Automatic System for the Removal of Artifacts from EEG Signals using Independent Component Analysis and Hurst Exponent

A thesis submitted in partial fulfillment of the requirements for the award of the degree of

B.Tech. in

INSTRUMENTATION AND CONTROL ENGINEERING

By

Lavanya K (110113042)

Pradeep A (110113060)

Senthil Hariharan A (110113082)

Surya N (110113087)

Arvind V (110113090)

DEPARTMENT OF
INSTRUMENTATION AND CONTROL ENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY
TIRUCHIRAPPALLI-620015

MAY 2017

BONAFIDE CERTIFICATE

This is to certify that the project titled **Automatic System for the Removal of Artifacts from EEG Signals using Independent Component Analysis and Hurst Exponent** is a bonafide record of the work done by

LAVANYA K (110113042)

PRADEEP A (110113060)

SENTHIL HARIHARAN A (110113082)

SURYA N (110113087)

ARVIND V(110113090)

in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **INSTRUMENTATION AND CONTROL ENGINEERING** of the **NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI**, during the year 2016-2017.

Ms.V Sridevi
Project Guide

Dr.B Vasuki
Head of the Department

Project Viva-voce held on _____

Internal Examiner

External Examiner

ABSTRACT

Epilepsy is a neurological disorder marked by sudden recurrent episodes of sensory disturbances, loss of consciousness or convolutions associated with abnormal electrical activity in the brain. Electroencephalogram (EEG) is a diagnostic tool used to pickup the electrical activity of the brain. The scalp EEG signals are mainly affected by artifacts which are generated by physiological and non-physiological activities, such as eye blinks, muscle movements and electrical noises respectively. The removal of these artifacts will improve the performance of the automatic epilepsy detection system.

This work focuses on the removal of the physiological artifacts with the help of Independent component analysis (ICA). The Independent component analysis is a statistical and computational technique for revealing hidden factors that underlie sets of random variables, measurements or signals. It decomposes the original signal into its component signals, arising from different sources. After the decomposition, the removal of unwanted signal is done by Hurst exponent with the values in the range of 0.58-0.66. The components that have values corresponding to artifact are removed and the signal is reconstructed. In order to validate the performance of the algorithm, a time-energy analysis of the signal is made for comparing the artifact-removed signal with the raw signal in order to improve the performance of the Epilepsy detection system.

Keywords: Epilepsy, EEG, artifact removal, Independent component analysis, Hurst exponent

ACKNOWLEDGEMENT

We express our sincere thanks and gratitude to **Mrs. V Sridevi, Assistant Professor,** Department of Instrumentation and Control Engineering, National Institute of Technology, Tiruchirappalli who guided us with valuable inputs continuously during our project work and enabled us to complete the work successfully.

We express our heartfelt thanks to **Dr. B Vasuki, Head of the Department**, Department of Instrumentation and Control Engineering, National Institute of Technology, Tiruchirappalli for allowing us to avail the facilities of the department. Our hearty thanks to the Department Project Evaluation Committee (DPEC), comprising of **Dr. K Srinivasan** and **Dr. K Dhanalakshmi** for their valuable suggestions during the reviews.

We are also thankful to the faculty and staff members of the Department of Instrumentation and Control Engineering, our individual parents and our friends for their constant support and help.

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter 1

Introduction

1.1 SEIZURES AND ITS EFFECTS

1.1.1 CENTRAL NERVOUS SYSTEM

The brain is the central hub for all voluntary and involuntary movements in your body. When abnormal signals interrupt the brain's normal functioning, you can have a seizure. Any type of epileptic seizure could potentially affect your memory, either during or after a seizure. If a patient has lots of seizures, memory problems might happen more often.

Some people have generalized seizures that affect all of the brain. Others have focal seizures (sometimes called partial seizures) that affect only part of the brain. It is possible to have both generalized and focal seizures. If a patient is affected by focal seizures, the way the seizures could affect the memory depends on where in the brain the seizures occurs.

The brain has two halves called hemispheres. Each half has four parts called lobes: the occipital, parietal, temporal and frontal lobes. Abnormalities in the temporal or frontal lobes of the brain are the most common reason for memory problems in people with epilepsy. The left temporal lobe is important for verbal memories such as learning names and remembering facts. If the seizures start in this area, the patient may have problems remembering words and might get stuck mid-sentence while speaking or

writing an exam. The right temporal lobe is important for visual memories like remembering a person's face or finding your way around a place. The frontal lobe is important for prospective memory. Seizures in this area can cause problems in remembering to do things in the future.

EFFECTS OF SEIZURE

After the active portion of a seizure, there is typically a period of confusion called the postictal period before a normal level of consciousness returns. This usually lasts 3 to 15 minutes but may last for hours. Other common symptoms include: feeling tired, headache, difficulty in speaking and abnormal behavior. Psychosis after a seizure is relatively common, occurring in between 6 to 10% of people. Often, people do not remember what occurred during this time.

1.2 ELECTROENCEPHALOGRAPHY

Electroencephalography(EEG) is a technique that records the electrical activity of the brain. During an EEG test, small electrodes (cup or disc type) are placed on the scalp. They pick up the brain's electrical signals and send them to a machine called electroencephalogram, which records the signals as wavy lines onto a computer screen or paper in the order of microvolts.

1.2.1 WORKING

The EEG is measured directly from the scalp or using depth probes. When neurons are activated, local current flows are produced. A potential of 60–70 mV with negative polarity may be recorded under the membrane of the cell body. This potential changes if an action potential travels along the fibre, which ends in an excitatory synapse. An excitatory post-synaptic potential (EPSP) occurs in the following neuron. Also, if two action potentials travel along the same fibre over a short distance, there will be a sum-

mation of EPSPs producing an action potential on the post-synaptic neuron, providing a certain threshold of membrane potential.

Large potential differences across a neuron can also be caused by action potentials(AP). APs are caused by an exchange of ions across the neuron membrane and it is a temporary change in the membrane potential that is transmitted along the axon. APs are initiated by many different types of stimuli; sensory nerves respond to many types of stimuli such as chemical, light, electricity, pressure, touch and stretching.

Only large populations of active neurons can generate electrical activity that are recordable on the head surface. Between electrode and neuronal layers, current penetrates through skin, skull and several other layers. Weak electrical signals detected by the scalp electrodes are massively amplified and then stored to computer memory.

1.2.2 BRAIN WAVES

Let us now take a brief look at the different brain waves that are recorded by the electroencephalogram. Brain patterns form wave shapes that are commonly sinusoidal. Usually, they are measured from peak to peak and normally range from 0.5 to 100mV in amplitude, which is about 100 times lower than ECG signals. Power spectrum is derived by Fourier transform of the raw EEG signal. The spectrum is continuous, ranging from 0 Hz up to one half of sampling frequency. The brain state of the individual may make certain frequencies more dominant. Brain waves have been categorized into four basic groups for obtaining basic brain patterns of individuals.

