

*An Automatic EEG Monitoring System
in the
Pediatric ICU*

Yixia Si

*Department of Biomedical Engineering
McGill University, Montreal*

*A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of
Master of Engineering*

*July, 1996
© Yixia Si*



National Library
of Canada

Bibliothèque nationale
du Canada

Acquisitions and
Bibliographic Services Branch

Direction des acquisitions et
des services bibliographiques

395 Wellington Street
Ottawa, Ontario
K1A 0N4

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file *Votre référence*

Our file *Notre référence*

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-612-19882-0

Canada

Abstract

A knowledge-based expert system was developed to automatically assess the level of EEG abnormality of pediatric patients monitored in the ICU. The system receives six hours of 8-channel EEGs and classifies the background EEG as one of seven abnormality levels.

A total of 188, six-hour EEGs were visually interpreted by a neurologist and used as training examples. Spectral band activity was computed; artifacts were rejected using a median filter with a hard-limiter. Quantitative variables reflecting amplitude, symmetry, Front/Back differentiation and time variability were then extracted based on the study of Pasupathy (1994). Relationships between quantitative measures and the neurologist's assessment of amplitude, symmetry and Front/Back differentiation were established. A two-layer neural network having the measures of EEG variability as input was created for variability evaluation. A single-layer network was constructed to give the integrative interpretation of EEG abnormality based on the neurologist's assessment of the four features. Suitable knowledge base and inference engine were also constructed.

Performance was tested using the rotation method of error estimation. 45% of testing instances were classified the same as the neurologist's interpretation. 46% were classified with an error of one abnormality level. Possible improvement and the clinical future of the system are discussed.

Résumé

On a développé un système expert à vase de connaissances pour déterminer automatiquement le niveau d'anormalité de l'EEG de malades sous monitoring dans l'unité pédiatrique de soins intensifs. Le système reçoit six heures d'un EEG de 8 voies et doit classer l'activité de fonds en sept niveaux d'anormalité.

Un total de 188 EEGs durant chacun 6 heures ont été interprétés visuellement par un neurologue et utilisés comme exemples d'apprentissage. On a calculé l'activité dans divers bandes de fréquence; les artéfacts ont été éliminés grâce à un filtre médian associé à un seuil d'amplitude. En se basant sur le travail de Pasupathy (1994), on a calculé des variables reflétant l'amplitude, la symétrie, la différenciation antéro-postérieure et la variabilité temporelle. On a déterminé les relations entre ces mesures et le jugement du neurologue quant à l'amplitude, la symétrie et la différenciation antéro-postérieure. On a utilisé un réseau neuronal à deux couches pour obtenir une mesure globale de variabilité temporelle à partir des mesures élémentaires de variabilité. L'interprétation complète de l'anormalité de l'EEG a été ensuite faite à l'aide d'un réseau neuronal à un niveau, ayant comme entrées les différentes mesures. On a aussi construit un système à vase de connaissances et un moteur d'inférence.

La performance globale a été évaluée par la méthode de rotation. 45% des échantillons sont classés au même niveau d'anormalité que celui donné par le neurologue, et 46% sont classés à un niveau d'anormalité en dessus ou en dessous. On discute des améliorations possibles et de la future utilisation clinique d'un tel système.

Acknowledgments

I would like to take this opportunity to thank Dr. Jean Gotman, Montreal Neurological Institute, for his supervision of my research project and for his help and support over the last two years. His invaluable comments and unending support have always inspired me to make progress in my project and helped me improve the quality of my thesis.

I would like to thank Dr. Bernard Rosenblatt, Montreal Children's Hospital, for reading the EEGs and for many useful discussions. I also appreciate the help extended by Dr. Gottesman, for his efforts to arrange the EEG recordings in the intensive care unit of the Montreal Children's Hospital.

I would also take this moment to extend my sincere thanks to Dr. Danny Flanagan for his helps of re-reading all the EEGs, his invaluable suggestions about my project and my thesis as well as his much abused patience while I tried to understand children's EEG.

As well I would like to thank Anitha K. Pasupathy for her talented former study which forms the basis of my project and my thesis. I also thank Carol Aksich for collecting and carefully documenting the EEG records for this project.

The help offered by Jing Zhang for computer-related problems and Rajeev Agarwal for helping me with writing the thesis are also deeply appreciated.

I should also acknowledge the support provided by the Medical Research Council of Canada and Stellate Systems through the university-industry program (MRC-UI-11514).

Finally, the love, understanding, and support my husband and my parents have always given me, have proven to be a source of strength for me over the years.

Contents

ABSTRACT	II
RÉSUMÉ.....	III
ACKNOWLEDGMENTS.....	IV
1 INTRODUCTION.....	I
1.1 NEUROPHYSIOLOGIC MONITORING IN THE ICU	1
1.2 ELECTROENCEPHALOGRAM.....	3
1.2.1 <i>Why use the EEG for Monitoring</i>	3
1.2.2 <i>Quantitative EEG</i>	6
1.2.3 <i>Abnormal EEG in Children</i>	8
1.3 PROJECT DEFINITION.....	12
2 DATA PRE-PROCESSING AND FEATURE EXTRACTION	15
2.1 DATA ACQUISITION AND PRE-PROCESSING	15
2.1.1 <i>Data Acquisition</i>	15
2.1.2 <i>Artifacts</i>	17
2.1.3 <i>Artifact Rejection</i>	20
2.2 FEATURE SELECTION.....	23
2.2.1 <i>Measure of Amplitude</i>	24
2.2.2 <i>Measure of Left/Right Symmetry</i>	26
2.2.3 <i>Measure of Front/Back Differentiation</i>	27
2.2.4 <i>Measure of Variability</i>	29
2.3 STATISTICAL ANALYSIS	32
2.3.1 <i>Population Distribution</i>	32

2.3.2	<i>T-Statistic</i>	37
2.3.3	<i>Assessment of Variability</i>	41
2.3.4	<i>Data Reduction</i>	42
3	KNOWLEDGE-BASED EXPERT SYSTEM	47
3.1	MACHINE LEARNING FROM EXAMPLES.....	48
3.2	FUZZY MEMBERSHIP LEARNING.....	52
3.2.1	<i>Fuzzy Sets Theory</i>	52
3.2.2	<i>Construction of Membership Function</i>	53
3.3	NEURAL NETWORK LEARNING.....	60
3.3.1	<i>Why Use the Neural Network Models</i>	61
3.3.2	<i>Learning in Single-Cell Models</i>	65
3.3.3	<i>Learning in Multi-Layer Models</i>	71
3.3.4	<i>A Network for Variability</i>	74
3.4	KNOWLEDGE BASE AND PERFORMANCE.....	76
4	RESULTS AND DISCUSSION	78
4.1	TRAINING AND TESTING DATA.....	78
4.2	RESULTS.....	80
4.2.1	<i>Results of Fuzzy Membership</i>	80
4.2.2	<i>Results of Final Evaluation Network</i>	83
4.2.3	<i>Results of EEG Variability Network</i>	85
4.2.4	<i>Results of the Entire System</i>	86
4.3	DISCUSSION.....	88
	REFERENCES	99

Chapter 1

Introduction

1.1 Neurophysiologic Monitoring in the ICU

In the present day intensive care units (ICU), while numerous catheters, transducers, digital readouts and alarms keep continuous vigil on a patient's heart, lungs and kidneys, the brains are "monitored" routinely by intermittent clinical observations, or "neuro-checks". Jordan discussed these examinations by stating (Jordan, 1995): "Even under the best circumstances and when carried out by expert personnel, 'neuro-checks' are discontinuous and subjective. All too often, they are unable to anticipate capricious clinical deterioration. Contrary to the goal of a true monitoring system, 'neuro-checks' must wait until after functional deterioration occurs for problems to be recognized. The constraints on this type of clinical monitoring become more serious when patients are sedated, medically paralyzed, or placed in barbiturate coma."

During the past few years, neurophysiologic monitoring in the ICUs has received increased attention. The number of bedside neurophysiologic monitoring methods have increased and thus have been the subjects of numerous clinical research studies. In addition to well-established intracranial pressure monitoring, several newer methods including continuous electroencephalography, evoked potentials, transcranial Doppler sonography, cerebral blood flow, and jugular venous oxygen saturation have been brought to the ICU bedside to measure different physiologic variables. They extend the

clinician's power of observation in order to detect physiologic abnormalities at a reversible stage.

A successful ICU monitoring system should meet the following criteria (Jordan, 1993):

- * be more sensitive and specific than clinical observation.
- * be non-invasive.
- * be easily operated and interpreted by non-experts.
- * be usable at the bedside.
- * not interfere with medical or nursing care of the patients.

Many neurointensivists share the opinion that the electroencephalography (EEG) can become an integral part of monitoring in the ICU (Emmerson and Chiappa, 1988). EEG recordings have been useful in the investigation of various disorders, subclinical seizures and coma. The technique itself is non-invasive, and advances in technology have made possible the collection, storage and analysis of continuously collected EEG. However, the raw EEG generates cumbersome amounts of data, and its complexity discourages interpretation by non-experts. Several methods have been developed to compress EEG data and simplify its interpretation with definite benefit, such as compressed spectral array (CSA) (Bickford et al., 1971), the cerebral function monitor (Maynard, 1979) and the topographic brain mapping (Nuwer, 1988).

EEG monitoring is currently in extensive use during carotid surgery (Chiappa et al., 1979; Cho et al., 1986) due to the fact that EEG changes correlate with regional ischemia during carotid artery endarterectomies (Trojaborg and Boysen, 1973). Continuous EEG monitoring for patient management has begun to be used routinely in a small number of

centers (Jordan, 1990, 1992). Jordan and Stringer (1991) have reported a decisive impact of EEG monitoring on clinical management decisions in 85% of the patients monitored subsequent to cardiac surgery. EEG monitoring of patients in coma has been performed which provides clues to the cause of coma as well as prognostic information (Cant and Shaw, 1984). EEG monitoring has also been used for monitoring barbiturate therapy for increased intra-cranial pressure with large benefit (Ropper and Rockoff, 1983).

1.2 Electroencephalogram

The electroencephalogram (EEG) represents the spatial and temporal summation of inhibitory and excitatory postsynaptic potentials which occur in the cerebral cortex (Jordan, 1995). It can be recorded from the scalp surface by electrodes attached to the scalp with conductive jelly. The arrangement of these electrodes on the surface of the scalp is based on an international standard of localization (the 10-20 system, Jasper, 1958). The EEG is recorded as a potential difference between pairs of electrodes and each such pair is referred to as a "channel". The combination of electrodes and channels examined at a particular point in time is referred to as a "montage". After amplification, the continuous analog EEG can be displayed using paper write-out and oscilloscopic display. In long-term EEG monitoring systems, computer-based digital EEG is normally implemented due to its efficient means of data storage and transmission and subsequent data processing ability.

1.2.1 Why use the EEG for Monitoring

There are several neurobiologic reasons for using the EEG as a monitor of brain activity (Jordan, 1993, 1995):

- *EEG is linked to cerebral metabolism.* The EEG represents the spatial and temporal summation of excitatory and inhibitory postsynaptic potentials and their modulation by subcortical projection fibers. At the cellular level, these interactions reflect the underlying state of cerebral metabolism, which in turn depends on multiple factors including synthesis of enzymes and energy (Siegel et al., 1989). A disturbance in one or more of these components will produce a disturbance in EEG. Hence the EEG is a highly sensitive, although non-specific, indicator of cerebral dysfunction.
- *The EEG is sensitive to ischemia and hypoxia, and detects neuronal dysfunction at a reversible stage.* The EEG is mainly generated by the cortex pyramidal neurons which are selectively vulnerable to hypoxia and ischemia. The EEG begins to deteriorate before cell membrane failure and irreversible reduction of ATP occurs (Heuser and Guggenberger, 1985). Astrup et al. (1981) have demonstrated that the EEG abnormalities occur before the cerebral blood flow (CBF) drops to 20-25 ml/100g/min, whereas synaptic activity continues until the CBF drops to 17 ml/100g/min, and that energy failure and cell death are not seen until the CBF falls to 10-12 ml/100g/min. These observations imply the presence of a "therapeutic window" in cerebral injury following the appearance of EEG abnormalities. Also, the correlation of EEG changes with cerebral ischemia during carotid artery endarterectomies has been observed by Trojaborg and Boysen (1973). This relationship forms the basis of the current use of EEG monitoring during carotid surgery (Chiappa et al., 1979; Cho et al., 1986).

- *EEG is the best available method for detecting seizure activity.* For many years, continuous EEG outside the ICU has been used in epilepsy units to detect and localize seizure activity (Gumnit, 1987). Engel (1989) reported a significant incidence, ranging from 10% to 27%, of acute seizures following head injury, spontaneous intracranial hemorrhages and ischemic stroke. A systematic study of continuous EEG monitoring among the patients in ICU has documented a surprisingly high incidence of nonconvulsive seizures (Jordan, 1992).
- *EEG correlates with cerebral topography.* The international 10-20 system for electrode placement establishes a reliable relationship between scalp electrode placement and underlying cerebral topography (Homan et al., 1987). This attribute has been de-emphasized because of the precise localization obtained by radiological methods such as CT and MRI. However, for the ICU patient, transport for imaging studies is logistically difficult and can be hazardous. The EEG provides useful information about cerebral localization at the bedside.

These attributes of the EEG have heightened interest in its application as an electrophysiologic monitor for the patients with acute brain injuries. In addition to these neurobiologic reasons, the technique of EEG recording from the scalp is non-invasive and EEG monitoring can be easily set up at the patient's bedside without any interference with patient care and medication. The interpretation of raw EEG data, however, requires skilled personnel who may not be available at all times in the ICU. Many researchers in the EEG monitoring field try to overcome this shortcoming by using different data

processing techniques. Among these techniques, quantitative EEG is the earliest and most widely used.

1.2.2 Quantitative EEG

A neurophysiologic EEG monitor should continuously record brain activity of the patients in the ICU over several hours. In order to detect the cerebral dysfunction at a reversible stage, on-line interpretation of the EEG is crucial. However, the presence of a neurologist onsite throughout the entire period of recording is quite impractical and the complexity of the EEG discourages the interpretation by non-expert. In practice, there is a considerable time lag between recording of the EEG and actual interpretation which largely reduces the effectiveness of EEG monitoring. Also, visual interpretation of long-term EEG recording is quite tedious and time consuming and it is difficult to evaluate the long-term trend of the EEG.

Quantitative EEG (QEEG) relies on the transformation of digital EEG signals into mathematically derived parameters. These parameters are then interpreted by statistical analyses based on comparison with data obtained from a population considered to be normative. The information of QEEG is displayed in a variety of formats including topographic scalp maps, bar graphs, CSA and density spectra array.

In addition to reducing data and simplifying the interpretation, QEEG can also facilitate the identification of intervals of physiologic sleep, gradual shifts in dominant frequency and long-term EEG trends of prognostic significance. In recent studies, considerable work has been done to estimate the sensitivity, reliability and usefulness of QEEG for the prognosis of nervous system disorders. In several situations, QEEG is

found to be more sensitive than conventional EEG. Nuwer (1988) has reported that QEEG using computer-assisted frequency analysis is more sensitive than conventional visual analysis in detecting abnormalities in focal cerebral ischemia. QEEG abnormalities were found in 94% of 54 patients with completed or partial stroke compared to only 60% found by visual EEG assessment. Also, Ritchdin et al. (1992) reported that QEEG was more sensitive than conventional EEG, magnetic resonance imaging or computed topography scanning as an indicator of cerebral dysfunction in systemic lupus erythematosus patients. QEEG is also found to be able to increase the sensitivity of some conventional techniques. Szeliés et al. (1994) compared quantitative topographical EEG with regional glucose metabolism measured by PET with respect to the sensitivity in classification of mild to moderate dementia. The results showed that combined use of QEEG and PET was more discriminating and reached higher diagnostic specificity than each test individually. Moreover, QEEG is demonstrated to be an indicator of brain perfusion and regional cerebral blood flow (Leuchter et al., 1994; Jibiki et al., 1994), and to be useful for the detection of encephalopathy and unilateral cerebral lesions (Jacobson et al., 1993; Salinsky et al., 1992).

Until now, most QEEG techniques are used in clinical laboratories. There are several limitations to using QEEG to monitor ICU patients and perform automatic diagnosis, such as abundant artifacts, fluctuating patient states of alertness, lack of normative ICU data, and difficulty in detecting spikes, brief seizures, burst suppression and periodic lateralized epileptiform discharges. These limitations, however, may be overcome and there are a number of efforts to do so. By new analytic procedures and signal processing techniques, artifacts can be recognized and often removed (Prior, 1973; Bickford et al.,

1971; Maynard and Jenkison, 1984). Spikes, seizures and lateralized epileptiform discharges can also be detected by successfully developed automatic seizure detection systems (Gotman, 1982, 1990).

The computer-assisted analysis based on QEEG has been used by several researchers, however the most sensitive variables to be monitored have not been determined. The tool most commonly used is the CSA designed by Bickford et al. (1971). Here the frequency spectra of the EEG activity are computed and plotted against a vertical time scale for successive epochs. Gross changes can therefore be identified visually. Maynard and Jenkison (1984) developed a system called the Cerebral Function Analyzing Monitor. This device bandpass filters (2-20Hz) the EEG, performs amplitude rectification and smoothing, and then computes five amplitude measures and the percentage of activity in nine frequency bands which are displayed as a function of time. The Vital Signs Monitoring System (Chiappa and Hoch, 1993) is another computer-assisted diagnostic tool. This system plots the trend of the EEG over several hours. Different physiological parameters are calculated and displayed such as peak, median power and spectral edge frequency, frequency bin activity totals, frequency bin ratios and intracranial pressure.

1.2.3 Abnormal EEG in Children

EEG abnormalities in children are classified by Lombroso (1985) into three main categories: (1) abnormalities of background activity, (2) ictal abnormalities, and (3) abnormalities of states and of maturation.

Background activity refers to the basic EEG rhythms which are present at all times. EEG background abnormalities correspond to deviations in amplitude and frequency

composition of these basic rhythms in different head regions. Lombroso (1985) reported that background abnormalities appear to be most suitable for diagnostic applications especially in long-term prognosis. Other studies also indicate that background EEG is a good indicator of prognosis subsequent to hypoxic-ischemic injury if recorded for sufficient duration in all states (Watanabe et al., 1980). The prognostic significance is influenced by the timing of the recording. Kayser-Gatchalian and Neundorfer (1980) demonstrated that recordings within the first 48 hours of the illness serve as the best prognostic indicators.

Ictal abnormalities are most often associated with seizures which appear as sharp wave and spikes, hypersynchronous rhythmic activity or as paroxysmal slow wave activity. In children, ictal abnormalities provide prognostic information as well and can be caused by the involvement of the central nervous system or by metabolic derangement.

The composition of EEG activity in children varies considerably with age. Children who exhibit sleep rhythms that are uncharacteristic of their age are considered to have abnormalities of organization in sleep states and maturation. This category consists of more subtle deviations, some of which are still being investigated and less well established.

Among the three categories of abnormalities discussed above, background abnormalities are the best prognostic indicator in long-term EEG monitoring especially subsequent to hypoxic-ischemic injury. An analysis of background abnormalities together with ictal abnormalities may give maximum prognostic information, although the processing of these two patterns may be independent. Background EEG abnormalities manifest themselves in several forms:

- *Inactive Pattern* - This is characterized by cerebral activities below 10 μ v almost continuously throughout the recording, and unreactive to stimulation. In the absence of a postictal state, notable hypothermia, high levels of drugs, or severe derangement in blood gases, this pattern carries a very unfavorable prognosis.
- *Burst Suppression* - Characterized by periods of inactive background (lower than 10-15 μ v) interrupted by synchronous or asynchronous bursts of activity. The bursts can be of high or, less often, low voltage, last 0.5-6s, contain irregular slow waves with or without interspersed sharp transients. In the absence of the factors mentioned above, a burst suppression pattern heralds an unfavorable outcome reaching high statistical predictability (Lombroso, 1985).
- *Low Voltage Pattern Through All States* - This exhibits mixed frequencies lower than 10-50 μ v more or less continuously and in all states. This pattern differs from inactive pattern only in the degree of voltage attenuation and from burst-suppression pattern because of the absence of bursts. If persistent and in the absence of scalp edema or technical faults (smearing of electrode paste, sweat, etc.), the prognosis of unfavorable clinical outcome reached statistically significant values in long-term prospective studies (Lombroso, 1974, 1978).
- *Interhemisphere Amplitude Asymmetry* - This consists of a persistent voltage asymmetry in background rhythms between hemispheres (Lombroso, 1980). Transitory or mild asymmetries are probably of no pathological significance. A persistent voltage asymmetry recognizable in various states could denote a depression in one hemisphere or large amplitude activity in the other, both of which

point to an underlying abnormality. These abnormalities can be intraparenchymal hemorrhages (Lombroso, 1981); pre- or postnatal vascular accidents; tumors or, rarely, congenital malformations. The diagnosis of this pattern needs to be very careful because the non-pathologic factors such as scalp edema or technical faults (asymmetric electrode placement, smearing of electrode paste, sweat, etc.) can produce similar asymmetries.

- *Monotonous Pattern* - This consists of an almost invariant diffuse pattern present at all times, poorly reactive to stimuli. Bricolo and coworkers (1973) have demonstrated that an invariant EEG carries a worse prognosis than a cycling (alternating) EEG. In another study, they also found that approximately 95% of the patients with a slow and monotonous CSA had unfavorable outcomes compared to only 30% of those with a changing CSA (Bricolo et al., 1978). Although the cycling EEG patterns only weakly correlate with clinical state of the patient, they still have significant independent prognostic value (Rumpl et al., 1979). To study spontaneous variations adequately, the EEG must be recorded over an extended period of time.
- *Absence of Frequency-Amplitude Gradient* - The EEG of a normal child shows a decrement in voltage from posterior to anterior head region, with an accompanying decrease in slow frequencies in the same posterior-anterior direction. There is a positive correlation between the severity of illness and the absence of this gradient (Slater and Torres, 1979). The same study also found that the presence of this

gradient is dependent on the patient's age. In most cases good gradients do not appear before 3 to 4 months of age and after 10 years old.

In addition to background abnormalities and seizures, clinical neurologists are most often interested in the trend of the EEG over long periods (at least several hours), especially subsequent to brain injury, since it is a good indicator of state of deterioration or recovery.

1.3 Project Definition

The aim of the project is to build a system for automated neurophysiological monitoring in the pediatric ICU. The system is to be used as a bedside EEG diagnostic tool for pediatric patients who may suffer neurologic dysfunction such as those subsequent to cardiac surgery, trauma or hemorrhage.

Such a monitor should mimic the neurologist's behavior, accepting several hours of raw EEG data as input and classifying it as normal or abnormal based on its characteristics. As background abnormalities appear to be the best diagnostic indicator in long-term EEG monitoring, several features describing the background abnormalities have been selected to be used by the monitor for interpretation of the EEG.

