National Library of Canada
Acquisitions and Bibliographic Services Branch
395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada
Direction des acquisitions et des services bibliographiques
395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Canada

Self-Adapting Algorithms for Seizure Detection During EEG Monitoring

Hao Qu
B.Sc., M. Sc.

Department of Electrical Engineering
McGill University
Montréal, Québec, Canada

A Thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Noember, 1994

ISBN 0-612-05778-X

Canada

Abstract

During long-term EEG monitoring of epileptic patients, seizure detection assists in selecting information important for diagnosis. We present a new generation of detection methods with self-adapting, more specifically patient-adapting, algorithms for two functions: (1) Reduction of false seizure detections (FSDs), thus increasing the sensitivity of detection. (2) Detection of seizure *onset*, thus providing a warning which is useful to patient and observers, allowing appropriate precautions and observations.

The self-adapting algorithm for reducing FSDs utilizes FSDs from one baseline monitoring session as template patterns. In subsequent sessions, events having a pattern similar to any template pattern are eliminated from the detection. A unique "similarity" measure was used to reflect the relation between two multichannel EEG patterns. An extensive test was done on twenty patients with 2600 hours of monitoring. Results show an average reduction in FSDs by 61% with a risk of missing seizures of 2.7%, comparing to the most commonly used method.

The self-adapting algorithm for seizure onset detection assumes one seizure has been recorded and uses that seizure and one set of non-seizure EEG to train a patient-specific classifier. By using special features and a modified nearest-neighbor classifier, this algorithm reached an onset detection rate of 100% with an average delay of 9.6 seconds after onset. The average false alarm rate was only 0.21/hour, making it an acceptable warning device. This test was done on 17 patients with 77 seizures.

In conclusion, our self-adapting algorithms make seizure detection more accurate and effective than was possible before. They are also efficient, practical and capable to work in real time.

Résumé

Au cours des sessions prolongées d'enregistrement EEG chez des patients épileptiques, la détection de crises aide à obtenir des renseignements diagnostiques. Nous présentons ici une nouvelle génération de méthodes de détection consistant en des *algorithmes auto-adaptatifs* ayant deux fonctions : (1) la réduction du nombres de fausses detections, permettant ainsi d'augmenter la sensibilité de la détection et (2) la détection du début des crises, procurant aux patients et observateurs un avertissement leur permettant de réagir et de d'ajuster leurs observations en conséquence.

L'algorithme auto-adaptatif pour la réduction des fausses detections utilise les fausses detections de la première session d'enregistrement comme pattern de base. Pendant les sessions suivantes, les événements dans lesquels on retrouve un pattern semblable à un des patterns de base sont éliminés de la détection. Une mesure unique de similarité a été utilisée pour déterminer la relation entre deux patterns apparaissant dans plusieurs voies d'EEG. Un examen approfondi a été mené auprès de vingt patients pour un total de 2600 heures d'enregistrement. Les résultats obtenus indiquent une réduction des fausses detections de 61% avec un risque de perte de vraies crises de 2,7%, en comparaison avec la méthode la plus couramment utilisée.

L'algorithme auto-adaptatif pour la détection du début des crises suppose qu'une crise a déjà été enregistrée et utilise cette crise ainsi qu'un échantillon représentatif de l'EEG sans crise pour former un *classificateur* particulier à chaque patient. À partir de caractéristiques spéciales et d'un *classificateur* modifié basé sur la méthode des plus

proches voisins, l'algorithme atteint un taux de détection du début des crises de 100% avec un délai moyen de 9,6 secondes après le début. Le taux moyen de fausses alarmes est seulement de 0,21 par heure, rendant la methode utilisable comme appareil d'avertissement. Nous avons evalué cette methode sur les EEGs de 17 patients ayant eu un total de 77 crises.

En conclusion, nos algorithmes *auto-adaptatifs* rendent la détection de crises plus précise et efficace qu'auparavant. Ils sont aussi plus efficaces, pratiques et capables de fonctionner en temps réel.

Acknowledgments

I would like to express my gratitude to my supervisor, Professor Jean Gotman, for his support, guidance and assistance. His encouragement and enthusiasm have contributed greatly to my understanding of the research process and the preparation of this thesis.

I am grateful to members of my thesis committee, Professor S. Morgera and Professor R. Poulsen for their invaluable advice.

Special thanks go to my colleagues and friends of the Montreal Neurological Institute for their support and friendship, in particular Benoit Farine. I am also thankful to technicians in the EEG lab at MNH for their help in the collection of patients' data, in particular I. Dube and N. Drouin.

But most of all, I could not accomplish anything without the love of my wonderful dad and mom, my caring sisters, Christine and Yanyun, and my supportive brother, Yong.

This work was supported by grant MA-10189 of the Medical Research Council of Canada to Dr. J. Gotman.

List of Figures

List of Tables

List of Abbreviations

ASZ	automatic seizure detection
EEG	Electroencephalogram
EEGer	Electroencephalographer
FFT	fast Fourier transform
FSD	false seizure detection
MAP	maximum a posteriori
MEZ	main energy zone
NN	nearest neighbour
PS	patient-specific
TSD	true seizure detection
QV	quality value

Claim of Originality

This dissertation contains following original contributions:

1. With respect to the methods used:
   a. A new definition of *similarity*: Different from other definitions of similarity, the new similarity defined in this proposal relies not only on the Euclidean distance of two patterns in a detection space, but also on probability distributions of their own class. A modification factor is also imposed to avoid negative effects of extreme cases. This measure is used to determine if a new pattern is similar enough to a previously stored pattern.
   b. The quantitative expression of *inter-channel information*: Traditionally, EEG channels are considered to be independent. In this proposal, a quantitative expression of distance between channels is presented. This makes it possible to take distance between channels as one of the dimensions in a detection space.
   c. *The average power in a main energy zone*: This feature is created to measure the concentration of energy of a section of EEG in the frequency domain. It reveals important characteristics of ictal EEG patterns. By using a logarithmic scale, this feature can make it easier to distinguish ictal EEGs from interictal EEGs.
   d. *Self-adapting algorithms*: The algorithms presented in this proposal can automatically adjust their classifiers to fit each patient's situation, as long as there is a known prototype of this patient's seizure as an input. These algorithms will be shown to be capable to reach the optimal

classifier for each patient without any human interaction. It makes these algorithms very efficient and practical.

2. With respect to the problems to be solved:

   a. The use of patient specific data to improve the performance of a method for the automatic detection of epileptic seizures has never been done.

   b. The very concept of an "early seizure detection device" capable of warning the patient or hospital staff that a seizure has just started is original and its implementation has never been attempted before, despite its obvious medical importance.

Table of Contents

Chapter 1:

Introduction

Since the electroencephalogram (EEG) was first recorded on humans by Berger in 1929 (Gloor, 1969), it has been playing a significant role in the diagnosis and the evaluation of treatments for brain related diseases (Gloor, 1985a). Among the main neurodiagnostic procedures, such as EEG, positron emission tomography, magnetic resonance measurements, EEG is the only one which can provide continuous recording of cerebral function over a long period of time. In studying epileptic patients, long-term EEG monitoring often provides information which is difficult or even impossible to obtain by any other means. For instance, seizures are the most important clinical and diagnostic features of the disease and they occur only rarely and unpredictably. Another epileptiform EEG abnormality, the spike, usually occurs more frequently than seizures, but it still occurs intermittently and unpredictably. EEG recording during a short period

of time is most often not sufficient to record a seizure or even a spike. Therefore, long-term EEG monitoring, which lasts from several hours to several days, even several weeks, can greatly increase the probability of recording seizures.

Monitoring an epileptic patient continuously over a long period of time, with an observer watching the patient's behavioral manifestations and EEG patterns simultaneously, is one way to capture all seizures and spikes. This procedure is, however, very labor intensive and expensive. Another way is to record the patient's EEG and behavior continuously but to review them after the monitoring is finished. The biggest disadvantage of this procedure is that it generates a large amount of data and review is therefore very time consuming.

During a long-term monitoring session of a patient with epilepsy, most data are redundant and sometimes irrelevant because most EEG patterns are repetitive and not specific to epilepsy. A data reduction method is therefore useful to decrease redundancy and extract specific features of interest for epilepsy. Seizure and spike detection methods are examples of data reduction methods useful in long-term monitoring of epileptic patients.

1 Motivations and Objectives of the Dissertation

The seizure detection method of Gotman (1982, 1990a) has been used for several years at the Montreal Neurological Hospital and in many other hospitals. Although this method can detect many seizures with a reasonable false alarm rate (Pauri et al. 1992), there is still much room for improvement. This dissertation aims at creating a new generation of seizure detection method having a much lower false alarm rate, a higher seizure detection rate and earlier seizure detection. A lower false alarm rate can lead to higher seizure detection rate because the detection sensitivity can thus be increased (Pauri et al. 1992). In addition, detecting seizures as soon as they start can give observers a chance to interact with patients to obtain information which cannot be acquired otherwise.

In the method of Gotman (1982, 1990a), a universal classifier is used to determine if a section of EEG is part of a seizure. Although there are some common characteristics to many seizures, seizures from different patients are different, as is the background EEG. Therefore, we think it is necessary to consider patient-specific information to create a different classifier for each patient so that the performance of seizure detection methods can be improved significantly. As a result, the main objectives in this dissertation are to use patient-specific information to develop a seizure detection method with (1) a much lower false seizure detections rate than the current method and consequently a better seizure detection sensitivity. (2) the ability to detect seizures early in their development.

1.1 Reduction of False Seizure Detection

The need for a low false alarm rate in long-term EEG monitoring is obvious: to reduce uninteresting and redundant data. At the Montreal Neurological Hospital, five patients are currently under long-term EEG monitoring every day; on average, a patient is monitored during 12 days. The false seizure detections (FSDs) cause unwanted data to be stored and therefore reduce the effectiveness of a seizure detection method. The FSD rate of the system of Gotman (1982, 1990a) was estimated to be 0.84/hour for scalp electrode patients and 1.35/hour for depth electrode patients (Gotman, 1990a). In another study (Pauri et al. 1992), the same seizure detection system was evaluated. Depending on the detection threshold, the average FSD rates were between 2.70/hour and 5.38/hour. It was confirmed in that study that a higher detection threshold results in a lower false alarm rate, as well as a lower seizure detection rate.

Whereas the FSD rate averaged over many patients is acceptable, it can become significantly higher and reach an unacceptable level in some patients. This usually happens when one or a small number of EEG patterns occur repeatedly in a given patient and cause many FSDs during several consecutive monitoring sessions. These FSD patterns vary from one patient to another. Thus, they can be called patient-specific FSD patterns.

As indicated in the study of Pauri et al. (1992), it is difficult to reduce FSD rates without reducing seizure detection rates because, by changing detection thresholds, lower FSD rates can only be achieved at the expense of lower seizure detection rates. For patient-specific FSDs, however, it may be possible to reduce FSDs with little effect

in detecting seizures. Because a patient is typically monitored for several days, we are proposing a monitoring system that can be "taught" to recognize and stop detecting EEG patterns clearly identified as FSDs early in the monitoring. However, the reduction of this kind of patient-specific FSDs should not result in a reduction in the ability to detect seizures. Reduction in false detections without major effects in detecting seizures is possible because we are no longer classifying all seizures from all non-seizures in all patients. Instead, we focus on eliminating EEG patterns similar to those pre-defined FSD patterns from each patient.

1.2 Seizure Onset Detection

The early part of seizures is always interesting to electroencephalographers (EEGers) since it reveals important information about the location in the brain of the epileptic focus. Moreover, interaction with a patient during the early part of a seizure, such as testing the patient's consciousness, can help to determine the type of seizure. Since observers are not watching a patient all the time during long-term monitoring, a warning signal when a seizure occurs would be very helpful in this situation so that observers can take appropriate actions.

Although we talk about a method of seizure onset detection, we actually mean detecting seizures *early*. This is because there is no specific pattern called "seizure onset" in all seizures. When a seizure occurs, it is possible to detect it only when enough information is processed, which may take at least a few seconds.

It is difficult to detect seizures early because seizures from different patients are different and the early pattern of a patient's seizure may look similar to non-seizure patterns in another patient. As a result, no on-line seizure onset detection method has been developed. Seizure detection methods only attempt to capture *prominent* seizure patterns rather than *early* seizure patterns. For instance, the seizure detection method of Gotman (1982, 1990a) captures seizures by using information over a long period of time before and after the current detection window. It is not appropriate to detect seizures early since, at the time a seizure pattern is found, the method has to wait a dozen seconds to increase the probability that the detection is correct.

It has been observed that most patients have one or sometimes two or three types of seizures which are repetitive. Seizures of each type are very much alike in terms of EEG patterns, including the early part of seizures. Although early EEG patterns in some seizures of a patient may be similar to some background EEG patterns of other patients, these seizures are usually different from the background EEG of the patient in whom they occur. We can therefore use the concept of "patient-specific seizure onsets".

According to the above observations, it is possible that a seizure can be detected soon after onset if (1) a template of this type of seizure was acquired in advance; (2) a match could be made between the template and the seizure. Detection of this kind of patient-specific seizure onset must not, however, cause frequent false alarms because they would annoy patients and observers, and all the detections would eventually be ignored.

2 Organization of the Dissertation

The dissertation has five chapters. Chapter 1 is an introduction. Motivations and objectives of the dissertation are presented.

In chapter 2, we first review some basic notions about EEG signals, EEG recording systems and abnormal EEG patterns, including spikes and seizures. The previous work on long-term EEG monitoring and automatic detection of spikes and seizures is then reviewed. Finally, some pattern recognition theory and applications are described, as they may relate to our applications.

In chapter 3, we present a pattern recognition system designed to reduce false seizure detections by automatically adapting to each patient's situation. A unique "similarity" measure is explained in detail. It reflects the relation between two EEG patterns in a detection space. A comprehensive evaluation of performance was done and is described in this chapter.

In chapter 4, the method for detecting seizures early is presented. After explaining the algorithm, we present comprehensive tests of detection rate, detection delay and the false alarm rate, showing promising results.

The last chapter, chapter 5, summarizes the work and provides possible directions for further development of seizure analysis and detection.

Chapter 2:

Literature Review

In this chapter, we will describe first the features of the EEG in epilepsy, then give a description of long-term EEG monitoring systems and of existing spike and seizure detection methods. Finally, we will present a review of some pattern recognition methods that are related to this project.

1 Characteristics of the EEG in Epilepsy

The EEG is a random and stochastic signal (Lopes da Silva 1987a). It is the recording of cortical neuronal activity. The recording can be done near the brain (scalp EEG), directly on the brain (cortical EEG) and within the brain (depth EEG) (Sharbrough 1990).

1.1 EEG Signal

The main generators of the EEG are cortical neurons, more particularly pyramidal neurons (Gloor 1985b). The amplitude of EEG signals depends on the distance between generators and a recording electrode, and on the tissues between the generators and the electrode. Since depth electrodes are usually closer to cortical neurons than scalp electrodes, the amplitude of EEGs recorded by depth electrodes is usually larger than by scalp electrodes. In addition, EEGs recorded by depth electrodes have less artifact than EEGs recorded by scalp electrodes because non-cerebral sources of electrical activity are also recorded at the scalp. This is particularly the case for electrical activity from scalp muscles and from the movement of eyeballs (The potential of the anterior part of the eyeball is different from that of the posterior part and its movement thus creates an electrical field). When EEGs are recorded with scalp electrodes, the amplitude is of the order of 20μV to 100μV, which is approximately one hundredth of that of the electrocardiogram (Morris and Luders 1985).

1.2 EEG Recording System

Electrodes in different locations pick up signals from a same source differently because the amplitude of the EEG is determined by tissues between electrodes and generators. In our study, we only study patients with scalp electrodes and depth electrodes. The different EEG recording systems for the different kinds of electrodes are described below.

1.2.1 Scalp Recording

In order to sample scalp electricai fields adequately, one must systematically place an appropriate number of electrodes (Reilly 1987; Sharbrough 1990). The distance between electrodes should reflect the spatial frequency of the EEG. In many situation 16 electrodes are sufficient, but it has been shown that for some patterns, up to 128 may be required (Gevins, 1994). Placing more electrodes results in a higher spatial resolution because the distances between electrodes decrease. However, placing more electrodes also results in more data to be processed and stored in the same period of time. The computation and storage capacity of the computer becomes the decisive factor in the determination of the number of electrodes to be placed. For instance, with the advance of computer technology, the number of electrodes in each patient has been increased from 16 channels to 32 channels and now even to 64 channels at the Montreal Neurological Hospital. The most common system of electrode placement for scalp electrodes is the international 10-20 system (Jasper 1958) which is used in our studies and the placement of the 19 electrodes is shown in figure 2-1. A more recent system (Chatrian et al 1988) has 81 electrodes placed a closer intervals, but it is rarely used because of its large number of electrodes.

1.2.2 Depth Recording

When a patient's epileptic focus cannot be determined from scalp recordings, the patient may undergo implantation of intracerebral electrodes because the location of the focus may be too far away from the scalp. The depth electrodes record the EEG directly from inside the brain. They are placed as close to the suspected epileptic focus as possible. Since every patient's epileptic focus is different, there is no standard system of depth electrode placement, which is therefore individualized. Figure 2-2 is an example of the electrode placement in a patient (Olivier et al. 1987).

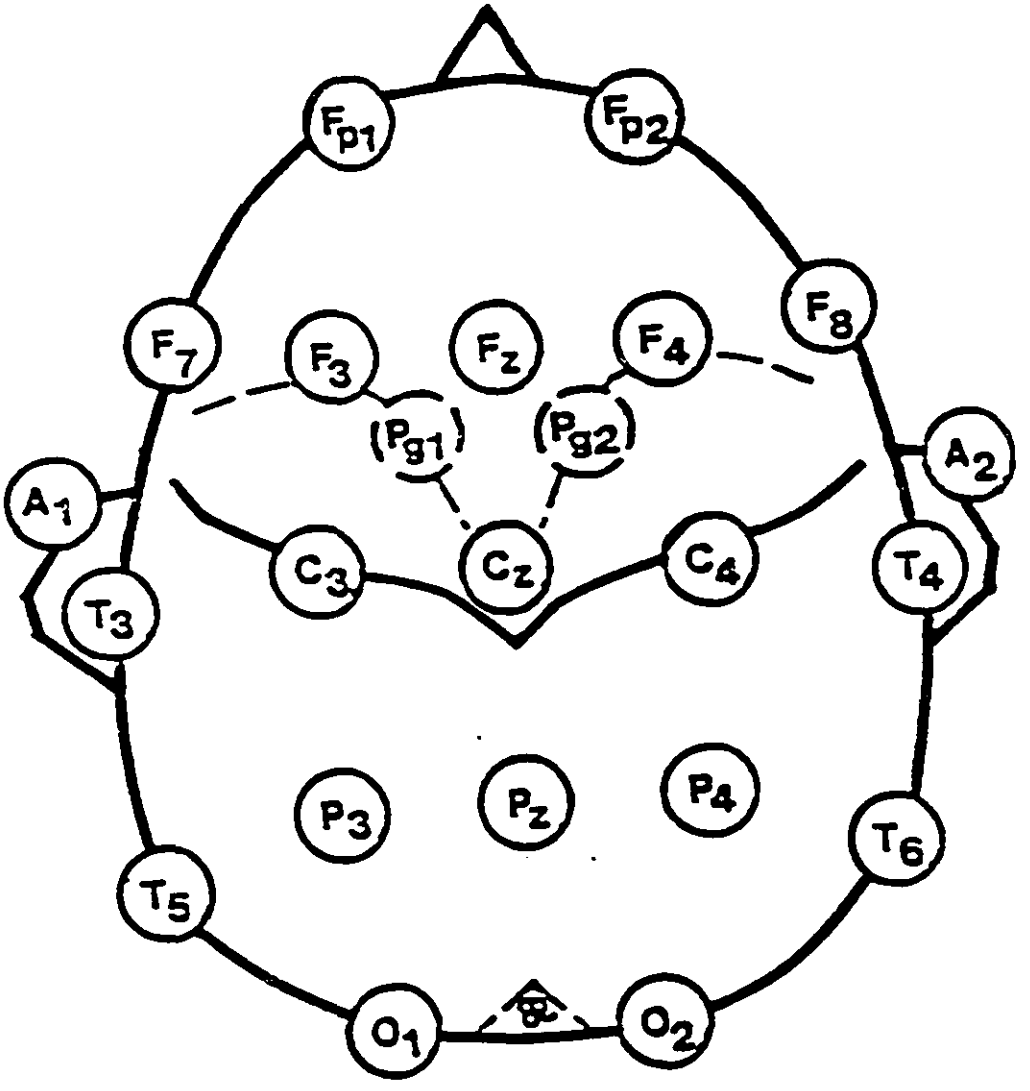


Figure 2-1: The international 10-20 system for determining locations of scalp electrodes. Schematic representation of the head seen from the top.

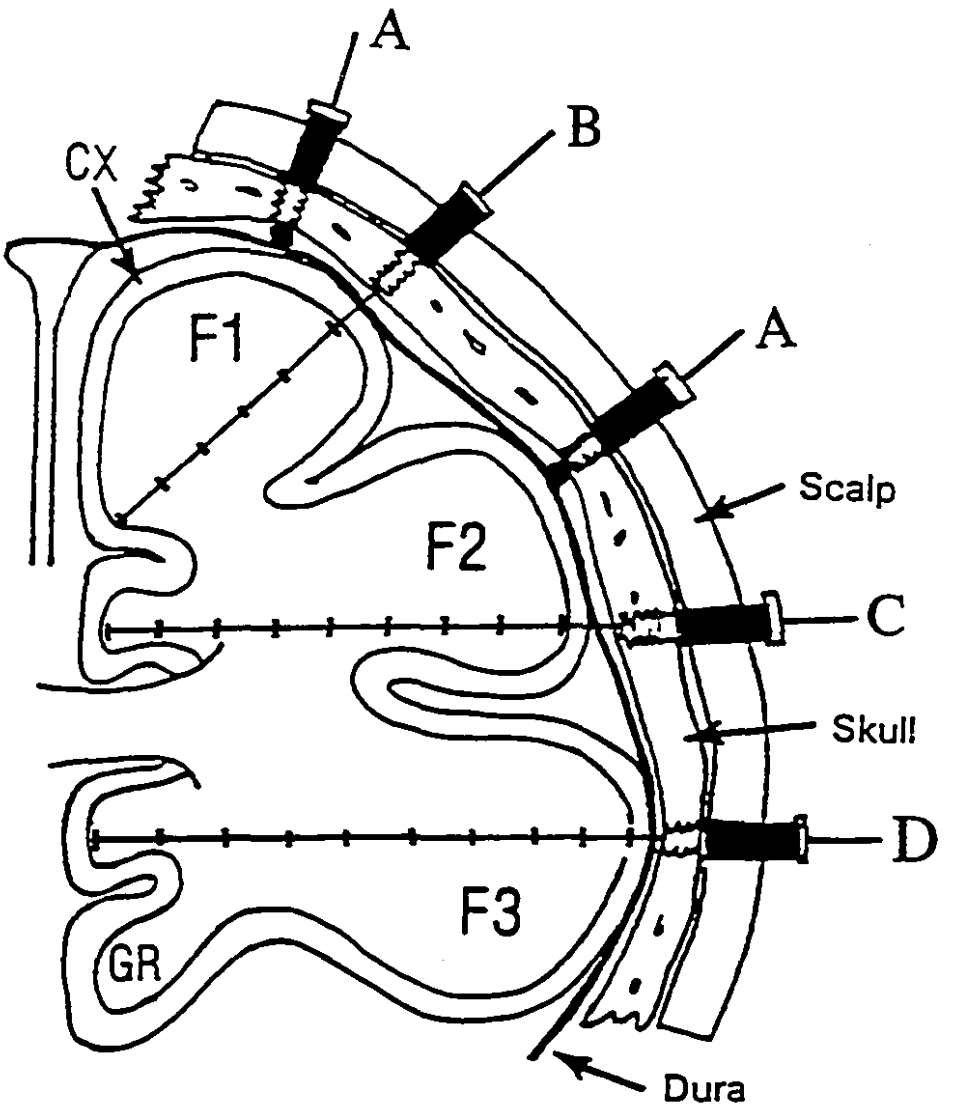


Figure 2-2: An example of the location of depth electrodes in a patient: Depth electrodes are implanted into the brain to record the EEG signal from regions inaccessible by scalp recordings and to be free of muscle artifact. The exact location of electrodes is determined by a frame with fixation pins and by magnetic resonance imaging. A are an epidural electrodes. B, C and D are intracerebral multicontact electrodes.

1.3 Methods of EEG Analysis

There are many methods to analyze the EEG signal. Visual interpretation by EEGers is the one that has been used since the EEG was first recorded. Computerized EEG analysis has developed rapidly in the last several decades. We review briefly below some of these methods.

1.3.1 The EEG Montage

In order to delineate the spatial distribution of a changing electrical field, an orderly arrangement (called a montage) of multiple channels is required. According to IFSECN (1974), the montage is "the particular arrangement by which a number of derivations are displayed simultaneously in an EEG recording." Montages can be divided into two categories: referential and bipolar. The referential montage displays the difference in potential between each electrode and a common reference electrode. Each channel in a bipolar montage displays the difference in potential between two different electrodes. Figure 2-3 is an example of montages for scalp electrodes recordings, while figure 2-4 is the one for depth electrodes recordings.

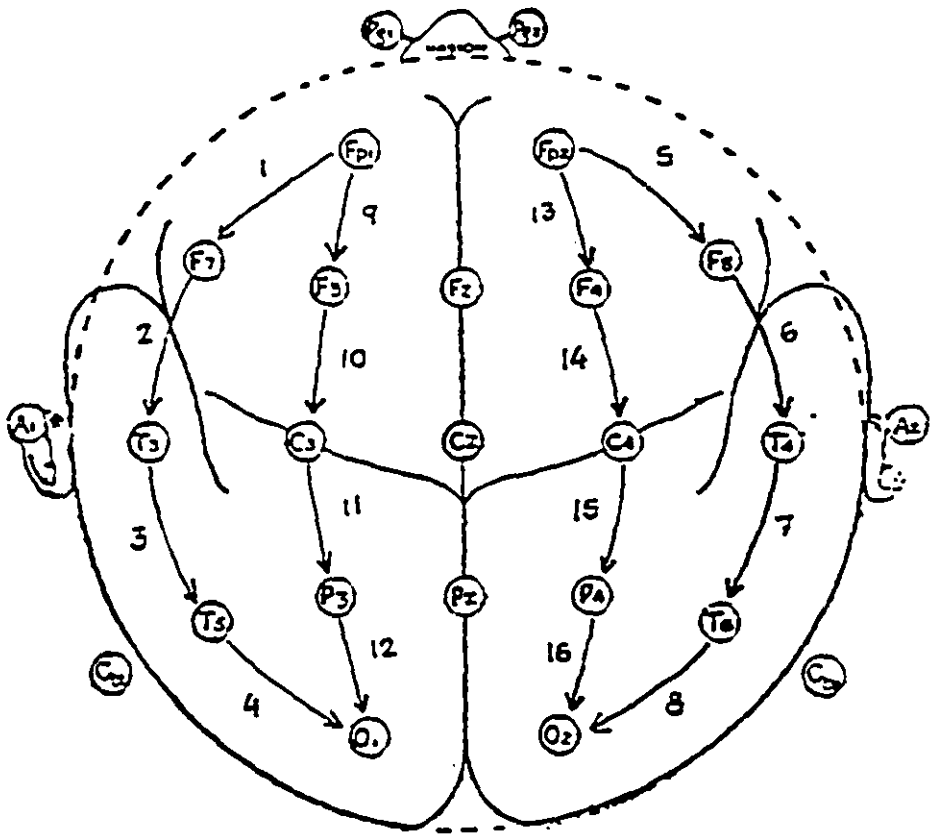


Figure 2-3: Bipolar montage for scalp electrode recording: Every channel of EEG is a subtraction of activity from two electrodes and the recording system is based on the international 10-20 system. For instance, channel 1 is the difference between the potential at electrodes Fp1 and F7.

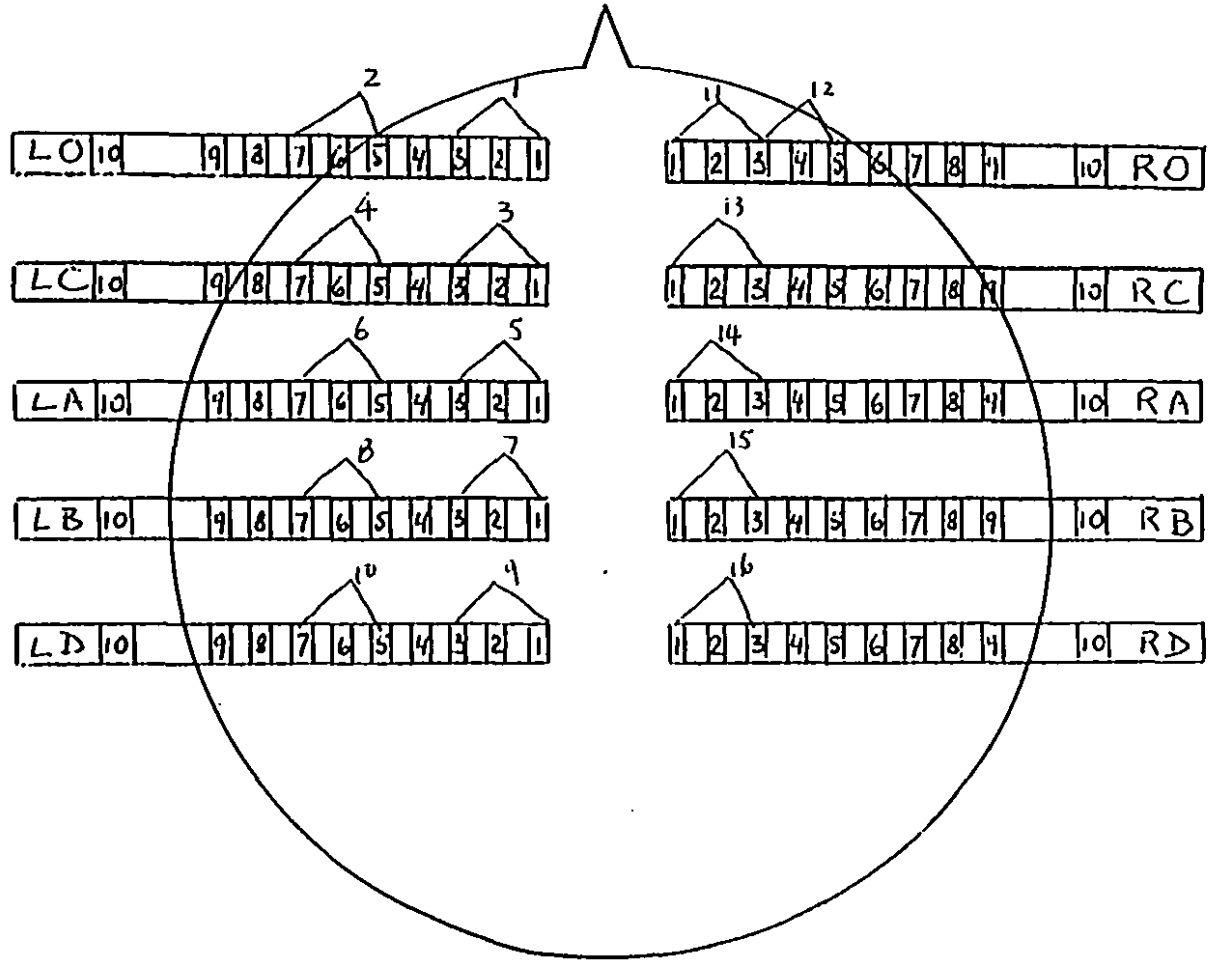


Figure 2-4: A montage for depth electrode recordings: This bipolar montage is specific for a particular patient. Every patient has a specific location of electrodes. LB, RB, LA, RA, etc. are the names of electrode stands. The number at each electrode indicates the actual contact location. The montage is a combination of pairs of electrodes.

1.3.2 Traditional Interpretation of the EEG

Traditionally, the EEG is interpreted visually by experienced EEGers for clinical purposes. This procedure has been used since the EEG was first recorded. For instance, the sleep EEG is usually classified into sleep stages so that a person's sleep normality can be determined. For epileptic patients, the EEG is usually reviewed visually to detect epileptic patterns, like spikes and seizures, so that the type and the origin of seizures can be determined. Some patients with migraine may have their EEG interpreted to rule out cerebral pathology (Niedermeyer 1987d). The interpretation of the EEG is also used in psychiatric disorders to determine states of mental retardation, attention deficit disorder, behavior disorders. (Small 1987). Visual interpretation of the EEG is, however, qualitative and subjective. When the recording is long, lasting many hours or even several days, the recording and interpretation becomes tedious, labor-intensive and expensive. Therefore, the computer is utilized to fulfill some parts of EEGers' tasks. Moreover, the computer can perform tasks which are not included in the traditional EEG analysis, such as making precise measurements of time relations between events in several channels, performing digital filtering without distortion, and statistically analyzing features of an EEG (Gotman 1990b),.

1.3.3 Computerized EEG Analysis

Since computerized EEG analysis is objective and quantitative, it has been used widely in extracting statistical features, spectral analysis, digital filtering and display of results.

Computerized analysis can extract statistical features of the EEG. For instance, parametric representation of the EEG, such as the autoregressive model for the study of epileptic patterns (as reviewed by Lopes da Silva 1987b), used a small number of

parameters to represent a section of EEG. Mimetic methods, like those of Gotman (1976) and Frost (1987), attempt to mimic the process of human interpretation of the EEG by finding waves comprising the various known EEG patterns: alpha waves, spindles, seizures and spike.

Spectral analysis is the quantitative method that has been used most commonly. Since the Fast Fourier Transform (FFT) was developed by Cooley and Tukey in 1965, it has been commonly used in the analysis of EEG background. For instance, Oken and Chiappa (1988) used spectral analysis to study variability among different frequency features, such as mean frequency, peak frequency and average power in background EEG. Techniques to perform spectral analysis of EEG background activities, as well as primary and postprocessing techniques currently used in clinical and experimental settings were reviewed by Dumermuth and Molinari (1987). Computerized EEG analysis can also quantitatively present the correlation between two channels of the EEG by computing cross-correlation (Gevins 1987b) and coherence (Brazier 1972; Gotman 1983).

Digital filtering is another utilization of computerized EEG analysis. Because digital filtering can filter the EEG without phase distortion, for instance with a finite impulse response filter, it has been used by Urbach and Pratt (1986) to distinguish the superposition of different waves which have distinctive frequency bands in the study of auditory evoked potential signals.

The computer is also used to display results of analysis of several EEG channels on a map of the head or brain so that spatial relationships can be easily observed. This kind of topographic display has been used to present ratio of EEG activities (Gotman 1981), number of spikes (Gotman 1976) or degree of normality (Matousek and Petersen 1973).

1.3.4 Segmentation of the EEG

The EEG is a non-stationary signal (Barlow and Dubinsky 1985). However, it can most often be considered time-invariant within 20 seconds (Lopes de Silva 1987b). In computerized EEG analysis, the EEG is divided into sequences of analysis windows so that computations can be done in each window. These windows are also called segments, or epochs. The EEG inside an epoch has to be considered time-invariant and wide-sense stationary so that some transforms, such as the Fourier transform, can be performed (Shanmugan and Breipohl 1988). Segmentation of the EEG should also take into account the characteristics of the particular EEG being analyzed. For instance, in the analysis of the EEG when an epileptic patient is not having a seizure, an epoch of under 10 seconds is adequate because the EEG in each epoch can be considered time-invariant (Cohen and Sances 1977). In the study of sleep, epochs of 30 seconds are commonly used (Barlow 1985) because most sleep patterns do not have a rapid change in a period of 30 seconds. It is not the case for studying the EEG during seizures. Because patterns during seizures, especially at seizure onset, have a more rapid change than interictal patterns, it is then better to use a short epoch so that the time-invariance and wide-sense stationarity can be considered appropriate. While most EEG analysis uses fix epoch lengths, Praetorius et al. (1977) used an autoregressive model to do automatic segmentation of the EEG. In that method, the EEG is first divided into sections lasting a few seconds. Each section is then described using an autoregressive model and compared with the subsequent one to determine if two sections can be accurately represented by the same model. If they do, these two sections are combined into one. The same procedure continues until one section of the EEG is no longer represented by the same autoregressive model. This method thus divides the EEG into epochs of similar characteristics. Although this segmentation method is good, it requires a lot of computations and therefore it can hardly be implemented to work on-line.

Moreover, this method may create different length of segments in different channels and therefore time alignment between segments in different channels becomes a problem.

1.4 Abnormal EEG Patterns in Epilepsy

The principal tasks of a clinical EEGer are to recognize particular waveform of diagnostic significance, such as spikes, sharp waves and delta waves, and to identify the likely location of their generators within the brain (Gloor 1985a). The scheme in figure 2-5 describes the overall way in which EEGers classify the EEG of a patient with epilepsy. Although in our studies only seizures are dealt with, characteristics of spikes are briefly described because spikes and seizures are highly related and are the patterns most characteristic of epilepsy.

1.4.1 Spike

The spike is an abnormal EEG waveform specific to epilepsy. Spikes occur randomly, with a rate ranging from one in a few seconds to one in a few hours. Wave morphology and spatial distribution of spikes can reveal information about the type of spikes and their sources. As a result, they are very important in the clinical diagnosis (Niedermeyer 1987b).

According to IFSECN (1974), the spike is "a transient, clearly distinguished from background activity, with pointed peak at conventional paper speeds and a duration from 20 to under 70 milliseconds, i.e., 1/50 to 1/14 sec, approximately. Main component is generally negative relative to other areas. Amplitude is variable." A spike can exist isolated, or several spikes may group together to form a burst, like 3/sec spikes-and-waves, which usually lasts a few seconds.