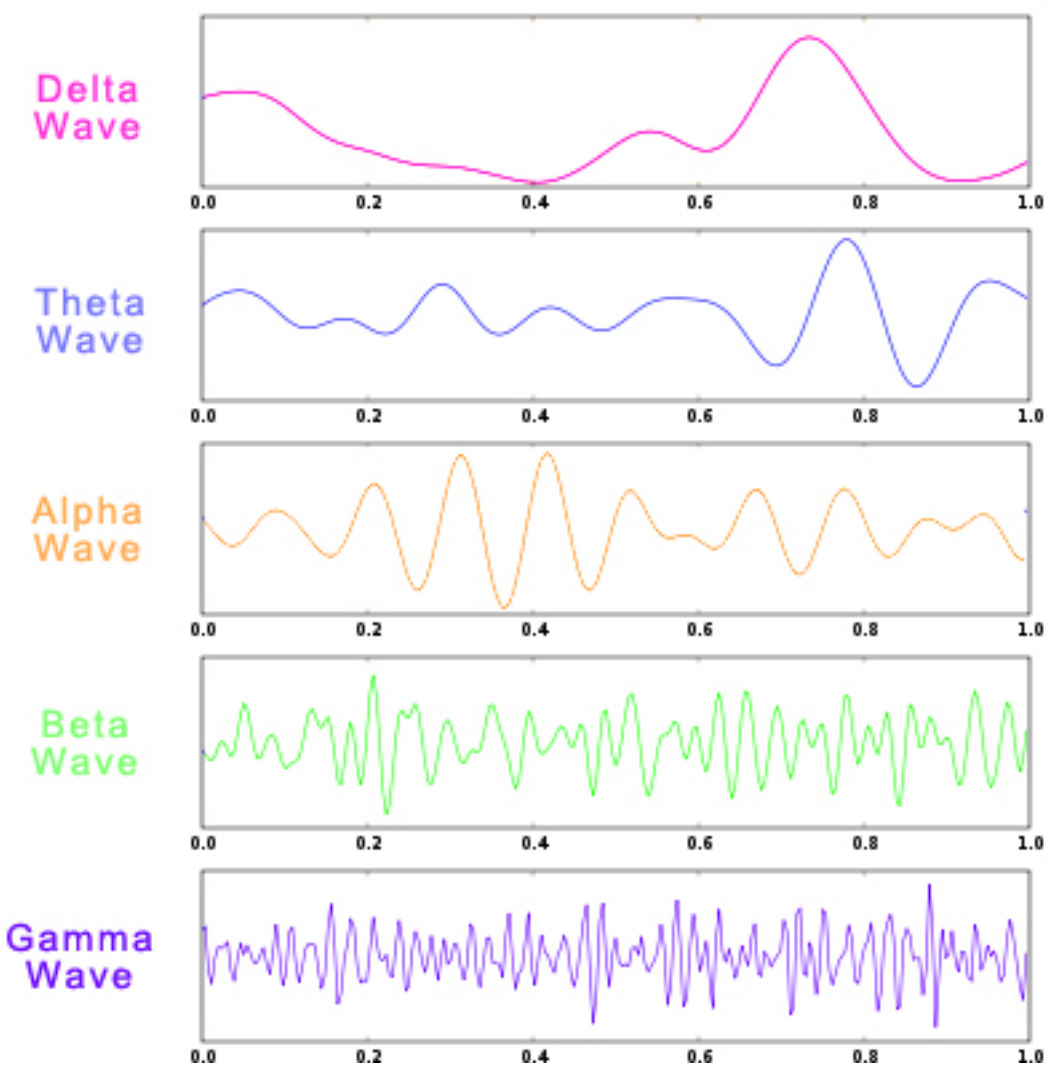

Figure 1.1: Brain Waves

The best-known and most extensively studied rhythm of the human brain is the normal alpha rhythm. Alpha can be usually observed better in the posterior and occipital regions with typical amplitude of about 50 microvolts (peak-peak). Alpha waves are usually attributed to summated dendrite potentials. During normal state of wakefulness with open eyes, beta waves are dominant. In relaxation or drowsiness, alpha activity rises and if sleep appears, power of lower frequency bands increase. Various regions of the brain do not emit the same brain wave frequency simultaneously. An EEG signal between electrodes placed on the scalp consists of many waves with different characteristics. A large amount of data received from even one single EEG recording presents a

WAVE	FREQUENCY (Hz)	AMPLITUDE (micro V)	REMARKS
Beta	14-30	2-20	The most common form of brain waves. Present during mental thought and activity
Alpha	8-13	20-60	Easily produced when quietly sitting in a relaxed position with eyes closed. Few people have trouble producing alpha waves. Alpha blockade occurs with mental activity exceptions in the case of mental arithmetic, archery, and golf putting
Theta	4-7	20-100	Believed to be more common in children than adults. A study found these waves to be related to displeasure, pleasure and drowsiness.
Delta	0.5-3.5	20-200	Found during periods of deep sleep in most people. Characterized by very irregular and slow wave patterns. Useful in detecting tumors and abnormal brain behaviors.
Gamma rays	36-44	3-5	Occur with sudden sensory stimuli.

Table 1.1: Brain Waves- Neurological signals

difficulty for interpretation. Individual's brain wave patterns are unique. In some cases, it is possible to distinguish persons only according to their typical brain activity.

1.2.3 ELECTRODE PLACEMENT

The 10/20 systems or International 10/20 system is a globally recognized method to describe the location of scalp electrodes. The system is based on the relationship between the location of an electrode and the underlying area of cerebral cortex. The head is divided into proportional distances from prominent skull landmarks (nasion, preauricular points, inion) to provide adequate coverage of all regions of the brain. Label 10-20 designates proportional distances in percents between ears and nose where points for electrodes are chosen. Thus, the number '10' and '20' refer to the fact that the distances between adjacent electrodes are either 10% or 20% of the total front-back or right –left distance of the skull. Each site has a letter to identify the lobe and a number to identify location.

Electrode	Lobe
F	Frontal
T	Temporal
C	Central
P	Parietal
O	Occipital

Table 1.2: Electrode Position

NOTE: No central lobe exists. The 'C' letter is used for identification purpose only. The 'z' (zero) refers to an electrode placed on the mid line. Even numbers (2,4,6,8) refer to electrode positions on the right hemisphere. Odd numbers(1,3,5,7) refer to electrode positions on the left hemisphere.

Four anatomical landmarks are used for the essential positioning of the electrodes: First, the nasion which is the point between the forehead and the nose; Second, the inion which is the lowest point of the skull from the back of the head and is normally indicated by a prominent bump; Third, the pre-auricular points anterior to the ear. Extra positions can be added by utilizing the spaces in between the existing 10/20 system.

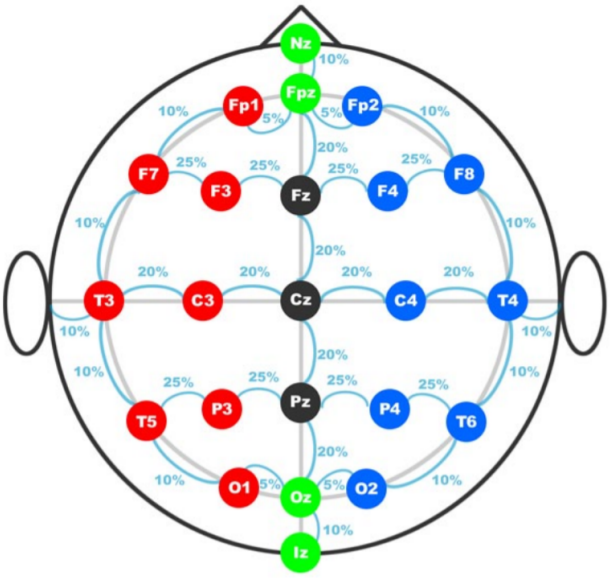

Figure 1.2: Electrode Location

As it is known from tomography, different brain areas may be related to different functions of the brain. Each scalp electrode is located near certain brain centers, e.g. F7 is located near centres for rational activities, Fz near intentional and motivational centers, F8 close to sources of emotional impulses. Locations near P3, P4, and Pz contribute to activity of perception and differentiation. Near T3 and T4 emotional processors are located, while at T5, T6 certain memory functions stand. Primary visual areas can be found below points O1 and O2.

Physical references can be chosen as vertex (Cz), linked-ears, linked-mastoids, ipsilateral-ear, contralateral-ear, C7 reference, bipolar references, and tip of the nose. Reference-free techniques are represented by common average reference, weighted average reference, and source derivation. Each technique has its own set of advantages and disadvantages. The choice of reference may produce topographic distortion if relatively electrically neutral area is not employed. Linking reference electrodes from two earlobes or mastoids reduces the likelihood of artificially inflating activity in one hemisphere.

Cz reference is advantageous when it is located in the middle among active electrodes. The combination of all active electrodes with reference and ground electrode compose channels. The general configuration is called montage. There are three montages they are Longitudinal bipolar, horizontal bipolar, referential. Bipolar is the potential difference between 2 active electrodes, and referential is potential difference between 1 active and 1 inactive electrode.

1.3 ARTIFACTS

EEG Artifacts are recorded signals that are non-cerebral in origin. They may be divided into one of two categories depending on their origin: physiological artifacts or non-physiological artifacts.

Physiological artifacts arise from a variety of body activities that are either due to:

1. *Movements*: movements of the head, body or scalp (e.g., pulsations of the scalp arteries, scalp muscle movement)

2. *Bioelectrical potentials*: from moving electrical potentials within the body (such as those produced by eye, tongue and pharyngeal muscle movement), or electrical potentials generated by the scalp muscles, heart or sweat glands, or

3. *Skin resistance changes*: due to sweat gland activity, perspiration and vasomotor activity.

Non-physiological artifacts arise from two main sources:

1. *External Electrical interference* from other power sources such as power lines or electrical equipment

2. *Internal electrical malfunctioning* of the recording system arising from recording electrodes, electrode positioning, cables, amplifiers, pen motors or the paper drive.

1.3.1 ARTIFACTS FROM THE PATIENT

One of the most common pitfalls of EEG interpretation is to mistakenly identify non-cerebral potentials as originating from the brain. The patient and record must be closely observed throughout the recording and notations made whenever artifacts occur.
In many instances, artefacts can be immediately recognised by applying the following two rules of spatial analysis:

1. Medium to high amplitude potentials that occur at only one electrode usually do not arise from the brain. Cortically generated potentials exhibit a physiological

distribution over the scalp characterized by a potential maximum that gradually drops off in voltage with increasing distance across the scalp. Therefore comma a prominent waveform that can only be recorder from one electrode is artefact until proven otherwise.

