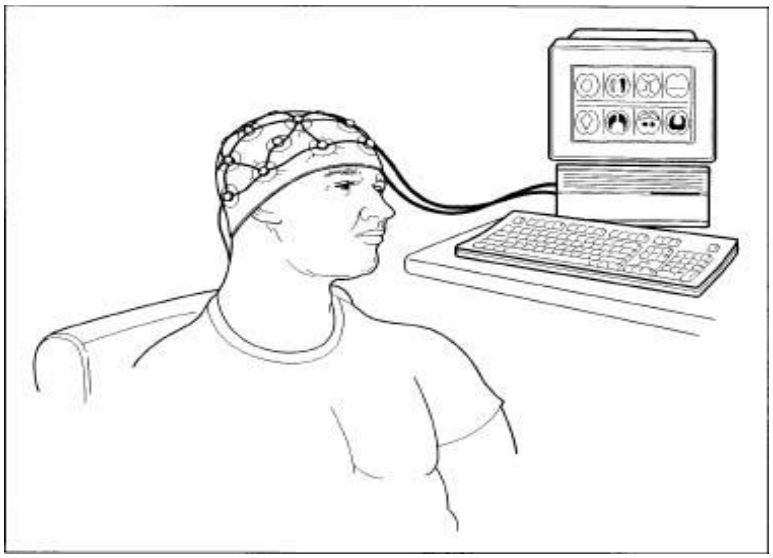
CHAPTER 1

INTRODUCTION

An Electroencephalogram (EEG) is a record of the electric signal generated by the cooperative action of brain cells, or more precisely, the time course of extracellular field potentials generated by their synchronous action. It is a test used to evaluate the electrical activity in the brain. Brain cells communicate with each other through electrical impulses. Electroencephalogram derives from the Greek words enkephalo (brain) and graphein (to write). EEG can be measured by means of electrodes placed on the scalp or directly on the cortex.



An EEG can be used to help detect potential problems associated with this activity. At 17^{th} -23^{rd} month of prenatal development neural activities get started in human brain and body's behavior can be analyzed by using EEG signals so this leads to the study of EEG signals.The brain is much complicated in structure, and contains abundant Information related to the human spirit and biological structures. Therefore, many researchers from kind of fields unceasingly extract and analyze the implicit information of EEG by all sorts of signal processing techniques. Electroencephalogram is the recorded electric potential from the exposed surface of the brain or from the

outer surface of the head. EEG recorded in the absence of an external stimulus is called spontaneous EEG.EEG generated as a response to external or internal stimulus is called an event-related potential (ERP).

1.1:HISTORY AND BACKGROUND OF EEG

Richard Caton (1842–1926) is regarded as the first scientist to investigate brain potentials. He worked on the exposed brains of cats and rabbits, measuring electric currents by means of a galvanometer, where a beam of light reflected from its mirror was projected onto a scale placed on a nearby wall. The results (presented in 1875) showed that ‘‘feeble currents of varying directions pass through the multiplier when the electrodes are placed at two points of the external surface, or one electrode on the gray matter and one on the surface of skull.’’ This observation can be regarded as a discovery of electroencephalographic activity.

Adolf Beck (1863–1939) also investigated spontaneous activity of the brains of rabbits and dogs. He was the first to discover (in 1890) the rhythmical oscillations of brain electrical activity. He also observed the disappearance of these oscillations when the eyes were stimulated with light, which was the first discovery of so-called ‘‘alpha blocking.’’ Later, his co-worker Napoleon Cybulski (1854– 1919) presented the electroencephalogram in a graphical form by applying a galvanometer with a photographic attachment and was the first to observe epileptic EEG activity in a dog elicited by an electric stimulation . In 1929, the first electroencephalogram was recorded from the surface of the human scalp by Hans Berger .In 1935 witnessed birth of the major fields of today’s clinical electroencephalography. F. Gibbs and H. Davis showed association of 3/sec spike-wave complexes in EEG with

epileptic absences and A. L. Loomis et al. studied human sleep patterns. Also in 1935, the first electroencephalograph (Grass Model I) started the era of contemporary EEG recording.