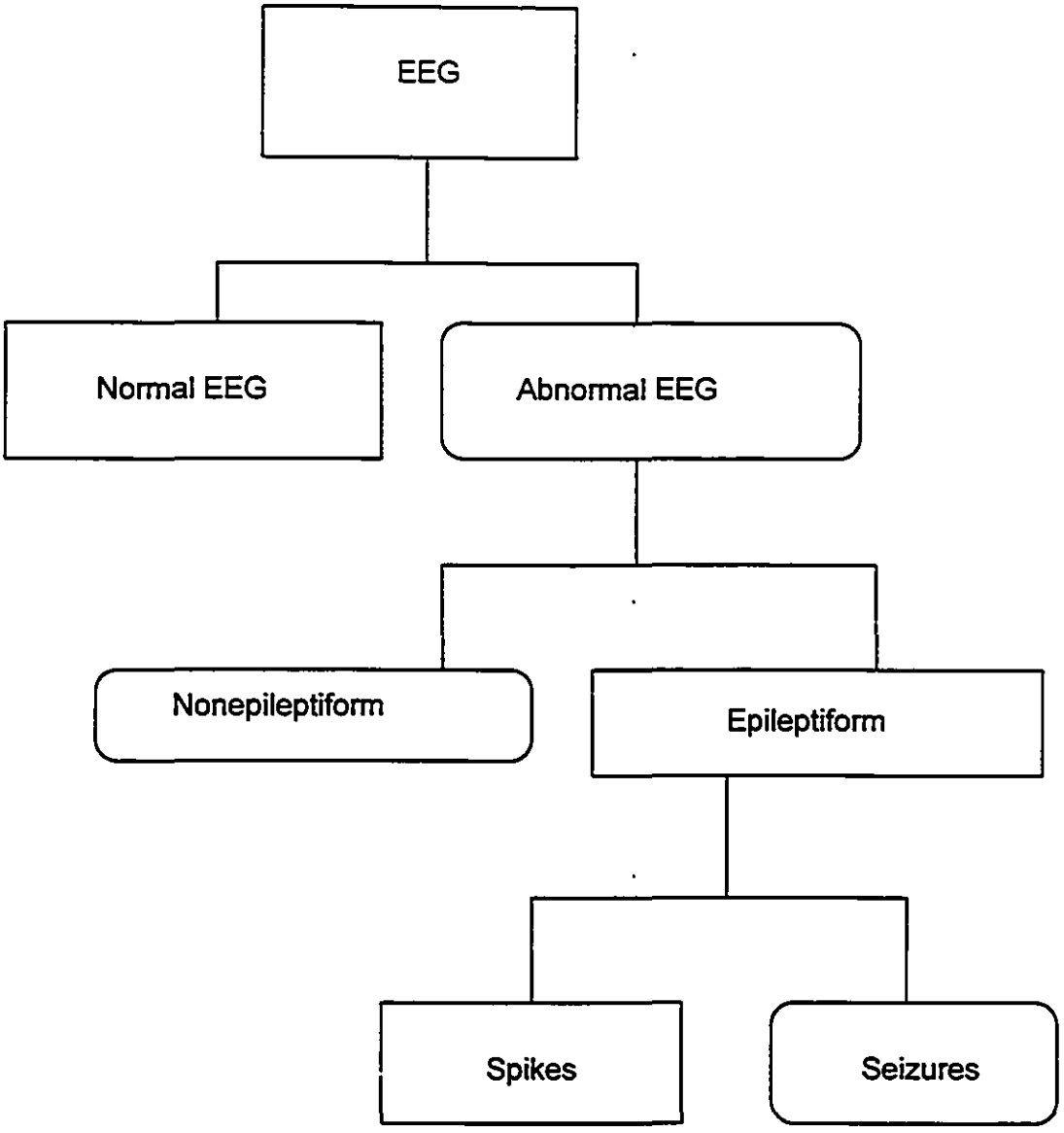


Figure 2-5: Scheme of EEG classification for epileptic patients by EEGers.

1.4.2 Ictal EEG

The ictal EEG is the EEG during seizures. Epileptic seizures are abnormal reactions of the brain caused by a number of diseases. The entire brain or parts of it may be involved and the extent of involvement largely determines the type of seizure (Niedermeyer 1987c). A seizure can last from a few seconds to dozens of minutes, but usually lasts about one or two minutes. There are two aspects of a seizure which are of interest to us: clinical manifestations and ictal EEG patterns. Most seizures have both clinical manifestations, which may include limb movements, absence of consciousness, screaming, or staring, and ictal EEG patterns, such as a sudden decrease in amplitude, or a sudden increase in amplitude and frequency . However, some seizures may have only clinical manifestations but no change in EEG from the background. This may be caused by the fact that the electrodes are too far away from the epileptic focus, the part of the brain where the seizure is generated. Some seizures may not have clinical manifestation but have significant epileptic discharges in the EEG. This kind of seizure is called "subclinical seizure". When a patient has focal seizures and seizures do not spread to areas of the brain that cause clinical manifestations, a subclinical seizure occurs.

In our studies, we only analyze the EEG. Therefore, all seizures in our studies have ictal EEG patterns. These seizures may or may not be accompanied by clinical manifestations. According to the IFSECN (1974), ictal EEG patterns are defined as: "phenomenon consisting of repetitive EEG discharges with relatively abrupt onset and termination and characteristic pattern of evolution, lasting at least several seconds. The component waves or complexes vary in form, frequency and topography. They are generally rhythmic and frequently display increasing amplitude and decreasing frequency during the same episode. When focal in onset, they tend to spread subsequently to other areas."

1.4.3 Interictal EEG

The EEG of an epileptic patient at a time when no seizure is taking place is called the *interictal* EEG. It is also referred to as background EEG. The interictal EEG comprises normal patterns, and abnormal patterns, such as spikes.

1.4.4 Relation between Spikes and Seizures

Although spikes may originate from several regions, most spikes are caused by the same epileptic foci as those which cause seizures. Therefore, the spatial distribution of spikes may give useful information about the localization of the focus. For instance, if a patient has 90% of his spikes in the right temporal region, it is likely that his seizures originate from the right temporal lobe. However, this is not always true and it is rarely possible to localize the epileptic focus of a patient by simply using one single measurement.

Spikes are most often isolated but sometimes can appear in bursts. When this kind of burst lasts less than two or three seconds, it will not be identified as a seizure according to the definition. When spike bursts last long enough, however, they could be called seizures. For instance, several 3/second spike-and-waves are considered as spikes only. When this kind of spikes and waves last longer than several seconds continuously, it could represent the EEG pattern of a so called "absence" seizure.

1.5 Characteristics of Ictal EEG

There is no stereotype pattern for all seizures. Most seizures however have some common characteristics, such as a rhythmic discharge of large amplitude or a low amplitude desynchronized EEG at onset, and repetitive spikes and irregular slow waves

later (Gotman 1985; Niedermeyer 1987c). The definition of a seizure still remains vague because seizure patterns are highly variable and some seizures may not have these characteristics while some may have only some of them. For instance, in terms of ictal EEG manifestations, the grand mal seizure is initiated by an abrupt loss of voltage of a few seconds duration. Then in patients with primary generalized epilepsy, several generalized bursts of polyspike-wave complexes may follow. Otherwise, rhythmic activity at about 10Hz with rapidly increasing amplitude will then dominate the EEG. About 10 seconds after the onset of a seizure, slower frequencies are noted, gradually slowing into a frequency range between 2Hz to 7 Hz. The EEG then becomes postictal with slow waves. For absence seizures, the ictal EEG is characterized by the generalized synchronous 3Hz spike wave discharge. This kind of spike wave discharge is maximal over the frontal midline and starts at a rate of around 4Hz, quickly slowing down to 3Hz and during the final phase of the seizure, slowing to about 2.5Hz. Onset and termination are abrupt and the seizure may be preceded and immediately followed by normal EEG activity (Niedermeyer 1987c).

Nevertheless, ictal EEG patterns can be differentiated from the interictal EEG patterns of a given patient (Sharbrough 1993). This characteristic is the most essential one in the identification of a seizure and was used in our studies. In addition, a patient may have several types of seizures. For a given type of seizure in a patient, the same ictal EEG patterns appear in the same electrodes because the ictal EEG comes from the same generators.

There have been documented disagreements between readers of the same EEG record for the presence of epileptic transients, as well as inconsistency in the same EEGer (Ktonas 1987). The same thing happens to the recognition of seizures, especially subclinical seizures, and the determination of seizure onset. This is because spikes and seizures are not well defined morphologically and leave many aspects of the

interpretation to EEGers. This increases the complexity of the problem of automatic detection.

2 Long-term EEG Monitoring

Long-term EEG monitoring, which is usually combined with video and audio monitoring, is one of the most important tools in the diagnosis of epilepsy and the understanding of its mechanism. This is because long-term EEG monitoring allows recording during long periods of time and thus allows capturing unpredictable EEG events, such as spikes and seizures. No other measurements, even with the most advanced methods like positron emission topography (PET) or magnetic resonance imaging (MRI), can offer the continuous monitoring provided by the EEG (Gloor 1985a). Long-term EEG monitoring can provide answers to several important questions in the diagnosis and therapeutic management of seizures. The first one is: does the patient have epilepsy or another episodic disturbance? This is one of the most important diagnostic questions. The second is: what kind of epilepsy does the patient have? Correct determination of the type of epilepsy can ensure that the optimal medical therapy is given to the patient. The third one is: which part of the brain is the focus of this patient's epilepsy? The answer to this question is essential to neurosurgeons who make a decision regarding which part of the brain is to be removed to stop medically intractable seizures. Long-term EEG monitoring can also be used to compare the result of a treatment by measuring seizure and spike frequency, as well as background EEG, before and after the treatment.

Although long-term EEG monitoring could theoretically be implemented as early as the EEG was first discovered in 1929, it has not been possible in practice until 1960s, when it was first used in sleep research. Long-term EEG monitoring is now mostly used in monitoring epileptic patients because epileptic events, like spikes and seizures, occur rarely and unpredictably. A typical EEG monitoring session lasts 10 to 24 hours and a patient may undergo several monitoring sessions. There are several drawbacks in the traditional EEG recording procedure, consisting of the EEG recorded on paper charts

and visually reviewed by EEGers. The first drawback is that it generates a large amount of paper. The second is that such a long EEG takes a long time to review and most of the recording is redundant. This makes it difficult for EEGers to be attentive to details all the time and therefore the accuracy of the interpretation suffers. The third drawback is that in order to catch all seizures, including those with electroencephalographic but no or very minor behavioral manifestation, it is necessary to have a person watching the patient and the EEG continuously. This is a labor intensive and expensive procedure. As a result, it would be useful to have a system which can replace this labor intensive procedure by detecting and presenting only the EEG patterns of diagnostic significance.

A typical long-term EEG monitoring system consists of electrodes, amplifiers, A/D conversion and computer software for recording, storage and processing. In addition, a video-audio recording system is used to record the patients' behavior simultaneously. Synchronization between EEG recording by computer and video-audio recording is realized by a system allowing to write on the video signal the time coming from the computer (Gotman et al. 1985).

At the Montreal Neurological Hospital, a long-term EEG monitoring system has been used for many years. The configuration is shown in figure 2-6. This system is also used in all data collection and analysis for the studies presented in this thesis.

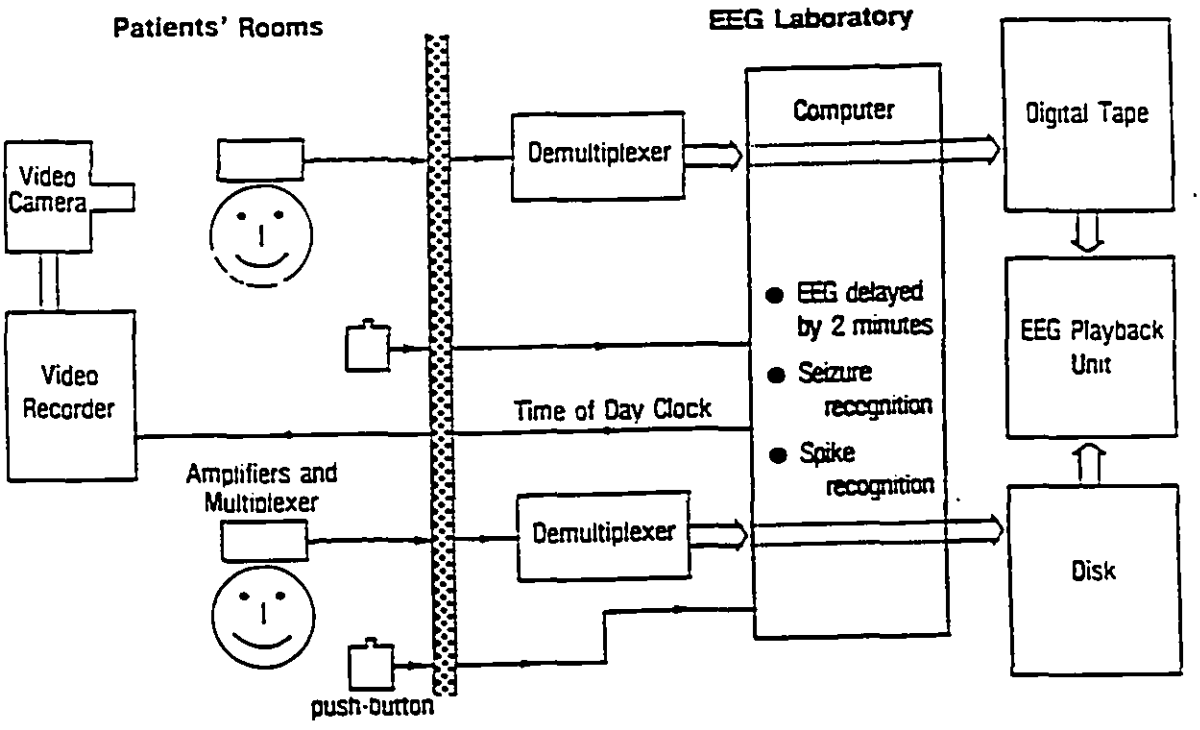


Figure 2-6: Block diagram of monitoring system at the Montreal Neurological Hospital. This system has been used in the hospital for many years. Today, the recording media have been changed from digital tapes to optical disks.

There are currently not many other types of long-term epilepsy EEG monitoring systems. Ives and Woods (1975) developed an ambulatory cassette EEG system to perform monitoring in routine laboratory EEGs and intensive inpatient monitoring. The system is portable, less cumbersome and also usable for monitoring outside the hospital.

A seizure and spike detection method has been used to analyze the EEG recorded in an ambulatory cassette by Ives (1994). The system can reduce a 24-hour recording into a much shorter length (<1 hour) so that EEGers can review only the EEG with spikes and seizures.

3 Automatic Spike and Seizure Detection

Seizure and spike detection methods are data reduction methods. During long-term monitoring, most of the EEG is not interesting to EEGers and therefore can be discarded. Only the EEG related to epileptic activities, spikes and seizures, is interesting and therefore should be retained. As a result, the first priority of a detection method is to keep as many spikes and seizures as possible. In other words, the detection rate should be as high as possible. False alarms will cause non-interesting data to be retained. Thus, false alarms should be as few as possible. Given the complexity of EEG patterns in epileptic patients, a trade-off between seizure detection rate and false alarm rate has to be considered.

3.1 Detection of Spikes

There are numerous spike detection methods, according to Ktonas (1987). They fall into five categories:

1.Orthogonal transform: The Fourier analysis was used by Principe and Smith (1982) for band filtering to separate sustained 3Hz spike and wave from background EEG.

2.Correlation methods, including matched filtering: Barlow and Dubinsky (1976) calculated a correlation coefficient between preselected EEG templates containing spikes and sharp waves and the EEG trace in order to detect new spikes and sharp waves.

3.Inverse filtering: Praetorius et al. (1977) used an autoregressive model to filter the EEG. If the prediction error exceeds a preset threshold, it means this section of EEG is no longer considered stationary. This may indicate a transient signal, which could be a spike or sharp wave.

4.Waveform decomposition: Gotman (1976, 1980) used a waveform decomposition method to break down the EEG into half-waves and detect spikes by measuring the amplitude, the duration and the sharpness of half-waves.

5.Discriminant analysis based on parameterization of the EEG: Five variables, including first and second derivatives, were used by Chik et al. (1977) in a linear discrimination function to detect spikes.

Glover et al. (1986, 1989) developed a microprocessor-based multichannel system to detect spikes. In that system, the comprehensive use of spatial and temporal information reduces false detections caused by a wide variety of artifacts in EEG recordings. Results show that the system is more reliable then those using less context information. The concept of interchannel information is used in our studies as well. Gotman and Wang (1991, 1992) significantly improved the method of Gotman by dividing the EEG into five states and applying different detection criteria to each state.

3.2 Detection of Seizures

There are not many seizure detection methods, in contrast to spike detection methods. This may be because seizure patterns are more variable and they only occur rarely. Moreover, EEG seizure patterns can be usually noticed by a technologist when they occur. However, when long-term EEG monitoring is performed in the absence of observers or there is a seizure with minor or no behavioral manifestations, a seizure detection method is very useful. It is impossible to detect all seizures by using only the EEG because some seizures start deep in the brain and may not spread to the location of electrodes (Gotman 1985; Sharbrough 1993). In addition, in scalp electrode patients, seizure patterns are sometimes obscured by muscle activity or artifacts.

Most seizures do not have significant seizure characteristics in their early part, but the prominent patterns gradually appear as seizures evolve. This is why most seizure

detection algorithms (Gotman 1982, 1990a; Murro et al. 1991; Liu et al. 1992; and Harding 1993) aim at detecting prominent seizure patterns rather than early seizure patterns which may not have prominent characteristics. In those algorithms, a seizure onset could be detected if the early seizure patterns are prominent. The detection can, however, only be reported about a dozen seconds later. This is because the system collects information during a dozen seconds to increase the probability of having a valid detection. When early seizure patterns are not prominent enough, those algorithms cannot detect them.

Different seizure detection methods use different lengths of epochs. However, all use short epochs (2 to 8 seconds) because most ictal patterns change more rapidly than most interictal EEG patterns. In the study of Hilfiker and Egli (1992), a 2-second epoch was used to study the evolution of rhythmic components. An epoch of two seconds was also used by Gotman (1982, 1990a). In other studies, Liu et al. (1992) used six seconds as an epoch to study neonatal seizures because neonatal seizures tend to have rhythmic discharges at a very low frequency (0.5Hz to 2Hz), and Murro et al. (1991) used 6.83 seconds as an epoch to analyze complex partial seizures. In another study of Murro et al. (1993), an epoch of 3.4 seconds was used to study the localization of temporal lobe seizures.

An automated seizure monitoring system for patients with intracerebral electrodes has been developed by Harding (1993). In that system, a real-time automatic seizure detection performs with an accuracy of 95% in detection rate. This was evaluated in 792 clinical and subclinical seizures during 1578 hours of monitoring. The false alarm rate of the system was estimated at 1.93/hour. However, this system is only used for patients with intracerebral electrodes.

Gotman (1982, 1990a) presented a method attempting to detect a variety of seizures. It stores EEGs which are detected as seizures for later review. This method is a mimetic one which simulates the way humans analyze the EEG. It uses waveform

decomposition to break waves into half-waves. By using average amplitude, average duration and coefficient of variation of halfwaves in an epoch of the EEG, it detects a seizure if several preset conditions are satisfied. It is currently used at the Montreal Neurological Hospital and in many other hospitals. It was formally evaluated by Gotman (1990a) and Pauri et al. (1992). In the evaluation of Pauri et al. (1992), depending on the detection thresholds, seizure detection rates vary from 47.8% to 81.4% and false seizure detection rates range from 2.70/hour to 5.38/hour. While there is room from improvement, this method is very practical and useful. Details of this detection method will be discussed in the next section.

Few reports have made a specific effort to reduce false seizure detections. It may be because most seizure detection methods have not been used on-line (Aziz et al. 1986, Murro et al. 1991, Liu et al. 1992) and therefore false detections are not causing major problems in EEG storage and review.

3.3 Seizure Detection Method of Gotman

In our attempts to reduce false seizure detections and detect seizure onsets, we used some features from the detection method of Gotman (1982, 1990a). These features have proven useful in the detection of seizures in patients with either scalp or depth electrodes. The waveform decomposition method will be explained first because it is the foundation of the method. Features used in the detection are then discussed. Finally, the detection criteria are presented.

3.3.1 Waveform Decomposition Method

The waveform decomposition method of Gotman performs an initial digital filtering to remove high frequency activity and then breaks down the EEG into

halfwaves. As shown in figure 2-7, the original EEG is first represented by a set of segments. A segment is the section between two consecutive extrema of amplitude. Since some small amplitude waves, like muscle and beta waves, may be superimposed on waves of longer duration, segments are regrouped into sequences, called halfwaves, by using some constraints. A sequence ends when a segment which may not belong to that sequence is found. Each halfwave is then characterized by its duration and amplitude.

3.3.2 Features

Before features are discussed, it is important to define how the concept of background is used in this method because features are often measured in relation to the background. In figure 2-8, it is shown that there are two backgrounds, background A and background B. A gap of 20 seconds is used to separate the current epoch from background A, which lasts 16 seconds. This is because some seizures start gradually and the gap can make the detection more reliable by not including the gradual onset in the background. The background B, which lasts 8 seconds, is used to make sure that an event lasts long enough to be a seizure. For each epoch and for each channel of EEG, six features have to be computed for the determination of a seizure detection.

1. Average amplitude of the current epoch: It is expressed as: $Avg_{curramp} = \frac{\sum_{i=1}^{N} CURRAMP_i}{N}$, where $CURRAMP_i$ is amplitude of halfwave i and N is the number of halfwaves in the current epoch.
2. Average duration of the current epoch: It is expressed as: $Avg_{DURCURR} = \frac{\sum_{i=1}^{N} DURCURR_i}{N}$ where $DURCURR_i$ is duration of halfwave i and N is the number of halfwaves in the current epoch.

3. Coefficient of variation of the current epoch: It is expressed as:

$COVA = \frac{\sum_{i=1}^{N}(DURCURR_i - AVG_{DURCURR})^2 \Big/ N}{AVG^2_{DURCURR}}$, where N is the number of halfwaves in the current epoch, $DURCURR_i$ is duration of halfwave i in the current epoch and $AVG_{DURCURR}$ is the average duration of halfwaves in the current epoch

4. Average amplitude of the background A: $Avg_{backAamp} = \frac{\sum_{i=1}^{N} BACKAAMP_i}{N}$, where $BACKAAMP_i$ is amplitude of halfwave i and N is the number of halfwaves in the background A.

5. Average duration of the waves in the background A: $Avg_{backAdur} = \frac{\sum_{i=1}^{N} BACKADUR_i}{N}$ where $BACKADUR_i$ is duration of halfwave i and N is the number of halfwaves in the background A.

6. Average amplitude of the background B: $Avg_{backBamp} = \frac{\sum_{i=1}^{N} BACKBAMP_i}{N}$, where $BACKABAMP_i$ is amplitude of halfwave i and N is the number of halfwaves in the background B.

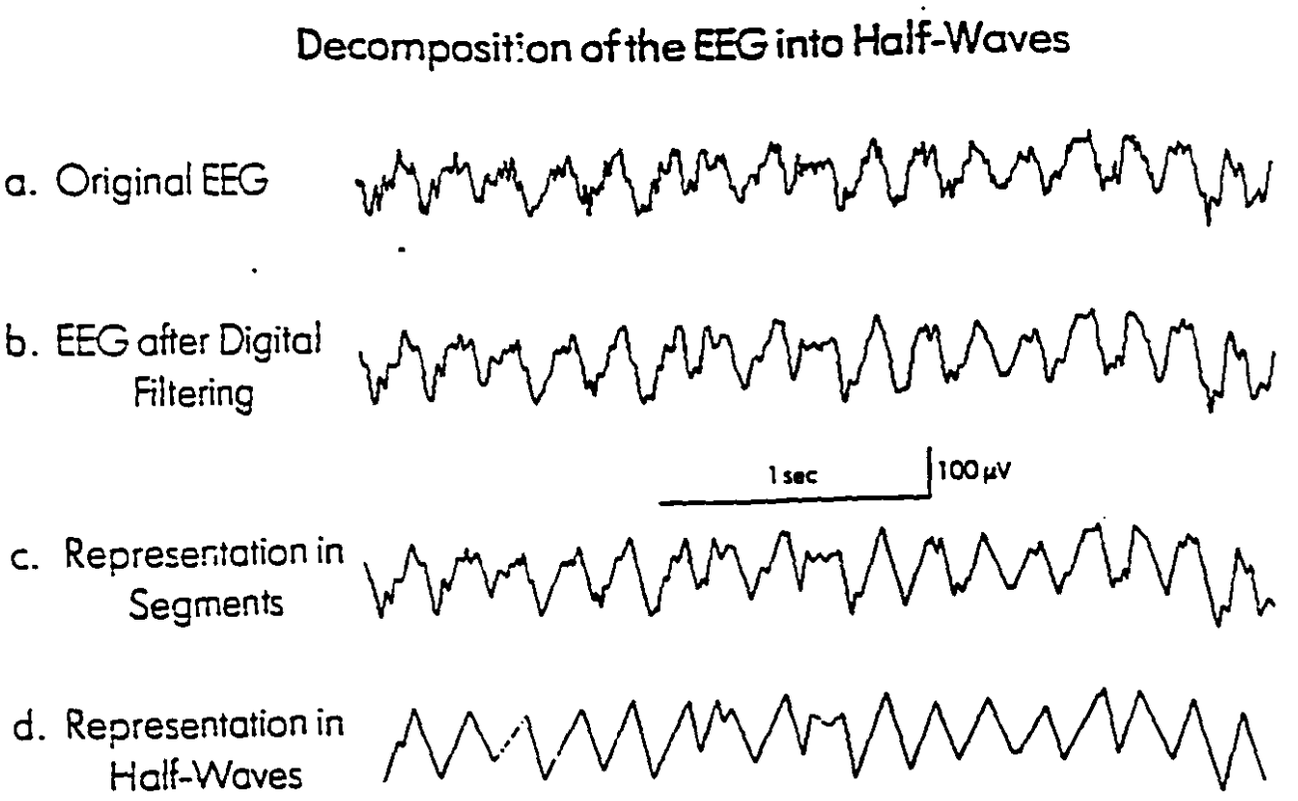


Figure 2-7: Waveform decomposition method of Gotman (1976): This figure illustrates the procedure of the decomposition of the EEG into half-waves. "a" is original EEG, "b" is EEG following digital filtering, "c" is EEG broken down into segments and "d" is EEG broken down into half-waves.

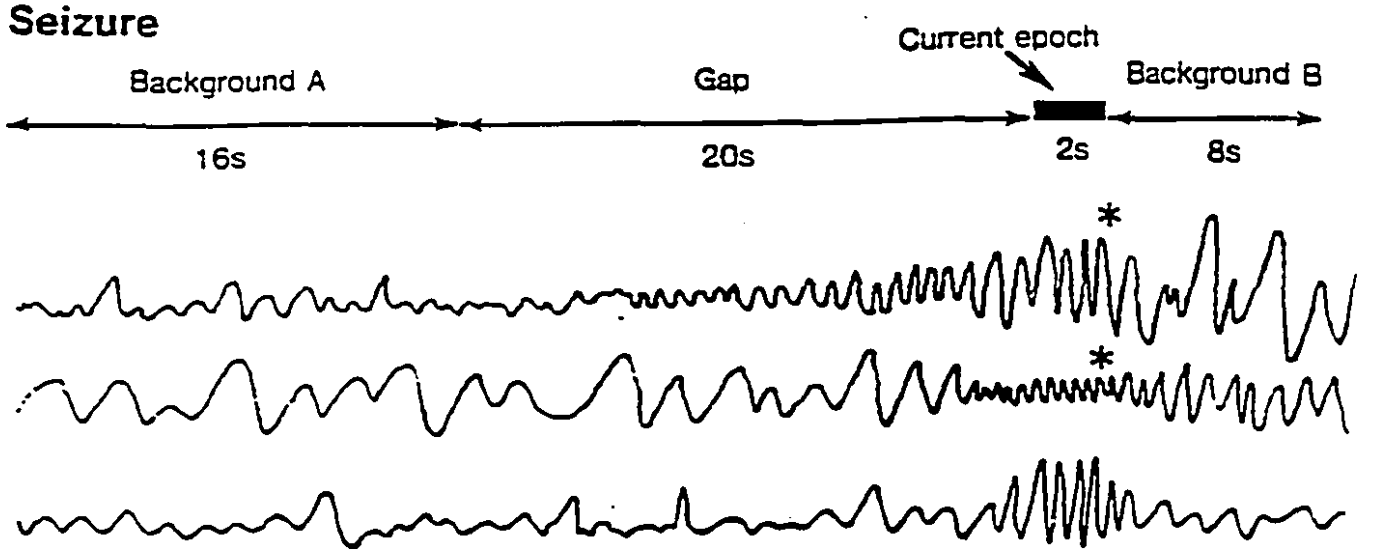


Figure 2-8: Schematic use of the background in the seizure detection method of Gotman (1982, 1990a). The gradual onset of a seizure is not included in the background because of the gap (20s). A detection takes place in the second line because the frequency during the current epoch is much higher than that of background A. In the third line, there is no detection because the activity in the background following the current epoch is too low (background B).

3.3.3 Detection Criteria

An epoch of EEG is declared a seizure pattern when the following criteria are satisfied.

1. The average amplitude of halfwaves in the current epoch is three times larger than that in background A, or the amplitude of halfwaves in the current epoch is the same or larger than that in background A, provided the average duration of halfwaves in the current epoch is one third shorter than that in the background A.
2. The average duration of halfwaves in the current epoch is between 25 and 150 milliseconds (roughly corresponding to frequencies of 20Hz and 3Hz).
3. The coefficient of variation of halfwaves is less than 0.6.
4. The average amplitude of halfwaves in the background B is at least 1.6 times that of the background A.
5. All above criteria have to be satisfied not only in the current epoch, but also in an adjacent epoch or the same epoch in a different channel.

3.4 Conclusion of Spike and Seizure Detection

Existing methods for the detection of spikes and seizures are primitive and far from perfect, despite the fact that some are in clinical use. While much effort has been made on the detection of spikes and more is currently being carried out, very few methods have been developed for the detection of seizures. As a result, seizure detection methods still detect many false events and miss many real events. There is certainly room for improvement if more sophisticated pattern recognition methods are used. It would be particularly useful to incorporate in the detection process information about a wide context, including information about patterns having taken place in the hours and days preceding a recording session. We review in the next section some pattern recognition

methods and their applications, specially the methods dealing with pattern classification.

4 Statistical Pattern Recognition and Applications

There are many approaches to pattern recognition, such as statistical, syntactic, heuristic and ad hoc methods. Among them, statistical methods have played a prominent role in the development of pattern recognition techniques (Devijver 1982). Statistical methods provide the proper framework for studying pattern recognition problems when the pattern-generating mechanism can be faithfully represented by a statistical model, and the goal of the recognition is to decide whether or not a given pattern belongs to some pre-specific class of patterns. Statistical methods have been widely applied to many fields, such as character recognition, medical diagnosis, automatic inspection, speaker recognition, etc...

Few problems are more challenging than decoding the meaning of the electrical activity of the human brain. The lack of sufficient knowledge of the origin and significance of the electrical activity of the brain is a fundamental obstacle to analyze and interpret the EEG. However, statistical methods are especially useful for extracting information from the human brain's electrical and magnetic fields (Gevins 1987a).

In this section, a general description of some statistical pattern recognition methods related to EEG analysis will be given first. Then, some of their applications will be presented.

4.1 Statistical Pattern Recognition Methods

The design of a typical pattern recognition system for EEG analysis comprises several basic steps: selection of data, feature extraction and selection, classification and finally the performance estimation. Due to the vast varieties of problems in EEG analysis, the design of each step should closely relate to the problems in each application.

4.1.1 Feature Extraction and Selection

Features can be extracted from many sources. For one dimensional signals, such as speech and EEG signals, features usually come from either time domain or frequency domain or both. In EEG analysis, features from spectral analysis (Walter 1963; Walter et al. 1966; Dumermuth and Fluhler, 1967; Dumermuth et al. 1970; Gotman et al. 1973) and linear prediction (Zetterberg 1969, Gersch 1970) are commonly used, while some waveform decomposition methods, zero-cross counts and various types of peak-picking methods are also used (Leader et al. 1967; Hjorth 1970; Gotman 1976). For instance, amplitude and average duration of waves (Gotman 1976, 1980) are features from a waveform decomposition method and they were used to detect spikes and seizures. Mean frequency (Walter et al. 1967), band power (Murro et al. 1993) and coherence (Walter et al. 1967; Sklar et al. 1973) are features from spectral analysis and they have also been used to detect seizures, as well as to classify sleep stages and evoked potentials.

Gasser (1977), Cohen and Sances (1977) and Huber et al. (1971) have studied the effect on the estimation of analytic quantities of departures from normality and nonstationaries of the EEG. As a practical matter, the consequences of violations of the assumption of stationarity and the necessity for a time-varying analysis should be assessed for each application. For instance, characterizing sleep stages lasting dozens of minutes, changes within a few seconds are not important. At the other extreme, a few dozens of milliseconds may be important in the study of evoked potentials (Gevins 1980). As a result, the length of the analysis window in each application should reflect the characteristics of the application so that stationary assumption can be considered satisfied. As indicated by Cohen and Sances (1977), an epoch of the normal EEG which

is shorter than 12 seconds can be considered a stationary random process based on mean-value and frequency-structure.

Feature selection is a step to maximize separability among different classes (Devijver and Kittle 1982). It is the least straight-forward part of an EEG pattern recognition study (Fu 1968; Mendel and Fu 1970; Foley 1972; Gray and Schicany 1972; Meisel 1972; Gonzalez and Thomason 1978). Different procedures produce different feature subsets, and there is no way of knowing that a chosen subset is the optimal one other than trying all possible subsets (Gevins 1980). Adding more features to the classification function will generally improve performance on the training data, but the generalization performance may actually decrease (Foley 1972). In most of the EEG analysis problems, stepwise discriminant analysis was used to select useful features (Walter et al. 1967; Berhout et al. 1969; Donchin et al. 1970; Sklar et al. 1973; Squires and Donchin 1976; John 1977; Sencaj et al. 1979; Horst and Donchin 1980; Yunck and Tuteur 1980), while in some cases non-linear trainable classification networks have been applied (Viglione 1970; Martin et al. 1972; Gevins et al. 1979a, 1979b). Since feature selection methods in the EEG analysis have an arbitrary aspect (Gevins 1980), each application should make the selection according to the situation. In some cases, a subset of features does not perform better in classifications than the whole feature set (Yunck and Tuteur 1980).

4.1.2. Classification

There are two types of statistical classification methods: parametric and non-parametric. Parametric classification represents data with a simple statistical model, such as the autoregressive model (Zetterberg 1969), and uses parameters of this model to do the classification. The assumption of statistical model of a signal depends on the mechanism of the generation of the signal. Many parametric statistical methods have

been used. For instance, Fukunaga and Kessell (1971) compared the error rates of a parametric Bayes' classifier with a non-parametric one in two sets of artificial data with the normal distribution. Yunck and Tuteur (1980) used the maximum a posteriori (MAP) classifiers to compare classification accuracy with non-parametric classifiers. Gath and Bar-on (1980) used autoregressive model to represent the EEG in classifying sleep stages.

In the case that a signal cannot be represented by a simple statistical model, a non-parametric classification should be used. The most commonly used non-parametric classification methods are the nearest-neighbor (NN) rule and the k-nearest-neighbor (k-NN) rule. The NN rule classifies data of an unknown class into the class in which the nearest data is. The k-NN rule assigns data of an unknown class into the class in which the majority of the k nearest data are (Fukunaga 1972; Duda and Hart 1973; Batchelor 1974; Devijver and Kittle 1982; James 1985). These non-parametric classification methods heavily depend on the local data distribution (one or k nearest data) instead of global data distribution as parametric classification does. The error rate of the NN classifiers is bounded between the Bayes error rate and twice of the Bayes error rate, and k-NN classifiers will approximate the Bayes error rate if k is chosen properly (Cover and Hart 1967; Bhattacharya et al. 1992).

It has been demonstrated by Yunck and Tuteur (1980) that the k-NN rule performs better than parametric classifiers, such as the MAP classifiers, in classifying the EEG into five categories according to different tasks. The tasks include rest, performing arithmetic exercises, listening to music, performing verbal exercises, listening to speech, performing pictorial exercises and viewing a film. As indicated in Gevins (1987a), non-linear and non-parametric classifiers have consistently performed better than linear and parametric ones in EEG analysis. This is possibly because the mechanism of the generation of the EEG is too complex to be described by a simple statistical model and

non-parametric methods do not make any assumption about the data distribution as parametric ones do.

In the NN classification rule, as well as other classification rules, the distance measure is one of the most important factors. There are several distance measures, such as the Euclidean distance, the city block distance, the Mahalanobis distance, the Minkoski r-distance, the square distance and the weighted distance (quadratic distance) (Batchelor 1974; Devijver and Kittle 1982). Among them, the weighted distance is expressed as: $\bar{D} = (\vec{X} - \bar{Y})\vec{M}(\vec{X} - \bar{Y})$, where the matrix $\vec{M}$ has different weights in different dimensions. This distance can reflect the importance of each dimension, and a weighted space is constituted by these dimensions (Bow 1984). We used a modified version of this type of distance in our studies. A similar weighted distance measure has been used in electromyography (Zhang et al. 1991).

4.1.3. Performance Estimation

The validation of a classifier is critical in the design of a pattern recognition system since it reveals the future performance of the classifier. A biased result may give misleading information. Therefore, all possible measures, including the collection of data and the evaluation procedure, have to be chosen carefully to make sure that results are as unbiased as possible.