2. Repetitive, irregular or rhythmical waveforms that appear simultaneously in unrelated head regions are usually not cerebral in origin. Evolving electrographic seizures or abrupt background abnormalities typically spread to involve adjacent electrodes. They do not jump to opposite ends of the head or to non-homologous (non-mirror image) areas of the opposite hemisphere.

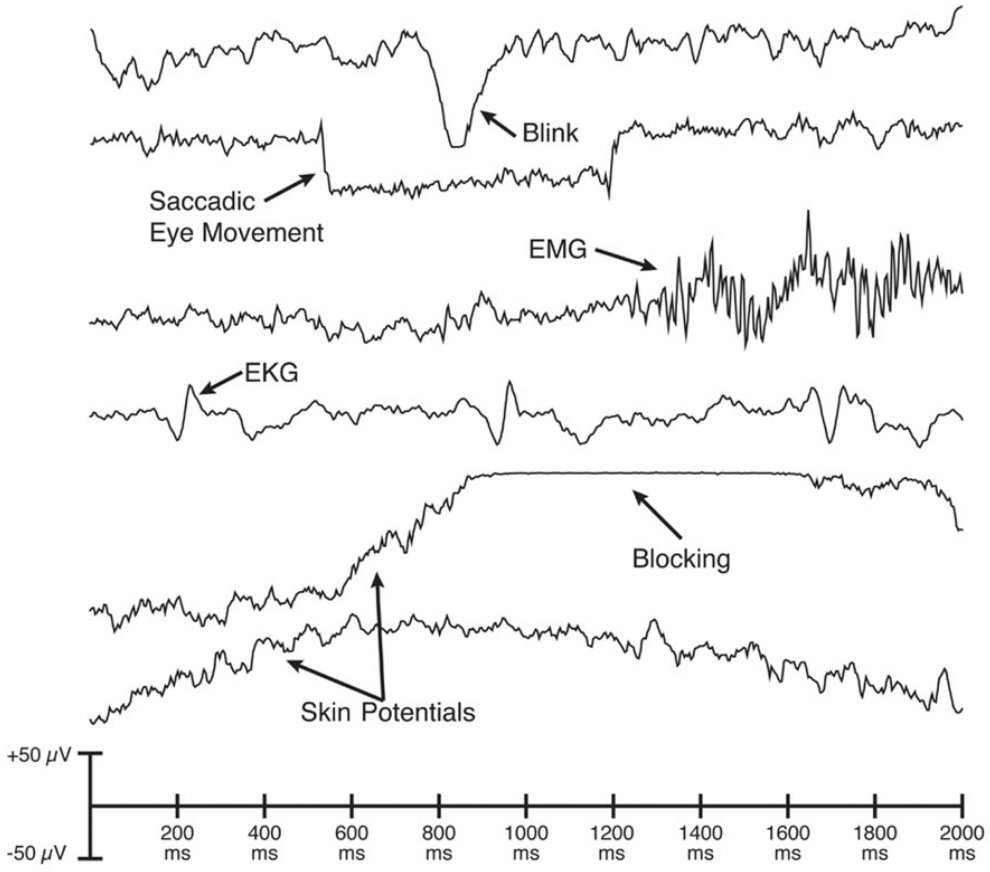

Figure 1.3: Common types of artifacts

1.3.2 BLINKING AND OTHER EYE MOVEMENTS

These movements cause potential changes that are picked up mainly by frontal electrodes, although they may extend into central and temporal electrodes. A simple but useful way of understanding eye movement artefact is to picture the front of the eye as a positive charge that either moves towards or away from the recording electrodes. The electrodes that record the largest potential change with vertical eye movements are Fp1 and Fp2 because they are placed directly above the eye. The electrodes that record the largest potential change with horizontal (lateral) eye movements are F7 and F8 because they are approximately lateral to the eyes. In a typical longitudinal bipolar montage an upward vertical eye movement (e.g., eye closure, eye blink) will produce a downward deflection in Fp1-F3 or Fp1-F7 because the positively charged cornea is moving towards Fp1 making it increasingly more positive.

Rapid eye movements may cause jagged artifacts. Muscle artefact may appear along with eye movements. Lateral eye movements may be preceded by a single sharp muscle potential sometimes referred to as a *lateral rectus spike*. Rarely, a lateral rectus spike in combination with the eye movement artefact may mimic abnormal epileptiform spike and wave activity.

Eye movement artifacts have long been believed to be due to movement of the eyeball which carries a steady electrical charge, the cornea being about 100mV positive with respect to the retina. However, it seems that movement of this corneoretinal dipole is not necessary to produce blink artifacts: movements of the lids across the eyeball can produce a similar artifact. Moreover, some low amplitude eye movement artifacts can be recorded even after removal of the eye including cornea and retina, suggesting that movement of residual membranes deep in the orbit can cause artifacts. Eye movement artifacts in the EEG can usually be identified by their frontal distribution, their symmetry and their characteristic shape.As a general rule, however, it is best to assume that activity in the alpha frequency range localized to the frontopolar head regions is eye

movement artifact until proven otherwise.

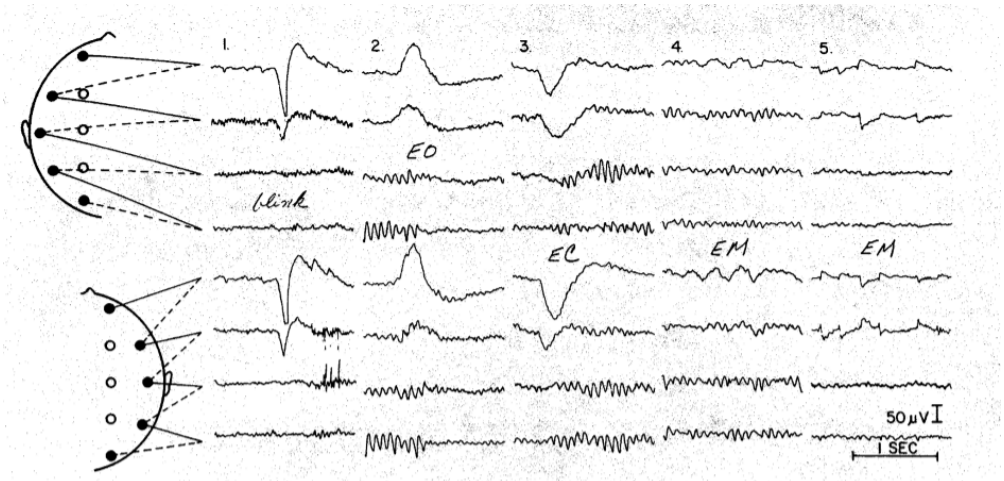

Figure 1.4: Eye blink artifacts and Electrode location (EO - eye open EC - eye close)

1.3.3 MUSCLE ARTIFACTS

Muscle activity causes very short duration potentials that usually occur in clusters or periodic runs. If they recur as discrete potentials with the same shape and in the same distribution, they may resemble cerebral spike discharges except that most cerebral spikes are of much longer duration than muscle action potentials. If they recur in rapid bursts of discharges, they can produce several different types of potentials that can merge and obscure the recording of cerebral activity.

Muscle artifacts from scalp and face muscles occur mainly in the frontal and temporal regions but may be recorded by electrodes nearly anywhere on the head. Reducing the settings of the high frequency filter will reduce the amplitude of these fast potentials, but will also change their form so that single muscle potentials may look more like spikes, and repetitive potentials may look like cerebral fast waves.

Muscle artefact, even if not related to recognizable movement by the patient, is usually easily identified by its shape and repetition. It can be reduced and often eliminated by asking the patient to relax, drop the jaw or open the mouth slightly, or change posi-

tion. Artifact from a single electrode can sometimes be stopped by gently pushing on that electrode, by stroking or massaging the skin near the electrode, or by reapplying the electrode. Reducing high frequency filter settings is only of limited value because of the aforementioned distortion.

A few specific conditions cause special electrographic patterns. Repetitive movements such as chewing, blinking or tremor may give rise to a combination of fast muscle and slow movement artifacts which may resemble cerebral discharges, especially if the combinations repeat with similar shape. Such rhythmical combinations may occur in tremor of Parkinson's disease, which is characterized by rhythmic 4-6 Hz waveforms that may appear in isolated electrodes or seem to have a physiological distribution consistent with cerebral activity.

Indeed, any mono-rhythmic theta pattern that occurs abruptly and seems isolated from the rest of the ongoing background activity should raise the question of tremor artifact. On the other hand, seizures may lead to muscle activity so that a recording electrode picks up mixtures of cerebral seizure activity and muscle activity caused by the seizure.

