151233

EEG SPIKE DETECTION USING TEMPLATE MATCHING ALGORITHM

A Thesis Submitted to the
Graduate School of Natural Applied Sciences of
Dokuz Eylül University
In Partial Fulfillment of the Requirements for
The Degree of Master of Science in Electrical Electronic Engineering

151233

by
Selim BÖLGEN

January, 2004
İZMİR

Ms.Sc. THESIS EXAMINATION RESULT FORM

We certify that we have read the thesis, entitled **"EEG SPIKE DETECTION USING TEMPLATE MATCHING ALGORITHM"** completed by **SELİM BÖLGEN** under supervision of **ASSIST. PROF.DR. GÜLDEN KÖKTÜRK** and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Yrd.Doç.Dr. Gülden Köktürk

Assist.Prof.Dr. Gülden KÖKTÜRK
Supervisor

Yrd.Doç.Dr. Mustafa A. Altınkaya

Committee Member

Yrd.Doç.Dr. Mehmet Kuntalp

Committee Member

Approved by the
Graduate School of Natural and Applied Sciences

Prof.Dr.Cahit Helvacı

ACKNOWLEDGMENTS

I would like to thanks to Prof.Dr. Barış BAKLAN for their supports about the improving and completing the thesis.

I would also like to thanks to Assist. Prof. Dr. Mehmet KUNTALP for his directions.

I want to express my gratefulness to my family for their great help.

Selim BÖLGEN

ABSTRACT

In this thesis, a program for EEG spike detection has been developed. This program is based on Template Matching Algorithm to detect EEG spikes. EEG signals with 19 channels 256 Hz sampling frequency and band-pass filtered between 1 and 70 Hz are used to test the programming performance. These EEG samples have been obtained from the Neurology Department of Dokuz Eylül University Hospital, İzmir, Turkey.

In the first stage 4 different templates were constituted from the EEG spikes. These templates are used by the EEG spike detection program to find spike/spikes in EEG signals using the Template Matching Algorithm. Furthermore a threshold value is assigned to make a decision. It is marked as spike if the algorithm produces a value less than threshold.

In addition to the program, other related methods and works have been studied. These are Statistical Classification, Syntactic and Structural Method, Neural Network, Wavelet Analysis, Deconvolution Methods.

The developed program was compared with other related works according to the sensitivity, specificity, selectivity and average detection rate properties.

Keywords: EEG, spike detection, Template Matching

ÖZET

Bu çalışmada EEG dikenlerinin bulunması için bir program geliştirilmiştir. EEG dikenlerinin bulunması için geliştirilecek programda Template Matching algoritması baz alınmıştır. Programın testi için 256 Hz örnekleme frekanslı, 1 -70 Hz band geçiren filtreleri 19 kanallı EEG işaretleri kullanılmıştır. Bu işaretler Dokuz Eylül Üniversitesi Uygulama ve Araştırma Hastahanesi İzmir, Turkiye 'den alınmıştır.

İlk aşamada EEG işaretlerinden 4 farklı diken oluştrulmuştur. Bu dikenler Template Matching Method'undaki eşleme özelliğini kullanarak EEG işaretlerindeki dikenleri bulunması için kullanılır. Ayrıca hangi işaretin diken olup olmadığına karar vermek için bir eşik değeri atanır. Eşik değerinin altında kalan işaretler spike olarak işaretlenir.

Bu programa ek olarak bu konuyla ilgili diğer metodlar ve çalışmalarda teorik olarak incelenmiştir. Bunlar, Statistical Classification, Syntactic and Structural Method, Neural Network, Wavelet Analysis, Deconvolution Method.

Geliştirilen program diğer çalışmalarla sensitivity, selectivity, specivity, ve average detection rates kriterlerine göre karşılaştırılmıştır.

Anahtar Sözcükler: EEG, Diken Bulma, Template Matching

CONTENTS

Chapter Three

TEMPLATE MATCHING METHOD

Chapter Four

OTHER PATTERN RECOGNITION TOOLS

Chapter Five
RELATED WORKS

Chapter Six
IMPLEMENTING TEMPLATE MATCHING ALGORITHM

Chapter Seven
REFERENCES

Appendix

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE
INTRODUCTION

1.1 Objective of the Study

Since the firsts recordings in humans performed in 1929, the EEG has become one of the most important diognastic tools in clinical neurophysiology, but up to now, EEG analysis still relies mostly on its visual inspection. Due to the fact that visual inspection is very subjective and hardly allows any statistical analysis or standardization, several methods were proposed in order to quantify the information of the EEG.

The objective of this study is to develop a program that will help neurologists to decide the spike waves by analysing the mental problems in the the EEG records. The mental disorder taken into consideration in the study is epilepsi. Spike waves of EEG signals have been analysed with Template Matching Algorithm. Template Matching Method may be defined as the pattern to be recognized is compared with a learned template allowing changes in scale. The detailed information abouth the methods used is given in the following chapter.

It has been studied to develop a program for the EEG spike detection using the Template Matching Method. Several methods have been developed previously, but up to now any of them has not certain solution for the spike detection. Spike waves have 20 -70 msec duration. Spikes can be located any place on the brain so each location creates spikes with differents shapes. These spikes have also same duration, shape is different. However muscle activities can be look like a spike. These activities calles “artifacts”. Because of these diffuculties each solution works with a performance.

The first contribution of this thesis research is a detailed clinical study, including patients from two different kinds of signal. The analysis has been repeated for several sweeps and channels of each signal and sensitivity selectivitiy specificity and avarage detection rate performance criterions are taken to discuss results.

In addition to Template Matching Algorithm, other basic pattern recognition tools which are statistical classification, syntactic and structural matching, neural network, wavelet analysis, deconvolution method, and related works have been studied to be able to compare with other solutions performance.

1.2 Outline of the Study

The study includes 6 chapters. Chapter 1 introduces the subject and explains the objective. Chapter 2 includes the EEG signals that should be given for a better understanding of the spike waves and other waves of the EEG. History of EEG, device, montage and conventions, electrode location, descriptors of EEG activity, abnormal EEG pattern are the subheadings of this chapter.

Chapter 3 gives detailed information abouth the methods that are used for analysis. There are two way for the method that are explained theorically.

Chapter 4 gives the information for the other pattern recognition tools. These are explained to understand the other ways which can be used for the spike detection. Statistical methods, syntactic and structural matching, neural network, wavelet analysis and deconvolution method are subheading of this chapter.

Chapter 5 gives the related works that are literature searchs for spike detection. They were developed a program to analysis the EEG Signals.

Chapter 6 gives the results of the application of the method to EEG records are given. The analysis includes the records of two patients with epilepsy.

Performance criterions are given for the sake of reliable discussion of results. The results are discussed in the discussions section.

There are two appendices added to the study. The source codes of the developed program for the EEG spike detection are given in Appendix I, and some outputs of the programs are added to Appendix II.

CHAPTER TWO
EEG SIGNALS

EEG is the abbreviation for electroencephalography. The aim of electroencephalography is to record and measure samples of electromagnetic fields during certain states and sequences of behaviour, in order to explain some of the mechanisms by which behaviour is generated. The electroencephalographic (EEG) signals represent electrical changes of the brain during its function. The digital signal processing of EEG, especially the spectral analysis, yields more meaningful information than the visual inspection of the EEG curve. However, there are still problems encountered like the artefacts during the registration, the individual differences among normal and among diseased brains, and the uncontrolled conditions of brain functions.

Modern advances in EEG have included what is referred to as digital EEG or dEEG. Here brain signals are similarly collected from the scalp and amplified but are fed into a computer (i.e., digitized) and then interpreted by viewing them not on paper but on the computer screen. Important advantages include storage of efficient digital media rather than on bulky paper. Another advantage is the ability to view the same EEG signals from different perspectives - paper affords only one view of a time period. A draw-back is that the computer screen may not afford the same clarity of image that is available on paper. Another advance is the speedier placement of electrodes by using an elastic cap with electrodes already imbedded. Careless use of this technology may result in improperly positioned electrode or poor electrode contact

Small, non-invasive electrodes (usually 16 to 32 in number) are placed upon a patient's scalp, after careful measurement by a trained technologist, with paste or glue like substance to hold them in place. Low voltage signals (5-500 micro volts) are amplified by the EEG machine and results are typically written by ink-fed pens on a moving paper strip chart. The resulting polygraphic strip chart, looking much like a multiple channel seismograph, is typically read by unaided visual inspection. The physician interpreting such a tracing is usually a neurologist with special training in EEG. Such an individual is often referred to as a neurophysiologist or electroencephalographer. Psychiatrists, neurosurgeons, and psychologists may also interpret EEGs but to do so, like neurologists, they require special EEG training. [http://www.ucdmc.ucdavis.edu/neurology/Patients/eeginfo.htm]

Typically the EEG is screened for features that stand out (transient responses) like the spike or spike and wave associated with epilepsy. Next the frequency or spectral content of the remaining EEG background is visually evaluated. There are four broad spectral band of clinical interest: delta (0-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), and beta (above 12 Hz). Not everyone agrees on the exact boundaries of these rhythms and many subdivide these bands, especially beta. Pathology typically increases slow activity (delta, theta) and diminishes fast activity (alpha, beta). Thus overlying a localized brain tumour one would expect increased slowing and decreased fast activity. Similarly following a global brain insult resulting in a global encephalopathy one might expect globally increased slowing and decreased fact activity. However, there are many exceptions to this oversimplified explanation. EEG interpretation requires considerable skill and often years of clinical experience. The mere determination of whether an EEG spectral band is normal, increased, or decreased may require years of experience

The electroencephalogram (EEG) is a complex, time-varying signal which requires sophisticated analysis techniques if clinically useful information is to be reliably extracted from it.

Rhythmic Activity	Normal Frequency	Where it can be found	Reactivity	Patho physiologies
Alpha	8 to 13 Hz Amplitude Typically 20 to 60 μV	The most prominent rhytm in the normal adult brain. Most prominent and occipital and parietal electrodes. Abouth 25% stronger over the right hemisphere	Fully present when a subject is mentally inactive, alert with eyes with closed. Distrupted by visual attentiveness. Almost totaly disappears when eyes oppened.	Slowing is considered a non-specific abnormality in metabolic, toxic and infectious conditions. Asymetries unilateral lesions. Lost of reactivity of a lesion in the temporal lobe. Loss of alpha brainstem lesion
Mu	7 to 11 Hz	Central electrodes	Does not react opening of eyes, show blocking before movement of contralateral hand	
Beta	18 to 30 Hz	Three basic types • Frontal Beta • Widespread • Posterior	• Frontal beta blocking by movement • Widespreadbeta often unreactive • Posterior beta shows reactivity to eye opening	Not clinically useful
Theta	4 to 7 Hz	In drowsy normal adult, in frontal and temporal regions.	Rare in EEG of awake adults local or lateralized. Theta indicates focal pathology, diffuse theta more generalized neurological syndrome.	
Delta	Less than 4 Hz	Dominant in infants and deep stages of adult sleep	Focal in pathologies	Polimorphic delta severe, accute, or ongoing injury to cortical neurons. Rhytmic discharge psycho physiologic dysfunction

Table 2.1 EEG Waves Properties

2.1 History of EEG

The history of electroencephalography began in 1874 when Caton an Englishman discovered evidences of electrical activity in the brains of living animals. He observed (1875) electrical fluctuations from the cortex of rabbits and monkeys, and he was convinced that these fluctuations were related to the functional activity of the brain.

Beck, in 1890, showed that the visual cortex of the dog produced large potential fluctuations when the eyes were illuminated, and that in the absence of stimulation smaller electrical fluctuations occurred which were not related to pulse or respiration. After the publication of Beck's paper the Vienna Academy opened a communication which had been deposited in 1883 by Fleischl von Marxow with instructions that it be kept sealed until a report appeared dealing with the electrical activity of the cortex. He probably was unwilling to risk his reputation on findings about which he was uncertain, yet he wished to get credit for his own work by adding corroborative data to evidence which some later worker might obtain. He said (1890, 1893) that he had recorded large potentials from the visual area when the animal's eyes were illuminated; that these potentials, which were abolished by chloroform and by cooling, could be obtained through the dura and even through the skull.

Gotch and Horsly (1892) reported on the electrical activity of the cortex as part of their study of the localization of function in the central nervous system. They believed that special cortical areas gave an electrical response to certain types of peripheral stimulation. The visual area responded to illumination of the eyes, and the cutaneous sensory area to stimulation of a sensory nerve. In 1892, Beck published a report with CYbulski showing that local injury to the cortex modifies its electrical activity and that stimulation of the leg of a dog produces a response in the contralateral cruciate area. Danilewsky (1891), Larionow (1898), and Trivus (1900) studied electrical response of the cortex to various types of peripheral stimuli. In 1904, Tchiriev concluded that fluctaations in potential recorded from the cortex are the result of pulsation of blood in the blood vessels, and have no direct bearing on nervous activity. In 1912, however, Kaufmann reaffirmed the conclusion of earliar workers that the electrical potentials from the cortex are modified by sensory stimulation, and are related to nervous function.

The Einthoven string galvanometer, a tremendous improvement over previous types of recording instruments, became generally available in 1906, and was first used for studies on the brain by Neminski He reported in 1903 that cortical potentials can be demonstrated when the sciatic nerve of a dog is stimulated. Cybulski and Macieszyna (1919) repeated with a string galvanometer the work done in collaboration with Beck (1892). Their results confirmed the earlier findings. In 1925, Neminski published a report on the electrical activity on the dog's brain in which he described the electrical activity of the cortexin much the same terms as those used today. He said that the *electrocerebrogram* (a word which he coined) consisted chiefly of spontaneous fluactuations in potential with a frequency of 10 of 15 per second, called *waves of the first order* and other faster fluctations with a frequency of 20 to 32 per second, called *waves of the second order*. He was able to obtain the *electrocrebrogram* from the cortex, the dura, or the outer surface of the skull.

Although these studies from 1874 onward were. with but few exceptions, in general accord, they had little or no effect on the main streams of contemporary research. The possibility of studying directly an electrical component of brain activity was overlooked by the leaders of neurology and neurophysiology for forty years (1893 to 1933). Hans Berger, however, did not overlook this possibility. In 1902 and in 1907 he recorded spontaneous fluctuations in potential from the brains of animals, but on neither occasion was he able to show that these fluctations were modified by sensory stimulation. In 1910, he tried once more, this time with a string galvanometer, but again with negative results as far as response to stimulation was concerned. In 1924, he attempted successfully to record the electrical activity of the *human* brain. The firs reports of these results, published in 1929, were greeted with incredulity. He continued his work, however, and by 1934 he had demonstrated that the brain of the man has an electrical beat; that this beat comes from neurons, not from blood vessels or connective tissue; and that it changes with age, with sensory stimulation, and with various changes in the physicochemical state of the body. He showed that normally this beat appears as a mixture of more or less sinusoidal fluctuations in voltage with a frequency of from one to sixty per second, and that the most easily discernible rhythm has a frequency of approximately ten per second. Waves of this frequency he called *alpha* waves, and demonstrated that they tend to disappear with attention. Faster waves, those having a frequency of 15 to 60 per second, he called *beta* waves. Because of the analogy between the electrical pulsations from the brain and the electrocardiogram, and because he disapproved of the mixture of Greek and Latin in Neminski's word *electrocerebrogram*, Berger called his record the elektreenkephalogramm, usually translated into English as *electroencephalogram.*

Almost all of Berger's observations have been reaffirmed by later workers, but such proof was not required, for his own work was well controlled and he furnished valid proof of his thesis that the electroencephalogram comes from the cortex and is related to psychic activity. Because he was a phychiatrist, he published most of his articles in psychiatric journals (1929 - 38), and tried to correlate his data with various psychiatric concepts. His titles were not especially informative; many articles were headed "Uebar das Elektreenkephalogramm des Menschen" and each contained a variety of data and much discussion. In 1938 he published a summary of hid observations and interpretations in monograph form.

If Berger had divided his work into discrete short studies, with satisfactory descriptive titles, and had avoided the psychiatric implications of his data, he might have been accepted more readily as a great *neurophysiologist*, but his actual achievement would have been less, for he could not have surveyed so completely the entire field of electroencephalography. His discovery of the correlation between the electrical activity of the cortex and psychic functions was truly revolutionary.

The spread of electroencephalography over the world was at first hampered by its unexpected place of origin and by the fact that electrophysiologist had been making rapid advances in exactly the opposite direction, i.e., by recording brief discharges in single fibers of peripheral nerves. To the confirmed "axonologist" the idea of recording from neurone aggregates was repugnant. But when Adrian, a Nobel Prize winner in electrophysiology, said that Berger was right and apologized for his own prolong doubts, electroencephalography became respectable.

Before this, However, Fischer (1933) had reported that the brains of animals when convulsed with drugs develop high voltage "Krampfströme." This crucial observations caught the attention of the Harvard Group and set them to recording the electroencephalogram in epilepsy (Gibbs, Davis and Lennox, 1935), a project for which Derbyshire had cleared the way (Davis, P. and Davis, H., 1940) by demonstrating to local sceptics that Berger's observations could be duplicated with apparatus primarily developed by studies on hearing. Grey Walter (1936) scored a great advance by showing that brain tumors can be localized by electroencephalographs through the unopened skull.