The collection of data has to ensure that it represents at least those properties of the data which are invariant from sample to sample and which distinguish the classes. It is certain that more data result in a better representation of all properties within the data. However, since it is difficult and expensive to collect neurophysiologic data, the sample size in EEG pattern recognition studies has usually been determined by the availability of data or other practical restrictions, rather than by the usual statistical power analysis (Gevins 1980).

There are many ways to estimate misclassifications (Toussaint 1974). Risk averaging is one of them. The resubstitution, holdout, leave-one-out and rotation methods are the four methods of error estimation by error counting (Devijver and Kittle 1982). They are non-parametric error estimators.

1. The resubstitution method uses the same set of data for training and for testing. This method has a major drawback: it may give a very misleading result, especially when the data set is small. For instance, if the NN rule is used as a classifier, there will never be any error by using the resubstitution method. This is because the decision boundary is trained by a set of data to classify all points into correct classes by using the NN rule. When the same set of data is used as testing data, the classifier is already perfect for this set of data and therefore no error will be found. It may not be the case if another set of data were used as testing data.

2. The holdout method divides data into two mutually exclusive sets and uses one of them for training a classifier and the other one for testing. This method makes poor use of data and gives pessimistic error estimation. (Devijver and Kittle 1982) However, when the data set is large enough, this method has a significant advantage in terms of computation costs because both training and testing need to be done only once. This method has been used in our studies.

3. The leave-one-out method consists in removing one sample from the data set and using this sample as testing data and others as training data. After this is done, the testing data is returned to the data set. Another data will then become a testing data, while the rest become a training set. This procedure will be repeated until all data become testing data once. This method gives an unbiased error estimate and makes full use of every single sample. It is particularly useful for small data sets. However, it increases the variance of the error estimation (Devijver and Kittle 1982) and has a high computation cost.

4. The rotation method is a combination of leave-one-out and holdout method. Instead of removing one sample from the data set, several samples are removed. This method is a compromise between the variance of the estimation and the computation costs. It can be used in both small and moderate size data sets. This method is chosen for the error estimation of probability of missing detection of seizures in our study of reduction of false detection of seizures.

In the study of physiological signals, validation of classifications is most often obtained with a completely independent data sample from the one used for the training of classifiers. However, because of the difficulties and expenses in data collection, the rotation or the leave-one-out methods are sometimes used for validation. Obviously, these methods can produce an unbiased estimate of the classification accuracy if this data set has not been used in the previous steps of analysis (Devijver and Kittle 1982). If the validation is done in a data set including different persons from those included in the training data set, the accuracy is obviously much more reliable. Nevertheless, when the number of persons used for training is adequate to represent the variability of the general population, and when the between-class differences are large, both ways of using the validation data sets produce the same results (Gevins 1980).

Error estimation by error counting has been used in many pattern recognition studies. For instance, Fukunaga and Kessell (1971) used the leave-one-out method to compare classification errors of two different classifiers: Bayes' classifier and nonparametric classifier using the Parzen approximation. The comparison was done on the mean error and standard deviation. In the study of Oliver et al (1979), a holdout experiment was carried out to test the average correct recognition rates (85%) of abnormal cells and an error rate (1%) on normal cells.

4.2 Applications of Statistical Pattern Recognition Methods

4.2.1 Application in EEG Analysis

Stationarity of the EEG has been studied by Cohen and Sances (1977). Time-varying value and time-varying frequency value of 104 clinical EEGs were used in the study and results showed that the EEG can be considered to be a stationary random process for epochs shorter than 12 seconds. For a probability of error of 10%, an epoch of 24 seconds of the EEG can be considered stationary too. The EEG can be treated as stationary for as long as 64 seconds with a probability of error of 35%.

Many parametric methods have been used in EEG analysis. The autoregressive model has been used to detect spikes (Zetterberg 1969; Lopes da Silva et al. 1975) and in sleep stages classification (Gath and Bar-on 1980, 1985). Correlation was used by Barlow and Dubinsky (1976) to detect spikes by template matching. Liu et al. (1992) used the autocorrelation function as a basic function to detect neonatal seizures. Larson and Walter (1970) used spectral analysis followed by stepwise discrimination analysis to classify sleep stages. Gath and Bar-on (1980,1985a) used fuzzy subset theory, fuzzy decision making and optimal fuzzy partition to classify sleep stages with linear prediction coefficients as features. In the study of Friedman and Jones (1984), a cluster analysis method, which used both the Euclidean distance to centroids of each cluster and structure features of sleep EEG, was used to classify three states, wakefulness, slow wave sleep and paradoxical sleep, in cats. A fuzzy clustering method, which uses the "k-nearest prototypes" method and adaptive segmentation of the EEG, has been also used to distinguish waking EEG with dominant alpha and low amplitude mixed frequency EEG as two different background stages in the evoked potential study by Gath et al. (1985)

Non-parametric methods are also popular in the analysis of the EEG. Leader et al. (1967) used a peak-picking method to determine the minimum and maximum of the EEG waves. The amplitude and duration of the EEG waves were then used in a step-

wise discrimination classifier to separate the EEG waves into 12 predefined categories. This method is similar but much simpler than the one of Gotman (1976). In order to separate subjectively stressful from non-stressful verbal stimuli, and to determine distinctive EEG responses to verbal stimuli of similar stress value differing only in semantic content, Berkhout et al. (1969) used intensity and bandwidth of auto-spectral analysis, and coherence pairs of channels as features. The classification was done by a step-wise discriminant function to separate the detection space into different regions and place each epoch of the EEG into one of them. It is interesting to notice that in this method the definition of the bandwidth has a similar concept to our main energy zone feature used in the detection of seizure onset (see chapter 4). The method achieved a 92% overall correctness in the classification.

The hold-out method has been widely used for the evaluation in many applications. Gevins et al. (1979a, 1979b) used this method to evaluate a nonlinear multivariate pattern recognition system and an 87.5% of accuracy was estimated. The system was used to distinguish the spatial distribution of the EEG patterns associated with several complex tasks, including Koh's block design, writing sentences, mental paper folding, and reading silently. Horst and Donchin (1980) also used the holdout method to estimate the correctness (87.8%) in classifying the evoked potentials elicited by a checkerboard presented to the upper or lower visual half-field.

4.2.2. Application in other Fields

We will only review briefly a few applications having some aspects in common with our problems.

Many applications use parametric classification methods. In the study of Durand et al. (1990), several classifiers were compared in the classification of spectra of heart sounds in patients with a porcine bioprosthetic valve implanted in the mitral position. In

this study, the Gaussian-Bayes model is performing slightly better than the NN rule (98% vs. 94%). Results of the study also show that among classifiers using the NN rule, the classifier using Mahalanobis distance performs better than the classifier using Euclidean distance. Automatic clustering and patient-specific classification of new patterns was used by Swenne et al. (1973) in the recognition of ventricular complexes during ECG-monitoring. In that study, it is assumed that all signal clusters can be described by ellipsoids. Human interaction is needed if a pattern is rejected by the automatic clustering method during the training of classifiers. The classification then determines if a new pattern belongs to the predetermined clusters in the detection space.

Non-parametric methods have been utilized in many applications too. In seismic wave interpretation, the NN rule has been used to classify exploration waves and earthquake waves (Chen 1982). In that study, autocovariance and autocorrelation were used as features and 89.2% correct recognition was achieved. In speech recognition, template matching has been widely used (De Mori 1982). Linear time warping is also commonly used in speech recognition for time alignment. Our studies use both template matching and time alignment between patterns, but they are different from the ones in speech recognition because of the differences between signals. The k-NN rule has been used in the multicategory classification of body surface potential maps (Reich et al. 1990). By using leave-one-out method, the method was evaluated in 123 patients belonging to four categories and the accuracy is 94% for normal patients, 88% for ischemia patients, 91% for myocardial infarction patients and 100% for left bundle branch block patients. Dube et al. (1988) designed an ECG monitoring system for ischemic patients by using a stepwise discriminant function to detect heart beats and ST-segment changes. Results indicated that the system can work reliably in adverse conditions.

Chapter 3:

Reduction of False Seizure Detection

1 Introduction

Long-term monitoring is an established procedure to record seizures in patients with intractable epilepsy. It is very tedious and expensive to have a person observe patients and EEGs continuously so that every single seizure can be recorded, including seizures with electroencephalographic but no behavioral manifestations. A computerized automatic seizure detection system can often detect and record seizures in the absence of observers or when patients do not notice their seizures (Gotman 1985, 1990a). However, it is unavoidable that some false seizure detections (FSDs) occur, and in some cases very frequently. A very high FSD rate causes a large amount of EEG to be stored and reviewed, and consequently reduces the practical value of seizure detection.

In the long-term EEG monitoring system used at the Montreal Neurological Hospital, the FSD rate was estimated to be 0.84/hour for scalp electrode patients and 1.35/hour for depth electrode patients (Gotman 1990a). This study was done by evaluating 241 recordings from 44 patients with scalp electrodes and 52 recordings from 5 patients with depth electrodes. In another study (Pauri et al. 1992), the same seizure detection system was evaluated on twelve patients with a total of 461 hours in monitoring. Depending on the detection threshold, the average FSD rates were between 2.70/hour and 5.38/hour. In that study, it was found that artifacts were the main cause of FSDs (80%) while normal and abnormal EEG patterns represented each 10% of FSDs.

Whereas the average FSD rate is acceptable, it can become significantly higher and reach unacceptable values in some patients. This usually happens when, in a given patient, one or a small number of patterns occur repeatedly and cause many FSDs during several consecutive monitoring sessions. These FSD patterns vary from one patient to another. Thus, they can be called patient-specific FSD patterns. Because a patient is typically monitored for several days, a monitoring system can be "taught" to recognize and stop detecting EEG patterns that have been identified early in the monitoring as FSDs.

Few reports have dealt with reducing FSDs. It may be because other seizure detection methods have not been used on-line (Aziz et al. 1986; Murro et al. 1991; Liu et al. 1992) and therefore FSDs are not causing major problems in EEG storage and review. Alternative solutions were tried before in our own system. For instance, if the EEG patterns causing FSDs occurred in one or a few channels only, the monitoring system could simply stop detecting any seizure from these channels. Another solution is to rise detection thresholds to reduce the FSD rate (Pauri et al. 1992). These relatively undiscriminating ways of eliminating FSDs may cause a serious problem: the true seizure detection rate is reduced, in some cases significantly (Pauri et al. 1992).

In order to eliminate patient-specific FSDs with a minimal probability of losing true seizures, we propose a method based on the learning of a trial session. If there are many FSDs in a patient in one monitoring session and most are caused by similar EEG patterns, this session becomes a trial session. EEG patterns of FSD during this session are collected in a set specific to this patient and labeled as the initial FSD set. In subsequent monitoring sessions of the same patient, an EEG pattern which is detected as a seizure but is similar to any pattern in the initial FSD set is regarded as a FSD and not reported as a seizure detection. In a given patient, the patterns of true seizures are not likely to be similar to the patterns of FSD; the probability of losing true seizures by this method is therefore reduced. One difficulty in the implementation of this method is that we must define a measure of similarity between two EEG patterns, in order to determine if a detection occurring one day is "similar" to false detections having occurred earlier.

The concept of learning about false detection from a training session could apply to any existing seizure detection method. We have selected to evaluate this concept on the seizure detection method of Gotman, which is therefore used as the basic seizure detection method on which we are grafting a patient-adaptive algorithm.

In this chapter, the algorithm will be discussed first, followed by the implementation procedure. Details of the evaluation procedure are given next, followed by results and discussions.

2 Algorithm

This pattern recognition system involves feature extraction, classifier design and classifier training. Before the design of the classifier is discussed, features used in this method are presented. Training of classifiers is then explained.

2.1 Feature Extraction and Normalization

Seven features are used in the seizure detection system of Gotman (Gotman 1982, 1990a). All these seven features are used in the recognition of patient-specific false detections. These features constitute the *detection space*, which is shared by both true and false seizure detections. Each detection is presented in the detection space as a seven-dimensional point. Before describing the features, some basic definitions have to be given.

(1) An epoch is defined as a section of EEG with a duration of 2 seconds.

(2) Halfwaves are basic segments of EEG waves and computed by using the waveform decomposition method of Gotman (1982).

(3) A section of background EEG is defined as between 36 seconds and 20 seconds before the current epoch, as defined in Gotman (1990a).

The seven features used in the method are defined as follows:

1. *Ratio of the average amplitude of halfwaves in the current epoch to the average amplitude of halfwaves in the background*: It is expressed as:

$$RAMP_{Curr/Back} = \frac{\sum_{i=1}^{N} AMPCURR_i \Big/ N}{\sum_{j=1}^{M} AMPBACK_j \Big/ M}$$

, where N and M are number of halfwaves in the current epoch and the background respectively. $AMPCURR_i$ is the amplitude of

a halfwave in current epoch and $AMPBACK_j$ is amplitude of a halfwave in the background.

2. *Average frequency of current epoch*: It is actually the inverse of twice the average duration of halfwaves in the current epoch. The formula is: $Freq_{Curr} = \frac{1}{2 \times \sum_{i=1}^{N} DURCURR_i / N}$, where N is the number of halfwaves in the current epoch and $DURCURR_i$ is the duration of a halfwave in the current epoch.
3. *Average frequency in the background*: It is the same as the average frequency of current epoch except it is computed in the background. The equation is: $Freq_{Back} = \frac{1}{2 \times \sum_{i=1}^{M} DURBACK_i / M}$.
4. *Ratio of the average amplitude of halfwaves in the eight seconds following the current epoch and the average amplitude of halfwaves in the background*: It is the same as the first feature except the average amplitude of halfwaves in the eight seconds following the current epoch is used instead that of the current epoch. It can be presented as: $RAMP_{Next/Back} = \frac{\sum_{i=1}^{N} AMPNEXT_i / N}{\sum_{j=1}^{M} AMPBACK_j / M}$
5. *Location of electrodes (channels):* It is an expression of electrode position in a montage. Since there is no simple quantitative expression for it, the presentation of this feature in the detection space is complex and will be explained later.
6. *Detection type*: There are four detection types for different kinds of EEG patterns. They are: slow waves, bursts, epileptic forms and fast activity. These four types are presented in the detection space as four values, 1, 2, 3 and 4 respectively. The difference between any two types, however, cannot be represented by the difference

of their values because these four types are unrelated and their relationship can only be either "same" or "different". As a result, the difference between any two types is expressed by a binary value, 0 or 1. The value 0 is used when two types are the same and 1 is used for any two different types.

7. *Coefficient of variation of duration of halfwaves in the current epoch*: It is expressed as: $COVA = \frac{\sum_{i=1}^{N}(DURCURR_i - AVG_{DURCURR})^2 / N}{AVG^2_{DURCURR}}$, where N is the number of halfwaves in the current epoch, $DURCURR_i$ is duration of a halfwave in the current epoch and $AVG_{DURCURR}$ is the average duration of halfwaves in the current epoch.

The expression of the location of electrodes can be done by labeling numbers to each electrode or naming them differently. However, to quantize the distance between any two electrodes in the detection space is very difficult. This is because two aspects have to be taken into consideration in the determination of distance between two electrodes. One is the physical distance between them and the other is the anatomic relation between two electrodes. From the physical distance point of view, all electrodes are located in nodes of a matrix-like shape (figure 2-3), according to the international standard 10-20 system. A distance between two electrodes can therefore be computed using the Manhattan distance (Devijver and Kittle 1982). A Manhattan distance unit here is defined as the distance between any two adjacent electrodes. From the point of view of anatomic relation between electrodes, the distance between two electrodes in the two cerebral hemispheres is larger than that in the same hemisphere because the two hemispheres often function independently of each other. In order to reflect these two aspects, we use two criteria to determine the distance between any two electrodes.

1. If two electrodes are located in the same hemisphere, their distance is their Manhattan distance. The typical distance between two electrodes is between 1 and 4.

2. If two electrodes are located in different hemispheres, their distance is their Manhattan distance plus a constant to reflect the anatomical distance between the hemispheres. The constant was set empirically at 10 in our study because it is large enough to separate two hemispheres in the detection space.

Since each feature, which serves as a dimension in the detection space, has its own physical unit, it is necessary to normalize them into a universal unit by considering the effects of each dimension in a seizure detection. This universal unit contains different physical distances in different dimensions. This normalization makes it possible to compare distances between any two points in the detection space.

2.2 Classifier Design

The most essential part of the classifier in this method is the similarity measure. Therefore, before the design of the classifier is presented, the similarity measure will be introduced and explained in details.

2.2.1 Similarity Measure

For a given patient with a trial EEG monitoring session in which there are many FSDs, FSD patterns from this session are called the *initial FSD set* and represented as prototype points in the detection space. Every point in the detection space represents an epoch of the EEG which has its seven features transformed into a universal unit as discussed above. As a result, the transformation of features will not be shown in all computations later because every dimension in the detection space has the same universal unit. When a new pattern is detected in a subsequent monitoring session, we

need to know if it is similar to any of the FSD patterns in the trial session and is therefore to be eliminated, or different from them and is therefore to be retained. We introduce a measure in the detection space called "similarity".

Similarity between two points in the detection space, one of them belonging to an initial FSD set, takes into account not only the Euclidean distance between the two points, but also the probability distribution of the initial FSD set. The point O_i represents a prototype point in an initial FSD set and is represented as a vector $\bar{O}_i$ in the detection space. Its probability in this initial FSD set is $\bar{P}_{O_i}$. $P_{O_{ij}}$ is the probability for dimension j at this point O_i. This value was computed by projecting all data into the jth dimension to form a probability distribution and $P_{O_{ij}}$ is the probability corresponding to the projection point of O_i in this dimension. A point N_m representing a new EEG pattern and represented as a vector $\bar{N}_m$ in the detection space, has a statistically weighted distance to $\bar{O}_i$ of $(\bar{O}_i - \bar{N}_m)^T \bar{M}^{-1} (\bar{O}_i - \bar{N}_m)$.

$SIMIL(\bar{O}_i, \bar{N}_m)$ is the value of similarity between $\bar{O}_i$ and $\bar{N}_m$. $G(\bar{O}_i, \bar{N}_m, \bar{P}_{O_i})$ is the main function to measure the similarity between $\bar{O}_i$ and $\bar{N}_m$ and $COR(\sum_{j=1}^{7} P_{O_{ij}})$ is a modification and normalization function which will be discussed later. The similarity between $\bar{O}_i$ and $\bar{N}_m$ is then defined as:

$$SIMIL(\bar{O}_i, \bar{N}_m) = G(\bar{O}_i, \bar{N}_m, \bar{P}_{O_i}) \times COR(\sum_{j=1}^{7} P_{O_{ij}}) \quad (1)$$

$$G(\bar{O}_i, \bar{N}_m, \bar{P}_{O_i}) = \frac{100}{(\bar{O}_i - \bar{N}_m)^T \bar{M}^{-1} (\bar{O}_i - \bar{N}_m) + 1} \quad (2)$$

$$= \frac{100}{\sum_{j=1}^{7} (O_{ij} - N_{mj})^2 \times (P_{O_{ij}}^2) + 1} \quad (3)$$

$$COR(\sum_{j=1}^{7} P_{O_{ij}}) = C \times (e - e^{\frac{1}{H \times \sum_{j=1}^{7} P_{O_{ij}} + 1}}) \quad (4)$$

Where: j indicates jth dimension,

$$\bar{O}_i = [O_{i1} \quad O_{i2} \quad O_{i3} \quad O_{i4} \quad O_{i5} \quad O_{i6} \quad O_{i7}]^T,$$

$$\bar{N}_m = [N_{m1} \quad N_{m2} \quad N_{m3} \quad N_{m4} \quad N_{m5} \quad N_{m6} \quad N_{m7}]^T,$$

$$C = \frac{1}{e-1},$$

$$\bar{M}^{-1} = \begin{bmatrix} P_{O_{i1}}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & P_{O_{i2}}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & P_{O_{i3}}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & P_{O_{i4}}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & P_{O_{i5}}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & P_{O_{i6}}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & P_{O_{i7}}^2 \end{bmatrix}.$$

H is a modification parameter and C is a normalization constant. The value of $SIMIL(\bar{O}_i, \bar{N}_m)$ ranges from 0 to 100, where 0 means no similarity and 100 means perfect similarity.

The function $G(\bar{O}_i, \bar{N}_m, \bar{P}_{O_i})$ is inversely proportional to the statistically weighted distance $(\bar{O}_i - \bar{N}_m)^T \bar{M}^{-1} (\bar{O}_i - \bar{N}_m)$ between points $\bar{O}_i$ and $\bar{N}_m$ as defined in equation (2) and (3). There are two reasons for this. First, the similarity should increase with a shorter Euclidean distance between two points in the detection space. In order words, the closer two points are, the more similarity they have. This is reflected in equation (3) where $\sum_{j=1}^{7} (O_{ij} - N_{mj})^2$ is the Euclidean distance between two points in the detection space. Secondly, the similarity monotonically increases when the square of the probability $P_{O_i}^2$, which is used in the $\bar{M}^{-1}$ matrix, decreases. The reason for taking $P_{O_i}^2$ into account is that this can make the similarity measure adapt to the data distribution. This is illustrated on figure 3-1. If two new detection points N_1 and N_2 have the same Euclidean distance to two false detection points F_1 and F_2, but F_1 is in a dense area (high probability P_{F_1}), and F_2 is in a sparse area (low probability P_{F_2}), then N_1 is less similar to F_1 than N_2 to F_2 (figure 3-1). In other words, if F is in a dense area, a high similarity between N and F is only possible if N is close to F; conversely if F is in a sparse area, a high similarity between N and F can be obtained even if N is not very close to F.

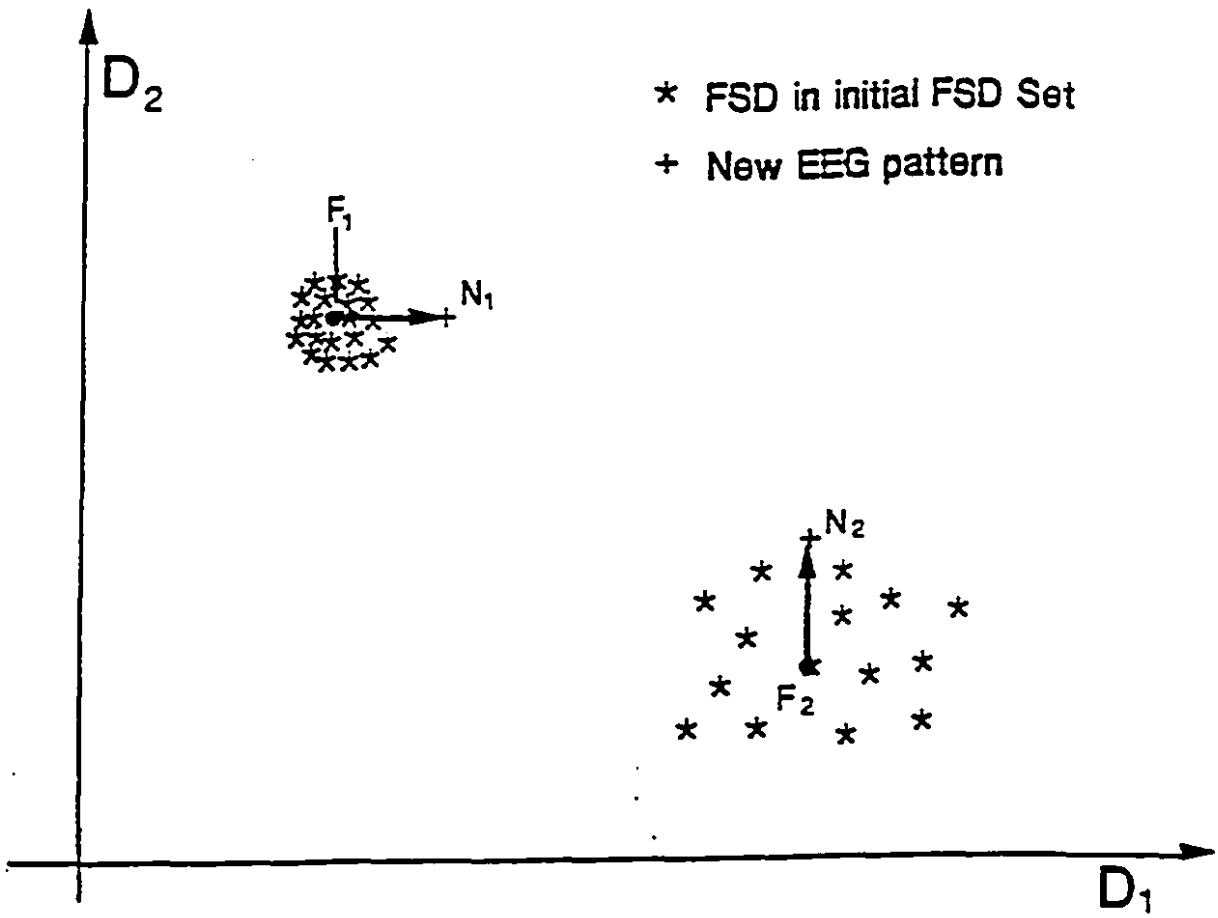


Figure 3-1: Use of probability density in the definition of similarity. F1 and F2 are points belonging to an initial FSD set. N1 and N2 represent two new EEG patterns. Although the Euclidean distance between F1 and N1 is the same as that between F2 and N2, the similarity between N1 and F1 is smaller than that between N2 and F2 because F1 is located in a denser area than F2. Axes represent any two of the seven dimensions.

The above logic causes some problems, however, when P_{O_i} is low for *all seven dimensions*. This is the case if $\bar{O}_i$ is a point which is isolated in the detection space. In this case, the value of the similarity function could be very high even for a point $\bar{N}_m$ at a large Euclidean distance from $\bar{O}_i$. We have therefore introduced a correction function $COR(\sum_{j=1}^{7} P_{O_{ij}})$ which depends on the sum of all seven probability density values for a point $\bar{O}_i$. As the sum of seven probabilities gets higher, the function becomes very close to 1. It thus does not affect the value of similarity defined by the function $G(\bar{O}_i, \bar{N}_m, \bar{P}_{O_i})$. As the sum of all seven probabilities becomes small, the function $COR(\sum_{j=1}^{7} P_{O_{ij}})$ decreases rapidly, thus reducing the value of similarity when the point $\bar{O}_i$ is nearly isolated. The parameter H was set empirically and can be adjusted to define what is meant here by a "small" or "large" sum of the probabilities. The function $COR(\sum_{j=1}^{7} P_{O_{ij}})$ is normalized from 0 to 1 and consequently the value of similarity remains between 0 and 100.

Empirical values, such as H or the weight of each feature, have been set according to experience. For instance, the range of the weight for each feature was set between 1 and 4. The most important feature, amplitude ratio, was assigned a weight at 4 and the least important feature, detection type, has a weight at 1. The settings, of course, may not be optimal. However, they reflect our best knowledge from experience.

The above similarity function allows the definition of a part of the detection space in which false detections from a particular patient are likely to be. This portion of space is called "false detection subspace" and is constituted by all the points which have a similarity to the FSDs of the trial session larger than a given threshold (figure 3-2). It is of course possible that some of the patient's true seizure patterns also fall within this "false detection subspace". In this case, such seizures will be lost. Unfortunately we do not know, after one monitoring session, all the possible seizure patterns of a particular

patient. It is possible, however, to represent in the detection space a variety of true seizure patterns from many patients. We can then determine how many fall within this patient's false detection subspace, thus giving an indication of the probability of losing seizures in this patient. If many do, it may be better to shrink the false detection subspace (by increasing the similarity threshold) in order to reduce the probability of losing true seizures. If very few true seizures fall within the false detection subspace, it may be possible to enlarge that subspace to increase the probability of eliminating false detections.

In the following sections, a question of terminology requires classification: (1) We call a "detection" or "seizure detection" an event detected by the original method of Gotman (1982, 1990a). The procedure described here results in the elimination of some of these detections: they are first detected by the classic method and then eliminated by the new method. (2) In the original method of Gotman (1982), a detection occurred when all the detection criteria were satisfied in one 2-second epoch of EEG in one channel. Each such detection is called here a *detection point*. An event can be detected several times if several epochs satisfy the detection criteria. The set of detection points relating to one event is called here a *detection section*. A large detection section may contain up to several hundred detection points because detections can be made in several epochs and in several channels. If a detection point is separated by more than 30 seconds from the previous detection point, it is considered the beginning of a new detection section. This makes sure that a single seizure corresponds to a single detection section. Although some seizures last more than 30 seconds, they are not separated into two or more detection sections because two detection points within a seizure are not usually separated by 30 seconds.

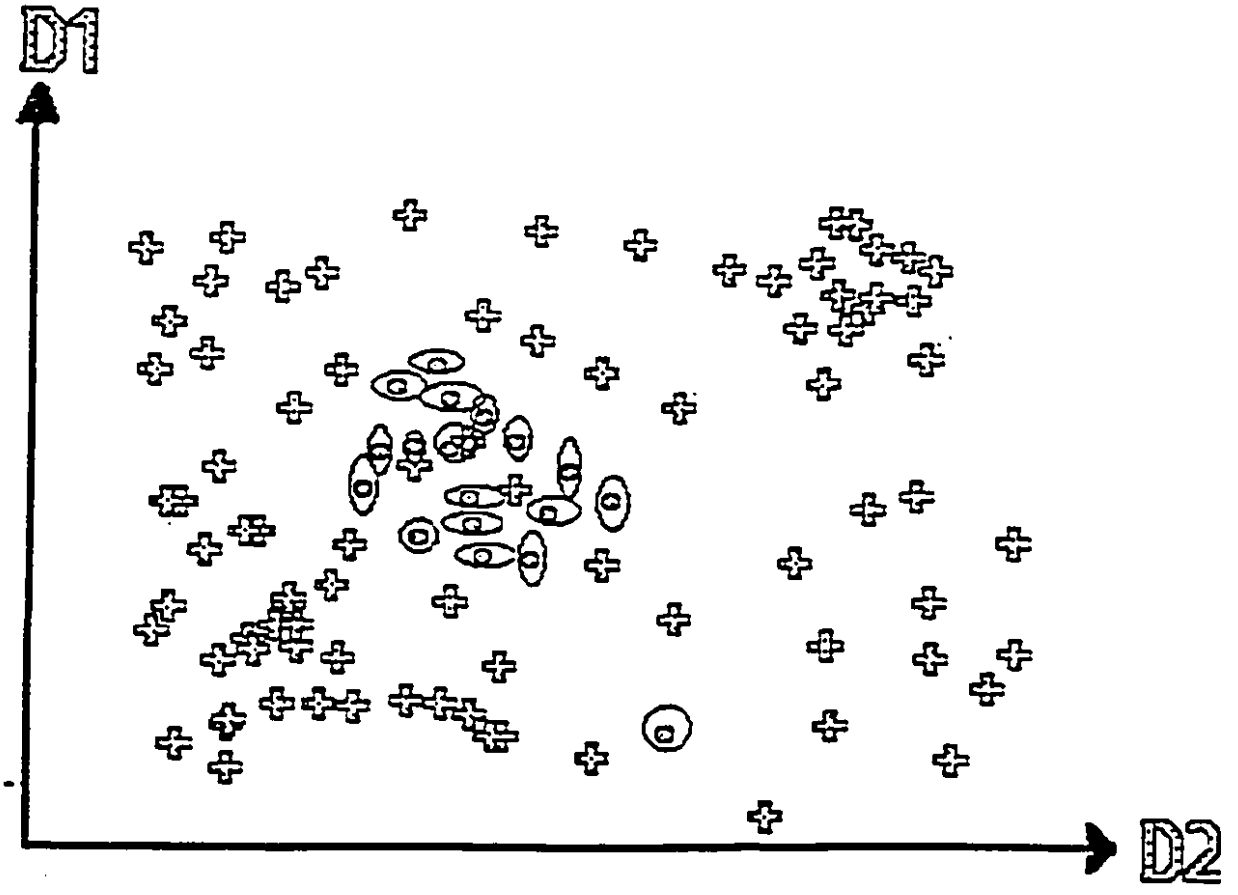


Figure 3-2: Description of the false detection subspace. Axes are two of the seven dimensions. For a given similarity threshold, boundaries of subspace are described by ellipses because they are determined by the probability distribution of the initial FSD set. The subspace corresponding to one FSD point covers more in the dimension where the probability distribution value is lower and less in the dimension where the value is higher.

2.2.2 Classifier

For a given similarity threshold T, a set of prototype points $\bar{O}_i$ and a new pattern $\bar{N}_j$, the classifier is defined as: $C(T,\bar{N}_j) = \max_i SIMIL(\bar{O}_i,\bar{N}_j) - T$

for all $\bar{O}_i$, where $SIMIL(\bar{O}_i,\bar{N}_m)$ is the similarity defined above.

$$\text{The classification rule is:} \begin{cases} \bar{N}_j \in \text{False Seizure Class} & \text{if } C(T,\bar{N}_j) \le 0 \\ \bar{N}_j \notin \text{False Seizure Class} & \text{Otherwise} \end{cases}$$

In other words, if $\bar{N}_j$ has a maximum similarity to an element $\bar{O}_i$ larger than the threshold T, $\bar{N}_j$ will be considered as an element of the false seizure class. If the similarity is smaller than T, $\bar{N}_j$ will not be classified into the false seizure class. All false seizure detections classified as elements of false seizure class will be eliminated and others are retained. True seizures which are classified as elements of the false seizure class are unfortunately lost.

2.3 Classifier Training

The goal of the training is to determine a patient-specific optimal *similarity threshold* for a classifier. It is easy to understand that the smaller the similarity threshold is, the more FSDs will be eliminated; and also the larger is the probability of losing seizures. As a result, a good compromise is to select a similarity threshold that leads to a high FSD eliminating rate and a reasonably low probability of losing seizures. This can be done by using FSDs from the second monitoring session to determine a predictive FSD eliminating rate of this patient in the future, and a seizure reference set to determine a predictive probability of losing seizures of this patient. Therefore, three sets of data are required to train a patient-specific classifier. They are: patient-specific false

seizure prototypes, true seizure reference data set and patient-specific false seizure detection training data set. Their definitions follow. Selection criteria of these data sets are discussed in the evaluation section.

1. *Patient-specific false seizure prototypes* are false seizure detections acquired from a trial monitoring session of a patient, and are called the initial FSD set. The trial monitoring session is usually the first monitoring session.
2. *True seizure reference data set* is a collection of true seizure patterns from a lot of patients. It can be used to represent possible seizure patterns in a patient. This reference set is a collection of seizures from a lot of patients instead of the patient currently under study because the patient may not have any seizure in the first monitoring session or only has too few seizures to form enough training data.
3. *False seizure detection training data set* is a collection of false seizure detections from the second monitoring session, which is the session subsequent to the trial session.

The true seizure reference data set and the false seizure detection training data set are used to determine how large should the area around the prototype points be to ensure a good elimination of false detection and avoid losing genuine seizures.

A patient-specific similarity threshold is determined by using the following optimal similarity threshold selection criterion function:

$$J(T) = \max_{T}\left(\frac{RateFSD(T)}{RateTSD(T)+K}\right).$$

where *Rate*FSD(T) is the patient-specific false detection eliminating rate at a given similarity threshold T and *Rate*TSD(T) is the percentage of seizures from the seizure reference set lost for a given distance threshold T. K is a constant to avoid J(T) becoming infinite when *Rate*TSD(T) is zero. The optimal T is the one with the highest value of J(T).

We will first illustrate in an example the selection of the optimal value of similarity. After the initial FSD set was acquired, results from the second monitoring session were obtained (figure 3-3). Curve 1 shows the percentage of FSDs eliminated from the second session as a function of different similarity thresholds. The percentage of seizures lost from the true seizure reference set is shown in curve 2. The ratio of percentage of FSDs eliminated to seizures lost from the seizure reference set is shown in curve 3 (a constant is added to the denominator of the ratio to avoid a division by zero when no seizure is lost; the constant remains the same for every patient). From curve 3, it is obvious that the best threshold of similarity is 74 because it corresponds to the highest ratio. That threshold was used for the five subsequent monitoring sessions in this patient, resulting in the elimination of 88% of false detection; 2.9% of the seizures of the true seizure reference set fell within this patient's false detection subspace, giving an indication of the probability that true seizures would be lost in this patient.