Chapter 2

Review Of Literature

2.1 OVERVIEW

Eye movements, eye blinks, muscle noise, heart signals, and electrical line noises are often responsible for large and distracting artifacts in EEG recordings. Though fixating subject's vision to a visual target reduces voluntary eye movements, involuntary eye movements such as blinks cannot be avoided. Most commonly, in research, EEG segments with artifacts exceeding an arbitrarily fixed threshold are rejected to improve the signal's quality. Though this method is widely applied in medical and research environment, it results in the loss of experimental data too. This becomes especially problematic if only a few epochs are available and artifacts such as blinks or movements occur too frequently. Moreover, this approach is inappropriate when working with the continuous EEG such as in real-time Brain Computer Interface (BCI) applications, and online mental state monitoring (T. Jung et al. [1]).

2.2 SUBTRACTION AND REGRESSION

Subtraction was the initially proposed methods for artifact rejection. In the case of artifact removal through subtraction, one would likely use a set of electrodes around the eye of the patient to receive accurate EOG readings. This would then be subtracted from the EEG signal to remove the noise. The problem with this techniques is that most of these reference signals are not uncorrelated with the input signals [T. Jung et al. [1]].

2.2.1 REGRESSION

Regression methods in either time or frequency domain depends on having a good regressing channel (e.g., EOG), and share an inherent weakness that spread of excitation from eye movements and EEG signals is bidirectional. Therefore, whenever regression-based artifact removal is performed, relevant EEG signals contained in the EOG channel(s) are also cancelled out in the "corrected" EEG channels along with the eye movement artifacts. Regression in the time or frequency domain (Kenemans et al. [2]), they concentrate mainly on removing ocular artifacts, may themselves introduce new artifacts into the EEG recording (Weerts and Lang [3])

Weerts and Lang [3] discusses about the overcompensating effect of regression in the time domain. Further, regression could also introduce new artifacts into the EEG signal. This overcompensation is attributed to the difference between the EOG-to-EEG transfer functions for blinks and saccades. Saccade artifacts arise from changes in orientation of the retinocorneal dipole, whereas blink artifacts arise from alterations in ocular conductance produced by contact of the eyelid with the cornea (Overton and Shagass [4]). The pickup of blink artifacts on the recording electrodes decreases rapidly with distance from the eyes, whereas the transfer of saccade artifacts decreases more slowly, so that at the vertex the effect of saccades on the EEG is about double that of blinks (Overton and Shagass [4]).

Regression in the frequency domain (Whitton et al. [5]; Woestenburg et al. [6]) can account for frequency-dependent transfer function differences from EOG to EEG, but is acausal and thus unsuitable for real-time applications. Kenemans et al. [2] proposed a time domain multiple-lag regression method capable of taking into account frequency- and phase-dependent differences in EOG-to-EEG transfer functions. Their method can be viewed as a causal time-domain equivalent of frequency-domain methods. However,

the method requires considerably more computation than its frequency-domain counterpart, and was not found to be better than simple time-domain regression in tests on actual EEG data (Kenemans et al. [2]).

2.3 OTHER STATISTICAL METHODS

Berg and Scherg [7] proposed a statistical technique for removing ocular artifacts using principal component analysis (PCA). First, they collected EEG and EOG signals simultaneously while the subject performed some standard eye movements and blinks. Then, a PCA of the variance in these calibration signals gave major components representing blinks and horizontal and vertical eye movements. Corrected EEG data could be obtained by removing these components through the simple inverse computation. They showed that ocular artifacts can be removed more effectively by the PCA method than by regression or by using spatiotemporal dipole models. Lagerlund et al. [8] showed that PCA methods cannot completely separate some artifacts from cerebral activity, especially when they both have comparable amplitudes.

Lagerlund et al. [8] uses a principle component analysis (PCA) based method for the removal of artifacts from an EEG signal (particularly eye-blink artifacts). This method separates a set of mixed signals into a set of linearly uncorrelated 'principal components' through finding orthogonal components with the largest possible variance from the remaining data, the component which most accurately represents the artifact is then removed and the data is reconstructed. Although this process is more effective than many earlier methods, it is limited in that it cannot completely separate artifacts from the remaining data due to the fact that orthogonality between EEG data and artifact is not guaranteed. This discrepancy is particularly apparent when the artifact and brain signals have similar amplitudes.

2.4 INDEPENDENT COMPONENT ANALYSIS

Jung et al. [1] and Makeig et al. [9] propose a promising method using independent component analysis (ICA) proposed by Bell and Sejnowski [10]; Comon,[11]; for data decomposition and separation of neuronal activity from artifacts (Fitzgibbon et al. [12]). The idea central to this method is that the EEG signal is a mixture of linearly independent source components (IC) that can be separated by ICA, visually examined, and classified as artifact or EEG signal components. Once the artifact components have been identified, they can be removed and the remaining EEG signal components can be projected back to the original time domain. This procedure yields the reconstruction of an artifact-free EEG. The visual classification in this case is done by inspecting the 2D scalp map projections of the ICs, called topoplots. Based on the topoplots' image pattern and on the experts' know-how, a decision on whether the IC is an artifact or an EEG signal component is made.(Radüntz T et al. [13]).

However, ICA cannot guarantee that some individual independent components (ICs) contain only noise and do not contain information about useful sources, especially in biomedical applications. Hence, the problem of detection and filtering of useful part of each independent component is still open, and additional tools are needed to solve it. In this work, we propose a noise reduction technique for multisensory signals described by model using the recently developed extension of improved cumulant independent component analysis (Tobias Blaschke and Laurenz Wiskott [14]) for the decomposition of EEG signals. The extended algorithm provides a much faster way to decompose the EEG signals into independent components. Once the independent time courses of different brain and artifact sources are extracted from the data, corrected EEG signals can be derived by eliminating the contributions of the artifacts. The automated detection of artifacts from independent components is performed through classification using Hurst exponent.

Chapter 3

Independant Component Analysis

Independent Component Analysis(ICA) is a statistical technique utilized to recover a set of independent component signals from a set of measured signals. It is assumed that each measured signal is a linear combination of the independent signals. Though the number of independent components could be arbitrary, for simplicity it is always assumed there are an equal number of measured and independent signals. The ICA carried out in this project deals with the separation of the different source signals that arise from the cerebral cavity, physiological and non-physiological artifacts. The following algorithm is carried out to separate the signals and remove the artifacts based on their non-gaussianity.

From statistical theory, the Central Limit Theorem states that the sum of several independent random variables tends towards a Gaussian distribution. For example, the sum of a pair of dice approximates a gaussian distribution with a mean of seven. This implies that if we can find a combination of the measured signals in the given output signal with minimal gaussian properties, then that will be one of the independent signals. From the obtained signal, one can easily find out the original independent components using simple mathematics. In order to find this, some way to measure the non-gaussianity of the input signal is needed. Kurtosis and Negentropy are few of the methods used to determine the non-gaussianity of the system.

3.1 KURTOSIS

Kurtosis is the classical method of measuring non-gaussianity. When data is pre-processed to have unit variance, Kurtosis is equal to the fourth moment of the data. In a general sense, Kurtosis measures the ”spikiness” of a distribution. Kurtosis is extremely simple to calculate, however, it is very sensitive to outliers in the data set. A single data point that is an outlier would change the entire data.

3.2 NEGENTROPHY

In a discrete signal, the entropy is equal to the sum of the products of probability of each event and the log of those probabilities. A value called differential entropy can be found for continuous function that uses the integral of the function times the log of the function. Negentropy is simply the differential entropy of a signal Y, minus the differential entropy of a gaussian signal with the same co-variance of Y. Negentropy is always positive and is zero only if the signal is pure gaussian. Although stable, Negentropy has computational complexity associated with it, which causes the system to take a longer time to find the gaussianity. Here, we use computational methods to calculate the Negentropy and hence reduce the input matrix to one that is made up of statistically independent components.

3.3 ICA ALGORITHM

Consider the following 2 sinusoidal signals of frequency 100Hz (A) and 200Hz (B).