Most of the EEGs used to build this system were recorded in the pediatric ICU (at Montreal Children's Hospital) to monitor the possible neurologic dysfunction of the cardiac patients subsequent to cardiopulmonary bypass. Due to the regional blood flow variation during surgery, these patients may suffer brain hypoxia-ischemia and may serve as a good test group for this system. The long-term EEGs of patients who are in coma or have seizures are also used to develop this system.

Chapter 2 first introduces the EEG recording protocols used in the Montreal Children's Hospital including montage, amplification, analog filter and sampling rate. Mathematical procedures are then presented which transform the raw EEG data into the frequency band arrays on which artifact rejection and feature extraction procedures are performed. Continuous EEG recording over long periods is subject to various kinds of artifact especially in the ICU where the patient care and medication are essential. Artifact rejection is therefore crucial for accurate EEG analysis. Common types of artifact are discussed in detail and illustrated. Various filters are designed and compared with each other based on their ability to reject artifact. Then the quantitative equivalents of selected qualitative features are extracted from the frequency band arrays of the EEG. The obtained parameters are further interpreted by statistical analysis based on a population considered to be "normative". The work discussed in this chapter is based on a study by Anitha K. Pasupathy (1994) and several modifications are introduced to improve the performance of the system.

Chapter 3 introduces the construction of a knowledge-based expert system which accepts the parameter values associated with the selected features and classifies the EEG into one of the several abnormality levels. The system is trained with examples of six-hour EEG sections interpreted by the neurologists. The feature extraction method described in chapter 2 is then used to calculate the quantitative values of the selected features for each six-hour EEG. By using fuzzy membership and neural network techniques, the system learns and updates its knowledge about the EEG classification by comparing its performance with the neurologists' assessment. The structure of the knowledge base and the inference engine are explained as well.

The performance of the automated monitor system is tested and described in chapter 4. The discussion of the possible improvement and the clinical future of the system are also included in this chapter.

Chapter 2 Data Pre-processing and Feature Extraction

This chapter describes the EEG data acquisition protocol, artifact rejection approach and feature extraction methods used to obtain the quantitative interpretations of the select features which describe the EEG background abnormalities. Most of the work in this chapter is based on a study by Anitha K. Pasupathy (1994) and several modifications are introduced to improve the performance of the system.

2.1 Data Acquisition and Pre-processing

2.1.1 Data Acquisition

The EEGs were recorded with an 8-channel montage called "Little H" in the intensive care unit of the Montreal Children's Hospital. As depicted in Figure-2.1, eleven electrodes were glued to the surface of the scalp. The EEG recorded from channels 1 and 4 corresponds to the activity from the frontal head regions of the left and right hemisphere respectively. Similarly, channels 2 and 5 represent central parietal head regions, channels 3 and 6 represent the posterior head regions and 7 and 8 represent the central temporal head regions.

After being picked up by the electrodes, the analog signal is amplified by a factor of 10000 and filtered by a low pass filter (cutoff frequency of 30Hz) to prevent aliasing and a high pass filter (cutoff frequency of 1Hz) to remove artifacts due to respiration.

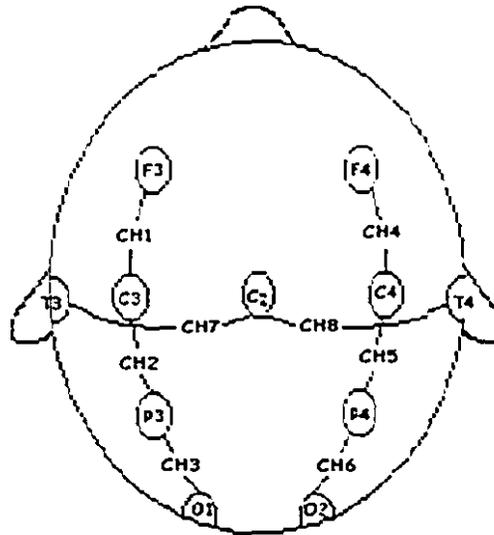


Figure-2.1 "Little H" montage

sweating, etc. Then, the analog signal is digitized at a 200 Hz sampling rate and displayed on a computer monitor using the software MONITOR (Stellate System, Montreal).

Visual interpretation of the EEG involves mainly the assessment of its frequency composition. The neurologists estimate visually the amount of EEG activity in the various frequency bands, delta (1-3 Hz), theta (3-7 Hz), alpha (7-14 Hz) and beta (14-30 Hz), compares and correlates them and arrives at a decision about the normality of the EEG. It would therefore be quite useful to transform the entire EEG into the frequency domain. Since analysis of the background activity is best achieved by studying long-term trends, it is preferable to analyze an extended EEG record lasting several hours. An 8-channel EEG recorded continuously over a period of six hours produces approximately 60 MB of data. Transformation of the EEG into the frequency domain has the added benefit of data reduction.

The frequency analysis of the EEG record is performed using a software called ECLIPSE (Stellate System, Montreal). The program divides the entire EEG recording

into epochs of 30 seconds, each of which is referred to as a 'page'. The Fast Fourier Transform is computed for every 512 sample points (2.56 seconds of EEG) and this amounts to eleven frequency domain distributions per channel per page. In fact only 28.16 of the 30 seconds are used. The frequency distribution for each page of a channel is obtained by averaging its 11 distributions. The band activity is then the average of the amplitudes of the activity within the frequency range of the corresponding band. The frequency ranges of the various bands are defined in Table-2.1.

Band Titles	Delta	Theta	Alpha	Total
Frequency Range(Hz)	1.17-3.13	3.52-7.03	7.42-13.28	1.17-13.28

Table-2.1 Frequency ranges of different activity bands.

The total band calculated here spans only up to 14Hz, although the frequency of genuine cerebral rhythms can be as high as 50Hz. The reason behind such a narrow band will be discussed in the section on Artifact Rejection. Once the "activity" of the EEG is computed, each channel is associated with four average amplitudes corresponding to the four frequency bands for every page of the record. The activity values for a frequency band are then plotted against time as a 'band array' for each channel. Activity computation by ECLIPSE reduces 6 hours of raw EEG data from 60MB to about 500KB.

2.1.2 Artifacts

Artifacts are frequent and often intractable during long-term EEG recordings in the ICU since the prime concern there is the support of patient's life and, particularly in the early hours after surgery, many procedures both diagnostic and therapeutic are in

progress. The effect of these artifacts on computer analysis is quite serious and their rejection is essential for authentic interpretation of the recording.

One common type of artifact is 60Hz power artifact, which is due to interference from nearby equipment or the ground loop. This 60Hz artifact also has harmonics at 120Hz, 180Hz, etc.. Since the signal is digitized at 200Hz and the analog anti-aliasing filter has not completely eliminated the frequency components above 100Hz, the activity at 120Hz aliases and appears as an activity peak at 80Hz and 180Hz activity aliases as 20Hz. The 20Hz artifact contaminates the beta band activity. Figure-2.2 illustrates 6 seconds of an EEG with 60Hz artifact most prominently visible in the left posterior head regions (channel p3-c1).

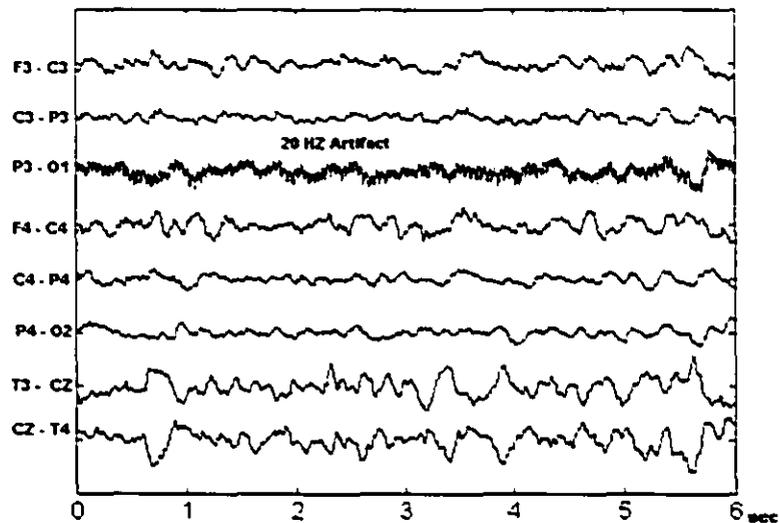


Figure-2.2 Six seconds of an 8-channel EEG recorded by using "little H" montage. Channel P₃-O₁ contains 60Hz artifact.

Another important source of artifact is patient movement, including body movement, contraction of the scalp muscles, blinking, chewing, coughing, swallowing and involuntary myoclonic jerks. Muscle contractions due to chewing, coughing and

swallowing produce broad band activity referred to as the Electromyogram (EMG) while the movement of the eyeball produces another kind of activity called Electro-oculogram (EOG), which is caused by the fact that the eyeball is an electrical dipole. Both EMG and EOG are quite useful for behavioral studies, however, these rhythms are not of interest and contaminate the EEG when monitoring cerebral function. The movement of the conductors that carry current from the scalp electrodes to the amplifier, caused by the body movement, can induce an electric current due to the earth's magnetic field. This induced current is comparable to the currents of cerebral origin and therefore contaminates the EEG signals. These patient-generated artifacts are usually spiky, containing sharp elements and very large amplitudes, and impulsive, not lasting for more than a few seconds in duration. Figure-2.3(a) illustrates an EEG contaminated by patient-generated artifact caused by the movement of the conductor of electrode P₃ and Figure-2.3(b) shows the total band activity of this EEG where the very large amplitude spiky components correspond to the patient generated artifact shown in (a).

Artifact due to poor electrode contact is quite inevitable during a long-term recording and an example is illustrated in Figure-2.4(a). During long-term EEG recording, the jelly that fills the gap between the electrode and the scalp to make electrical contact could dry up and this may impair the contact. Artifact due to poor contact is characterized by low frequency and moderately high amplitude. Such artifacts however, are not impulsive and remain until the electrode is glued again. Figure-2.4(b) shows the total band array corresponding to the EEG with poor electrode contact depicted in Figure-2.4(a).

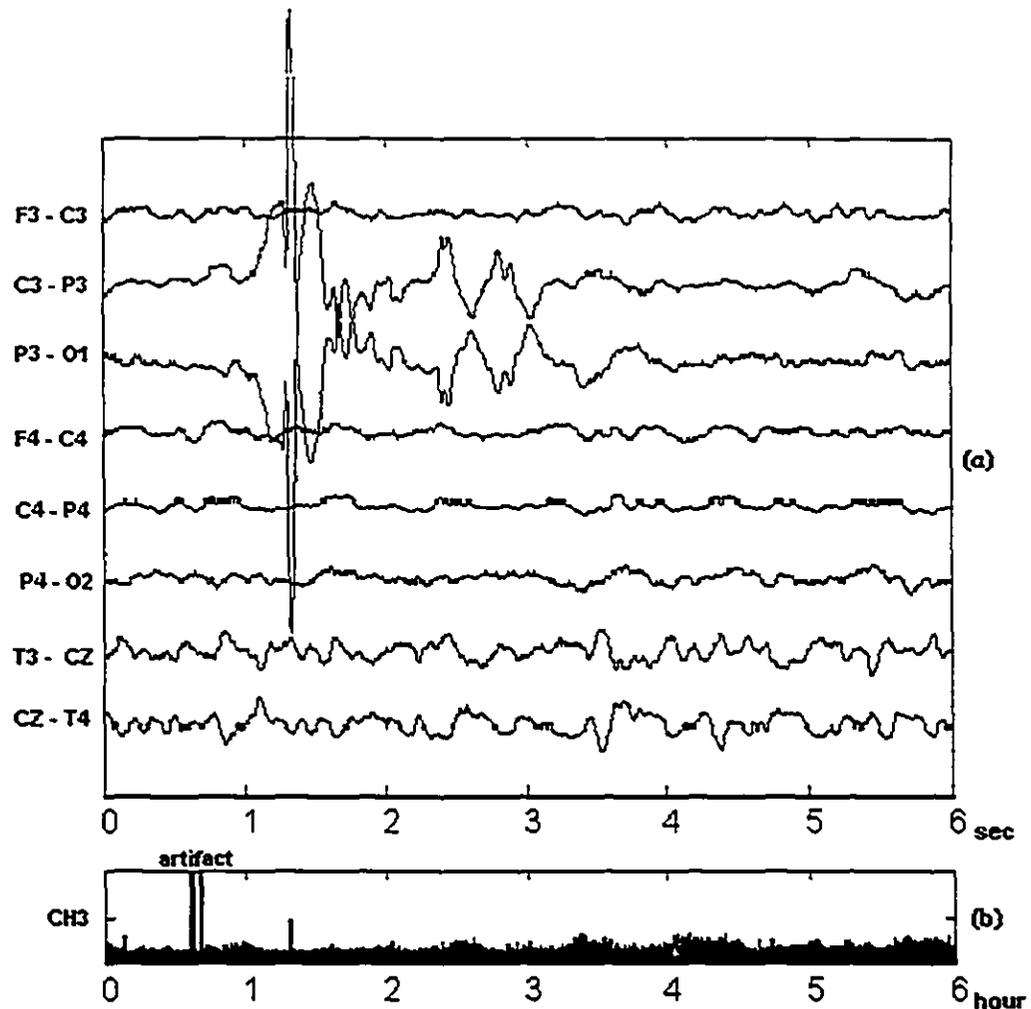


Figure-2.3 Patient-generated artifact. (a) Six seconds 8-channel EEG: artifact can be seen in channels C_3-P_3 and P_3-O_1 due to the movement of the conductor of common electrode P_3 ; (b) The total band activity of channel P_3-O_1 in which the sharp spike corresponds to the artifact shown in (a).

2.1.3 Artifact Rejection

As discussed earlier, the aim of this diagnostic tool is to determine the gross state of the EEG on the basis of its frequency composition over several hours. It is quite unnecessary, therefore, to identify and reject specific artifactual EEG waves. Artifact identification and rejection from the frequency band arrays would be sufficient for our purpose. It is important to realize that artifact identification can be done visually only by

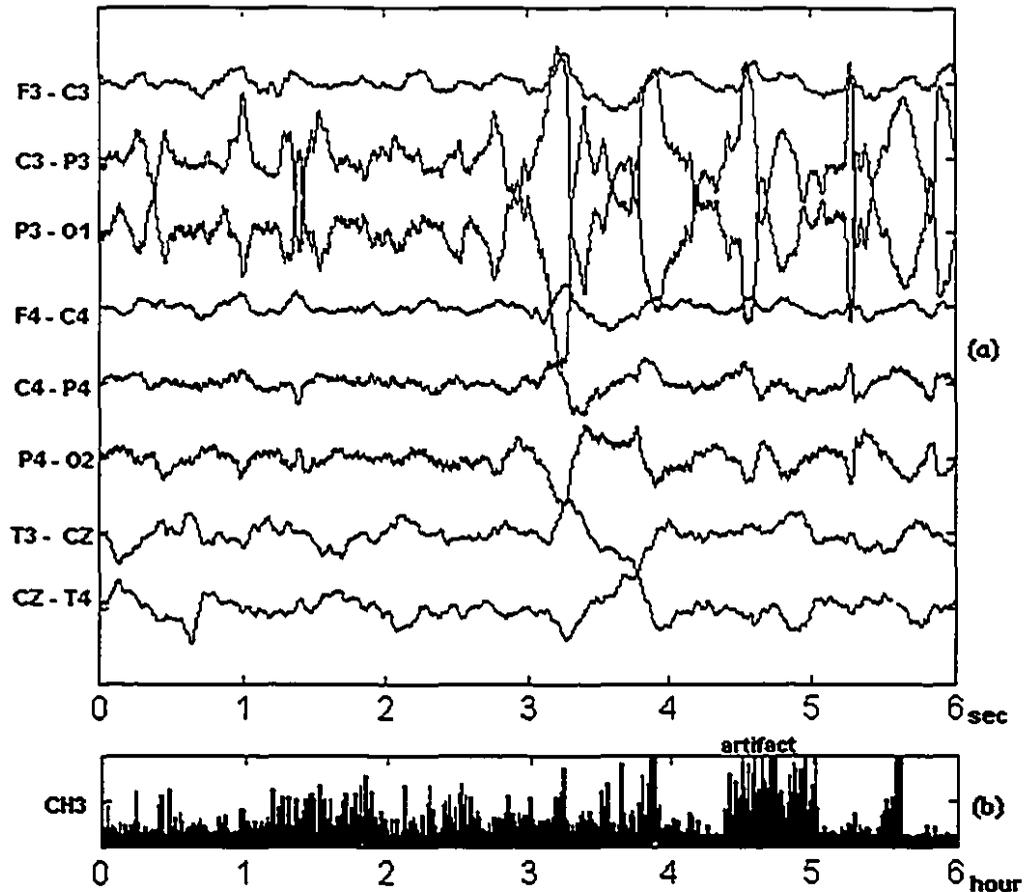


Figure-2.4 Sustained artifact. (a) six seconds 8-channel EEG: artifact is observable in channels C_3-P_3 and P_3-O_1 , due to the poor electrode contact; (b) The total band arrays of channel P_3-O_1 , in which the marked activities correspond to the artifact shown in (a).

examining the raw EEG record in the time domain. Suppression of artifacts, however, can be done in the frequency domain.

The frequency range of background EEG of cerebral origin is 0-30Hz. Faster rhythms of cerebral origin up to 50Hz may be present but usually correspond to seizure activity. The beta frequency band (14 to 30 Hz) is commonly affected by medication, administration of anesthetics and 20Hz alias of 60Hz noise. To remove the effect of the non-pathological factors, the 60Hz artifact and its harmonics, the total band analyzed is from 0-14Hz. Since most background EEG abnormalities manifest themselves in the

delta and theta frequency ranges, such a definition of the total band is not likely to distort the interpretation in a significant way.

To reject the artifact due to patient movement and poor electrode contact, a filter capable of rejecting both impulsive very high amplitude artifact and sustained moderately high amplitude artifact is required. Today, linear filters are used extensively. However, they fail to perform well with both the impulsive noise component and the sustained noise component. A non-linear filter called the median filter has been used successfully for the impulsive noise component. It replaces the value at each point by the median of the signal values in some finite neighborhood about that point. The performance of a 5-point median filter on an EEG with impulsive artifact and on one with sustained artifact is illustrated in Figure-2.5.

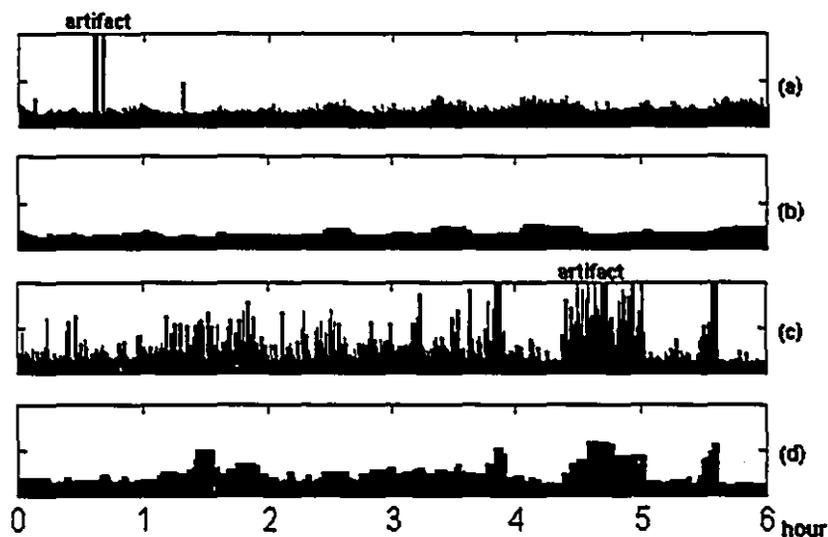


Figure-2.5 Artifact rejection by median filter. (a) The total band array of an EEG containing impulsive artifact; (b) The band array shown in (a) filtered by a 5-point median filter; (c) The total band array of an EEG containing sustained artifact; (d) The band array shown in (c) filtered by a 5-point median filter.

It is clear that the median filter is efficient at rejecting impulsive noise components, but fails to remove the sustained artifact. Sustained artifacts in EEG recording can continue over several hours. The median filter fails to recognize them when the entire neighborhood is artifactual and similar. Therefore a hard threshold should be used.

A median filter with hard-limiter threshold performs well in the rejection of sustained artifact. The data was first filtered with a 5-point median filter. If the median of any window of total band activity is greater than a threshold, then the activities in all frequency bands of that particular epoch for that channel are replaced by the respective average band activities averaged up to that point in time. The choice of the threshold is quite critical for the performance of the filter. Extensive EEG review indicates that genuine fluctuations of cerebral origin in frequency band activity are rarely greater than 25% of the averaged background activity while the amplitude of most artifacts is at least 1.5 times the average EEG background. Therefore, the threshold chosen was 1.5 times the average EEG activity. Figure-2.6 depicts the performance of a median filter with a hard-limiter. Clearly, the hard-limiter threshold rejects both sustained and impulsive artifacts quite effectively. It needs to be mentioned here that the user should ensure that at least several minutes of “clean” EEG are present at the beginning of the recording so that the filter is able to recognize the artifact by comparison with this reference.

2.2 Feature Selection

Once artifacts have been suppressed or rejected, the relevant features that will provide information about the EEG background can be extracted. From the discussion on abnormal EEG pattern, it is apparent that the nature of background EEG activity is

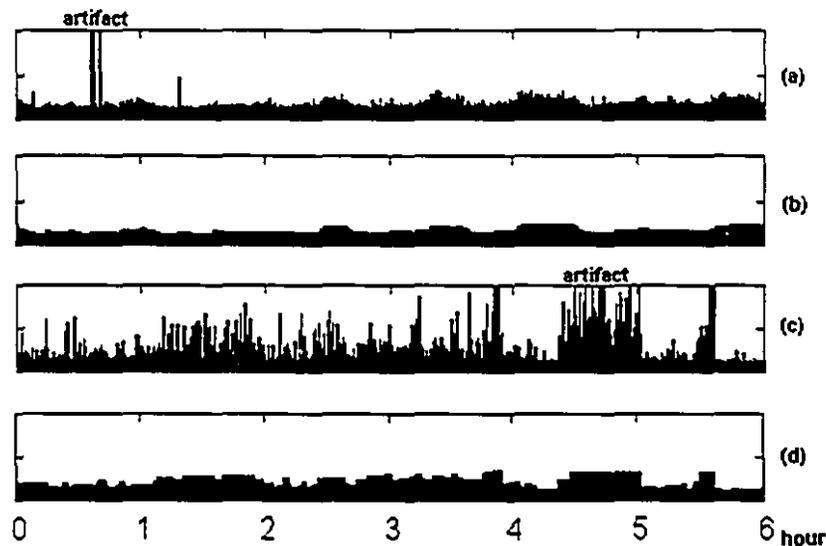


Figure-2.6 Artifact rejection by median filter with hard-limiter. (a) The total band array of an EEG containing impulsive artifact; (b) The band array shown in (a) filtered by a 5-point median filter with hard-limiter of 1.5 avg; (c) The total band array of an EEG containing sustained artifact; (d) The band array shown in (c) filtered by a 5-point median filter with hard-limiter of 1.5 avg.

primarily assessed on the basis of four aspects, namely, amplitude, Left/Right symmetry, Anterior/Posterior gradient and variability of the EEG.