2.2 EEG Device

Although all new EEG devices are digital, analog EEGs are still used for clinical studies (Brenner and Scheuer, 1998). These devices did not differ radically from the Berger's EEG. An analog EEG consists of a differential pre-amplifier, an adjustable post-amplifier, a filter array and a plotter. A digital EEG has a similar pre and post-amplifier, but the amplified signal is converted to a digital sample. These samples are recorded onto a digital media, from where they can either be plotted or processed further. Digital EEGs sample the signal at a pre-set rate and accuracy. Nyquist's rule dictates that the signal must be sampled at twice the frequency of the fastest component. Although component over 100 Hz are not detectable from the scalp, commercial EEG devices over-sample the signal. Common sampling rates are between 500 and 1000 Hz. The DC component of the signal is removed with a highpass filter. A prevalent cut-off frequency for this filter is 0.1 Hz. The sampling accuracy of commercial devices is usually 12 or 16 bits per sample. EEG data is recorded from multiple electrode locations (channels), simultaneously. The number of channels is usually a power of 2. Clinical recordings use up to 64 channel. Experimental studies have been conducted with 256 channels. Measurements commonly include non-EEG channels that record information about muscle activity and eye movement. EEG equipment is inexpensive compared to other research instruments, for example positron emission tomography (PET), single photon emission computed tomography (SPECT), magneto encephalography (MEG) and functional magnetic resonance imaging (fMRI). EEG has a better temporal resolution than PET, SPECT or fMRI. PET and SPECT imaging require the use of radioactive isotopes, which are expensive and require special care in administering. MEG devices have the same spatial and temporal capabilities as EEG. MEG devices require shielded rooms and helium cooling. MEG devices are excruciatingly expensive. Skull and scalp resistively effect the EEG waveform, while they are almost invisible to the MEG. Research is being conducted on minimising these effects in the EEG (Babiloni et al., 1997).

2.3 Montages and conventions

The terminology and conventions of modern electroencephalography is plagued with remnants of old analog technology. A good example would be Peter Manu's article (1994) were he uses the term cycles per second instead of Hz. Before the advent of the transistor, EEG amplifiers used vacuum tubes. It was customary to connect the active electrodes to grid 1 and the ground or indifferent electrode to grid 2. The vacuum tube functioned as a differential amplifier. The amplifier was connected to a pen chart in a fashion that when the active electrode was more electronegative than the indifferent electrode, the pen would make an upward deflection. The illogical reverse convention has survived to this day. The term indifferent electrode is also a misnomer since the indifferent electrode has an equal effect on the amplifier. The electrode leads have an established colour scheme.

Indifferent electrodes are white and active electrodes are black or coloured (Kiloh, McComas and Osselton, 1972). Since the pre-amplifier of the EEG requires two inputs, EEG measurements can not be made from a single point. The configuration of electrodes and the manner in which they are connected to the amplifier is defined in the montage. When potential is measured between sets of 2 electrodes the montage is called bipolar. The channel name reveals both electrodes, for instance a measurement with the positive electrode at Fz and the negative at Cz would be called Fz-Cz (Jasper, 1958). Often a common reference is chosen for all of the channels. This sort of measurement is called unipolar. The reference electrode is the indifferent input to all of the pre-amplifiers and it is often designated with a white lead. The location of the reference electrode should be at a place that is far away from all of the active electrodes. Frequently an electrode clip is attached to both earlobes and the electrodes are connected to each other. This forms a virtual electrode that is equally sensitive to both hemispheres. This reference is appropriately named linked earlobes. Unfortunately the linked ear lobes change the current flow and potential distribution of the head (Regan, 1989). The chin, neck and mastoids are also common locations for the reference. The reference electrode should not pickup excessive cerebral activity, muscle artefacts or ECG waveform.

The channels in the unipolar montage are named using only the active electrode. When the measurement scheme is explained the location of the reference electrode is mentioned separately. An average reference can be formed by combing all of the active electrodes to a single point. Large resistors are placed between the active electrodes and the reference. All of the pre-amplifiers use this average reference as their negative input. This configuration is often called the Goldman-Offner montage after its inventors. This system assumes that the potentials will cancel each other out and form a stable reference. Global activities, especially rhythmic activities, affect a large amount of electrodes causing the reference potential to pickup this signal. This causes some of the channels to produce a misleadingly small signal when the reference and active potentials cancel each other out. Nevertheless, average references produce good results with spatially localised events. (Kiloh, McComas and Osselton, 1972). Although all of the EEG channels do not have to use the same configuration, mixed montages are rare. Most studies are polygraphic, meaning non-cerebral measurements are taken along with the EEG. Muscular and ocular measurements are often bipolar regardless of the EEG montage.

2.4 Electrode Location

The location of the electrodes should be tailored to suit the study. Electrode locations can either be equidistant or relative. In order to ensure reproducibility of studies and to ease exchanging of results, electrode location sets have been standardised.

An equidistant grid is convenient for the spatial study of a local event or activity in the brain, but it is not reproducible for subjects with different head sizes The queen square system has equidistant electrodes covering the parietal and occipital areas (Regan, 1989). This system has not found widespread acceptance. The International Federation of Societies for Electroencephalography and Clinical Neurophysiology recommended a system that relies on the relative distances between fixed landmarks on the head (Jasper, 1958). In this system, the electrode placement remains constant for different head sizes. The relative distances are 10 % and 20 % of the ear to ear distance and the inion to nasion distance. Therefore this system is called the international 10-20 system and it pinpoints the location of 21 electrodes. The electrodes are arranged in row according to the lobe in which they reside. The centre electrode is marked with a z, for instance the middle frontal electrode is Fz. The other electrodes are numbered from the middle to sides, so that all of the odd numbers are on the left side and all of the even numbers are on the right side. The electrodes in the temporal lobe are placed in an anterior-posterior direction and do not have a centre electrode. The original 10-20 system skipped numbers so that electrode could be placed in between the existing numbers. The American EEG Society devised a guideline for the correct placement of these intermediate electrodes. This combinatorial nomenclature places electrodes between both the anterior posterior rows and the medial-lateral rows. This system also has an electrode on the nasion and inion producing a total of 75 electrode places (Niedermeyer and Lopes Da Silva, 1993)

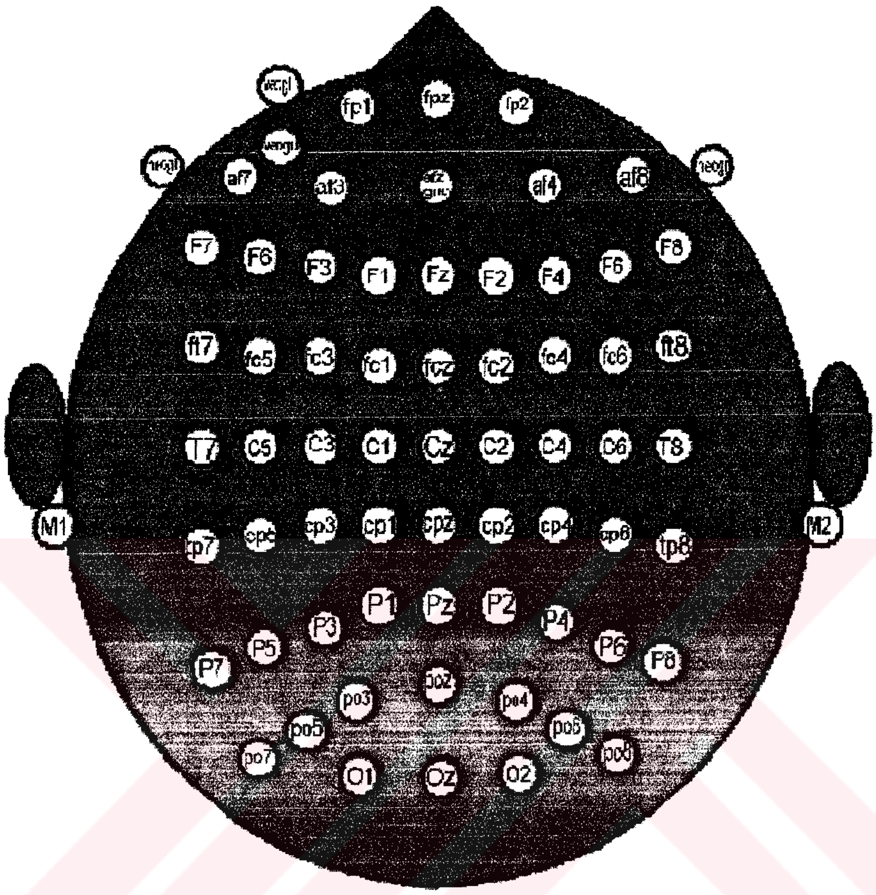


Figure 2.1 , The combinatorial nomenclature. (1999. 64 Channel Electrocap Layout. NeuroScan Labs. Virgina, Neurosoft Inc. [http://neurocog.psy.tufts.edu/images/ERP_components.gif]

]

2.5 Descriptors of EEG activity

2.5.1 Wave Forms

Wave form or *shape* are simple term used to describe the *configuration* or *morphology* of a wave. Any change in the difference of the electrical potential between two recording electrodes is called a *wave,* regardless of its form. Any wave or sequence of waves is called *activity.* Many waves are regular , i.e they have a fairly uniform appearance due to symmetrical rising and falling phases (Fig 2.2, Part 1). Some regular waves are similar to sine waves and a re called *sinusoidal* (Fig 2.2, Part 2). While other regular waves may be arch-shaped or saw-toothed. *Irregular waves* have uneven shapes and durations (Fig 2.2, Part 4).

A *monophasic* wave is a single deflection either up or down from the baseline. A *diphasic* wave has two components on oppisite sides of the baseline while a *triphasic* wave has three components of different direction. These terms do not indicate whether a wave has positive or negative electrical polarity nor whether it was recorded with bipolar or referential electrode montages .

A *transient* is an event which clearly stands out against the background. It consist of either a single wave or a complex, i.e. a sequence of two or more waves which have characteristic form or recur with a fairly consistent shape (Fig 2.2, Part 5).

A *sharp transient* is a wave form of any duration which has a pointed peak at conventional EEG recording speed. Sharply contoured waveforms which are not abnormal epileptic waveforms are often referred to sharp transient. *Epileptiform* is a term used to describe EEG patterns that are identical to those that have been specifically associated with seizures or epilepsy. Epileptiform patterns usually consist of apiculate wave forms referred to as *spikes* or *sharpe waves.* A spike is a sharply contoured wave form with a duration of 20- 70 msec (Fig 2.2, Part 7). A sharp wave has a duration of 70 – 200 msec and may not be as sharply contoured waveforms that: (1) appear as part of the background rhythm (e.g. , mu rhythm); (2) appear at different times either in isolation or as part of the background rhythm (e.g., wicket spikes; (3) demonstrate a varying morphology; or (4) only occur once in the entire record are often referred to as sharp transients because they have less significance in the diagnosis of seizure disorders than do stereotyped spikes or sharp waves.

A spike may be followed by a slow wave and form a *spike and wave complex* (Fig 2.2, Part 8) which may repeat at regular intervals. Spike and wave complexes recurring at rates below 3 Hz are called slow spike and wave complexes. A sharp wave may be followed by a slow wave and form a *sharp and slow wave complex;* complexes of this kind usually last longer than a third of a second and therefore do not repeat at rates over 3 Hz. In some cases, two or more spikes occur in sequence, forming *multiple spike complexes* also called *polyspike complexes* (Fig 2.2, Part 9). These complexes may be followed by a slow wave and thus form part of a *multiple–spike-and- slow-wave complex or polyspike–and-slow–wave complex* (Fig 2.2, Part 10). Spikes recorded in the EEG should not be confused with action potentials of single nevre cells which are recorded through microelectrodes inserted into the brain and which last only abouth 1 msec (1.1.3); thay too are often called spikes but are never observed in the surface EEG.

Single spikes and sharp waves, and complexes which contain spikes and sharp waves and last for less than a few seconds are called *interictal epileptiform activity*; longer lasting activity of this type and of some other types is referred to as a *seizure* pattern ot *ictal pattern*. Although seizure patterns are often associated with clinical seizure manifestations, yhey may occur without such correlates and are then called *subclinical seizure patterns*. Both interictal and ictal patterns are here called *epileptiform* in contrast to the definition in the glossary of the International Federation of Socieities for Electroencephalography and Clinical Neurophysiology which uses the term "epileptiform" only for interictal patterns.

A paroxysm or a *paroxysmal discharge* consist of one or more waves which begin abruptly, reach maximum amplitud rapidly, and disappear suddenly. These waves clearly stand out against the background, are usually abnormal, anda re often seen in epileptiform patterns. Paroxysms often consist of complexes (Fig 2.2 Part 5), but not all complexes begin and end abruptly, and not all paroxysms recur with a similar shape.

It is important to note that although the terms spike(s), sharp wave(s), paroxysms and paroxysmal discharges are often used to describe epileptiform patterns, they are not synonmous with epileptiform activity. Thus, if epileptiform activity seen is considered to be present, then the term epileptiform must be added to any other descriptive terms used. For example, if the activity seen is considered to be epileptiform then in the interpretation of the EEG it would be incorrect to simply state: The EEG is abnormal due to the presence of spikes localized to the left anterior temporal lobe. Instead it should be rephrased as: The EEG is abnormal due to the presence of epileptiform spikes localized to the left temporal lobe.

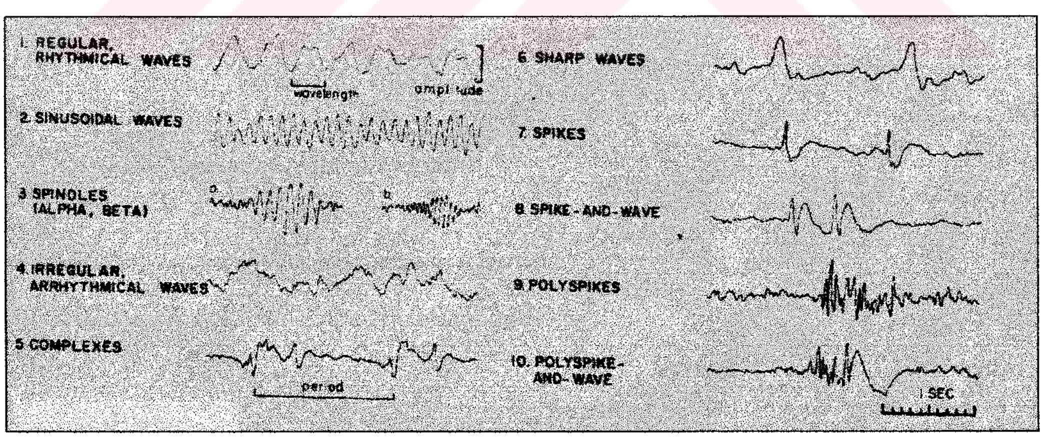


Fig 2.2 Characteristic wave forms

2.5.2 Repetetion

Repetition of waves may be rhythmical or arrhythmical. Rhythmical repetitive waves have similar intervals between individual waves; they are usually regular and often sinusoidal in shape (Fig 2.2, Parts 1-3). Spindles are groups of rhythmical repetitive waves which gradually increase and then decrease in amplitude (Fig. 2.2, Part 3). Rhythmical repetitive waves were formerly called "monorhythmic" or "monomorphic". Arrhythmical repetitive waves are characterized by variable, irregular intervals between individual waves (Fig 1.2 Part 4). They can be considered to be a sequence of waves of different frequency. They often have irregular shape. Arrhtthmical irregular waves were formerly called "polyrhythmic" or "polymorphic".

2.5.3. Frequency

Frequency refers to the number of times a repetitive wave recurs in one second. A wave completing three cycles in one second is called a wave of 3 Hertz (Hz) or of 3 per sec. The frequency of single or repetitive waves can be determined by measuring the duration of an individual wave, the *wavelenght* (Fig 2.2 Part 1), and calculating the reciprocal. Single waves and complexes may be repeat at intervals longer than the wavelength anda re then called "periodic", the period being the time interval between them (Fig 2.2 Part 5).

The frequency of EEG waves is often divided into four groups or *frequency bands,* namely:

- Delta frequency band: under 4 Hz
- Theta frequency band: from 4 to under 8 Hz
- Alpha frequency band: from 8 to 13 Hz
- Beta frequency band: over 13 Hz (Fig. 9. 2).

These divisions are somewhat arbitrary; many EEGs contain waves of frequencies extending across the boundaries of the bands, for instance waves of 3 -5 Hz. Nonetheless, the frequency bands help to set apart the most important criterion for asssesing abnormality in clinical EEG. Althoug waves under 8 Hz are commaly called *slow waves* and waves over 13 Hz are commonly called *fast waves*, it is more accurate and therefore preferable to either state the frequency of the activity(e.g., 3-5 Hz) or describe it according to the frequency band(s) it occupies (e.g., delta and theta activity). Activity that is less than ½ Hz or more than 20 Hz is of limited clinical utility in routine scalp recordings because it is often unclear if such activity is of cerebral origin.

2.5.4 Amplitude

Amplitude of EEG waves is measured in microvolts (μV). It is determined by measuring and comparing the total vertical distance of a wave (Fig 2.2, Part 1) to the height of a calibration signal recorded at the same gain and filter settings. Thus, if the height of an EEG wave measures 14 mm and a calibration signal of 50 μV measures 7 mm, the amplitude of the wave is 100 μV . If the sensitivity of the amplifier is known to be 7 μV */mm,* a wave of 7 mm can be inferred to have an amplitude of 50 μV without direct comparison with a caibration pulse. Amplitude should not be expressed in terms of the height of the pen deflection because this varies with the amplifier settings.

In clinical EEG, amplitude is often not reported in terms of microvolts but described loosely as low (under 20 μV) *medium or moderate* (20-50 μV), or *high* (over 50 μV). However, these terms often are used to describe amplitude of certain waves relative to that of other waves in the same record. For instance, a wave of 60 μV cannot be said to be of high amplitude if it occurs on a background of 40-50 μV .

An important abnormality is *asymmetry* of amplitude of the activity that is recorded simultaneously from corresponding parts of the two sides of the head.

Even slight differences of amplitude may be of clinical importance if they persist; this is true especially of the adult EEG, with the exception of the alpha rhythm.

Differences of amplitude are sometimes caused by factors outside the brain, especially by unequal spacing and impedance of the recording electrodes; the technician should therefore verify correct electrode placement and impedances before accepting that abnormal amplitude is genuine.

2.5.5 Distribution

Distribution refers to the occurence of electrical activity recorded by electrodes positioned over different parts of the head. EEG patterns may appear over large areas on both sides of the head, over one side only, or in a restricted, small area.

2.5.5.1 Widespread Diffuse or Generalized Distrubution

Widespread diffuse or generalized distrubution refers to activity which occours at the same time over most or all of the head. Generalized activity may have a clear maximum within its field of disribution, recognized by the highest amplitude in referential recordings and by phase reversal in bipolar recordings from that area.