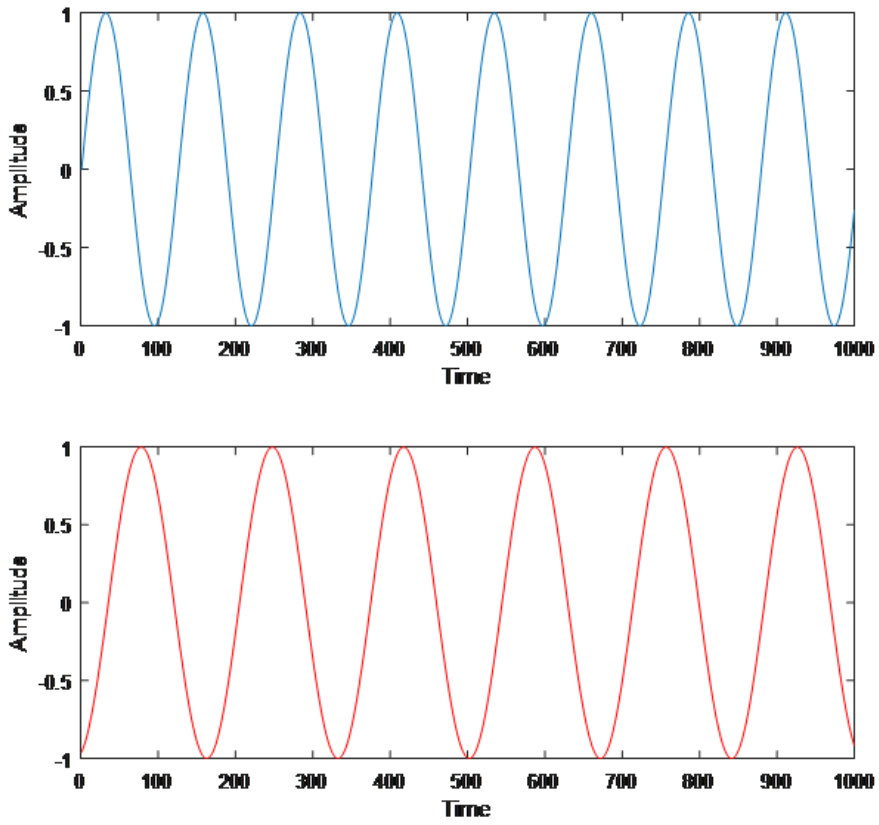

Figure 3.1: sinusoidal signals of 100Hz (top) and 200Hz (bottom) frequency

We then mix linearly these two sources. The top curve is equal to A minus twice B and the bottom the linear combination is 1.73*A +3.41*B.

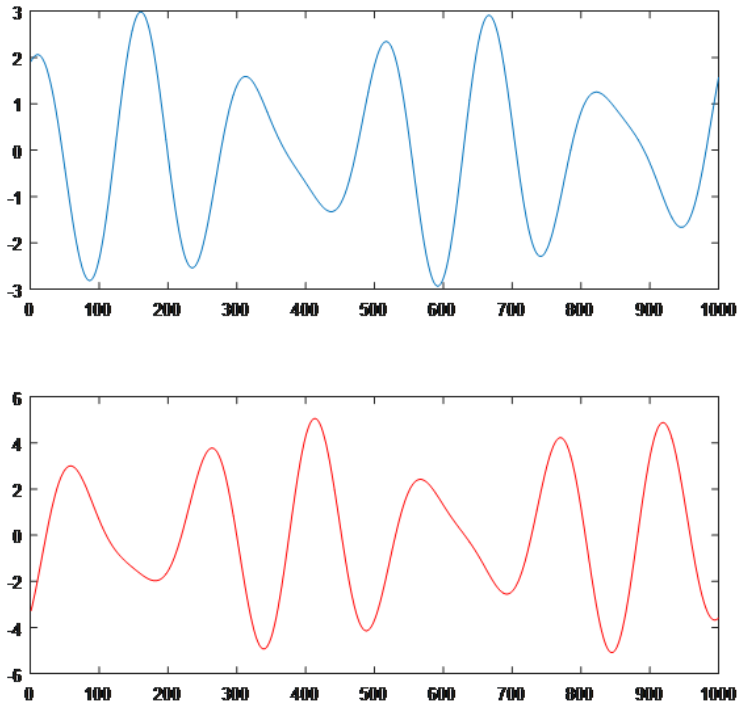

Figure 3.2: Mixed signals

We make a plot such that the value of A is the abscissa of the data point and the value of B is their ordinates.

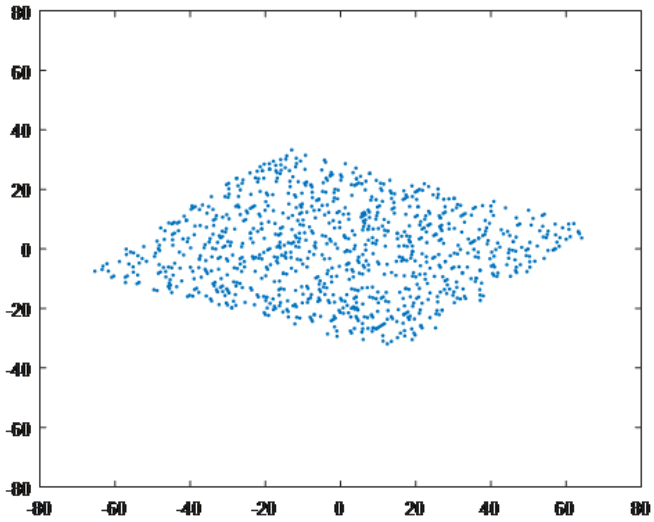

Figure 3.3: Scatter plot of linear mixtures of A and B

Then if we whiten the two linear mixtures, we get the following plot

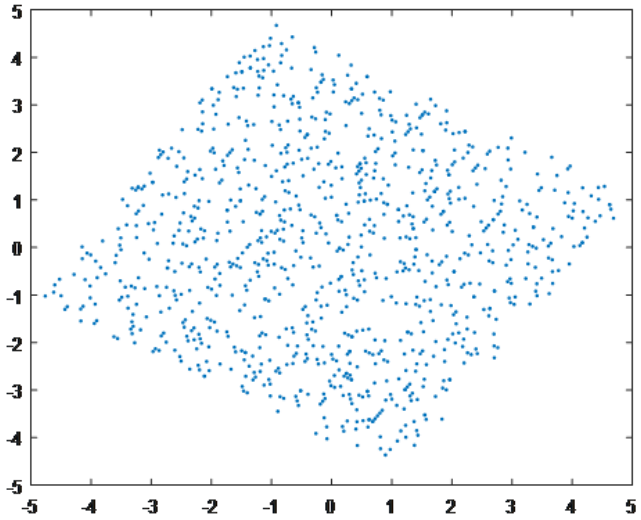

Figure 3.4: Scatter plot of linear mixtures of A and B after whitening

the variance on both axis is now equal and the correlation of the projection of the data on both axis is 0 (meaning that the covariance matrix is diagonal and that all the diagonal elements are equal). Then applying ICA only mean to "rotate" this representation back to the original A and B axis space. Intuitively you can imagine that ICA rotates the whitened matrix back to the original (A,B) space. The following graphs show the scatter plot with their projections along X and Y axis before and after the application of ICA.