To study spontaneous variability over time and assess long-term trends of the EEG, it is necessary to analyze an extended EEG record. However, to be most useful, it is important to obtain information about an abnormal EEG pattern as early as possible. Taking both these factors into consideration, it was decided that the classification of the EEG would be done based on 6 hours of recording. The mathematical derivations of the quantitative features extracted from an EEG prior to classification are discussed below.

2.2.1 Measure of Amplitude

A depressed EEG record is characterized by low amplitude values, which can be reflected in the total band activity of the EEG, as depicted in Figure-2.7. Since sustained rather than impulsive amplitude abnormalities are evaluated, the amplitude measure is

mathematically derived as the logarithm of the average over a 5-minute period of the total band activity. In a band array, this represents the logarithm of the average of 10 points since basic calculations are made every 30 seconds. The reason for using the logarithm of the average rather than a simple average is explained in the section on statistical analysis. As a total, an amplitude measure extracted every 5 minutes for 8 channels over 6 hours amounts to $12 \times 6 \times 8$ or 576 values.

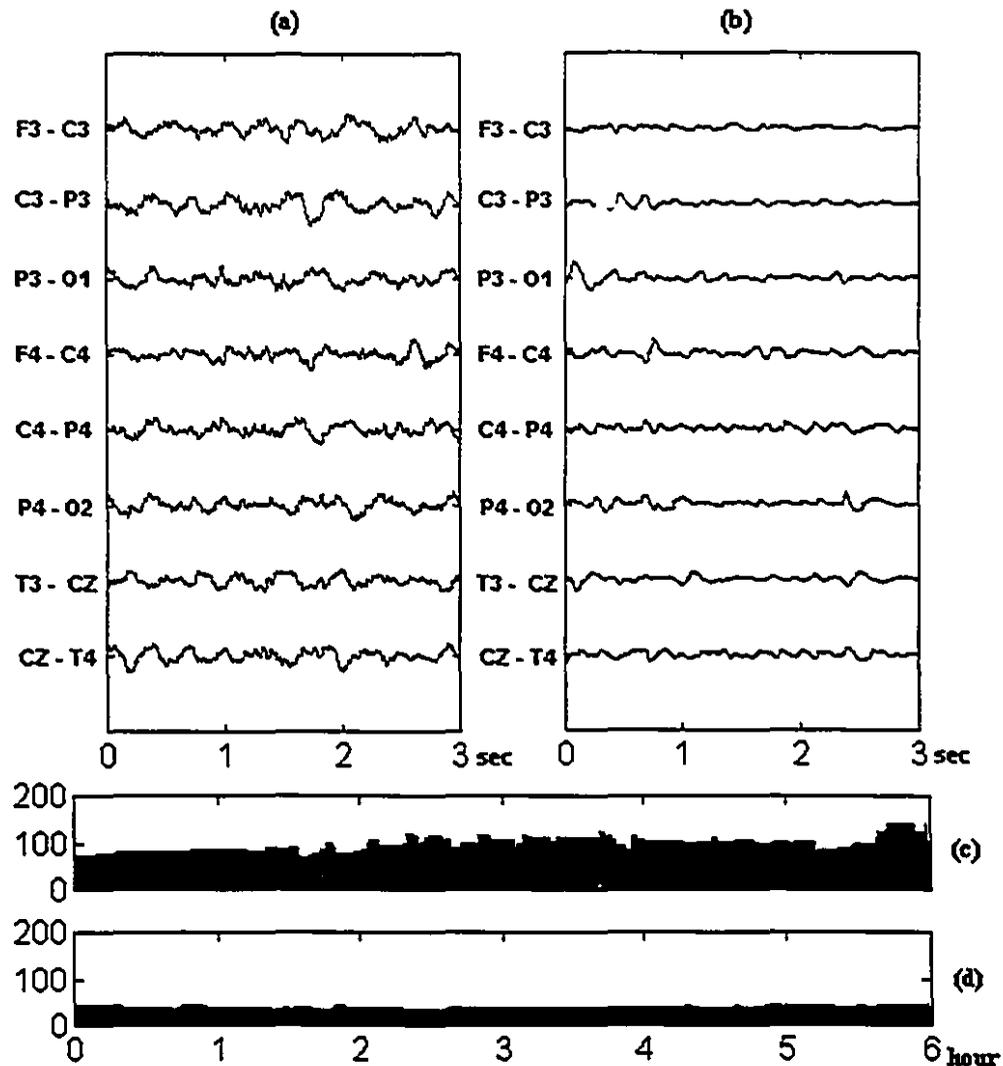


Figure-2.7 Amplitude depression. (a) Three seconds 8-channel EEG with normal amplitude; (b) Three seconds 8-channel EEG with depressed amplitude; (c) Total band activity of channel P₃-O₁ of the EEG shown in (a); (d) Total band activity of channel P₃-O₁ of the EEG shown in (b).

2.2.2 Measure of Left-Right Symmetry

Figure-2.8 illustrates two EEG sections from two patients, one with Left-Right symmetry and one without. Their total band activities are shown in Figure-2.9. A simple ratio of activity between corresponding channels of the left and right hemispheres could quantify the level of symmetry. Since a ratio value closer to "1" implies a symmetrical EEG pattern, the extent of asymmetry is proportional to the absolute difference between 1 and the ratio value. The formula for the Left/Right symmetry is given as:

$$R = \log\left(\frac{\text{Averaged activity of channel in left hemisphere}}{\text{Averaged activity of the corresponding channel in right hemisphere}}\right).$$

Here, averaged activity represents averaged total band activity over a 5-minute period.

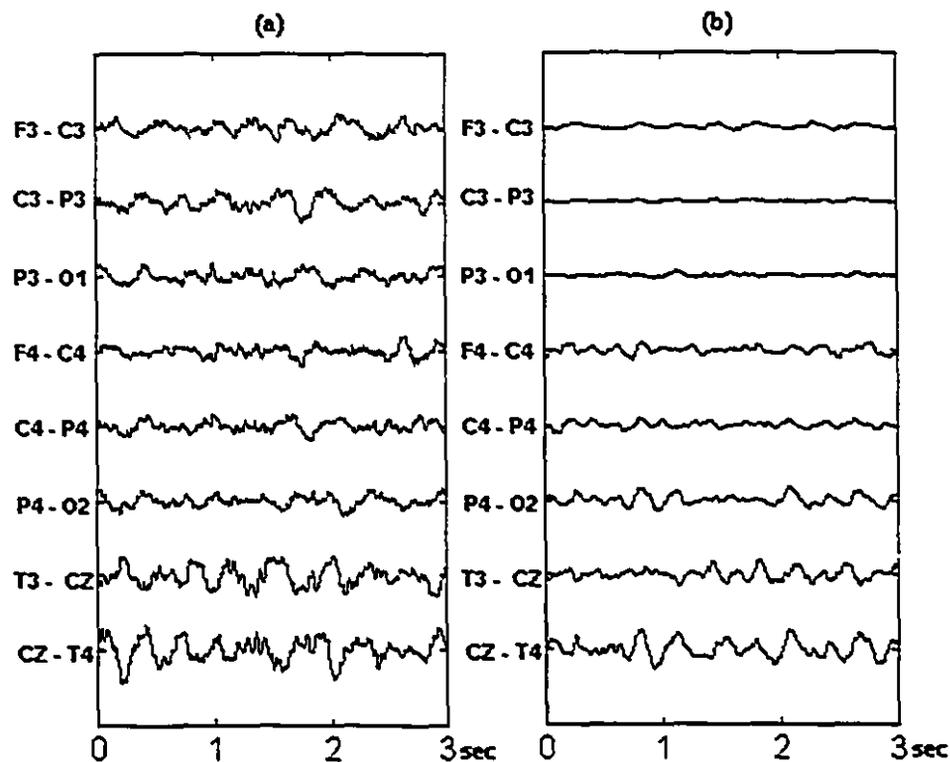


Figure-2.8 Left-Right symmetry. (a) Three seconds 8-channel EEG with Left-Right symmetry; (b) Three seconds 8-channel EEG of asymmetrical. The activity of left hemisphere is depressed.

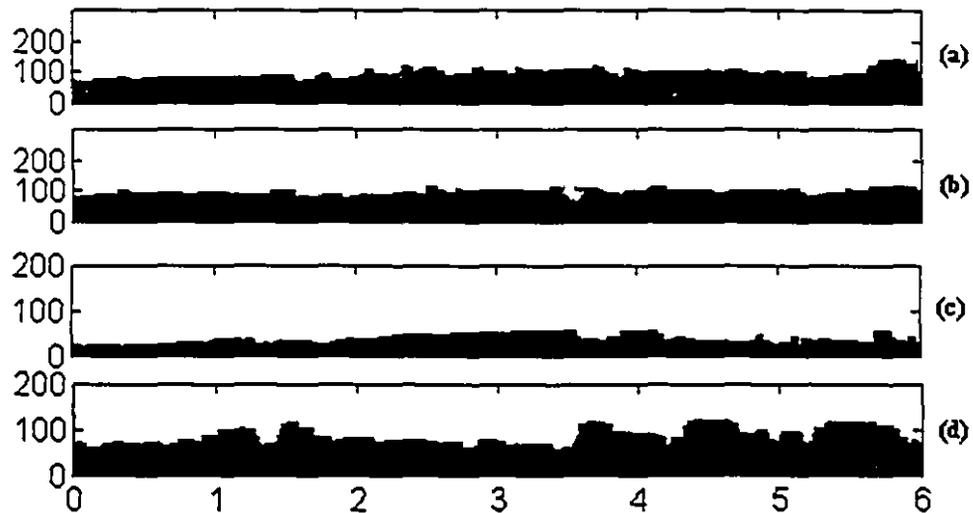


Figure-2.9 Total band activities of the EEGs in Figure-2.8. (a) (b) are the activities of channel P_3-O_1 and channel P_4-O_2 of the EEG shown in Figure-2.8(a); (c) (d) are the activities of channel P_3-O_1 and channel P_4-O_2 of the EEG shown in Figure-2.8(b).

Four Left/Right symmetry measures corresponding to the four pairs of channels symmetrically located in the two hemispheres are extracted every 5 minutes of an EEG record. A quantified 6 hour EEG record is thus associated with 72 Left/Right symmetry values for each of the four channel pairs, or 288 values.

2.2.3 Measure of Front/Back Differentiation

As explained in chapter 1, a normal pediatric EEG record is associated with a gradient in amplitude of activity in the low frequencies with amplitude decreasing in the posterior to anterior direction. This is referred to as Front/Back differentiation. Figure-2.10 illustrates two EEG records and their delta band activities, one with Front/Back differentiation and one without. A ratio of the delta band activity of the posterior channel to that of the anterior channel of the same hemisphere reflects effectively the extent of their differentiation. Two such measures are extracted for every EEG record

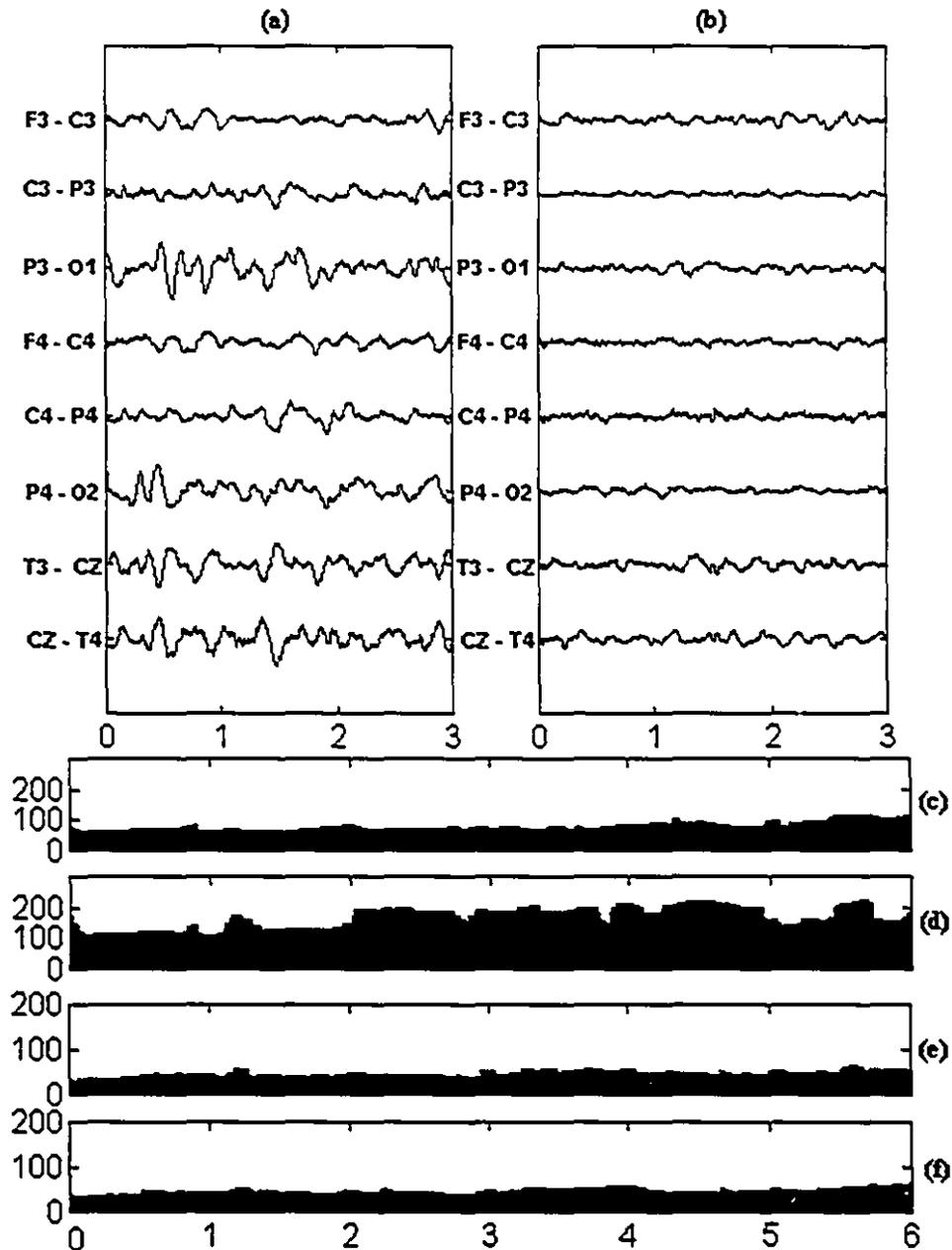


Figure-2.10 Front/Back differentiation.(a)Three seconds EEG with F/B dif.; (b)Three seconds EEG without F/B dif.; (c)(d)Delta band activities of channel F_3-C_3 and channel P_3-O_1 , of the EEG shown in (a); (e)(f)Delta band activities of channel F_3-C_3 and channel P_3-O_1 , of the EEG shown in (b).

corresponding to the two hemispheres. The parameter is derived using the formula given below.

$$R = \log\left(\frac{\text{Averaged delta activity of posterior head region}}{\text{Averaged delta activity of anterior head region of the same hemisphere}}\right).$$

Averaged delta activity in the formula above corresponds to an average over a 5-minute period. For each hemisphere, the Front/Back differentiation extracted from every 6-hour EEG record is described by 72 values. The total number of values is therefore 144.

2.2.4 Measure of Variability

This measure assesses the extent of spontaneous cycling in an EEG record. The delta band activity of the posterior channels of the two hemispheres is most reflective of a cycling EEG.

Figure-2.11 illustrates three posterior delta band activities subsequent to artifact rejection, one with spontaneous variability (a) and two others (b, c) without. Spontaneous alteration in Figure-2.11 (a) appears in the band array as humps that are significantly higher in the amplitude and extend over a substantial period. Figure-2.11 (c) shows an absolutely flat band activity with no big humps. A solitary large hump such as the one presented in Figure-2.11 (b) is not representative of variability either. Therefore to quantify variability, the number, duration and height of the humps are to be quantified.

The band array signals can be considered as alternating signals superimposed on a DC signal, the DC component being the averaged band activity over the entire duration. By subtracting the DC component, the quantification of the number, height and duration of the humps can be accomplished by computing the zero-crossing rate and the energy of the alternating signal.

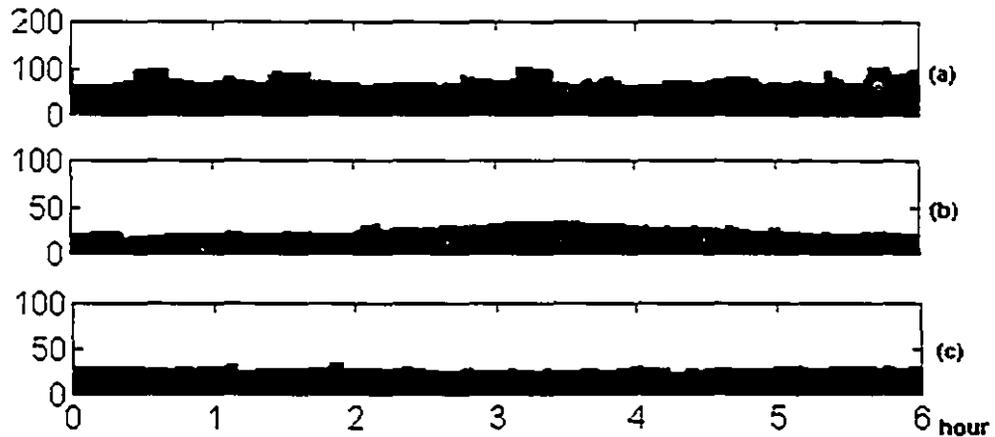


Figure-2.11 Delta band activity of the left posterior channel of three EEGs. (a) Showing a good number of broad humps which implies a normal variability; (b) Having only one big hump which represents an abnormal variability; (c) Showing an absolutely flat activity which indicates an abnormal variability;

Figure-2.12 (a) shows a delta band array of an EEG record, having continually increased amplitude along the entire duration. Such a continual increase is often encountered in patients after cardiac surgery. The line illustrates the position of the DC component. It is apparent that, by simply subtracting the DC component, the information about the spontaneous humps can't be acquired by calculating the zero-crossing rate and energy of change, because of the linear amplitude trend. In this case, the linear trend can be fitted and subtracted from the band array instead of the DC component to remove its influence on both zero-crossing rate and energy of change, as depicted in Figure-2.12 (b). It needs to be mentioned here that for the recordings which have no amplitude trend (increasing or decreasing) the fitted linear trend is equivalent to the DC component.

As the name implies, the zero-crossing rate is a count of the number of times the signal crosses the time axis after subtraction of the trend, which reflects the number of humps. Since we are only interested in the humps that extend over a substantial period, a

5-point linear averaging filter is used to smooth out the humps in the band array that lasts for less than a couple of minutes.



Figure-2.12 The delta band activity of channel p3-o1 of an EEG having continually increasing amplitude along the six-hour duration. (a) The mean of the six-hour activity doesn't represent the base around which the EEG activity alternates; (b) The linear trend of the six-hour activity represents the base around which the EEG activity alternates.

As mentioned before, not only the number but also the height and duration of the humps need to be quantified. The extent of the humps is ascertained by computing the energy of change of the alternating signal using the formula below:

$$E = \log\left(\frac{\sum_{i=1}^{720} x_i^2}{[\text{Average band activity}]^2}\right)$$

Where "x_i" denotes the amplitude of each point of the band array after the fitted linear trend has been removed from it. There are 720, 30-second epochs in 6 hours.

From the discussion above, it is evident that in order to quantify variability of an EEG, both zero-crossing rate and energy of change are essential. A zero-crossing rate and an energy of change measure are extracted from the posterior channel of each of the two

hemispheres for delta activity of the 6-hour EEG record and used for further interpretation.

2.3 Statistical Analysis

Following feature extraction, the 500KB activity file is reduced to a set of 1012 data values, 576 of which describe its amplitude, 288 describe the Left/Right symmetry, 144 describe the Front/Back differentiation and 4 describe the variability of the EEG. For the quantitative interpretation of different features of the EEG, these values need to be reduced further. Unfortunately, the EEG expert is unable to help with the definition of the quantitative boundaries of the normal and abnormal categories for all of the measures, since all these values are extracted from the frequency domain. Therefore it would be preferable to provide the various measures of the EEG as quantitative assessments indicating normality of the corresponding features, rather than as a series of numerical estimates that would be difficult for a clinician to interpret.

Interpretation of quantitative EEG measures, as described above, is usually performed by statistical analysis comparing results to values obtained from a normal control population. The selection of a representative normal control population is crucial for good system performance. For this project, since it is impractical to get real "normal" subjects to undergo long-term EEG monitoring in the ICU, a group of ten post-cardiac surgery patients varying from six months to 12 years old with normal post-operative long-term EEG recordings and normal short-term neurological outcome were carefully chosen to form the control population.

2.3.1 Population Distribution

As seen in previous sections, all the features extracted from the frequency band arrays of the EEG are defined as the logarithm of a certain estimate rather than the estimate itself. The distribution of the estimates such as average amplitude do not follow the characteristics of a normal distribution. This may either be due to the biological mechanism generating the EEG or due to rigid boundaries associated with the estimates themselves. To facilitate transformation of these distributions toward the normal distribution, the logarithm of the estimate is used as the parameter definition (Gasser et al., 1982).

Amplitude Data values from the ten patients of the control population amounting to 2016 points are used to construct frequency distributions of amplitude for each of the eight channels, as illustrated in Figure-2.13. Skewness and kurtosis of these distributions are similar to that of the normal distribution. Averages and standard deviations (STD) of the various distributions appear in Table-2.2.

The last two amplitude measures represent the central temporal head regions. The inter-electrode distance for these two channels is twice that of the other channels and hence the amplitude averages are higher than those of the other channels.

Head Region (Channel)	Average	STD.
Left Anterior	4.43	0.37
Left Central Parietal	4.41	0.31
Left Posterior	4.62	0.41
Right Anterior	4.48	0.33
Right Central Parietal	4.39	0.32
Right Posterior	4.63	0.36
Left Central Temporal	5.10	0.33
Right Central Temporal	5.12	0.32

Table-2.2 Average and standard deviation values of the eight amplitude distributions of the control population.