1.2:NEUROPHYSIOLOGICAL BASIS OF EEG

In the brain, two main classes of cells exist: nervous cells, called neurons, and glial cells as shown in figure. In both, the resting potential is approximately -80mV, with the inside cells being negative.

The difference of potentials across a cell membrane comes from the difference of concentration of cations : K^+, Na+, anions Cl-, and large organic anions. Ca++ions are less abundant, but they have an important regulatory role. The potential difference is maintained by the active

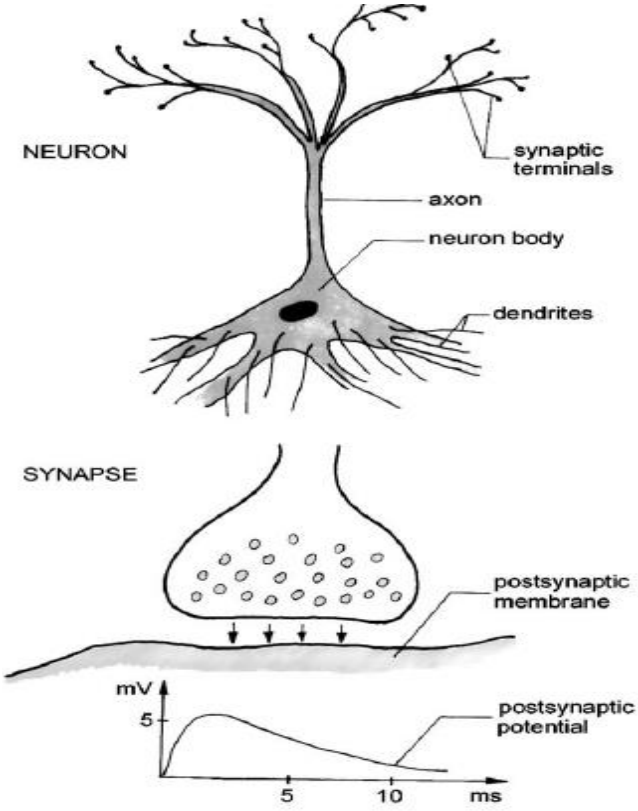


Figure 1.2 :neuron and glial cells

transport of cations K+ to the inside of the cell and Na+ to the outside, using the energy supplied through metabolic processes.

Electric activity of neurons is manifested by generation of action potentials and postsynaptic potentials (PSP).Action potentials occur when the electrical

excitation of the membrane exceeds a threshold. Postsynaptic potentials are subthreshold phenomena.

The generation of action potentials is connected with rapid increase of permeability for Na+ ions. Their influx in the cell causes a rapid increase of the potential inside the cell and the change of polarity of the inside of the neuron from negative to positive (about +30 mV). A subsequent increase of membrane permeability to K+ ions (leading to their outflow from the cell), and a decrease of permeability for Na+ ions makes the inside of the cell negative again with respect to the surrounding medium. In this way, action potential of characteristic spike-like shape (duration about 1 ms) is created. It obeys the ‘‘all or nothing’’ rule: for supra-threshold stimuli, a pulse of a constant amplitude is generated; for sub-threshold excitation, the neuron doesn’t fire .

PSPs are connected with the phenomena occurring on the postsynaptic membrane. When action potential arrives at the synapse, it secretes a chemical substance, called mediator or transmitter, which causes a change in the permeability of the postsynaptic membrane of the next neuron. As a result, ions traverse the membrane and a difference in potentials across the membrane is created.

When the negativity inside the neuron is decreased (e.g., by the influx of Na+ ions), the possibility of firing is higher and an excitatory postsynaptic potential (EPSP) is generated. An inhibitory postsynaptic potential (IPSP) is created when the negativity inside the neuron is increased and the neuron becomes hyperpolarized.

Unlike the action potential, the PSPs are graded potentials, their amplitudes are proportional to the amount of secreted mediator, which depends on the excitation of the input neuron. Postsynaptic potentials typically have

amplitudes of 5–10mV and a time span 10–50 msec. In order to obtain supra-threshold excitation, the amplitudes of many postsynaptic potentials have to be superimposed in the soma of a neuron.