2.5.5.2 Lateralized Distribution

Lateralized distribution refers to activity which appears only or mostly on one side of the head. Lateralized activity is abnormal and suggest a cerebral abnormality either on the side where abnormal activity is presen tor on the side where normal activity is absent. Some normal patterns may appear on one side of the head at one time and then occur on the other side a few seconds or minutes later.

2.5.5.3 Focal Activity

Focal activity is that which is restricted to one or few electrodes over an area of the head. Some of the neighboring electrodes may pick up the same activity with lower amplitude. This restricted distribution must be distunguished from a wide or generalized distribution which may have a maximum in one area. This distinctions is important especially with regard to abnormal slow waves and to sharp waves. Criterion which sometimes help in this distinctions are that *focal slow waves* often have a lower frequency at the area of maximum amplitude wheres generalized slow waves do not. *Focal sharp waves* with a tendency to spread can often be distunguished from generalized sharp waves with a local maximum of amplitude by their greater persistence at the focus.

Activity arising from a single unilateral focus is always abnormal. Activity from a midline focus or from two foci located symmetrically in the two hemispheres may be part of a normal pattern.

2.5.6 Phase Relation

Phase refers to timing and polarity of components of waves in one or more channels. Waves of different frequency may occur in different channels so that the troughs and peaks occur at the same time; these waves are said to be *in phase*. If they do not coincide in this manner, they are said to be *out of phase*. The phase difference may be expressed in terms of phase angle. For instance, peaks pointing in opposite directions are said to be 180 out of phase. Such a "*phase reversal*" is the major indicatorof the origin of EEG potentials in bipolar recordings. In a single channel, phase refers to the time relationship between different components of a rhythm; for instance, the peak of a sine wave is said to "lead" the preceding crossing of the zero line by 90 and to "lag" behind the next following peak by 360

2.5.7 Timing

Timing of waves in different areas of the head may be similar or different. The terms "simultaneous" and "synchronous" are used to indicate that two events occured at the same time. These terms are usually with the same meaning, but "synchronous" is sometimes used to denote precise coincidence while "simultaneous" may be used more broadly to indicate the coincidence that is recognizable only imprecisely within the limits of the relatively slow recording speed of the EEG. The eye can hardly distunguish a horizantal difference of less than 1mm between corresponding points on two waves even in neighboring channels. A horizantal distance of only 1 mm corresponds to a time difference of 33 msec at the conventional EEG recording speed. The resolution of time relations deteriorates if more distance channels are compared and if the writing units are not perfectly aligned; because of the curvilinear movement of the pens, synchronous excursions of different amplitude seem to have occured at different times.

Waves which occred at the same time on both sides of the head are called "*bilateraly synchrous*" or "*bisynchronous*". These terms consider mainly the relationship between the two sides but not necessarily that on the same side; thus bilaterally synchronous waves may be out of phase in the same hemisphere. In some instances, waves are delayed against each other by the same amount in succesive channels which record activity from electrodes placed from the front to the back of the head,giving the impression that these waves spread from front to back. For instance, this type of delay can be seen in triphasic waves of metabolic encephalopathies.

Waves which occur in different channels without constant time relation to each other called "asynchronous". This is usually implies that the waves are present in different areas at the same time even though they do not fall in phase with each or do not have the same frequency. If waves occur in one area at one time and in other areas at another time, they are usually said to be "*independent*" ; for instance spikes in both temporal lobes may occur bisynchronously or independently; each case has different implications regarding a possible triggering mechanism.

2.5.8 Persistence

Persistence describes how often a wave or pattern occurs during a recording. Some waves occur only occasionally or intermittently either in the form of a single wave or trains of waves; other waves are present through most or all of the recording. The persistence of waves can be estimated by measuring the proportion of time during which these waves appear. This is called the *index* For instance, a delta index of 20 % means that delta activity was present during one-fifth of a recording. Because the clinical importance of EEG paterns often depends not only on their persistence but also on their amplitude , the persistence and amplitude are often described together in terms of their *quantity, amount or prominence.* The term "abundance" previously used to describe this combination of persistence and amplitude, is now obsolote.

Single waves and complexes may occur with a high, moderate or low persistance or incidence; the persistence of these events is best expressed as their average number in one second or one minute. They may occur periodically or at irregular intervals. Irregular and infrequent occurence is sometimes called "*sporadic*". The terms "random" and "diffuse" should not be used to describe persistence of EEG patterns.

2.5.9 Reactivity

Reactivity refers to changes which can be prodeced in some normal abnormal patterns by various maneuvers. Some patterns are induced or increase, diminished or blocked by opening or closing the eyes, hyperventilation, photic or sensory stimuli, changes in levels of alertness, movements or other maneuvers. Abnormal slow waves in toxic and metabolic encephalopathies are often diminished by alerting and enhanced by hyperventilation and drowsiness whereas abnormal slow waves şeen in cases of structural lesions usually show a less attenuation or blocking during alerting maneuvers.

Thus, a recording should not be considered complete unless at least simple alerting maneuvers have been performed to demonstrate the effects of arousal on the EEG. These maneuvers include eye opening and closing (this may be passively performed for infants or other individuals who cannot respond to verbal commands) and questions testing memory and simple calculations. If the patient is unable to respond to verbal commands then vigorous auditory and tactile stimulation should be applied. These maneuvers will also help clarify if background slowing is actually presen tor if the patient was merely excessively drowsy during the recording.

2.6 Abnormal EEG Patterns

2.6.1 Definition of the Abnormal EEG

An EEG is abnormal if it contains A. Epileptiform activity, B. Slow waves, C. Amplitude abnormalities, or D. Certain patterns resembling normal activity but deviating from it in frequency, reactivity, distribution or other features. In most abnormal EEGs, the abnormal patterns do not entirely replace normal activity: they appear only intermittently, only in some channels, or only superimposed on a normal background

The most important EEG abnormalities can be divided into the following basic abnormal EEG patterns which are discussed in the subsequent chapters.

A. Epileptiform Activity

1. Localized epileptiform activity
2. Generalized epileptiform activity
3. Special epileptiform activity

B. Slow waves

1. Localized slow waves
2. Generalized asynchronous slow waves
3. Bilaterally synchronous slow waves

C. Amplitude abnormalities

1. Localized amplitud changes: Asymmetries
2. Generalized amplitude changes

D. Deviations from normal patterns.

2.6.2 Correlation Between Abnormal EEG Patterns, General Cerebral Pathalogy and Specific Neurological Diseases

Each of the basic abnormal EEG patterns listed above can be caused by one or a few types of cerebral abnormalities. The abnormalities are characterized by their irritative or destructive character and by their cortical, subcortical and epicortical location. The four major categories of abnormal EEG patterns are the subect of Table 1.2 – Basic Epileptiform patterns, 1.3 – Basic patterns of slow wave abnormalities, 1.4 – Basic patterns of abnormal amplitude and 25.1 – Deviations from normal patterns. The epileptiform patterns listed in Table 1.2; the patterns of abnormal amplitude listed in Table 1.4

2.6.2.1 Epileptiform Activity

Epileptiform activity is outlined in Table 1.2. Local epileptiform activity is usually due to a focal irritative lesion of the cerebral cortex; in infants, such activity may be the result of widespread lesions or of toxic, metabolic or electrolytic abnormalities whereas some children have local spikes without any detectable cerebral lesions. Generilized epileptiform activity is either not associated with demonstrable lesions or associated with a variety of conditions which increase the excitability of subcortical centers, of wide parts of the cerebral cortex, or of both. Special epileptiform patterns have a great variety of pathalogical correlates.

Basic Patterns	General Pathological Correlates	Examples of Specific Diseases
1. *Local epileptiform Activity*	(1) Chronic local cortical lesions	Cortical scars after strokes and injuries, tumours; with or without recurring partial seizures:symtomatic epilepsy
	(2) Acute local cortical lesions	Acute strokes, head injuries; with or without partial seizures
	(3) In young infants	
	(a) Widespread structural damage	Perinatal injury, anoxia, ischemia; with or without partial, uni – or bilateral seizures.
	(b) Toxic, metabolic, electrolytic abnormalities	Hypoglycemia, pyridoxine deficiency, pheny lketonuria; with or without seizures as above
	(4) Children without detectable lesion	Beningn epilepsy of childhood with partial seizures
2. *Generalized Epileptiform activity*	(1) No detectable abnormaliy	Idiopathic epilepsu with primary generalized seizures
	(2) Diffuse cortical and subcordical disordes:	
	(a) Structural	
	(aa) Acute damage	Acute anoxia, head injury, encaphalitis; with or without primary generalized seizures
	(bb) Chronic diseases	Postanoxic and postraumatic generalized cerebral damage, myoclonus epilepsy; with or without primary generalized seizures
	(b) Toxic, metabolic, endocrine, electrolytic disorders	Hypoglycemia, renal encephalopathy, alcohol with drawal; with or without primary generalized seizures during the disorder
3. *Special Epileptiform patterns*,		
3.1 Infantile and Juvenile patterns of multifocal and generalized spikes	Widespread structural or metabolic cerebral disease; patterns are more specific for age than for cause	Pre-, peri- and postnatal injury, cerebromaculer degeneration, tuberous sclerosis, phenylketonuria, leukodystrophies; with or without partial or generalized seizures
3.2 Periodic complexes	Acute or subacute, fairly widespread cerebral damage or metabolic derangements	Fresh cerebral infarcts, Jacob-Creutzfeldt disease, subacute sclerosing panencephalitis, barbiturate intoxication, herpes simplex encephalitis, metabolic encephalopathies; with or without myoclonus
3.3 Ictal patterns without spikes and sharp waves	No common pathological correlate	Certain partial complex seizures, tonic seizures, neonatanal seizures, absence seizures, epilpsia partialis continua
3.4 Epileptiform patterns with-out known pathological correlates and without seizures	No detectable abnormality	No known diseases or seizures

Table 2.2 Basic Epileptiform patterns: Pathological and clinical correlates

2.6.2.2 Slow Wave Abnormalities

Local slow waves are offen due to circumscribed damage of the white matter of the hemispheres with or without involvement of the cortex. Generalized asynchronous show waves suggest a widespread disturbance of cerebral function, often due to greater involvement of subcortical white matter than of the cerebral cortex. Bisynchronous slow waves are often due to widespread involvement of deep midline structures; this may be due to structural damage or to metabolic, toxic or endocrine disorders.

Basic Patterns	General Pathological Correlates	Examples of Specific Diseases
1 . Local Slow waves	(1) Local structural damage of	
	(a) Subcortical white matter	Strokes, tumors, abscesses
	(b) Thalamus	As above
	(2) Local disorders of cerebral blood flow or metabolism	Transient ischemic attacks, migraine, postictal condition
2. Generalized asynchronous slow waves	(1) No detectable abnormality in some cases of mild or moderate slow waves	No known disease, in 10 – 15 % of normal adults
	(2) Widespread structural damage including subcortical white matter	Widespread degenerative and cere-brovascular disease
	(3) Generalized disorders of cerebral function	Acute anoxia, syncope, coma postictal condition
3.Bilaterally synchronous slow waves	Deep midline grey matter involvemnt by	
	(1) Diffuse diseases damaging subcortical and cortical grey matter more than white matter	Presenile dementia, progressive supranuclear palsy
	(2) Local structural lesions which directly involve or compress, distort or render ischemic deep midline structures of the mesencephalon, diencephalon, mesial and orbital parts of frontal lobe	Tumours, strokes at or near the bottom of the anterior, middle or posterior fossa
	(3) Metabolic, toxic, and endocrine encephalopathies	Hepatic, renal, hypoparathyroid encephalopathies

Table 2.3 Basic Patterns of slow waves abnormalities: Pathological and clinical correlates

2.6.2.3 Amplitude Changes

Amplitude changes are described in Table 1.4. Local reductions of amplitude are often due either to superficial lesions which reduce the electrical potentials generated in the cortex or to material that is interposed between cortex and recording electrodes and interferes with the electrical conduction of cortical potentials to the recording electrodes; a local increase of amplitude often results from skull defects. Generalized reductions of amplitude are due either to a widespread decrease of the production of electrocortical potentials or to a generalized increase of the conducting media between cortex and recording electrodes.

Basic Patterns	General Pathological Correlates	Examples of Specific Diseases
1. Local differences of amplitude (asymmetries)	(1) Locally decreased EEG production	
	(a)Structural cortical damage	Cortical infarct, contusion
	(b) Disorder of cortical function	Cortical transient ischemia, migraine
	(2) Local change of media between cortex and recording electrode	
	(a) Increase	Subdural hematoma, subgaleal hematoma
	(b) Decrease	Surgical skull defect
2. Generalized changes of amplitude	(1) Generally decreased EEG production	
	(a) No detectable abnormality in some cases of mild or moderate reduction	No known disease, in 5 – 10 % of normal adults
	(b) Structural diseases of cerebral cortex	Huntington's chorea, postanoxic encephalopathy
	(c) Disorders of cortical function	Hypothyroidism, acute anoxia,hypothermia, intoxications, anxiety, postictal
	(2) Bilateral increase of media between cortex and recording electrodes	Subdural hematoma

Table 2.4 Basic patterns of abnormal amplitude: Pathological and clinical correlates

CHAPTER THREE

TEMPLATE MATCHING METHOD

3.1 Introduction

Template matching is a technique used to isolate certain features in an image. These features can be single pixels, lines, edges or complete objects. It is in a sense the same as filtering, only differing in the goal. Perhaps it is easier to look upon a image, convolved with a template window, as a correlation between an image and that window. The result will be an image with high values where there is a strong correlation (that is, where the template matches part of the image) and low values elsewhere.

In pattern recognition, two major issues are feature extraction and distance measure definition. Distance measure definitions have been used widely by the the Template Matching Algorithm.

The template matching technique relies on the use of a basis template that is compared to the signal. Using one of several transforms (like SSE, convolution), the basis template is used to create a measure of error (or of difference) agains the input signal.

Because template-matching functions use a basis template to transform the signal, the resulting measure of error is clearly sensitive to changes in the shape and nature of the template. However there is a reasonable amount of research to show that not just the shape of the template is important. The length of the template can have a significant effect on a template-matching algorithm's performance. There often exist an exponential relation ship between the length of the template and the number of false positives.

Template Matching is the term given to the process of detecting an event buried in a signal by comparing it to a predifened "template". The goal is to locate possible events in the signal that "closely resemble" the template. In practice there are two basic methods to determine "how close" any given section of a signal is to the template. They are Summed of Square Error (SSE), Convolution (CONV) and Correlation.

3.2 Summed Square Error Template Matching

We have some (small) image of a specific pattern or object we want to find. We will use the term template. A measure of match between a template t, and a window w, of another image is the squared errors summed over all pixels in the window:

$$Err = \Sigma(t - w)^2 = \Sigma t^2 + \Sigma w^2 - 2\Sigma wt$$

The product and exponentiations on the right hand side are inner products, meaning direct pixel wise operations. In order to check every where, we have to calculate this error for each position of the template in a sliding window. The first term on the left, is a constant for all positions of the sliding window. The second term is equivalent to a convolution of the pixel wise squared image with a uniform template of window size. The last term in the above equation involving a sliding window w, corresponds to a correlation.

Summed Square Error measures the squared difference between each point in the template and each corresponding point in the signal section being compared. This method is good at detecting similar trends (signal increasing or decreasing in the same direction and at the same rate), however it is sensitive to baseline (DC or very low frequency) shifts. Thus it is common to remove the mean from the template and the section being compared before performing the error measurement.

3.3 Convolution and Correlation

The convolution and correlation product between to continues functions f(x) and g(x) is defined:

$$(f * g)(t) = \int_{-\infty}^{\infty} f(x) \cdot g(t - x) dx$$

$$(f \circ g)(t) = \int_{-\infty}^{\infty} f(x) \cdot g(t + x) dx$$

By replacing the integral with a sum we get the discrete analogs, which we will use. It is straightforward to generalize to higher dimensions. The relevance of these products is there use in Template Matching. If we want to look for a specific a template, in an image, we can do so by comparing each individual window in the large image to the template. If the difference is small, we have a good match. This procedure of comparing with a sliding window is captured by the correlation. The only difference between convolution and correlation is a sign. In the following we will mostly talk about convolution, but keep in mind that it is identical to correlation up to a mirroring. The reason for this is the Convolution Theorem.

3.3.1 The convolution theorem

The theorem states that convolution in the spatial domain is equivalent to multiplication in the Fourier domain:

$$\overline{f * g} = \overline{f} \cdot \overline{g}$$

Proof:

The bar over an expression means the Fourier Transform Since the Fourier transform is its own .inverse up to a constant the reverse relation is also true. Multiplication in the spatial domain is equivalent to convolution in the Fourier domain.

There is a similar theorem for correlation:

$$\overline{f \circ g} = \overline{f}^* \cdot \overline{g}$$

The only difference is the complex conjugate on the right hand side.

One reason that this theorem is useful is the fact that there exists a very fast divide and conquer algorithm for convolution FFT. Another reason is its' usefulness in various proofs calculations and arguments. For instance, since the Fourier transform is linear, it is easily read off that convolution is commutative, associative and distributes over sums. Since the group of Gaussian Functions is closed under Fourier transformation and multiplication, it also follows that the group is closed under convolution. This fact alone brings you halfway through the proof of the central limit theorem.

Template convolution can be thought of as a filter rather than an error measure. It convolves the template with a section of the signal. It seeks to amplify the areas of the signal that are correlated to template. Because convolution is a multiplication process in the frequency domain, convolving the template with the signal can be viewed multiplying the frequency components of the template with the corresponding frequency components of the signal. This has the effect of amplifying only those portions of the signal that resemble the template.

CHAPTER FOUR

OTHER PATTERN RECOGNITION TOOLS

4.1 Statistical Methods

In statistical approaches, pattern data is represented by a feature vector which is used as input to some classifier or decision process. Features may characterize global form (area, elongatedness, major axis orientation) or local elements (corners, characteristic points). Shapes are viewed as points in shape feature space. For effective recognition, the requirement is to choose features such that patterns of the same class are tightly clustered in N dimensional space corresponding to N features, and patterns of different classes are in other tightly clustered regions well separated from each other (Duda and Hart 1973).