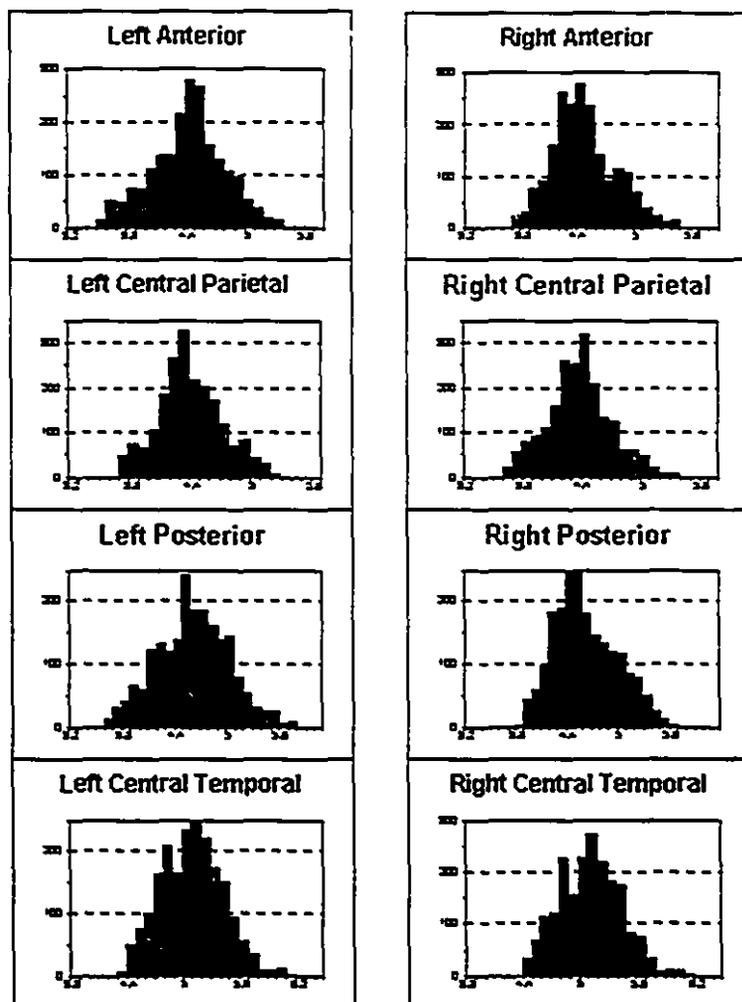


Figure-2.13 Distributions of the eight amplitude measures of the control population.

Symmetry Data values from the same ten patients are used to construct the four frequency distributions for the Left/Right symmetry measures. These distributions also are close to normal distributions, as shown in Figure-2.14. The averages and standard deviations of the four distributions are presented in Table-2.3.

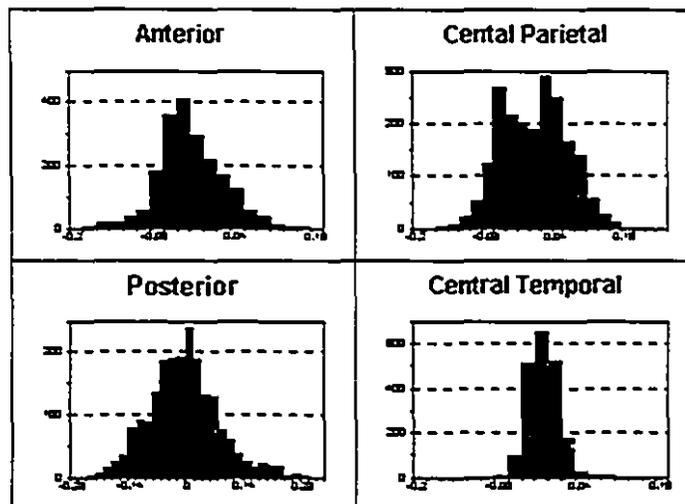


Figure-2.14 Distributions of the four symmetry measures of the control population.

Head Region	Average	STD.
Anterior	-0.018	0.049
Central Parietal	0.009	0.056
Posterior	-0.002	0.083
Central Temporal	-0.008	0.022

Table-2.3 Average and standard deviation values of the four symmetry distributions of the control population.

The distribution from the channel pair that monitors the central-temporal head region has a much smaller range as compared to the other distributions. The reason is that the two channels of this pair have a common electrode, C_z , and the difference in activity between hemispheres is therefore much less, when compared to the other channel pairs.

Front/Back Differentiation As discussed in the section on abnormal EEG, the normal pediatric EEG is reported to present significant Front/Back differentiation when the patient's age falls in the range of 4-month to 10-year. Beyond that range, this Front/Back differentiation is unimportant. Among the ten control subjects, the EEG recordings of the patients whose age is in the above range are selected to form the control distributions for

the Front/Back differentiation. Skewness and kurtosis of these distributions consisting of 1800 data values are also similar to that of the normal distribution. The distributions and the statistical values are presented in Figure-2.15 and Table-2.4 respectively.

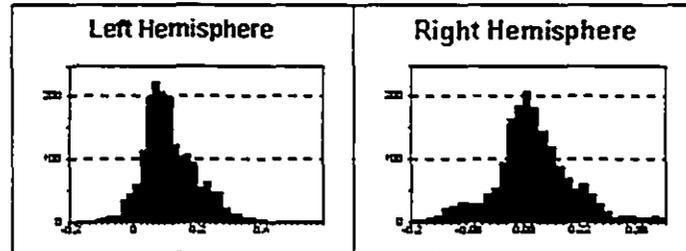


Figure-2.15 Distributions of the two Front/Back differentiation measures of the control population.

Hemisphere	Average	STD.
Left	0.13	0.11
Right	0.11	0.12

Table-2.4 Average and standard deviation values of the two Front/Back differentiation distributions of the control population.

Variability Unlike the other three measures that are extracted once every five minutes, the variability parameters are extracted only once in six hours. Hence a very small sample set is available to form the control distributions. Figure-2.16 illustrates the distributions constructed by 28 values from the delta band activity. Table-2.5 lists the average and standard deviation values of these distributions.

Parameters	Mean	STD.
Zero-Crossing Rate: Left	17.72	3.17
Zero-Crossing Rate: Right	17.96	3.19
Energy Exchange: Left	-3.64	1.17
Energy Exchange: Right	-3.79	1.25

Table-2.5 Average and standard deviation values of the four variability distributions of the control population.

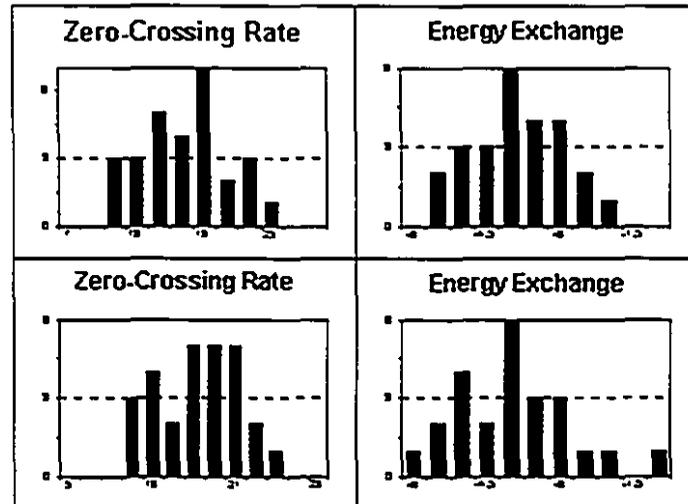


Figure-2.15 Distributions of the four variability measures of the control population

2.3.2 T-Statistic

Eighteen variables are measured from a six-hour EEG recording, eight for amplitude, four for symmetry, two for Front/Back differentiation and four for variability. Each of the eight amplitude measures and six ratio measures is associated with a distribution composed of 72 data values. The amplitude and ratio distributions of a new “normal” six-hour EEG could be expected to be quite similar to the corresponding distributions of the control population. An abnormal EEG, on the other hand, would have distributions quite different from that of the control group.

A measure of the degree of similarity between the distributions of the control population and the EEG being analyzed would give an estimate of the normality of the feature concerned. A t-statistic measures the level of similarity between distributions by comparing their means and may be used to compare the distributions of each new six-hour EEG with the control population. Normally, the t-test statistic can be applied only to distributions that satisfy two conditions: independent samples and normal distributions.

For the problem at hand, the two samples are absolutely independent of each other since they are from different patients and do not influence each other. The distributions of the control population obtained for amplitude, symmetry and Front/Back differentiation are within limits of a normal distribution as discussed earlier. The corresponding distributions of most patients have also been found to conform with the characteristics of a normal distribution. The t-statistic can therefore be applied to obtain the information about similarity of these distributions.

Suppose that independent random samples of sizes of n_1 and n_2 are taken from two normally distributed populations with means μ_1 and μ_2 respectively. Let x_1 and s_1 represent the sample mean and standard deviation of the samples from the population with mean μ_1 and x_2 and s_2 the sample mean and standard deviation of the sample from the population with mean μ_2 . Then the random variable

$$t = \frac{(x_1 - x_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

has approximately the t-distribution with degree of freedom given by

$$df = \frac{(s_1^2 / n_1 + s_2^2 / n_2)}{\frac{(s_1^2 / n_1)^2}{n_1 - 1} + \frac{(s_2^2 / n_2)^2}{n_2 - 1}}$$

rounded to the nearest integer. The t-curve is symmetric about '0' and extends indefinitely in both directions. It approaches a normal curve as the number of degrees of freedom gets larger.

To test the similarity of the means of two populations, we can set a hypothesis H_0 : $\mu_1 = \mu_2$. The random variable

$$t = \frac{x_1 - x_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

is called a t-statistic. A t-statistic value of 0 implies that the hypothesis is absolutely true and the greater its absolute value the smaller is the probability that the two means are equal.

Subsequent to feature extraction from a new EEG to be interpreted, the means and standard deviations of the 14 parameters (8 for amplitude, 4 for symmetry and 2 for Front/Back differentiation) are calculated and the corresponding t-statistic values are computed. If the t-statistic is greater than zero then the mean of the control population is less than that of the EEG being analyzed. Similarly, if the t-statistic is less than zero the mean of the control population is greater than that of the EEG being evaluated.

The range of t-statistic values extends from $-\infty$ to $+\infty$. Its value represents the level of normality of the measure associated with it. For amplitude, depression in EEG activity is an abnormality and amplitudes greater than or equal to that of the control population represent normal EEG. Therefore a t-statistic value greater than or equal to zero implies amplitude normality for the corresponding amplitude parameter and the probability of normality is assigned a value of 1.0. It has been observed that a t-statistic value less than -100 indicates a very severe depression and is therefore assigned a 0.0 probability of amplitude normality. Then the t-statistic value from -100 to 0 is linearly mapped onto a probability range from 0.0 to 1.0. The probability of normality of an amplitude parameter is thus defined as:

$$P_{Nor} = \begin{cases} 1.0, & t \geq 0 \\ (100+t)/100, & -100 < t < 0 \\ 0.0, & t \leq -100 \end{cases}$$

The symmetry parameters compare activity of the left hemisphere to that of the right hemisphere. A t-statistic value of 0 for a symmetry parameter indicates that the EEG being analyzed is symmetrical and is hence associated with a probability of normality of 1.0. On the other hand, a t-statistic value greater or less than zero implies that either left side is more active than the right side or the right side is more active than the left side and the EEG is asymmetrical. It has been observed that an absolute t-statistic value greater than 80 indicates a very severe asymmetry and is therefore assigned a probability of normality of 0.0. Thus the absolute t-value range from 0 to 80 is linearly mapped onto a probability range of 1.0 to 0.0, as described below:

$$P_{Nor} = \begin{cases} 0.0, & t \leq -80 \\ (80-|t|)/80, & -80 < t < 80 \\ 0.0, & t \geq 80 \end{cases}$$

In the case of Front/Back differentiation measures, a t-statistic value greater than or equal to 0 indicates the front back ratio parameter of the EEG analyzed has a greater mean than the control population and implies an equal or greater differentiation. This is considered normal and associated with a probability value of 1.0. On the contrary, a t-value less than 0 indicates an insufficiency in differentiation in activity between posterior and anterior head regions. Once again it was observed that a t-statistic less than -50 denotes a very severe lack of differentiation and the probability of normality is set to 0.0. The probability of normality of Front/Back differentiation is described as:

$$P_{Nor} = \begin{cases} 1.0, & t \geq 0.0 \\ (50+t)/50, & -50 < t < 0 \\ 0.0, & t \leq -50 \end{cases}$$

2.3.3 Assessment of Variability

Unlike the amplitude and ratio parameters which are summarized by data distributions for each new EEG record, variability is characterized by four values each one summarizing a different aspect of variability, the zero-crossing rate and energy of change of the delta activity of the two hemispheres. As it is not a distribution, the t-test cannot be used to assess it. The data values themselves can be assessed by comparing them with the corresponding distributions of the control population.

A measure of the relative standing of the data values within the distributions of the control population may be treated as an estimate of the level of normality. For instance, a zero-crossing rate greater than all sample points of the control population implies normality in the zero-crossing rate and a value less than all sample points of the control population implies abnormality. Such a measure could be arrived at by computing the z-score of the parameter value.

Any normal curve is defined by its two parameters: μ , the mean about which it is symmetrical and σ , its standard deviation. The percentage of the area that lies to the left of a given value x compared to the total area under a normal curve gives a measure of the relative standing of the data value x within this normal curve. This percentage value is equal to the area that lies to the left of a value z under the standard normal curve (which, by definition, has a mean of 0 and standard deviation of 1, and the total area under the

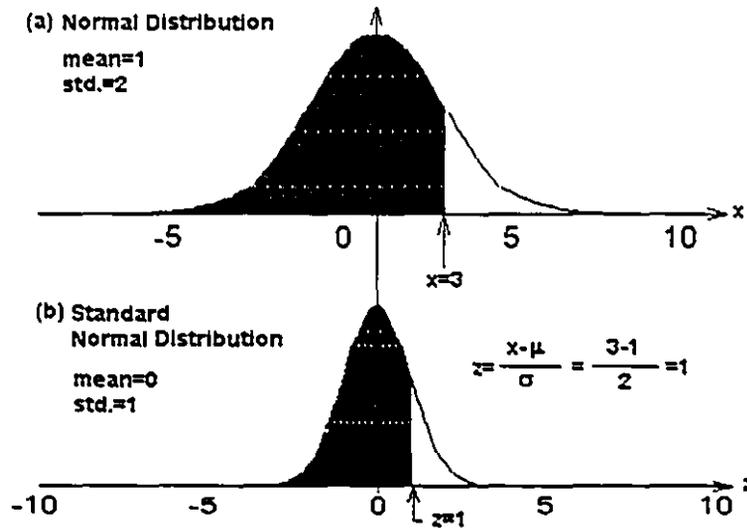


Figure-2.17 Z-score transformation. (a) A normal curve with mean of 1 and std of 2. The shaded area is 84% of the total area under the curve; (b) The standard normal curve. The shaded area is 0.84. Since the total area under the standard normal curve equals to 1, the percentage of the shaded area is also 84%, as in (a).

curve is 1) where $z = \frac{x - \mu}{\sigma}$, as illustrated in Figure-2.17, and can be read from a standard table.

If we assume the four measures of the control population, two zero-crossing rates and two energy exchanges, follow normal distributions, the z-score method described above may be used to evaluate the normality of these parameters. Subsequent to computation of the z-score, the two zero-crossing rate values and the two energy exchange values are converted to four probability values of normality.

2.3.4 Data Reduction

Following the statistical analysis described above, an EEG is characterized by 18 probability values of normality: 8 for amplitude, 4 for symmetry, 2 for Front/Back differentiation and 4 for variability. An automated learning machine could be built to learn the EEG classification knowledge based on these measures. However, the efficiency

of such a system is dependent on the number of variables provided to it and is better if fewer variables are input to the system. Therefore it would be useful to minimize the number of variables provided to such a learning device.

For amplitude, the eight measures are probability values of normality based on the measurement of the average of the total band activity of a six hour EEG on different head regions. This average can also be performed spatially since depression in several channels is worse than depression in only one channel. Therefore the eight probability values are averaged to form one value which indicates the overall normality of the amplitude. Similarly, the Front/Back differentiation measures corresponding to the two hemispheres are averaged to provide evaluation of the normality of the Front/Back differentiation. In symmetry, however, the situation is a little more complicated. Ordinarily, the neurologist evaluates an EEG with one severely asymmetrical pair of channels as more critical than one with several mildly asymmetrical pairs of channels. The simple average may therefore not be enough to represent the overall evaluation of the symmetry. Here, the overall estimate of normality of the symmetry is achieved by

$$P_{Overall} = \frac{1}{2}(P_{Sev} + \frac{1}{3} \sum P_{Other})$$

Where P_{Sev} is the probability value of the most asymmetrical channel pair and P_{Other} is the value of the other three pairs of channels.

An EEG is said to be variable if it is characterized by a good zero-crossing rate AND a good energy of change. The normality of variability is therefore indicated by the magnitude of the lower of the two parameter values as

$$P_{nor} = \min\{\text{probability of normality of zero-crossing rate}, \\ \text{probability of normality of energy exchange}\}.$$

The two variability parameters corresponding to the two hemispheres are then averaged to provide an overall measure of the variability of the EEG.

Front/Back differentiation is a feature largely influenced by the patient's age. A probability value of normality of 0.0 for the Front/Back differentiation of an EEG record would be associated with normal differentiation if the patient is younger than 4 months or older than 10 years. Therefore, it would be preferable to transform these probability values based on the patient's age before entering the next step. Figure-2.18 illustrates a significance function of Front/Back differentiation, $S(\text{age})$, related to the patient's age, which is established upon the project expert's interpretation of published data (Slater and Torres, 1979). The transformation is described by:

$$P_{New} = \begin{cases} 1.0, & \text{age} < 4_month \\ \lceil P_{Old} / S(\text{age}) \rceil_{1.0}, & 4_month \leq \text{age} \leq 10_year \\ 1.0, & \text{age} > 10_year \end{cases}$$

Where $\lceil \rceil_{1.0}$ means that the value inside $\lceil \rceil$ would be set to 1.0 if greater than 1.0.

As mentioned earlier, the work described in this chapter is based on a former study of Pasupathy (1994). Several modifications were introduced to improve system performance: (1) The trend instead of the DC component was subtracted from the frequency band array in order to calculate the zero-crossing rate and the energy exchange for time variability of the EEG; (2) When reducing the four probability values of symmetry to one, more weight is given to the most severely asymmetrical channel pair

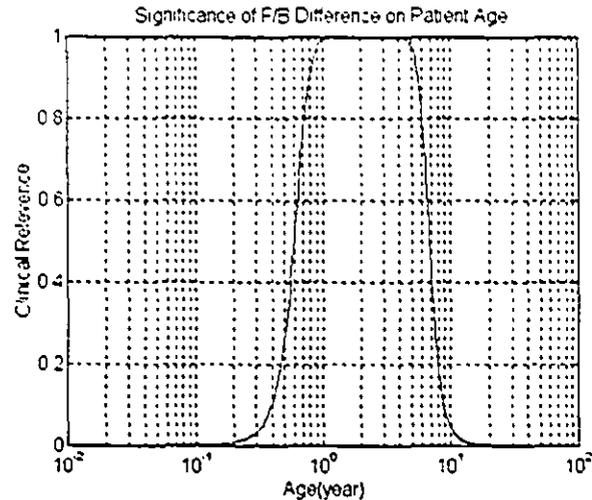


Figure-2.13 Significance of the Front/Back differentiation based on the patient's age

instead of calculating the average; (3) The age factor was introduced to transform the probability value calculated for Front/Back differentiation based on the patient's age.

The data processing and feature extraction methods described in this chapter are summarized as follow. Firstly, the raw EEG is split into 30-second epochs, each of which is called a 'page'. The frequency band arrays are then calculated for each six-hour EEG. The frequency distribution for each page of a channel is computed by Fast Fourier Transform. For each frequency band, an activity value is calculated for each page of a channel from the frequency distributions. Therefore, a particular band array of each channel of a six-hour EEG contains 720 values corresponding to 720 pages. Then, these frequency band arrays are filtered by median filter with a hard-limiter to reject or suppress artifacts. Thirdly, eighteen measures are calculated for the four features from the frequency band arrays, 8 for amplitude, 4 for symmetry, 2 for Front/Back differentiation and 4 for variability. For a six-hour EEG, each of the 14 measures for amplitude, symmetry and Front/Back differentiation gives a distribution composed of 72 values after

a 5-minute average, whereas only one value is calculated for each of the four variability measures. Next, statistical analysis methods (t-statistic or z-score) are used to compare these distributions and values with the distributions of the control population and produce eighteen probability values of normality for the eighteen measures. Finally, these eighteen probability values are further reduced to 4, and a six hour EEG is characterized by 4 probability values of normality of the amplitude, symmetry, Front/Back differentiation and variability. This information is to be further interpreted to classify the EEG as being normal, mildly abnormal, moderately abnormal or severely abnormal. This is done by building an automated learning machine and will be discussed in the following chapter.

Chapter 3

Knowledge-Based Expert System

This chapter describes the method used to build a knowledge-based expert system which accepts the four indices, amplitude, symmetry, Front/Back differentiation and variability measures of an EEG as inputs, and then classifies the EEG as one of the four categories, normal, mildly abnormal, moderately abnormal or severely abnormal (Figure-3.1).

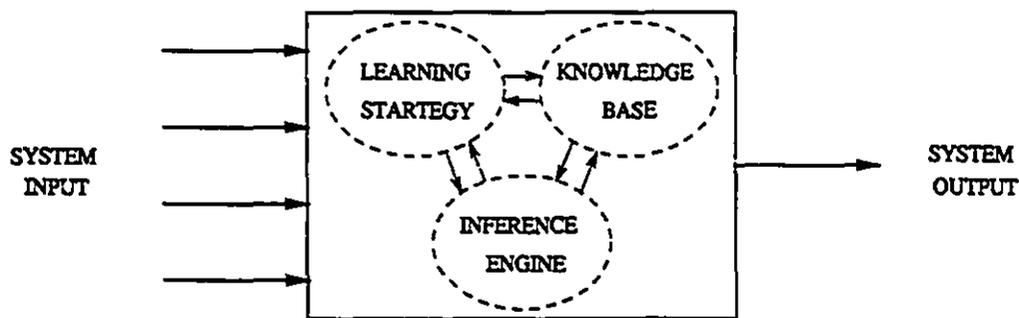


Figure-3.1 A knowledge-based expert system serving as an EEG interpreter.

Such a system should mimic the neurologist's criteria for classification. In other words, it should have the ability to learn the knowledge embedded in the neurologist's decision-making about the level of EEG abnormality based on its various characteristics.