A neuron can have very abundant arborizations , making up to 10,000 synaptic junctions with other neurons (in the human brain, about 1011 neurons exist).The electrical activity of neurons generates currents along the cell membrane in the intra- and extracellular spaces, producing an electric field conforming approximately to that of a dipole. Macroscopic observation of this electric field requires the synchronization of electrical activity of a large number of dipoles oriented in parallel.

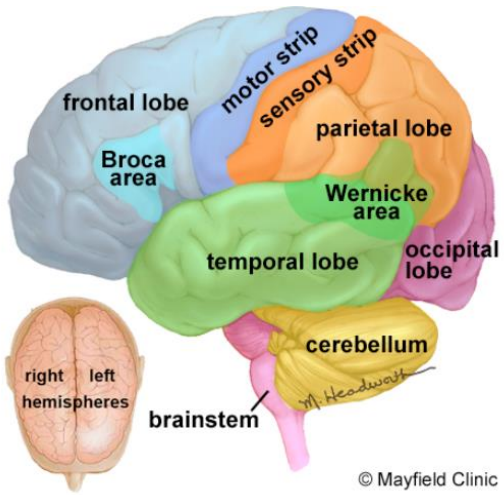
Indeed, pyramidal cells of the cortex are, to a large degree, parallel and, moreover, they are synchronized by virtue of common feeding by thalamo cortical connections.The condition of synchrony is fulfilled by the PSPs, which are relatively long in duration. The contribution from action potentials to the electric field measured extracranially is negligible.

EEG comes from the summation of synchronously generated postsynaptic potentials. The contribution to the electric field of neurons acting synchronously is approximately proportional to their number, and, for those firing nonsynchronously, as a square root of their number. For example, if the electrode records action of 108neurons (which is typical for scalp electrode) and 1% of them are acting synchronously, their contribution will be 100 times.

1.3:HUMAN BRAIN

The human brain is an amazing three-pound organ that controls all functions of the body, interprets information from the outside world, and embodies the essence of the mind and soul. Intelligence, creativity, emotion, and memory are a few of the many things governed by the brain. Protected within the skull, the brain is composed of the cerebrum, cerebellum, and brainstem. The brainstem acts as a relay center connecting the cerebrum and cerebellum to the spinal cord. The brain receives information through our five senses:
sight, smell, touch, taste, and hearing - often many at one time. It assembles the messages in a way that has meaning for us, and can store that information in our memory. The brain controls our thoughts, memory and speech, movement of the arms and legs, and the function of many organs within our body. It also determines how we respond to stressful situations (such as taking a test, losing a job, or suffering an illness) by regulating our heart and breathing rate.

The brain is composed of the cerebrum, cerebellum, and brainstem (Fig. 3).



• The **cerebrum** is the largest part of the brain and is composed of right and left hemispheres. It performs higher functions like interpreting touch, vision and hearing, as well as speech, reasoning, emotions, learning, and fine control of movement.

• The **cerebellum** is located under the cerebrum. Its function is to coordinate muscle movements,maintain posture, and balance.

Fig 1.3:Human Brain

• The **brainstem** includes the midbrain, pons, and medulla. It acts as a relay center connecting the cerebrum and cerebellum to the spinal cord. It performs many automatic functions such as breathing, heart rate, body temperature, wake and sleep cycles, digestion, sneezing, coughing, vomiting, and swallowing. Ten of the twelve cranial nerves originate in the brainstem. The surface of the cerebrum has a folded appearance called the cortex. The cortex contains about 70% of the 100 billion nerve cells. The nerve cell bodies color the cortex grey-brown giving it its name – gray matter (Fig.4). Beneath the cortex are long connecting fibers between neurons, called axons, which make up the white matter.The folding of the cortex increases the brain's surface area allowing more neurons to fit inside the skull and enabling higher functions. Each fold is called a gyrus, and each groove between folds is called a sulcus. There are names for the folds and grooves that help define specific brain regions.