A key problem in statistical methods is the reduction of the dimensionality of the feature vector. This may be accomplished by a feature selection process, in which low significance features are deleted, or by a feature space transformation method, or both. Classically, a particular class was represented by a template with matching against templates; this matching was considered to be intractable for large numbers of objects due to the need to compare with inputs which have been rotated, scaled, partly occluded, non-rigidly transformed, or presented under varying lighting conditions. Recent schemes employing normalization (the RBF networks underlying Chorus) and interactions among multiple well chosen prototypes, or the sophisticated weighting of a large feature set (Mel 1997) have overcome this to some extent. Another approach to the use of features is to create a transformed representation space on the basis of correlations among the dimensions to enhance cluster tightness and inter-class separation. Feed-forward supervised networks, or competitive networks such as self organizing maps can use feature vectors as input, and via training transform the features into activation levels in a set of network elements corresponding to classes. Decision methods may generally be classed as non-parametric or parametric (Leedham 1991). Non-parametric methods include linear discriminant functions, minimum distance classifiers, and nearest neighbour classifiers. The most widely used parametric decision rule is the Bayes classifier. The main distinction from non-parametric methods is that the decision rule involves class conditional densities and a priori probabilities of occurrence of classes. Bayesian classifiers are particularly important with large object databases, where setting classifier decision boundaries properly and defining the optimal feature set are crucial for good recognition performances.

The description of statistical pattern recognition methods presented here thus far has been in general terms, applicable to any data set. Recognition of object shapes in a statistical framework poses additional problems unique to this class of data. Non-rigid objects are composed of parts which can assume different poses – human and animal figures are good examples.

The changing projections of three dimensional objects seen from different viewpoints constitute the stimulus identity problem. Different features and feature conjunctions will be present in each view. This problem has been addressed by geometric methods seeking invariants (treated in the next section), or by neural Networks exploiting regularities in the changing distributions of raw features (the Chorus RBF ensemble approach). Recently, however, progress on stimulus equivalence within a "raw feature" paradigm has been demonstrated, by careful design. Mel, describing the design goals for a recent high performance feature based system (Mel 1997), notes the following expectations on feature sets to overcome these problems:

1. Features should be large in number; sparsely occupied high dimensional representations are most robust to noise.
2. Features should be useful; they may be sensitive to object quality (occlusion, poor lighting) but should be robust in the face of pose or configuration changes.

2. Features should be useful; they may be sensitive to object quality (occlusion, poor lighting) but should be robust in the face of pose or configuration changes.

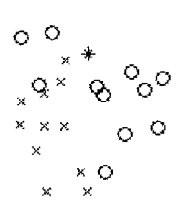
3. They should be dominated by spatially local features; this is particularly important for non-rigid objects, which preserve local but not global structure in any particular view.

4. They should be driven by multiple visual cues to maximize discrimination, represent diverse objects, and buffer representation against degradation which affects different cues (feature channels) more or less severely.

The use of these principles led to the creation of his SEEMORE system, which achieves recognition rates above 90% in a 100 object world, even for scrambled images. The high performance achieved with these first order 9 feature channels is interpreted by Mel to support the idea that a simple feature space is all that is needed and attempts to extract structural information or otherwise "bind" collections of features may be unnecessary for biological systems. It is easy, however, to construct images with identical first order statistics which will fool such a system but are readily distinguished by humans. It seems likely that some of SEEMORE's recognition success depends on diversity in first order statistics of the object world, along with limited use of second order statistics for some feature channels. [http://www.well.com/user/demaris/ch2.pdf]

4.1.1 Statistical Classification

The blue crosses are data points samples from one class, and the red circles are sampled from another class. The green asterisk is the new point or sample to which we must assign class membership.



The figure shows several things. If the new object is not identical to any of the samples, we can not know for sure, simply because our samples only carry partial information of the classes they are choices, and there no uniquely correct way of doing this. For instance, the green point is much nearer the centre of the red circles than the blue crosses. The nearest neighbour on the other hand is a blue cross. If we have no other information it is our choice according to which rules are classify.

It is also noticed that classes can overlap. Even though our classes in reality might be truly distinct, the measurements/data we have at our hand, might not be sufficient.

4.1.2 Classify According to Smallest Distance

If the new point can have any location, it is needed to assign a class number to any point in space. To do so, it is sufficient to define the borders between different areas. This is often done via a distance measure. Given a distance to every class, it is an obvious choice to classify according to the nearest. The boarders are then the subspaces in which distances to two different classes are identical. This means that the classification is done by defining a distance measure and solving a set of equations. There is no loss of generality in this procedure, but off course there might be cases where another approach is more attractive.

In the case of only two classes it will be denoted the boarder subspace as the separator and the difference between the two distances D, it will be denoted the discriminator. Thus the separator is the solution to the equation Δ=0.

4.1.2.1 Distance measures

In this section several distance measures will be discussed. Let dist denote our distance, whatever the measure might be.

- A linear measure
- A quadratic measure
- Interpolation based measure.

4.1.2.2 Mahalanobis distance (MHD).

It will be focused on the quadratic measure. When nothing else is mentioned it is assumed that the covariance matrix is positive semidefine. In other words all eigenvalues are real and bigger than zero. Under a Gaussian model the straightforward choice is to chose the distance measure imposed by the metric :

$$\text{dist} = (X-\mu)'\sum{}^{-1}(X-\mu)$$

This is called Mahalanobis distance. Eigenproblems can be written as composition of a rotation matrix and a diagonal matrix. $\sum^{-1}$ = RLR. This means that if it is shifted from X to the rotated coordinates $Y = (X-\mu)'R$, it has: dist = Y'LY, and the point is now that since L is diagonal this can be written as the sum:

$$dist = \sum_{1=1}^{d} \frac{y_i^2}{\lambda_i}$$

λ_i is the i'th eigenvalue. It is now clear why the eigenvalues is not allowed to be zero.

4.1.2.2.1 The two class problem

The two-class discriminator in the Mahalanobi case is:

$$\Delta = (X-\mu_1)'\sum{}_1^{-1}(X-\mu_1) - (X-\mu_2)'\sum{}^{-1}(X-\mu_2)$$

This can be rewritten

$$\Delta = (X-\mu)'\sum{}^{-1}(X-\mu) + k$$

Where

$$\sum^{-1} = \sum_{1}^{-1} - \sum_{2}^{-1}$$

$$\mu = \sum(\sum_{1}^{-1}\mu_1 - \sum_{2}^{-1}\mu_2)$$

$$\mathrm{k} = \mu_1'\sum_{1}^{-1}\mu_1 - \mu_2'\sum_{2}^{-1}\mu_2 - \mu'\sum^{-1}\mu$$

To understand the implications of this it is useful to have a look at two dimensional examples. First note that the new metric of the separator $\sum_{1}^{-1} - \sum_{2}^{-1}$ $\eta\alpha\sigma$ real eigenvalues. In the generic case they will be non zero, but they can be negative. There are two distinct cases, they can have same sign andopposite signs. Both cases are shown on the figures beneath. On the left figure eigenvalues has same sign and the separator is an ellipse. On the right figure signs are opposite and the separator is a hyperbola. The hyperbola of course continues towards infinity and only part of it is visible in the figure window. Note that the area contained by the left leg of the hyperbola is classified as belonging to the "red" class. To the eye this might at first seem counter intuitive, but the explanation is that the "red" class has bigger variance in the left right direction than the "blue" class.

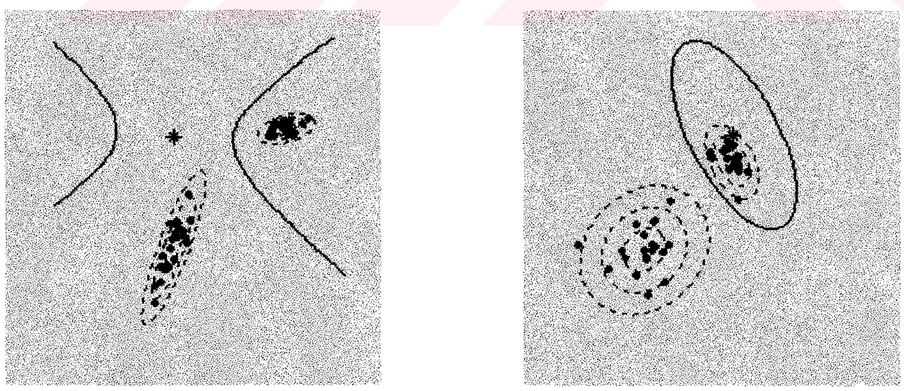


Fig 4.1 Eigenvalues and separators

4.1.2.2.2 Fast Approximate Estimate of MHD. Dimensionally Reduction

The transformation to the Principal components can be attractive because it tells us if variation is small in some directions. If the task is to represent information in a sparse manner, this is useful because we can choose to specify only the coordinates in the directions of large variation. Neglecting the coordinates in directions of small variation only introduces small errors. This procedure is often referred to as dimensionality reduction.

In the case of classification this procedure also can be relevant. Calculating Mahalanobis distance has a time complexity of d squared. If it can be done reasonably with an estimate based on only m eigencoordinates the time complexity reduces to md. The key to finding the approximation is to note, that the total variance V of the coordinates of X (around μ) $\sum x_i$ is equal to $\sum y_i$. Furthermore the maximum likelihood estimate of this is Trace $(\sum) = \sum \lambda_i$. Using this:

$$\sum_{i=1}^{d} \frac{y_i^2}{\lambda_i} = \sum_{i=1}^{m} \frac{y_i^2}{\lambda_i} + \sum_{i=m+1}^{d} \frac{y_i^2}{\lambda_i} \cong$$

$$\sum_{i=1}^{m} \frac{y_i^2}{\lambda_i} + (d-m) \cdot \frac{V - \sum_{i=1}^{m} y_i^2}{V - \sum_{i=1}^{m} \lambda_i}$$

where the sum in the second term has been split over nominator and denominator. The sum over the eigenvalues in the nominator has been replaced by their mean value. This approximation is good if their variation is small or the nominator is small.

4.1.2.2.3 Choice of Eigenvector Subset

The approximation is valid for any subset of the eigenvectors. For compression it is choose the large eigenvalues, but in classification the small eigenvalues might be a better choice, because directions of small variations are the useful ones when it is wanted to distinguish between classes. However as shown in the section: Reliability of covariance estimate and eigenspectrum you should be careful when using the small eigenvalues.

4.1.2.2.4 Prescaling of Data

The performance of the above approximation formula can be improved by a Prescaling. The units of measurement for the individual coordinate in the data vector X is a matter of free choice. The PCA analyses will 100% eliminate the difference between any two choices, but the approximation formula is better off, when all variables are scaled to have same variance (one), meaning that the diagonal of the covariance matrix is a vector containing all ones. The relative variation of the estimates is biggest for the smallest eigenvalues.

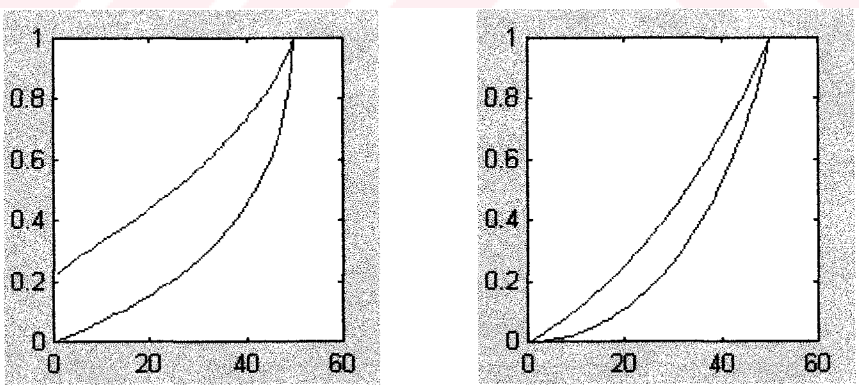


Fig 4.2 Effect of prescaling on artificial data

The figures on the left show the effect of prescaling on artificial data. The left figure shows the eigenvalue spectrums, and the right figure shows the accumulated spectrums. The blue curve is with prescaling, and the red curve is without. The conclusion is (in this example) that the spectrum is more flat in the top end. The consequence of this is, that a distance estimate based on the small eigenvalues, is less error prone after prescaling, because a mean value approximates the remaining spectrum better.

4.1.2.3 Comment

The benefit of prescaling might not be overwhelming. It is easy to construct examples where the effect is vanishing. It is very difficult to say anything in general, because it all depends on the metric in the actual case. My suggestion would be simply to check in each case. The cost is that all data vectors have to multiply by a scaling vector. This is the same cost as calculating one principal component. So if you are able to get the desired precision in you distance estimate by fewer principal components, then it's worth it.

4.1.2.4 A linear distance measure

In the previous section it was shown how we could classify by projecting data on a set of eigenvectors or alternatively calculate the distance directly via the full metric: dist $=X' \sum^{-1} X$. It is of course possible to design faster but possible poorer procedures. One method is to use the projection, on only one direction in space. The critical point in this strategy is to calculate the direction best suited for this purpose. A solution to this problem is presented here.

First the problem has to be stated. The straightforward suggestion is: which linear separator would minimize the probability of misclassification. It turns out that it's difficult to find an effective algorithm that produces the answer. Instead we will use an alternative formulation that can attack analytically.

Find the direction upon which the overlap between projections of the two classes is minimal.

The overlap W between to probability densities f(x) and g(x) i define as :

$$\Omega = \int_{-\infty}^{\infty} f(x) \cdot g(x) dx$$

In the case of two one dimensional Gaussians N($\mu_1 \sigma_1$), N($\mu_2 \sigma_2$) the overlap is

$$\Omega = \exp\left(-\frac{1}{2}\frac{\mu^2}{\sigma^2} + \ln(2\pi\sigma) \right)$$

$$\mu = \mu_1 - \mu_2 \qquad \sigma^2 = \sigma_1^2 + \sigma_2^2$$

The projection of the vector μ upon some direction vector n, is μ 'n. The variance in a specific direction n given the metric $\sum$, is n $\sum$ n. Inserting these projections into the above equation gives the problem of maximizing the expression:

$$\frac{n'Mn}{n'\Sigma n} + \ln(n'\Sigma n)$$

$$M = \mu\mu' \qquad \Sigma = \Sigma_1 + \Sigma_2$$

When $\sum$ is positive definite, it can be decomposed into B'B, and by substituting x=B'n the expression writes: x'(B^{-1})'M(B^{-1})x, and maximizing this is equivalent to the eigenvalue problem :

$$B'^{-1} MB^{-1}x = \lambda x \Leftrightarrow$$

$$Mx = \lambda \Sigma x$$

The lower expression is recognized as a generalized eigenvalue problem. The solutions has to be retransformed to the original system by n=B'^{-1}x.

4.1.2.5 Comments

The procedure can be understood simply in geometrical terms. First it is coordinated transform our system to get an isotropic metric. Then x'$\sum$ x equals one, and we get an ordinary eigenproblem, which solutions has to be retransformed.

The matrix M is an outer product $\mu\mu'$ (not the inner product $\mu\mu'$) of the vector μ by itself. M only has rank one, and there is only one eigenvector with a non zero eigenvalue. This is our best separating direction.

4.1.2.6 Position of the separator

The two figures shows hyper planes (full blue lines in 2D) having the best separating direction as normal vectors. The figures are analog to the two figures in the case of quadratic discriminate. The quadratic separators have been shown in dotted black as well for comparison.

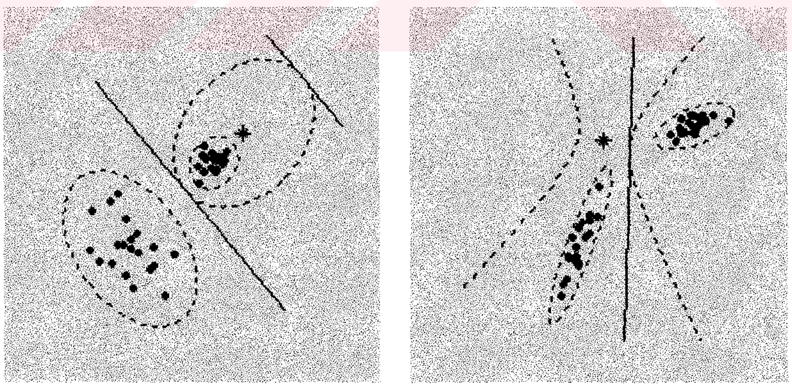


Fig 4.3 Mahalanobi Separator

As shown beneath, it turns out that separating via the best single direction, requires the use of two parallel hyper planes. At first that might seem surprising, but on the right figure it is clear, that it reflects the properties of the Mahalanobi separator. On the right figure we only see one, the other one being far out to the left. When one of the planes is located very far from the classes, it does not have much practical significance. As a limiting case one plane can be located in infinity, but this is non generic

Both figures make it clear, that we need to specify where to locate the separators. So far we have found its normal, but there is still one translational degree of freedom. If we classify according to MHD in the one-dimensional case of to Gaussians N(μ_1 σ_1), N(μ_2 σ_2), the criterion is :

$$\frac{(x-\mu_1)^2}{\sigma_1^2} > \frac{(x-\mu_2)^2}{\sigma_2^2} \Leftrightarrow$$

$$\frac{\sigma_1\mu_2 - \sigma_2\mu_1}{\sigma_1 - \sigma_2} < x < \frac{\sigma_1\mu_2 + \sigma_2\mu_1}{\sigma_1 + \sigma_2}$$

For $\sigma_1 < \sigma_2$ otherwise the solution is outside the roots. As above the values of m and s are the projections of the d-dimensional analogs μ and $\sum$. If the direction vector is n, $\mu = \mu' n$ and $\sigma = n' \sum n$

4.1.2.7 Conclusion

A method has been given to calculate hyper planes (subspaces of dimension d-1) with high classification performance. Only d+1 parameters has to be estimated. The MHD is potentially a stronger classifier but relies on d(d-1)/2 parameters and is much more sensible to weak statistics.