This system should also be able to organize the learned knowledge in a proper way and perform the classification task under a suitable inference engine.

3.1 Machine Learning from Examples

The ability to learn is one of the most fundamental attributes of intelligence, which makes it an important concern for cognitive psychology, artificial intelligence, information science, pattern recognition, education, and related disciplines. The field of machine learning, which crosses these disciplines, studies the computational processes that underlie the learning in both human and machines.

When starting a specific learning problem, an important consideration is the degree of supervision available. In some cases, a tutor or domain expert gives the learner direct instructions about how things should be done or some kind of feedback about the appropriateness of its performance. Such *supervised learning* problems contrast sharply with *unsupervised learning* tasks, in which these instruction or feedback are absent. The less supervision the expert can provide, the more inference work the learner should perform. Below are listed several strategies which exhibit a trade-off in the amount of effort required of the learner and of the teacher:

- * Rote learning and direct implanting of new knowledge.
- * Learning from instruction.
- * Learning by analogy.
- * Learning from examples.
- * Learning from observation and discovery.

Direct implanting of knowledge into a system needs a well-developed and highly-organized knowledge source, such as a spelling checking system, in which the knowledge is implanted directly from a dictionary. Evidently, such a knowledge source would not be available for classifying background EEG signals.

Both learning from instructions and learning by analogy need the tutor's direct instructions, which are precise and quantified knowledge pieces that teach the learner to perform a task. Since neurologists commonly evaluate EEGs through visual interpretation, and classification criteria are neither precise nor quantified, it is not feasible for the neurologists to provide this kind of direct instructions for the learning process.

Learning from observation and discovery is sometimes named unsupervised learning due to the lack of teacher. Normally it achieves the learning purpose by clustering the data into similarity groups. Such unsupervised learning problems are by themselves hard to learn and should not be used if any form of supervision is available.

Another way to teach a system how to perform a task is by presenting it with positive and negative examples of the concept to be learned. The learner induces a general concept description that describes all of the positive examples and none of the counterexamples. This strategy is termed learning from examples. For many learning issues, examples, as highly specific pieces of knowledge, can be designed for the tutors to provide their supervision smoothly and lead the learning strategy to an appropriate point between the two extremes of effort that learner and tutor should give. As we can see, learning from

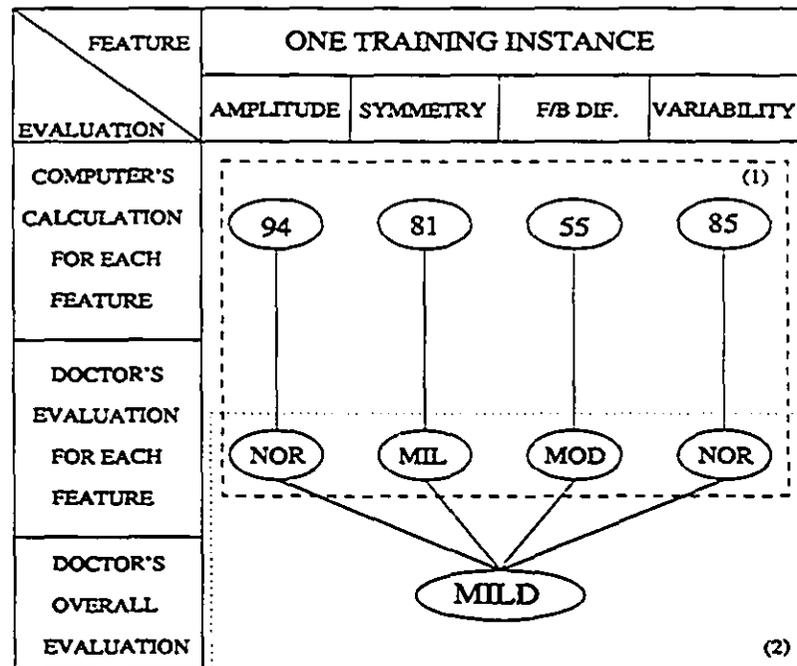


Figure-3.2 The structure of the examples.

examples is a more fitting scheme for our learning issue and the proposed example's structure is depicted in Figure-3.2.

For each of the samples, a six-hour EEG record is quantified by four probability values of normality varying from 0.0 to 1.0 calculated for the four features by the computer. To simplify further interpretation, we transform these probability values to integers from 0 to 100 by multiplying them by 100. The neurologist is asked to classify each of the four features and the overall EEG into one of the four categories: normal (NOR), mildly abnormal (MIL), moderately abnormal (MOD) and severely abnormal (SEV), as illustrated in Figure-3.2. In total, 188 six-hour EEG sections were acquired from the long term EEG recordings obtained in the ICU of the Montreal Children's Hospital and prepared to serve as the examples. Suitable automated techniques can be developed to extract the required knowledge implied in these examples. From Figure-3.2 we can see

that the entire learning system may include two parts. The first part is designed to transform the computer's measures for the four attributes into the doctor's interpretation of these attributes. Since the quantified measures indicate the normality levels of different attributes, with 100 implying complete normality and 0 implying maximal abnormality, it is reasonable to assume that these measures should be highly correlated with the doctor's interpretation of these features if the feature extraction methods are effective. Table-3.1 lists the correlation values calculated by using 4,3,2 and 1 to represent the doctor's classification of normal, mildly abnormal, moderately abnormal and severely abnormal. The correlation levels of amplitude, symmetry and Front/Back differentiation are quite high and therefore the feature extraction methods of these attributes are efficient. On the contrary, the correlation value of variability is seriously low, which indicates that the feature extraction method used for it does not accurately reflect the doctor's assessment and needs to be revised. The transformation approach of variability is therefore distinctive from that of the other three, and the details will be discussed in section 3.2 and 3.3.4 respectively. The second part of the system, which is described in section 3.3, is implemented to evaluate the effects of different interpretations of different attributes on the final classification.

Attributes	Amplitude	Symmetry	Differentiation	Variability
Correlation	0.89	0.82	0.86	0.32

Table-3.1 Correlation values of the extracted probability of normality values of the examples with the doctor's evaluation for them.

When the neurologists evaluate EEG recordings, they often have some difficulty deciding which one of two adjunct classes an attribute belongs to. Therefore they may

give an interpretation like "the overall EEG is mildly to moderately abnormal". After discussing with the neurologists, we decided to accept the three intermediate classes of normal-mild, mild-moderate and moderate-severe. We are therefore dealing with a total of seven classes of abnormality.

3.2 Fuzzy Membership Learning

The aim of this step is to find the relationships which lead the precise probability value of normality calculated for each of the first three features to a classification similar to that of the expert. If a six-hour EEG section has a computer output for amplitude of 75, what would the expert say about its amplitude, normal or mildly abnormal? And what is the probability that it is called normal? How can we model the computer's intrinsic precision together with the human's vague knowledge about EEG?

3.2.1 Fuzzy Sets Theory

Most of our traditional tools for formal modeling, reasoning and computing are crisp, deterministic and precise in character. But in many areas such as expert system, pattern recognition and artificial intelligence, two complications arise (Zimmermann, 1985):

- Real situations are very often not crisp and deterministic and they cannot be described precisely.
- The complete description of a real system often could require far more detailed data than a human being could ever recognize simultaneously, process and understand.

Zadeh, who established the fuzzy set theory, writes (1965):" The notion of a fuzzy set provides a convenient point of departure for the construction of a conceptual framework

which parallels in many respects the framework used in the case of ordinary sets, but is more general than the latter and, potentially, may prove to have a much wider scope of applicability, particularly in the fields of pattern classification and information processing. Essentially, such a framework provides a natural way of dealing with problems in which the source of imprecision is the absence of sharply defined criteria of class membership.”

At a more elementary level, one can consider fuzzy set theory to be a generalization of ordinary set theory: the theory of collections of things. A fuzzy subset of some universe U is a collection of objects from U such that with each object is associated a degree of membership. The degree of membership is always a real number between zero and one, and it measures the extent to which an element is in a fuzzy set, or in ordinary set theoretic terms, it measures the plausibility of an element being in a particular set.

3.2.2 Construction of Membership Function

The theory of fuzzy sets provides a strict mathematical framework in which vague conceptual phenomena can be precisely and thoroughly studied. In this framework the computer's precise output may be interpreted in the terms of fuzzy propositions by means of membership functions.

Interpretation of Membership When choosing the method for constructing a membership function, the interpretation of membership is of utmost importance. The most common and well-studied is a probabilistic interpretation based on the notion of frequency probability (Borisov et al., 1982). A frequency probability $\widetilde{P}(A)$ is a number around which the frequency of the occurrence of the event A $V_N(A) = K_N(A)/N$

oscillates, where $K_N(A)$ is the number of occurrence of A in N trials. By definition,

$P(A) = \lim_{N \rightarrow \infty} V_N(A)$. The membership function of the elements $u \in U$ in the set S is

identified with the probability density $\mu_S(u) = P(\chi_S(u) = 1) = P_S(u)$, where

$$\chi_S(u) = \begin{cases} 1, \dots u \in S \\ 0, \dots u \notin S \end{cases} .$$

Furthermore, subjective probability (Raiffa, 1968), deterministic (Allen, 1974), and possibility-oriented (Borisov et al., 1982) interpretations of membership are also studied in depth and widely implemented.

Basic Methods of Constructing Membership Function Two groups of methods of constructing the membership function of a fuzzy set based on expert estimates can be identified: direct and indirect (Blishun, 1989).

Direct methods are characterized by the fact that the expert directly assigns the rules of determining the values of a membership function which characterizes the notion. Direct methods are used as a rule for describing notions which possess measurable properties such as height, weight and volume. In this case it is convenient to assign values of degrees of membership directly. Methods based on a probabilistic interpretation of membership function $\mu_S(u) = P_S(u)$, i.e. as a probability that the object $u \in U$ will be assigned to a set which characterizes the notion S are also viewed as direct. A direct assignment of membership may be biased. For example, there is a subjective tendency to shift the estimates toward the end-points of a scale. Hence, direct measurements based on direct determination of membership should only be used in the case when the errors are either insignificant or improbable.

In indirect methods, the values of the membership function are selected in a manner that satisfies *a priori* stated conditions. The expert information is only an initial information that needs further processing. Additional conditions could be imposed on the form of the information obtained as well as on the processing procedure. Indirect methods are based on weaker requirements for the expert information. For example, the complex notion "beauty", unlike "length" or "height", possesses no universal measurable properties which would define the concept. In these cases, only ranking measurements based on pairwise comparisons of objects are used. Indirect methods are more time consuming but are more robust to distortions in the initial information.

Moreover, the membership function may reflect the opinion of a group of experts as well as the opinion of a single expert, so that at least four groups of methods are available: direct and indirect for a single expert, and direct and indirect for a group of experts. For a detailed analysis of constructing a membership function, other characteristics such as procedure of collecting the initial data, type of measurements and type of membership scale are also used.

Practical Implementation For our problem, the amplitude, symmetry and Front/Back differentiation of the EEG are measurable unlike the complex notion of "beauty". In fact, the measurement of these three features has already been done by using feature extraction methods which act as the ruler in the case of "height", and the probability values of normality are just the result of the measurement. However, unlike the "height" case, the EEG expert is unfamiliar with these results and therefore cannot assign the value of the membership function directly. A mathematical model is established to automatically

estimate the probability density function $P_S(u)$ from statistical data, which makes our method of constructing the membership function lie somewhere between the direct and the indirect group. Two EEG experts evaluated the EEG sections. But since their work is not independent from each other, it is better to consider it as a single expert case.

Assume that we want to build a membership curve for the fuzzy subset "normal amplitude". In this case, the computer's calculation for the EEG's amplitude is defined as the empirical objects U on an interval of $[0, 100]$, and the fuzzy subset notion "normal amplitude" is represented by S . Using the probabilistic interpretation based on the frequency probability, the membership function is $\mu_S(u) = P(\chi_S(u) = 1) = P_S(u)$.

Consider that there are 150 six-hour EEG sections used as training examples, each of which has the data structure shown in Figure 3.1. In order to build the membership curve for "normal amplitude", the data used here is the 150 computer calculations for EEGs' amplitude and the corresponding expert's classifications. The detailed algorithm for evaluating the frequency probability $P(\chi_S(u)=1)$ is:

- (a) Divide the range of the computer's value, from 0 to 100, into twelve bins: ≤ 0 , 1-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-99, ≥ 100 .
- (b) For every bin B_i ($i \in [0,11]$), calculate the number of training examples (AN_i) having a computer's value of amplitude in bin B_i . For example, if among 150 training examples, there are 17 for which the computer's calculations of amplitude are in the range (80,90], then $AN_9=17$.
- (c) For every bin B_i ($i \in [0,11]$), in the training examples which account for AN_i , calculate the number of events (ANN_i) for which the expert's evaluation of

amplitude is normal. If in the 17 examples above, 16 are classified as having normal amplitude, then $ANN_9=16$.

- (d) For every bin B_i ($i \in [0, 11]$), calculate its frequency probability (FP_i) for "normal amplitude": $FP_i = ANN_i / AN_i$. For the example above, FP_9 of normal amplitude = $16 / 17 \approx 0.94$.

Table 3.2 shows the AN_i , ANN_i and FP_i calculated for "normal amplitude".

B_i	<0	1-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-99	>100
AN_i	4	3	5	14	21	13	20	13	14	17	11	15
ANN_i	0	0	0	0	2	5	10	8	12	16	11	15
FP_i	0.00	0.00	0.00	0.00	0.095	0.385	0.5	0.615	0.857	0.941	1.00	1.00

Table-3.2 The AN_i , ANN_i and FP_i values calculated for the subset "normal amplitude".

By plotting FP_i against the twelve bins, we acquire the primitive membership curve (Figure-3.3) for "normal amplitude".

Since the calculated $FP_i(S)$ is the estimate of $P(\chi_S(i)=1)$ and equals $P(\chi_S(i)=1)$ when the sample number $N \rightarrow \infty$, noise is inevitably introduced into this primitive membership curve. Therefore it is reasonable to expect a better membership function by using a well-defined reference function to fit the primitive curve.

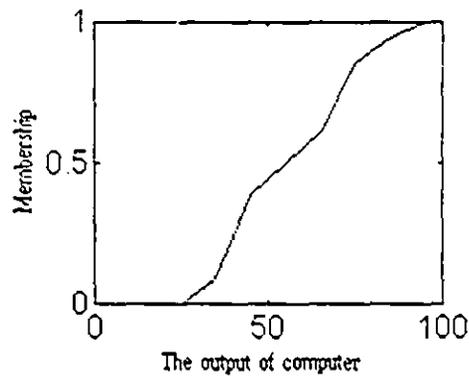


Figure-3.3 Primitive membership curve for "normal-amplitude".

So far, many membership functions have been proposed, however, there is still no systematized conclusion about when and how to use which function, since the choice of the membership function is largely application dependent. We propose to use reference functions of the S-type, which resembles the normal probability distribution (Rommelfanger, 1993). The most often used S-shaped function is the exponential function

$$\mu_S(u) = \frac{c}{1 + \exp^{-a(u-b)}}$$

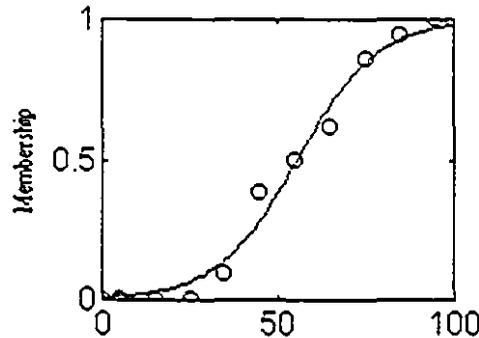
in which b is the inflexion point, a is the slope of the function and c is the scale.

Figure-3.4 shows the original membership curve and its fitted exponential function by using Levenberg-Marquardt nonlinear optimization method (Press et al., 1992).

Unimodal functions, which will be encountered in "mild abnormal" or "medium abnormal" cases, can also be described by representing the increasing and decreasing part separately. Figure-3.5 illustrates this optimization process step by step. First, separate the unimodal curve at the point of maximum value; Second, fit the increasing part and the

decreasing part separately: Third, combine the two fitted exponential curves as:

$$U_c = \text{Min}[U_i, U_d].$$



The output of computer

Figure-3.4 Using exponential function to fit the original membership curve

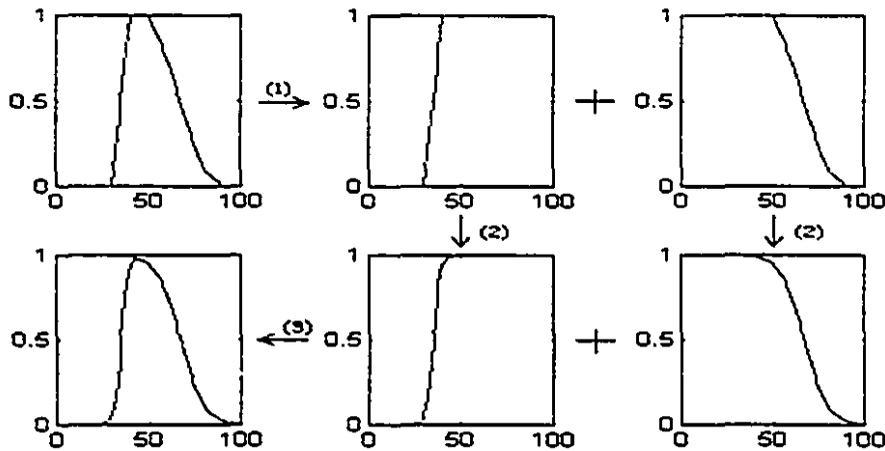


Figure-3.5 Optimization steps for unimodal functions.

For each attribute, we only construct the fuzzy membership functions for the four main classes, not the three intermediate classes. The reason is that 188 training examples are too small to construct membership functions for seven classes, and the embedded noise may make the construction impossible. Therefore, any attribute evaluated as an

intermediate class will be set to the nearest lower class. For instance, if the amplitude of an example is evaluated as mild-normal, it would be transformed to mild to implement this strategy.

3.3 Neural Network Learning

When a neurologist gives an overall evaluation for the EEG, a complex process takes place in his or her mind. A moderate amplitude may be given more weight than a moderate Front/Back differentiation and may carry a larger influence on the final interpretation. Again, he may think the normal symmetry and mildly abnormal symmetry are almost the same, but a severely abnormal symmetry will be totally different. Although all these thoughts and processes obtained from the experience are stored in his mind, it is nearly impossible to explain it verbally in a formal way. We need a system which can automatically learn and encode these thought processes from the training examples and then serve as a classifier.

Before going further, let us discuss a little the issue of noise. Noise, as the probabilistic introduction of errors into data, is inevitable in real-world applications. We call a set of training examples noisy if a probabilistic process is involved in their generation. Noisy training examples are sometimes contradictory. Commonly, a contradiction means that there exist two training examples with the same inputs but different outputs. In our situation, this contradiction has another kind of meaning. Assume there are two training examples $\{E_1, C_1\}$ and $\{E_2, C_2\}$ in which the four feature's classifications of E_1 are better than those of E_2 , so the final evaluation C_1 should be better than or equal to C_2 , otherwise there could be a contradiction. Unfortunately, our training data is very noisy

and these two kinds of contradiction can be found in our training instances. A contradictory training data set is sometimes called nonseparable training examples due to the fact that you cannot separate the class C_1 from C_2 based on these training instances regardless of which method you choose. Therefore when selecting the learning strategy, the ability to handle noise and nonseparable training data becomes an important factor.

3.3.1 Why Use the Neural Network Models

Normally, researchers in learning from examples associate themselves with one of three paradigms, inductive learning (especially rule induction), instance based learning and neural networks (Langley, 1996). In order to evaluate the appropriateness of these three techniques to the task at hand, we have to ponder several detailed issues, such as noise-tolerance, representation, ease of retraining and the amount of training data.

Rule induction employs a statistical evaluation function to incorporate attributes into condition-action rules, decision trees, or similar logical knowledge structure (Langley, 1996). Most early learning algorithms in the rule induction framework assumed that every instance was correctly classified as positive or negative with respect to the desired concept (Koloder et al., 1985). These strategies had almost no noise-tolerance and a single contradictory example would throw the entire learning sequence into confusion. New systems try to fix this drawback by labeling the instances with certainty factors or setting production rules with different reliabilities. The possibility of implementing such techniques in this EEG classifier, however, is very poor because it is extremely difficult for the neurologists to give certainty factors to their evaluations. On the other hand, since the learned knowledge is represented as production rules, retraining procedure will

include reorganization of these condition-action rules, and setting reliability to the rules may make the burden of retraining even more serious.

Instance-based learning simply stores training instances in memory and applies them to new situations through matching techniques. The basic requirement for an instance-based system is a big, pattern-rich training database. For our EEG classification approach, there are four description attributes each of which contains four levels of interpretation, so the total pattern number desired is $4^4=256$. But in practice, we only have 188 training instances. Subtracting the ones that have identical descriptions (their classifications may be different), the remaining patterns reside in 30-40, only one-seventh of the total pattern number. With such a small portion of patterns, it is difficult to implement this paradigm for the task at hand. Furthermore, instance-based learning also has problems in dealing with the contradictory training examples, as discussed above.

Compared with rule induction and instance-based learning, neural network learning has more advantages for our task. Briefly stated, a neural network model (also called a connectionist model) consists of units that are arranged in layers and interconnected by weighted connections, as pictured in Figure-3.6. These connections are used to pass the data through the units. The network learns the input patterns by changing the values of their weights. With suitable weights the network can model any computable function.

Perhaps the strongest appeal of the neural network models is their suitability for machine learning. Machine learning in connectionist models consists of adjusting connection weights to improve performance of a network. This is a very simple and

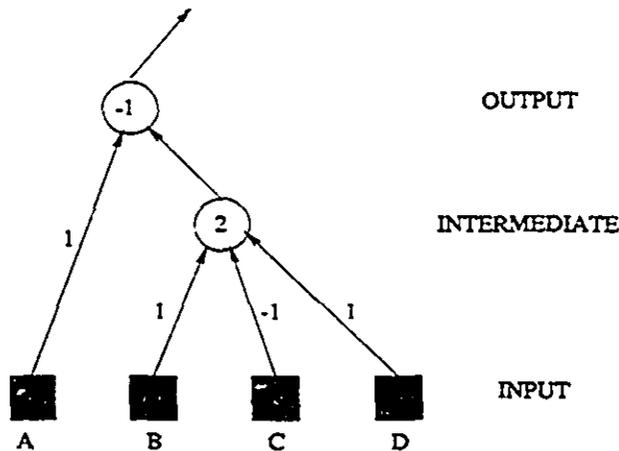


Figure-3.6 A feedforward, two layer network.

pleasing formulation of the learning problem. It abstracts out essentials of learning and invites the construction of general algorithms.

