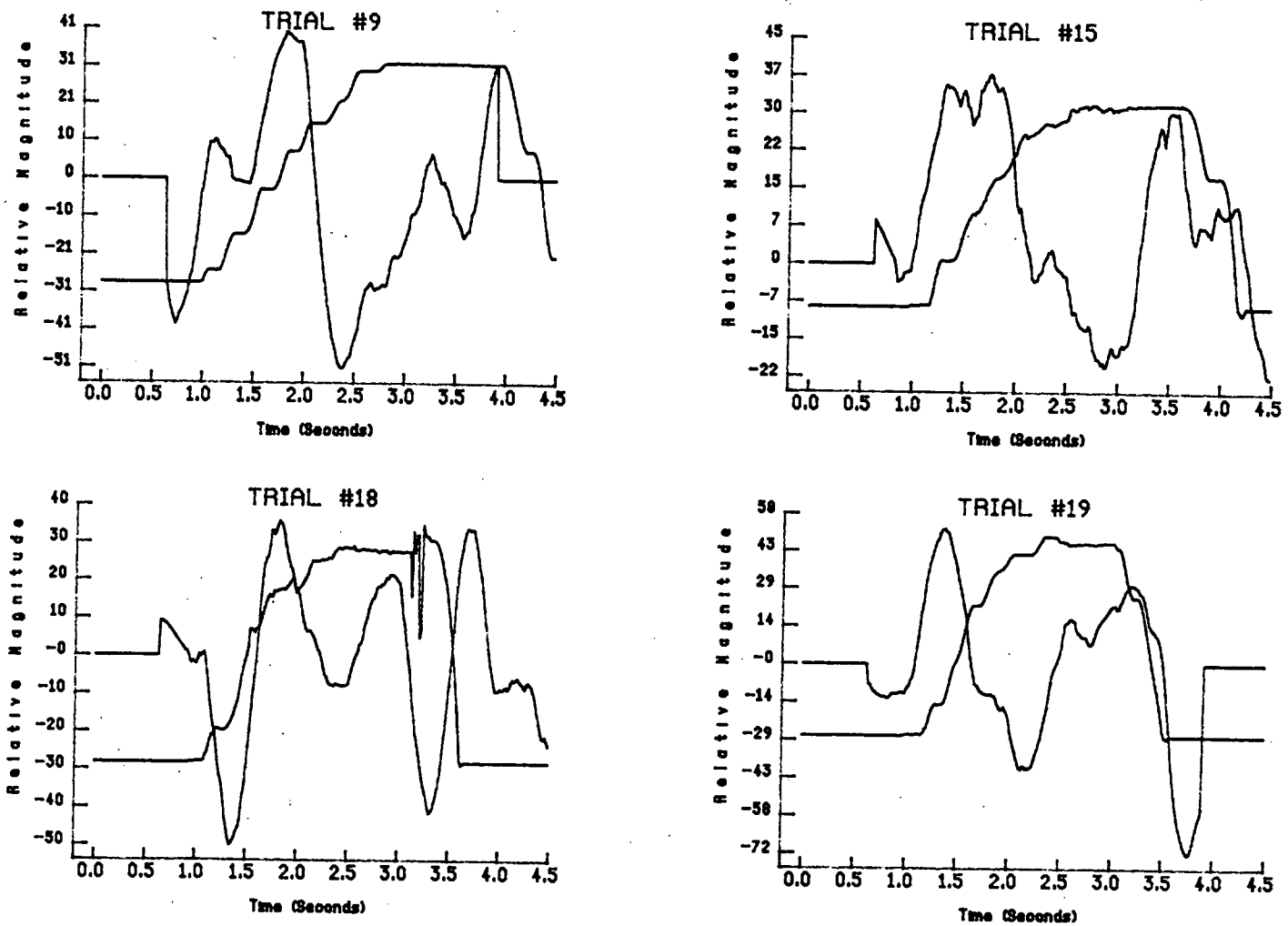


Figure 4.13b

SINGLE TRIAL GM2 OUTLIER PATTERNS
WITH CORRESPONDING THUMB MOVEMENT

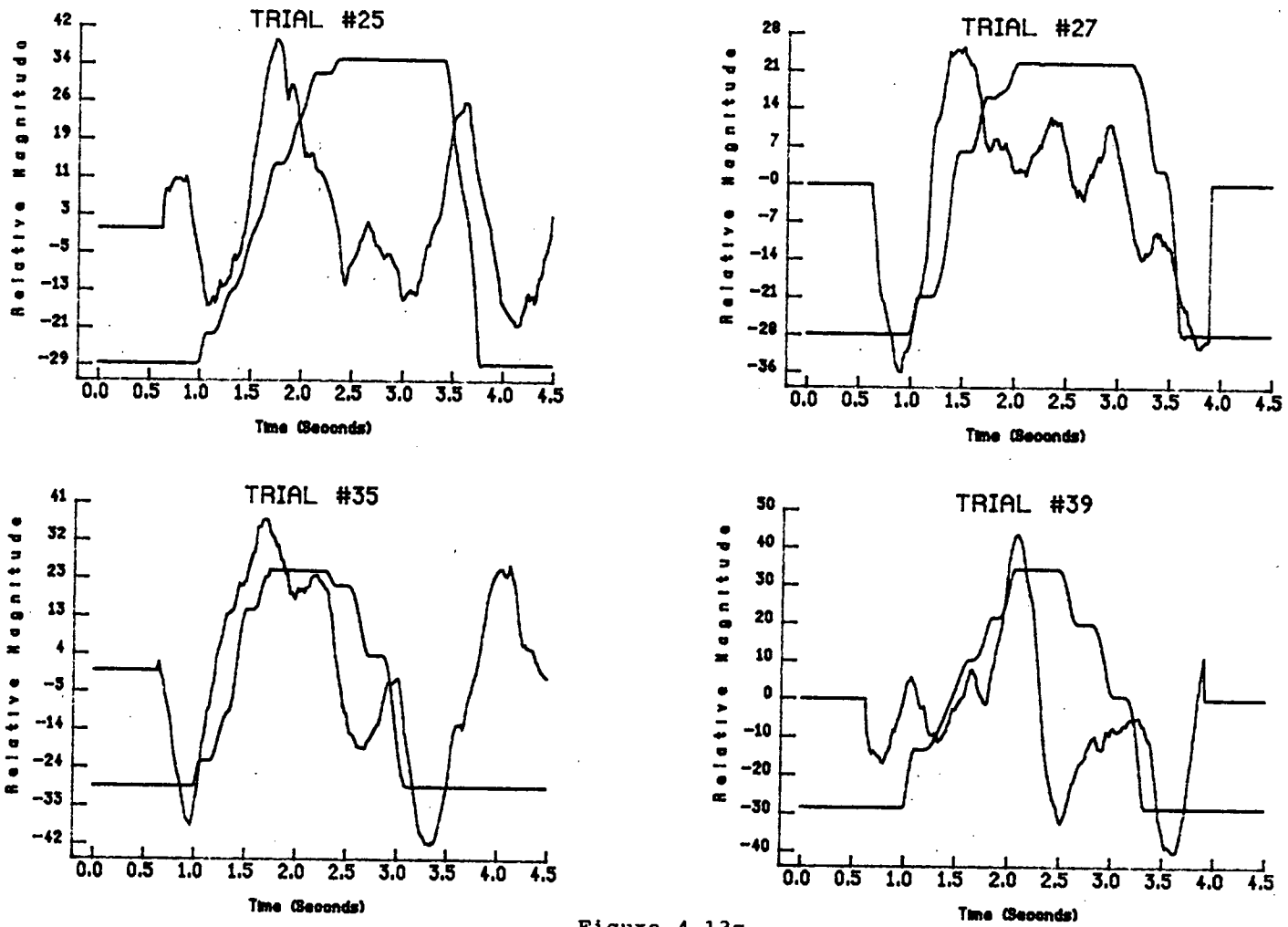


Figure 4.13c

SINGLE TRIAL GM2 OUTLIER PATTERNS WITH CORRESPONDING THUMB MOVEMENT

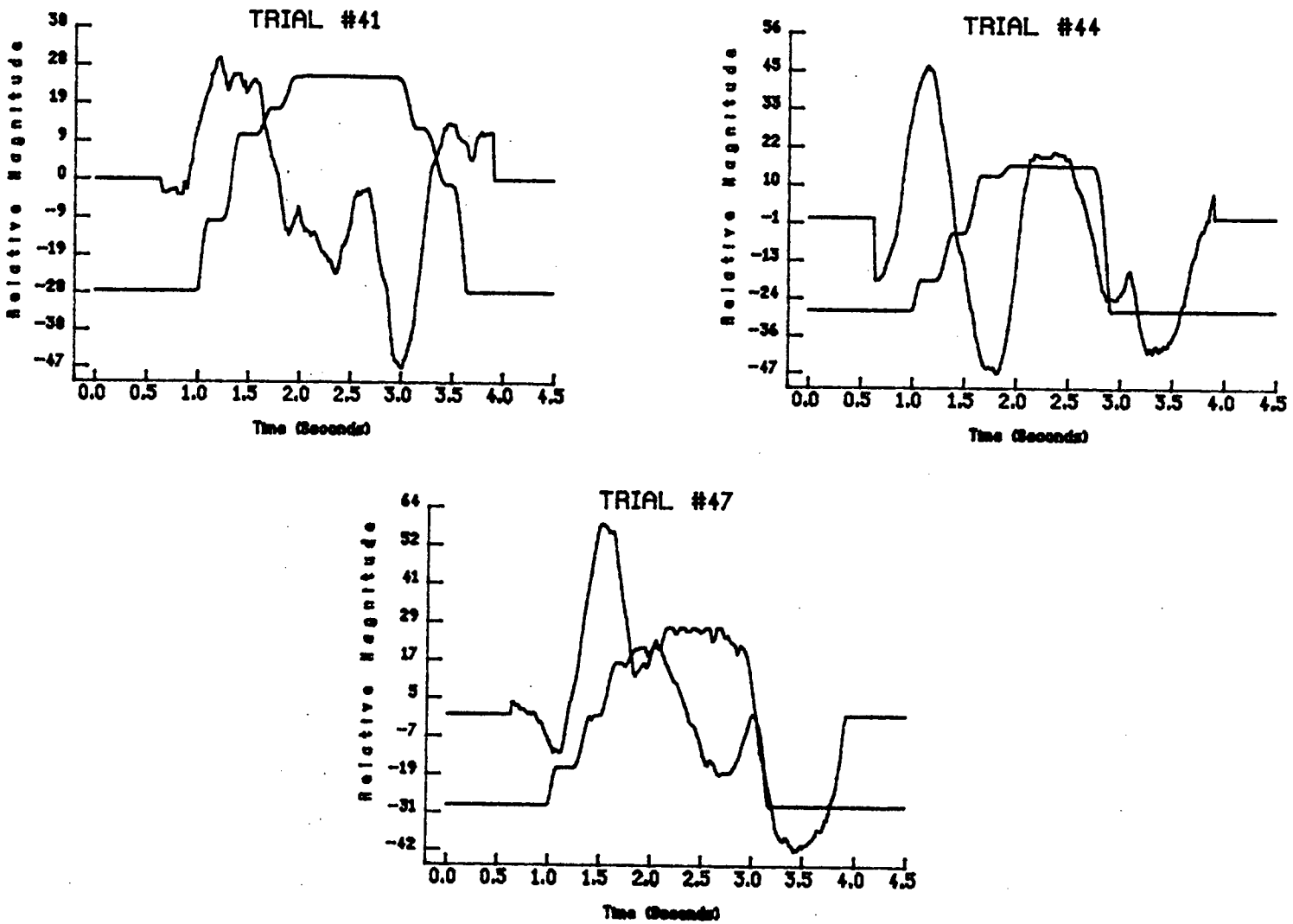