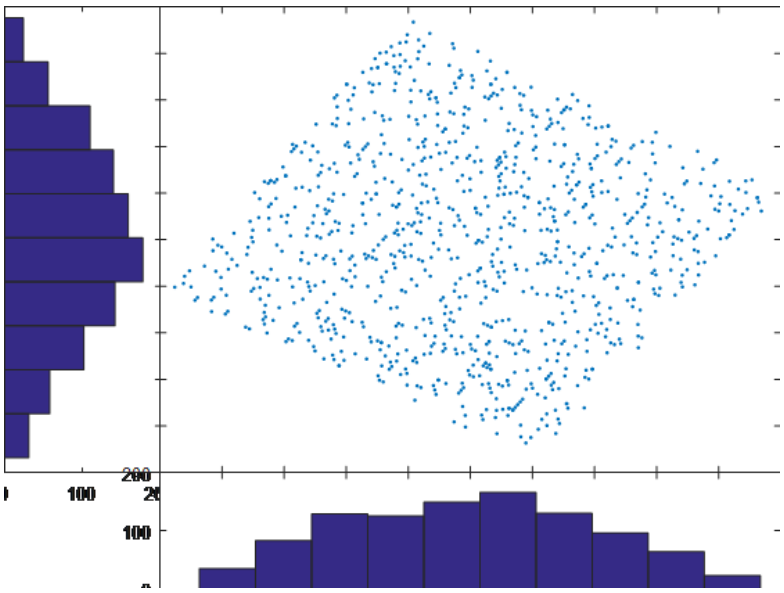

Figure 3.5: Scatter plot of linear mixtures of A and B after whitening

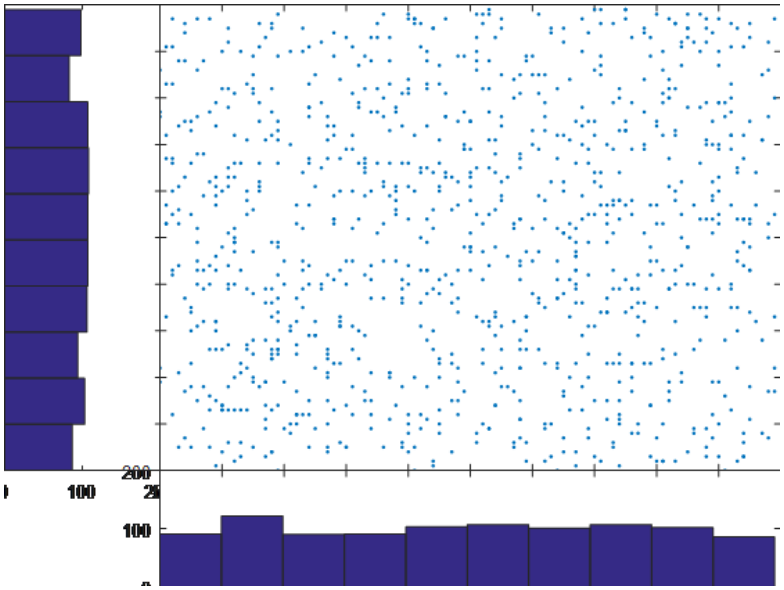

Figure 3.6: Scatter plot of linear mixtures of A and B after ICA

STEP 1: Centering

$$X = X' - E\{X\} \tag{3.1}$$

STEP 2: Whitening

$$Y = VX \tag{3.2}$$

Here, we select V which is a linear transformation from X to Y such that the correlation in the input data is reduced.

Here Y is such that

$$E\{YY'\} = I \tag{3.3}$$

Taking $Cov\{X\} = C$, That is $E\{XX'\} = C$ the above said condition can be achieved by taking $V = C^{-}1/2$

$$E\{YY'\} = E\{VXX'V'\} = C^{-1/2}CC^{-1/2} = I \tag{3.4}$$

STEP 3: Computing Rotation

Objective: To increase the non-gaussianity of the system thereby increasing the independence among the components

$$Obj(W) = \sum_{i=1}^{T} G(W^T X) - \lambda(W^T W - I) \tag{3.5}$$

$$\delta Obj = X g(W^T X)^T - W = 0 \tag{3.6}$$

Where g( ) is the derivative of G( ), W is the rotation matrix and is lagrange Multiplier

Algorithm for rotation:

Input: Y (Random initialization W)

Iterate the below till convergence

$$S = W^T Y \tag{3.7}$$

$$W = Y g(S)^T \tag{3.8}$$

$$W = W(W^T W) - 1 \tag{3.9}$$

Outputs are W,S

3.4 HURST EXPONENT

First, we apply ICA to obtain the independent components(ICs). The independent components have to be further classified to artifact carrying signals and useful signals. Clearly, there are several possibilities for classification of signals based on a "usefulness" criterion. Most commonly implemented method is to visually identify and analyze the independent component for segments containing EOG like pulses and reject them. But that system is labour intensive as an expert is required to perform artifact rejection manually. There also exists the need to specify the sense of "usefulness" in EEG-analysis application. This is not a trivial problem, for it may vary with each case. However, we always can be certain that if a signal has no temporal structure or is independent identically distributed (i.i.d.) it can give us no information for analysis, except the information that the signal is unpredictable. Such a signal can be rejected from the analysis. But, electrical noise and artifacts can still have a temporal structure and hence, we need to define a more general measure of randomness. Here we discuss the Hurst exponent as a measure of "usefulness" of the signal.

The Hurst exponent H is one of the parameter used to characterize a time-series data. (Hurst et al.) developed the (R/S) analysis for a time-series $y(t)$, where t represents the instant of time. The range R is defined as the difference between the maximum and the minimum accumulated value.

$$R(t) = \max_{1\leq t\leq T}\{Y(t,T)\} - \min_{1\leq t\leq T} Y(t,T) \tag{3.10}$$

where

$$Y(t,T) = \sum_{t=1}^{T}[y(t) - < y(t) >] \tag{3.11}$$

and secondly, the standard deviation S was estimated from the observed value $y(t)$

$$S = (\frac{1}{T}(\sum_{t=1}^{T}[y(t) - < y(t) >]^2))^{1/2} \tag{3.12}$$

Hurst found that the ratio R/S is very well described for a large number of phenomena by the following nonlinear empirical relation:

$$\frac{R}{S} = (cT)^H \tag{3.13}$$

where T is the number of samples, c is some constant, and H is the Hurst exponent in the range from 0 to 1. The Hurst exponent of value 0.5 corresponds to a time series that is truly random (e.g.,Brown noise). The Hurst exponent of $0 < H < 0.5$ exhibits the so-called anti-persistent behavior; e.g., white uniformly distributed noise has H = 0.15 approx. At the limit of H = 0, the time series must change direction every sample. On the other hand, the Hurst exponent of $0.5 < H < 1$ describes a temporally persistent or trend-reinforcing time series. Similarly, a straight line with nonzero slope will have the Hurst exponent of 1. It was found by many researchers that the Hurst exponent H has a value equal to 0.70–0.76 for many natural, economic, and human phenomena. It was observed that EOG artifact fall under the range of 0.58-0.66. Hence components with Hurst exponent within this range can be rejected safely.

Chapter 4

Methodology and Results

4.1 INTRODUCTION

This work focuses on building an automatic system that processes the given EEG data using statistical methods, identifying noise-laden independent components of the EEG signal and removing them, and finally reconstructing the components to get back the original data with minimum artifacts affecting the important data in the signal. The system is designed with the help of MATLAB(2015) software and with special toolkits including BioSIGPLOT for the process. This chapter covers the various steps taken by the system and the analysis of the results obtained in the process.

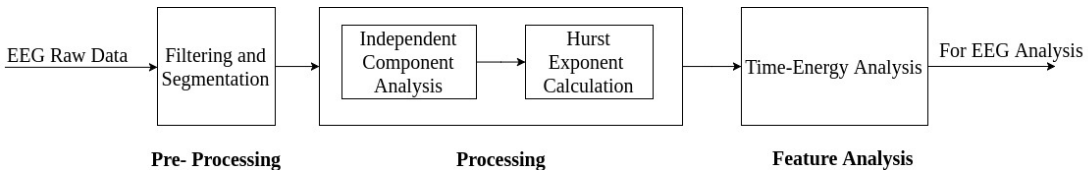

Figure 4.1: Block Diagram

4.2 DATASET

A continuous 15 minute 21-channel EEG dataset with a sampling frequency of 400Hz is considered for the artifact analysis. The EEG dataset has known regions which suffer from artifacts caused by electrical line noise, EOG and motion artifacts. The pre-processing and processing stages of the dataset has been detailed below.

4.3 PRE-PROCESSING

The major operations performed in the pre-processing stage are filtering and data segmentation. The input raw EEG data is affected by the low frequency noise (<0.3Hz), electrical line noises and its harmonics. A notch filter is designed to reject the line noise of 50Hz and its harmonics till 200Hz. The filter has been designed with a notch width of 0.1 Hz. Further, the low frequency components are discarded using a high pass filter.

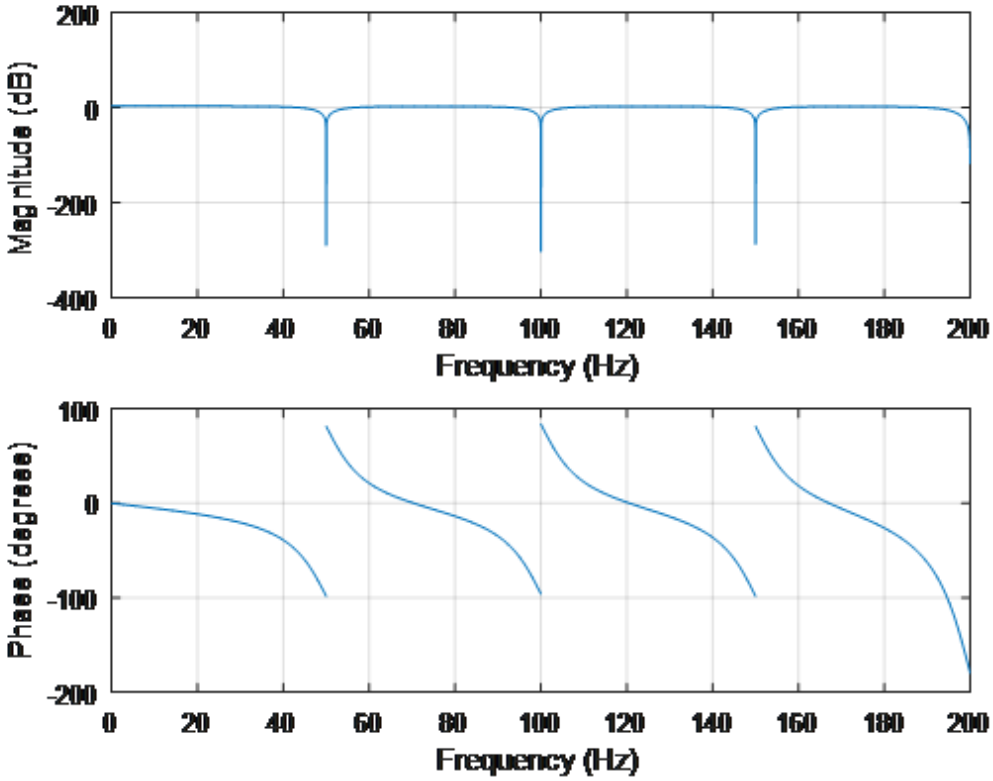

Figure 4.2: Bode plot - notch filter

Rejecting independent components with artifact-laden signals also cause loss of important neurological data, since the ICA is not completely efficient in dissolving the raw EEG data into artifact and neurological components. The main idea behind segmenting the input raw EEG data into smaller components is to minimize the loss of useful data in segments not affected by EOG signals. Here we have segmented the input EEG dataset into 100 second segments to obtain a trade-off between efficient ICA decomposition and prevent loss of useful data.