Neural network models can be especially good at handling noise. A lot of learning algorithms for network training are able to learn knowledge from contradictory training data without any extra rules, which makes them handle the noise in the training examples in a very natural way. Moreover, since cells can easily examine large numbers of inputs, they tend to be less sensitive to on-line noise; the greater number of correct input variables can outvote the fewer number of incorrect input values.

Another major attraction of neural network models is that they can serve as knowledge bases for classification expert systems. Most importantly, learning algorithms allow us to generate knowledge bases automatically from training examples. This is particularly convenient for retraining the system because the old knowledge base can be simply replaced by new network weights without any extra effort to reorganize the knowledge base.

The neural network learning scope is usually arranged into two fields, single-layer learning problem and multi-layer learning problem. This classification is based upon a definition about separable function and nonseparable function. Assume there exists a Boolean function which can describe the relationship between the inputs and the outputs of a network, then this Boolean function is called a separable function if it can be computed by a single-cell linear discriminant model; otherwise it is a nonseparable function. Most of the common Boolean functions such as AND, OR, NOT or selector are separable functions, whereas the XOR function is a nonseparable function. Figure-3.7 depicts the case space of these two functions. The left picture shows a separable function space which can be separated by a line, whereas the right illustrates a nonseparable function space.

In real applications, sometimes it is hard to tell whether a system function is separable or nonseparable and whether the single-layer model or the multi-layer model should be chosen. If we are primarily interested in constructing a model from training data that will generalize well, then the only reason to use a multi-layer network is to fit the training data better than would be possible with a single-layer model so that better generalization might be achieved. Therefore if a single-layer model fits the training data as well as a multi-layer network, then the single-layer model should be used. In this case the single-layer model will be faster, simpler and quicker to retrain.

In the next two sections, both a single-layer model and a multi-layer model are created and trained by using corresponding learning algorithms, and the results will be discussed in the next chapter.

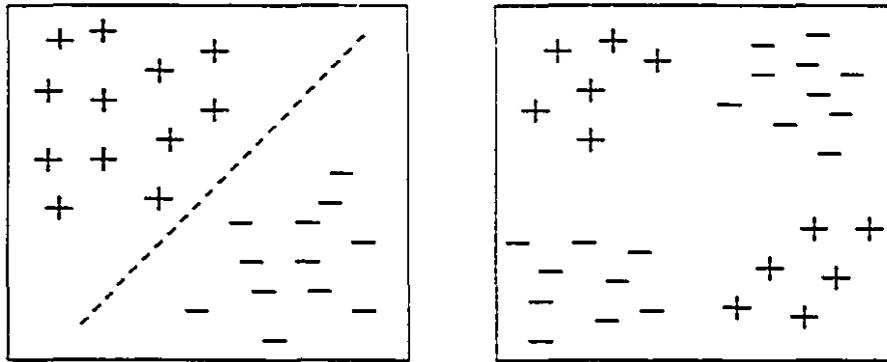


Figure-3.7 Separable and nonseparable functions. The left picture shows a separable function space which can be separated by a line, whereas the right illustrates a nonseparable function space.

3.3.2 Learning in Single-Cell Models

In connectionist models, a k -layer model is defined as a network where cells are grouped into $k+1$ subsets (layers) such that cells are connected only to cells in the next layer. A single-cell, linear discriminant model is a single-layer model (without intermediate cells) in which all the input cells are connected to a single output cell. We can construct our single-cell model as pictured in Figure-3.8.

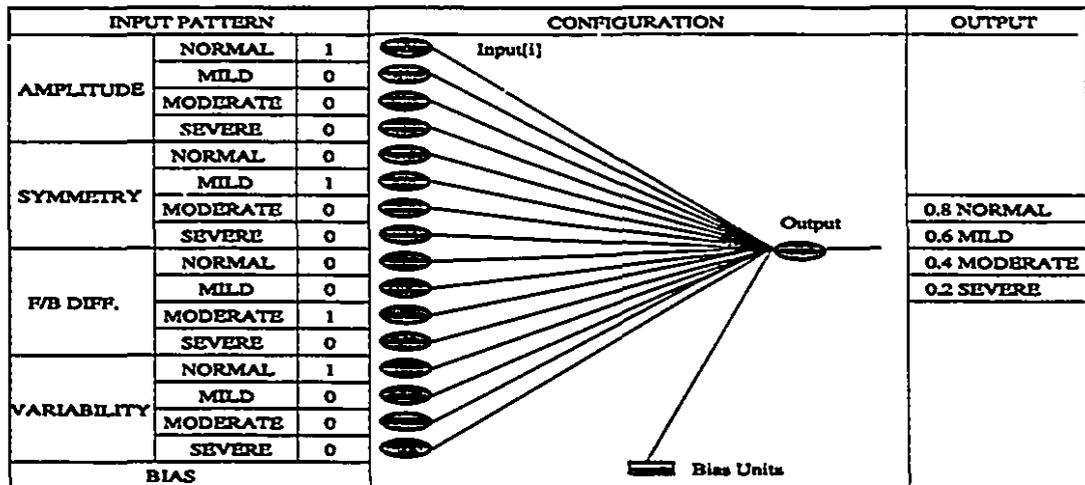


Figure-3.8 The structure of the single cell network.

Perceptron Learning and LMS Perceptron learning is the most important "early" learning algorithm. This algorithm requests that all the inputs, outputs and weights are integral and updates the weights by using $W^* = W \pm E^k$, in which E^k is the input of one training example, W is the old weight vector and W^* is the revised weight vector. The perceptron convergence theorem (Minsky, 1961) proves that for a finite set of separable training examples, E , the perceptron learning algorithm will produce a set of weights, W , that correctly classifies all $E^k \in E$ after a finite number of change steps. The perceptron learning algorithm has limited power since it can only be used for separable training data sets. This shortcoming is fixed by the pocket algorithm with ratchet (Gallant, 1986) which is a recent extension of perceptron learning and shows a better ability to handle nonseparable training data.

The least mean square algorithm (LMS), also called Widrow-Hoff rule, is another important single-cell learning algorithm which tries to find an optimal set of weights, W , that minimizes the mean squared error (MSE) based on the gradient descent algorithm. Compared to the perceptron learning algorithm, the LMS has several advantages:

- It can do a better job with nonseparable data (but not necessarily better than the pocket algorithm).
- It generalizes the training data rather than only finding a solution, as illustrated in Figure-3.9 (a).
- It improves the speed of learning by adjusting the magnitude of the change.
- It is better suited for output values that are continuous.

Of course the LMS algorithm also has its drawbacks. A major disadvantage of LMS is that it overemphasizes the far-out cases, as shown in Figure-3.9 (b).

Since our training examples are nonseparable and the outputs of the examples are continuous, the Least Mean Square algorithm is more suitable than the Perceptron learning algorithm. The gradient descent algorithm, as the basis of the LMS, is discussed in the next section.

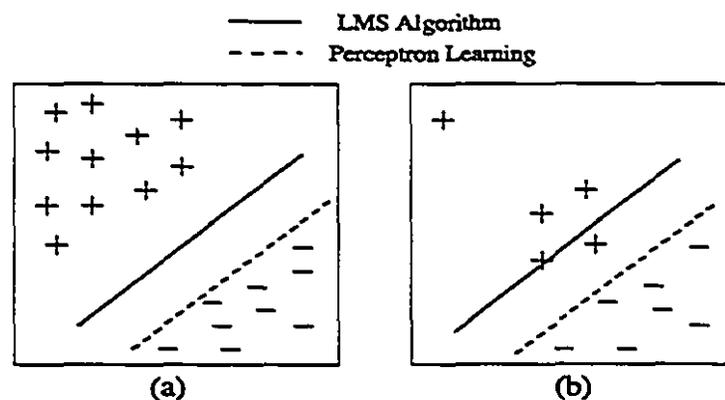


Figure-3.9 Merits and demerits of Perceptron algorithm and LMS algorithm. (a) LMS algorithm generalizes better than perceptron algorithm; (b) LMS algorithm is more sensitive to the far-out samples than perceptron algorithm.

Gradient Descent Gradient descent is a basic technique that plays an important part in most connectionist learning algorithms, especially LMS and backpropagation.

Suppose we have a differentiable function ε that takes a set of network weights $W = \langle w_{p+1,0} \dots w_{p+1,p} \rangle$ and the training examples for a p-input single-cell problem, and produces a measure of the error, $\varepsilon(W)$, for those weights. The error surface gives the error for every set of weight values, W . Because the error function is differentiable, we can compute its multi-dimensional derivative (or gradient) vector

$$\nabla \varepsilon = \left\langle \frac{\partial \varepsilon}{\partial w_{p+1,0}} \dots \frac{\partial \varepsilon}{\partial w_{p+1,p}} \right\rangle$$

at any point in weight space. The gradient gives the direction in weight space that would result in maximum increase of the error when an infinitesimally small weight change is made in that direction. Letting ρ be a small positive number, we can compute the revised weight vector W^* by $W^* = W - \rho \nabla \varepsilon(W)$. This immediately suggests a learning algorithm: start at some arbitrary set of weights, W , and continue to evaluate $-\nabla \varepsilon(W)$ while taking small steps in that direction until some stopping criterion has been reached.

Least Mean Square Assume there are N training examples $\{E^k\}$ with corresponding correct output $\{C^k\}$, then the mean squared error, ε , is given by

$$\varepsilon(W) = \frac{1}{N} \sum_{k=1}^N (W \cdot E^k - C^k)^2 = (W \cdot E - C)^2. \text{ The gradient vector}$$

$$\nabla \varepsilon = \left\langle \frac{\partial \varepsilon}{\partial w_{p+1,0}} \dots \frac{\partial \varepsilon}{\partial w_{p+1,p}} \right\rangle$$

$$\frac{\partial \varepsilon}{\partial w_{p+1,j}} = 2(W \cdot E - C)E_j = 2(S - C)E_j,$$

where S is the weighted sum for the output cell. Because we must specify a step size, ρ , for the gradient descent, we can absorb the factor of 2 into ρ . Then the final form of the Widrow-Hoff algorithm is $W^* = W + \rho(C - S)E$. This updating procedure is also called the delta rule because δ is sometimes used to represent the error, $(C-S)$.

Gradient descent algorithms are not guaranteed to find a global minimum, even if they converge. Sometimes following the derivative will lead to a local minimum, but not the

global minimum. A very nice property of LMS is that there is no local minimum to worry about, since it can be proven that the LMS algorithm must converge to a global minimum (Widrow and Hoff, 1986).

Below is listed the detailed training algorithm for the single-cell model pictured in Figure-3.8.

- 1) Choose a small positive value for ρ , the step size, and assign randomly selected small initial weights $\{W_{i,j}\}$ to the 16 inputs.
- 2) Repeat until changes in the mean squared error, ϵ , become sufficiently small.
 - 2.1) Take the next training examples, E , with correct output C .
 - 2.2) Calculate the weighted sum, $S = W_0 + \sum_{j=1}^p W_{p+1,j} E_j$
 - 2.3) Update weights: $W_{p+1,j}^* = W_{p+1,j} + \rho(C - S)E_j$
 - 2.4) Update step size: $\rho^* = \rho * \zeta$, where $\zeta = 0.9999$.

For the algorithm above, several practical considerations need to be mentioned:

- Input values. If the neurologist's classification for the amplitude of one training example is normal, then the cell that represents normal-amplitude will be fired, having input value of +1, while the other three cells that represent mild, moderate, and severe-amplitude will not be fired, having input values of 0. If this evaluation is mild-normal, then both the cells that represent normal-amplitude and mild-amplitude will be half fired, with values of 0.5.
- Output values. The output value is in the interval [0.2, 0.8], with 0.2 representing severe, 0.4 representing moderate, 0.6 representing mild, and 0.8 representing

normal. If the doctor's classification for one training instance is mild-normal, its output will be represented as 0.7.

- When to update. By definition, $\nabla \mathcal{E}$ is computed for all training examples before updating the weights. But both practice and theory indicate that estimating $\nabla \mathcal{E}$ from a single training example and updating the weights is more efficient (LeCun et al., 1989).
- Iterative learning. Normally there are two kinds of training process: one-shot learning and iterative learning. One-shot learning examines each training example only once in the course of computing a set of weights for the network, while iterative learning examines each training example many times, either cycling through the set of the examples or choosing one at random for each cycle. In our algorithms, we use the iterative learning at random which is more powerful than one-shot learning, but of course slower.
- Initial weights. Initial weight values are set to small, randomly generated positive and negative quantities in the interval $[-2/p, 2/p]$ where p is the number of the input cells.
- Choice of ρ . Appropriate choice for ρ is problem-specific. But it is probably best to keep it less than 0.1. Here we set the initial $\rho = 1/(p+1)$, where p is the number of input cells. We also change ρ for each iteration by a factor $\zeta = 0.9999$. This will make ρ decrease when the error decreases, leading to asymptotic convergence, the characteristic of gradient descent. The decreasing factor ζ is obtained by manual intervention.

- Bias, W_0 , is the constant term when computing $S = W_0 + \sum_{j=1}^p W_{p+1,j} E_j^k$, which can greatly increase the separating power. To calculate the bias, simply add an additional 0th input, then the bias is exactly the same as a weight from an input that is always +1.
- Where to stop. The changes in the mean squared error is monitored during the training process and the algorithm stops when this change is less than 0.005 during 1000 iterations.

3.3.3 Learning in Multi-Layer Models

Backpropagation, rediscovered by Rumelhart, Hinton, and Williams (1986), is currently the most important and most widely used algorithm for multi-layer connectionist learning. Its rapid rise in popularity has been a major factor in the resurgence of neural networks.

Backpropagation is an algorithm for learning in feedforward networks using mean squared error and gradient descent. Compared with the LMS algorithm, backpropagation uses activation functions and works on multiple layers, whereas LMS operates on a single-cell model with no activation functions.

We start with a feedforward network with a set of weights $W_{i,j}$ and a set of training examples $\{E^k, C^k\}$. The activations for all the intermediate cells and the output cells can be computed as:

$$S_i = \sum_j W_{i,j} u_j$$

$$u_i = f(S_i)$$

where the activation function, $f(x)$, is

$$f(x) = \frac{1}{1 + e^{-x}}.$$

The derivative of this sigmoid function evaluated at x is

$$\begin{aligned} f'(x) &= \frac{d}{dx} (1 + e^{-x})^{-1} \\ &= -(1 + e^{-x})^{-2} e^{-x} (-1) \\ &= \frac{1}{1 + e^{-x}} \left(1 - \frac{1}{1 + e^{-x}}\right) \\ \therefore f'(S_i) &= u_i (1 - u_i). \end{aligned}$$

For the gradient vector $\nabla \varepsilon = \langle \partial v_{i,j} \rangle$, we can prove that (Gallant, 1993).

$$\frac{\partial \varepsilon}{\partial v_{i,j}} = -\delta_i u_i$$

where $\delta_i = \begin{cases} (C_i - u_i) f'(S_i) & \text{if } u_i \text{ is an output unit} \\ (\sum_{m>i} W_{m,i} \delta_m) f'(S_i) & \text{for other units.} \end{cases}$

The detailed algorithm is as follows:.

- 1) Choose a small positive value for ρ , the step size, and assign randomly selected small initial weights $\{W_{i,j}\}$ to all cells.
- 2) Repeat until changes in the mean squared error, ε , become sufficiently small.
 - 2.1) Take the next training examples, E , with correct output C .
 - 2.2) *Forward propagation step*: Make an input-output pass through the network to compute weighted sums, S_i , and activations, $u_i = f(S_i)$, for every cell.

2.3) *Backward propagation step:* Starting with the outputs, make an output-input pass through the output and intermediate cells computing:

$$f'(S_i) = u_i(1 - u_i).$$

$$\delta_i = \begin{cases} (C_i - u_i)f'(S_i) & \text{if } u_i \text{ is an output unit} \\ (\sum_{m:m>i} W_{m,i}\delta_m)f'(S_i) & \text{for other units.} \end{cases}$$

2.4) Update weights: $w_{i,j}^* = w_{i,j} + \rho\delta_i u_j$.

2.5) Update step size: $\rho^* = \rho * \zeta$, where $\zeta = 0.99995$.

The practical considerations are similar to those encountered in LMS described in the last section, except that because of the usage of the activation function $f(x)$, the backpropagation error, ε , can have local minima, so we are no longer sure of finding a solution with a global minimum squared error. This is unavoidable, and the only easy countermeasure is to try starting again with different initial random weights.

Most applications of backpropagation to free-network problems make use of a 2-layer network. A major reason for this is that intermediate cells that are not directly connected to output cells will have very small weight changes and will learn very slowly. Another reason for restricting attention to such configurations is that now the only remaining topology decision is how many intermediate cells to use. Here, we also construct a 2-layer network with 16 inputs and one output for our learning problem (Figure-3.10). Different numbers of intermediate cells are tried for this system and their generalization ability is discussed in the next chapter.

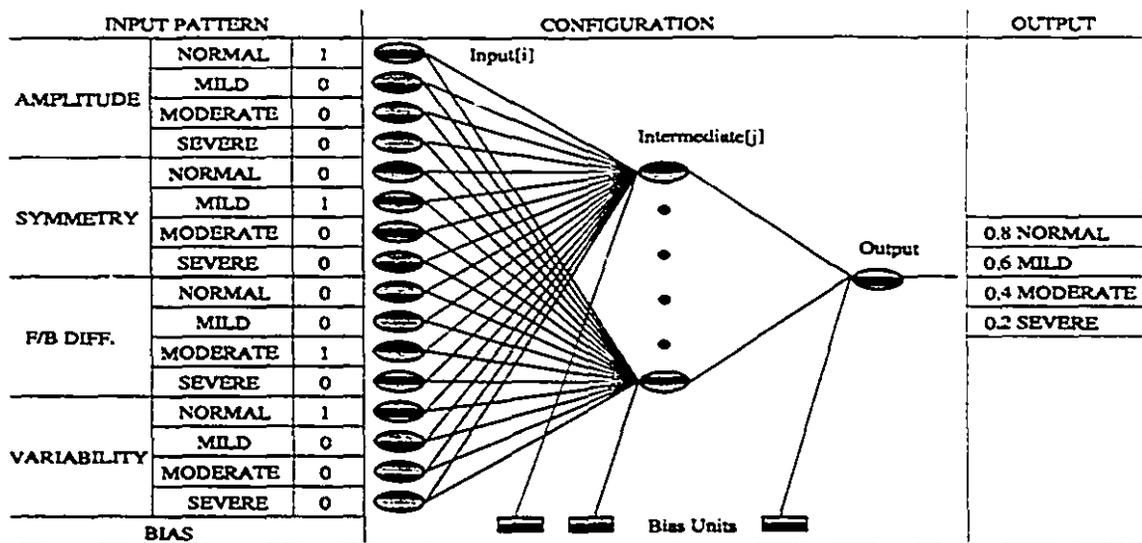


Figure-3.10 The structure of the two-layer network. The number of the intermediate cells is unknown and will be determined upon their generalization abilities.

3.3.4 A Network for Variability

In chapter 2, we discussed the feature extraction method used for variability. Two quantities were calculated for each hemisphere: the zero-crossing rate and the energy of change in the delta band array of the posterior head region. Probability values of normality of different parameters are arrived at by using the z-score method. Then the four probability values are reduced to one by choosing the smaller one of the two quantities in one hemisphere and averaging the values of the two hemispheres. However, at the beginning of this chapter, it was shown that this probability value fails to represent the degree of variability that was assigned by the neurologists. One reason for this failure is that in addition to the delta band activity, the alpha band activity may also reflect the variability. Lack of analysis of this band may therefore partially influence the results. After extensive observation, we realized that neither delta band nor alpha band can reflect

the variability with their own zero-crossing rate and energy exchange parameters. So, there may exist some interaction among the activities of these two bands which reflects the variability of the EEG but is difficult to acquire by observation. When calculating the probability values of normality of different quantities, we assume that the larger the quantity value, the greater the probability of normality of that parameter. For instance, a zero-crossing number greater than all sample points of the control population would imply a 100% normality of zero-crossing rate. But this assumption maybe not correct. The largest probability of normality for the zero-crossing rate of the delta band array over six hours may lie in a particular range, say, 17-18 times, and the further a zero-crossing number is away from this range the smaller the probability of normality may be. In this case, it is very difficult to calculate the probability value of normality for the zero-crossing rate and energy exchange quantities since it is hard to tell where this normal range lies. Our problem is that we have eight quantities: zero-crossing rate and energy exchange of delta and alpha band activity of two hemispheres which are thought to reflect the variability of a six-hour EEG in some way, either by themselves or by their interactions. We try to determine what the relationships between the eight parameters and the variability of the EEG are. To solve this problem, we can construct another 2-layer network which uses the eight quantities as the inputs and the expert's classification of variability as the output and train the network by using the examples. Through the training process, the relationships between the eight parameters and the variability of the EEG will be encoded in the weights of the network automatically. The network is trained with the backpropagation algorithm introduced in the last section and the result will be shown in the next chapter.

3.4 Knowledge Base and Performance

Our knowledge base includes four parts: age factor for front/back differentiation, fuzzy membership functions for amplitude, symmetry, and F/B differentiation, network for variability (network B), and network for final evaluation (network A). Except for the age factor, which is obtained from the neurologist's interpretation of published data (Slater and Torres, 1979), the knowledge in the other three parts is learned from the training examples. Fortunately, both our learning strategies for fuzzy membership and neural networks are such that they allow us to generate the knowledge base during the learning process without any form of reorganization.

Figure-3.11 pictures the entire system's structure and shows an example to illustrate its performance. Assume there is a six-hour EEG section from a patient who is eleven months old. After processing by artifact rejection and feature extraction strategies, the values calculated for amplitude, symmetry and F/B differentiation are 77, 82 and 49. Subsequently, the age factor, which is 0.74 corresponding to the patient's age of 11 months (0.9 years), is obtained by checking the age factor database, and then the Front/Back differentiation value 49 is divided by 0.74 to get the transformed value of 66. Then, the fuzzy membership base is checked to obtain the membership values for 77, 82 and 66. These membership values are then applied directly as the corresponding inputs of network A. Next, network B employs the 8 parameters of variability as the inputs and calculates its output as 0.7; since 0.7 is half way between normal and mildly abnormal. The last four cells of network A are set to 0.5, 0.5, 0, 0. Finally, the overall result, 0.7, is calculated from network A, which classifies this six-hour EEG recording as mild-normal.