Figure 4.13d

SINGLE TRIAL FEATURE DEFINITION

SAMPLE THUMB MOVEMENT WITH STANDARD GM2 OUTLIER PATTERN FOR SUBJECT #1

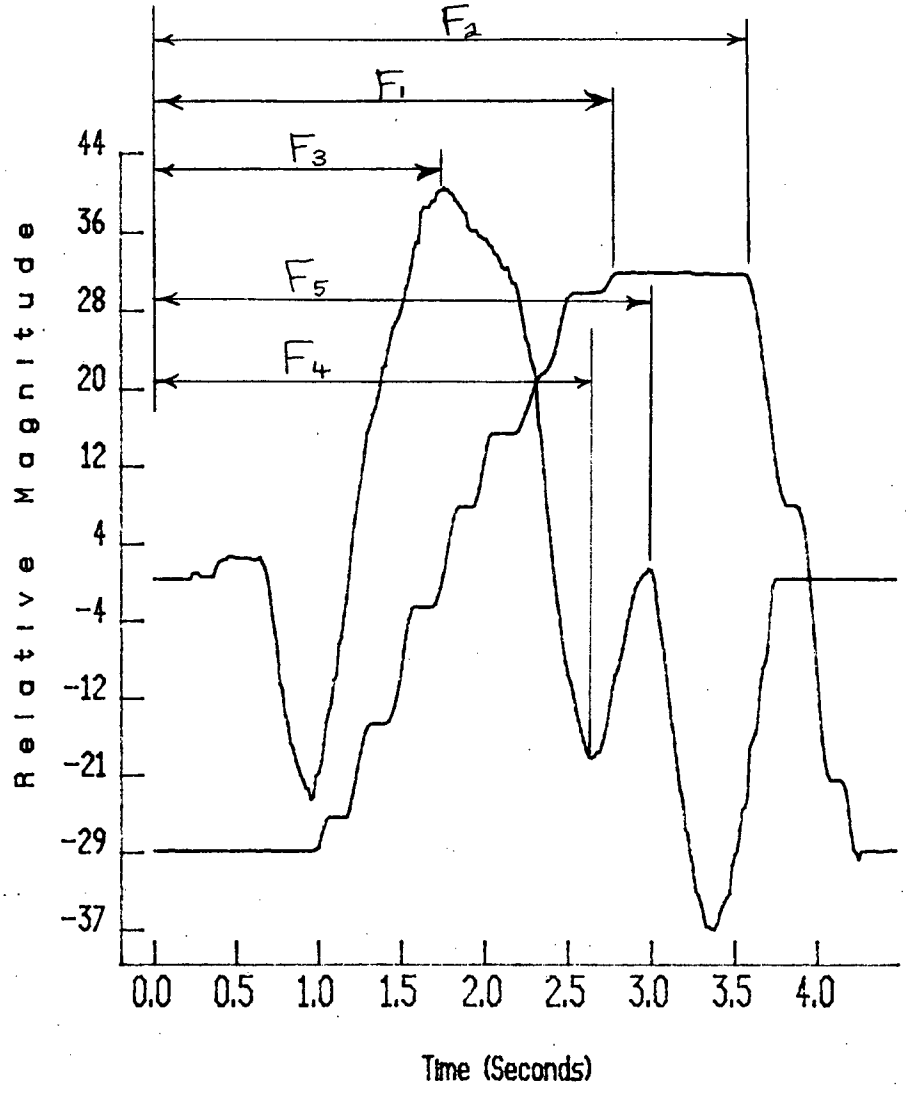


Figure 4.14

Feature 3: Time from epoch onset to the first dominant (greatest amplitude) positive peak in the outlier pattern.