4.4 PROCESSING

The filtered EEG data segment was observed to contain EOG artifacts which are characterized by low frequency and high amplitude pulse.

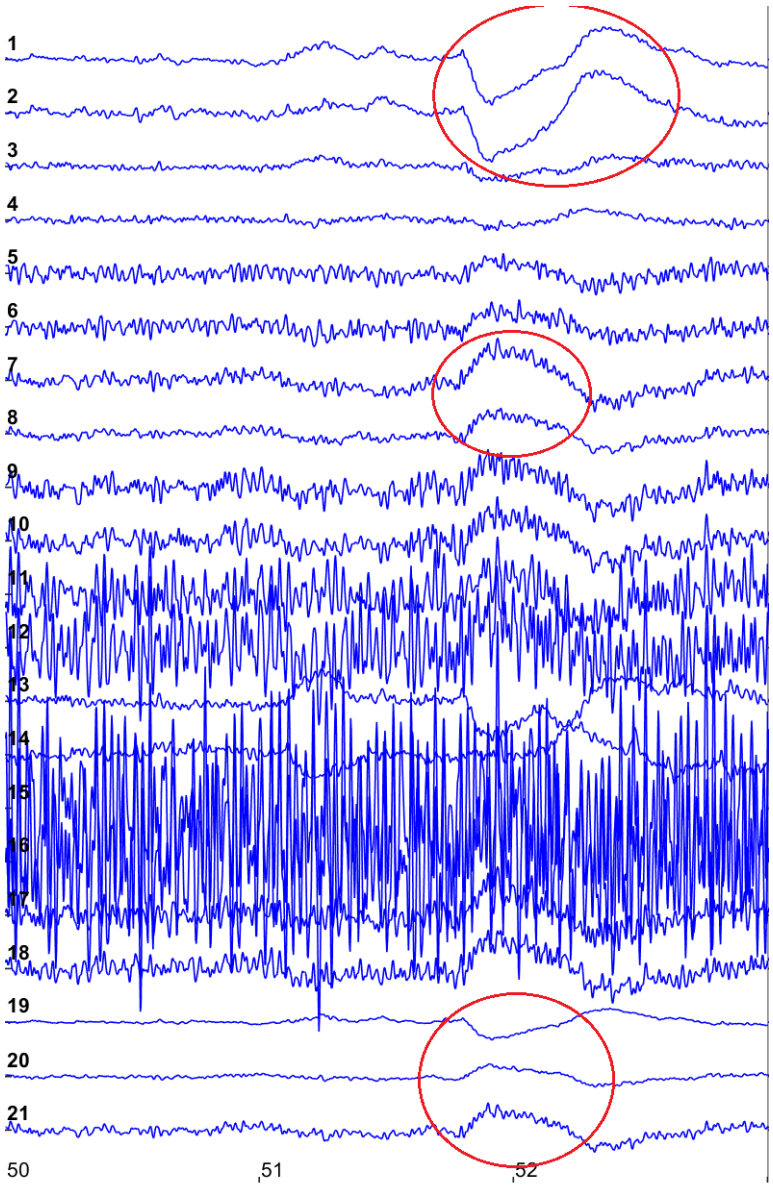

Figure 4.3: 21-channel raw EEG data laden with EOG artifact

The filtered and segmented data is decomposed using Improved Cumulant ICA (CUBICA). The Improved Cumulant ICA is an extension of the ICA algorithm, which provides faster decomposition as compared to the commonly used FastICA algorithm. The independent components are not comparable in amplitude to the raw EEG signal. This is because the original amplitude can be realised only on multiplication with the mixing

matrix. Analyzing the independent components visually we observe the certain independent components contain EOG pulses in accordance with the time in the original signal.

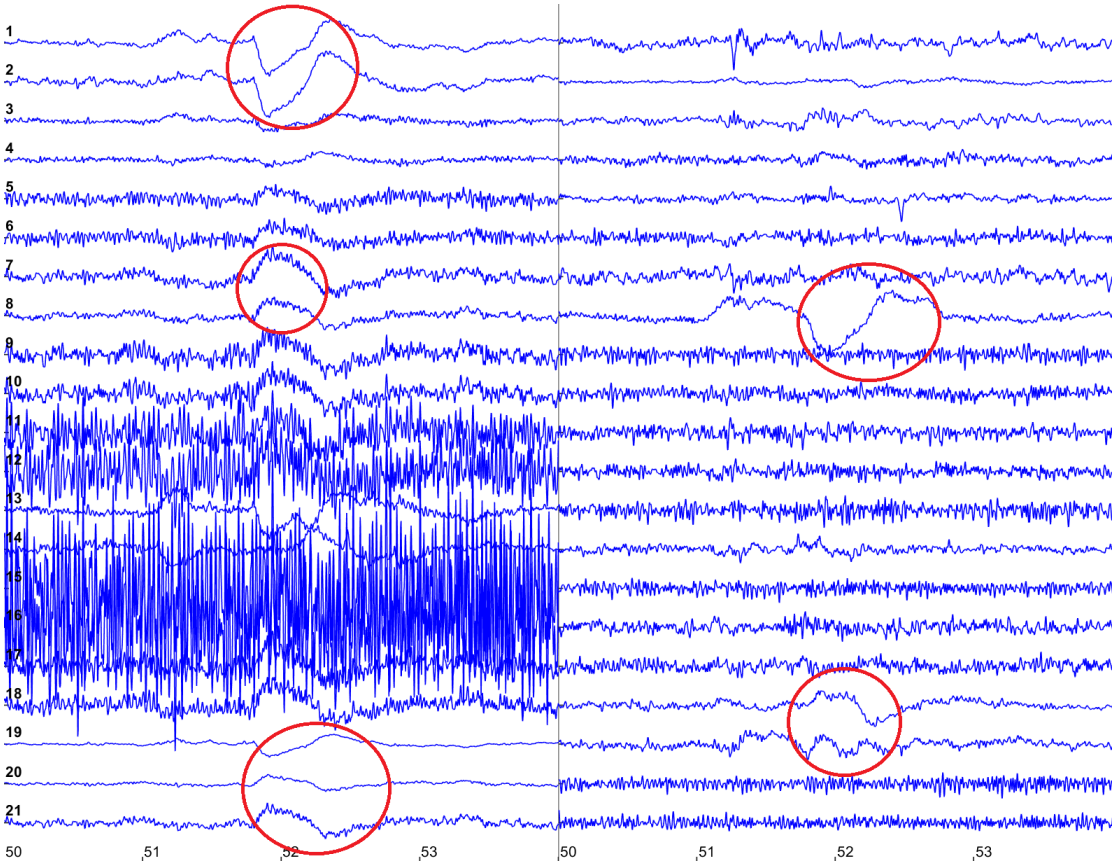

Figure 4.4: EEG data segment and corresponding components

Since this process of analyzing components visually is labour intensive, we propose to utilize Hurst analysis to automate the process. The decomposed signals are analyzed for their usefulness using the Hurst exponent. It was observed that independent components with Hurst exponents in the range of 0.58-0.66 contain more artifact data as compared to useful EEG data.

Component Number	Hurst Exponent
1	0.5721
2	0.7422
3	0.6068
4	0.5758
5	0.6620
6	0.5567
7	0.5730
8	0.6580
9	0.5315
10	0.5498
11	0.5419
12	0.4933
13	0.4845
14	0.6125
15	0.4805
16	0.5349
17	0.5702
18	0.6387
19	0.6111
20	0.4525
21	0.4461

Table 4.1: Tabulation of Independent components and their Hurst exponents

From the tabulation the components 3, 8, 14, 18, 19 have Hurst exponents in the range between 0.58-0.66. These components are classified as artifacts and their corresponding columns in the mixing matrix are rejected. The clean EEG dataset can be obtained by multiplying the pruned ICA and pruned mixing matrix.