4.1.3 Classification by Iterative Linear Separation

In stead of trying to use a "one shot" classifier, it is possible to use several linear classifiers in an iterative setting. It can be relaxed our classification in the sense, that we at first iteration only accept classifications far one the safe side, and leave the rest for next iteration.

This is showed on the figure at the left. In stead of positioning our separator planes at the optimal locations, (shown dotted) we translate them to leave some parts of space unclassified. In the training face the remaining unclassified data can now be used to calculate a second linear classifier and so on. This also has the strength that non-Gaussian Classes that are entangled as on the figure can be handled better than by MHD. The drawback is that there is no obvious way of automatically calculating the alternative "safe" positions of hyper planes. But since we are working with one-dimensional directions of separation, we can monitor the projections of the classes, and set the positions interactively in the training phase.

4.1.3.1 Interpolation

A third way of defining distance measures is via interpolation. If it is viewed each class as a function in space and it has samples of a class, then it can be used any kind of interpolation/extrapolation scheme to assign values to other points in space. This can be done for each class, and it can then classify any point according to the class that has the highest interpolated function value in that point. [http://www.diku.dk/undervisning/1999f/f99.134/Classification.html#Mahalanobis%20distance%20(MHD)]

4.2 Structural or Syntactic Methods

The other major family of classic pattern recognition approaches, chiefly developed for image or shape processing applications are structural or syntactic methods (Pavlidis 1977). Here, the input image must first be segmented into primitives; the primitives must be recognized, and spatial or topological relationships between these primitives extracted. Finally, with this information, a syntactic analysis and classification on that basis can proceed. None of these problems are trivial. Within computer vision, structural methods based on raw image data have been largely superseded by related methods which capture structural information implicitly by multi-scale representations or by deformations. In the psychological examination of human vision, structural approaches still command a good deal of support. In part, this is 9 First order features implies that no information on the spatial proximity of other features is present. Second order features would capture adjacencies of feature pairs at one or more scales, with increasing high order features preserving this trend. Due to the fact that task specific or language mediated descriptions of objects offer evidence that compositional representations are used. Statistical approaches and feedforward neural networks have been problematic in regard to this issue. Compositionality is essentially the separability of the components of a composite representation, i.e. the ability to use or talk about them independently after the formation of that representation (Van Gelder 1990). Recurrent networks have been demonstrated to exhibit a so called functional compositionality, in which tree structures can be represented and their constituent parts derived (Pollack 1990).

4.3 Neural Network

4.3.1 Introduction

Many claims have been made concerning the importance of neural networks (NNs) as a paradigm shift in modelling both nature and the central processes of Information Technology including, most directly, Artificial Intelligence problem domains. To this stage, NNs have been applied to many and varied areas of inquiry from the control of chemical plants, through to pattern recognition, and many biological domains. Further, most proponents do not claim that NNs literally correspond to what actually occurs in the human brain, but there is a general belief of a loose correspondence between the actual computational units used and the response properties of individual neurons. Added to this, there is a belief that the inherent parallelism of NNs is consistent with brain function and that the use of fundamental numerical computations, in contrast to symbolic or declarative representations, is representative of the “hardware" of intelligent behaviour. This thesis is not aimed at challenging these claims.

It is important to note that most past NN formulations have a few central features in common. They are:

1. Problems are solved by the determination of weights or states in an extremely high-dimensional state space.
2. Learning, parameter estimation, and information processing are all inherently parallel.
3. No further constraints on the system are required, since the NN learns weights which are necessary and sufficient to predict behaviour.

This indicates that most applications of NNs are based upon the assumption that solutions to problems can be obtained by using generic technologies which essentially search high-dimensional state spaces, and so require no additional knowledge about the system under analysis. Examples of this abound in the literature, where the input and output level responses are defined by discrete nodes and their transducer functions, and at least one hidden layer is introduced. Further, NNs usually employ Supervised Learning which, in one sense, is a form of constraining the system and weight estimation processes. However, it is usually "black-box" in so far as it, per se, makes no assumptions about the processing characteristics and desired properties of the system not explicit in the training samples. For example, most traditional NN approaches to pattern recognition lack explicit shift, rotation and scale invariances, as the NNs are not modelled with such characteristics in mind. If such invariances arise it is due to the presence of appropriately shifted, rotated, or scaled example patterns in the training data set. [e.g Tebelski and Waibel, 1990; Waibel, Jain, McNair, Saito, Hauptmann and Tebelski, 1991] This can result in dramatically improved generalization of classification performance to patterns not present in the training data, and representation in considerably lower-dimensional state spaces. Perhaps most importantly, model-based NNs can be constructed so that they are guaranteed to have responses which are invariant under certain transformations of the input data. Such networks can be trained with very much smaller training sets, since it is no longer necessary to provide examples of transformed versions of the input prototypes. This, coupled with the reduction in the dimensionality of the parameter space, means that training such model-based NNs is often much less computationally-expensive than the traditional alternative. To attain these goals we first define the classical formulation, compare it to past technologies, and then develop the model-based formulation.

4.3.2 Classical Feed-forward Neural Network

For a classical feed-forward NN, the input x_j to a given "neuron", j, is defined by

$$x_j = \sum_i y_i w_{ij}$$

and the output is defined by the logistic function

$$y = \frac{1}{1+e^{-x}}$$

This process is implemented in NNs with at least three layers: an input layer, and output layer and one or more hidden layers. In classification or recognition problems, the input layer is defined by an array of nodes which constitute a sampled version of the input signal. The output layer is defined by a set of nodes each corresponding to a class, pattern type, or category. Connections w_{ij} between nodes are usually restricted to layers above and below a given layer (see Figure 3.4). For a 3-layer NN with 100, 20, and 10 nodes for the three layers, respectively, this results in 100 x 20 + 20 x 10 = 2200 connections. Each such connection has a weight to be estimated in accordance with the desired input-output (I-O) characteristics.

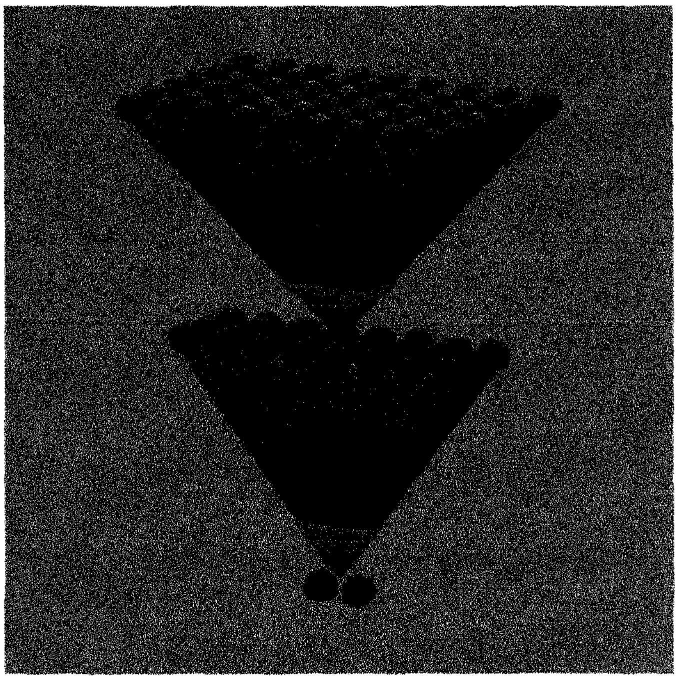


Figure 4.4: A classical feed-forward network. The weighting function between all pairs of levels is of type M.1.

For recognition or classification problems using Supervised Learning, the NN is trained to reproduce given outputs from known input examples as accurately as possible. That is, the learning problem is defined by estimating the w_{ij} such that a given output error, or “cost", function E is minimized. The most commonly used cost function is the sum of the squares of the errors of the output nodes:

$$\min_{w_{ij}} \left\{ E = \frac{1}{2} \sum_{l} \sum_{c} (y_{lc} - t_{lc})^2 \right\}$$

where w_{ij} refers to all connection weights over all levels of the network, c indexes the input-output exemplar pairs, l indexes the output nodes, y_{lc} is the actual output, and t_{lc} is the desired output.

Though described in a variety of ways, learning techniques use either deterministic gradient descent (in particular, back propagation [Rumelhart, Hinton and Williams, 1986b; Rumelhart, Hinton and Williams, 1986a; Plaut and Hinton, 1987] or stochastic relaxation methods [e.g. Metropolis, Rosenbluth, Rosenbluth, Teller and Teller, 1953; Kirkpatrick, Jr. and Vecchi, 1983; Aarts and Korst, 1989] to nd a set of weights (a state of the network). Solution or convergence times can be improved with various acceleration techniques [e.g. Fahlman, 1988].

We are not so much concerned with these search techniques but, rather, the definition of the state space being searched. The state space dimensionality is very high since it is defined by the total number of independent connection weights, which here includes every connection. Consequently, there are likely to be many local minima available to the solution technique for the training examples, but do not characterize the problem in general. The problems associated with search time (number of iterations for convergence, etc.) are also significant. Further, under the classical NN perspective, no further constraints are added to such estimation problems as that would challenge the idea that relations and rules can be “discovered" automatically through the minimization or search for global minima of above equation. Here, we propose to retain the notion that states can indeed be discovered by the learning procedure, but to constrain the search procedure by modelling the network to explicitly include our knowledge of the important features of the data, or desired invariant properties.

This modelling can take a variety of forms. The network is frequently divided into a number of modules. These modules can be designed to perform component sub-tasks of the overall classification problem. Within a module, the weights may be constrained in some way: some weights may be set by hand; others may be constrained by a functional relationship; the module may be trained to perform a mapping independent of the classification training data. First, however, this view must be put in perspective. To this end, we will first relate the parameter estimation problem in NNs to similar problems in Signal Processing and Principal Components Analysis.

4.3.3 A Matrix View of Neural Networks

We first note that, for a given set of connections, Equations 1.1 and 1.2 can be written in matrix form as:

$$X_n = f\{WY_m\}$$

Where f corresponds to the logistic nonlinear transducer (Equation 1.2), W to the connection matrix of size n _ m (for m \source" and n \destination" nodes), and X and Y to the source and destination vectors corresponding to the pair-wise NN node layers. This form points to the essential idea that the own of information from one level to the next corresponds to a transformation, defined by the connection matrix, which maps m-dimensional signals into n-dimensional ones, where, usually, $n < m$. The nonlinear transducer is used to map input values monotonically to the range (0; 1). The net effect is that the NN procedure endeavours to discover a mapping which satisfies an overt constraint like least squares (Equation 1.3) or others.

Consequently, the best form of W is that of an orthogonal mapping of rank equal to the true dimensionality of the input - as measured, for example, by the rank of the input signal correlation matrix. Hidden units (units in neither the input nor output layers) correspond to components of the eigenstructure in the mappings, albeit constrained by the optimization goal. Except for the nonlinear transducer function, the use of hidden units in NNs is a way of determining the eigenstructure, or principal components, of the data. There must be at least as many hidden units (spread throughout the hidden layers) as there are non-zero eigenvalues in the input data correlation matrix, since it is necessary to span a vector space of the same dimensionality as the signal samples. The hidden units, by definition, extract the important features in the signals which, in more traditional signal processing, are extracted through the eigenvectors of the signal autocorrelation or autocovariance matrices [Ade, 1983; Ahmed and Rao, 1975].

Since many applications of NNs lie in the area of classification, it is interesting to note that, when the final output layer of such a cascaded NN is a set of nodes corresponding to classes or categories, we can interpret the system as a form of Discriminant Function Analysis. The Discriminant Analysis model is based upon the determination of decision hyper planes which lie between sample class means and which maximize the between-class and minimize the within-class variance from their projections onto such planes. Discriminant functions are linear equations of the form:

$$z = \sum_{i=1}^{m} a_i y_i + a_0$$

and class membership is determined by the value of the function. For example, for a two-class classification problem, class membership is determined by the sign of z. For the n-class problem Equation 1.5 generalizes to the matrix form:

$$Z=DY$$

where the n-class vector Z determines the weights associated with the data, Y , from each of the classes. D corresponds to the set of discriminant function coefficients which define each class. The crucial difference between NNs and classical Discriminant Function Analysis is that the decision boundaries of NNs are non-linear, due to the transducer f.

Similar formulations for classification have occurred in the Pattern Recognition literature where, for example, the Least Squares Minimum Distance Classifier [Ahmed and Rao, 1975] attempts to find a mapping, in matrix form, which transforms samples in feature space into points in "decision space" (whose dimensionality corresponds to the number of classes) and satisfies the constraint that samples from the same class should be mapped as close as possible to each other, and as far away as possible from other class samples, in decision space.

However, NNs differ from these well-known past techniques, even in their traditional form, by binding the representation and processing characteristics together. Although all these techniques have similar aims and structures, the NN formulation integrates the processing characteristics with the decision processes in one network, which is represented, in general, by a set of cascaded transformations. The use of nonlinear transducers and layers of differing sizes has the disadvantage that analytic solutions are difficult, particularly since the dimensionality of the representation is of high order, but permits the network to distinguish between classes which are not linearly separable.

Our aim is to preserve this binding of process with representation (feature space), but to extend the NN philosophy to include more explicit constraints on the network geometry and connection weights. The resulting systems behave similarly to traditional NNs, but have two main advantages. First, it is possible to construct networks that are constrained to respond to features of the input data that are known a priori to characterize the task or to have desired invariances, rather than hoping that the training data will cause the optimization technique to find a set of weights with these properties. Secondly, this allows the dimensionality of the system to be reduced, which can reduce the chance of finding a local minimum that characterizes the training data but not the general task. We call this extension Model-Based Neural Networks (MBNN).

The use of MBNNs allows a network to be constructed in which the supervisor's knowledge of the task to be performed is used to specify, partially or completely, the roles of some hidden units, or of whole hidden layers or modules, in advance. Thus the supervisor's knowledge of which features of the training data are significant for the task is incorporated into the network geometry and connection weighting functions, serving as a constraint on the state space searched.

4.4 Wavelet Analysis of Epileptic Spikes

Interictal spikes and sharp waves in human EEG are characteristic signatures of epilepsy. These potentials originate as a result of synchronous, pathological discharge of many neurons. The reliable detection of such potentials has been the long-standing problem in EEG analysis, especially after long-term monitoring became common in investigation of epileptic patients. The traditional definition of a spike is based on its amplitude, duration, sharpness, and emergence from its background. However, spike detection systems built solely around this definition are not reliable due to the presence of numerous transients and artifacts. It is used wavelet transform to analyse the properties of EEG manifestations of epilepsy. It is demonstrated that the behaviour of wavelet transform of epileptic spikes across scales can constitute the foundation of a relatively simple yet effective detection algorithm.

The wavelet transform is an integral transform for which the set of basis functions, known as wavelets, are well localized both in time and frequency. Moreover, the wavelet basis can be constructed from a single function $\varphi(t)$ by means of translation and dilation:

$$\varphi_{a;t_0} = \varphi(\frac{t-t_0}{a}).$$

$\varphi(t)$ is commonly referred to as the mother function or analysing wavelet. The wavelet transform of function $h(t)$ is defined as

$$W(a,t_0) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} h(t)\varphi^*_{a;t_0}\, dt,$$

Where $\varphi^*(t)$ is referred the complex conjugate of $\varphi(t)$. The continuous wavelet transform of a discretе time series $\{h_i\}_{i=0}^{N-1}$ of length N and equal spacing ∂t is defined as

$$W_n(a) = \sqrt{\frac{\partial t}{a}} \sum_{n'=0}^{N-1} h_{n'} \varphi^* \left[\frac{(n' - n)\partial t}{a} \right].$$

The above convolution can be evaluated for any of N values of the time index n. However, by choosing all N successive time index values, the convolution theorem allows us to calculate all N convolutions simultaneously in Fourier space using a discrete Fourier transform (DFT). The DFT of $\{h_i\}_{i=0}^{N-1}$ is

$$\hat{h}_k = \frac{1}{N} \sum_{n=0}^{N-1} h_n e^{-2\pi i k n / N},$$

Where $k = 0, ..., N-1$ is the frequency index. If one notes that the Fourier transform of a function $\varphi(t/a)$ is $|a|\hat{\varphi}(af)$ then by the convolution theorem

$$W_n(a) = \sqrt{a\delta t} \sum_{k=0}^{N-1} h_n \hat{\varphi}^*(af_k) e^{2\pi i f_k n \delta t}$$

Frequencies f_k are defined in the conventional way. Using and a Standard fast Fourier transform (FFT) routine it is possible to efficiently calculate the continuous wavelet transform (for a given scale a) at all n simultaneously. It should be emphasized that formally above equation does not yield the discrete linear convolution corresponding to equation $W_n(a)$ but rather a discrete circular convolution in which the shift $n' - n$ is taken modulo N. However, in the context of this work, this problem does not give rise to any numerical difficulties. This is because, for purely practical reasons, the beginning and the end of the analysed part of data stream are not taken into account during the EEG spike detection.

From a plethora of available mother wavelets, It is employed the Mexican hat

$$\varphi(t) = \frac{2}{\sqrt{3}} \pi^{-1/4} (1 - t^2) e^{-t^2/2}$$

which is particularly suitable for studying epileptic events.

In the top panel of Fig 3.5 it is presented two pieces of the EEG recording joined at approximately $t = 1s$. The digital 19-channel recording sampled at 240 Hz was obtained from a juvenile epileptic patient according to the international 10-20 standard with the reference average electrode. The epileptic spike in this figure (marked by the arrow) is followed by two artifacts. The bottom panel of Fig.1 displays the contour map of the absolute value of Mexican hat wavelet coefficients $W(a,t_0)$. It is apparent that the red prominent ridges correspond to the position of either spike or the motion artifacts. What is most important, for small scales, a, the values of the wavelet coefficients for the spike's ridge are much larger than those for the artifacts. The peak value along the spike ridge corresponds to $a = 7$. In sharp contrast, for the range of scales used in Fig. 3.5. The absolute value of coefficients $W(a,t_0)$ for the artifacts grows monotonically with a.