Feature 4: Time from epoch onset to the first negative peak in the outlier pattern after feature 3, that has a minimum of 5 units magnitude peak-to-trough difference.

Feature 5: Time from epoch onset to the next positive peak in the outlier pattern after feature 4, that has a minimum of 20 units magnitude peak-to-trough difference on both sides of the peak.

There was an expectation resulting from the earlier conventional study by Grunewald and Grunewald-Zuberbier [7] and from observations taken from Figure 4.13 that feature 1 would be particularly related to features 3 and 4 whereas feature 2 would be particularly related to feature 5. The sample correlation coefficients between all of the features from Subject #1 were calculated and are summarized in Table 4.1. These results show that the correlation coefficients between the features that were expected to be particularly related are the strongest, with coefficient values all greater than 0.77. Hence, this demonstrates that there is a strong consistent relationship between features in the thumb movement and features in the single trial outlier pattern. In particular, the relationship between features in the outlier pattern and in the thumb movement was examined using the z-test for the difference between correlations calculated on dependent samples (see Steiger [54]). The results from these tests are also summarized in Table 4.1. They show that features 3 and 4 correlated with feature 1 significantly

TABLE 4.1
SINGLE TRIAL FEATURE STATISTICS

Feature Correlation Matrix

	feature 1	feature 2	feature 3	feature 4	feature 5
feature 1	1.0				
feature 2	0.76	1.0			
feature 3	0.78	0.51	1.0		
feature 4	0.88	0.69	0.71	1.0	
feature 5	0.60	0.80	0.41	0.57	1.0

z-test on the Difference Between Correlations
Calculated on Dependent Samples

	Correlation Coefficients		z	p (one-sided)
	feature 4	feature 5		
feature 1	0.78	0.51	1.91	0.029
feature 2	0.88	0.69	1.67	0.048
feature 3	0.60	0.80	1.55	0.060

more strongly than with feature 2 ($p < 0.05$). As expected, the correlation between feature 5 and feature 2 was larger than that between feature 5 and feature 1, but this difference achieved only a marginal level of significance.

4.8 Application of Dynamic Time Warping to Outlier Patterns

All the initial work with actual EEG was carried out on the data from one subject, referred to as Subject #1. These initial investigations, revealed that the use of dynamic time warping (DTW) provided the best quantitative measure of performance for the single trial processing method compared to the other previous analysis. Hence, DTW analysis was ultimately applied to all four of the subjects included in this study. Specific results are summarized in the following subsections.

4.8.1 Standard Outlier Patterns using Dynamic Time Warping

DTW as described by Roberts et al. [55] was used to obtain standard (template) representative single trial active outlier patterns for each subject. The time warping procedure attempts to best match waveform A to waveform B by shifting, expanding or contracting the time scale of waveform A in such a manner that minimizes the "cost" of warping. The cost of warping is based on the following cost function [55]

$$C = \frac{1}{Q(A,B,W) - \lambda P(w)} \qquad 4.15$$

where w is the warped time function (warped time axis) used to warp waveform A, Q is the correlation between warped waveform A and waveform B, P

is a penalty function and λ is the penalty coefficient. The penalty function is nonlinear such that the penalty for large expansions or contractions is proportionately much higher than the penalty for small expansions or contractions. The λ coefficient is a tuning parameter that directly effects how expensive it is to warp. In all the DTW applications used in this study a λ=75.0 was utilized because it was found by trial and error that this value produced reasonable warped waveforms; smaller values of produced λ extreme warpings whereas larger values of λ produced warpings that were only shifted in time and contained very limited expansions or contractions.

A standard pattern for each of the four subjects was achieved by using the procedure recommended by Roberts et al. [55]. The set of active single trial patterns was warped against each pattern in that set. The pattern that produced the lowest mean cost and variance across the set was then selected as the best representative pattern of that set. The standard pattern was then constructed by averaging together all the patterns in the set after being warped to the above selected pattern. Plots of the standard patterns for all four subjects using LSQ, GM and GM2 parameter estimation are provided in Figure 4.15.

4.8.2 DTW Cost Statistics on Individual Subjects

Once a standard active case pattern was obtained for each subject, it was warped against the 15 trials of active outlier patterns and the 15 trials of idle outlier patterns for each subject. Each time a pattern is warped to the standard pattern a cost value is produced. This cost reflects how well the single trial pattern "fit" the standard pattern. The lower the cost the