Frontal lobe
• Personality, behavior, emotions
• Judgment, planning, problem solving
• Speech: speaking and writing (Broca's area)
• Body movement (motor strip)
• Intelligence, concentration, self awareness

Parietal lobe
• Interprets language, words
• Sense of touch, pain, temperature (sensory strip)
• Interprets signals from vision, hearing, motor, sensory and memory
• Spatial and visual perception

Occipital lobe
• Interprets vision (color, light, movement)

Temporal lobe
• Understanding language (Wernicke's area)
• Memory
• Hearing
• Sequencing and organization

The right and left hemispheres of the brain are joined by a bundle of fibers called the corpus callosum that delivers messages from one side to the other. Each hemisphere controls the opposite side of the body. If a brain tumor is located on the right side of the brain, your left arm or leg may be weak or paralyzed. Not all functions of the hemispheres are shared. In general, the left hemisphere controls speech, comprehension, arithmetic, and writing. The right hemisphere controls creativity, spatial ability, artistic,and musical skills. The left hemisphere is dominant in hand use and language in about 92% of people.

1.4: BRAIN WAVES CLASSIFICATION

EEG brain waves have been classified into four basic group delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz),beta (13–30 Hz), and gamma (above 30 Hz). Gamma components are difficult to record by scalp electrodes and their frequency does not exceed 45 Hz; in ECoG components, up to 100 Hz, or even higher, may be registered. The contribution of different rhythms to the EEG depends on the age and behavioral state of the subject, mainly the level of alertness.

Considerable inter subject differences in EEG characteristics also exist. EEG pattern is influenced by neuro-pathological conditions, metabolic disorders, and drug action. Their description are as follow

a) **Delta rhythm** is a predominant feature in EEGs recorded during deep sleep. In this stage, delta waves usually have large amplitudes (75–200 mV) and show strong coherence all over the scalp.

b) **Theta rhythm**s rarely occur in adult humans. However, they are predominant in rodents; in this case, the frequency range is broader (4–12 Hz) and waves have a high amplitude and characteristic sawtooth shape. It is hypothesized that theta rhythms in rodents serve as a gating mechanism in the information transfer between the brain structures. In

humans, activity in the theta band may occur in emotional or some cognitive states; it can be also connected with the slowing of alpha rhythms caused by pathology.

c) Alpha rhythms are predominant during wakefulness and are most pronounced in the posterior regions of the head. They are best observed when the eyes are closed and the subject is in a relaxed state. They are blocked or attenuated by attention (especially visual) and by mental effort. Mu rhythms have a frequency band similar to alpha, but their topography and physiological significance are different. They are related to the function of motor cortex and are prevalent in the central part of the head. Mu rhythms are blocked by motor functions.

d) **Beta rhythm** is characteristic for the states of increased alertness and focused attention, as was shown in several animal and human studies.

e) **Gamma rhythm** is connected with information processing(e.g., recognition of sensory stimuli) and the onset of voluntary movements. In general, it can be summarized that the slowest cortical rhythms are related to an idle brain and the fastest to information processing.