The question arises to whether the behaviour of the wavelet transform as a function of scale can be used to develop a reliable detection algorithm. The first step in this direction is to use the normalized wavelet power

$$w(a,t_0) = W^2(a,t_0)/\sigma^2$$

instead of the wavelet coefficients to reduce the dependence on the amplitude of the EEG recording. In the above formula σ^2 is the variance of the portion of the signal being analyzed (typically it is used pieces of length 1024 for EEG tracings sampled at 240 Hz). In actual numerical calculations it is preferred to use the square of $w(a,t_0)$ to merely increase the range of values analysed the spike detection process. In Fig. 3.5 w^2 for the signal used in Fig 3.5 is plotted for three scales $A = 3$, $B = 7$ and $C = 20$.

In the most straightforward approach, it is identified an EEG transient potential as a simple or isolated epileptic spike if and only if:

- the value of w^2 at a = *7* is greater than a predetermined threshold value T_1 ,
- *the square of normalized wavelet power decreases from scale a = 7 to a = 20,*
- the value of w^2 at a = *3* is greater than a predetermined threshold value T_2.

The threshold values T_1 and T_2 may be considered as the model's parameters which can be adjusted to achieve the desired *sensitivity* (the ratio of detected epileptic events to the total number of epileptic events present in the analysed EEG tracing) and *selectivity* (the ratio of epileptic events to the total number of events marked by the algorithm as epileptic spikes).

While its simple algorithm is quite effective for simple spikes such as one shown in Fig.3.5 it fails for the common case of an epileptic spike accompanied by a slow wave with comparable amplitude. The example of such complex is given in Fig. 3.5 (a) . The overlap of the negative tail of the Mexican hat with the slow wave yield the inherently low values of w^2 at scale A (panel (b)) and scale B (panel (c)) as compared to those characteristic of the "isolated" spike. Nevertheless, the normalized wavelet power does decrease from scale *B to C*. Consequently, in the same vein as the argument we presented above, it can be developed an algorithm which detects the epileptic spike in the vicinity of a slow wave by calculating the following linear combination of wavelet transforms:

$$\widetilde{W}(a,t_0) = c_1 W(a,t_0) + c_2 W(a_s,t_0+\tau)$$

and checking whether the square of corresponding normalized power $\widetilde{w}(a,t_0) = \widetilde{W}^2(a,t_0)/\sigma^2$ at scales a = *3* and a = *7* exceed the threshold value $\widetilde{T}_1$ and $\widetilde{T}_2$, respectively. The second term of the above equation allows us to detect the slow wave, which follows the spike. The parameters a_s and τ are chosen to maximize the overlap of the wavelet with the slow wave. For the Mexican hat is used $a_s = 28$ and $\tau = 0.125s$. By varying the values of coefficients c_1 and c_2, it is possible to control the relative contribution of the spike and the slow wave to the linear combination.

For testing purposes, it is built up the database of artifacts and spikes. It is made available some of these EEG tracings along with the examples of the numerical calculations. While the analysis of the pieces of EEG recordings such as those shown Fig. 3.5 is essential in determining the generic properties of the epileptic events, it can hardly reflect the difficulties one can encounter in interpretation of clinical EEG. Therefore it had been selected four *challenging* EEG tracings with 340 epileptic events. The algorithm described marked 356 events out of which 239 turned out to be the epileptic events. Thus the sensitivity of the algorithm was 70 % and its selectivity was equal to 67%. Then it was analysed the same tracings with the leading commercial spike detector developed by the Persyst Development Corporation. This software marked 654 events out of which 268 were epileptic events. Thus slightly better sensitivity of 79%was achieved at the expense of the low 41% selectivity.

The goal of wavelet analysis of the two types of spikes, presented in this chapter, was to elucidate the approach to epileptic events detection, which explicitly hinges on the behaviour of wavelet power spectrum of EEG signal *across* scales and not merely on its values. Thus, this approach is distinct not only from the detection algorithms based upon discrete multiresolation representations of EEG recordings but also from the method developed by Senhadji and Wendling which employs continuous wavelet transform.

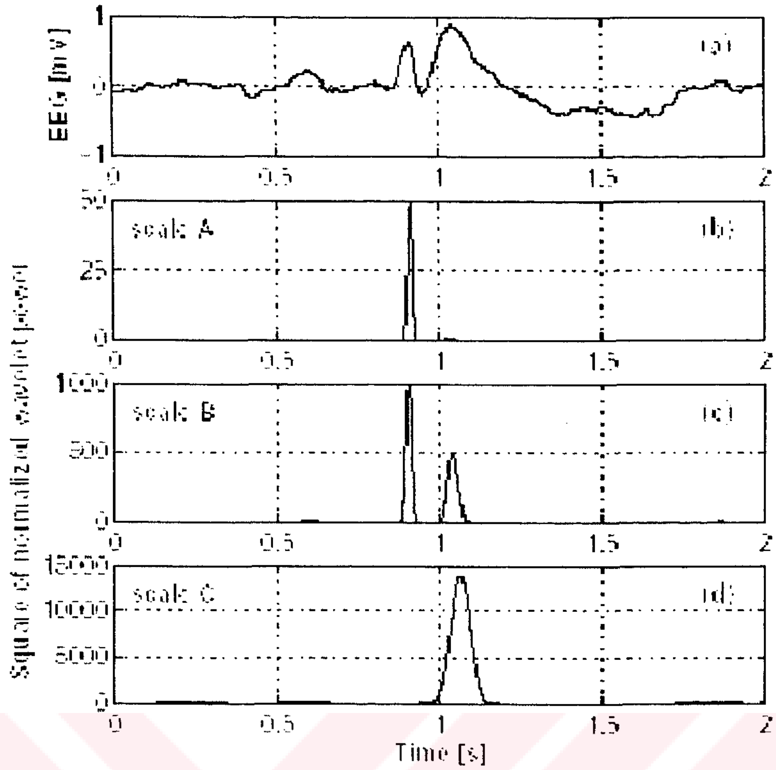


Fig 4.5 (a) Epileptic spike – slow wave complex. The amplitude of the slow wave is comparable to that of the spike. The square of normalized wavelet power for this signal is shown in panels (b) – (d) for three different scales $A<B<C$

4.5 EEG Spike Detection Using Deconvolution

The method of predictive deconvolution is employed for spike detection in EEG signals (Robinbson. E.A. 1958 Mohandas 1981). Based on experiments on real EEG segments, it has been shown that spikes in the frequency range 16-60 Hertz, which is common in EEG signal, can effectively detected by these methods. Very sharp spikes can also be detected. The effectiveness of the proposed algorithms for EEG spike detection is compared using quantitative measures such as signal to noise ratio improvement factor.

4.5.1 Deconvolution and Spike Detection

The recorded EEG signal is usually corrupted by noise so that if (z(k)) is the noise free EEG signal, the recorded EEG can be written as :

$$y(k) = z(k) + v(k)$$

where (v(k)) is a noise sequence assumed to be uncorrelated. The signal (z(k)) available at the skull is influenced by physiological changes in the brain. These changes in patients with epilepsy, or with injuries of the brain are in the form of sudden electric discharges. These discharges appear as spikes in the recorded EEG signal. It is of clinical interest to detect these spikes, which will help in the diognosis of epilepsy, brain damage, etc. The detection of such random spikes is not easily done since these spikes may be hidden in the background activity or masked by the noise present in the observations. The purpose of deconvolution is, then, to enhance the resolution of these spikes, given the EEG record (y(k)) with suspected spikes present in it.

A quantitative measure to indicate the resolution of these spikes against the EEG background can be defined and evaluated (Mohandas 1981). The signal to noise improvement factor (SNRIF) can be defined as:

$$SNRIF = \frac{\left(\sum_{i=M1}^{M2}[e(i)]^2\right) \Big/ \left(\sum_{i=1}^{N}[e(i)]^2\right)}{\sum_{i=M1}^{M2}[y(i)]^2 \Big/ \sum_{i=1}^{N}[y(i)]^2}$$

where M1 and M2 mark the beginning and end of a small window in which the spike appears and N is the total number of sample the recorded EEG segment and e(k) is the prediction error which is the same as the convolved output. Greater the SNRIF better will be the resolution of the spikes against the background.

4.5.2 Algorithms for Deconvolutions

Deconvolution or inverse filtering is the process of estimating the input function, given the output of the system and the system weighting sequence. The method of deconvolution has been extensively used in seismic data processing (Robinson. E.A 1957, Mohandas. K. P. 1981) in which, the input to the system is assumed to be a random reflection coefficient sequence. In the absence of information regarding the system weighting sequence (seismic wavelet), the input sequence is estimated using the predictive deconvolution approach of Robinson. The situation in EEG signal spike detection problem is quite similar in the sense that only the EEG signal output is available, system function is not known and the input signal to be estimated is a random spike time of occurrence are not known a priori. A time series model fitted to the real EEG signal can remove the predictable part of the signal leaving the prediction error as an estimate of the input spike signal.

In some of the earlier studies in EEG modelling autoregressive models have been extensively used. Some attempts have also been made to describe an EEG signal as an output of an ARMA process.

4.5.3 Nonrecursive Autoregressive Models

The current value of y(k) is expressed is the weighted sum of pasat values of y(k)'s plus a random shock.

$$y(k) = a_1.y(k-1) + ... + a_p.y(k-p) + e(k)$$

The parameters of the model can be estimated nonrecursively using Yule-Walker equations. The single step predicted value of y(k) will then be :

$$\hat{y}(k) = \hat{a}_1.y(k-1) + ... + a_p.y(\hat{k}-p)$$

where $\hat{a}_i$'s are the values of the estimated parameters. Then the prediction error e (k) is given by:

$$e(k) = y(k) - \hat{y}(k)$$

4.5.4 Recursive ARMA Predictors

In a manner similar to above equation it is possible to express the current value of y(k) as weighted sum of the p previous values of y and q previous values of the white noise sequence (e(k)). This model is more general and in known as the mixed autoregressive moving average (ARMA) model. In this case:

$$y(k) = \sum_{i=1}^{p} a_i y(k-i) + \sum_{j=1}^{q} b_j e(k-j)$$

A Kalman-type algorithm can be used to estimate the ARMA model parameters. Since the parameters are recursively updated with each incoming data points, such an algorithm can track slow variations in the parameters and hence weak nonstationarities in the data.

4.5.5 State Variable Models

Use of state variable models for deconvolution as an efficient alternative to AR/ARMA models has been demonstrated (Mohandas 1981. Mahalanobis. Et.al 1981) from the known values of the output sequence, and the estimated values of the auto-covariance function. An innovations' model can be identified. The model is described by:

$$\hat{x}(k/k+1) = F\,\hat{x}(k/k+1) + K.e(k)$$

$$\hat{y}(k) = H\,\hat{x}(k/k-1) + e(k)$$

where $\hat{x}(k/k-1)$ is the predicted estimate of x(k) based on the measurements y(0), y(1),...,y(k-1) ; e(k) is the zero mean innovations' sequence with unknown covariance Q and K represents the asymptotic value of the Kalman predictor gain.

The details of the model parameter identification are given elsewhere (Tse & Weinert 1975, Mahalanobis. Et. Al 1981).

This method, even though nonrecursive yields comparable performance at far less computational effort. The determinant ratio test (Woodside 1971) can be used I determining the system order.

4.5.6 Effectiveness of Different Algorithms

The comparison of the performance of these algorithms has been made quantitavely and qualitatively using simulation experiments were conducted on several segments of real EEG data collected from a normal patient. As there were no apparent spikes or other abnormalities in the recorded EEG, spikes of known amplitude and frequency were superimposed on this. In the result presented in the paper, specifically, we have chosen a sharp spike and another of frequency 32 Hz for illustrating the efficacy of the proposed methods for spike detection.

The quantitative comparison of the performance of the algorithms is based on the SNRIF defined in above equation and the qualitative comparison based on the plots of the deconvolved EEG segments. In table I is presented the SNRIF obtained using the different methods of spike detection under study. As would be expected, the recursive methods, and the methods based on ARMA models yield marginally better SNRIF in comparison with the one based on an AR model as hitherto used in such studies. Recursive models are apparently better for detection of sharp spikes. However, the numbers parenthesis of these recursive methods may become prohibitive as the model parameters are updated with every incoming data point. The state variable models are computationally superior alternative to AR modelling since this method yields comparable results at a lesser computational cost. Only difficulty in implementing the innovations' model identification algorithm is the need for determining the minimal order for the model since an incorrect model order may result in the divergence of the parameter.It is shown the plots of the EEG segments with the spikes introduced and after deconvolution by these methods. In each case the enhancement of the resolution of the spikes is quite evident. The comparatively superior performance of the recursive method in the processing of real EEG data is once again confirmed.

4.5.8 Conclusions

It has been demonstrated that ARMA models are better suited for the description of EEG real data. The method of predictive deconvolution can be employed for the detection of spikes present in the EEG segments. Both recursive and nonrecursive deconvolution techniques have been attempted on several segments of EEG real data. Efficacy of the methods had been compared with the help of quantitative and qualitative criterion for comparison. It is observed that recursive models yield better resolution of the spikes against the background activity and the noise at a higher computational effort. The nonrecursive innovations model (state model) is a better alternative to the AR modelling technique that has been used for modelling EEG segments. These desirable feature of the technique presented in this study is that non of them require any a priori assumptions except that of stationary of the data and elaborate steps are involved in the determination of a threshold based on ad hoc assumptions for the detection of spikes in a real EEG segment.

Further, recursive methods can be used for on-line detection of spikes in an EEG segment. It is expected that the proposed methods can help significantly in the reduction of subjectivity in the detection of spikes in EEG segments, which are symptoms of brain injury thromboses tumour and epilepsy.

A SPIKE OF 32 HZ INTRODUCED

Window No	**Non Recursive Method**		**Recursive Method**	
	AR(4)	**State**	**AR(4)**	**ARMA (4,1)**
1	2,11	1,42(2)	3,39	3,88
2	6,60	6,29(5)	3,96	5,65
3	1,79	2,84(5)	2,76	1,9
4	13,60	14,40(4)	18,70	16,28
5	7,66	9,13(4)	11,51	11,31

2 B SHARP SPIKE INTRODUCED

1	2,38	1,71(2)	4,29	4,35
2	8,68	9,17(4)	7,58	7,54
3	2,51	2,56(4)	3,99	2,62
4	17,40	19,96(4)	25,89	24,08
5	7,33	7,24(3)	11,24	10,90

TABLE 4.1

Numbers within parenthesis in Col. 3 indicate the system order identified.

CHAPTER FIVE
RELATED WORKS

5.1 Related Works

Several other groups are developing methods for eeg spike detection. Here we present a general summary of the available literature, followed by a more detailed review of three published papers in this field.

It is difficult to compare results from different papers reporting different methods but there is a set of standard statistical measures in addition to accuracy. In this review, and in presenting our own results, two of these measures shall be used - sensitivity and specificity.

These are defined as:

$$\text{sensitivity} = \frac{\text{correctly detected positives}}{\text{total actual positives}}$$

$$\text{specificity} = \frac{\text{correctly detected negatives}}{\text{total actual negatives}}$$

In this context a `positive' is a detected seizure and a `negative' is a detected non-seizure.

5.1.1 Literature Search

Automatic analysis of the human EEG for assisting in the diagnosis of epilepsy started in the early 1970s. In 1973 Prior et al. described a system which identified tonic-clonic seizures by detecting a large increase in EEG signal amplitude followed by a clear decrease, accompanied by high levels of muscle activity. Ives et al. applied amplitude discrimination to the sum of all 16 channels of EEG and was successful in detecting large seizures . Moving away from software methods, Babb et al. presented an electronic circuit to perform seizure detection.

From these beginnings the analysis of epileptic EEG has progressed in two main directions:

1. Seizure detection
2. Inter-ictal event detection

Some recent work has also been done on automatic systems to locate the origin of focalseizures.

Ambulatory EEG recording achieved widespread usage in the mid-1980s . Prior to this, most work in epileptic EEG analysis concentrated on the second area | the detection of inter-ictal events. This bias was probably due to the fact that long-term monitoring would take place in a dedicated unit; in these conditions it is relatively easy to identify seizures when they occur. Ambulatory recording, however, introduces the need to scan large amounts of EEG data for seizures, a time-consuming operation.

In 1982, Gotman presented a computerised system designed to detect a variety of different types of seizure. This method has been updated several times but remained basically unchanged over the 1980s. A more detailed description of some recent work by
Gotman. Interest in the issue of seizure detection has resurfaced during the '90s with a number of papers published by different research groups.

Neural network detection systems have been proposed by a number of researchers including Weng and Pradhan. Pradhan uses the raw EEG as input to a neural network while Weng uses the features proposed by Gotman with an adaptive structure neural network, but his results show a poor false detection rate (7 per hour).

In a very recent paper Osorio et al. have applied a wavelet transform to ECoG recordings . The wavelet scale used corresponds approximately to a 5{40 Hz band-pass filter. The output is squared, median filtered and finally compared with a background measure.

Perfect sensitivity and specificity is given for 125 seizure and 205 non-seizure examples. In a new development the paper also claims the ability to predict seizure onset | in 92% of the seizures investigated the detection took place shortly before the point marked by a clinician. This method claims to be generic, but on close inspection appears to be specific to a single group of patients. Detector parameters have been optimised with respect to this group and no testing has been performed on previously unseen individuals.

5.1.2 Qu and Gotman

Gotman was one of the first researchers to explore the possibilities of using an automated system to detect epileptic events in the human EEG - his first paper on the subject was published in 1976. He has published papers on similar topics on a regular basis since then and is cited in almost every piece of work in this and related areas. In a recent paper Qu and Gotman propose the use of a nearest neighbour classifier on EEG features extracted in both the time and frequency domains to detect the onset of epileptic seizures. Five features are used to characterise each 2.56 second epoch of EEG

1. Average wave amplitude
2. Average wave duration
3. Coeficient of variation of wave duration
4. Dominant frequency
5. Average power in the main energy zone

The first three of these features are calculated from Gotman's wave decomposition method which breaks the EEG down into half-waves and performs some smoothing. Feature vectors from both ictal and non-ictal EEG epochs are used as templates in the classifier. New EEG patterns are classified according to the closest template vector in feature space (although a threshold parameter is included in the definition to allow the system to be biased towards either ictal or non-ictal classification).