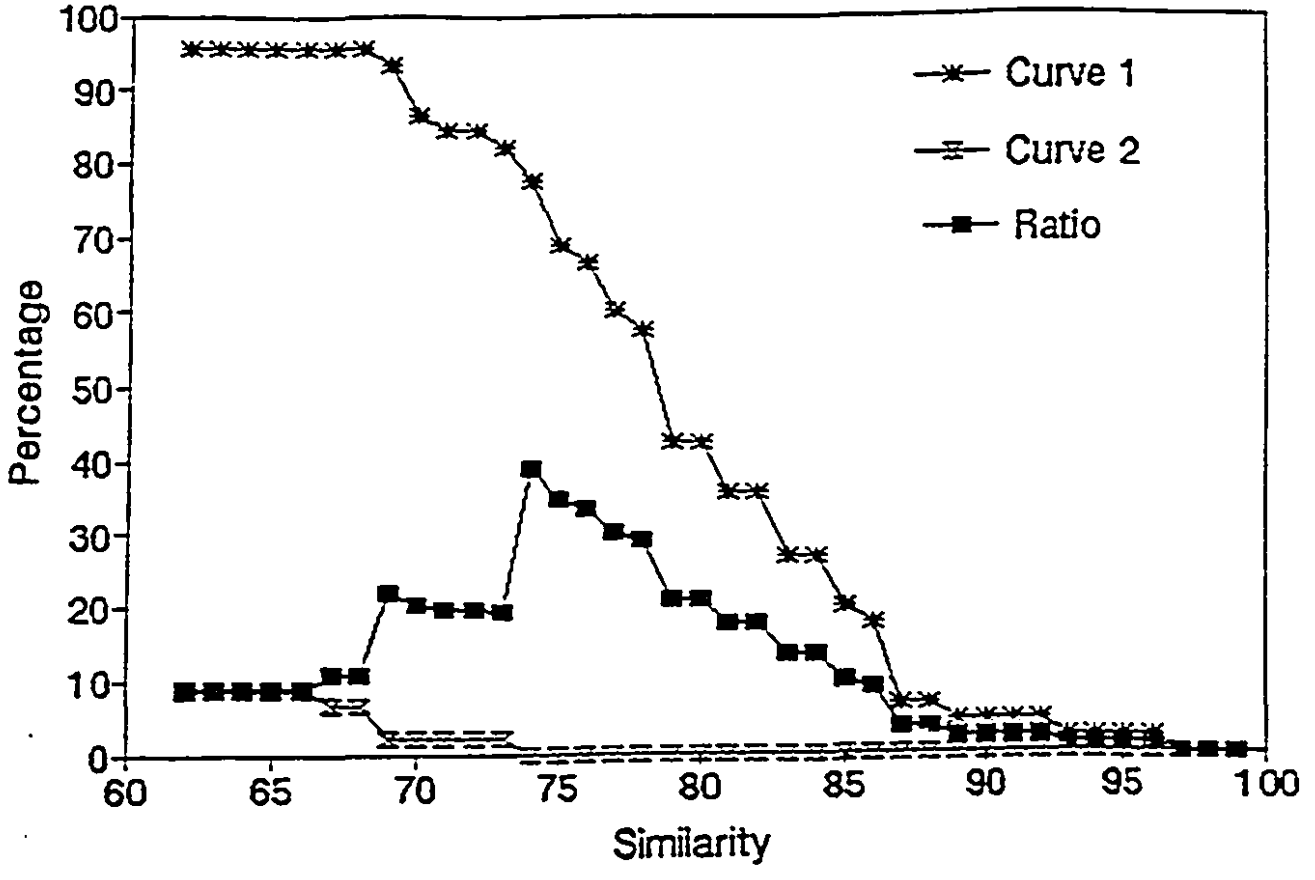


Figure 3-3: Selection of the optimum similarity threshold. The X axis represents the similarity threshold. Curve 1 is RateFSD(T): it shows the percentage of FSDs from the FSD training data eliminated as a function of similarity threshold. Curve 2 is RateTSD(T): it represents the percentage of lost seizures from the seizure reference set. The last curve is J(T): it is the ratio of percentage of FSDs eliminated to percentage of lost seizures (a constant is added to the denominator of the ratio to avoid a division by zero when no seizure is lost). The ratio curve has a maximum when the similarity is 74; this becomes the threshold for this patient in subsequent monitoring sessions.

3 Implementation

The implementation of the method requires only few human interventions, which makes it easy to use. When many FSDs occur in a patient for two monitoring sessions, a patient-specific classifier with the optimal similarity threshold can be trained automatically by simply providing FSDs. Details of the implementation procedures follow.

1. All false seizure detections from the first monitoring session should be collected as the original patient-specific false detection prototypes.
2. When false detections are available from the next monitoring session, they are used to train a patient-specific classifier to maximize the probability of eliminating false seizure detections in the future. In addition, a set of true seizures collected from a lot of patients are used to train the classifier to minimize the probability of missing seizures in the future monitoring sessions of the patient.
3. For all following monitoring sessions, every detection has to pass the classifier. The ones which belong to the false seizure class will be eliminated. The rest are retained.

Figure 3-4 illustrates the implementation procedure of our method, including the use of FSDs and the seizure reference set, and the training of a classifier. FSDs from the first monitoring session are used to form the false seizure prototype points. FSDs from the second session, prototypes and the seizure reference set are all used to train a classifier. This classifier is then used in the third and subsequent monitoring sessions of the same patient to reduce the patient-specific FSDs. More details about the implementation can be seen in Appendix A in the form of a block diagram.

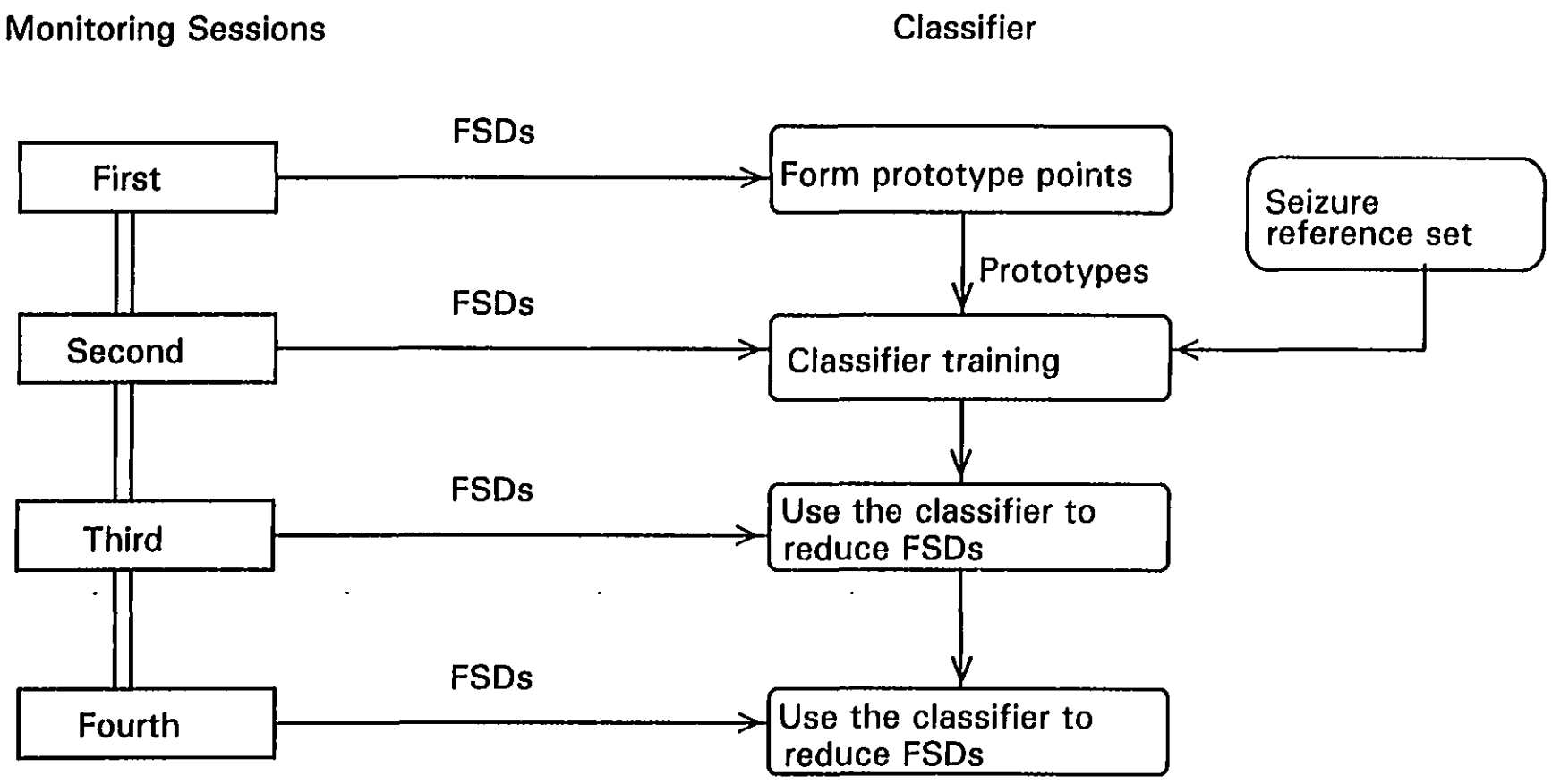


Figure 3-4: Implementation procedure of our method: FSDs from the first monitoring session provide false detection prototype points, which are used together with FSDs from the second session and the seizure reference set to train a patient-specific classifier. The classifier is then used to reduce FSDs during subsequent monitoring sessions.

4 Evaluation

The evaluation was done on data from 20 patients. They belong to two groups: 10 patients with depth electrodes and 10 patients with scalp electrodes. The implementation procedure discussed above was used to evaluate the reduction in false detections (figure 3-4). Therefore, for each patient, FSDs from the first monitoring session were used to form the false detection prototypes; FSDs from the second monitoring session and a seizure reference set were used to train a patient-specific classifier; and the rest of the monitoring sessions were used for testing. The seizure reference set was also used to evaluate the probability of losing seizures by using the rotation method.

The collection of both training and testing data will be discussed first. The evaluations of the reduction of FSDs and the probability of losing seizures are explained in detail next.

4.1 Data Collection

In order to evaluate our method, we collected the following data: (1) false detections in several monitoring sessions of each patient to constitute false detection prototypes, to train classifiers, and for the evaluation of the reduction of FSDs. (2) true seizure detections to form the seizure reference set for the training of classifiers and the evaluation of the probability of losing seizures. At the Montreal Neurological Hospital, a bipolar montage of 32 channels for scalp electrodes recordings and one of 16 channels from depth recordings are used clinically for seizure detection. In order to truly evaluate the practical usage of the method, we used the same montages in all our evaluations.

4.1.1 False Detections:

False detections came from twenty patients who were selected from the 114 patients who were subjected to long-term monitoring at the Montreal Neurological Hospital from January 1991 to April 1992. Two criteria were used in the selection of these patients:

(1) Patients had to be monitored from 4 to 11 consecutive days. For each patient, all selected monitoring sessions had to have the same montage and the same detection thresholds.

(2) The FSD rate had to be high: patients with depth electrodes had to have a FSD rate higher than 2/hour and patients with scalp electrodes a rate higher than 1.5/hour.

The reason for the use of a different FSD rate in different group of patients is that usually the FSD rates are higher in patients with depth electrodes than in patients with scalp electrodes. Although all EEGs were recorded with 32 channels, seizure detection was performed in 16 channels for patients with depth electrodes, and in 32 channels for patients with scalp electrodes.

In 10 patients with depth electrodes, 64 monitoring sessions were collected with 5,029 FSDs (detection sections) during 1,301 monitoring hours. The average FSD rate was 3.86/hour. In 10 patients with scalp electrodes, there were 70 monitoring sessions with 4,195 FSDs during 1,325 monitoring hours. The average FSD rate was 3.17/hour.

4.1.2 True Seizure Detections:

True seizure detections are used to constitute the *true seizure reference set*. Two seizure reference sets are needed for two groups of patients: patients with depth electrodes and patients with scalp electrodes. This is because seizure patterns from

patients with depth electrodes are different from those with scalp electrodes, as we discussed in the chapter 2 "Literature Review".

All seizures in both seizure reference sets satisfy three criteria: (1) They are all small seizures which have a small number (10 or less) of detection points (each detection point represents an epoch of the EEG which is detected as a seizure by the classic seizure detection method). A seizure including a large number of detection points is not very likely to be eliminated by our method since it is unlikely that all of its detection points would belong to the false detection subspace. For this reason, we selected only small seizures as the seizures at risk to be lost by our method. We estimated that small seizures represent approximately half of all the seizures detected by our system. (2) They include a large variety of seizure patterns so that they can be as representative as possible. (3) They do not come from the 20 selected patients for the evaluation of the reduction of FSDs. This is because patients rarely have seizures in the first one or two monitoring sessions in the practical clinical environment. As a result, only seizures from other patients can be used to train classifiers. Moreover, since each seizure reference set was collected in advance and then used to train classifiers for all patients in each group later on, no seizure from these 20 selected patients could be included in the seizure reference sets.

Forty-four seizures were obtained from 13 patients with scalp electrodes, with an average of 4 seizures and maximum of 10 seizures per patient. This set of data is called seizure set A. Forty-nine seizures were collected from 10 patients with depth electrodes, with an average of 5 seizures and maximum of 10 seizures per patient. This set of data is named seizure set B.

4.2 Estimation of FSD Elimination Rate

The holdout method was used to estimate the FSD elimination rate because it is possible for us to obtain a large amount of FSDs from several monitoring sessions of each patient. Another reason is that we want to test our method in a practical clinical situation since the acquisition of FSDs has its time sequence. Our method is designed to function as soon as enough FSDs are acquired from the first two monitoring sessions. If the rotation method were used, some FSDs acquired from monitoring sessions later than the first two would be used to train the classifier and FSDs from the first two sessions would be used as testing data. This does not represent the actual clinical environment. Therefore, the rotation method is not suitable in this situation.

In the evaluation, only FSDs from the first two monitoring sessions were used to form false detection prototypes and to train the classifier for each patient, and the rest were used for testing, as indicated in figure 3-4. For instance, if there were four monitoring sessions in a patient, the FSDs from the first monitoring session represents the false detection prototype points; the FSDs from the second session are used together with a seizure reference set to train a classifier specific to the patient. FSDs from the remaining two monitoring sessions then passed the classifier to test the reduction in false detection rate. On average, each classifier was tested on FSDs acquired during about 100 hours of monitoring. Results are shown in the "Results" section..

4.3 Estimation of Error Rate

In the estimation of the error rate (the probability of eliminating genuine seizures), the rotation method is used. The seizure reference set is used for both training and testing. There are four reasons: (1) The error rate could be estimated by counting seizures eliminated by our method in the same monitoring sessions as for the evaluation

of the reduction of false detections. However, the number of seizures occurring in those sessions is too small to ensure a small variance in the result. We therefore had to turn to the seizure reference set to estimate the error rate because they contain many seizures with a large variety of patterns. (2) Since we do not know the characteristics of the seizures of a particular patient at the time of the first monitoring session, we use a large group of seizures as a representation for these unknown patterns. However, because of the difficulties in the collection of seizures, the number of seizures included in each seizure reference set is still not very large (42 seizures for scalp electrode patients and 49 for depth electrode patients). Therefore, the holdout method is not suitable. The rotation method can achieve a better, unbiased estimation when the data set is small (Devijver and Kittle, 1982). (3) Since the seizures are pooled together, the sequence of occurrence is no longer a problem as it is in the evaluation of the reduction of FSDs. (4) If the substitution method were used, results might be misleading since our classifier is a non-parametric one and the data set is too small (Devijver and Kittle 1982).

In order to use the rotation method, each seizure reference set has to be divided into subsets first. The criterion for the division is: each subset contains all the seizures from one patient only. As a result, the seizure reference set A (scalp electrodes) was divided into 13 subsets, while there are 10 subsets in the seizure reference set B (depth electrodes). Details about the division of subsets, the rotation sequence and the averaging method are discussed below.

The evaluation of the error rate was carried out in the procedure described in figure 3-5. A classifier is provided with a set of false detection prototype points from a patient, is trained with a set of FSD training data of the same patient and with a portion of a seizure reference set (seizure training data), and then tested with the other portion of the seizure reference set (seizure testing data). In each rotation, seizure training data and seizure testing data are changed, while prototypes and FSD training data remain the

same. Of course, when another classifier (corresponding to another patient) is used for the evaluation, both prototypes and FSD training data will be changed.

4.3.1 Scalp Electrodes

The seizure reference set A was divided into 13 subsets because seizures come from 13 patients (N=13) with scalp electrodes. The number of seizure in each subset is shown below.

Subset	1	2	3	4	5	6	7	8	9	10	11	12	1
SEZ	1	1	1	10	10	1	4	1	4	2	1	2	4

By using these seizure subsets, each classifier has to be trained by twelve seizure subsets and tested on the one left for each rotation. For instance, if a classifier, which is named classifier #1, is trained by data containing subsets #2 to #13 and tested on subset #1, an error rate e_1 is obtained. Then after the same classifier is trained by the other 12 subsets and tested on subset #2, another error rate e_2 is obtained. When each subset has been the testing data once and only once, a total of 13 error rates has been obtained. The average error rate for classifier j is therefore: $E_{Aj} = \frac{\sum_{i=1}^{N} e_i}{N}$.

The same procedure applies to the other classifiers and 10 error rates can be obtained (P=10) because there are 10 patients in this group. The average error rate of classifiers trained and tested by seizure set A is: $E_A = \frac{\sum_{j=1}^{P} E_{Aj}}{P}$ and the standard deviation is: $SD_A = \sqrt{\frac{\sum_{j=1}^{P} (E_A - E_{Aj})^2}{P}}$.

4.3.2 Depth Electrodes

The seizure reference set B was divided into to 10 subsets (N=10) because seizures were contributed by 10 patients with depth electrodes. The number of seizures in each subset can be seen below.

Subset	1	2	3	4	5	6	7	8	9	10
SEZ	10	7	6	6	4	2	1	3	4	6

By using the same procedure as in the section of scalp electrodes, the classifier j has an average error rate at: $E_{Bj} = \frac{\sum_{i=1}^{N} e_i}{N}$ and N=10.

Since there are 10 classifiers (P=10, one classifier for each patient) in this group, 10 error rates can be obtained. The average error rate of this set of classifiers is: $E_B = \frac{\sum_{j=1}^{P} E_{Bj}}{P}$ and the standard deviation is: $SD_B = \sqrt{\frac{\sum_{j=1}^{P} (E_B - E_{Bj})^2}{P}}$.

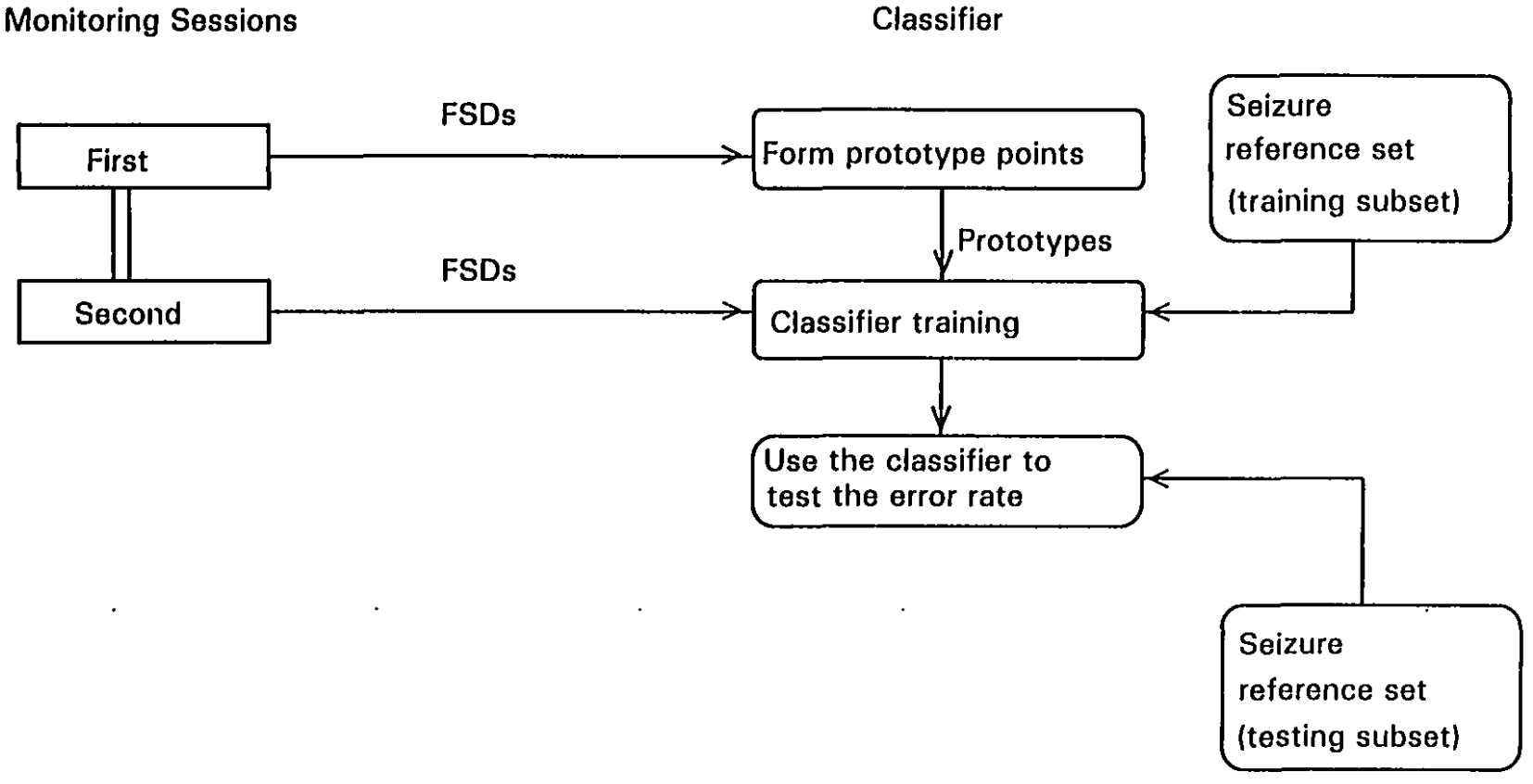


Figure 3-5: Diagram illustrating the evaluation of the probability of losing seizures: For a particular patient, the classifier specific to him is trained with a same set of prototype points and the same FSD training data, but with different seizure reference training subset in different rotations. A seizure reference testing subset, which changes in each rotation, is used to test the error rate.

5 Results

After the evaluation proccdures were executed, results were computed and will be discussed in this section. False detection elimination ra'es will be presented first, followed by probability of losing seizures. Some examples of results will be shown at the last.

5.1 False Detection Elimination Rate

A total of 20 patients with FSDs acquired from 2,0681 hours of monitoring were used in the evaluation of the reduction of false detections. The average false detection rate in the classic seizure detection method of Gotman (1982, 1990a) is 3.25/hour in these monitoring sessions. By using our method, the false alarm rate was reduced to 1.26/hour in the same set of data. Average results from all the patients indicated that most FSDs (61%) could be eliminated. There was a higher FSD elimination rate for depth recordings (71%) than that for scalp recordings (50%). This is probably because most FSDs from depth recordings are more similar to each other than FSDs from scalp recordings. Results are shown in detail in table 3-1.

Table 3-1: Results of the false detection elimination rate

	Depth	Scalp	Total
Number of patients	10	10	20
FSDs acquired from (hours)	997.5	1,070.6	2,068.1
Original FSD rate (/hour)	3.63	2.90	3.25
New FSD rate (/hour)	1.07	1.45	1.26
Elimination rate	.70.6%	50.0%	61.1%

5.2 Probability of Losing Seizures (Error Rate)

Results are shown in tables 3-2 and 3-3. Each "cls" represents the classifier corresponding to one patient. In each column starting with "cls", the error rate of each rotation is presented, with the average error rate for this classifier in the last row. For instance, in table 3-2 for patients with scalp electrodes, classifier #1, which is specifically for patient #1, has an error rate in each of 13 rotations. In all these rotations, this classifier has the same false detection prototypes, and is trained with the same FSD training data (figure 3-5). However, it is trained with different partitions of the seizure reference set and tested on the other, according to the details discussed in the section of "Evaluation". The average error rate for this classifier is 7.72%. A similar procedure was applied to the other classifiers in this group, as well as classifiers in the group of patients with depth electrodes. The number of seizures used in the testing set for each rotation is shown in the second column from the left. The overall average error rate and its standard deviation are shown in the last row of the last two columns.

The average error rate is 2.78% for patients with scalp electrodes and 2.58% for patients with depth electrodes. This indicates that the method has a very low error rate.

In other words, the method has a very small probability of eliminating a true seizure detection.

5.3 Examples:

Examples of FSDs which are successfully eliminated are shown in figure 3-6. Pattern A is typical alpha activity and pattern B is an artifact. The EEG pattern C is a rhythmic burst of unknown significance in the hippocampus. It occurred in a patient every night and caused a lot of FSDs. Such patient-specific FSDs can be frequent in long-term EEG monitoring and they were largely eliminated by our method.

A small number of seizures from the reference set were lost. The reason is that these seizures were small and also similar in some of their electrographic characteristics to the FSDs of the patient. An example is shown in figure 3-7. Pattern A is a seizure in the reference set and pattern B is a FSD from the initial FSD set in a patient: if that seizure had occurred in that patient, it would have been lost as a result of our procedure.

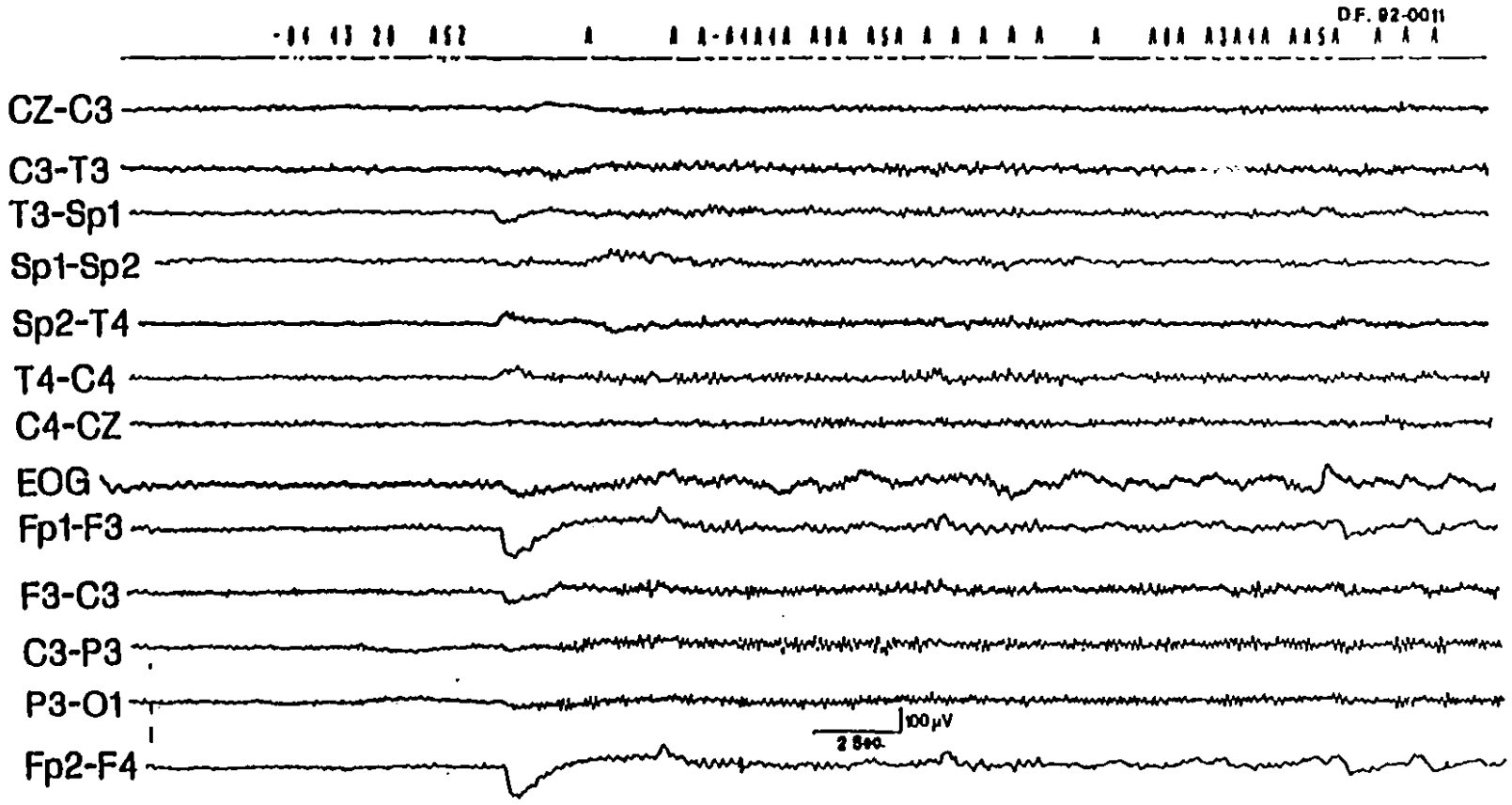
Not all FSDs were eliminated by using the algorithm. An example is shown in figure 3-8. This is a FSD pattern in a new monitoring session and it is very different from most of the FSD patterns from the initial FSD set (figure 3-6, pattern C) of the same patient. When a FSD pattern is not similar to the patterns found in the initial FSD set, it cannot be eliminated by our method.

Table 3-2: Results of the error rate from patients with scalp electrodes: "R" means rotation, "Err" means error rate, "tst" means number of seizures in a seizure reference testing set, "cls" means classifier, "Avg" means average and "SD" means standard deviation. The same abbreviations apply to table 3-3.

	tst	cls1	cls2	cls3	cls4	cls5	cls6	cls7	cls8	cls9	cls10	Avg	SD
R1	1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%		
R2	1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%		
R3	1	100%	0%	0%	0%	100%	100%	0%	0%	0%	0%		
R4	10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%		
R5	10	0%	10%	10%	20%	0%	0%	10%	10%	0%	0%		
R6	1.	0%	0%	.0%	0%	0%.	0%	0%	0%	0%	0%		
R7	4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%		
R8	1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%		
R9	4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%		
R10	2	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%		
R11	1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%		
R12	2	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%		
R13	4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%		
Err		*7.72%*	*0.77%*	*0.77%*	*1.54%*	*7.72%*	*7.72%*	*0.77%*	*0.77%*	*0.0%*	*0.0%*	*2.78%*	*3.26%*

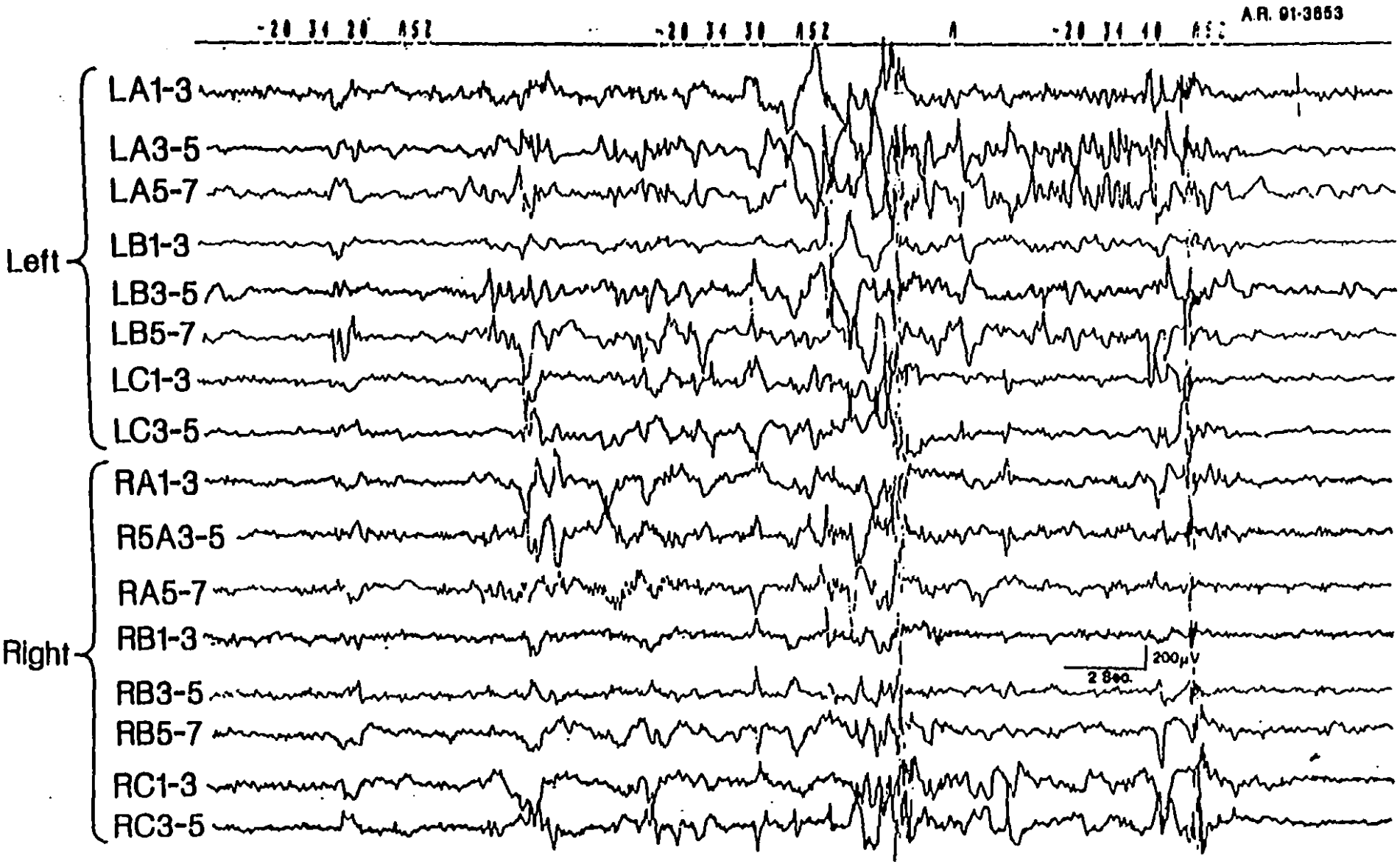
Table 3-3: Results of the error rate from patients with depth electrodes

	tst	cls1	cls2	cls3	cls4	cls5	cls6	cls7	cls8	cls9	cls10	Avg	SD
R1	10	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%		
R2	7	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%		
R3	6	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%		
R4	6	0%	0%	0%	0%	0%	0%	0%	0%	16.6%	0%		
R5	4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%		
R6	2	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%		
R7	1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%		
R8	3	0%	0%	33.3%	0%	0%	0%	0%	0%	0%	0%		
R9	4	0%	25%	0%	0%	0%	50%	25%	25%	0%	0%		
R10	6	0%	0%	0%	16.7%	16.7%	0%	0%	0%	16.6%	33.3%		
Err		*0%*	*2.50%*	*3.33%*	*1.67%*	*1.67%*	*5.00%*	*2.50%*	*2.50%*	*3.33%*	*3.33%*	*2.58%*	*0.96%*



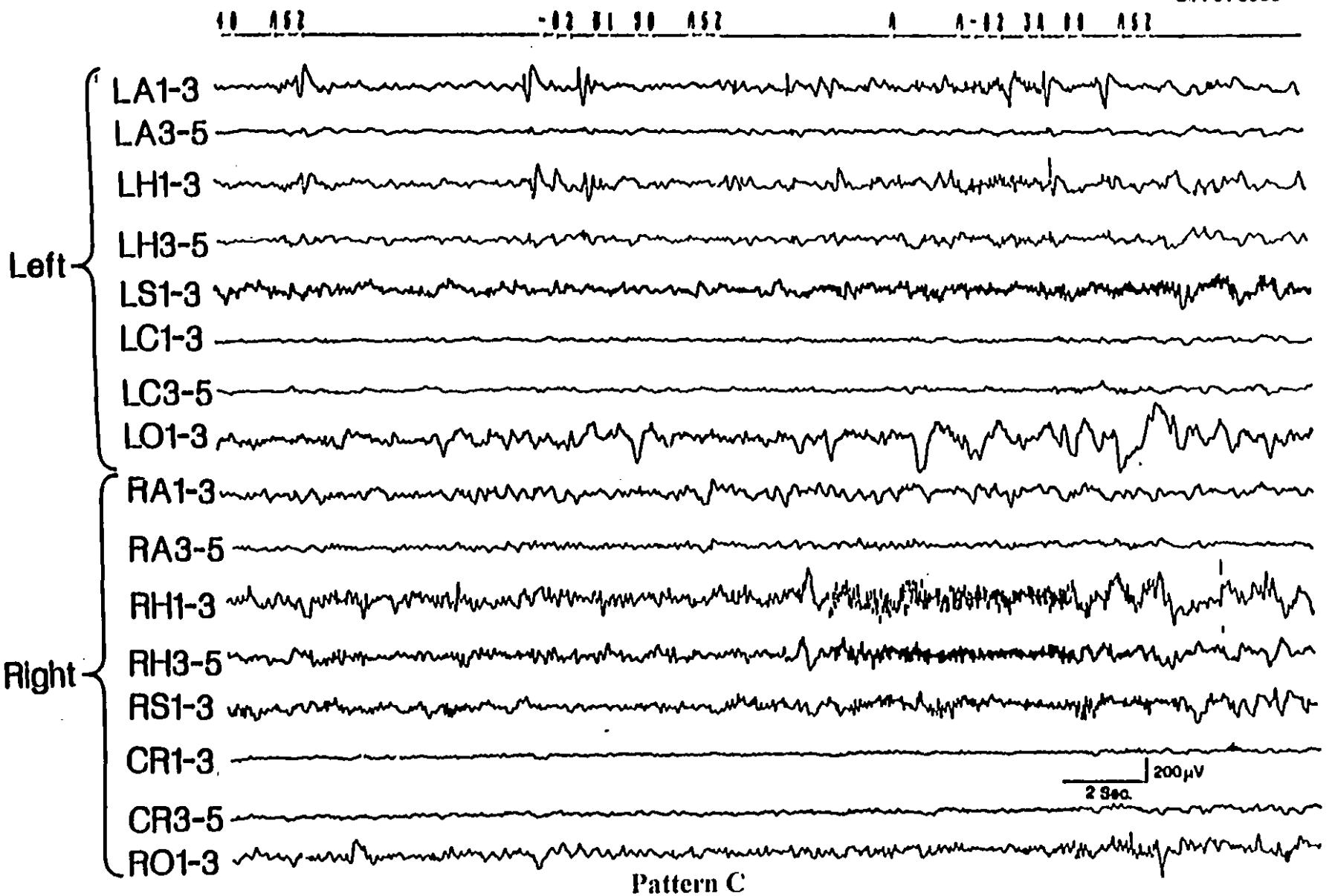
Pattern A

Figure 3-6: Examples of FSDs eliminated. These are EEG patterns which caused a lot of FSDs in some patients; most of them were eliminated. Pattern A is a typical alpha activity pattern and pattern B is an artifact. Pattern C is a rhythmic burst of unknown significance occurring in the hippocampus. On the top of the figure, "ASZ" indicates this section is an automatic seizure detection section and "A" indicates the time a detection occurred. The time of day is also indicated. These symbols stay the same in all following figures.