STANDARD OUTLIER PATTERNS
GM2 MODEL PARAMETERS

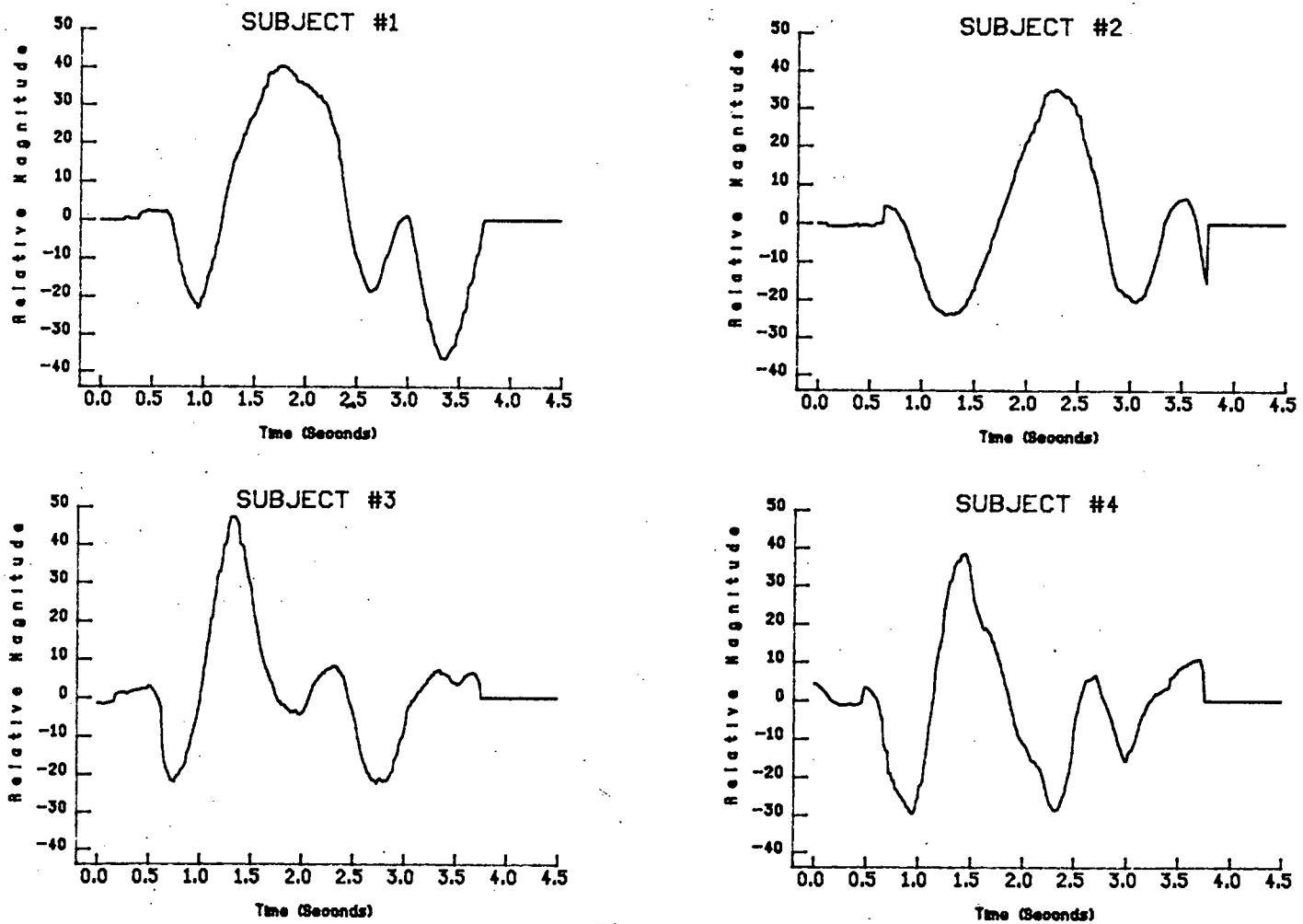


Figure 4.15a

STANDARD OUTLIER PATTERNS
GM MODEL PARAMETERS

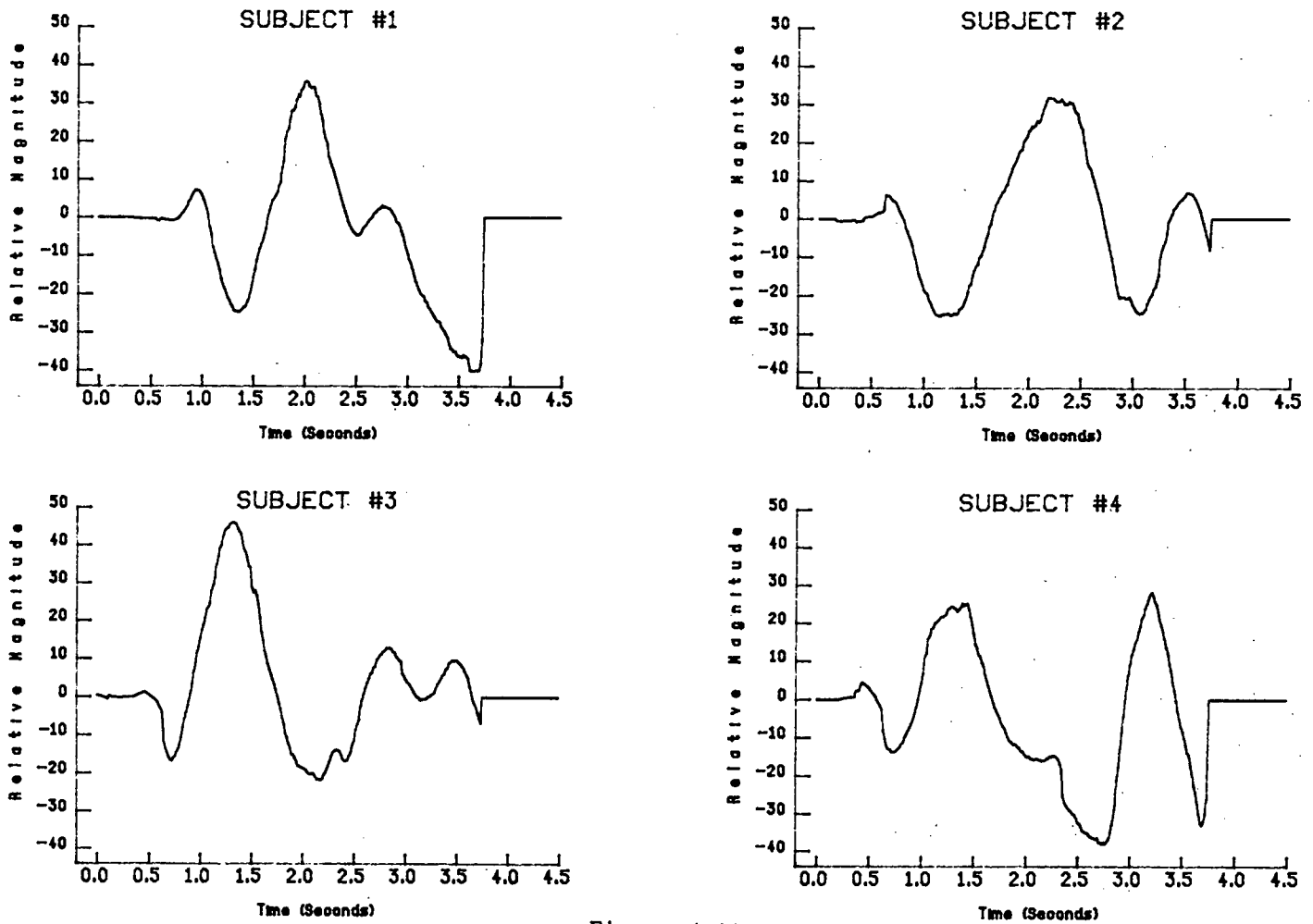


Figure 4.15b

STANDARD OUTLIER PATTERNS
LSQ MODEL PARAMETERS

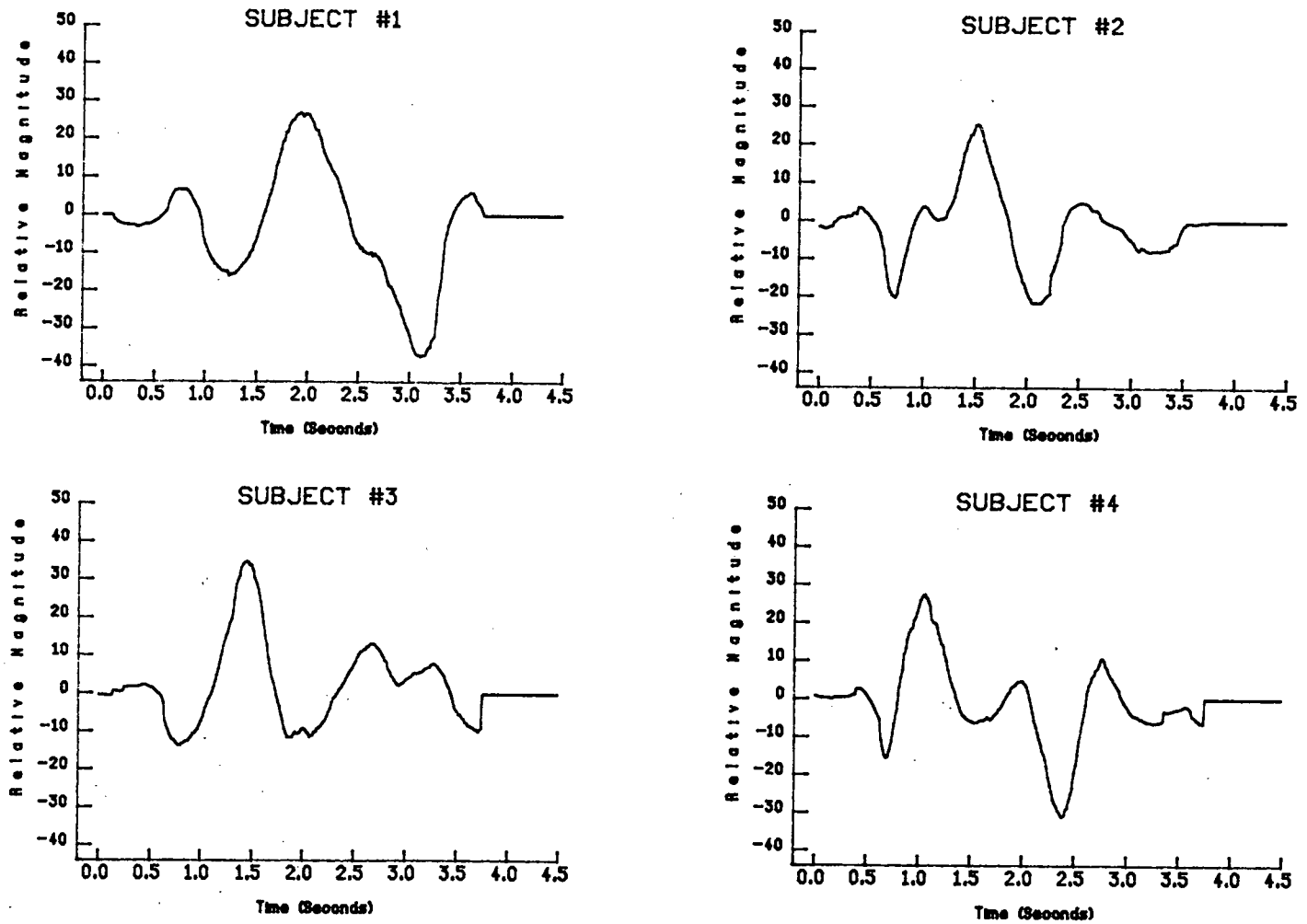


Figure 4.15c

better the fit. Therefore, it would be expected that the costs in the active cases would be smaller than the costs in the idle case. A t-test designed to test the difference between two means, given by [56]

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2 + s_2^2}{N}}} \qquad 4.16$$

$$df = 2N - 2 = 28$$

where $\bar{x}_1$ and $\bar{x}_2$ are the sample means, s_1^2 and s_2^2 are the sample variances, and N is the number of active and idle cases, was applied to the mean costs for the active and idle cases. The results of this test for GM2, GM, and LSQ model parameters for all four subjects are summarized in Table 4.2. This test shows that in the GM2 case the difference between the means is highly statistically significant ($p < .001$). The mean differences in the GM and LSQ case are also statistically significant (ranging from $p < .001$ to $p < .01$) except for Subject #4 with LSQ parameters where the difference was not significant. These results strongly support the expectation that the costs from the active case would be smaller than the costs from the idle case. In turn, this implies that on average the active case patterns fit the standard patterns much better than the idle case patterns.

4.8.3 Grouped DTW Cost Statistics

The average active and idle cost values from the four subjects were considered together to provide inferential statistics about the actual population (for all possible subjects), of mean differences between idle and active cases. For each subject, the difference between the average idle cost (over the 15 idle trials) and the average active cost (over the 15 active trials) was evaluated. The differences appear in Table 4.3 and form

TABLE 4.2
DIFFERENCE BETWEEN IDLE AND ACTIVE WARPING COSTS
t-Test Results for the Difference
Between Means: df=28

SUBJECT	DIFFERENCE BETWEEN THE MEANS	TWO-SIDED t-VALUE	p <
GM2 MODEL			
1	13.3	6.05	0.001
2	16.4	5.04	0.001
3	11.6	4.66	0.001
4	11.5	6.25	0.001
GM MODEL			
1	8.7	4.18	0.001
2	12.7	3.74	0.001
3	6.1	2.93	0.01
4	4.9	3.35	0.01
LSQ MODEL			
1	12.7	4.08	0.001
2	17.2	3.65	0.002
3	12.9	2.96	0.01
4	3.5	0.65	X

the basis of the inference to be made for all possible subjects. This was carried out by considering the null hypothesis that the mean difference of the actual population is zero and then applying a statistical test to determine whether that hypothesis should be rejected. The statistical test was a t-test designed to test the difference between two means with correlated (paired) samples. It is given by [50]

$$t = \frac{MD}{SE_{MD}} \qquad df = N - 1 = 3 \qquad 4.17$$

where MD is the sample mean difference (mean of the paired differences), SE_{MD} is the standard error of this sample mean difference and N equals the number of mean differences (number of subjects). The standard error of the sample mean difference is given by

$$SE_{MD} = \frac{s_d}{\sqrt{N}} \qquad 4.18$$

where s_d is the sample standard deviation of the mean difference. With three degrees of freedom a 95% confidence interval for the mean difference is given by [50]

$$CI_{95} = MD \pm 3.18(SE_{MD}) \qquad 4.19$$