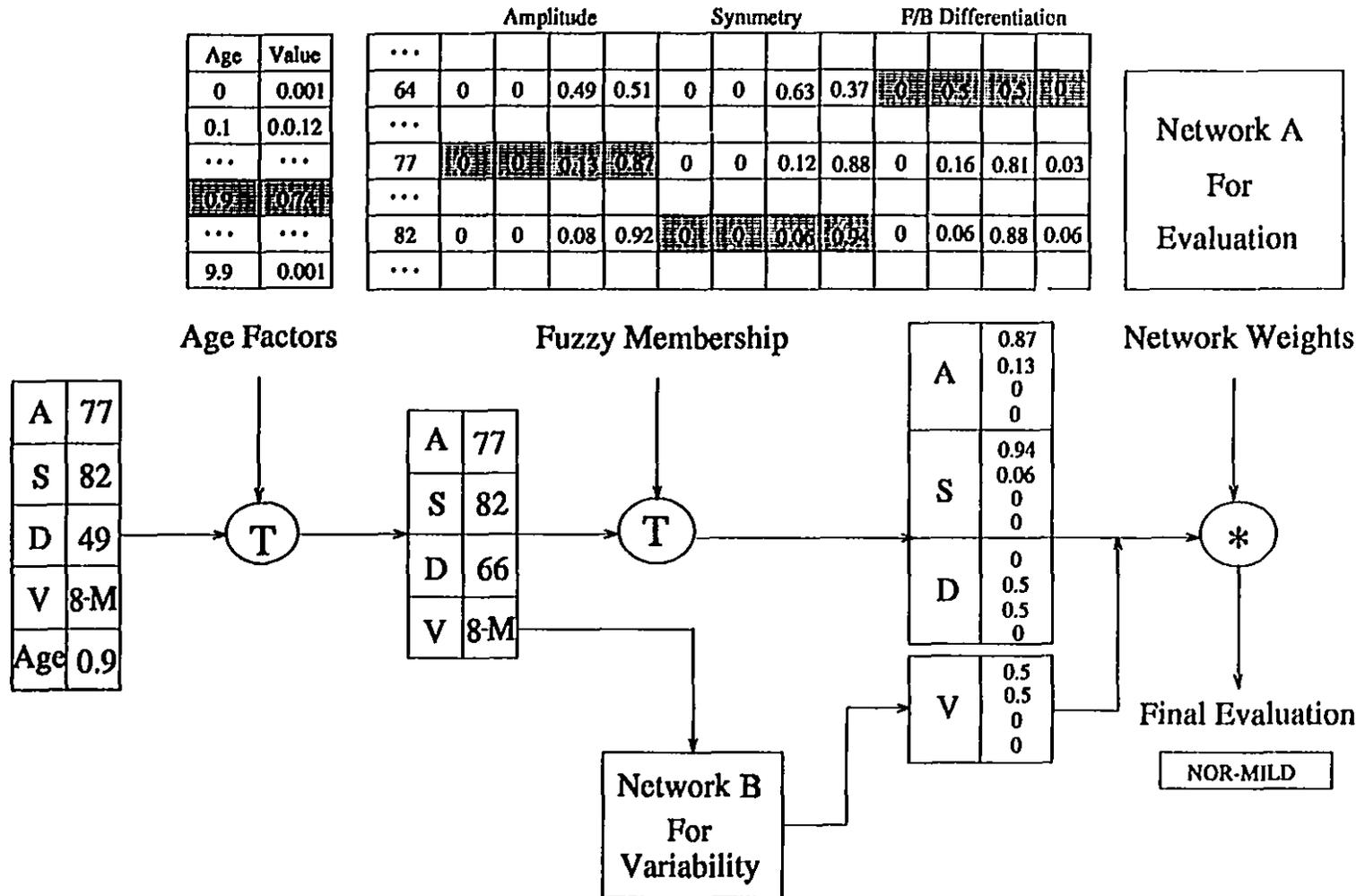


Figure-3.11 Inference Engine

Chapter 4

Results and Discussion

4.1 Training and Testing Data

Seventy four long-term EEGs recorded at the Montreal Children's Hospital were used as the training and testing data for the expert monitoring system. Among them, sixty nine were recorded in the intensive care unit after corrective cardiac surgery, and five were recorded at another monitoring unit for suspected seizures or coma. The recordings were interpreted by two neurologists and every six-hour section was graded for each of its four features and evaluated as one of seven abnormality levels. Every one of the seventy four EEGs yields one to three sections depending on the length of the recording. In all, a set of 188 sections was available from the seventy four EEGs for training and testing the expert system. There are several ways in which a data set can be divided into training and testing subsets, each of which has its own merits and demerits (Devijver, 1982).

One technique, which is called the resubstitution method, uses the entire data set to train the system and the same data set to test it as well. Normally, this technique is used with a perceptron-like learning algorithm since one essential property of the perceptron learning algorithms is that they converge to one hyperplane that correctly classifies all the training data. Obviously, this method is not suitable for our learning problem in which the

training data set is nonseparable and the performance results are not replicated by independent testing data.

The holdout method, the most obvious alternative to the resubstitution method, requires the data to be partitioned into two mutually exclusive subsets and uses one for testing and the other for training the expert system. This scheme suffers from making poor use of the data since a learning system trained on a larger data set will, in general, perform better than one that is trained on a smaller data set, especially when the sample-size is small or incomplete.

Leave-one-out is a method that tries to compensate for this demerit of the holdout method and make efficient use of the available data. By this technique, if the sample size is N , the system is trained with $(N-1)$ samples and tested with the last one. This is carried out N times until all the samples have been used for testing. Here for each run almost the entire sample-set is used for training and ultimately all samples are used in the tests, though each run consists of independent training and testing sets. However, the extensive computation involved in the N training sessions is a big drawback of this technique.

The rotation method is a compromise between the holdout and leave-one-out methods. For this method, the N samples are divided into r sets with N/r samples each. In each run one of the r sets serves as the testing set while all the others are used to train the learning machine. The performance of the system is then arrived at by calculating its added performance for the r runs. The rotation method reduces both the poor data utilization inherent to the holdout method and the computational complexity associated with the leave-one-out method.

In this project, the rotation method of performance estimation was used. The 188 samples were divided into six subsets with 31 or 32 samples each. For each series, five subsets were used to train each part of the system and one was used to test. Six training and testing runs were carried out and the performance was the summation of the six runs.

4.2 Results

4.2.1 Results of Fuzzy Membership

Six training and testing runs were carried out to train the system and test the performance. Figures-4.1, 4.2 and 4.3 demonstrate the training process and the membership functions obtained for amplitude, symmetry and Front/Back differentiation in training series 1. In each figure, the first four graphs delineate the estimated frequency probabilities for twelve bins ('o') and the fitted exponential functions for the four classes: normal, mild, moderate and severe; the fifth graph outlines the combination of the four fitted curves which partitions the total value space into four successive but overlapping subspaces.

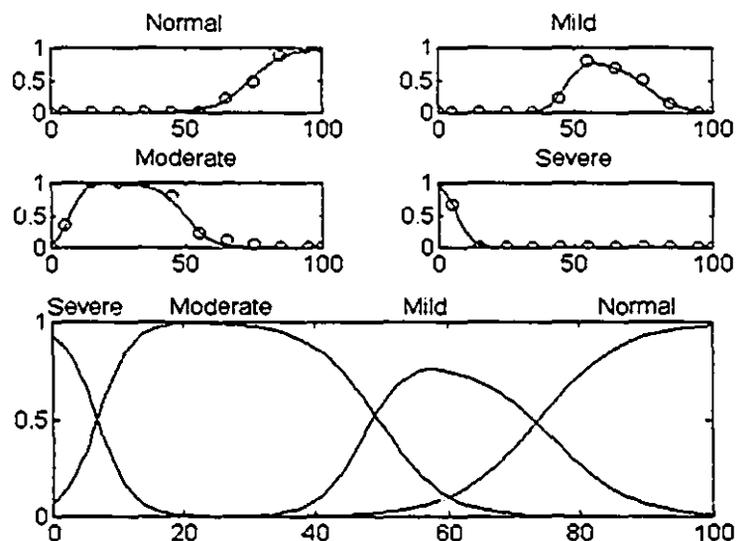


Figure-4.1 Amplitude Membership Functions.

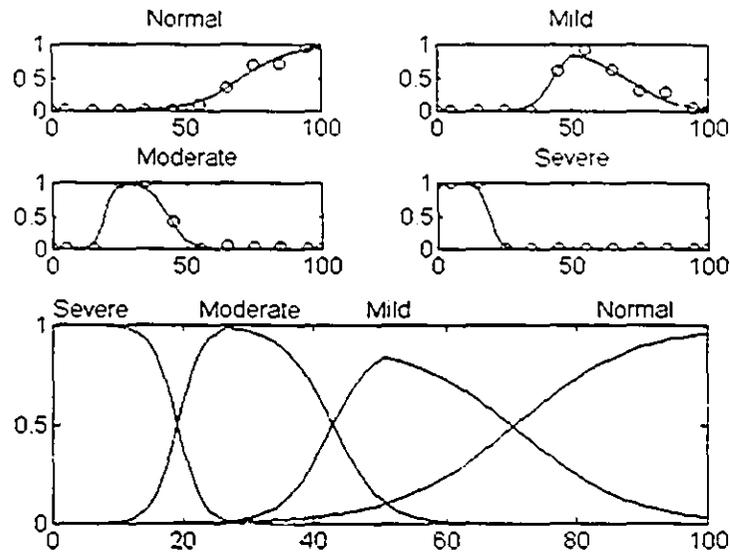


Figure-4.2 Symmetry Membership Functions.

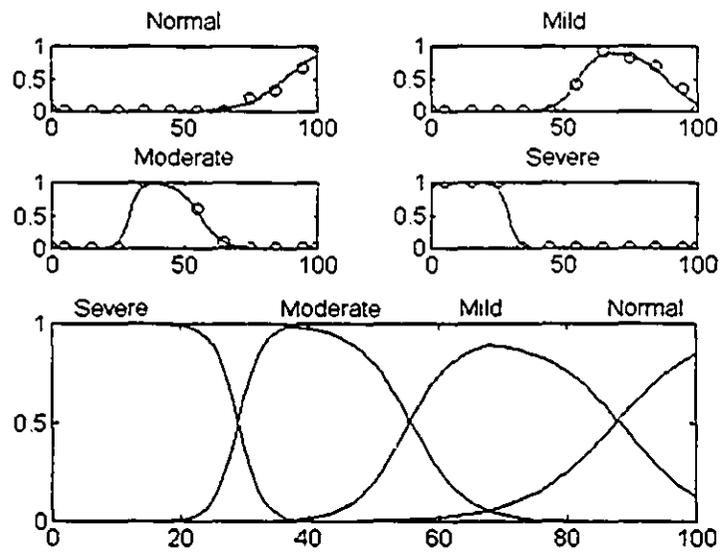


Figure-4.3 F/B Differentiation Membership Functions.

From these figures, it can be seen that the four abnormality levels are not equally distributed in the value space of the computer's output. Some of these subspaces are narrower and some are wider. Moreover, the increase and decrease slopes of most bell-shaped membership functions, as the membership curve obtained for moderately

abnormal Front/Back differentiation, are not symmetrical. All of these may indicate that the neurologist's assessment of abnormality for these features is a nonlinear process. Besides, in all the three features, the normal subspace overlaps the moderate subspace which means that the noise levels between the computer's quantitative measures and the neurologist's evaluation of the three features are high.

Through these fuzzy membership functions, the computer's calculation for each feature has been matched to the doctor's classification. The accuracy of this matching is examined by using corresponding testing examples of different training sessions and the summation of the outcomes is listed in Table-4.1. The "same level" categories present the percentage of the examples for which the fuzzy membership functions' classifications are the same as the doctor's ones. The "less than half level" categories present the percentage of the examples that the fuzzy membership functions' results are only half level away from the doctor's classifications (for example, one is mild and another is mild-normal) plus the examples that have the same classification.

	Amplitude		Symmetry		F/B Dif.	
	Test	Train	Test	Train	Test	Train
Same Level	61%	62%	55%	56%	65%	68%
<Half Level	90%	90%	90%	91%	86%	86%

Table-4.1 Evaluation of the transformation ability of the fuzzy membership functions.

For all of the three features, the results are similar. For testing data, about 55%-65% of the instances are appraised as the same class as the neurologist's evaluation, and only less than 15% are more than half level away from the desired outputs. It needs to be noted that there is almost no difference between the results of the training and the testing data.

4.2.2 Results of Final Evaluation Network

The artificial neural network (ANN) models shown in Figure-3.8 and Figure-3.10 were trained with the Least Mean Square and backpropagation algorithms. The generalization ability of different network models for the training and the testing data is revealed in Table-4.2, in which the neurologist's interpretation for the four features serves as input, and the output of the network is compared with the doctor's assessment of the overall EEG abnormality. The neural network architectures are expressed as strings, showing the number of inputs, number of nodes in hidden layer and number of outputs. The variable Φ indicates the average of mean squares of the differences between the actual and the desired outputs at the end of training over six runs. Most of the models achieve 70% correct classification for both training and testing data, and over 95% are within half level distance from the neurologist's outputs.

Comparing the 2-layer models with the single-layer model (the first row in Table-4.2), even the best of the 2-layer models can only achieve the same generalization ability as the single-layer model does, despite the iteration number of the 2-layer models being ten times that of the single-layer model. The learning performance of these two kinds of models is compared in Figure-4.4. Evidently, the single-layer model converges much

faster than the 8-intermediate-cell 2-layer model. It is concluded that the single-layer models are more suitable than the multi-layer models for this learning problem.

ANN	Weight Number	Iteration	Φ	Train		Test	
				Same Level	<Half Level	Same Level	<Half Level
16-1	16	10000	0.0036	70%	98%	69%	97%
16-2-1	34	100000	0.0043	68%	97%	69%	97%
16-3-1	51	100000	0.0042	67%	98%	67%	96%
16-4-1	68	100000	0.0043	67%	98%	68%	96%
16-5-1	85	100000	0.0042	69%	98%	69%	97%
16-6-1	102	100000	0.0042	67%	98%	66%	97%
16-7-1	119	100000	0.0043	67%	97%	67%	96%
16-8-1	136	100000	0.0039	68%	98%	68%	97%
16-12-1	204	100000	0.0042	68%	98%	67%	97%
16-16-1	272	100000	0.0040	69%	98%	66%	97%
16-20-1	340	100000	0.0041	66%	98%	66%	97%
16-24-1	408	100000	0.0042	68%	98%	65%	96%
16-28-1	476	100000	0.0041	68%	98%	67%	97%
16-32-1	544	100000	0.0039	69%	98%	67%	97%

Table-4.2 The generalization ability of different networks for the final evaluation on the training and testing data.

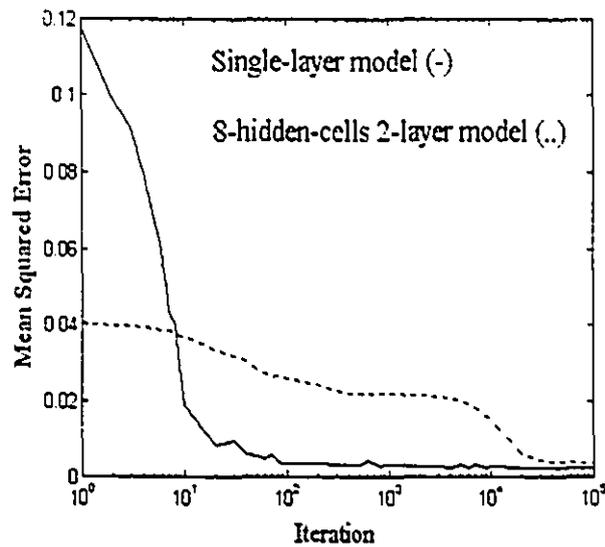


Figure-4.4 Comparison of the convergence velocity of the single-layer and the two-layer networks.

4.2.3 Results of EEG Variability Network

Similar to that pictured in Figure-3.10, the architecture of the network for variability has eight inputs, one-layer hidden cells and one output. The training results (Table-4.3) reveal a reduced accuracy of classification by comparison with the other three features (Table-4.1). The statistical data in Table-4.3 shows that the correct classification percentage is less than 20% in all the models while the parallel values in table-4.1 are in 50%-60%. Also only about 40% of the testing instances are within half level range from their desired output against the 85% in the other three features.

The reasons for the present drop of accuracy in variability classification are complex and the most important one seems to be that the criterion used by the neurologists for classifying variability is different from that extracted by computer. The details of this issue and the possible solution will be discussed in section 4.3.

ANN	Train			Test		
	Same Level	<Half Level	<One Level	Same Level	<Half Level	<One Level
8-2-1	20%	39%	89%	20%	39%	89%
8-3-1	20%	39%	89%	20%	39%	89%
8-4-1	19%	40%	89%	20%	40%	89%
8-5-1	20%	39%	89%	20%	39%	89%
8-6-1	20%	39%	89%	20%	40%	89%
8-7-1	19%	41%	89%	19%	42%	89%
8-8-1	20%	39%	89%	20%	39%	89%
8-10-1	20%	39%	89%	20%	39%	89%
8-12-1	19%	41%	89%	19%	42%	90%
8-14-1	20%	39%	89%	20%	39%	89%
8-16-1	20%	39%	89%	20%	39%	89%

Table-4.3 The generalization ability of different networks for the EEG variability on the training and testing data.

4.2.4 Results of the Entire System

Until now, all parts of the system, their structures, implementation methods and effectiveness have been presented. The capability of the entire expert system built for background EEG classification was evaluated as the cumulative result of the test data of the six runs (Table-4.4), in which the quantitative measures of the test instances were used as the input of the inference engine depicted in Figure-3.11, and the output of the system is compared with the neurologist's interpretation. The single layer model and the

12-intermediate-cell 2-layer model were used as the final evaluation network and the EEG variability network of the system respectively.

Results	Doctor						
Computer	Normal	Nor-Mil	Mild	Mil-Mod	Moderate	Mod-Sev	Severe
Normal	5	2	1	0	0	0	0
Nor-Mil	37	1	20	1	3	0	0
Mild	4	9	2	6	6	1	0
Mil-Mod	0	1	6	3	7	0	1
Moderate	0	0	0	2	4	2	1
Mod-Sev	0	0	0	0	4	2	0
Severe	0	0	0	0	0	0	1

Table-4.4 Comparison of classification results from the automatic method and the human expert.

In the 7 * 7 matrix presented above, rows represents the system's classification for the testing instances and the columns correspond to the neurologists' evaluation. The dark shaded cells in the main diagonal represent the concordance of the two kinds of classification, and the light shaded cells stand for the instances whose system's outputs are half a level away from the expert's classification. As we can see, the results are very encouraging and most of the instances fall in the shaded area. Table-4.5 shows the statistical assessment of the results above. Almost 40% of the total testing instances are classified as the same class as the neurologists'. If we add the cases with a half level deviation, the total percentage of instances which are considered as having acceptable

results reaches 90%. In contrast, the instances whose classification error exceeds one class are less than 3%.

Degree	Percent
Same Level	40%
Half Level	90%
One Level	97%

Table-4.5 Percentage interpretation of the results presented in Table-4.4.

4.3 Discussion

Many systems have been developed to assist EEG monitoring by extracting different kinds of features from the EEG. But in most cases, the users are expected to interpret these features by themselves. This would largely impede the utility of these systems in the ICU during long-term EEG monitoring where expertise may not be present all the time. The system developed in this project tries to remove this drawback by automating the entire process of EEG interpretation.

In the system, artifacts were identified and rejected from the frequency band arrays of the EEG. The main reason for performing artifact rejection in the frequency domain rather than in the time domain is that the features used for EEG interpretation in this system were extracted from the frequency domain to represent gross state of the EEG background over several hours. Therefore, it is not necessary to identify and reject specific artifactual EEG waves. Furthermore, it is quite difficult to anticipate the

waveforms of the various artifacts and it is computationally economical to reject artifacts in the frequency domain (Barlow, 1986).

Since the qualitative EEG features used by the neurologist include not only amplitude but also frequency information about the EEG, the quantitative measures of these features were extracted from the frequency band arrays of the EEG. The measures extracted for amplitude, symmetry and Front/Back differentiation by using t-statistic analysis based on a control population quantify the EEG quite effectively, and the probability values of normality obtained were highly correlated with the expert's visual interpretation. By using fuzzy membership techniques, these probability values were then mapped into the neurologist's assessment for each of the three features. S-shaped membership functions were chosen to represent the distributions of the neurologist's levels of normal, mildly abnormal, moderately abnormal or severely abnormal, since they resemble the normal probability distribution and may reflect the pattern of human language and thought. For further development, other membership functions could also be tested and compared in order to find a best fitted function type. For every normality level of each feature, a histogram was computed and fitted to S-shaped functions through error optimization methods. The unimodal functions were described by fitting both the increasing and decreasing parts to S-shaped functions rather than fitting it to a bell-shaped function. This is based on the consideration that these unimodal functions may not be symmetrical.

As mentioned before, the probability values of normality obtained for EEG variability over time by using the z-score method could not demonstrate the abnormalities observed by the neurologist on visual analysis. In order to solve this problem, a two-layer neural

network was established to classify EEG variability based on the measures of zero-crossing rates and energy exchanges of the delta band array. The classification ability of this subsystem was presented in Table-4.3. Its classification error is quite high when compared with that of the other three features. The reasons of this low accuracy are complex. EEG variability over time is defined as the extent of spontaneous cycling in an EEG record. Many studies (Bergamasco et al., 1968; Bricolo et al., 1973; Garrel et al., 1979; Rimpl et al., 1980; Karnaze et al., 1982) indicated a clear correlation between the presence of this spontaneous fluctuations and a favorable clinical outcome. The subjects of most of these studies were post-traumatic comatose patients aged at least several years. Most of our subjects, however, are post cardiac surgery infants who suffered pain and received sedatives during the recording. The influences of the patient's age, sedatives and the unstable sleep induced by the pain on this spontaneous alteration pattern still remain unclear and may affect the doctor's judgment. Furthermore, most of the studies mentioned above used this spontaneous cycling together with typical polygraphic sleep patterns although some of them considered it as an independent prognostic value. Due to these factors, when evaluating the variability of the EEG samples, our project experts prefer to combine the EEG variability with the typical polygraphic sleep patterns such as spindles, k-complex and vertex sharp waves. Automatically identifying these waves from long-term EEG is, however, very difficult, either in the time domain or in the frequency domain. After further investigation, we found a close correlation between level of the EEG depression and presence or absence of sleep cycle features based on the neurologist's interpretation. In fact, for our training examples, the cross-correlation of the neurologists' classification for amplitude and the combined feature of EEG variability

and sleep cycle is as high as 0.8. A similar finding was reported by Watanabe and coworkers (1980). In their study, they indicated that the classification of the background EEG depression was well correlated with the disturbance of sleep cycle in the full-term newborns with perinatal hypoxia. It may be reasonable, therefore, to rebuild the system by using the three features of amplitude, symmetry and Front/Back differentiation, removing the variability. Table-4.6 presents the testing results of the system trained by three features in which a twelve-input single-layer network similar to that shown in Figure-3.8 was used as the final evaluation network. It performs better than the system trained by four features and increases the percentage of correct classification by almost 5% (Table-4.7).