A.R. 91-3853
Left
LA1-3
LA3-5
LA5-7
LB1-3
LB3-5
LB5-7
LC1-3
LC3-5
Right
RA1-3
R5A3-5
RA5-7
RB1-3
RB3-5
RB5-7
RC1-3
RC3-5
200μV
2 Sec.

Pattern B

B.F. 91-3506



Pattern C

6 Discussion

We presented a procedure for the reduction in false detections in an automatic seizure detection method used during long-term monitoring of epileptic patients. It is difficult to reduce false detections significantly with a method applicable to all patients because seizure patterns as well as the main causes of false detection are quite different from patient to patient. The same pattern can be a cause for frequent false detections in one patient but be characteristic of a seizure in another. Since automatic seizure detection is most useful in the context of long-term monitoring sessions which can last one, two or several weeks, we decided to use information gained in the first one or two days of monitoring to learn about the patterns specific to a particular patient.

Instead of teaching the system about a patient's seizures, we elected to teach it about the patient's false detections. In our experience, when false detections are frequent it is usually because one paroxysmal pattern tends to repeat often. We developed a measure of similarity between a new pattern and the patterns of false detections in each patient to determine whether the new pattern resembles known false detections. We have applied this measure of similarity to the seizure detection method of Gotman (1982, 1990a) but it could also be applied to any seizure detection method for which each detection is characterized by a set of variables.

This algorithm can eliminate not only FSDs caused by normal or abnormal EEG patterns, but also artifacts, as long as these patterns are persisting for several monitoring sessions. This is particularly important because artefact cause more than 80% of FSDs (Pauri et al. 1992). In figure 3-6, we can see that artefact, normal and abnormal EEG patterns which cause a lot of FSDs can be eliminated well. Some FSDs are not eliminated yet. The main reason is that these FSDs have EEG patterns which were not included in the initial FSD set from the trial session (as shown in figure 3-8). This

problem could be solved by adding new FSD patterns, whenever they occur, to the initial FSD set.

A small number of seizures were lost. The main characteristics of these seizures are (1) they are relatively similar to some FSDs patterns in an initial FSD set from the patient. An example can be seen in figure 3-7. (2) These seizures have a few detection points only. It is much more likely to lose a seizure with only a few detection points than seizures with many detection points. The results show that the probability of losing small seizures is tolerable (below 3%). Although we have not measured it, the probability of losing larger seizures is certainly much lower. Since a significant reduction of FSD rates could lead to the use of lower detection thresholds and consequently to a rise of seizure detection sensitivity, this small price appears justified. The increase of seizure detection rate by lowering detections thresholds (Pauri et al. 1992) can be much higher than the rate of the loss of seizures caused by this method.

Although this method was evaluated off-line, it can be implemented on-line to work with detection programs because it does not require a lot of memory space and the computations are not very complex. If there are M detection points in an initial FSD set, a maximum of M distance computations are required to determine if a new detection point belongs to the FSD subspace. Since M can be typically of the order of several hundred, the calculation is not too long.

Although there are not many papers about seizure detection, several aspects of this field can be discussed. For instance, even with our method, there is still a false detection rate at about 1.3/hour despite the fact that our method reduced the rate by an average of 61%. This rate could be reduced further so that lower detection thresholds could be used and consequently more seizures could be detected. Although the estimated probability of losing seizures in our method is very low (below 3%), the actual probability could be even lower because we only took small seizures into account in the error estimation, and larger seizures are much less likely to be lost than small seizures. As a result, the

optimal distance coefficient could be higher than we have now: more FSDs could be eliminated, while the probability to lose a seizure would remain low and acceptable. This can be done by changing the constant K in the optimal similarity selection.

Reduction of FSDs can be achieved by our method, and it could be achieved by other methods as well. One way is to use a patient-adaptive algorithm to vary the detection thresholds in a detection program. When FSDs are rare in a certain period of time, the algorithm could lower the detection threshold so that more seizures and more FSDs would be detected, as proven by Pauri et al. (1992). However, when FSDs are frequent, the algorithm could raise the threshold according to the FSD patterns and therefore FSD rate will decrease with the disadvantage that some seizures may be missed. As long as the algorithm balances the trade-off between the elimination of FSDs and the probability of missing detecting seizures, the overall performance of the detection method could improve. One of the important aspects of such an algorithm is that it would keep the FSD rate at a constant level for all patients at all time. This will make full use of the storage capacity of the computer. If it is used with our method, the system can have a better seizure detection rate than the original one or the one with our method alone. Another aspect is that this algorithm is an unsupervised one and therefore even the minimum human interventions, as required by our method, would not be necessary any more.

We used an orthognal space for the detection and assume the independence between features. Although we have not tested their independence, we used different features to describe different characteristics of the EEG and therefore it is reasonable to assume their independence. Moreover, the satisfactory performance of the method proves the usefulness and effectiveness of the features. Of course, the independence of features should be tested systematically and future performance and efficiency could be improved by using a better set of features.

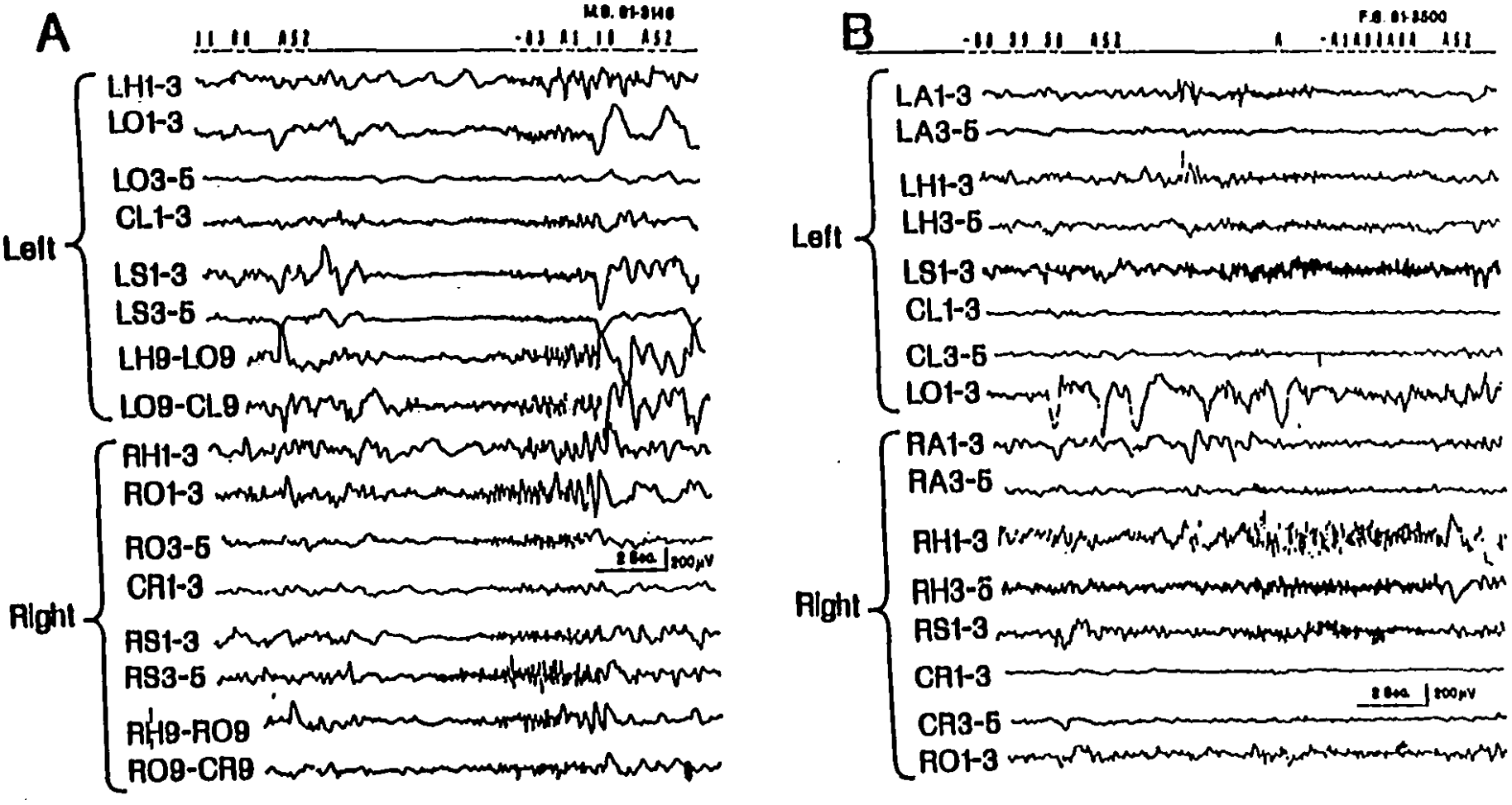


Figure 3-7: Example of a seizure lost because it is similar to a FSD: Pattern A is a seizure from the seizure reference set and Pattern B is a FSD from the initial FSD set of a patient. In figure B the detection took place in channel RH1-RH3, LS1-LS3 and RH3-RH5; in figure A the detection was in channel RS3-RS5. The similarity between the detected pattern in A and the false detection in B resulted in the elimination of the seizure in A.

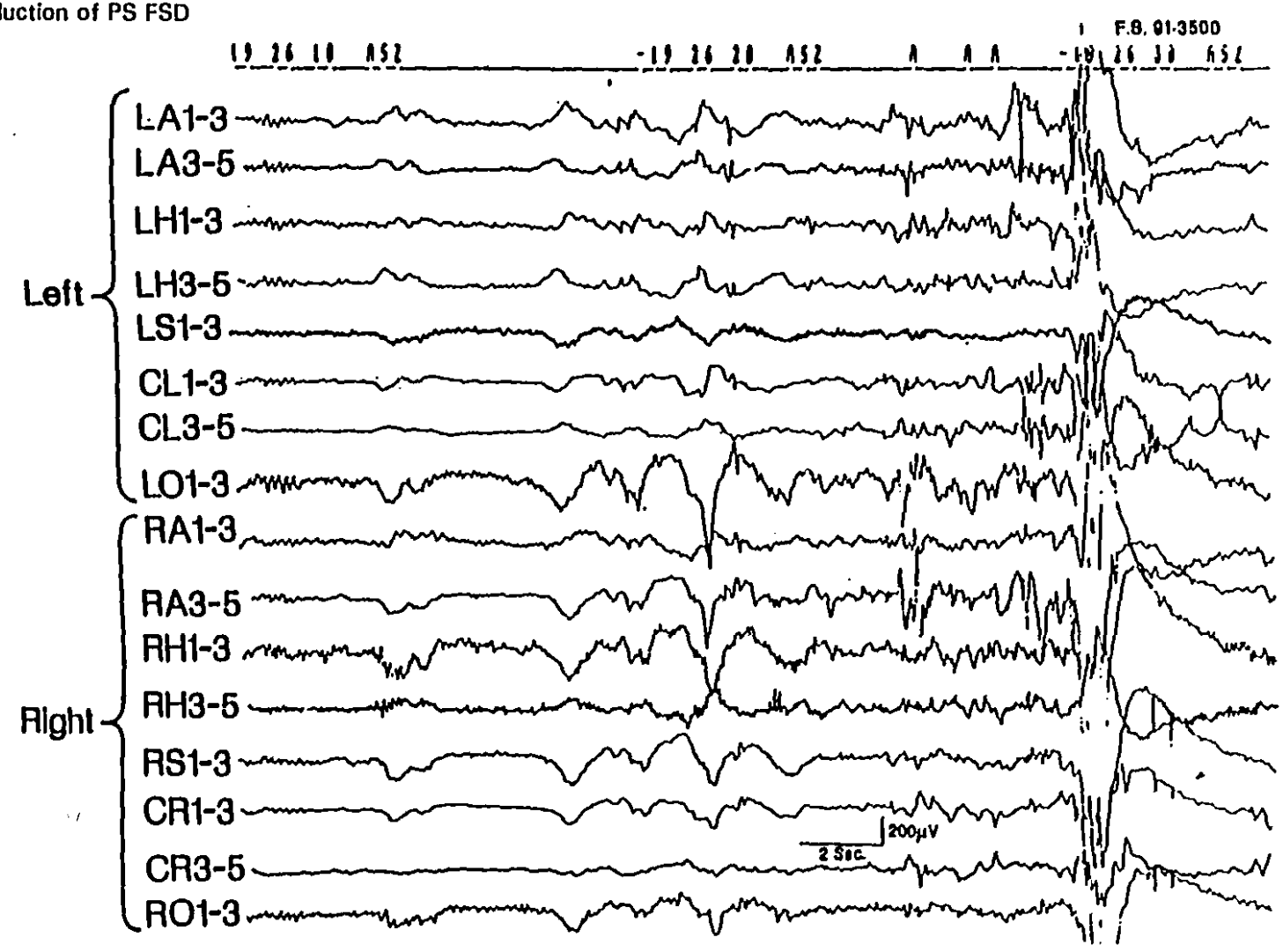


Figure 3-8: Example of a FSD retained. This EEG pattern of a FSD was detected and not eliminated because it is different from the same patient's FSDs which occurred during the trial monitoring session (figure 3-6, pattern C).

Chapter 4:

Seizure Onset Detection

1 Introduction

Changes in the EEG at the onset of a seizure are of considerable interest because of the information they reveal about an epileptic focus. Detection of seizures at an early stage may allow precautions to be taken to avoid danger to the patient, and may improve observation of early behavioral changes, and allow behavioral testing to better define the anatomical structures involved in the epileptic focus. Once a seizure is fully developed, this subtle information about location may be unavailable.

Interaction with an epileptic patient during the early part of a seizure is very important. As a part of the normal procedure for caring for an epileptic patient in the hospital (Engel 1989), the observation during the ictal phase includes: (1) the type and anatomical distribution of movements at the beginning of the seizure; (2) initial alterations in consciousness; (3) responsiveness and memory during the ictal event, and

(4) postictal neurological and mental deficits. These observations can help to determine the type of seizure. For instance, complex partial seizures have symptoms similar to those of simple partial seizures, with an additional impairment of consciousness at seizure onset. The consciousness of a patient cannot be known if there is no interaction between observer and patient. In addition, the earlier the interaction takes place after seizure onset, the better the information to determine the type of seizure. Many patients are not aware of their seizures; as a result, a seizure warning system is very useful to give observers a chance to interact with patients as early as possible.

There are three difficulties in the detection of all seizure onset: (1) The early seizure pattern is highly variable from a patient to another. (2) Although it is sometimes abrupt, the seizure onset often involves small changes and evolves into a full seizure pattern during a period of about 20 seconds. (3) The seizure onset in one patient can be very similar to non-seizure EEG patterns of other patients, especially when artifacts and muscle activity are involved. This is why the seizure onset is more difficult to detect than prominent seizure patterns occurring later in the seizure. This is also why most seizure detection algorithms (Gotman 1982, 1990a; Murro et al. 1991; Liu, et al. 1992) aim at detecting prominent seizure patterns rather than early seizure patterns. In those algorithms, a seizure onset might be detected if there is a clear, abrupt pattern at onset, but the detection can only be reported about a dozen seconds later. This is because the system waits during this time to make sure the detection is valid. When early seizure patterns are not prominent enough, those algorithms cannot detect them.

The following observations may help design a system for seizure onset detection: In most patients, one or several types of seizures tend to repeat all the time. This is reasonable since the sources that cause seizures in a patient usually do not change during the usual period of observation and therefore similar seizures tend to occur repeatedly in a patient. Although this phenomenon has not been proven by a large scale study, it has been our conclusion based on extensive clinical experience. In these cases, the seizures of

each type are very similar to each other, including the seizure onset. For a particular patient, the seizure onset is most often distinguishable from the EEG background, although it may resemble some background patterns of other patients, as mentioned above.

On the basis of these observations, we propose a method to detect patient-specific seizure onsets by using template matching. Once a seizure occurs in a patient, it can be memorized. During prolonged EEG monitoring sessions subsequent to the recording of that seizure, if an EEG pattern has a good match to the stored seizure, it will be reported immediately as a seizure onset. However, extreme caution has to be taken to avoid false alarms because frequent alarms will annoy the staff and patient, and be ignored. In addition to detecting early seizure onset, this method has the potential of detecting some seizures that are missed by traditional detection methods: if one seizure is missed by a detection method, similar seizures which occur later are also likely to be missed. With the template matching method, if one of this kind of seizures is somehow captured, all other similar seizures will probably be detected.

Although it could appear that such a method could be used for standard seizure detection, this is not the case because standard detection must be able to detect all kinds of seizures, not just seizures having a known pattern.

2 Method

This method relies on the availability of one seizure for each patient. A patient-specific classifier is trained by using this seizure to detect the onset of subsequent occurrences of similar seizures. The algorithm will be discussed first, followed by the implementation procedure and the evaluation of the method.

The method uses six features to reveal important characteristics of ictal EEG patterns. After a seizure and some interictal EEG are acquired from a patient, a classifier, which is specific to the patient, is trained with the patient's data only. In subsequent EEG monitoring sessions of the same patient, the classifier is used to determine if a seizure onset occurs. If it does, an alarm is triggered.

2.1 Algorithm

Features used in this method are discussed first, followed by the criteria in selecting a template. The distance measure is then presented and finally the design and training of the classifier are discussed.

2.1.1 Feature Extraction

2.1.1.1 Epoch Selection

The EEG is broken down into sections, or epochs, for the purpose of feature selection. The length of an epoch depends on the type of application. In sleep research, for instance, the analysis is done in epochs of 30 seconds (Gath and Bar-on 1985) or longer (Friedman and Jones 1984). In the study of seizures, epochs of 2 seconds (Gotman 1982; Hilfiker and Egli 1992), 6 seconds (Liu et al. 1992) or 6.83 seconds

(Murro et al 1991) have been used. The reason that shorter epochs are used in seizure analysis than in sleep research is that seizure patterns change faster than sleep stages. In our method, an epoch of 2.56 seconds is used for three reasons: (1) It is just long enough to capture statistical characteristics of EEG and short enough to capture the evolution of seizures. (2) Since the aim of the project is to detect seizure onsets as soon as possible, a short epoch length may increase the chance of early detection. (3) Since the EEG is digitized at a sampling rate of 200Hz, an epoch of 2.56 seconds contains 512 samples. It is a convenient length to compute the Fast Fourier Transform (FFT).

2.1.1.2 Features

There is a total of six features in this method, including a special one to describe spatial information about electrodes. These six features are: average wave amplitude, average wave duration, coefficient of variation of wave duration, dominant frequency, average power in a main energy zone, and the location feature.

1. Average wave amplitude in one epoch: The waveform decomposition method of Gotman (1982) basically divides the EEG into halfwaves. The amplitude and duration of each halfwave can therefore be measured. The average wave amplitude is the average amplitude of halfwaves in one epoch. It is expressed as: $AvgAmp = \frac{\sum_{i=1}^{N} Amp_i}{N}$, where N is the number of halfwaves in one epoch and Amp_i is the amplitude of a halfwave.

2. Average wave duration in one epoch: It is expressed as: $AvgDur = \frac{\sum_{i=1}^{N} Dur_i}{N}$, where N is the number of halfwaves in one epoch and Dur_i is the duration of a halfwave.

3. Coefficient of variation of wave duration in one epoch: It is expressed as: $COVA = \frac{\sum_{i=1}^{N}(Dur_i - \overline{Dur})^2}{N \times \overline{Dur}^2}$, where N is the number of halfwaves in one epoch, Dur_i is the duration of a halfwave and $\overline{Dur}$ is the average duration of halfwaves in one epoch. This feature reflects the regularity of the duration of halfwaves in one epoch.

4. Dominant frequency: To every peak in a spectrum corresponds a peak frequency. Two other frequencies can be defined in relation to this peak: one is in the rising slope and the other is in the falling slope, and they correspond to amplitudes equal to half the amplitude of the peak. These two frequencies define a frequency band called *full width half maximum band* of the peak. Among all peaks in a spectrum, the peak which has the largest average power in its full width half maximum band is called the dominant peak. The dominant frequency is defined as the peak frequency of the dominant peak.

5. Average power in the main energy zone: The main energy zone is a frequency band that centers at the average frequency and contains 80% of the total energy in a spectrum. The *average power in the main energy zone* is used to reflect the concentration of energy in a spectrum. If the power in a spectrum concentrates in one area, the main energy zone is narrow and the average power within it is large.

   The ictal EEG has a special characteristic: the main frequency in similar seizures varies more in the high frequency zone and less in the low frequency zone. In other words, if a seizure has a dominant frequency at 20Hz, a similar seizure could have its dominant frequency at 21Hz or at 19Hz. However, if a seizure has its dominant frequency at 3Hz, a similar seizure is unlikely to have its dominant frequency at 4Hz. In order to reflect this characteristic, a logarithmic scale is used in the frequency axis, i.e., F(f)=logf. All frequency

bands are measured according to F(f) instead of frequency. In this way, a same frequency difference in the lower frequency zone has a greater weight than in the higher frequency zone because F(f) is a logarithmic mapping of frequency.

6. Location feature: This feature contains the positions of electrodes where a seizure onset occurs. Our method attempts to detect seizures with similar patterns *in the same brain regions*. This characteristic is very important in the determination of a seizure and its onset. The location feature is translated in the classification into a requirement that the seizure onset occurs in the same channels as that of the template seizure.

2.1.1.3 Rationale for Using the Features

The first three features come from the waveform decomposition method of Gotman (1976, 1982). They represent the most important characteristics of the EEG and they were used in the seizure detection system of Gotman (1982, 1990a). They have proven useful in the detection of seizures. Other features used in the seizure detection method of Gotman (1982), such as information about the EEG before or after the current epoch (the context), are not necessary because our method uses template matching and therefore only information about the current epoch should be compared with the template.

Two features, dominant frequency and average power in the main energy zone, represent important characteristics of the EEG in the frequency domain. Similar features were used in the system of Murro et al. (1991) to detect complex partial seizures.

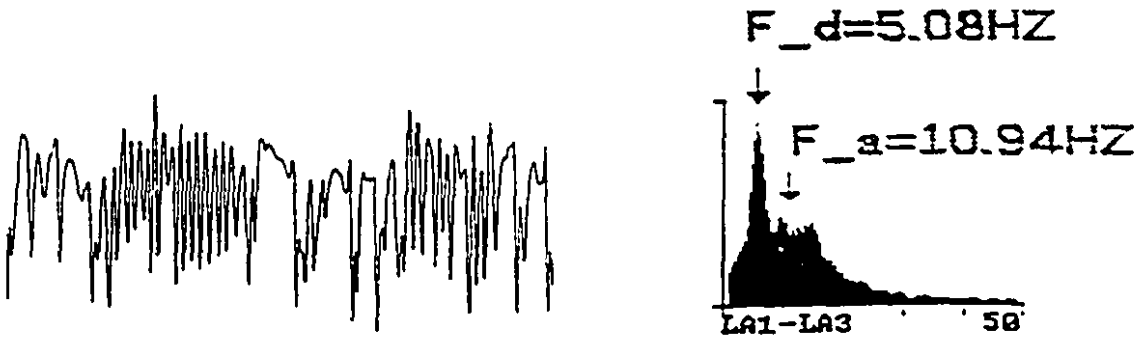
The reason for using the dominant frequency instead of the more common average frequency is that average frequency is a gross estimation of all frequency components. In ictal EEGs, especially in the prominent part of a seizure, there is usually a main frequency component because the EEG tends to be rhythmic, as defined in IFSECN

(1974). However, ictal EEG can also be accompanied by frequency components of noise, such as those of muscle activity or baseline movement. This is why the dominant frequency is useful in terms of reflecting this characteristic of ictal EEG, while the average frequency mixes ictal characteristics with those of noise. In addition, similar seizures have similar main frequency components, as shown in figure 4-1, even though some other activities, such like slow waves and the noises, can vary from one seizure to another. In this case, the dominant frequencies are the same in two similar seizures while the average frequencies are different due to the different slow activities.

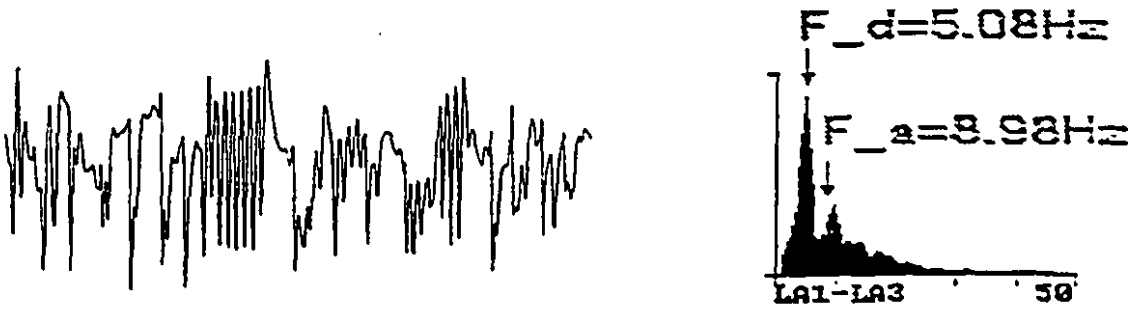
The main energy zone is defined to reflect the rhythmicity and the frequency content of the rhythmicity in an epoch of EEG: when the EEG is rhythmic, its energy concentrates in a small frequency region. This creates a small main energy zone and the average power in the main energy zone is high. The location of the main energy zone indicates the frequency content of the rhythmicity. As shown in figure 4-2a, for an epoch of ictal EEG, the spectrum shows a narrow main energy zone because of rhythmic activity. For another epoch of EEG from a similar seizure of the same patient, as shown in figure 4-2b, in the same frequency band as the main energy zone of the template, the average power in the band are very similar to that of figure 4-2a. Figure 4-2c is an example of an epoch of interictal EEG of the same patient from the same electrode. Its spectrum shows a very different pattern in terms of the average power in the same zone as the main energy zone of the template. Figure 4-2d shows an example of interictal rhythmic activity. Although its main energy zone is also small and the average power within is large, it can be distinguished from figure 4-2a and figure 4-2b because the location of the main energy zone is different. In other words, in the same frequency band as the main energy zone of the template, the average power in this band in figure 4-2d is much smaller than that in the template. Therefore, patterns in figure 4-2d and figure 4-2a are distinguishable even though both of them are rhythmic. As a result, the average

power in a main energy zone is a useful feature to distinguish the rhythmic activity in a seizure from both interictal non-rhythmic activity and interictal rhythmic activity.

The last feature in our method is the location feature. This feature imposes spatial restraints in our detection method. Spatial information has been used in other EEG pattern detection systems (Gotman 1982; Glover et al. 1989). It has been proven that the use of both temporal and spatial information can provide better results in the detection of EEG patterns than the use of temporal information alone. In our false seizure reduction method, the spatial information was also used to make detection more reliable. Unlike the other five features, which are characteristics of EEG morphology, electrode positions are the characteristic of physical location and they are difficult to express in the detection space, especially in the case of intracerebral electrodes. Although distances between electrodes were a dimension of the detection space in the reduction of false seizure detections, this is not applicable in the onset detection because similar false seizure patterns can occur in electrodes close to the ones in the prototype patterns while similar seizure onsets always occur in the same electrodes as in the template. For instance, artefact, which are the main cause of false detections (Pauri et al. 1992), do not necessarily occur in exactly the same electrodes.



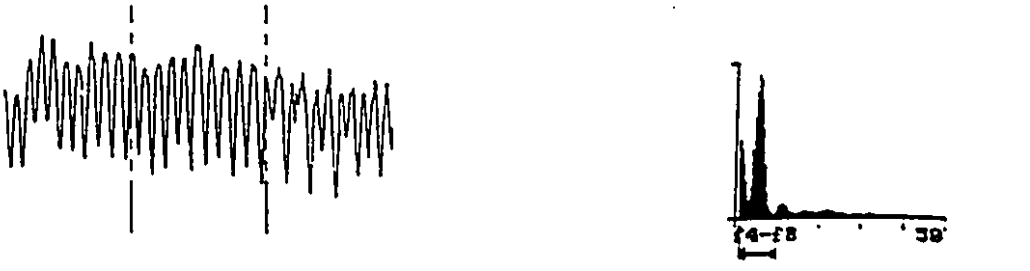
Pattern A



Pattern B

Figure 4-1: Dominant frequency and average frequency: Pattern A and pattern B are EEG and their spectra of two similar seizures. Although these two seizures are similar, seizure A has more higher frequency components and therefore its average frequency (10.94 Hz) is different from the one in seizure B (8.98 Hz). However, the dominant frequencies of these two patterns remain the same (5.08 Hz). This illustrates how the dominant frequency is a better feature than the average frequency in terms of representing major frequency components of the seizure.

Pattern A



Pattern B

Pattern C

Pattern D

Figure 4-2: (see next page for captions)

Figure 4-2: Examples of using average power in a main energy zone as a feature: Pattern A is a seizure EEG and a spectrum of one epoch. In the figure, MEZ stands for main energy zone. The rhythmicity of EEG is represented in the spectrum as an area of concentrated energy, i.e., a major peak. The main energy zone in this case is small (1Hz to 10 Hz) and the average power in it is high. Pattern B is a seizure similar to pattern A from a same patient. The two spectra show great similarity in terms of main energy zones and average power within them. Pattern C is a section of interictal EEG and its spectrum. Since there is not much rhythmicity in the EEG, the average power within the same main energy zone as the one in the template (pattern A) is low. For an interictal rhythmic activity (pattern D), the average power in its own main energy zone may be similar to the one in the template. However, since its main energy zone is different from that of the template, the average power in the same energy zone as the one in the template is lower than that in the template. This figure shows that average power within a main energy zone is a good feature to reflect rhythmicity and frequency of the seizure. As a result, it can be used as one of features to distinguish ictal EEG from interictal rhythmic and non-rhythmic EEG.

2.1.2 Template Selection

Template selection has two steps: selection of template EEG and the generation of template points. A seizure can last as much as several minutes, or as little as a few seconds. In this section, we describe how a part of seizure is chosen as a template and how a template is represented in the detection space.

2.1.2.1 Selection of Template EEG

A template is a portion of a selected seizure. A template should contain the early part of the seizure so that the detection can be made as early as possible. Also, a template must have a reasonable length to increase the probability of detecting seizures in case the very early part of the seizure cannot be detected. Selecting the template also includes the selection of the location feature: a set of channels should be selected from the template. The template selection criteria are the following.

1. A template EEG starts at seizure onset.

2. A template EEG ends 20 seconds after the onset or at the end of a seizure, whichever occurs first.

3. The set of channels in which the seizure onset occurs is selected as the location feature.

Template selection must be done visually by an EEGer. The reasons for choosing 20 seconds as the length of a template are: (1) Most seizures last more than 20 seconds. (2) Twenty seconds is normally long enough to catch the predominant patterns of a seizure. (3) A template shorter than 20 seconds may decrease the chance of detecting onsets, and a detection 20 seconds after the onset is not considered too late. (4) Although a detection 20 seconds later than onset is still considered better than no detection at all, a

longer template creates more template points, as explained below. It therefore increases the probability of causing false alarms, as well as increasing the computation burden.

2.1.2.2 Generation of Template Points

A template EEG is divided into epochs of 2.56 seconds. Every epoch generates a template point in the detection space after the five features are extracted from this epoch. As a result, a template EEG is represented in the detection space as a set of template points, one for each epoch. We will discuss the question of how to divide the template into epochs.

Since our method compares EEG patterns epoch by epoch, the division into epochs of the template and of a new EEG in which we are looking for a match to the template affects the probability of finding a match. For instance, as shown in figure 4-3a, pattern A is a template seizure and is divided into epochs from the beginning of the seizure. Pattern B is a new seizure which is randomly divided into epochs since the beginning of the seizure is unknown at the time of the search. In this case, the two seizures do not have a good time alignment and therefore a match may not be found or may be found late after the onset, although these two seizures are similar. If pattern B is divided into epochs starting at the beginning of the seizure, a match can be made correctly. As a result, the time alignment between a template seizure and a new seizure is important.

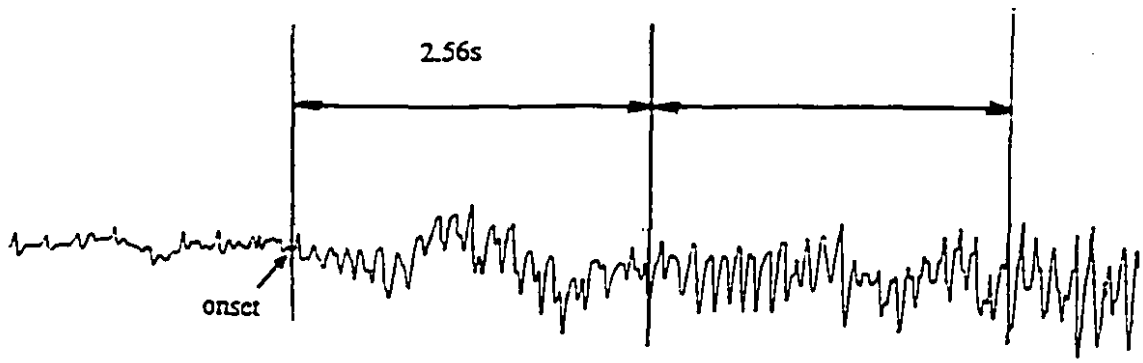
The first idea to solve the time alignment problem could be to divide a new EEG into epochs starting at every sample (figure 4-3b). The advantage is that there are always some epochs with the best possible time alignment with template epochs because at least one of the epochs starts from the beginning of the new seizure. However, there is a major disadvantage: the computational burden is too high, because every epoch in a new recording has to be processed on-line and some of features have to be extracted from the frequency domain. Doing the FFT in each epoch starting from each sample is too time

consuming because this is a multichannel recording and the computation power of personal computers is too small: there is a sample every 5 milliseconds in each channel. Therefore, this method is not suitable to solve the time alignment problem.

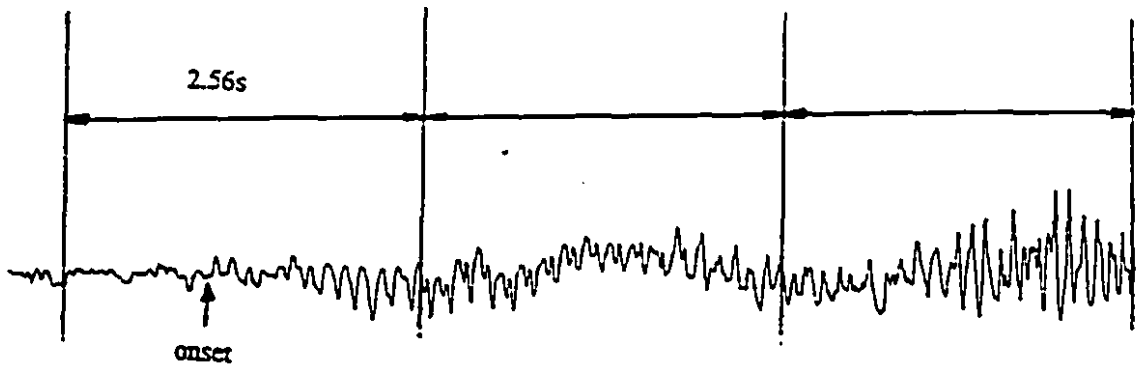
The other way to solve is problem is to divide the template seizure into as many epochs as possible so that, independently of the kind of division in a new EEG, there is always one epoch in the template having a good time alignment with epochs in a new seizure. This can be done by dividing the template seizure into epochs starting from every sample, as shown in figure 4-3c. There are two advantages in doing so: (1) The best time alignment can be obtained since this division has the same time alignment resolution as the previous method, which is also the best possible time alignment. (2) The feature extraction of template epochs is done off-line and only once, and the heavy computational load resulting from many epochs is not a problem. Nevertheless, there are two disadvantages: (1) There is too much overlap between epochs starting at every sample in the template. EEG patterns, even ictal patterns, do not change rapidly enough to alter significantly the statistical properties of epochs lasting 2.56 seconds and starting at 5 milliseconds intervals. (2) This method creates many points in the detection space (4,000 points for a 20-second template) and every new epoch of EEG being analyzed has to be compared to all template points in the detection space. This results in an important computational burden for on-line processing.