The group statistics for GM2, GM, and LSQ parameter estimation are summarized in Table 4.3. In the GM2 case the statistics imply that, even given this small sample of four subjects, the hypothesis that the mean difference of the actual population of mean differences is zero is strongly rejected ($p < .002$). Alternatively, in terms of confidence intervals, it was found that with 95% confidence the interval of 9.7 to 16.8 contains the actual mean difference value. The implications are similar but less significant in the GM and LSQ cases. These results imply that the mean cost differences between active and idle trials based on all possible subjects is highly unlikely to be zero.

TABLE 4.3
GROUP STATISTICS

Mean Cost Difference Between
Active and Idle Cases

	GM2 MODELING	GM MODELING	LSQ MODELING
Subject #			
1	13.3	8.7	12.7
2	16.4	12.7	17.2
3	11.6	6.1	12.9
4	11.5	4.9	3.5
Mean Cost	13.2	8.1	11.6
Std. Dev.	2.28	3.45	5.77
Std. Error	1.14	1.73	2.88

Summary for GM2 Modeling

Mean Difference = 13.2
two sided t-test
$t = 11.58$: H_0 rejected $p < 0.002$
95% Confidence Interval: 9.6 to 16.8

Summary for GM Modeling

Mean Difference = 8.1
two sided t-test
$t = 4.68$: H_0 rejected $p < 0.05$
95% Confidence Interval: 2.6 to 13.6

Summary for LSQ Modeling

Mean Difference = 11.6
two sided t-test
$t = 4.03$: H_0 rejected $p < 0.05$
95% Confidence Interval: 2.4 to 20.8

TABLE 4.4
BAYESIAN CLASSIFICATION
Assuming Probabilities of Either Case Occurring are Equal

a) Cost of Misclassification set Equal

SUBJECT	BOUNDARY VALUE	ACTIVE CLASSIFIED CORRECTLY	FALSE POSITIVE
GM2 MODEL			
1	11.35	14/15	2/15
2	19.88	15/15	1/15
3	19.10	13/15	5/15
4	17.05	14/15	3/15
		56/60=93%	11/60=18%
GM MODEL			
1	15.28	12/15	4/15
2	19.10	15/15	4/15
3	16.93	12/15	8/15
4	13.26	13/15	5/15
		52/60=87%	21/60=35%
LSQ MODEL			
1	22.58	13/15	4/15
2	41.06	14/15	5/15
3	34.08	13/15	7/15
4	20.52	5/15	2/15
		45/60=75%	18/60=30%

b) Cost of Misclassifying an Idle Case as Active Set to be 5 Times Greater

GM2 MODEL			
1	9.18	14/15	0/15
2	19.88	12/15	1/15
3	19.10	10/15	0/15
4	17.05	12/15	0/15
		48/60=80%	1/60=1.7%
GM MODEL			
1	8.57	9/15	0/15
2	15.58	11/15	2/15
3	7.53	2/15	0/15
4	6.38	0/15	0/15
		22/60=37%	2/60=3.3%
LSQ MODEL			
1	12.22	4/15	1/15
2	28.68	5/15	2/15
3	21.35	7/15	3/15
4	2.92	0/15	0/15
		16/60=26%	6/60=10%

4.8.4 Bayesian Classification of Active Cases versus Idle Cases

A Bayesian classifier was applied to the cost values to classify active cases verses idle cases. It was assumed that the cost values had a Gaussian distribution. This classification was carried out under two different conditions: cost of misclassification set equal and cost of misclassifying an idle case as an active case set to be five times greater. In both conditions the probabilities of either case occurring were set equal. The results of the classification under both conditions across the four subjects for GM2, GM and LSQ are provided in table 4.4. Under the first condition 93% of the GM2 active cases were correctly classified with 18% of the idle cases being classified as active. The number of false positives is fairly high if the active cases are going to be used in a control application. Hence, the second classifying condition was carried out were the decision boundary was moved closer to the active case mean by increasing the cost of a false positive. In this case the percentage of GM2 active cases correctly classified was reduced to a still very respectable 80% but in so doing reduced the false positives to a very low 1.7% (only one idle case out of sixty was classified incorrectly).