Results	Doctor						
Computer	Normal	Nor-Mil	Mild	Mil-Mod	Moderate	Mod-Sev	Severe
Normal		8	3	0	0	0	0
Nor-Mil	16		23	3	3	0	0
Mild	3	9		3	2	1	0
Mil-Mod	0	1	6		7	0	1
Moderate	0	0	0	8		1	1
Mod-Sev	0	0	0	0	5		0
Severe	0	0	0	0	0	1	

Table-4.6 Results of the system trained only by amplitude, symmetry and Front/Back differentiation.

Degree	Percent
Same Level	45%
Half Level	91%
One Level	97%

Table-4.7 Percentage interpretation of the results presented in Table-4.6.

A single-layer neural network was used for evaluation of the overall normality level of the EEG based on the classifications of different features. By considering both the generalization ability and computational efficiency, this single-layer network was proven to be more suitable than the two-layer networks. This may reflect that a linear function can be used to represent the relationship between the inputs and the output of the network. Twelve inputs that represent every normality class of the three features were used (sixteen for four features) in this network because different classifications for one feature may influence nonlinearly the final evaluation. If using three inputs instead of twelve (one for each feature), the relationship between the inputs and the output of the network may no longer be linear.

The whole system contains no training data specific values in the learning algorithms. Therefore, the exact same procedures could be repeated for another training data set or for a different type of abnormality. This is convenient for retraining the system in further development.

By comparison to Pasupathy's study (1994), the results presented in Table-4.6 are quite encouraging. The correct classification rate increases from 23% to 45%. The instances classified with an error exceeding one class (here the class means the four main

classes) decrease from 35% to 3%. This improved performance may be due to following reasons: (1) Available training data increases from 60 six-hour EEG sections to a total of 188. This not only trains the system more completely but also expands the scope of technology which can be implemented. (2) Fuzzy membership and neural network techniques handle the noises better than the version space technique used by Pasupathy, which is very crucial in this system due to the fact that the training data is extremely noisy. (3) The whole system is divided into several subsystems, such as fuzzy membership functions, a network for variability assessment and a network for final classification. This may simplify the task of each step and facilitate the error detection in an earlier stage. (4) The age factor is introduced in Front/Back differentiation assessment, which was not emphasized by Pasupathy.

Table-4.6 shows that most of the examples are classified as normal or mildly abnormal. In fact, about eighty percent of the total six-hour EEG instances used in this study show normal or mildly abnormal background activity while only twenty percent were assessed as moderately or severely abnormal. This biased training data results in an incomplete learning of the assessment of severely abnormal EEG patterns in both fuzzy membership functions and neural network weights. The constructed system, therefore, exhibits an increased misclassification rate from normal EEG pattern to severely abnormal EEG pattern.

The weights obtained for the 12-input single-layer network are listed in Table-4.8. We can see that the differences of the weights between the successive nodes that represent different amplitude level are greater than that of symmetry and Front/Back

differentiation. It reveals that the amplitude, which indicates the degree of the EEG depression, is a more important feature than the symmetry and the Front/Back differentiation and a variation in amplitude will greatly influence the neurologists' assessment of the EEG abnormality. This may be due to the fact that amplitude directly represents two abnormal EEG background patterns: electro-cerebral inactive pattern and low voltage pattern through all states. As discussed in the section on abnormal EEG of chapter 1, if persistent, both of these two patterns are well correlated with unfavorable clinical outcome. Moreover, since low voltage pattern differs from the electro-cerebral inactive pattern only in the degree of amplitude attenuation, these pattern can easily be discriminated by the system using amplitude classification. Sometimes, burst-suppression, another important background EEG pattern, can also be reported as generalized depression by this amplitude measure. The reported degree of depression is altered depending on the height and duration of the bursts, the duration of the suppression and the persistency of the burst-suppression pattern. This representation of burst-suppression by amplitude measure is therefore very imprecise and far from the clinical requirement. In recent clinical studies, more and more researchers showed that the EEG burst suppression pattern was highly predictive of unfavorable outcome (Lombroso, 1985; Rowe et al., 1985; Holmes et al., 1993), and it is very crucial to add another measure that could precisely identify this pattern.

Table-4.8 also shows that the weight differences between the symmetrical and mildly asymmetrical nodes are much smaller than that between the moderately asymmetrical and severely asymmetrical nodes. This increased influence on the assessment of EEG

abnormality may reflect that transient or mild asymmetries are probably of no pathological significance, as reported by Lombroso (1985). The symmetry measure used in this system is based on the amplitude, but in some cases the symmetry in frequency is also of interest (Shagass et al., 1988). A symmetry measure in both amplitude and frequency may therefore be more representative and useful although the frequency measurement is easily disturbed by high frequency components induced by sedatives, such as the midazolam that is often used in our patients.

Feature	Class	Run1	Run2	Run3	Run4	Run5	Run6
	Bias	0.215	0.222	0.200	0.214	0.206	0.207
Amplitude	Normal	0.297	0.295	0.306	0.308	0.292	0.291
	Mild	0.140	0.145	0.148	0.145	0.138	0.143
	Moderate	-0.045	-0.040	-0.036	-0.038	-0.045	-0.045
	Severe	-0.181	-0.217	-0.231	-0.215	-0.231	-0.224
Symmetry	Normal	0.145	0.139	0.156	0.176	0.140	0.154
	Mild	0.118	0.105	0.115	0.142	0.117	0.119
	Moderate	0.043	0.021	0.036	0.087	0.009	0.036
	Severe	-0.152	-0.094	-0.080	-0.120	-0.029	-0.083
F/B Diff.	Normal	0.108	0.115	0.120	0.083	0.123	0.117
	Mild	0.058	0.068	0.070	0.033	0.066	0.067
	Moderate	0.087	0.023	0.032	0.051	0.024	0.022
	Severe	-0.014	0.005	0.006	0.028	-0.016	0.012

Table-4.8 Weights obtained for the 12-input single-layer networks after training in the six runs.

There is another important EEG pattern called generalized high-amplitude slow-wave activity which needs to be incorporated into this system. Chiappa (1979) reported that generalized slowing is one of the first few signs of ischemic brain damage. In methodology, this excessive high-voltage slow activity can be characterized as increased low frequency components together with high amplitude. But since only three of the EEGs recorded at the Montreal Children's Hospital were considered as showing excessive high-voltage slow activity, the quantification of these attributes was not feasible.

When interpreting pediatric EEG, it is important for the neurologists to take the patients' age into account since the EEG of children varies extensively from neonatal up to about 10 years old. As mentioned in the section on Front-Back differentiation, an EEG pattern considered normal for a 3-month-old may be an abnormal pattern for a 3-year-old child. Ideally, it would be better to define age intervals, such as <4 month, 4month-1year, 1-4 years, 4-10 years and >10 years within which the characteristics of a normal EEG recording are not expected to vary extensively. Independent "normal" control populations for the various age intervals could then be established and the system could be trained and tested for each of these intervals. However, in practice, the realization of this aim will inevitably encounter the difficulty that recordings for every age interval have to be collected.

Besides background abnormalities, seizures also provide important prognostic information. Although some studies reported that epileptiform activity was not as predictive of outcomes as background activity (Holmes et al., 1982).

electroencephalographic seizures, whether associated with clinical manifestations or not, are highly correlated with poor outcomes (Rome et al., 1985). Recent long-term EEG monitoring revealed that seizures among patients with acute brain injury are far more common than previously recognized. Jordan (1992) reported a significant incidence, up to 35%, of seizures in an analysis of 124 monitored patients. Lowenstein and Aminoff (1992) described 47 comatose patients, in 80% of whom EEGs suggested nonconvulsive seizures. Therefore, the combined use of this background EEG monitoring system together with computer-based automatic seizure detection systems (Gotman, 1982, 1990) may provide more useful diagnostic information.

The present monitoring system was built to receive six-hour raw EEGs as input and classify them to different background abnormality levels. The reason for using a six-hour EEG section as the analyzing unit is because it provides information about the EEG trend and this is of clinical interest. The assessment of the EEG trend is sometimes used for high risk patients to determine the adequate depth of anesthesia (Halimi et al., 1990). More often, the EEG trend is monitored for changes in the CNS function and condition during surgery. In a study of 92 patients monitored by an EEG-trend-analyzer during carotid surgery, EEG-trend-analyzer showed a sensitivity of 100% for ischemic events (Loeprecht et al., 1985). The EEG trend monitoring is particular useful in postsurgical intensive care to determine whether the patient is undergoing deterioration or recovery. To be most useful, the final monitoring system should be able to provide the information about background activity, trend and seizure occurrence of the EEG recorded in the last six hours and update these information every hour.

In clinical use, non-pathological factors such as medication or patient's anesthetic level need also be incorporated into the monitoring system. Drug-induced amplitude attenuation, burst suppression, or seizures have to be excluded. High frequency components introduced by sedatives should also be considered carefully.

With the valuable earlier work of Anitha K. Pasupathy, an effective knowledge-based EEG monitor system in pediatric intensive care unit has been constructed. The overall system performance is quite encouraging. For further development, more long-term EEGs should be collected to train the system, especially the severely abnormal ones. Inclusion of few more quantitative measures and small age intervals suggested above would also enhance the system performance.

References

- Allen AD. Measuring the empirical properties of sets. *IEEE Transaction on Systems, Man and Cybernetics* 1974; 4:No.1.
- Astrup J, Simon L, Siesjö BK. Thresholds in cerebral ischemias—the ischemic penumbra. Editorial. *Stroke* 1981; 12:723-725.
- Barlow JS. Artifact processing (rejection and minimization) in EEG data processing. In: Lopes da Silva FH, Storm van Leeuwen W, Rémond A, eds. *Clinical Applications of Computer Analysis of EEG and other Neurophysiological Signals*. Amsterdam: Elsevier Science Publisher B.V. 1986:15-62.
- Bergamasco B, Bergamini L, Doriguzzi T. EEG sleep pattern as a prognostic criterion in post-traumatic coma. *Electroencephalograph and Clinical Neurophysiology* 1968; 24:374-377.
- Bickford RG, Fleming NI, Dillinger TW. Compression of EEG data by isometric power spectral plots. *Electroencephalography and Clinical Neurophysiology* 1971; 31:632.
- Blishun AF. Comparative Analysis of Methods of Measuring Fuzziness. *English Translation on Tekhnicheskaya Kibernetika* 1989; 1:110-126.
- Borisov AN. *Models of Decision Making Based on a Linguistic Variable*. Zinatne, Riga, 1982.

- Bricolo A, Turella G, Ore GD, Terzian H. A proposal for the EEG evaluation of acute traumatic coma in neurosurgical practice. *Electroencephalography and Clinical Neurophysiology* 1973; 34:789.
- Bricolo A, Turrazzi S, Faccioli F, Odonizzi F, Sciarretta G, Erculiani P. Clinical application of compressed spectral array in long-term EEG monitoring of comatose patients. *Electroencephalography and Clinical Neurophysiology* 1978; 45:211-225.
- Cant BR, Shaw NA. Monitoring by compressed spectral array in prolonged coma. *Neurology* 1984; 35:39.
- Chiappa KH, Burke SR, Young RR. Results of electroencephalographic monitoring during 367 endarterectomies. Use of a dedicated mini computer. *Stroke* 1979; 10:381-388.
- Chiappa KH, Hoch DB. Electrophysiologic monitoring. In: Ropper AH, eds. *Neurological and neurosurgical intensive care*. 3rd ed. New York: Raven Press, 1993:171-183.
- Cho I, Smullens SN, Stretz LJ, Fariello RG. The value of intraoperative EEG monitoring during carotid endarterectomy. *Annals of Neurology* 1986; 20:508-512.
- Devijver PA. *Pattern Recognition: A statistical approach*. Englewood Cliffs, NJ: Prentice Hall International, 1982.
- Emmerson RG, Chiappa KH. Electrophysiologic Monitoring. In: Ropper AH, Kennedy SK, Zervas NT, eds. *Neurologic and Neurosurgical Intensive Care*, 2nd ed. Rockville, MD: Aspen, 1988:13.

- Engel J. *Seizures and epilepsy*. Philadelphia: F.A. Davis, 1989:112-134. (Contemporary Neurology Series.)
- Gallant SI. Optimal linear discriminants. In: *Processings Eighth International Conference on Pattern Recognition*, Paris, France 1986: 849-852.
- Gallant SI. *Neural Network Learning and Expert System*. Cambridge: MIT Press, 1993.
- Garrel S, Maynard R, Barge M, Nicolet A, Reymond F. EEG and clinical study of 107 patients with acute severe traumatic comas. *Revue d Electroencephalographie et de Neurophysiologie Clinique* 1979; 9(2):139-148.
- Gasser T, Bacher P, Mocks J. Transformation towards the normal distribution of broad band spectral parameters of the EEG. *Electroencephalography and Clinical Neurophysiology* 1982; 53(1):119-124.
- Gotman J. Automatic recognition of epileptic seizures in the EEG. *Electroencephalography and Clinical Neurophysiology* 1982; 54:530-540.
- Gotman J. Automatic seizure detection: improvements in evaluation. *Electroencephalography and Clinical Neurophysiology* 1990; 76:317-324.
- Gumnit RJ, ed. *Intensive neurodiagnostic monitoring*. New York: Raven Press, 1987. (Advances in neurology, vol 46.)
- Halimi P, Gozal Y, Cphen M, Gozal D. Computerized electroencephalographic monitoring in anesthesia. *Cahiers d Anesthesiologie* 1990; 38(5):309-317.

- Heuser D, Guggenberger H. Ionic changes in brain ischemia and alteration produced by drugs. A symposium on brain ischemia. *British Journal of Anaesthesiology* 1985; 57:23-33.
- Holmes GL, Rowe J, Hafford J, Schmidt R, Testa M. Prognostic value of the electroencephalogram in neonatal asphyxia. *Electroencephalograph and Clinical Neurophysiology* 1982; 53(1):60-72.
- Holmes GL, Lombroso CT. Prognostic value of background patterns in the neonatal EEG. *Journal of Clinical Neurophysiology* 1993;10(3):323-352.
- Homan RW, Herman J, Purdy P. Cerebral location of International 10-20 System electrode placement. *Electroencephalography and Clinical Neurophysiology* 1987; 66:376-382.
- Jacobson SA, Leuchter AF, Walter DO. Conventional and quantitative EEG in the diagnosis of delirium among the elderly. *Journal of Neurology, Neurosurgery and Psychiatry* 1993; 56:153-158.
- Jasper HH. The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology* 1958;10:371-375.
- Jibiki I, Kurokawa K, Fukushima T, Kido H, Yamaguchi N, Matsuda H, Hisada K. Correlations between quantitative EEG and regional cerebral blood flow in patients with partial epilepsy. *Neuropsychobiology* 1994; 30(1):46-52.
- Jordan KG. Continuous EEG monitoring in the neurological intensive care unit. *Neurology* 1990; 40(Suppl 1):180.

- Jordan KG, Stringer WA. Correlative xenon-enhanced CT cerebral flow (XeCTCBF) and EEG to functionally stratify acute cerebral infarction. *Neurology* 1991; 41(Suppl 1):336.
- Jordan KG. Nonconvulsive status epilepticus in the neuro-ICU detected by continuous EEG monitoring. *Neurology* 1992; 42(Suppl 1):194.
- Jordan KG. Continuous EEG and evoked potential monitoring in the neuroscience intensive care unit. *Journal of Clinical neurophysiology* 1993; 10(4):445-475.
- Jordan KG. Neurophysiologic monitoring in the neuroscience intensive care unit. *Neurologic Clinics* 1995; 13(3):579-626.
- Karnaze DS, Marshall LF, Bickford RG. EEG monitoring of clinical coma: the compressed spectral array. *Neurology* 1982; 32(3):289-292.
- Kayser-Gatchalian MC, Neundorfer B. The prognostic value of EEG in ischemic cerebral insults. *Electroencephalography and Clinical Neurophysiology* 1980; 49:608-617.
- Koloder J, Simpson R, Sycara K. A process model of case-based reasoning in problem solving. In: Joshi A, ed. *Proceedings IJCAI-85*, Los Angeles, CA, 1985; 284-290.
- Langley P. *Element of Machine Learning*. San Francisco: Morgan Kaufmann Publisher, 1996.
- LeCan Y, Boser B, Denker J, Henderson D, Howard R, Hubbard W, Jackel L. Backpropagation applied to handwritten zip code recognition. *Neural Computation* 1, 1989; 4:541-551.

- Leuchter AF, Cook IA, Mena I, Punkin JJ, Cummings JL, Newton TF, Migneco O, Lufkin RB, Walter DO, Lachenbrush PA. Assessment of cerebral perfusion using quantitative EEG cordance. *Psychiatry Research* 1994; 55(3):141-152.
- Loeprecht H, Wolfle K, Heudorfer J, Reich H. Can EEG monitoring with the Trend Analyzer replace stump pressure measurement in carotid surgery? *Langenbecks Archiv fur chirurgie* 1985; 366:333-338.
- Lombroso CT. Seizures in the newborn period. In: Vincken PJ, Bruyn GW, eds. *Handbook of clinical neurology*, vol 15. Amsterdam: North-Holland, 1974:189-218.
- Lombroso CT. Convulsive disorders in newborns. In: Thompson RA, Green JR, eds. *Pediatric neurology and neurosurgery*. New York: Spectrum, 1978:205-239.
- Lombroso CT. Normal and abnormal EEGs in full-term neonates. In: Henry CE, ed. *Current Clinical Neurophysiology*. Update on EEG and evoked potentials. New York: Elsevier / North-Holland, 1980:83-150.
- Lombroso CT. Intracranial hemorrhages in the newborn. A prospective clinical and electrophysiological study of 37 cases. In: *Proceedings of the IYDP commemorative international symposium on developmental disabilities*. Tokyo: ISBN Elsevier Science Publishing Co., 1981:251-256.
- Lombroso CT. Neonatal polygraphy in full-term and premature infants: a review of normal and abnormal findings. *Journal of Clinical Neurophysiology* 1985; 2(2):105-155.

- Lowenstein DH, Aminoff MJ. Clinical and EEG features of status epilepticus in comatose patients. *Neurology* 1992; 42:100-104.
- Lundgren JP. Computerized EEG: Applications and interventions. *Journal of Neuroscience and Nursing* 1990; 22(2):100-112.
- Maynard DE. Development of the CFM: the cerebral function analyzing monitoring (CFAM). *Annals of Anesthesiol Fr* 1979; 3:253-255.
- Maynard DE, Jenkison JL. The cerebral function analysing monitor. Initial clinical experience, application and further development. *Anaesthesia* 1984; 39:678-690.
- Minsky M. Steps toward Artificial Intelligence. In: *Proceeding IRE* 1961; 49:8-30.
- Nuwer MR. Quantitative EEG: I. Techniques and problems of frequency analysis and topographic mapping in clinical settings. *Journal of Clinical Neurophysiology* 1988; 5:1-85.
- Pasupathy AK. *An expert System for EEG Monitoring in the Pediatric ICU*. Master thesis, McGill, 1994.
- Press WH, Teukolsky SA, Vetterling WT, Flannery BP, eds. *Neumerical Recipes in C*. 2nd ed. Cambridge: Cambridge University Press, 1992:683-688.
- Prior PF. *The EEG in acute cerebral anoxia*. 1973.
- Raiffa H. *Decision Analysis*. Addison-Wesley, Reading, MA, 1968.
- Ritchdin CT, Chabot RJ, Alper K, Buyon J, Belmont HM, Roubey R, Abramson SB. Quantitative electroencephalography: A new approach to the diagnosis of cerebral

- dysfunction in systemic lupus erythematosus. *Arthritis and Rheumatism* 1992; 35(11):1330-1342.
- Ropper A, Rockoff MA. Treatment of intracranial hypertension. In: Ropper AH, Kennedy SF, eds. *Neurological and Neurosurgical Intensive Care*. Baltimore: University Park Press, 1983:138.
- Rose A, Lombroso CT. Neonatal seizure states. *Pediatrics* 1970; 45:404-425.
- Rowe JC, Holmes GL, Hafford J, Baboval D, Robinson S, Philipps A, Rosenkrantz T, Raye J. Prognostic value of the electroencephalogram in term and preterm infants following neonatal seizures. *Electroencephalograph and Clinical Neurophysiology* 1985;60(3):183-196.
- Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL, eds. *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, vol. 1. Cambridge: MIT Press, 1986.
- Rumpl E, Lorenzi E, Hackl J, Gerstanbrand F, Hengl W. The EEG at different stages of acute secondary traumatic midbrain and bulbar brain syndromes. *Electroencephalography and Clinical Neurophysiology* 1979; 46:487-497.
- Rumpl E. Electro-neurological correlations in early stages of post-traumatic comatose states. II The EEG at the transition stage to, and at the full stage of the traumatic apallic syndrome. *EEG-EMG Zeitschrift für Elektroenzephalographie Elektromyographie und Verwandte Gebiete* 1980; 11(1):43-50.

- Salinsky MC, Oken BS, Kramer RE, Morchead L. A comparison of quantitative EEG frequency analysis and conventional EEG in patients with focal brain lesions. *Electroencephalography and Clinical Neurophysiology* 1992; 83(6):358-366.
- Shagass C, Roemer RA, Josiassen RC. Some quantitative EEG findings in unmedicated and medicated major depressives. *Neuropsychobiology* 1988; 19(4):169-175.
- Siegel G, Agranoff B, Albers RW, Motinoff P, eds. *Basic Neurochemistry*, 4th ed. New York: Raven Press, 1989.
- Slater GE, Torres F. Frequency-Amplitude Gradient: A new parameter for interpreting pediatric sleep EEGs. *Archives of Neurology* 1979; 36:465-470.
- Szelies B, Mielke R, Herholz K, Heiss WD. Quantitative topographical EEG compared to FDG PET for classification of vascular and degenerative dementia. *Electroencephalography and Clinical Neurophysiology* 1994; 91(2):131-139.
- Trojaborg W, Boysen G. Relation between EEG, regional cerebral blood flow and internal carotid artery pressure during carotid endarterectomy. *Electroencephalography and Clinical Neurophysiology* 1973; 34:61-69.
- Watanabe K, Miyazaki S, Hara K, Hakamada S. Behavioural state cycles, background EEGs and prognosis of newborns with perinatal hypoxia. *Electroencephalography and Clinical Neurophysiology* 1980; 49:618-625.
- Widrow B, Hoff ME. Adaptive switching circuits. In: *Institute of Radio Engineers, Western Electronic Show and Convention, Convention Record* 1986; 4:96-104.

Zadeh L.A. Fuzzy Sets. *Information and Control* 1965: 8:338-342.

Zimmermann HJ. *Fuzzy Set Theory and Its Applications*. Hingham: Kluwer-Nijhoff
Publisher, 1985.