The division of template seizure into epochs is therefore done by compromising computational load for on-line processing and the precision of time alignment. We consider that the template seizure can be divided into epochs starting every 320 milliseconds (figure 4-3d); 320 milliseconds is a convenient value because it corresponds to 64 samples, which is the basic processing unit in our programs. There are two reasons for the choice of 320 milliseconds: (1) For a template seizure and a new seizure, a mismatch is unlikely to occur when two epochs of seizure EEG have a time alignment difference at 320 milliseconds. This is because changes in the EEG within 320

milliseconds rarely changes dramatically the statistical characteristics of a 2.56-second epoch of EEG. (2) In comparison with the division of epochs starting every sample, this division reduces the number of template points by a factor of 64, thus reducing by the same factor the on-line processing time for classification. This helps in accommodating the limited computational power of personal computers



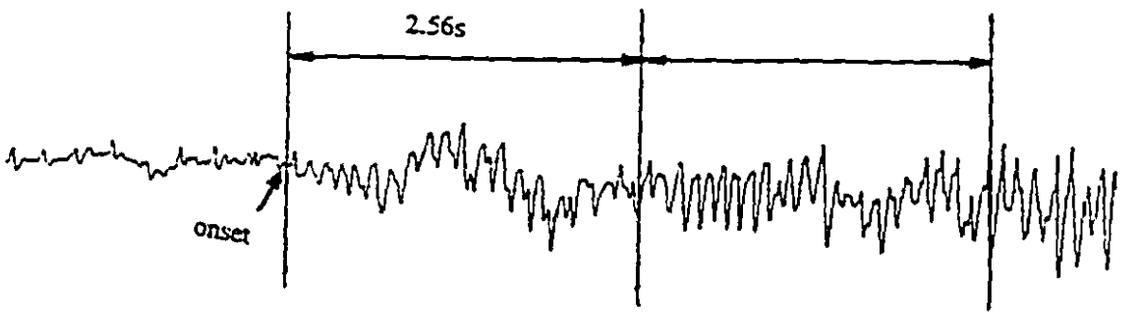
Pattern A (Template)



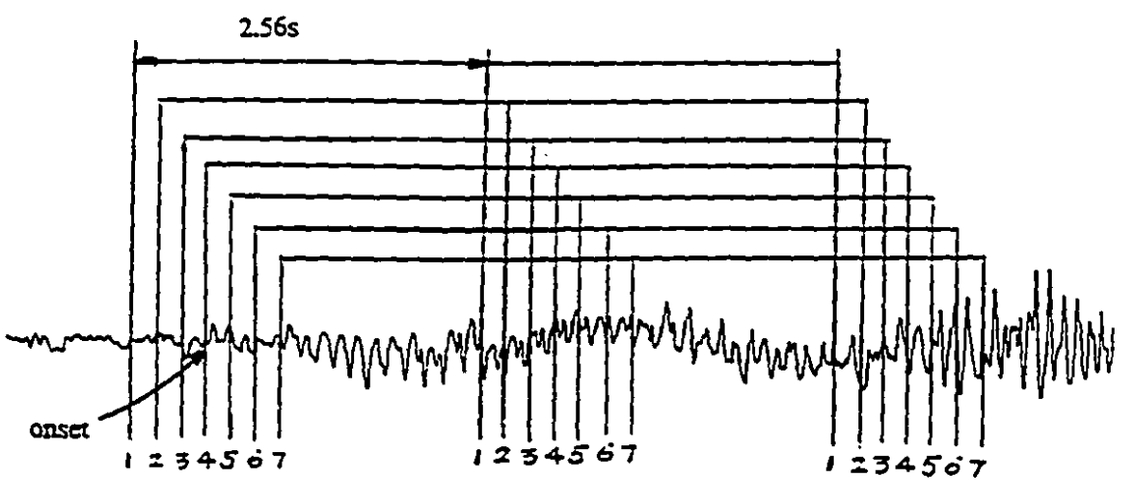
Pattern B (New Seizure)

Figure 4-3A

Figure 4-3: Epoch division and time alignment: Pattern A is a template and pattern B is a new coming seizure. In figure 4-3a, the template and the new seizure are divided in adjacent epochs. In this case, a time alignment problem arises and a mismatch is likely to occur. (to be continued in the next page)



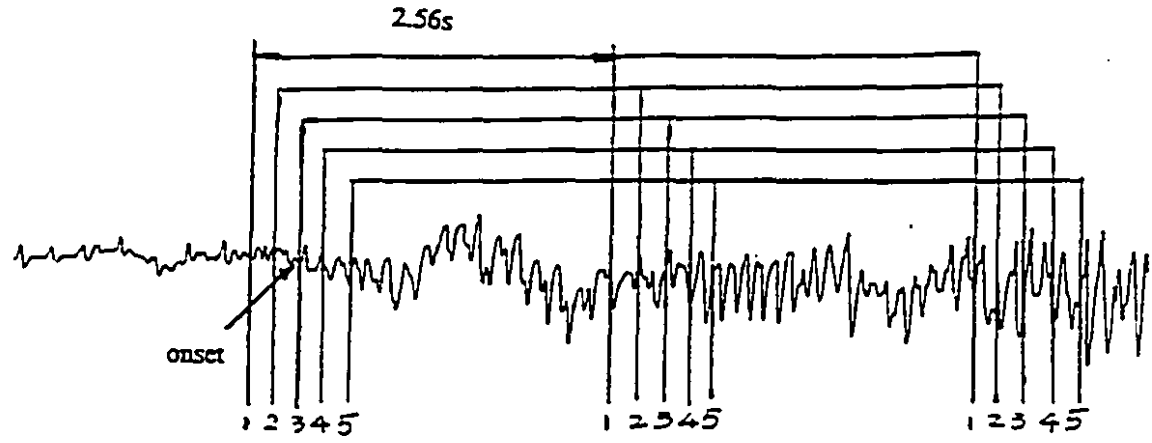
Pattern A (Template)



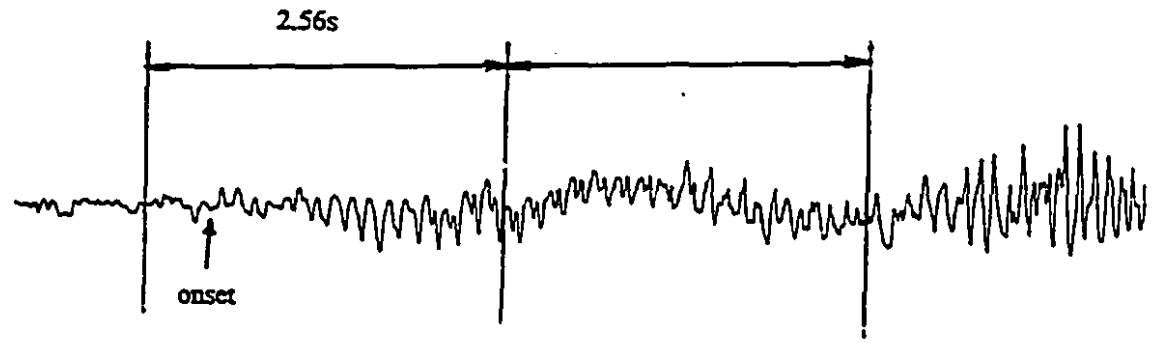
Pattern B (New Seizure)

Figure 4-3B

Figure4-3: (continue from the last page) In figure 4-3b, the *new seizure* is divided into a large number of overlapping epochs and therefore the time alignment problem can be solved. (to be continued to the next page)



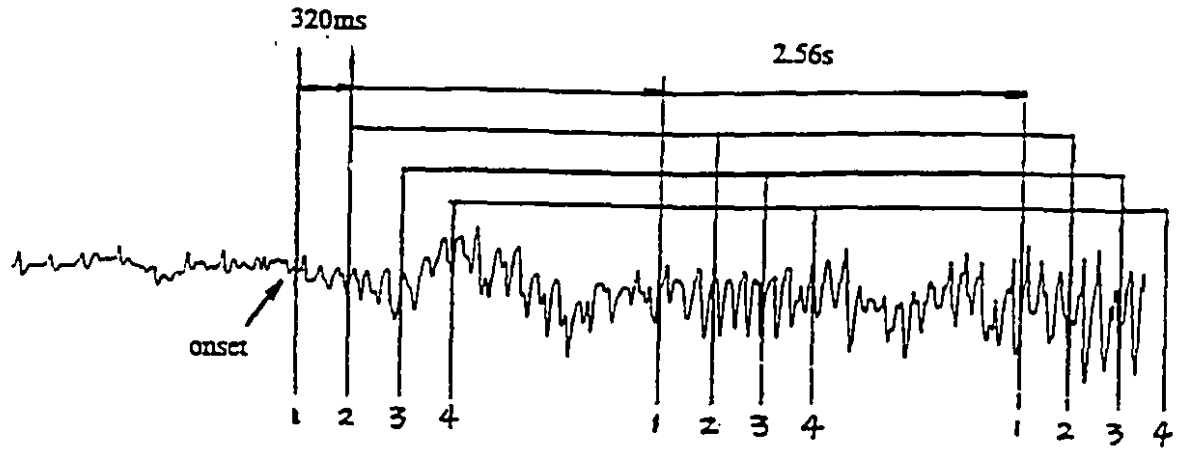
Pattern A (Template)



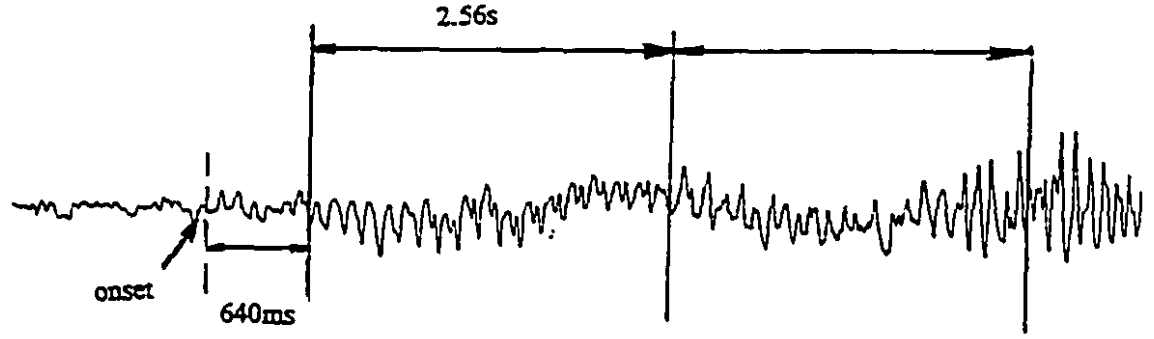
Pattern B (New Seizure)

Figure 4-3C

Figure 4-3: (continued from the last page) Figure 4-3c shows another way to solve the time alignment problem by dividing the *template* into a large number of epochs. Both solutions required a large amount of computation. (to be continued to the next page)



Pattern A (Template)



Pattern B (New Seizure)

Figure 4-3D

Figure 4-3: (continued from the last page) One way to solve the time alignment problem without increasing too much the computational burden is to divide the template into epochs starting every 64 samples, or 320 milliseconds, as shown in figure 4-3d. In this case, even if the division of epochs in the new seizure does not start at the onset, a good match still occurs between the first epoch of the new seizure and the first epoch in the division number 3 in the template because both epochs start 640 milliseconds after onset.

2.1.3 Distance Measure

We have to define the distance between two points of the detection space so that we can measure how close the features of a new epoch are from those of the template epochs. A weighted distance is used to reflect the different effects of each feature in a detection. The distance between two points, $\vec{A}$ and $\vec{B}$, in the detection space is defined as: $D(\vec{A},\vec{B}) = \left\|(\vec{A}-\vec{B})\vec{M}^{-1}(\vec{A}-\vec{B})\right\|$, where

$$\vec{M}^{-1} = \begin{pmatrix} M_1 & 0 & 0 & 0 & 0 \\ 0 & M_2 & 0 & 0 & 0 \\ 0 & 0 & M_3 & 0 & 0 \\ 0 & 0 & 0 & M_4 & 0 \\ 0 & 0 & 0 & 0 & M_5 \end{pmatrix}$$ and M_i is the weight for each dimension.

The difference between the Euclidean distance and this distance is that each feature is not considered equally weighted in this measure. This is because some features are more important in the determination of onsets than others. The weight in each feature also acts as a normalization factor to convert different physical units in different dimensions into a universal unit so that comparisons of distances can be performed. In this method, all M_i are set empirically.

Although the distance measure looks similar to the Mahalanobis distance (Devijver and Kittle 1982), there are some differences. The first one is that the distance measure is used to measure distance between two points in a detection space, while the Mahalanobis distance is used to reflect the distance between two classes by using the mean of each class. The second difference is that the matrix $\vec{M}_i$ in our distance measure is the weight of each dimension and it is a covariance matrix in the Mahalanobis distance.

2.1.4 Classifier Design and Training

2.1.4.1 The Classifier

For a new EEG pattern, designated as $\vec{N}_j$, a set of template points, $\vec{P}_i$, and a set of interictal EEG points, a patient-specific classifier can be expressed as:

$$\begin{cases} \vec{N}_j \text{ is a possible seizure onset} & \text{if } D(\vec{P}_i, \vec{N}_j) < T \times D_{NN_i} \text{ for all } i \\ \vec{N}_j \text{ is NOT a seizure onset} & \text{if } D(\vec{P}_i, \vec{N}_j) \geq T \times D_{NN_i} \text{ for all } i \end{cases}$$

where $D(\vec{P}_i, \vec{N}_j)$ is the distance defined above, T is a distance threshold coefficient for all template point $\vec{P}_i$, D_{NN_i} is the distance between a template point $\vec{P}_i$ and its nearest interictal EEG point.

If $\vec{N}_j$ is a candidate seizure onset, we must still check whether it occurs in a channel selected as part of the location feature (the location feature was discussed above; it represents the channels involved in seizure onset). In any patient, two seizures are never absolutely identical. The variability between seizure onsets in some channels could be such that a mismatch could occur, even though the seizure onsets are quite similar. We decided therefore to require a match in at least half of the channels in the location feature. As a result, in each epoch of EEG, if 50% of channels selected in the location feature have matches with the template at the same time, it is considered that a seizure onset is detected.

The distance threshold coefficient T, when it is smaller or equal to 0.5, is used to bias the system to have a small chance of causing false alarms, at the expense of detecting fewer seizure onsets or detecting them later. This is because the classification boundary tends to be closer to template points when T is smaller than 0.5 (figure 4-4).

When T is larger than 0.5, the opposite results may occur because the classification boundary gets closer to interictal EEG points.

This classifier can be considered a modified nearest-neighbor (NN) classifier. As in the NN rule, each template point $\vec{P}_i$ has a distance to its nearest interictal EEG point. Some $\vec{P}_i$ are more similar to some interictal EEG points than other $\vec{P}_i$ and therefore have a small distance to their nearest neighbors. We use smaller distance thresholds for these $\vec{P}_i$. This allows the classifier to adjust itself to accommodate the situation of each template point and thus be able to detect seizure onsets more accurately with less probability of causing false alarms. The modification to the NN rule in our method is that the distance between a template point and the classification boundary around it is a constant in all directions (figure 4-4). In other words, the classification boundary in our method is a five-dimensional ball centered at a template point and with a radius of $T \times D_{NN_i}$. This makes the classification boundary of the method closer to the template point even when T is 0.5 than that of NN rule (figure 4-4). This reflects the bias we want to place in our method. Another reason for the modification is that there are much fewer template points than interictal EEG points, the difference being of the order of several hundred times. Therefore, it is better to be conservative by shrinking the decision boundary toward template points. This results in having a lower probability of misclassifying interictal EEGs as early seizure EEGs, at the expense of having a higher probability of missing seizures. The last reason for the modification is that the computation cost of our method is much less than that in the NN rule. For M template points and N interictal EEG points, a NN classifier needs M+N distance computations and M+N comparisons to classify a new point, while our method only needs M distance computations and M comparisons. Since N is of the order of several hundred times M, the reduction in computation in our method is considerable.

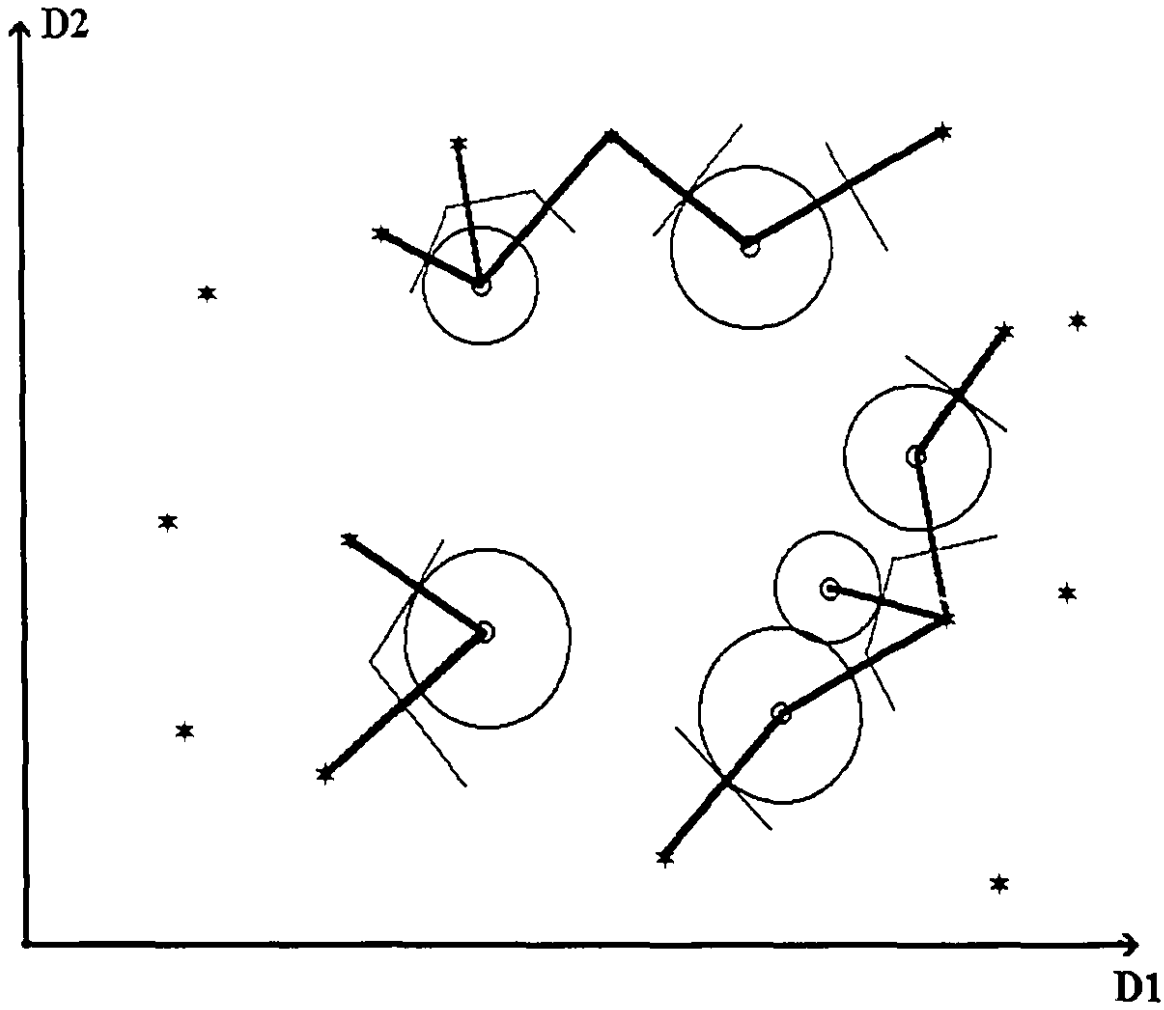


Figure 4-4: Discrimination hyperplane of the NN rule and that of the current method: "*" represents interictal EEG points and "o" represents template points. D1 and D2 are two of the dimensions used in the detection space. The bold straight lines which connect template points and interictal EEG points indicate the distance between these points. The thin straight lines mark the hyperplane of the NN rule. The circles show the classification boundary of our method with a distance threshold coefficient of 0.5. It is obvious that the classification boundary of our method is always closer to template points than that of the NN rule when T' is equal to 0.5.

2.1.4.2 Training the Classifier

As indicated in figure 4-5, a classifier is trained with seizure template and interical EEG. The classifier should be trained with appropriate amount of data so that it can achieve a high seizure onset detection rate and a low false alarm rate. It is always the case that the more training data there are, the better a classifier can be trained. In our method, the more seizures are used for training, the more accurately a classifier can detect onsets; the more interictal EEG is used for training, the lower the false alarm rate will be. However, more ictal data as training data could result in more false alarms, while more interictal data as training data could result in a lower probability of detecting onsets. Therefore, we carried out experiments to determine the most appropriate amount of training data for our classifier, in particular the balance between the number of seizures and the amount of interictal data. One important limitation is the availability of seizures, a factor which must also be considered in the determination of the amount of training data. Experiments and results are explained in the "Results" section.

Our method is based on patient-specific information by using patient-specific ictal and interictal data as training data. In order to assess whether our method is optimal, we designed two additional experiments. One is to use semi-patient-specific information, which includes patient-specific ictal information and non-patient-specific interictal information, to train classifiers. The other one is to use both non-patient-specific ictal and interictal data to train classifiers. Experiments and results are explained in the "Results" section.

2.1.4.3 Relabeling Template Points

It can happen that seizure patterns, when divided into epochs, are not distinguishable from some interictal EEG patterns of the same patient. This is because

the identification of a seizure by an EEGer takes into account the evolution of the EEG. In other words, every epoch of the seizure may be considered as part of a seizure because of the pattern in neighboring epochs, because of the evolution of the seizure pattern. This is not the case in our method because each epoch is considered independently from neighboring epochs. Therefore, some seizure patterns could cause false alarms because similar EEG patterns can be found in the interictal EEG. In order to solve this problem, we relabeled some template points as interictal EEG points during classifier training. We set a relabeling threshold empirically for all classifiers. If a template point has a distance to its nearest interictal EEG point smaller than the relabeling threshold, this template point is relabeled as an interictal EEG point. This will, of course, delay the possible onset detection due to the elimination of template points. However, trading earlier onset detection for a lower false alarm rate is justified because a low false alarm rate is a high priority.

Theoretically, it is possible but very unlikely that all template points are relabeled if all EEG patterns in the template can be found in the interictal EEG of the same patient. This means that the seizure onset patterns are not distinguishable from the interictal EEG patterns in the patient and therefore the computer can either detect it at the expense of many false alarms or is not able to detect it at all.

We did not relabel interictal EEG points that are very close to template points as template points because this would increase the false alarm rate.

2.1.5 The Quality Measure of the Classifier

The aims of this project are to detect as many seizures as possible, as early as possible and with as few false alarms as possible. In order to compare different classifiers, a measure has to be defined to reflect the overall quality of the classifier. The quality of a classifier is determined by three factors: detection rate, detection delay and

false alarm rate. A measure indicating the quality of a classifier should reflect these three factors appropriately. We called "quality value", or "QV", the measure of quality. It is defined as: $QV = \frac{C}{(R_{fa} + 0.2) \times (T_{dly} \times P_{det} + 60 \times P_{msd})}$, where C is a constant to scale the value of QV, R_{fa} is the number of false alarms per hour, T_{dly} is the average delay of onset detections in seconds, P_{det} is the percentage of seizures detected and P_{msd} is the percentage of seizures missed. There are several reasons for using this formula.

(1) QV monotonically increases when R_{fa} decreases. It means the lower the false alarm rate is, the better the classifier is.

(2) $T_{dly} \times P_{det} + 60 \times P_{msd}$ is actually the weighted average of the delay of onset detections. While T_{dly} is the average delay of onset detections, 60 seconds is used as a delay for seizures missed by the method. We decided that a detection 60 seconds after the onset is equivalent to missing the seizure, in terms of the benefits of seizure onset detection. The weights for the average delay of detected or missed seizures are the percentages of detected and missed seizures. Consequently, the weighted average delay reflects the quality of a classifier with respect to detection delay.

(3) As a result of (2), QV is inversely proportional to the weighted average of the delay of onset detections. This reflects the fact that the shorter the onset detection delay, the better the classifier.

(4) The constant "0.2 false-alarm/hour" is added to R_{fa} in the formula of QV. In our method, the false alarm rate is low and it usually ranges from 0 to 0.5. A false alarm rate higher than 0.5 false-alarm/hour is considered unacceptable. Without the constant, a change in R_{fa} from 0.01 to 0.02 results in the reduction of QV by half, while a change from 0.2 to 0.3 only reduces QV by 33%, although in the latter case the false alarm rate increases much more significantly than the previous one. This constant is therefore necessary to make sure that the change of R_{fa} has an appropriate effect on the QV.

(5) The constant C is empirically set to 10 because this scales the value of QV to an easy reading range. The constant does not change the meaning of the results, but only affects the readability of the results.

(6) We carefully considered the range of each factor, how their changes affect the quality of classifiers and the trade-off between factors. We used two constants, 0.2 and 60, to scale the effect of each factor in the QV appropriately. As a result, the risk that extreme values of one of the factors results in an aberrant QV has been greatly reduced, if not eliminated.Inconclusion, QV is justified in terms of the combination of all the aspects of our classifier. The higher the value of QV, the better the classifier. Among different classifiers, the classifier with the highest QV should be the best.

2.2 Implementation

The implementation is simple if the number of human interventions is limited. In our implementation, only one human intervention is required: selection of the template EEG and location feature. This human intervention is indispensable because only a human operator can accurately provide this vital information. The details of the implementation procedures are described below, and also illustrated in figure 4-5.

1. When a seizure occurs in a patient during a long-term EEG monitoring session, a section of this seizure will be stored as the seizure onset template of this patient according to the criteria mentioned earlier. The template points are generated automatically by extracting features from epochs of the template.

2. The template EEG should be reviewed to determine the location feature, i.e. the channels in which the seizure onset takes place.

3. A set of interictal EEG training data has to be collected. This is done automatically, according to criteria discussed in the section on data collection.

4. The template EEG and the interictal EEG training data are used to train the classifier

5. The patient-specific classifier is then used to detect seizure onsets during long-term EEG monitoring. Each epoch of new EEG should pass through the classifier. If a seizure onset is detected, a warning signal will be given so that observers can take appropriate action.

More details about the implementation can be seen in Appendix B in the form of a block diagram.

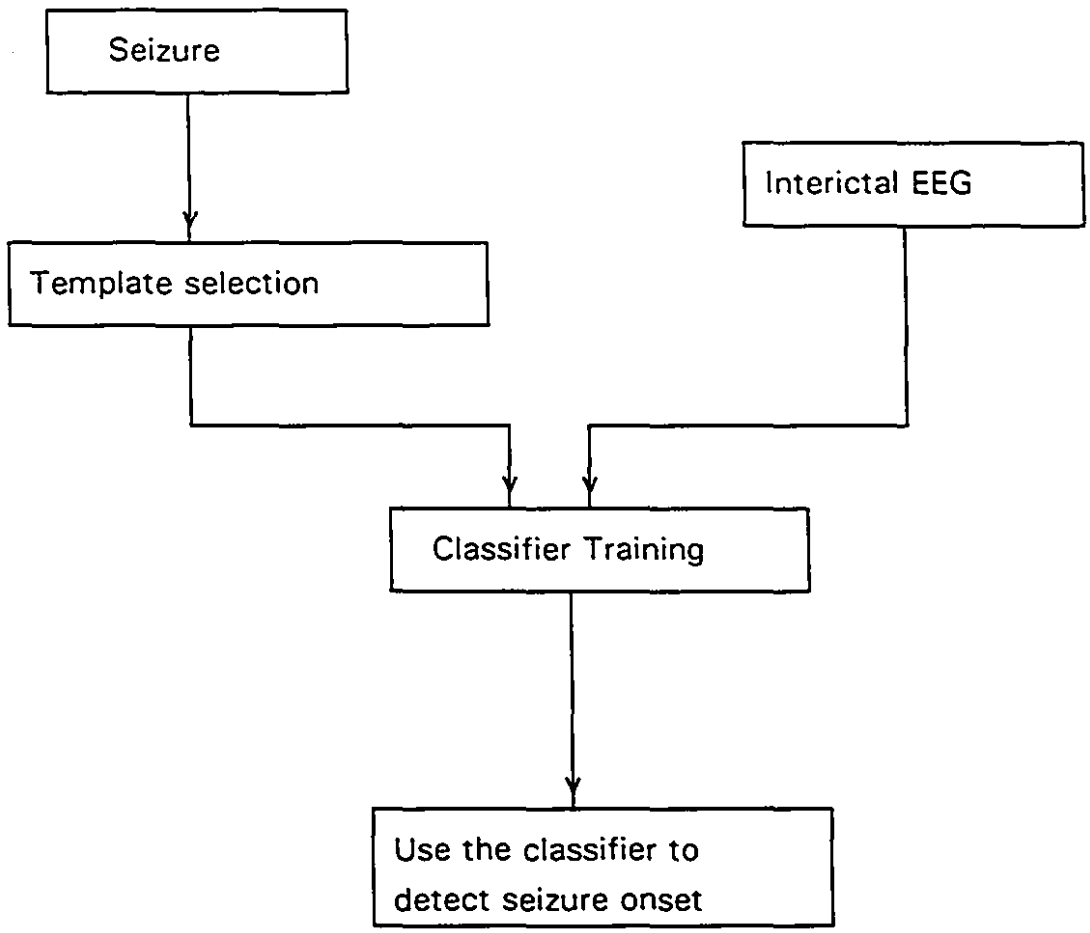


Figure 4-5: Implementation procedure of our seizure onset detection method: After a seizure is captured, a human intervention has to be made to select the template, as well as the location feature. A set of interictal EEG should be collected at the same time. Both sets of data are used to train a patient-specific classifier. This classifier is then used to detect seizure onsets in subsequent monitoring sessions.

2.3 Subjects

2.3.1 Criteria in Subject Selection

Subjects for this study were selected from patients undergoing long-term EEG monitoring at the Montreal Neurological Hospital from January 1993 to August 1993. A total of 24 types of seizures coming from 17 patients was selected. All these patients satisfy the following conditions:

1 Each patient has at least one type of seizure. Some patients may have up to three types.

2. For each type of seizure there are three to seven similar seizures from the same patient.

We need at least three seizures to test the consistency of the method. We limited the total number of seizures per type to seven to avoid giving too much weight to that type in the overall average result

2.3.2 Data Collection

In order to train and test classifiers, interictal EEGs, as well as ictal EEGs, have to be collected. We selected the holdout method for error estimation (Devijver and Kittle 1982), in which the data are separated into two mutually exclusive sets, so that one of them is used to train classifiers and the other is used for testing. The reason for using the holdout method is that it is not very difficult to acquire a large size of interictal EEGs in long-term EEG monitoring. The holdout method for estimating error rates of classifiers has the advantage of having less variance in the estimation, but the disadvantage of

requiring a large data set (Devijver and Kittle 1982). Although there are not many seizures, we still use the holdout method for the estimation of onset detection rate for two reasons: (1) This method is better than the resubstitution method, especially when the NN-rule-like classifiers are used (Devijver and Kittle 1982). (2) Although the rotation method may look suitable for this situation, the associated computational load and complexity made it impractical for our purpose.

2.3.2.1 Training Data

Only one seizure, which serves as a template, and a set of interictal EEGs are used as training data for each patient. Interictal EEGs are easy to obtain. The interictal EEG training data should represent as many as possible of the interictal EEG patterns present in a patient. The more interictal EEG data are used for the training, the less chance false alarms will occur. Practically, a set of interictal EEGs sampled evenly over a period of 24 hours is suitable. There are three reasons for this: (1) The main factors affecting the EEG, sleep and level of activity, have a period of 24 hours. (2) Although a data set with the complete 24-hour EEG has all the interictal EEG patterns, a sample rate of one minute every twenty to thirty minutes is practically suitable. Of course, the longer sample results in a better interictal EEG training data set, but also results in greater computation and storage cost. (3) According to the hospital records at Montreal Neurological Hospital from January 1 1993 to December 31 1993, as shown in table 4-1, there is on average about one seizure per day per patient for both depth electrode patients and scalp electrode patients. This is the *a priori* knowledge of seizure frequency. As a result, using one seizure and a set of interictal EEG from 24 hours to train a patient-specific classifier appears justified. This seizure frequency can also be interpreted as two seizures per 48 hours. This means that using two seizures and interictal EEG from 48 hours to train a classifier is also justified.

Table 4-1: Statistics of patients' records from Montreal Neurological Hospital

Time period	January 1 1993 to December 31 1993		
Electrodes	Depth	Scalp	Total
Total number of patients	17	135	152
Total number of monitoring days	344	1036	1380
Total number of seizures	301	1032	1333
Average number of seizures per monitoring day	0.875	0.996	0.966

2.3.2.2 Testing Data

Two types of EEGs were collected as testing data: interictal EEGs and seizures similar to template seizures. Interictal EEGs were collected to test false alarm rates, while seizures were used to determine the seizure onset detection rates as well as delays in onset detection. Since the false alarm rate is very low in this method, a lot of interictal EEGs are needed so that the variance of the false alarm rate estimates can be small. This is also the reason we are using the holdout method, since it gives a smaller variance than those determined by the leave-one out method and the rotation method (Devijver and Kittle 1982).

3 Results

EEGs were recorded according to the clinical protocol of the Montreal Neurological Hospital. They included 32 or 64 electrodes and therefore there are 32 or 64 channels in each referential montage. Seizure detection, however, was performed with a bipolar montage. All patients with scalp electrodes used 32-channel bipolar montages and patients with depth electrodes used 16-channel bipolar montages. In the 12 patients with scalp electrodes, the average number of seizures per type is 3.9. In the 12 seizure types from patients with depth electrodes, there are 4.5 seizures per type. The average length of interictal EEGs used for training is 0.7 hours and 2.7 hours for testing. All training and testing data required approximately 4 gigabytes of memory.

The results show that this method can detect seizure onsets accurately and quickly. It can also detect some seizures missed by the classic seizure detection method. This proves that it is possible to have a reliable patient-specific on-line seizure onset detection system.

We first compare different kinds of classifiers with different parameters. The results for the best classifier for all patients are presented at the end.

3.1 Comparisons Among Different Classifiers

The performance of classifiers can be affected by the distance threshold coefficient and the training data. In order to find out the optimal classifier, we designed three groups of experiments to compare results: (1) classifiers with different distance threshold coefficients (2) classifiers with different amounts of training data (3) classifiers with different proportions of patient-specific and non patient-specific information.

3.1.1 Classifiers with Different Distance Threshold Coefficients

There is one major parameter to be adjusted in the design of a classifier in this method: the distance threshold coefficient T. As mentioned before, the distance threshold coefficient determines if the classification boundary is closer to template points or closer to interictal EEG points. When the classification boundary is closer to template points, there will be fewer false alarms, as well as fewer onset detections. In order to find the optimal value of the distance threshold coefficient, we tried classifiers with five distance threshold coefficients and compared the QV of these five classifiers. The implementation procedures of the experiments are the same as the one illustrated in figure 4-5 by using a seizure as the template and a set of interictal EEG of the same patient sampled from 24 hours of monitoring. The only difference among classifiers in this group of experiments is that their distance threshold coefficients are different. The results are shown in table 4-2. Also, figure 4-6 presents the change of QV according to distance threshold coefficients in the case of scalp electrode patients, depth electrode patients and all the patients.

Theoretically, we should try many distance thresholds with a small increment so that the optimal threshold can be determined precisely. However, because computation and memory required in the experiment are huge, it takes about one month of a personal computer's time to compute results from each distance threshold coefficient. Therefore, we can only try five distance threshold coefficients.

From table 4-2 and figure 4-6, it is noticeable that the classifier with the distance threshold coefficient of 0.5 is the best one because the QV values are always the highest.

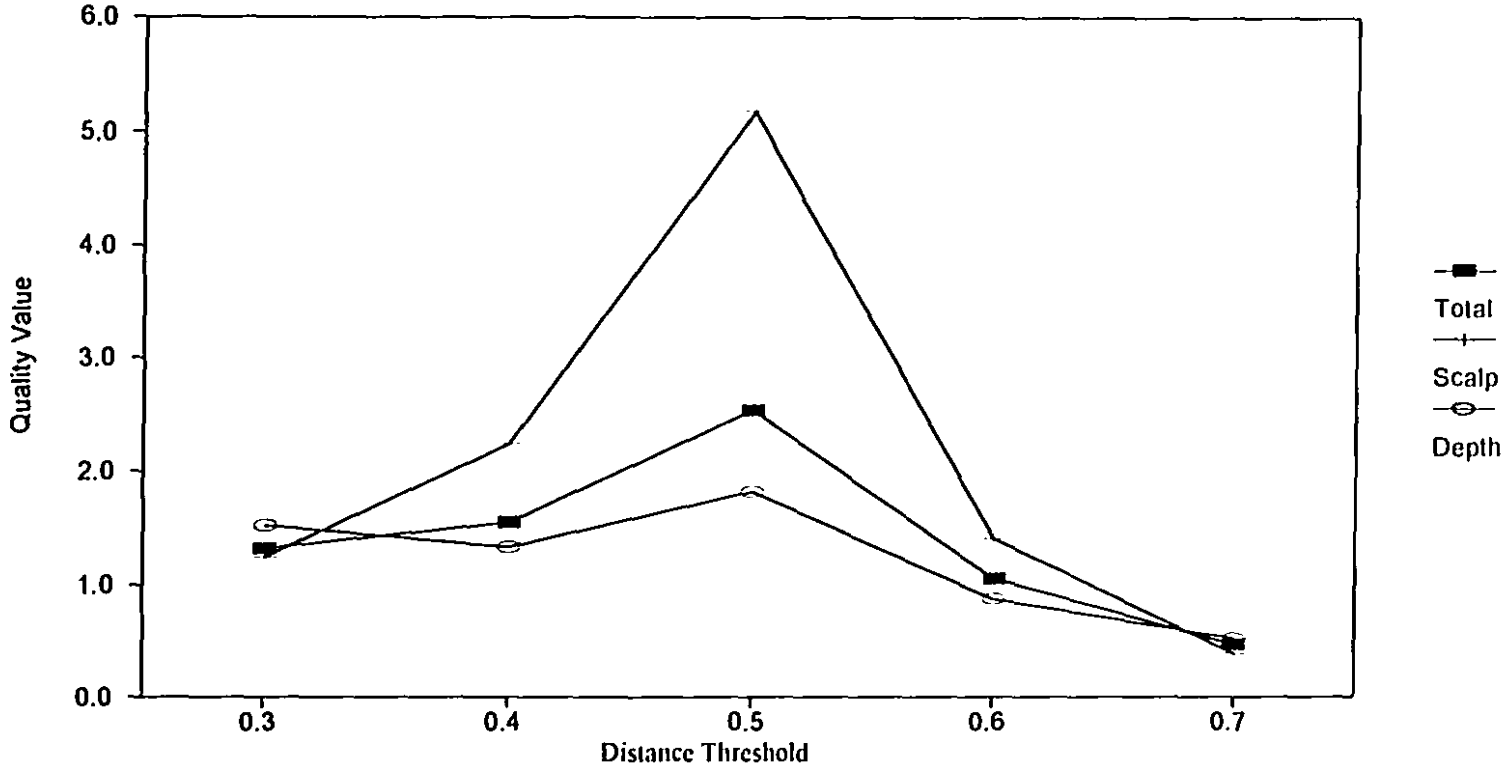


Figure 4-6: Quality values of classifiers with different distance threshold coefficient: It can be seen that classifiers with a distance threshold coefficient of 0.5 have the highest QV. Results are the same for classifiers designed for patients with scalp electrodes and for patients with depth electrodes.