The results using GM and LSQ outlier patterns under the first condition were fairly good in classifying active cases correctly but both had a very high percentage of false positives. Under the second condition, however, the classification performance using GM and LSQ outlier patterns fell off dramatically. This is perhaps the best demonstration of the superiority of utilizing GM2 parameter estimates in the single trial processing method.

4.8.5 Cross Validation and Intersubject Reliability Using DTW

A combined measure of the cross validation of a standard pattern on a unrelated set of outlier patterns and the intersubject reliability of the standard patterns was obtained by applying DTW with the standard pattern from one subject to the outlier patterns from a different subject. The GM2 standard patterns from each subject were applied to the GM2 outlier patterns of each of the other subjects. This resulted in twelve addition sets of active and idle costs. The t-test for the difference between two sample means was applied, in a similar manner as in Subsection 4.8.2, and the difference was highly significant ($p < 0.001$) in every case. In addition, in a similar manner as in subsection 4.8.4, Bayesian classification with the cost of misclassifying an idle case as an active case set to be five times greater was carried out and the results are given in Table 4.5. On average the percent correct using cross-matched standard patterns fell off by about 16% when compared to the percent correct using matched standard patterns for subjects 1,3 and 4 while it stayed exactly the same for subject 2. The percentage of false positives was almost exactly same for both the matched and cross-matched cases. The overall average of active cases correctly classified using standard patterns from one subject on the outlier patterns from the other subjects was 67% with only a 3% overall average of false positives. These results strongly indicate that the standard patterns do cross validate on data that was not used in the construction of these patterns and that these patterns provide substantial intersubject reliability.

TABLE 4.5
SUMMARY OF BAYESIAN CROSS-MATCHED CLASSIFICATION

Cost of Classifying an Idle Case as an Active Case
Set to be Five Times Greater

Standard Pattern #	Data from Subject # 1		2		3		4	
	C	F	C	F	C	F	C	F
1	14/15	0/15	12/15	1/15	9/15	0/15	8/15	0/15
2	11/15	0/15	12/15	1/15	8/15	0/15	11/15	0/15
3	12/15	0/15	12/15	1/15	10/15	0/15	9/15	0/15
4	11/15	1/15	12/15	1/15	6/15	0/15	12/15	0/15

C = Active Trial Classified Correctly F = False Positive

AVERAGED RESULTS

Subject Number	With Matched Standard Pattern: Active Correct	With Matched Standard Pattern: False Positive	With Cross-Matched Standard Pattern: Active Correct	With Cross-Matched Standard Pattern: False Positive
1	93%	0%	75%	2%
2	80%	7%	80%	7%
3	66%	0%	51%	0%
4	80%	0%	62%	0%
OVERALL AVERAGE	80%	2%	67%	3%

CHAPTER 5

CONCLUSION

5.1 Summary of Major Results and Related Conclusions

AR Parameter Estimation

Simulation studies on AR parameter estimation demonstrated that the robust general maximum likelihood (GM) methods performed almost as well as the least squares (LSQ) method on Gaussian processes and significantly better on additive outlier (AO) contaminated Gaussian processes. Amongst the GM methods the GM2 method provided the best performance. However, it should be noted that the GM2 estimates are also the most computationally expensive to calculate. In terms of the single trial outlier processing method, the most important finding from these simulation studies is that given the AO model (see equation 3.37) the robust estimation methods, in particular the GM2 method, demonstrated a strong ability to model the process x_i without being unduly influenced by the additive outliers v_i.

Neurological Premise and Outlier Extraction

The basic neurological premise for the single trial processing method is that event related potentials have an additive outlier effect on the ongoing EEG process. If the outlier content could be extracted from the resulting overall combined process, then single trial event related information could be obtained. Simulation studies demonstrated that a robust signal estimator, which utilizes robustly estimated AR model parameters, has a distinct ability to extract a significant amount of the additive outlier

content from a contaminated process. This extraction process was found to be most effective when GM2 parameter estimates were utilized. This result should be expected since the GM2 parameter estimation is based on an estimated cleaned signal x_i and hence, it has the best opportunity to provide a good estimated model representing the actual process x_i. The better the model of x_i the better the expected performance of the outlier extraction process.

Spectral Analysis of EEG

AR spectral analysis provided a great deal of insight into both the EEG signal itself and into the application of AR modeling to the EEG signal. EEG data used in this thesis work was collected during an active task involving motor activity and an idle task not involving motor activity. Spectral analysis demonstrated that the signal characteristics of these EEG signals were typically changing at a relatively rapid rate. Hence, an EEG segment size as small as practical parameter estimation considerations would allow was utilized. Ultimately, in the single trial processing method, the EEG epochs were broken down into 1.5 second segments offset by 0.75 seconds. AR spectral analysis was also the key tool in determining appropriate AR model orders. It was found that orders in the range of 12 to 14 were suitable for idle task EEG while orders in the range of 20 to 24 were suitable for active task EEG. Finally, the study of prewhitened AR spectral analysis was useful in providing some insight into how well the AR model represented the information in the EEG signal. This study indicated that, given the selection of an appropriate model order, the AR model does represent much of the information contained in a short segment of EEG.

Single Trial Outlier Processing

Initial investigation into the single trial outlier processing method was carried out on the EEG data from one subject. It was shown, through conventional averaging analysis, that the cleaned active task EEG did not contain any significant motor related potentials. This result indicated that much of the motor related activity had been extracted from the active task EEG signal by the application of the cleaning process. By using the outlier information extracted from the active EEG, single trial outlier patterns were produced. These patterns had strong similarities to previous results using conventional averaging techniques over many trials of active EEG. By averaging active trial outlier patterns together it was demonstrated that much of the information was consistent across active trials whereas, in contrast, the average of idle case patterns showed that there was no significant information that was consistent across idle trials. In addition, this averaging also demonstrated that there was a significant amount of information in the active task patterns that was unique to individual trials which was lost when the patterns were averaged together. It is, therefore, expected that this same loss of information is occurring in the conventional averaging method of EEG analysis.

It was shown that consistent features in the active outlier patterns were strongly correlated with features in the thumb movement. This analysis demonstrated that there was a strong relationship between the information in the outlier patterns and events in the thumb movements on a trial by trial basis.

It was found in the initial investigations that dynamic time warping

(DTW) analysis provided the best quantitative information on the performance of the single trial processing method. Hence, DTW analysis was applied to all four of the subjects used in this initial investigation into the single trial processing method. Through the application of DTW, standard representative active single trial outlier patterns for each subject were obtained. The outlier patterns from both the active trials and the idle trials were warped against the standard patterns. With each warping an associated cost value was obtained which reflected how well the outlier patterns fit the standard pattern. These cost values revealed that there was a highly statistically significant ($p < .001$) difference between the idle and active mean costs across all four subjects with outlier patterns derived with GM2 parameter estimates. The cost values from the four subjects pooled together in a group, demonstrated that the mean of the actual population of mean differences between active and idle cases was highly ($p < .002$) unlikely to be equal zero.