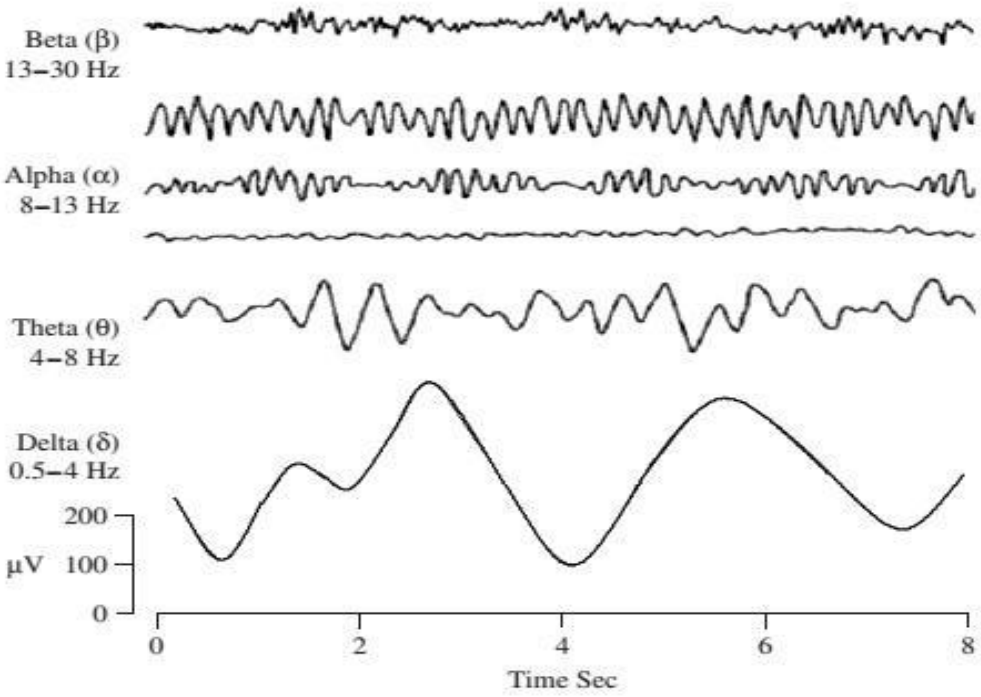


Figure 1.4 found in EEG signal

BRAINWAVE TYPE	FREQUENCY RANGE	MENTAL CONDITION
DELTA	0.1Hz TO 3 Hz	Deep ,dreamless sleep ,non-REM sleep ,unconscious
THETA	4 Hz TO 7 Hz	Intuitive ,creative ,recall ,fantasy ,imaginary ,dream
ALPHA	8 Hz TO 12Hz	Relaxed ,but not drowsy ,tranquil ,conscious
LOW BETA	12 Hz TO 15Hz	Formerly SMR ,relaxed yet focused ,integrated
MIDRANGE BETA	16 Hz TO 20Hz	Thinking ,aware of self & surroundings
HIGH BETA	21 Hz TO 30Hz	Alertness ,agitation
GAMMA	30 Hz TO 100Hz	Motor functions ,higher mental activity

Table 1.1: Different rythms of EEG signal

1.5:ELECTRODE POSITIONING

The International Federation of Societies for Electroencephalography and Clinical Neurophysiology has recommended the conventional electrode setting (also called 10–20) for 21 electrodes (excluding the earlobe electrodes), as depicted in Figure 1.6 [17]. Often the earlobe electrode A1 connected to the left earlobe and the earlobe electrode A2 connected to the right electrode are used as the reference electrodes. The 10–20 system avoids both eyeball placement and considers some constant distances by using specific anatomic landmarks from which the measurement would be made and then uses 10 or 20 % of that specified distance as the electrode interval. The odd and even electrodes are on the left and right respectively. The rest of the electrodes are placed in between the above electrodes equidistance between them, for setting a larger number of electrodes using the above conventional system. For example, C1is placed between C3and Cz.

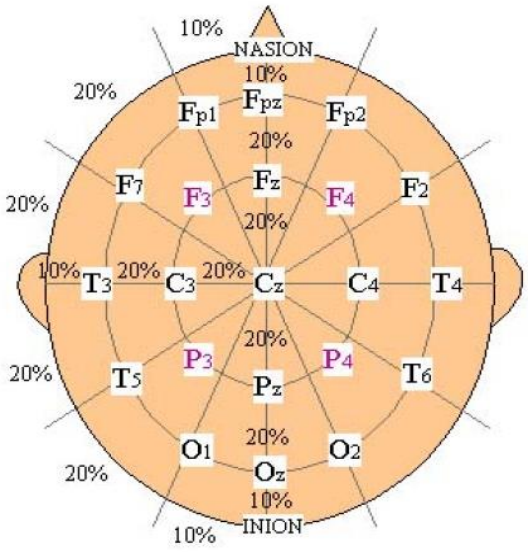


Figure: 1.5 Conventional 10–20 EEG electrode positions

For the measurement of EOG, ECG, and EMG of the eyelid and eye surrounding muscles, extra electrodes are sometimes used. A single channel may be used, in some applications such as ERP analysis and brain computer interfacing. The position of the corresponding electrode has to be well determined.