Table 4-2 Results of classifiers with different distance threshold coefficients.

Distance threshold coefficient		0.3	0.4	0.5	0.6	0.7
Seizure detection rate (%)	Scalp	37	74	100	100	100
	Depth	76	90	100	100	100
	Total	58	83	100	100	100
Delay (s)	Scalp	6.1	9.1	9.5	9.2	8.8
	Depth	10.8	10.7	9.6	9.3	8.3
	Total	8.7	9.9	9.6	9.2	8.5
False alarm rate (/h)	Scalp	0	0	0.03	0.57	2.7
	Depth	0.09	0.28	0.37	1.02	2.1
	Total	0.05	0.15	0.21	0.82	2.3
QV	Scalp	1.25	2.24	4.58	1.41	0.39
	Depth	1.53	1.33	1.83	0.88	0.52
	Total	1.32	1.55	2.54	1.07	0.47

3.1.2 Classifiers with Different Amounts of Training Data

The amounts of training data will certainly affect the quality of a classifier. Theoretically, the more training data there are, the better a classifier can be trained. In our method, a modified NN rule is used for the classification. As shown in table 4-1, the probability of seizure is about one seizure per day per patient. As a result, in our experiment to compare classifiers with different training data sets, we maintained the right proportion of seizure points in the detection space by selecting two sets of training data: one is the first available seizure and interictal EEG sampled from 24 hours (set 1)

and the other is the first two available seizures and interictal EEG sampled from 48 hours (set 2). The distance threshold coefficient used in this experiment is 0.5 because the evidence seen in the last section suggests that this is the best one. The implementation procedure of this group of experiments is similar to the one in figure 4-5. The difference only exists in the amount of seizures selected as templates (one seizure or two) and the amount of interictal EEGs (EEGs sampled from 24 hours or 48 hours) used to train classifiers. These experiments were carried out in patients with depth electrodes because only this set of patients has enough interictal EEGs for both training and testing. The results are shown in table 4-3.

Table 4-3: Results of classifiers trained with different amount of data.

Training data set	One seizure and EEG from 24 hours	Two seizures and EEG from 48 hours
Detection rate (%)	100	100
Weighted delay (s)	9.6	9.2
False alarm rate (/h)	0.37	0.20
Quality Value (QV)	1.83	2.73

As shown in table 4-3, the classifier trained with two seizures and interictal EEG sampled from 48 hours performs better. This verifies the fact that increasing the amount of training data results in a better classifier. In this case, more template seizures result in a shorter detection delay; the more interictal EEGs are used for the training, the smaller the number of false alarms.

3.1.3 Classifiers with Patient-specific, Semi-patient-specific and Non Patient-specific Information

We carried out three experiments to explore the effects of patient-specific and non-patient-specific data as training data on the performance of classifiers. Our method was designed to use patient-specific information, which includes seizures and interictal EEGs from one single patient, to train classifiers and use it to detect seizure onsets in the same patient. It is reasonable to think that a classifier trained with information from more than one person could possibly better than the one trained with information from one patient only, because more training data could result in a better performance in a classifier. We therefore first design an experiment to train classifiers with semi-patient-specific information which includes seizures from a patient A and interical EEGs from many patients, and utilize the classifier to detect seizures of the patient A. In order to fully explore this issue, another experiment was designed to train classifiers with non patient-specific information only which consists of both seizures and interictal EEGs from many patients, and use it to detect seizures of a new patient. Details of experiments are explained later.

The experiment for using totally patient-specific information was done by training the classifier with the first available seizure and interictal EEG sampled during 24 hours in each patient. Details of this experiments were discussed in the previous section. With a distance threshold coefficient at 0.5, the results are shown in table 4-4.

3.1.3.1 Classifiers with Semi-patient-specific Information

The experiment for training the classifier with semi-patient-specific information includes one seizure for a patient and interictal EEGs from many patients. It was implemented in the same way as in figure 4-5, except that interictal EEG training data

come from many patients instead of a specific patient whose first seizure is used to train the same classifier.

This experiment could be carried out in the patients with scalp electrodes only because this set of patients has the same detection montage. This makes it possible to compare EEGs between different patients because all EEGs were recorded from the same electrodes. This is not the case for patients with depth electrodes because these patients have individualized electrode placement. As a result, patients with depth electrodes are not included in this experiment. We combined interictal EEGs of all patients except one and used them to train a classifier for that left-over patient. Every patient's classifier was trained in this way. With the distance threshold coefficient set at 1.2, as shown in table 4-4, there is an average false alarm rate at 0.04/hour. With this distance threshold coefficient, the seizure onset detection rate is 48.6% and the average weighted delay is 34.3 seconds. This gives the classifier a QV of 1.22.

We only use a distance coefficient of 1.2 in this experiment instead of any other value for the following reasons: (1) This is a very computationally and memory intensive experiment: computing results for one distance threshold coefficient takes about one month of personal computer time. Therefore, it is very difficult to search the best distance coefficient in this experiment as it was done in the first experiment. (2) The goal of the experiment is to compare performance among classifiers. If the classifier with patient-specific information and a distance threshold coefficient of 0.5 is called "classifier A", the classifier in this experiment can be called "classifier B". As shown in table 4-4, classifier B with a distance threshold coefficient of 1.2 has a higher false alarm rate and a lower onset detection rate than classifier A, and therefore is worse than classifier A. As explained earlier, both onset detection rate and false alarm rate increase when the distance threshold coefficient increases, and vice versa. Classifier B with a distance threshold coefficient higher than 1.2 will not have a better result than classifier A because the onset detection rate in classifier A is maximum (100%) and the false alarm

rate in classifier B already surpasses that of classifier A. If classifier B has a distance threshold coefficient lower than 1.2, its QV has an upper bound at the QV of classifier B with a distance coefficient of 1.2 but for which the false alarm rate is set to zero. Even for classifier B with upper-bound, the QV is only 1.46, still much lower than classifier A. As a result, knowing the performance of classifier B with the distance threshold coefficient of 1.2, it can be concluded that classifier B is worse than classifier A. More details about the proof of this conclusion can be seen in Appendix C.

3.1.3.2 Classifiers with Non Patient-specific Information

We also designed an experiment to train classifiers with non patient-specific information. Only patients with scalp electrodes were used in this experiment because only this group of patient has the same montage so that their EEGs can be combined into a single training data set. Since there are twelve patients in this group, we used template seizures of eleven patients and their interictal EEGs sampled from 24 hours as training data sets. A non patient-specific classifier is then trained with these training sets. In other words, as in figure 4-5, the seizure training data contain first seizures of eleven patients; the interictal EEG training data consist of the same eleven patients' interictal EEGs sampled from 24 hours in each patient. This classifier is then used to evaluate the onset detection rate and the false alarm rate on the seizures and interictal EEGs from the patient who is not included in the training sets. This classifier is therefore a non patient-specific one because this classifier was not trained with any information coming from the patient whose data are used for the testing. The same procedure was applied to every patient until each patient's data have been a testing set once and only once. A distance threshold coefficient of 1.0 instead of 1.2 was used because of the same reasons described in the previous experiment. This classifier is much worse than the classifier in the first experiment because it only has the QV value at 0.46.

Results can be seen in table 4-4. It shows that classifiers trained with patient-specific information have better results than classifiers trained with either semi-patient-specific or non-patient-specific information.

Table 4-4: Comparison of classifiers with patient-specific, semi-patient-specific and non patient-specific information

Classifier	PS	Semi-PS	Non PS
Template seizures come from	one patient	one patient	many patients
Interictal EEGs come from	same patient	many patients	many patients
Interictal EEG sampled from	24 hours	264 hours	264 hours
Number of training seizures	1	1	11
Distance threshold coefficient	0.5	1.2	1.0
Detection rate	100%	48.6%	11.4%
Weighted delay	9.6s	34.3s	53.9s
false alarm rate	0.03/h	0.04/h	0.20/h
Quality Value (QV)	4.58	1.22	0.46

3.2 The Best Classifier

From all above classifiers, we find that the best classifier has a distance threshold coefficient of 0.5 and it is trained by two seizures and interictal EEG sampled from a period of 48 hours. However, the availability of seizures has to be considered as a factor because seizures are rare events. A classifier trained by one seizure and interictal EEG from 24 hours is therefore more practical while maintaining a high quality. This classifier can detect 100% of seizures with an average detection delay of 9.6 seconds after seizure

onset. Although the classical seizure detection method of Gotman (1982, 1990a) was not meant for seizure onset detection, one can compare the results of the two methods to get an idea of the order of magnitude of the difference because there is no existing onset detection method for direct comparison. The classical method gave a seizure detection rate of 87% and an average onset detection delay of 22.4 seconds. The average false alarm rate for our method is 0.21/hour, compared to an average of 2.2/hour in the classical method. The details are shown in table 4-5.

Some seizure onsets can be detected very early if they are abrupt. Figure 4-7A shows a template seizure with an abrupt onset. In figure 4-7B, the seizure is from the same patient and its onset was reported 2.5 seconds after the onset occurred. The delay was caused by the detection epoch which has a length of 2.56 seconds. From figure 4-7C, we can see a template seizure which evolves gradually after the onset. The earliest seizure pattern is hard to distinguish from some interictal EEGs from the same patient. This is why an onset detection was reported only 9.9 seconds after the onset of a similar seizure (figure 4-7D). These examples illustrate that abrupt seizure onsets are easier to detect than gradual ones.

Table 4-5: Results of the best classifier

	Scalp	Depth	Total
Number of types of seizures	12	12	24
Number of patients	12	5	17
Number of seizures tested	35	42	77
ASZ[1] detection rate	83%	90%	87%
Seizure onset detection rate	**100%**	**100%**	**100%**
ASZ detection delay (sec)	19.7	23.6	22.4
Onset detections delay (sec)	**9.5**	**9.6**	**9.6**
False alarm rate of ASZ (/h)	1.8	2.5	2.2
False alarm rate (/h)	**0.03**	**0.37**	**0.21**
Interictal EEG tested (hrs)	29.7	35.2	64.9

[1] ASZ means classic automatic seizure detection method of Gotman (1982, 1990).

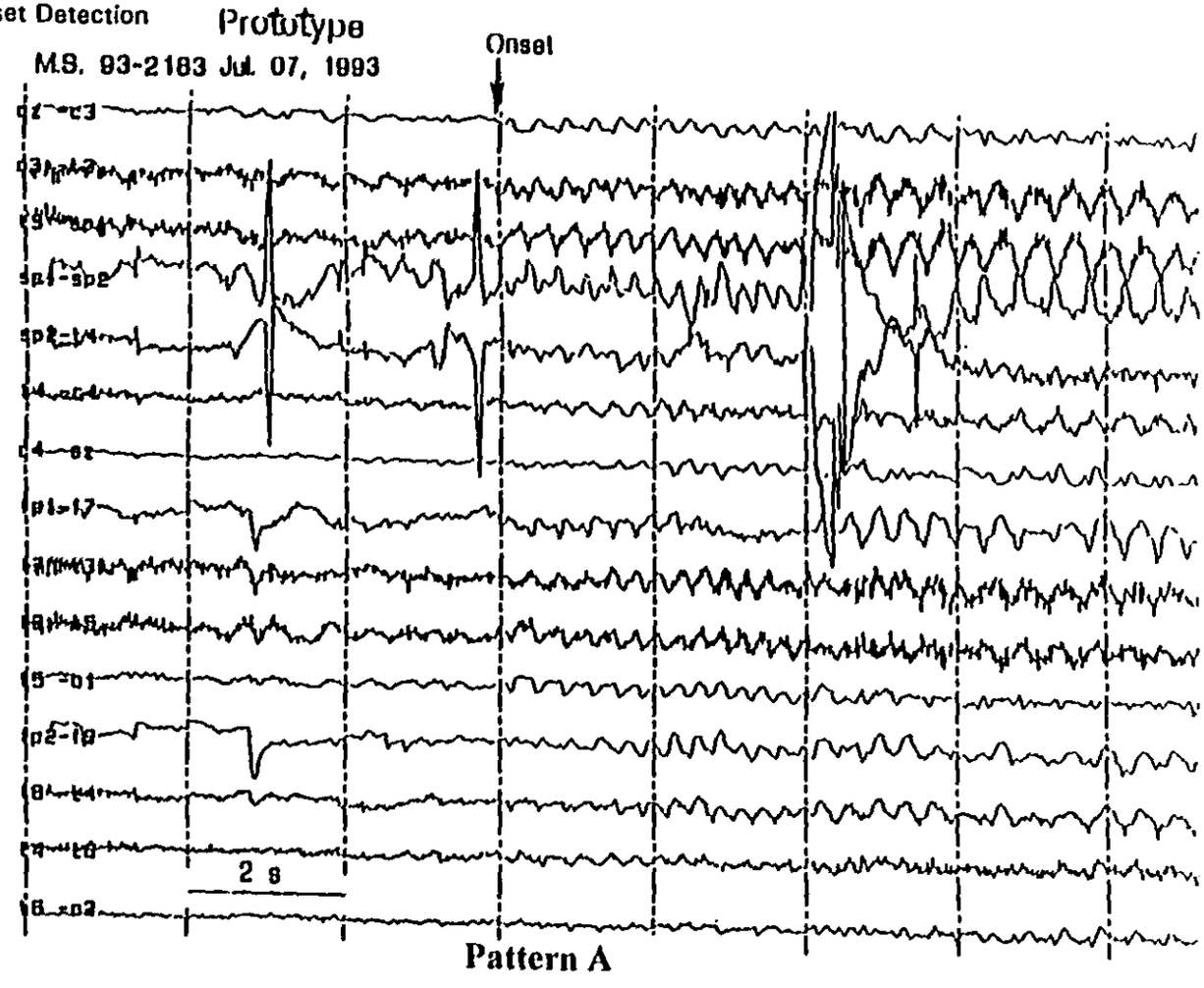
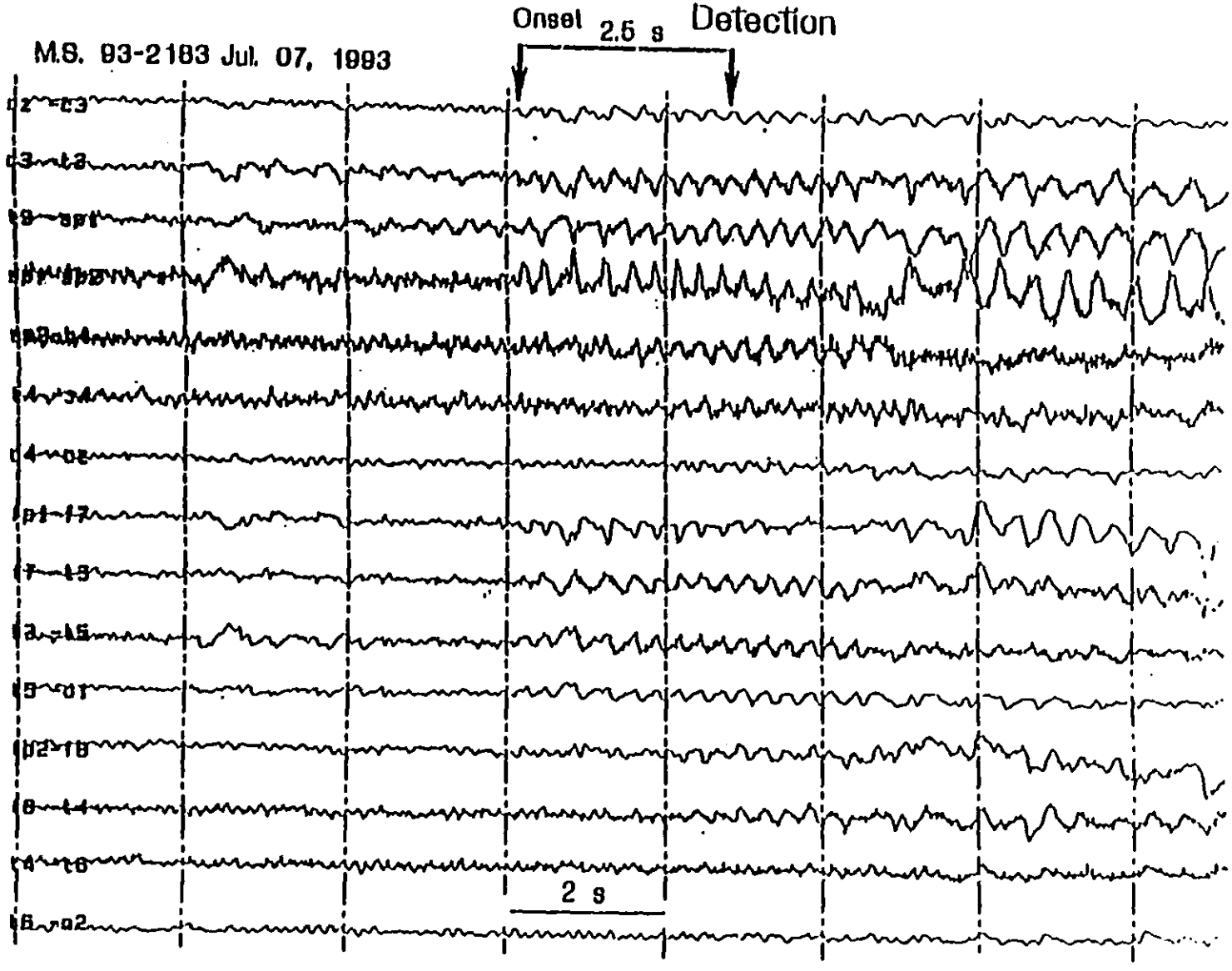
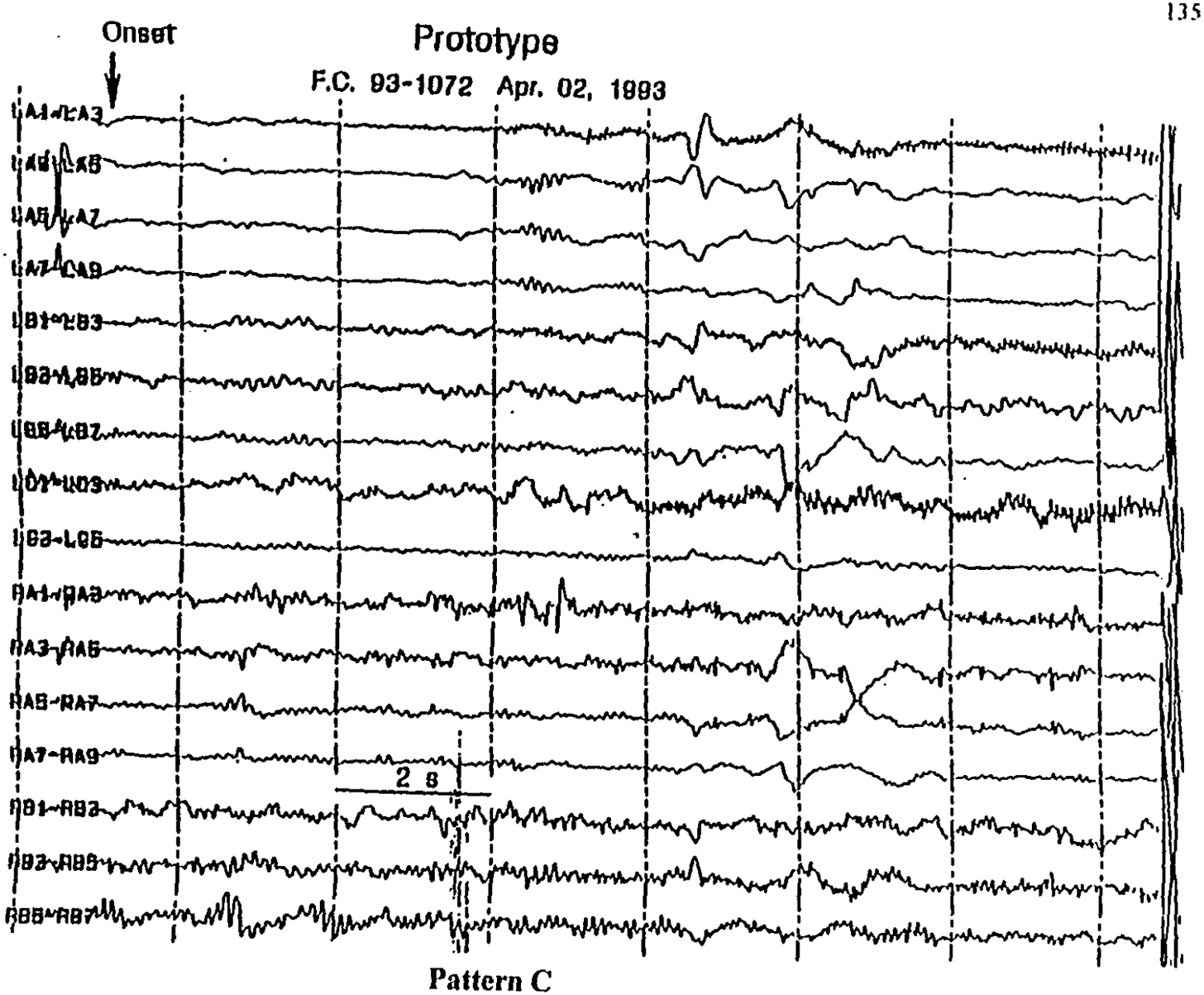


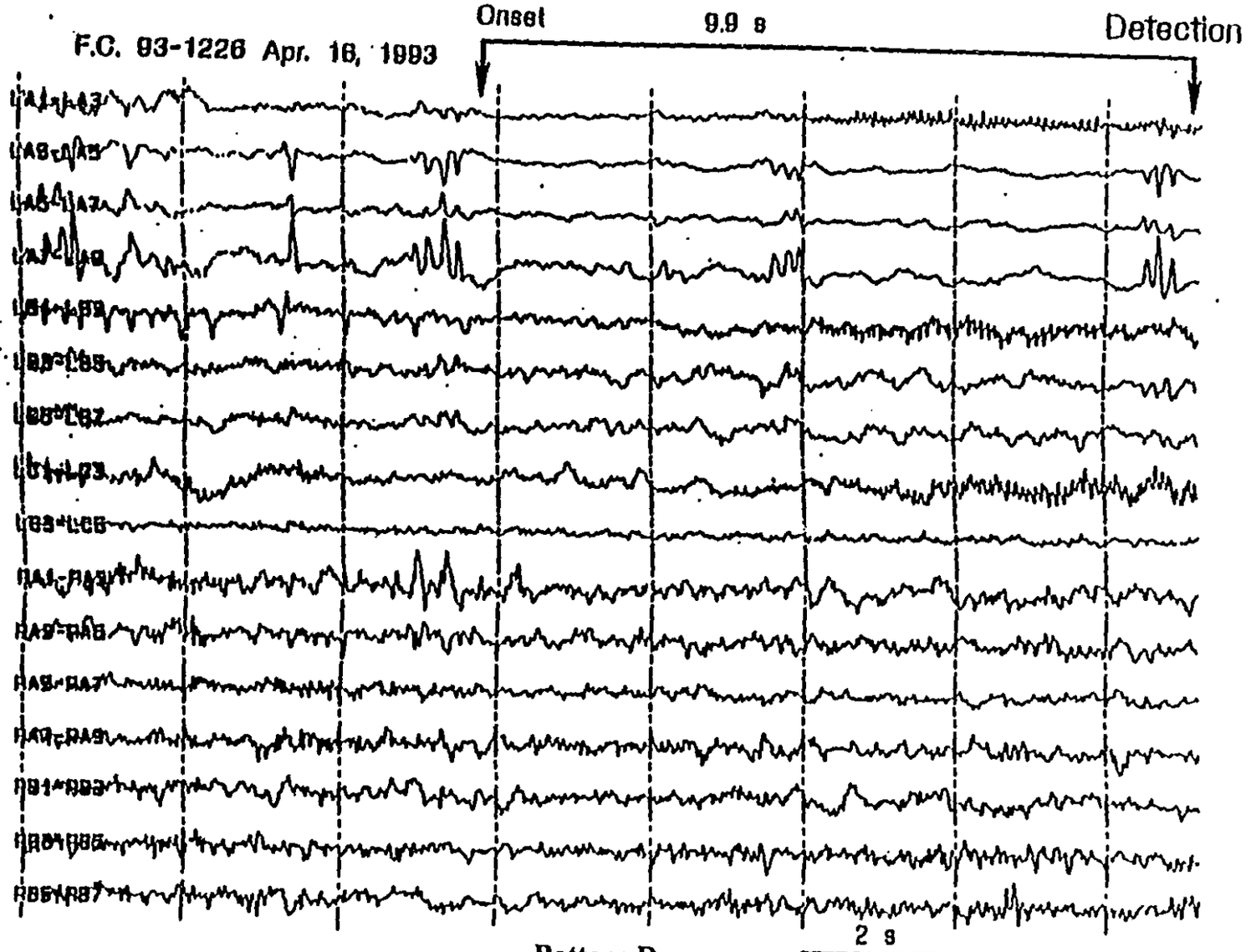
Figure 4-7: Detected seizure onsets and their templates: Pattern A is a template seizure with an abrupt onset. Pattern B is a similar seizure from the same patient and the onset detection was made 2.5 seconds after the onset. Pattern C is another template seizure with a gradual evolution after the onset. Pattern D is a similar seizure and the onset detection was made 9.9 seconds after the onset because of the gradual evolution of the seizure pattern.



Pattern B


Onset
Prototype
F.C. 93-1072 Apr. 02, 1993
LA1-LA3
LA3-LA5
LA5-LA7
LA7-LA9
LB1-LB3
LB3-LB5
LB5-LB7
LO1-LO3
LO3-LO5
RA1-RA3
RA3-RA5
RA5-RA7
RA7-RA9
RB1-RB3
RB3-RB5
RB5-RB7
2 s

Pattern C



Pattern D

4 Discussion

The results show that this method is able to detect patient-specific seizure onsets accurately, reliably and quickly. Large sets of interictal EEGs sampled evenly from a very long period of time ensure that the estimation of false alarm rates is accurate and unbiased. It can be noticed that there is not a large difference in results between scalp electrode patients and depth electrode patients except for false alarm rates, which are more frequent in depth electrode patients. This is also the case for the classic seizure detection method. This is because the EEG of patients with depth electrodes has a larger variety of patterns and greater dynamic range than the EEG of patients with scalp electrodes.

4.1 Motivations and Goals

During long-term EEG monitoring, a patient may have a seizure at any time. In most cases, if the patient does not feel the seizure coming, seizures can only be noticed when behavioral manifestations or prominent EEG patterns occur and the patient is under constant observation. Since constant observation of a patient is a tiresome and expensive procedure, a seizure onset warning device allows observers to interact with the patient when a seizure occurs, thus revealing some important information, such as memory and speech ability. The problem is particularly serious for apparently subclinical seizures or seizures with minor behavioral manifestation: these kinds of seizures are usually missed by observers and no interaction takes place during them. A system that can give a warning signal when a seizure onset occurs is therefore very useful in this situation. Since this system is designed as an independent on-line detection system with high

accuracy, it could possibly be used as a warning device outside the hospital, such as in an ambulance, or while monitoring at home.

The goal of this method is to detect seizures with a high seizure detection rate, a short detection delay from onset and a low false alarm rate. It is impossible to obtain a perfect performance in all these conditions at the same time. A compromise has to be made. Among these three goals, a low false alarm rate is the most important one. A high false alarm rate ruins the effectiveness of the device because warnings will be ignored by observers if most of them are false. With a reasonably low false alarm rate, a high seizure detection rate is the next thing to be considered. Missing a seizure is worse than detecting a seizure with a longer delay from its onset. A short detection delay should be considered when the system maintains a low false alarm rate and a high seizure detection rate. In our method, using the distance threshold coefficient and relabeling some template points are measures used to reduce false alarm rates, all at the expense of longer detection delays and possibly lower seizure detection rates.

In addition to making early seizure onset detections, this method also detects some seizures that are missed altogether by other detection methods: if a seizure is missed by the standard seizure detection method, similar seizures occurring later are also likely to be missed. By using our method, if one of this kind of seizures is somehow captured, all other similar seizures will also be detected.

4.2 Training of Classifiers

In data selection for the method, 12 types of seizures are used from patients with scalp electrodes and 12 in patients with depth electrodes. This is to reduce the variance of results between patients with different electrodes. Each type of seizures contains three to seven seizures. There are two reasons for this selection criterion: (1) If there are fewer than three seizures, there are not enough seizures to test the consistency of the

performance of a classifier in a patient. (2) Too many seizures of one type will give a higher weight to this type when results are averaged from all types. In this case, results may be misleading.

It is shown in the results that the amount of training data affects the quality of classifiers. When both seizure and interictal training data are used appropriately to reflect a priori knowledge of the frequency of seizures in patients, which is one seizure with interictal EEG from 24 hours or twice as much in both, classifiers can be well trained. In contrast, using too much seizure training data, or too much interictal training data, will result in deterioration of the overall performance of classifiers even though one aspect of the performance is improved. This is because there are three aspects in the determination of classifier performance. Improving one of them may deteriorate others. The amount of seizure and interictal training data have to be balanced. From our experiments, a seizure with interictal EEG from 24 hours is the minimum training set. Any training data set having a multiple of this minimum training set in both seizure and interictal data will train classifiers better.

The amount of EEG used to train a classifier determines the quality of the classifier. When a training data set contains all interictal EEG patterns of a patient, the quality of the classifier with respect to false alarm rate is the highest. Obviously, putting all interictal EEGs available into the training data set is the best way. However, this is unnecessary because most of them are redundant data. The most important thing in selecting the interictal data is not the size of data, but the period of time during which the data are sampled because the longer that period, the more variety in the EEG patterns likely to be included. For instance, one hour of EEG acquired continuously does not include as many kinds of EEG patterns as one hour of EEG made from samples taken evenly throughout a period of 24 hours. Because of the human biological clock, most EEG patterns have a repetition rate of 24 hours or less. Although some interictal EEG patterns may occur or disappear from one day to the next, most interictal EEG patterns

can be found within a day. Therefore, getting interictal EEG from a period of 24 hours is a minimum requirement to represent most interictal EEG patterns. Since a high sampling rate results in too much data and a low sampling rate takes the risk of being less representative, a compromise has to be made. We used a sampling rate of one minute every thirty minutes to collect interictal EEGs in an interval of 24 hours for these two reasons: (1) This will create a set of EEG with a total length at 0.8 hours which is a reasonable size for processing. (2) This set of data can be considered representative of most interictal EEG patterns of a patient.

It is possible to retrain a classifier with new interictal EEGs if it appears that they have changed during the monitoring session. By doing so, the classifier is kept up to date with respect to possible slow changes occurring over several days.

4.3 Method

False alarms occur only when patterns in interictal EEGs are similar to some patterns in the template seizure. If an onset occurs in a total of N channels, an interictal event is unlikely to trigger a false alarm because it is unlikely that EEG patterns in half of N channels have matches with the template at the same time. It is likely though that in one of N channels at some time an EEG pattern has a match with the template. As a result, the location feature plays an important role in the elimination of false alarms. In some patients with depth electrodes, seizures may be very focal and include only one or two channels. In terms of the constraint of the location feature, as long as there is a match in one of two selected channels, the constraint is satisfied. In this case, a false alarm is more likely to occur because there is a high probability that the constraint can be satisfied. Although the location feature is very important in our method, it also relies on the right features, practical distance measure and a conservative approach in the selection

of the classification boundary. All these steps combine to make our system accurate and reliable.

Our method is totally based on patient-specific information to detect patient-specific events. We designed experiments to compare classifiers trained with patient-specific, semi-patient-specific and non-patient-specific information. Results indicate that the use of patient-specific information provides the best performance. This provides confirmation of the validity of our original approach.

The method was implemented and tested off-line. However, our method has taken into account memory management and computational load so that it may be eventually implemented on-line. For instance, the classifier uses the modified NN rule rather than the k-NN rule. This is not just because both can achieve the similar error rate in the recognition of EEG pattern (Gevins 1987a), but also because the k-NN rule needs much more memory and computation, as discussed in the section on classifiers.

Our method has one major shortcoming: it only detects seizures similar to the template. In epilepsy monitoring, one wants to explore as many kinds of seizures as possible. Therefore, our method cannot replace traditional unbiased seizure detection which aims at recording all kinds of seizures. Our method, however, can be used together with traditional seizure detection methods to detect possibly many seizures and detect some known ones as early as possible.

Although there is no seizure onset detection method with which to compare our method, there are some methods in speaker recognition which aim at solving a similar problem. For instance, the speech signal is an one-dimensional signal, similar to the EEG signal. Speaker recognition uses pre-recorded speeches from speakers as templates and compares them to a new speech when speaker recognition is required. Soong et al. (1985) used short-time linear predictive coding vectors as feature vectors, and a vector quantization codebook to efficiently characterize the short-time spectral features of a speaker, and a minimum distance (distortion) classification rule to recognize a speaker

according to pre-recorded speeches from N speakers. A minimum distortion is computed for each speaker by comparing this speaker's spectral features with other speakers' spectral features. Colombi et al. (1993) compared a similar method to a neural network approach in the recognition of speakers. His results showed that the neural network approach performs as well as the other method in clean environments and better in noisy environments The speaker recognition problem is similar to our problems because our goal is to distinguish a seizure similar to a pre-recorded one (template) from non-seizure patterns and patterns of other kinds of seizures. Different from EEG signal which is a multichannel signal, speech signal is a single channel signal. That is the reason we used the inter-channel information and there is no similar method in speaker recognition. Both false positive and negative detection in our method are very serious because false alarms will annoy staff people and missing a seizure means the loss of important information about a patient. However, in speaker recognition, false negative detection is not very serious because the speaker can try again, but false positive detection is very serious because a wrong person may be identified.

Our method is not restrained to detect only one type of seizure per patient. It can be used to detect multiple types of seizures in a patient. This has been demonstrated in our evaluation of patients with depth electrodes. In this group of patients, some patients have up to three types of seizures. In that situation, each type of seizures has its own classifier and therefore detecting three types of seizures in a patient means running three classifiers concurrently. There is a significant negative effect of such a situation: the false alarm rate will increase, possibly by as much as the number of seizure types.

In our evaluation, we only selected similar seizures from 17 patients. Although this set of data is quite representative, it would be better if the data set was larger so that results could be more reliable. In addition, we did not select patients with two similar seizures only because there are not enough data for testing in these patients. In practice, these patients can also use our method because, as long as there is a template, a patient

can start using our method to make onset detection. As a result, our evaluation may not include all possible cases.

4.4 Future Improvements

Our method needs an operator to determine the location feature and the time of onset. If this step can be replaced by a program, the whole system will become fully automatic and therefore no human interaction will be needed. Since this would be done off-line, it is possible to develop a more sophisticated method to analyze a seizure and determine its onset because the computational cost is no longer a serious threat to the implementation.

It would be helpful to re-evaluate our method on a group of unselected patients. Our method was only evaluated in 17 selected patients. If we redo our evaluation in a practical clinical situation without any restrictions, several interesting questions can be answered: (1) What is the percentage of all kinds of seizures which can be detected? (2) What is the percentage of patients who can benefit from our method? (3) In what sense can seizures be called "similar" in our method?

Our method pays much attention to keep the false alarms as few as possible. When a patient has a very low false alarm rate, it will be possible to increase the distance threshold coefficient so that onset detection delay can be shorter at the expense of possibly a higher false alarm rate, as shown in results. Therefore, a false alarm rate related adapting algorithm could be utilized to keep false alarm rate at a constant low level, for instance one in 24 hours, by changing the distance threshold coefficient for each patient at different times. This algorithm will make the system reliable with a constant level of false alarm rate, while the onset detection can be as early as possible.

In our method, all features are computed from the current epoch only. This contributes to false alarms since some isolated patterns in the background may be similar

to some patterns in templates. These false alarms could be avoided if there is a new feature containing the information about the evolution of seizures to distinguish those isolated patterns similar to the template from real seizures. In this case, only patterns, similar to the template and with evolution characteristics similar to the template seizure, will be detected as seizures. This will decrease the probability of false alarms significantly However, since this feature considers the evolution of a seizure, it needs more than one epoch to measure. This will delay the possible detection of a seizure and therefore is a shortcoming in the onset detection. The best system should use our method together with a method using the evolution feature. Since in some cases our method cannot detect onsets early enough, this system will decrease the delay of the onset detection in these cases with few chances to cause additional false alarms.