Bayesian classification was applied to the warping cost values to classify active patterns versus idle patterns. It was found that with the cost of misclassifying an idle case as an active case set to be five times greater, 80% of the GM2 active patterns were classified correctly while only 1.7% of the idle cases were incorrectly classified as active. This analysis also demonstrated the superiority of utilizing GM2 parameter estimates because with the utilization of GM and LSQ parameter estimates the classification performance fell off dramatically.

Bayesian classification was also applied to the cost values obtained from using standard patterns from one subject on the outlier patterns from the other subjects. These results strongly indicated that the standard

patterns do cross validate on data that was not used in the construction of these patterns and that these patterns provide substantial intersubject reliability.

In conclusion, the pursuance of robust methods to deal with the ranging Gaussian properties of the EEG signal led to the development of a single trial processing method based on utilizing outlier information. The validity of this processing method to extract event related information from active task EEG has been established through the results obtained from the investigations undertaken in this thesis work.

5.2 Areas for Future Investigation

There are many areas involved with the possible further improvement of the single trial processing method that should be pursued in future investigations. Some of the most important recommended areas are:

1) Pursue models that will better represent the underlying signal since the processing method is fundamentally based on these models. This may involve the application of different types of models such as those based on orthonormal functions. Regardless of the type of model employed, the accuracy of the estimated parameters is an important issue. In terms of using GM estimation on AR models, further work could be carried out to improve the parameter estimates. One particular aspect to consider is the utilization of methods that are more robust than MEM estimation, such as L_1 based methods, to provide the starting estimates in the GM iteration procedure.

2) Study the effects of varying the segment size to determine the length that is best suited for the representation of the underlying EEG process. For instance, it may be found that the signal characteristics of the underlying process are changing slowly enough such that the modeling of longer segments would allow for an improved representation of the underlying process.

3) Further investigate the outlier detection process to determine ways in which the performance could be improved. One specific area to consider, as suggested by Martin [45], is to utilize a cleaning process that makes use of both forward and backward prediction in the estimation of the signal x_i.

4) Peruse alternatives to the current method of processing the outlier information. The current method is relatively unsophisticated in that it is simply a careful smoothing of the extracted outlier information. Investigation into other approaches of processing this information may prove to be beneficial in revealing additional information that may be contained in the outlier data.

Future empirical EEG experimentation should be carried out on the single trial processing method. The initial goals of these investigations should be to further validate the method on a new set of EEG data. A recommended paradigm would be to collect a training set of active trials for the construction of a standard pattern. Then collect a set of intermixed active and idle trials, perhaps at the discretion of the subject, on which the classification performance of the processing method could be further evaluated. Later goals of these investigations should be oriented to using the processing method to learn more about motor potentials, particularly the

motor potentials rom disabled persons, so that when utilized in a control application these potentials can be taken advantage of in the most appropriate manner.

Finally, investigations on making the single trial processing method work in real time must be undertaken. As it stands, the method is very computationally intensive. Although some improvements could undoubtedly be made in the efficiency of these computations, the most significant advances towards this goal would likely be in the implementation of some or all of the component processes in specialized hardware.

5.3 Significant Contributions

The signal processing method developed in this thesis work is a significant contribution. Modeling the underlying signal and then extracting the outlier content from the underlying signal is a unique approach to deriving very low level and relatively short event type information from an ongoing process. The path that led to the development of this processing method, also led to the understanding that ranging levels of Gaussianity in the EEG signal requires that serious consideration be given to the use of robust methods in the future application of various types of EEG signal processing. Also, a successful approach to the selection of appropriate AR model orders, the selection of appropriate EEG segment lengths and the assessment of the relative ability of AR models to represent the EEG signal were established through the studies on AR spectral analysis.

The ability to consistently acquire event related information from a single trial is an important contribution to the field of EEG signal

analysis. In addition, the neurological premise involving the way in which event related information is contained in the overall EEG signal is established as a viable model. This model should be considered when attempting to understand event related potentials and their relationship to ongoing EEG processes.

Finally, the work in this thesis, taken as a whole, represents an important contribution towards the ultimate goal of harnessing EEG signals for control applications. It overcomes perhaps one of the greatest obstacles by providing the framework for the extraction of useful information from single trial EEG.

REFERENCES

1. Westerkamp,J.J. and Aunon,J.I., "Optimum Multielectrode A Posteriori estimates of Single-Response Evoked Potentials", IEEE Trans. Biomedical Engineering, Vol. BME-34, No. 1, Jan., 1987, pp 13-22.

2. Heinze,H.J., Kunkel,H., "ARMA - Filtering of Evoked Potentials", Methods of Information in Medicine, 1984, Vol. 23, pp 29-36.

3. Vidal,J.J., "Real-Time Detection of Brain Events in EEG", Proceedings of the IEEE, Vol. 65, No. 5, 1977, pp 633-641.

4. Kornhuber,H.H. und Deecke,L., "Hirnpotentialanderungen beim Menschen vor und nach Willkurbewegungen, Dargestellt mit Magnetbandspeicherung und Ruckwartsanalyse", Pflugers Arch. ges. Physiol., 1964, 281:52.

5. Landau,B., Essential Human Anatomy and Physiology, Scott Foresman Company, 1976, pp 227.

6. Brunia,C.H.M. and Van den Bosch,W.E.J., "Movement-related Slow Potentials. I. A Contrast Between Finger and Foot Movements in Right-handed Subjects", Electroencephalography and Clinical Neurophysiology, 1984, 57: 515-527.

7. Grunewald,G. and Grunewald-Zuberbier,E., "Cerebral Potentials During Voluntary Ramp Movements in Aiming Tasks", Tutorials in ERP Research: Endogenous Components, A.W.K. Gaillard and W. Ritter (eds.), North-Holland Publishing Company, 1983.

8. Vaughan, Jr., H.G., Costa, L.D. and Ritter, W., "Topography of the Human Motor Potential", Electroencephalography and Clinical Neurphysiology, vol. 25, pp. 1-10, 1968.

9. Vaughan, Jr.,H.G., Bossom,J. and Gross,E.G., "Cortical Motor Potentials in Monkeys Before and After Upper Limb Deafferentation", Exp. Neurol., 1970, 26: 253-262.

10. Papakostopoulos,D., "A No-stimulas, No-response, Event-related Potential of the Human Cortex", Electroencephalography and Clinical Neurophysiology, 1980, 48: 622-638.

11. Persson,J., "Comments on Estimations and Tests of EEG Amplitude Distributions", Electroencephalography and Clinical Neurophysiology, 1974, 37: 309-313.

12. McEwen,J.A. and Anderson,G.B., "Modeling the Stationarity and Gaussianity of Spontaneous Electroencephalographic Activity", IEEE Trans. Biomedical Engineering, Vol. BME-22 No.5, 1975, pp 361-369.

13. McGillem,C.D., Aunon,J.I. and Childers,D.G., "Signal Processing in Evoked Potential Research: Applications of Filtering and Pattern Recognition", CRC Critical Reviews in Bioengineering, 1981, pp 225-265.

14. Persson,J., "Comments on 'Modeling the Stationarity and Gaussianity of Spontaneous Electroencephalographic Activity'", IEEE Trans. Biomedical Engineering, Vol. BME-24, 1977, p 302.