In addition to this variant on the nearest neighbour classifier, the system uses a location parameter to impose some spatial restrictions on where seizures should be detected. In efiect this requires seizures to be detected in several spatially close EEG channels before the whole system will signal it.

Using patient specific classifiers the paper claims a 100% detection rate with a false-alarm rate of 0.02 per hour. However, attempts to use this method as a generic detection system (training on one set of patients for use on a difierent set) gave very poor results. Although the patient-specific results in the paper are excellent there is little information about the types of seizure which are being detected. In addition the features extracted from the EEG are based on a method which is over 20 years old, developed at a time when computing power was not suficient to perform more complicated frequency domain analysis.

5.1.3 Gabor et al.

In their paper Gabor, Leach and Dowla state their aim to detect 85% of seizures with a false positive rate of 1 per hour or less using a generic system (compare this with the results given by Gotman for a specific detector).

Frequency domain features were extracted from the EEG data by passing the signal through a matched wavelet transform filter followed by a 256 point FFT. The resulting 256 coeficients were used as a feature vector for all further analysis. Features extracted from 98 examples of ictal EEG were used to train a self-organising map used for classification.

Testing was carried out using 62 seizures from 22 patients and the system identified 90% of the seizures with an average false positive rate of 0.71 per hour.

Although the results from this detection system are impressive there are a number of questions as to the methods employed. Given the dimensionality of the feature space the number of seizure examples used for training is more than likely to be insuficient. More importantly, however, is that the numerous empirical detection parameters employed were adjusted to optimize the performance of the detection algorithm". It appears that the method has not been tested on a previously unseen set of recordings without tuning.

5.1.4 Webber et al.

The Webber paper is the only one of the three reviewed here to use a standard multi-layer perceptron (MLP) structure neural network to classify the input EEG feature vector. Their network is a 31-30-8 structure with the 31 input features being various statistical measures of each two second EEG epoch. The input feature vectors were classified into eight groups including small seizure, large seizure, and normal. The neural network was trained using 8000 feature vectors equally distributed among the output classes. A separate set of 4000 vectors was used as a validation set. Due to the amounts of data available for training the feature vectors were extracted from EEG segments which overlapped by a considerable margin. Whether this overlap will afiect the validity of the results is debatable. Testing on recordings from 50 patients not used in the original training gave a sensitivity of 76% with an average false positive rate of 1 per hour. However, many of the results tabulated in the paper are taken from the training data set.

The use of 31 statistical features to characterise the EEG appears to be unnecessarily complicated since many of the values are highly correlated. This issue is raised in the paper, but no attempt is made to reduce the size of the feature vector. Finally a suggestion is made that increasing the number of EEG classes at the neural network output will improve the method. This runs against common sense as an increase in the complexity of the classification task is instead likely to make the problem much harder.

5.1.5 Tarassenko et al.

Tarassenko considered both time-domain parameters and frequency-domain parameters for the characterization of the EEG signal. In addition Tarassenko *et. al.* Have proposed that the duration of the spike with the maximum slope within an epoch be used as a feature. Gotman has suggested that the average of the parameters from the preceding five epochs should be used to normalize the paarmeters of the present epoch. Tarassenko et. al. Have employed these criterion in their design.

The frequency-domain parameters are determined by the AR modeling of the EEG. Tarassenko et.al. have pointed out that AR modeling overcomes the problem of traditional Discreate Fourier Transform (DFT), in which DFT produces a large number of coefficients from the signal, while ar modeling can effectively madel the signal with just a few coefficients. It has been reported that a model order six is suitable. It is also suggested that the "reflections coefficients" (which are produced as part of the Levinson – Durbin recursive procedure) should be used instead of the AR coefficients. Further, the prediction error of reflection coefficients can be used as an additional feature.

5.1.6 Kalayçi et. al.

Kalayci and Ozdamar suggested that the accuracy of the classifier could be improved by enlarging the input window size of the classifier. However the training of such neural network may become cumbersome and unrealistic for real-life systems. As a result, WT was used to generate a finite number of features. The WT overcomes the limitation of the short time Fourier transform by perfoming a multi resolution analysis of the signal. With the proper choice of mother wavelet the morphology of spikes can be effectively represented by the wavelet coeeficients. Kalayci and Ozdamar has attempted to use Daubechies-4 and Daubechies -20 mother wavelet (with the central eight coefficients from scale 3 and 4 as a feature vector) is an effective method for representing spikes in the EEG.

5.1.7 Ozdamar et. al.

While it is attractive to use parametric methods to extract features from the EEG data to identify spikes, Ozdamar and Kalayci have examined the efficacy of using raw EEG as the input to an ANN, to detect transient Epileptiform discharges in the EEG recordings. It has been shown by Ozdamar and Kalayci the raw EEG can be succesfully used to train ANNs and detect epileptogenic discharges with a high rate of success. It is suggested by them that parameterizing the signal may limit the efficiency of the system. Further it is argued that during training the ANN selects its own features for optimal detection and thus, the method does not use preselected features as in a traditional rule-based system.

5.1.8 Conclusions of Review

Although the research bias towards automatic detection of inter-ictal events is now being redressed, some of the more recent work in the area seems to have been carried out with unsatisfactory test protocols. Work on neural network detectors appears to have sufiered particularly badly in this way. In addition, with the exception of work using the wavelet transform, much of the feature extraction from the EEG uses techniques which fail to utilisethe potential of modern computing technology. The power of desktop computers is now such that sophisticated signal analysis tasks may be carried out on large amounts of input data at very high speed.

CHAPTER SIX

IMPLEMENTING TEMPLATE MATCHING ALGORITHM

6.1 Data Acquisition

The EEG data used in this study were acquired from 20 epileptic patients who had been under evaluation and treatment in the Neurology Department of Dokuz Eylül University Hospital, İzmir, Turkey. Data were obtained from a clinical EEG monitoring system, which stores continuous EEG data on its hard disk. EEG data were acquired with Ag/AgCl disk electrodes placed using the 10-20 international electrode placement system. The recordings were obtained from 19 channels with 256 Hz sampling frequency and band-pass filtered between 1 and 70 Hz. Data were then stored on both a hard disk and an optical disk.

6.2 Preparation of Templates

A convention adopted throughout this study is to use epileptiform discharge (ED) respectively to refer to epileptiform activity on a single channel and to refer to activity, which is simultaneously seen across two or more channels, actually on all channels, as epileptiform events (EV). Spike also refers to ED. All calculated performances throughout this study are determined in single channel.

20 EEG records were obtained. Two of them contain generalized epileptiform activity. The remaining EEG records contain focal epileptiform activity. The total EEG length is 11 h 6 min (average 22.1 min) and the age of the patients varies from 2 to 69 years (average 28 years). First, all EEG records had been previously seen independently by two EEGers and labelled as spike as a single channel epileptiform activity throughout the 19 channels.

It was used 20 EEG records. These EEG records have an average length of 22.7 min and total length of 7 h 18 min. The patients have an average age of 29 years. 216 EVs are determined by EEGers. The patients have an average age of 26 years.

Real EEG spikes were used to create the templates, which use to compare with in EEG segments to match process in the program. These spikes have been chosen from EEG segments by neurologist. Four different templates created. One of template is created from two spikes. These spikes are similer to each other. It was taken the average of the two spikes to create the template. Other three templates are real spikes. The four templates were created because these templates are differing from each other according to the amplitude, time duration, polarization properties. Each template was created from 41 points. Peak of the spike was located at the center of the templates. And then, it was created by taking 20 points from left, 20 points from right of the center of the spike.

Finally four templates were created with 41 points to use in the program. These template units can be increased or decreased according to the spike types. A spike has one form but these forms can be changed according to the patience. Such as amplitude, if the spike takes from the temporal lobe the amplitude will be high voltage but form still is same.

6.3 Matching Algorithm

Template Matching is the term given to the process of detecting an event buried in a signal by comparing it to a predifened "template". The goal is to locate possible events in the signal that "closely resemble" the template. In practice the methods which Summed of Square Error (SSE) to determine "how close" any given section of a signal is to the template.

The method first estimates the template by detecting the most prominent spikes. These template/templates are the shape that will be used by the rest of the algorithm to compare the signal against (using one of the two previous mentioned measures.) A threshold is then appilied to the output of the measured template function, much like a normal threshold detector, the result of which is used to identify detection events. The key difference between a threshold detector and a template based detector is over which function the threshold is applied. With normal threshold detectors, if the amplitude of the incoming data exceeds a threshold, an event detection is declared. With a template based detector if the output of the template error measure crosses a threshold, an event detection is declared.

In order to implement spike detection program, we need to select a suitable segment of EEG Data. In our experiment, the implementing data were selected through a short sampling window and all EEG signals were visually examined by qualified EEG technologist. 8 Segments were created from 2 patience because if two EEG segment use on the algorithm directly, it can not be readable format in the matlab for the neurologists to obtain the results. These real EEG segments were separeted to 8 segments because of that reason A neurologist's decision regarding a spike and wave complex (or normal EEG segment) was used as the gold standard. It was chosen as a template window with 41 points of the spike form. This width is that effective way to cover all spike forms. A spike has a pointed peak and duration of 20 to 70 msec. Using this criterion, 41 points use as template signal. Peak point of the spike placed center of the spike.

In order to assess the performance of the matching, it was selected EEG segments containing spikes and/or slow waves spike and wave complex, artifacts and background normal EEG. Two different pathological EEG Signals together with 19 channels were used to measure the sensitivity, specificity, selectivity and average detection rate of the algorithm.

6.4 Results

The system is evaluated using 19 channel clinical EEG records of 20 epileptic subjects. Two of them are used for testing purposes. The proposed system has been developed using MATLAB 6.0. The tests are performed on a Pentium Celeron 400 MHz PC computer. Detection procedure is performed off-line on data stored on hard disk.

Algorithm	True Detects	False Detects	Missed Events	SEN	SPE	SEL	ADR
TemplateMatching Template width = 41	258	282	16	94.16	70.36	47.77	82.26

Table 6.1 Test results

In the evaluation process, the false detection rate per hour is also calculated, which is a method to determine the performance of the system (Table 6.1). By definition, false detection rate is the number of false detections per hour. The false detection rate is an important measure of the performance of a detection system as it gives an indication of usefulness of a detection system will be in routine clinical applications. In addition, the measure false detection rate per hour of EEG can be used to place the reported performance of the system into context when considering the length of EEG records used in the test sets.

Single Channel		%	True Detects	False Detects	Missed Events
Sensitivity	The Best	100	15	0	0
	The worst	30	3	0	7
Selectivity	The Best	100	15	0	0
	The worst	0	0	12	0

Table 6.2 The best and worst results of the program

Table 6.1 gives the performance of the system. At each stage, the measures of the True detects, false detects, Missed events are indicated for 2 EEG records with 19 channels overall. It can be seen that the template matching results in the highest sensitivity for all EEG records with a sensitivity of 100%. On the other hand, however, it results in the lowest selectivity of 0 %. This shows that, most of spike activities are detected with very little selectivity and a too high false detection.

At the below table was mentioned the results of the developed algorithm and others algorithm

<table>
<tr><th colspan="2">Algorithm</th><th>SEN</th><th>SPE</th><th>SEL</th><th>ADR</th></tr>
<tr><td colspan="2">Taressenko</td><td>81.51</td><td>96.21</td><td>86.89</td><td>88.86</td></tr>
<tr><td colspan="2">Webber</td><td>83.17</td><td>97.66</td><td>91.75</td><td>90.42</td></tr>
<tr><td colspan="2">Kalayci</td><td>79.44</td><td>96.26</td><td>86.59</td><td>87.85</td></tr>
<tr><td colspan="2">Ozdamar</td><td>80.40</td><td>96.36</td><td>87.04</td><td>88.38</td></tr>
<tr><td colspan="6">Template Matching Algorithm</td></tr>
<tr><td rowspan="2">Min 1 template match condition</td><td>% 90 correlation</td><td>89.82</td><td>90.12</td><td>71.42</td><td>89.97</td></tr>
<tr><td>% 95 correlation</td><td>78.53</td><td>92.68</td><td>82.75</td><td>85.60</td></tr>
<tr><td rowspan="2">Min 2 templates match condition</td><td>% 90 correlation</td><td>82.71</td><td>94.37</td><td>83.15</td><td>88.54</td></tr>
<tr><td>% 95 correlation</td><td>80.13</td><td>95.7</td><td>86.39</td><td>87.91</td></tr>
</table>

Table 6.3 Comparing the results

Above results were written as global training method for other methods.

1) True Positive (TP): The Template Matching identifies a spike pattern that was labelled as a spike by the expert.

2) True Negative (TN): The Template Matching algorithm and the expert both agree that the pattern is normal.

3) False Positive (FP): The detection of a spike in an EEG segment that was labelled as normal by the expert.

4) False Negative (FN): The Template Matching has missed a spike that the expert has identified in that segment.

The performance of the classifier is also assessed in terms of sensitivity, specificity, and selectivity, as follows.

1) Sensitivity (SEN): A measure of the ability of the classifier to detect spikes

$$\text{Sensitivity} = \frac{TP}{TP + FN} x100\%$$

2) Specificity (SPE): A measure of the ability of the classifier to specify normal activities

$$\text{Specificity} = \frac{TN}{TN + FP} x100\%$$

3) Selectivity (SEL): A measure of the ability of the classifier to reject false detection of spikes

$$\text{Selectivity} = \frac{TP}{TP + FP} x100\%$$

4) Average Detection Rate (ADR): The average of sensitivity and specificity

$$\text{Average Detection Rate} = \frac{Sensitivity + Specificity}{2} x100\%$$

6.5 Discussions

In this thesis, it introduces a program based on Template Matching Method for the spikes detection in a single-channel electroencephalographic signal. In the first stage, four different templates were created from the spikes in EEG signals. This study shows that the template matching algorithm can proved good results for spike detection if the template spikes can be constituted well. The problem with that algorithm is that there are many different forms of spikes and artifacts therefore different templates should be used. The program detected a lot of false positive because of the artifacts. So the programing's selectivity decreased. This solution is not acceptable in real, but according to the neurologists important criterion is missing events. Missing spikes should be zero to use the program in the clinical applications. Other signals such as false positive can be eliminated manually in clinical applications. Developed program's averaged sensitivity is 94.16 %. This result can be improved if the artefacts can be eliminated and the template spikes created perfectly. Spikes are identified with 41 points. In this study four templates were used and tested. Another problem EMG artifact was detected as spike because detected artifacts have similar properties with the templates signals. For the normal waves, template matching algorithm produced good results. In addition to that the algorithm was be able to detecte most of the spikes.

Comparison of developed program with other detection systems given in the literature is difficult due to the wide range of measurements used for evaluating the performance. For example, Webber et al. have tested their system on the parameterized EEG records obtained from 10 patients, and reported satisfactory sensitivity and selectivity values (both at 74%) by using mimetic and ANN methods. When using raw EEG instead of parameters, they have obtained low sensitivity and selectivity values (both at 46%). Özdamar et al. have reported similarly good results for sensitivity (90%), but selectivity is relatively low (69%). Dingle et al. have given a very good result for false detection rate per hour (0) and selectivity (100%), although the sensitivity is relatively low (53%).

James et al. have also reported very good results for the selectivity (82%) and false detection rate per hour (7), but a relatively low sensitivity (55%).

. When the program compares with the systems mentioned above, it can be seen that it achieves very good sensitivity but because of the false positive value selectivity is not good according to the other results.

In conclusion, this study introduces a program based on Template Matching Algorithm for the spike detection that will contribute to the clinical applications with its good sensitivity levels. The proposed approach accomplishes template creating, matching algorithm of the single channel EEG. Comparison with other successful methods shows that, Template Matching is very useful procedure achieves a significant improvement in terms of sensitivity. But not on the selectivity and false detection rate.

CHAPTER SEVEN

REFERENCES

http://www.ucdmc.ucdavis.edu/neurology/Patients/eeginfo.htm

http://fhdno2.tch.harvard.edu/www/qeeg/qeegintro.html

http://web1.tch.harvard.edu/cfapps/A2ZtopicDisplay.cfm?Topic=Electroencephalograms%20(EEG%20or%20Brain%20Wave%20Tests)

http://www.eng.ox.ac.uk/World/Research/Summary/B-Neural.html

http://biocybernaut.com/tutorial/brainwave.html

http://www.diku.dk/undervisning/1999f/f99.134/Classification.html#Mahalanobis%20distance%20(MHD)

http://www.diku.dk/undervisning/1999f/f99.134/Convolution_and_Correlation.html

http://www.brunel.ac.uk/research/qi/nn/home.html

http://www.statsoftinc.com/textbook/stneunet.html

http://www.well.com/user/demaris/ch2.pdf

www.sciam.com/1096issue/1096lustedbox2.html

http://neurocog.psy.tufts.edu/images/ERP_components.gif http://www.inns.org/

http://www.bethel.edu/college/dept/mathcs/Neural/summary.htm

http://www.csis.hku.hk/~sfwong/Presentation/NNPR.pdf

http://www.ru.ac.za/academic/departments/statistics/textbook/stneunet.html

Brenner, R.P. & Scheuer M.L. 1998. Cross-country digital EEG survey 1998. Journal of Clinical Neurophysiology. Vol 15(6), p. 485-488

Babiloni, F., Babiloni, C., Carducci, F., Fattorini, L., Anello, C., Onorati, P. &Urbano, A. 1997. High resolution EEG: a new model-dependent spatial deblurring
method using a realistically-shaped MR-constructed subject's head model. Electroencephalography and Clinical Neurophysiology. Vol. 102, p. 6980.

Manu, P. 1994. Alpha-delta sleep in patients with a chief complaint of chronic fatigue. Southern Medical Journal. Vol. 87(4), p. 465470.