In the following diagram we can observe that the EOG pulses have been successfully removed from the raw data. Due to the presence of high amplitude EOG pulses in the data the energy of the channel increases at those instances of time. We perform time-energy analysis to verify that the artifacts have been successfully removed. The random peaks in the time-energy curve is found to have reduced after the artifact removal process.

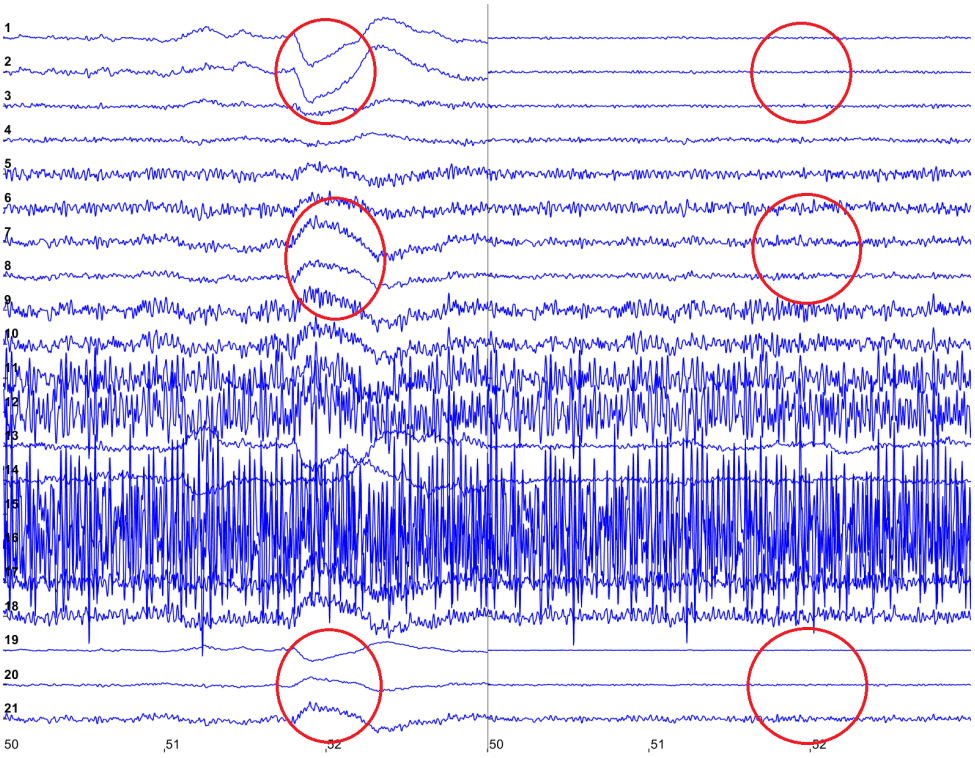

Figure 4.5: EEG data before (left) and after(right) artifact removal

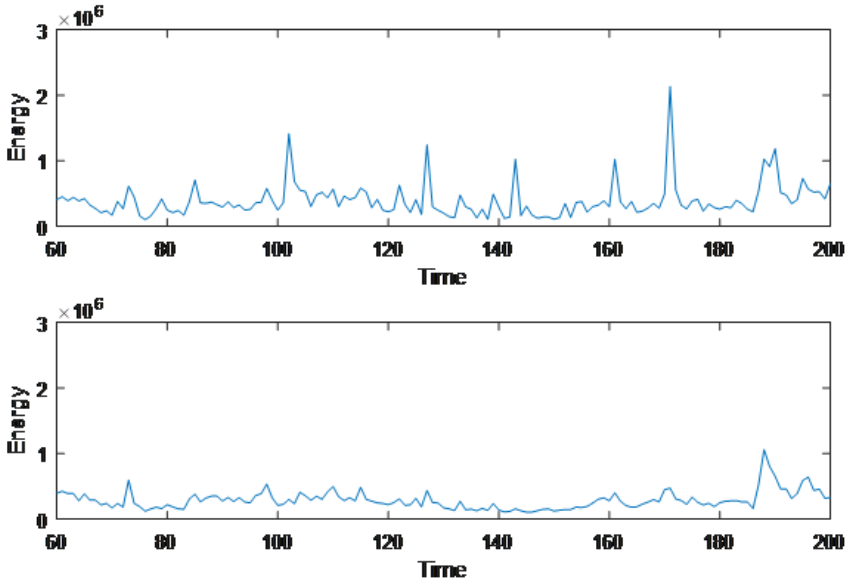

Figure 4.6: Time-energy plot before (top) and after (bottom) artifact removal

The above methodology was repeated for various data sets, each of which are known to have artifacts in them. Complete removal of the artifact signals were seen in all cases and the data obtained was visually better than the original.

Chapter 5

Conclusion and Future work

5.1 CONCLUSION

An automatic system that decomposes the given EEG signals with Independent component analysis, identifies components that have potential artifacts from the eye and muscles with the help of Hurst exponent and removes them from the component set is designed. The system also recombines the components to give an artifact-less signal. This work is an advancement of existing implementations of ICA for filtering of EEG data. The methodology adopted in this work is simple and effective for the analysis of the signal. The system was tested for its working and efficiency using different datasets with known cases of artifacts and seizure data to good results.

5.2 SCOPE FOR FUTURE WORK

The system is capable of analyzing pre-existing EEG data at present. Future inclusions to the system may see to the online analysis of the EEG data with improved accuracy and precision. Online artifact removal in EEG could be very useful in the case of Brain Computer Interface (BCI) where real-time brain signals are utilized to control computer, drones etc.

- Wavelet ICA can be implemented in place of Cumulant ICA, which could make the processing pseudo-realtime in nature.
- Topography plots can also be simultaneously analyzed using image processing to reinforce the results obtained from the Hurst exponent evaluation.

Bibliography

[1] Tzyy-Ping Jung, Scott Makeig, Colin Humphries, Te-Won Lee, Martin J. McKeown, Vicente Iragui, and Terrence J. Sejnowski. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37(2):163–178, 2000.

[2] J. Leon Kenemans, Peter C.M. Molenaar, Marinus N. Verbaten, and Jef L. Slangen. Removal of the ocular artifact from the eeg: A comparison of time and frequency domain methods with simulated and real data. *Psychophysiology*, 28(1):114–121, 1991.

[3] Theodore C. Weerts and Peter J. Lang. The effects of eye fixation and stimulus and response location on the contingent negative variation (CNV). *Biological Psychology*, 1(1):1–19, jan 1973.

[4] D A Overton and C Shagass. Distribution of eye movement and eyeblink potentials over the scalp. *Electroencephalography and clinical neurophysiology*, 27(5):546, Nov 1969.

[5] Joel L Whitton, Frank Lue, and Harvey Moldofsky. A spectral method for removing eye movement artifacts from the EEG. *Electroencephalography and Clinical Neurophysiology*, 44(6):735–741, jun 1978.

[6] J.C. Woestenburg, M.N. Verbaten, and J.L. Slangen. The removal of the eye-movement artifact from the EEG by regression analysis in the frequency domain. *Biological Psychology*, 16(1-2):127–147, feb 1983.

[7] Patrick Berg and Michael Scherg. A multiple source approach to the correction of eye artifacts. *Electroencephalography and Clinical Neurophysiology*, 90(3):229–241, mar 1994.

[8] Terrence D. Lagerlund, Frank W. Sharbrough, and Neil E. Busacker. Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition. *Journal of Clinical Neurophysiology*, 14(1):73–82, jan 1997.

[9] Jörn Anemüller, Terrence J Sejnowski, and Scott Makeig. Complex independent component analysis of frequency-domain electroencephalographic data. *Neural networks : the official journal of the International Neural Network Society*, 16(9):1311–23, Nov 2003.

[10] Anthony J. Bell and Terrence J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.*, 7(6):1129–1159, November 1995.

[11] Pierre Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, apr 1994.

[12] S P. Fitzgibbon, D M. W. Powers, K J. Pope, and C R. Clark. Removal of EEG noise and artifact using blind source separation. *Journal of Clinical Neurophysiology*, 24(3):232–243, jun 2007.

[13] T. Radüntz, J. Scouten, O. Hochmuth, and B. Meffert. EEG artifact elimination by extraction of ICA-component features using image processing algorithms. *Journal of Neuroscience Methods*, 243:84–93, mar 2015.

[14] T. Blaschke and L. Wiskott. CuBICA: Independent component analysis by simultaneous third- and fourth-order cumulant diagonalization. *IEEE Transactions on Signal Processing*, 52(5):1250–1256, may 2004.