15. Weiss,M.C., "A New Test for EEG Gaussian Amplitude Distribution", Proc. IEEE Frontiers of Engineering in Health Care, 1979, pp 309-313.

16. Elul,R., in Progress in Biomedical Research, Fogeland L.J. and George,F.W., eds., Vol. 1, p 131, Spartan Books, 1967.

17. Elul,R., in Data Acquisition and Processing in Biology and Medicine, Enslein,K., ed., Vol. 5, p 93, Pergaman Press, 1968.

18. Elul,R., "Gaussian Behavior of the Electroencephalogram: Changes during Performance of Mental Task", Science, Vol. 164, 1969, pp 328-331.

19. Elul,R., Internat. Rev. Neurobiol., 15, p 227, 1972.

20. Seigel,A., "Stochastic Aspects of the Generation of The Electroencephalgram", J. Theor. Biol-92, 1981, pp 317-339.

21. Bernstein, S., Math. Ann., vol. 97, p. 1, 1927.

22. Anninos,P., Zenone,S. and Elul,R., "Artifical Neural Nets: Dependence of the EEG Amplitude's Probability Distribution on Statistical Parameters", J. Theor. Biol-103, 1983, pp 339-348.

23. Kay,S.M. and Marple,S.L., "Spectrum Analysis - A Modern Perspective", Proceedings of the IEEE, Vol.69, No.11, 1981.

24. Beamish,N. and Priestly,M.B., "A study of autoregressive and window spectral estimation," Appl. Statist., vol.30, no.1, pp.41-58, 1981.

25. Jansen,B.H., Bourne,J.R. and Ward,J.W., "Autoregressive Estimation of Short Segment Spectra for Computerized EEG Analysis", IEEE Trans. Biomedical Engineering, Vol. BME-28 No. 9, 1981, pp 630-641.

26. Box,G.E.P. and Jenkins,G.M., Time Series Analysis: Forcasting and Control, Holden-Day, 1970.

27. Kavech,M. and Cooper,G.R., "An Empirical Investigation of the Properties of the Autoregressive Spectral Estimator", IEEE Trans. Inform. Theory, Vol. IT-22, 1976, pp 313-323.

28. Andrews,D.F., "The Robustness of Residual Displays", in Robustness in Statistics, Launer,R.L. and Wilkinson,G.N. eds., Academic Press, 1979.

29. Chambers,R.L. and Heathcote,C.R., "On the Estimation of Slope and the Identification of Outliers in Linear Regression", Biometrika-68, 1, 1981, pp 21-33.

30. Brown,R.G., Introduction to Random Signal Analysis and Kalman Filtering, John Wiley & Sons, 1983.

31. Goodwin,G.C. and Sin,K.S., Adaptive Filtering, Prediction, and Control, Prentice-Hall, 1984, pp.224-288.

32. Jansen,B.H., "Analysis of biomedical signals by means of linear modeling," CRC Critical Reviews in Biomedical Engineering, vol. 12(4), pp.343-392, 1985.

33. Isaksson,A. and Wennberg,A., "Spectral properties of nonstationary EEG signals evaluated by means of Kalman filtering. Application to a vigilance test," in Quantitative Analytic Studies in Epilepsy, P. Kellaway and I. Peterson, Eds., Raven Press, New York, 1976.

34. Ulrych, T.J. and Bishop, T.N., "Maximum Entropy Spectral Analysis and Autoregressive Decomposition", Rev. Geophysics Space Phys., vol. 13, pp. 183-200, 1975.

35. Burg,J.P., "Maximum Entropy Spectral Analysis", Proceedings of 37th Meeting of the Society of Exploration Geophysicists, Oct. 31, 1967.

36. Anderson,N., "On the Calculation of Filter Coefficients for Maximum Entropy Spectral Analysis", Geophysics, Vol. 39, No.1, Feb., 1974, pp. 69-72.

37. Hurvich,C.M., "Data-Dependent Spectral Windows: Generalizing the Classical Framework to Include Maximum Entropy Estimates", Technometrics, Vol. 28, No.3, Aug., 1986, pp 259-268.

38. van den Bos,A., "Alternative interpretation of maximum entropy spectral analysis," IEEE Trans. Inform. Theory, vol.IT-17, pp.493-494, 1971.

39. Box,G.E.P. and Muller,M.E., "A Note on Generation of Normal Deviates", AMS, Vol. 28, 1958, pp 610-611.

40. Lewis,T.G. and Payne,W.H., "Generalized Feedback Shift Registor Pseudo-random Number Algorithm", Journal of the Association for Computing Machinery, Vol. 20, No.3, 1973, pp 456-468.

41. Smith,W.D. and Lager,D.L., "Evaluation of simple algorithms for spectral parameter analysis of the electroencephalogram," IEEE Trans. Biomed. Eng., vol.BME-33, no.3, pp.352-358, 1986.

42. Kleiner,B., Martin,R.D. and Thomson,D.J., "Robust Estimation of Power Spectra", J. Royal Statist. Soc., Ser.B, Vol. 41 No.3, 1979, pp 313-351.

43. Hogg, R.V., "An Introduction to Robust Estimation", in Robustness in Statistics, Launer,R.L. and Wilkinson,G.N. eds., Academic Press, 1979.

44. Huber,P.J, "Robust Estimation of Location Parameters", Ann. Math. Statist., 35, 1964, pp 73-101.

45. Martin,R.D. and Thomson,D.J., "Robust-Resistant Spectrum Estimation", Proceedings of the IEEE, Vol. 70, No.9, 1982, pp 1097-1115.

46. Ziemer,R.E., and Tranter,W.H., Principles of Communications: Systems, Modulation, and Noise, Houghton Mifflin Company, Boston, 1976, p.57 and p.224.

47. Martin,R.D., "Robust Estimation for Time Series Autoregression", in Robustness in Statistics, Launer,R.L. and Wilkinson,G.N. eds., Academic Press, 1979.

48. Martin,R.D., "Robust Estimation of Autoregressive Models", in Institue of Mathematical Statistics Directions in Time Series, Brillinger,D.R. and Tiao,G.C., 1978.

49. Jasper, H.H., The Ten-Twenty Electrode Placement System of the International Federation", Electroencephalography and Clinical Neurophysiology, vol. 10, pp. 371-375, 1958.

50. Isaksson,A., Wennberg,A. and Zetterberg,L.,"Computer analysis of EEG signals with parametric models," Proc. IEEE, vol.69, no.4, pp.451-461, 1981.

51. Martin,R.D. and Zeh,J.E., "Determining the Character of Time Series Outliers", Proceedings of the Amer. Statist. Assoc., Business and Economics Section, 1977, pp. 818-823.

52. Papoulis,A., "Minimum-bias windows for high-resolution spectral estimates', IEEE Trans. Inform. Theory, vol.IT-19, no.1, pp.9-12, 1973.

53. Birch,G.E., Lawrence,P.D., Lind,J.C. and Hare,R.D., "Application of Prewhitening to AR Spectral Estimation of EEG", IEEE Trans. Biomedical Engineering, In press, 1988.

54. Steiger, J.H., "Tests for Comparing Elements of a Correlation Matrix", Psychological Bulletin, vol. 87, no. 2, pp. 245-251, 1980.

55. Roberts,K., Lawrence.P., Eisen,A. and Hoirch,M., "Enhancement and Dynamic Time Warping of Somatosensory Evoked Potential Components Applied to Patients with Multiple Sclerosis", IEEE Trans. Biomedical Engineering, Vol. BME-34, No. 6, June 1987, pp. 397-405.

56. Glass,G.V. and Hopkins,K.D., Statistical Methods in Education and Psychology, Second Edition, Prentice-Hall, 1984.