Kiloh, L.G., McComas, A.J. & Osselton, J.W. 1972. Clinical electroencephalography (3 rd ed.) London, Butterworth.

Jasper, H. 1958. The ten twenty electrode system of the international federation. Electroencephalography and Clinical Neurophysiology. Vol. 10, p. 371375.

Regan, D. 1989. Human brain electrophysiology - evoked potentials and evoked magnetic fields in science and medicine. New York, Elsevier Science Publishing.

Niedermeyer, E. & Lopes Da Silva, F.H. 1993. Electroencephalography. (3 rd ed) Maryland, Williams and Wilkins, Inc.

Template Matching: Matched Spatial Filters and beyond R. Brunelli [1], T. Poggio [2] 1 Istituto per la Ricerra Scientifica e Tecnological – 38050 Povo, Trente, Italy 2 Artificial Intelligence Laboratory Massachusetts Instute of Technology, Cambridge, Massachusetts 02139 ,USA

Training Algorithms for Robust face Recognition using a Template-matching Approach Xiaoyan Mu, Mehmet Artýklar, Metin Artýklar, and Mohamad H. Hassoun , Paul Watta

Faster Template Matching without FFT, Kimmo Fredrikson and Esko Okkonen, Department of Computer Science, University of Helsinki PO Box 26, FIN-00014 Helsinki, Finland

A simple and efficient template matching algorithm, Frederic Jurie and Michel Dhome

Applied Multivariate Methods for Data Analysis Dallas E. Johnson

Duda, R. O. and P. E. Hart (1973). Pattern Classification and Scene Analysis. New York: JohnWiley & Sons.

Van Gelder, T. (1990) Compositionality: A connectionist variation on a classical theme. Cognitive Science

Lewicki, P., Hill, T., & Czyzewska, M. (1992). Nonconscious acquisition of information. American Psychologist, 47, 796-801

Spehlmann's EEG Primer, Second Revised and Enlarged Edition ,Bruce J. FISCH, 1991 ELSEVIER Amsterdam, New York, Oxford.

Atlas of EEG, Volume One Methodology and Controls, Frederic A. GIBBS, M. D: and Erna L. GIBBS, Second Edition Massachusetts, U.S.A

EEG Spike Detection Using Deconvolution, National Systems Conference 1989, IIT, Kharagpur, Dec 1989, pp 59-62,(co-author: Santhosh Mathew)

APPENDIX I

Spike Detection Programs

1 Minumum 2 templates match condition with % 90 correlation

```
w1 = load ('patient.txt');  % load the eeg signal from the c hard drive
 w = w1 (:,channel_number);
 t1 = load ('spike1.txt');
 t2 = load ('spike2.txt');
 t3 = load ('spike3.txt');
 t4 = load ('spike5.txt');
 w = double(w);
[m n] = size (w);
 d={t1,t2,t3,t4};
 plot (w)
 xlabel ('Time (msec)')
 ylabel ('Amplitude (microvolts)')
 title ('EEG Signal')
 [k l] = size (t1);
   for J = 1:   n-l+1
     for I = 1 : m-k+1
        wt = w ( I : I+k-1 , J : J+l-1 );
        a = 0 ;
         for s = 1 : 4
           d{1,s} = double(d{1,s});
           [k l] = size (d{1,s});
           Err = sum ( sum ( d{1,s}.^2 ) + sum ( wt.^2 ) - 2 * sum (wt.*d{1,s}) );


if  (Err < 800500)
```

```
a = a+1;
if (a >=2)

    cor = (sum (d{1,s}.*wt) -(sum (d{1,s})*sum (wt)) / k ) / ( ((sum
(d{1,s}.^2)- ( (sum(d{1,s}))^2 ) / k)^0.5)*(( sum (wt.^2)- ( (sum(wt))^2 ) /
k)^0.5)) ;
    if (cor >= 0.9)
Hold on
plot (I : I+k-1 , J : J+l-1,'r+')
Hold off

      end
    end
  end
 end
end
end
```

2 Minumum 2 templates match condition with % 95 correlation

```
w1 = load ('patient.txt');  % load the eeg signal from the c hard drive
 w = w1 (:,channel_number);
 t1 = load ('spike1.txt');
 t2 = load ('spike2.txt');
 t3 = load ('spike3.txt');
 t4 = load ('spike5.txt');
 w = double(w);
[m n] = size (w);
 d={t1,t2,t3,t4};
 plot (w)
 xlabel ('Time (msec)')
 ylabel ('Amplitude (microvolts)')
 title ('EEG Signal')
 [k l] = size (t1);
   for J = 1:   n-l+1
     for I = 1 : m-k+1
        wt = w ( I : I+k-1 , J : J+l-1 );
        a = 0 ;
         for s = 1 : 4
           d{1,s} = double(d{1,s});
           [k l] = size (d{1,s});
           Err = sum ( sum ( d{1,s}.^2 ) + sum ( wt.^2 ) - 2 * sum (wt.*d{1,s}) );



if  (Err < 800500)
           a = a+1;


             if (a >=2)
```

```
cor = (sum (d{1,s}.*wt) -(sum (d{1,s})*sum (wt)) / k ) / ( ((sum
(d{1,s}.^2)- ( (sum(d{1,s}))^2 ) / k)^0.5)*(( sum (wt.^2)- ( (sum(wt))^2 ) /
k)^0.5)) ;
            if (cor >= 0.95)
    Hold on
    plot (I : I+k-1 , J : J+l-1,'r+')
    Hold off


          end
        end
      end
    end
  end
end
```

3 Minumum 1 template match condition with % 90 correlation

```
w1 = load ('patient.txt'); % load the eeg signal from the c hard drive
 w = w1 (:,channel_number);
 t1 = load ('spike1.txt');
 t2 = load ('spike2.txt');
 t3 = load ('spike3.txt');
t4 = load ('spike5.txt');
 w = double(w);
[m n] = size (w);
d={t1,t2,t3,t4};

 plot (w)
 xlabel ('Time (msec)')
 ylabel ('Amplitude (microvolts)')
 title ('EEG Signal')

 for s = 1 : 4

    d{1,s} = double(d{1,s});
    [k l] = size (d{1,s});

    for J = 1:   n-l+1

       for I = 1 : m-k+1

          wt = w ( I : I+k-1 , J : J+l-1 );
          Err = sum ( sum ( d{1,s}.^2 ) + sum ( wt.^2 ) - 2 * sum (wt.*d{1,s}) );
```

```
if (Err < 100500)
   cor = (sum (d{1,s}.*wt) - (sum (d{1,s})*sum (wt)) / k / ( ((sum (d{1,s}.^2) - ( (sum(d{1,s}))^2 ) / k) ^0.5)*(( sum (wt.^2)- ( (sum(wt))^2 ) / k)^0.5) ) ;

         if (cor >= 0.9)

            Hold on
            plot (I : I+k-1 , J : J+l-1,'r+')
            Hold off

         end
       end
     end
   end
end
```

4 Minumum 1 template match condition with % 95 correlation

```
w1 = load ('patient.txt'); % load the eeg signal from the c hard drive
 w = w1 (:,channel_number);
 t1 = load ('spike1.txt');
 t2 = load ('spike2.txt');
 t3 = load ('spike3.txt');
t4 = load ('spike5.txt');
 w = double(w);
[m n] = size (w);
d={t1,t2,t3,t4};

 plot (w)
 xlabel ('Time (msec)')
 ylabel ('Amplitude (microvolts)')
 title ('EEG Signal')

 for s = 1 : 4

    d{1,s} = double(d{1,s});
    [k l] = size (d{1,s});

    for J = 1:   n-l+1

        for I = 1 : m-k+1

            wt = w ( I : I+k-1 , J : J+l-1 );
            Err = sum ( sum ( d{1,s}.^2 ) + sum ( wt.^2 ) - 2 * sum (wt.*d{1,s}) );

                if  (Err < 100500)
```

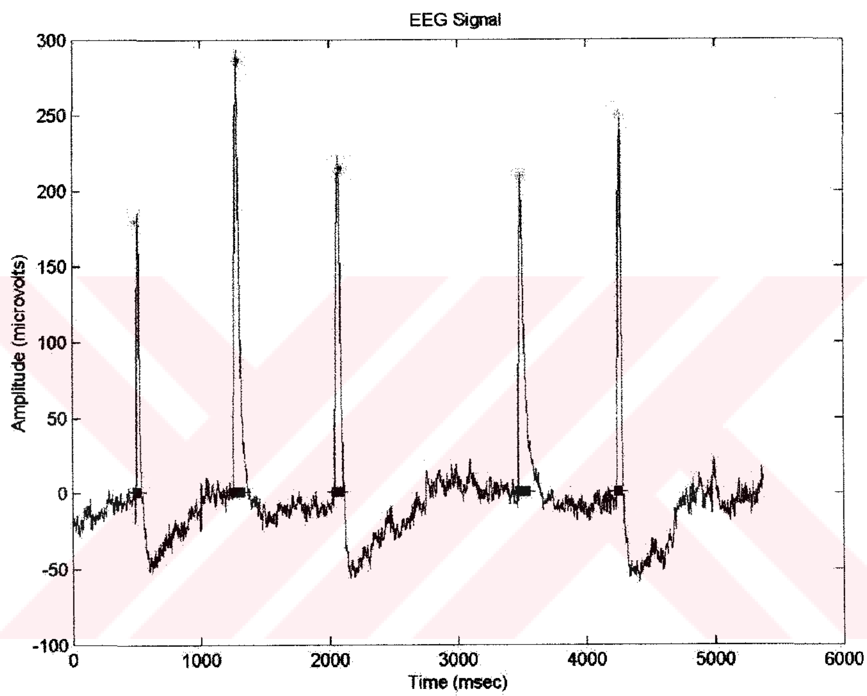
```
cor = (sum (d{1,s}.*wt) - (sum (d{1,s})*sum (wt)) / k / ( ((sum (d{1,s}.^2) - ( (sum(d{1,s}))^2 ) / k) ^0.5)*(( sum (wt.^2)- ( (sum(wt))^2 ) / k)^0.5) ) ;

                if (cor >= 0.95)
                    Hold on
                    plot (I : I+k-1 , J : J+l-1,'r+')
                    Hold off
                end
            end
        end
    end
end
```

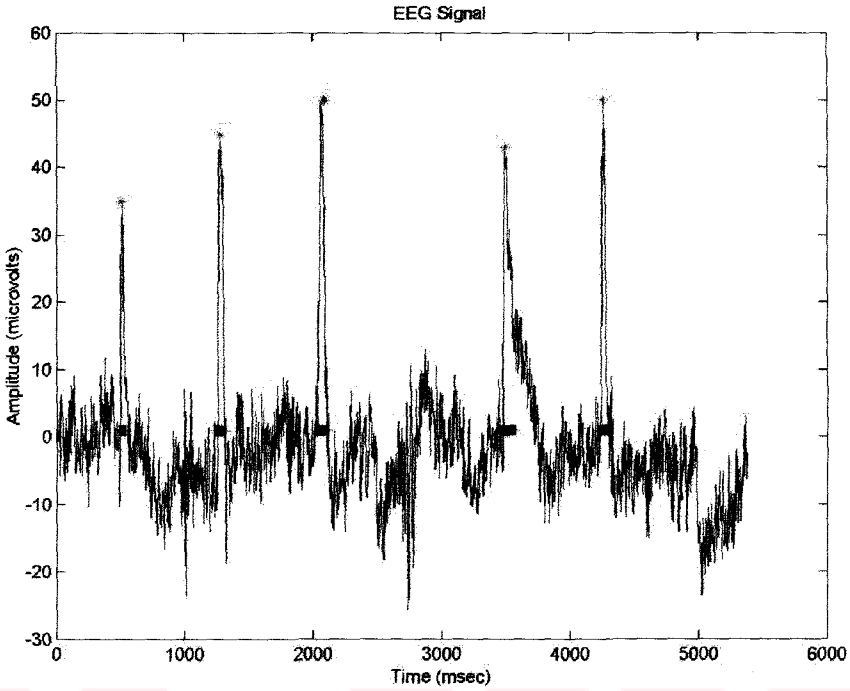
APPENDIX II

Detected Spikes

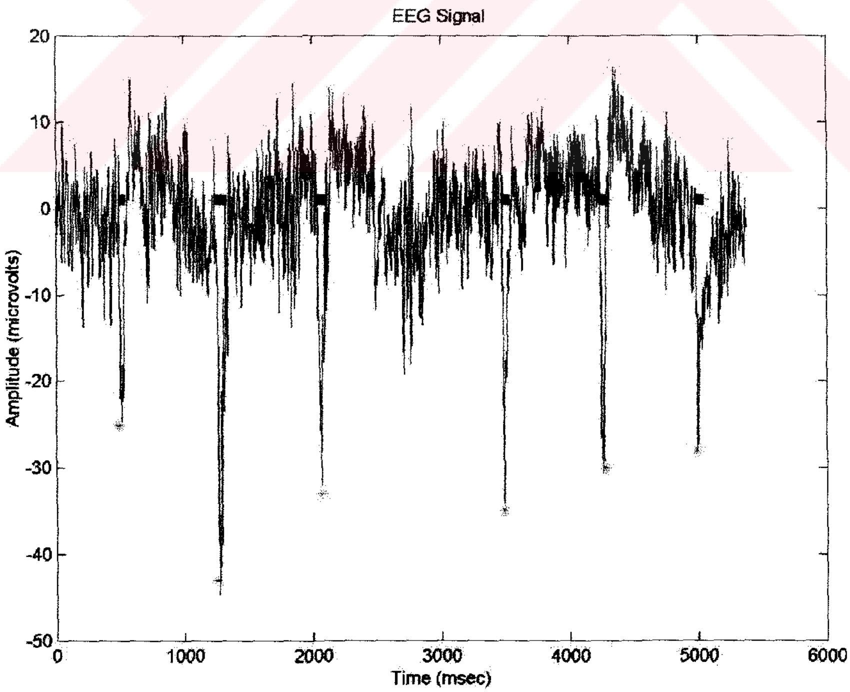
Some results were listed in below as JPEG format. Red points state the spikes in the EEG segment and green points state the real spikes which is identified by the expert.



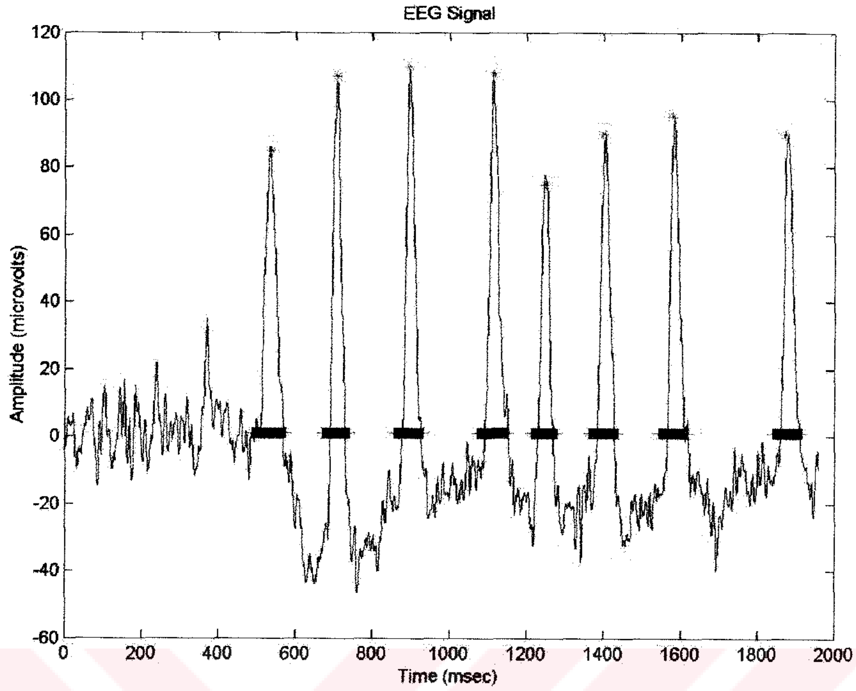
Patient I - Channel FP1



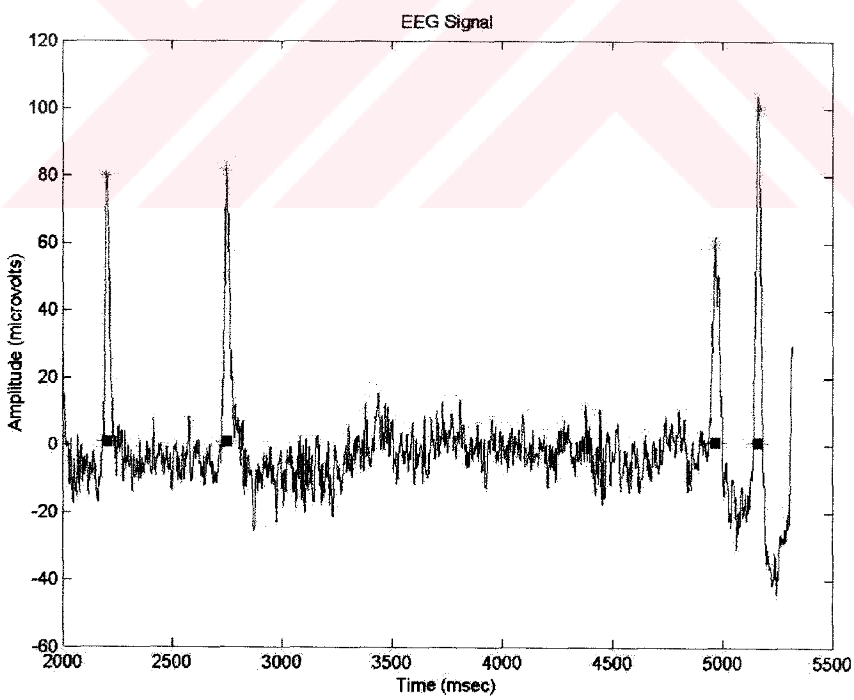
Patient I – Channel F3



Patient I – Channel T5



Patient II - Channel FP1



Patient II - Channel FP1