Chapter 5:

Conclusion

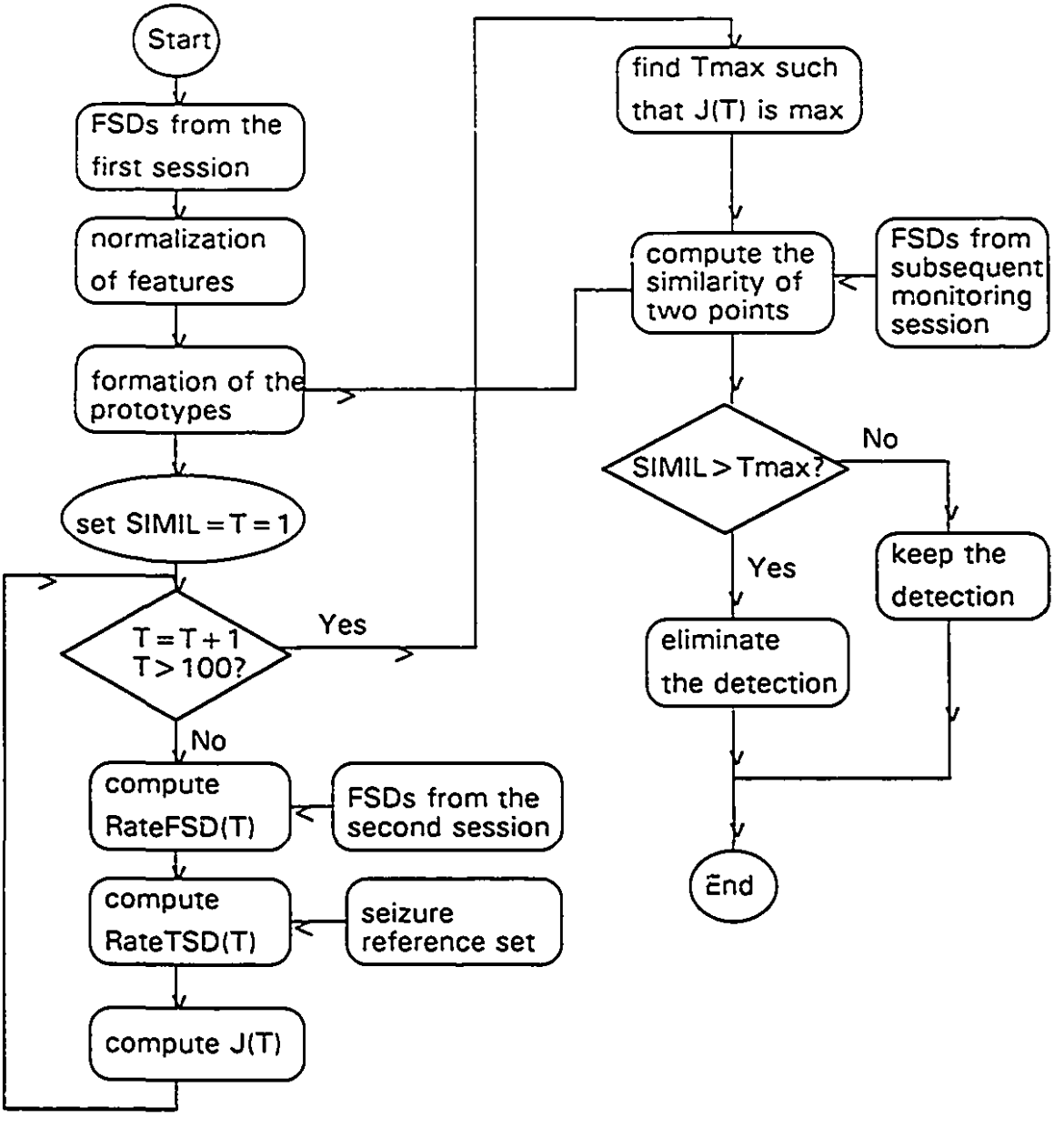
Our work represent a new approach to the field of long-term epilepsy monitoring, particularly in the area of the reduction of false seizure detections and early detection of seizures. Since ictal EEG patterns are highly variable from patient to patient, a universal algorithm to detect all seizures without causing many false detections is very difficult to achieve. An attempt has been made by Gotman (1982, 1990), but the performance of this system has much room for improvement (Pauri, 1992). In some patients, the ictal EEG patterns, as well as interictal EEG patterns, tend to repeat frequently during long-term EEG monitoring. Therefore, patient-specific algorithms can focus better on certain patient-specific patterns. By using a new "similarity" measure and a patient-specific classifier, the method for reducing false seizure detections increases the performance of seizure detection greatly in terms of sensitivity and accuracy. A comprehensive evaluation has been done by using the holdout method and the rotation method.

Although on-line seizure *onset* detection has never been attempted before because of the diversity of onset patterns, our algorithm for detecting seizure onsets indicates that a high accuracy of detection, a reasonable short delay and a low false alarm rate are possible. It has also been shown that a modified NN rule can perform well in onset detection while it has a minimum computational cost. Therefore, this method is able to perform in real time to serve as a warning device.

Because there is a large variety in the EEG among patients, a universal classifier can hardly perform well in the detection of abnormal EEG patterns, such as spikes and seizures. It has been illustrated by our methods that a classifier which has more patient-specific information performs better. Gotman (1990) indicated that, by extracting the context from a few seconds to approximately one minute, his seizure detection method performed better than before. Our methods used patient-specific information obtained hours and even days preceding the recording and therefore performed very well. This brings the computer method closer to human interpretation of the EEG since the EEGer uses, consciously or unconsciously, information from past recordings when interpreting an EEG. This concept can certainly be used in the detection of other EEG patterns, such as spikes and seizures in children and infants, or other physiological signals, such as abnormal ECG patterns. The main disadvantage of the concept is that some patient-specific information has to be acquired before the system can start working. This is certainly a major problem during short-term recordings, but it is much less critical during long-term monitoring.

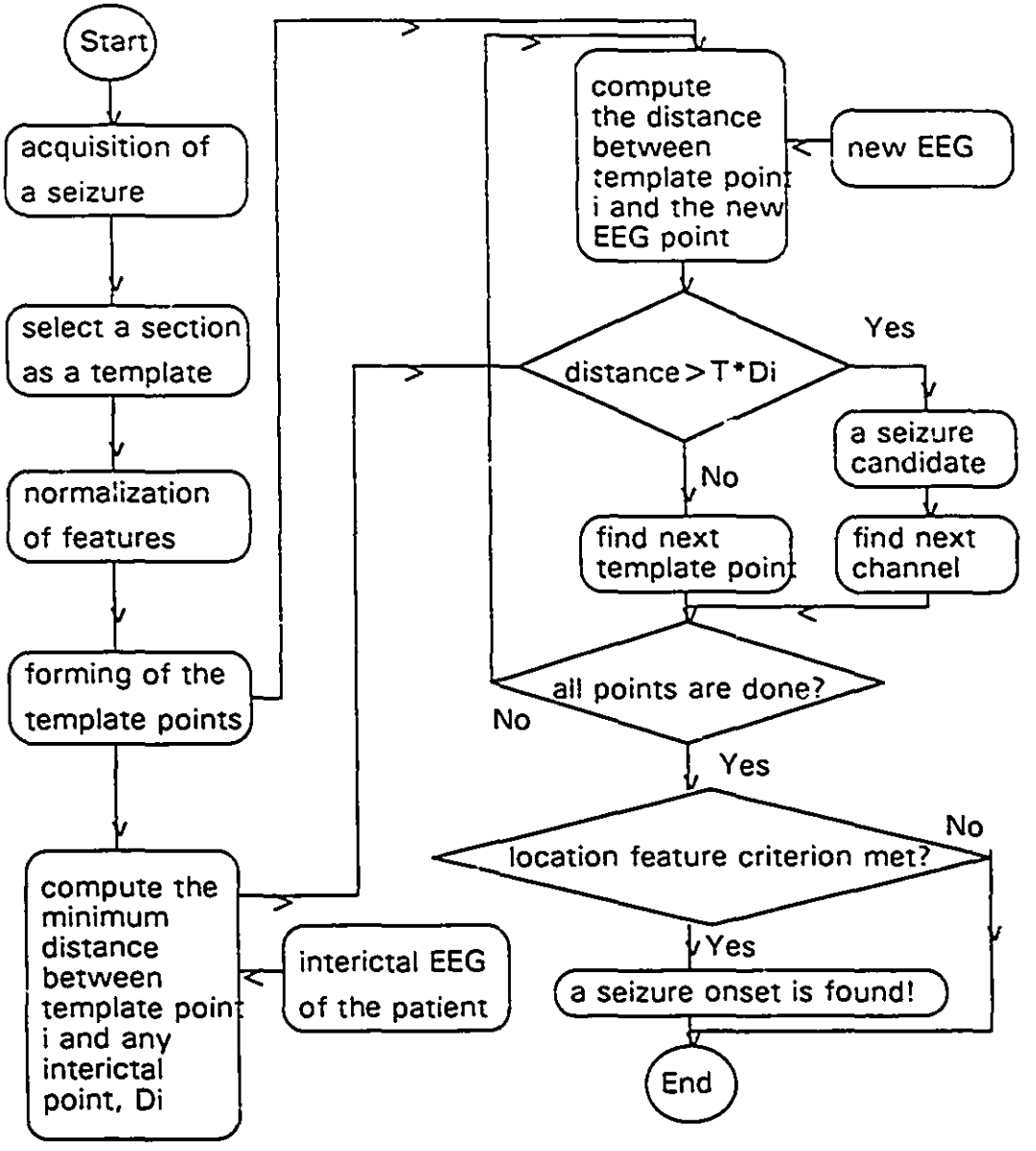
Appendix A:

Block Diagram of the Algorithm for the Reduction of False Detections



Appendix B:

Block Diagram of the Algorithm for the Detection of Seizure Onsets



Appendix C:

Comparison of the Quality of Classifiers

Let us start with the following known conditions:

(1) $QV = \frac{C}{(R_{fa} + 0.2) \times (T_{dly} \times P_{dct} + 60 \times P_{msd})}$.

(2) T is the distance coefficient.

(3) Both R_{fa} and P_{dct} monotonically increase with T.

(4) $P_{msd} = 1 - P_{dct}$.

(5) Weighted delay $D = T_{dly} \times P_{dct} + 60 \times P_{msd}$.

(6) Classifier A has R_{fa}=0.03, P_{dct}=1, T_{dly}=9.6 and WD=9.6 when T=0.5 and therefore has a QV of 4.58.

(7) Classifier B has R_{fa}=0.04, P_{dct}=0.486, T_{dly}=7.11 and WD=34.3 when T=1.2 and therefore has a QV of 1.22.

We need to prove that classifier B with the optimal T cannot perform better than classifier A with T=0.5.

Now let us look at all three possible cases:

Case 1: classifier A with T=0.5 and classifier B with T=1.2

As we have computed, the QV of classifier A is 4.58, which is higher than that of classifier B, 1.22.

Case 2: classifier A with T=0.5 and classifier B with T<1.2

As we mentioned in the above conditions, both R_{fa} and P_{dct} will decrease when T decreases. In order to find the highest possible upper bound for QV in this case, we set R_{fa} to the lowest possible value, 0; P_{dct} to the highest possible value, 0.486 (the value

corresponding to T=1.2, since lower values of T result in lower values of P_{dcr}); and T_{dly} to the lowest possible value, 0. With all these settings, the upper bound of QV of classifier B is only 1.62, which is far lower than classifier A with T=0.5.

Case 3: classifier A with T=0.5 and classifier B with T>1.2

We already know that both R_{fa} and P_{dcr} increase when T increases. As a result, when T>1.2, R_{fa} becomes larger, although it is already larger than the R_{fa} of classifier A with T=0.5. Since P_{dcr} has the highest possible value, 1, it is also the highest upper bound for the P_{dcr} of classifier B. Assuming classifier B could possibly reduce the weighted delay WD to the level of that of classifier A without increasing R_{fa}, the QV would be 4.3. This extremely unlikely high value for the QV of classifier B is still lower than that of classifier A with T=0.5.

By analyzing all three possible cases, we can conclude here that classifier A with T=0.5 always has a higher QV than classifier B with any value of T.

Bibliography

[1] S. S. Aziz, S. J. Wallace, J. F. Murphy, C. P. Q. Sainsbury, and O. P. Gray, "Cotside EEG monitoring using computerized spectral analysis," *Archives of disease in Childhood*, vol. 61, pp. 242-246, 1986.

[2] B. G. Batchelor, "*Practical approach to pattern classification.*". New York: Plenum, 1974.

[3] J. S. Barlow and J. Dubinsky, "Some computer approaches to continuous automatic clinical EEG monitoring," in: *Qualitative studies in Epilepsy*, P. Kellaway and I. Petersen, Eds. New York: Raven, 1976, pp. 309-327

[4] J. S. Barlow, "Method of Analysis of Nonstationary EEGs with Emphasis on Segmentation Techniques: A Comparative Review," *Journal of Clinical Neurophysiology*, vol. 2(3), pp. 267-304, 1985.

[5] J. Berkhout, D. O. Walter, and R. Adey, "Alternations of the human electroencephalogram induced by stressful verbal activity," *Electroencephalgraphy and clinical neurophysiology*, vol. 27, pp. 457-469, 1969.

[6] B. Bhattacharya, R. Poulsen, and G. Toussaint, "Application of Proximity Graphs to Editing Nearest Neighbor Decision Rules," Techinical report No. SOCS92.19, McGill University, December 1992.

[7] S. T. Bow, "*Pattern recognition, application to large data-set problems.*" New York: Marcel Dekker, 1984.

[8] M. A. B. Brazier, "Spread of seizure discharges in epilepsy: Anatomical and electrophysiological considerations," *Experimental Neurophysiology*, vol. 36, pp. 263-272, 1972

[9] G. E. Chatrian, E. Lettich, and P. L. Nelson, "Modified Nomenclature for the "10%" Electrode System," *Journal of Clinical Neurophysiology*, vol 5(2), pp. 183-186, 1988

[10] C. H. Chen, "Application of pattern recognition to seismic wave interpretation," in *Application of pattern recognition*, K. S. Fu, Editor. Boca Raton. CRC, 1982, pp. 107-119.

[11] L. Chik, R. J. Sokol, and M. G. Rosen, "Computer interpreted detail electroencephalogram: sharp wave detection and classification of infants for one year neurological outcome," *Electroencephalography and clinical neurophysiology*, vol. 42, pp. 745-753, 1977

[12] B. A. Cohen, and A. Sances, "Stationary of the human electroencephalogram," *Medical and biological engineering and computing*, vol. 15, pp. 513-518, 1977

[13] J. M. Colombi, T. R. Anderson, S. K. Rogers, D. W. Ruck, and G. T. Warhola, "Auditory model representation and comparison for speaker recognition," *IEEE international conference on neural networks*, pp. 1914-1917, 1993

[14] J.W. Cooley, and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Mathematics of computation*, vol. 19, pp. 297-301, 1965

[15] T. M. Cover, and P.E. Hart, "Nearest Neighbor Pattern Classification," *IEEE transaction on information theory*, vol. IT-13, No.1, pp. 21-27, 1967

[16] R De Mori, "Automatic speech recognition," in *Application of pattern recognition*, K S Fu, editor, Boca Raton: CRC, 1982, pp. 237-265.

[17] P A Devijver, and J. Kittle, "*Patter recognition: a statistical approach.*" London: Prentice-Hall, 1982.

[18] P. A. Devijver, "Statistical pattern recognition," in *Application of pattern recognition*, K.S. Fu, editor, Boca Raton: CRC, 1982, pp. 15-36.

[19] E. Donchin, E. Callaway, III and R. Jones, "Auditory evoked potential variability in schizophrenia,. II. The application of discriminant analysis," *Electroencephalgraphy and clinical neurophysiology*, vol. 29, pp. 429-440, 1970.

[20] B. Dube, A. -R. LeBlanc, J. L. Dutoy, D. Derome and R. Cardinal, "PC-based St-segment monitoring with the VCG," *IEEE engineering in medicine and biology, tenth annual conference*, vol. 4, pp. 1768-1770, 1988.

[21] R. O. Duda, and P. E. Hart, "*Pattern classification and scene analysis.*" New York: Wiley, 1973.

[22] G. Dumermuth, and H. Fluhler, "Some modern aspects in numerical spectrum analysis of multichannel electroencephalographic data," *Medical and biological engineering*, vol. 5, pp. 319-331, 1967.

[23] G. Dumermuth, P.J. Huber, B. Kleiner, and T. Gasser, "Numerical analysis of electroencephalographic data," *IEEE transaction on audio electroacoustic*, vol. AU-18, No.4, pp. 404-411, 1970.

[24] G. Dumermuth, and L. Molinari, "Spectral analysis of EEG background activity," in: *Handbook of electroencephalography and clinic neurophysiology*, vol. 1: Methods of analysis of brain electrical and

Magnetic signals, A. Gevins, and A Remond, Eds. Amsterdam: Elsevier/North-Holland Biomedical Press, 1987, pp. 85-130.

[25] L. G Durand, M. Blanchard, G. Cloutier, H. N. Sabbah, and P. D. Stein, "Comparison of pattern recognition methods for computer-assisted classification of spectra of heart sound in patients with a porcine bioprosthetic valve implanted in the mitral position," *IEEE trans. on biomedical engineering*, vol. 37, No. 12, pp. 1121-1129, 1990.

[26] J. Engel, Jr., "*Seizures and epilepsy*." Philadelphia: Davis, 1989.

[27] D. Foley, "Consideration of sample and feature size," *IEEE transaction on information theory*, vol.IT-18, pp. 618-626, 1972.

[28] L. Friedman, B. E. Jones, "Study of sleep-wakefulness states by computer graphics and cluster analysis before and after lesions of the pontine tegmentum in the cat," *Electroencephalgraphy and clinical neurophysiology*, vol. 57, pp. 43-56, 1984.

[29] J. D. Frost, Jr., "Mimetic techniques," in: *Handbook of electroencephalography and clinic neurophysiology*, vol. 1: Methods of analysis of brain electrical and Magnetic signals, A. Gevins, and A Remond, Eds. Amsterdam: Elsevier/North-Holland Biomedical Press, 1987, pp. 195-206.

[30] K. S. Fu, "*Sequential machines in pattern recognition and machine learning*." New York: Academic, 1968.

[31] K. Fukunaga and D. Kessell, "Estimation of Classification Error," *IEEE transaction on computers*, vol. c-20, No.12, pp. 1521-1527, 1971.

[32] K. Fukunaga, "*Introduction to statistical pattern recognition*." New York: Academic, 1972.

[33] T. Gasser, "General characteristics of the EEG as a signal," in: *EEG informatic A didactic review of methods and applications of EEG data processing*, A. Remond, Editor, Amsterdam: Elsevier/North-Holland Biomedical Press, 1977, pp. 37-54.

[34] I. Gath and E. Bar-on, "Computerized method for scoring of polygraphic sleep recordings," *Computer programs in Biomedicine*, vol. 11, pp. 217-223, 1980.

[35] I. Gath and E. Bar-on, "Sequential fuzzy clustering of sleep EEG recordings," in: *Methods of sleep research*, A. Kubicki and B. Hermann, Eds. New York: Gustav Fischer, 1985, pp. 55-64.

[36] I. Gath and E. Bar-on, and D. Lehmann, "Automatic classification of visual evoked responses," *Computer methods and programs in biomedicine*, vol. 20, pp. 17-22, 1985.

[37] W. Gersh, "Spectral analysis of EEG's by autoregressive decomposition of time series," *Mathematical biosciences*, vol.7, pp. 205-222, 1970.

[38] A. S. Gevins, G. Zeitlin, C. Yingling, J. Doyle, M. Dedon, J. Henderson, R. Schaffer, J. Roumasset, and C. Yeager, "EEG patterns during 'cognitive' tasks: Part 1, methodology and analysis of complex behaviors," *Electroencephalgraphy and clinical neurophysiology*, vol. 47, pp. 693-703, 1979a.

[39] A. S. Gevins, C. Zeitlin, J. Doyle, R. Schaffer, and E. Callaway, "EEG patterns during 'cognitive' tasks: Part 2, analysis of controlled tasks," *Electroencephalgraphy and clinical neurophysiology*, vol. 47, pp. 704-710, 1979b.

[40] A. S. Gevins, "Pattern recognition of human brain electrical potentials," *IEEE transactions on pattern analysis and machine intelligence*, vol. PAMI-2, No.5, pp. 383-404, 1980

[41] A. S. Gevins, "Statistical pattern recognition," in: *Handbook of electroencephalography and clinic neurophysiology*, vol. 1: Methods of analysis of brain electrical and Magnetic signals, A. Gevins, and A Remond, Eds. Amsterdam: Elsevier/North-Holland Biomedical Press, 1987a, pp. 541-582

[42] A. S. Gevins, "Correlation analysis," in: *Handbook of electroencephalography and clinic neurophysiology*, vol. 1: Methods of analysis of brain electrical and Magnetic signals, A. Gevins, and A Remond, Eds. Amsterdam: Elsevier/North-Holland Biomedical Press, 1987b, pp. 171-193

[43] A. S. Gevins, J. Le, N. K. Martin, P. Brickett, J. Desmond, and B. Reutter, "High resolution EEG: 124-channel recording, spatial deblurring and MRI integration methods," *Electroencephalography and clinical neurophysiology*, vol. 90, pp. 337-358, 1994.

[44] P. Gloor, "Hans Berger on the electroencephalogram of man," *Electroencephalgraphy and clinical neurophysiology*, suppl. 28, pp. 1-25, 1969.

[45] P. Gloor, "General introduction," in: *Long-term monitoring in epilepsy*, J. Gotman, J. R. Ives and P. Gloor, Eds. *Electroencephalgraphy and clinical neurophysiology*, suppl. 37, pp. xiii-xx, 1985a.

[46] P. Gloor, "Neuronal Generators and the Problems of Localization in Electroencephalography: Application of Volume Conductor Theory to

Electroencephalography," *Journal of Clinical Neurophysiology*, vol. 2(4), pp. 327-354, 1985b.

[47] J. R. Glover, P. Y. Ktonas, N. Raghavan, J. M. Urunuela, S. S. Velamuri, and E L. Reilly, "A Multichannel Signal Processor of the Detection of Epileptogenic Sharp Transients in the EEG," *IEEE transactions on biomedical engineering*, vol. BME-33, No. 12, pp. 1121-1128, 1986.

[48] J. R. Glover, N. Raghavan, P. Y. Ktonas, and J. D. Frost, "Context-based Automated Detection of Epileptogenic Sharp Transients in the EEG: Elimination of False Positives," *IEEE transactions on biomedical engineering*, vol. BME-36, No. 5, pp. 519-527, 1989.

[49] R. Gonzalez, and M. Thomason, "Syntactic pattern recognition, an introduction," Reading, MA: Addison-Welsey, 1978.

[50] J. Gotman, D. R. Skuce, C. J. Thompson, P. Gloor, J. R. Ives, and W. F. Ray, "Clinic applications of spectral analysis and extraction of features from electroencephalograms with slow waves in adult patients," *Electroencephalography and clinic neurophysiology*, vol. 35, pp. 225-235, 1973.

[51] J. Gotman, "Automatic Recognition and Quantification of Interictal Epileptic Activity in the Human Scalp EEG," *Electroencephalgraphy and clinical neurophysiology*, vol. 41, pp. 513-529, 1976.

[52] J. Gotman, "Quantitative Measurements of Epileptic Spike Morphology in the Human," *Electroencephalgraphy and clinical neurophysiology*, vol. 48, pp. 551-557, 1980.

[53] J. Gotman, "Interhemispheric relations during bilateral spike-and-wave activity," *Epilepsia*, vol. 22, pp. 453-466, 1981.

[54] J. Gotman, "Automatic recognition of epileptic seizure in the EEG," *Electroencephalgraphy and clinical neurophysiology*, vol. 54, pp. 530-540, 1982.

[55] J. Gotman, "Measurement of small time differences between EEG channels: Method and application to epileptic seizure propagation," *Electroencephalography and clinic neurophysiology*, vol. 56, pp. 501-514, 1983.

[56] J. Gotman, J. R. Ives, P. Gloor, L.F. Quesney, and P. Bergsma, "Monitoring at the Montreal Neurological Institute," in: *Long-Term Monitoring in Epilepsy*. J. Gotman, J. R. Ives and P. Gloor, Eds. *Electroencephalgraphy and clinical neurophysiology*, Suppl.37, pp. 327-340, 1985.

[57] J. Gotman, "Seizure recognition and analysis," in: *Long-Term Monitoring in Epilepsy*. J. Gotman, J. R. Ives and P. Gloor, Eds. *Electroencephalgraphy and clinical neurophysiology*, Suppl.37, pp. 133-145, 1985.

[58] J. Gotman, "Automatic seizure detection: improvements and evaluation," *Electroencephalgraphy and clinical neurophysiology*, vol. 76, pp. 317-324, 1990a.

[59] J. Gotman, "The use of computers in analysis and display of EEG and evoked potentials," in: *Current practice of clinic electroencephalography*, D. D. Daly, and T.A. Pedley, Eds. Second edition, New York: Raven, 1990b, pp. 51-83.

[60] J. Gotman, and L. Y. Wang, "State-dependent spike detection: concepts and preliminary results," *Electroencephalgraphy and clinical neurophysiology*, vol. 79, pp. 11-19, 1991.

[61] J. Gotman, and L. Y. Wang, "State dependent spike detection: validation," *Electroencephalgraphy and clinical neurophysiology*, vol. 83, pp. 12-18, 1992.

[62] H. L. Gray, and W. R. Schicany, "*The generalized Jackknife Statistic*," New York: Deikker, 1972, pp. 115-136

[63] G.W. Harding, "An Automated Seizure Monitoring System for Patients with Indwelling Recording Electrodes," *Electroencephalgraphy and clinical neurophysiology*, vol. 86, pp. 428-437, 1993.

[64] P. Hilfiker, and M. Egli, "Detection and Evolution of Rhythmic Components in Ictal EEG Using Short Segment Spectra and Discriminant Analysis," *Electroencephalgraphy and clinical neurophysiology*, vol. 82, pp. 255-265, 1992.

[65] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalgraphy and clinical neurophysiology*, vol. 29, pp. 306-310, 1970.

[66] R. Horst, and E. Donchin, "Beyond averaging II: Single-trial classification of exogenous event-related potentials using step-wise discriminant analysis," *Electroencephalgraphy and clinical neurophysiology*, vol. 48, pp. 113-126, 1980.

[67] P. J. Huber, B. Kleiner, T. Gasser, and G. Dumermuth, "Statistical methods for investing phase relations in stationary stochastic process," *IEEE transactions on audio electroacoustic*, vol. AU-19, pp. 78-96, 1971.

[68] IFSECN, "A glossary of terms commonly used by clinical electroencephalographers," *Electroencephalgraphy and clinical neurophysiology*, vol. 37, pp. 538-548, 1974.

[69] J. Ives, and J. F. Woods, "4-channel 24 hours cassette recorder for long-term EEG monitoring of ambulatory patients," *Electroencephalography and clinical neurophysiology*, vol. 39, pp. 88-92, 1975.

[70] J. Ives, "Analysis of ambulatory outpatient EEG recordings," *Proceedings of the fifth international Cleveland Clinic-Bethel epilepsy symposium*, pp. 19, May, 1994.

[71] M. James, "Categorical Variables and Non-parametric Methods," in: *classification algorithms*, M. James, editor, New York: Wiley, 1985, pp. 149-173.

[72] H. Jasper, "Report of the committee on methods of clinical examination in electroencephalography," *Electroencephalography and clinical neurophysiology*, vol. 10, pp. 370-375, 1958.

[73] E. John, "*Functional neuroscience, vol.2, Neurometrics*," E. John and R. Thatcher, Eds., Hillsdale: Erlbaum, 1977.

[74] P.Y. Ktonas, "Automated spike and sharp wave (SSW) detection," in: *Handbook of electroencephalography and clinic neurophysiology*, vol. 1: Methods of analysis of brain electrical and Magnetic signals, A. Gevins, and A Remond, Eds. Amsterdam: Elsevier/North-Holland Biomedical Press, 1987, pp. 211-242

[75] L. E. Larsen, and D. O. Walter, "On automatic methods of sleep staging by EEG spectra," *Electroencephalography and clinical neurophysiology*, vol. 28, pp. 459-467, 1970.

[76] H. S. Leader, R. Cohen, A. L. Weihrer and C. A. Caceres, "Pattern reading of the clinical electroencephalogram with a digital computer," *Electroencephalgraphy and clinical neurophysiology*, vol. 23, pp. 566-570, 1967.

[77] A. Liu, J S Hahn, G. P. Heldt and R. W. Coen, "Detection of neonatal seizures through computerized EEG analysis," *Electroencephalgraphy and clinical neurophysiology*, vol. 82, pp. 30-37, 1992.

[78] F. Lopes da Silva, A. Kijk, and H. Smiths, "Detection of nonstationaries in EEGs using the autoregressive model. An application to EEGs of epileptics," in: *CERN: computer EEG analysis*, G. Dolce and H. Kunkel, Eds. New York: Gustav Fischer, 1975, pp. 180-199.

[79] F. Lopes da Silva, "EEG analysis: Theory and Practice," in: *Electroencephalography, Basic Principles, Clinical Applications and Related Fields*. Ernst Niedermeyer, Fernando Lopes da Silva, Eds. Second Edition, Baltimore-Munich: Urban&Schwarzenberg, 1987a, pp. 871-897

[80] F Lopes da Silva and N. J. I. Mars, "Parametric methods in EEG analysis," in: *Handbook of electroencephalography and clinic neurophysiology*, vol. 1: Methods of analysis of brain electrical and Magnetic signals, A. Gevins, and A Remond, Eds. Amsterdam: Elsevier/North-Holland Biomedical Press, 1987b, pp. 243-260

[81] M. D. Low, "Psychology, psychophysiology, and the EEG," in: *Electroencephalography, Basic Principles, Clinical Applications and Related Fields*. Ernst Niedermeyer, Fernando Lopes da Silva, Eds. Second Edition, Baltimore-Munich: Urban&Schwarzenberg, 1987, pp. 541-548.

[82] W. B. Martin, L. Johnson, S. Viglione, P. J. Naitoh and J. Moses, "Pattern recognition of ECG-EOG as a technique for all night sleep stage scoring," *Electroencephalgraphy and clinical neurophysiology*, vol. 32, pp. 417-427, 1972.

[83] M. Matousek and I. Petersen, "Automatic evaluation of EEG background activity by means of age-dependent EEG quotients,"

Electroencephalography and clinical neurophysiology, vol. 35, pp. 603-612, 1973

[84] W. S. Meisel, "*Computer-oriented approaches to pattern recognition*," New York: Academic, 1972.

[85] J. M. Mendel and K. S. Fu, "*Adaptive, learning and pattern recognition system*," New York: Academic, 1970.

[86] H. H. Morris, III and H. Luders, "Electrodes," in: *Long-Term Monitoring in Epilepsy*. J. Gotman, J. R. Ives and P. Gloor, Eds. *Electroencephalgraphy and clinical neurophysiology*, Suppl.37, pp. 3-26, 1985.

[87] A. M. Murro, D. W. King, J. R. Smith, B. B. Gallagher, H. F. Flanigin and K. Meador, "Computerized seizure detection of complex partial seizures," *Electroencephalgraphy and clinical neurophysiology*, vol. 79, pp. 330-333, 1991.

[88] A. M. Murro, Y. D. Park, D. W. King, B. B. Gallagher, J. R. Smith, K. J. Meador and W. Littelton, "Localization of Temporal Lobe Seizures with Quantitative EEG," *Electroencephalgraphy and clinical neurophysiology*, vol. 86, pp. 88-93, 1993.

[89] E. Niedermeyer, "The Normal EEG of the Waking Adult," in: *Electroencephalography, Basic Principles, Clinical Applications and Related Fields*. Ernst Niedermeyer, Fernando Lopes da Silva, Eds. Second Edition, Baltimore-Munich: Urban&Schwarzenberg, 1987a, pp. 97-118.

[90] E. Niedermeyer, "Abnormal EEG Patterns," in: *Electroencephalography, Basic Principles, Clinical Applications and Related Fields*. Ernst Niedermeyer, Fernando Lopes da Silva, Eds. Second Edition, Baltimore-Munich: Urban&Schwarzenberg, 1987b, pp. 183-208.

[91] E. Niedermeyer, "Epileptic Seizure Disorders," in: *Electroencephalography, Basic Principles, Clinical Applications and Related Fields*. Ernst Niedermeyer, Fernando Lopes da Silva, Eds. Second Edition, Baltimore-Munich: Urban&Schwarzenberg, 1987c, pp. 405-510.

[92] E. Niedermeyer, "The EEG in patients with migraine and other forms of headache," in: *Electroencephalography, Basic Principles, Clinical Applications and Related Fields*. Ernst Niedermeyer, Fernando Lopes da Silva, Eds. Second Edition, Baltimore-Munich: Urban&Schwarzenberg, 1987d, pp. 517-522.

[93] B. S. Oken and K. H. Chiappa, "Short-term variability in EEG frequency analysis," *Electroencephalography and clinical neurophysiology*, vol. 69, pp. 191-198, 1988.

[94] L. H. Oliver, R. S. Poulsen, G. T. Toussaint and C. Louis, "Classification of Atypical Cells in the Automatic Cytoscreening for Cervical Cancer," *Pattern recognition*, vol.11, pp. 205-212, 1979.

[95] A. Olivier, E. Marchand, T. Peter, and J. Tyler, "Depth Electrode Implantation at the Montreal Neurological Institute and Hospital," in: *Surgical Treatment of the Epilepsy*, Engel, J. Jr., editor, New York: Raven, 1987.

[96] F. Pauri, F. Pierelli, G. E. Chatrian, and W. W. Erdly, "Long-term EEG-video-audio monitoring: computer detection of focal EEG seizure patterns," *Electroencephalgraphy and clinical neurophysiology*, vol. 82, pp. 1-9, 1992.

[97] J. C. Principle and J. R. Smith, "Microcomputer-based system for the detection and quantification of petit mal epilepsy," *Computer in biology and medicine*, vol 12, pp. 87-95, 1982.

[98] H. M. Praetorius, G. Bodenstein, and O. D. Cruetzfeldt, "Adaptive segmentation of EEG records: A new approach to automatic EEG analysis," *Electroencephalgraphy and clinical neurophysiology*, vol. 42, pp. 84-94, 1977.

[99] E. L. Reilly, "EEG Recording and Operation of the Apparatus," in: *Electroencephalography, Basic Principles, Clinical Applications and Related Fields.* Ernst Niedermeyer, Fernando Lopes da Silva, Eds. Second Edition, Baltimore-Munich: Urban&Schwarzenberg, 1987, pp. 57-78.

[100] Y. Reich, C. W. Thomas, Y. H. Pao, J. Liebman, and Y. Rudy, "Multicategory classification of body surface potential maps," *IEEE transaction on biomedical engineering*, vol. 37, No. 12, pp. 945-955, 1990.

[101] R. W. Sencaj, J. I. Aunon, and C. D. McGillem, "Discrimination among visual stimuli by classification of their single evoked potentials," *Medical and biological engineering and computing*, vol. 17, pp. 391-396, 1979.

[102] K. S. Shanmugan and A. M. Breipohl, "*Random Signals: Detection, Estimation, and Data Analysis*," New York: Wiley, 1988.

[103] F. W. Sharbrough, "Electrical fields and recording techiniques," in: *Current practice of clinic electroencephalography*, D. D. Daly, and T.A. Pedley, Eds. Second edition, New York: Raven, 1990, pp. 29-49.

[104] F. W. Sharbrough, "Scalp-recorded ictal patterns in focal epilepsy," *Journal of Clinical Neurophysiology*, vol. 10(3), pp. 262-267, 1993

[105] B. Sklar, J. Hanley, and W. W. Simmons, "A computer analysis of EEG spectral signature from normal and dyslexic children," *IEEE transaction on biomedical engineering*, vol. BME-20, pp. 20-26, January, 1973.

[106] J. G. Small, "Psychiatric disorders and EEG," in: *Electroencephalography, Basic Principles, Clinical Applications and Related Fields.* Ernst

Niedermeyer, Fernando Lopes da Silva, Eds. Second Edition, Baltimore-Munich: Urban&Schwarzenberg, 1987, pp. 523-540.

[107] F. K. Soong, A. E. Rosenberg, L. R. Rabiner, and B. H. Juang, "A vector quantization approach to speaker recognition," *Proc. ICASSP*, pp. 387-390, 1985.

[108] K. Squires and E. Donchin, "Beyond averaging: the use of discriminant functions to recognize event related potentials elicited by single auditory stimuli," *Electroencephalgraphy and clinical neurophysiology*, vol. 41, pp. 449-459, 1976.

[109] C. A. Swenne, J. H. Van Bemmel, S. J. Hengeveld, and M. Hermans, "Pattern Recognition for ECG-monitoring: An Interactive Method for the Classification of Vertricular Complexes." *Computers and biomedical research*, vol. 5, pp. 150-160, 1973.

[110] G. T. Toussaint, "Bibliography on estimation of misclassification," *IEEE transaction on information theory*, vol.20, No. 4, pp. 472-479, 1974.

[111] O. Urbach and H. Pratt, "Application of finite impulse response digital filters to auditory brain-stem evoked potentials," *Electroencephalography and clinical neurophysiology*, vol. 64, pp. 269-273, 1986.

[112] S. S. Viglione, "Application of pattern recognition technology," in: *Adaptive, learning and pattern recognition*, J. M. Mendel and K. S. Fu, Eds. New York: Academic, 1970, pp. 115-161.

[113] D. O. Walter, "Spectral analysis for electroencephalograms: Mathematical determination of neurophysiological relationships from records of limited duration," *Experimental neurology*, vol. 8, pp. 155-181, 1963.

[114] D. O. Walter, J. M. Rhodes, B. S. Brown, and W. R. Adey, "Comprehensive spectral analysis of human EEG generators in posterior

cerebral regions," *Electroencephalgraphy and clinical neurophysiology*, vol. 20, pp. 224-331, 1966.

[115] D. O. Walter, J. M. Rhodes, and W. R. Adey, "Discriminating among states of consciousness by EEG measurements; A study of four subjects," *Electroencephalgraphy and clinical neurophysiology*, vol. 22, pp. 22-29, 1967.

[116] T. P. Yunck and F. B. Tuteur, "Comparison of decision rules for automatic EEG classification," *IEEE transactions on pattern analysis and machine intelligence*, vol. PAMI-2, No.5, pp. 420-428, 1980.

[117] L. H. Zetterberg, "Estimation of Parameters for a Linear Difference Equation with Application to EEG Analysis," *Mathematical Biosciences*, vol. 5, pp. 227-275, 1969.

[118] L. Q. Zhang, R. Shiavi, M. Hunt, and J. J. Chen, "Clustering analysis and pattern discrimination of EMG linear envelopes," *IEEE transactions on biomedical engineering*, vol. 38, No.8, pp. 777-784, August, 1991.