1.6:EEG SIGNAL CONDITIONING

The raw EEG signals have amplitudes of the order of µvolts and contain frequency components of up to 300 Hz. The signals have to be amplified before the ADC and filtered, either before or after the ADC in order to reduce the noise and make the signals suitable for processing and visualization to retain the effective information. The filters are designed in such a way so that it does not introduce any change or distortion to the signals. To remove the disturbing very low frequency components such as those of breathing, high pass filters with a cut-off frequency of usually less than 0.5 Hz are used. By using low pass filters with a cut-off, frequency of approximately 50–70 Hz, high-frequency noise is mitigated.

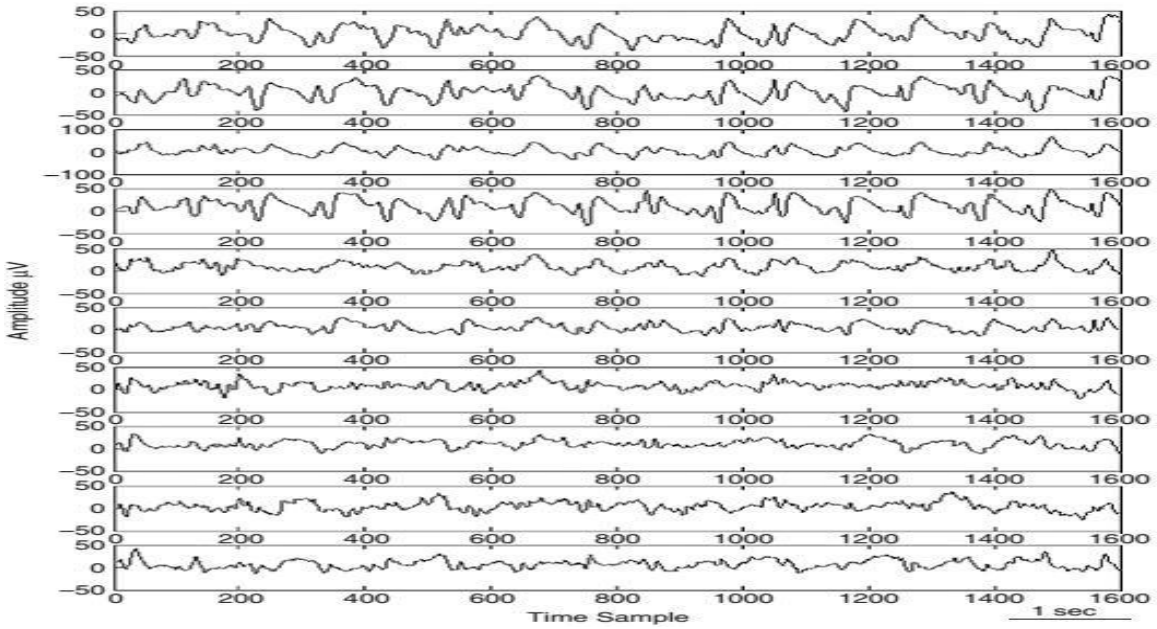


Figure 1.6:EEG recording of a normal adult

Notch filters with a null frequency of 50 Hz are often necessary to ensure perfect rejection of the strong 50 Hz power supply. In this case the sampling frequency can be as low as twice the bandwidth commonly used by most EEG systems. The commonly used sampling frequencies for

EEG recordings are 100, 250, 500, 1000, and 2000 samples/s. The main artefacts can be divided into patient-related (physiological) and system artefacts. The patient-related or internal artefacts are body movement-related, EMG, ECG (and pulsation), EOG, ballistocardiogram, and sweating. The system artefacts are 50/60 Hz power supply interference, impedance fluctuation, cable defects, electrical noise from the electronic components, and unbalanced impedances of the electrodes. Often in the preprocessing stage these artifacts are mitigated and the informative information is restored.

1.7:ARTIFACTS

Contamination of EEG data can occur at many points during the recording process. Most of the artifacts considered here are biologically generated by sources external to the brain. Improving technology can decrease externally generated artifacts, such as line noise, but biological artifact signals must be removed after the recoding process. Figure 1.2 shows waveforms of some of the most common EEG artifacts, each of which is discussed below.

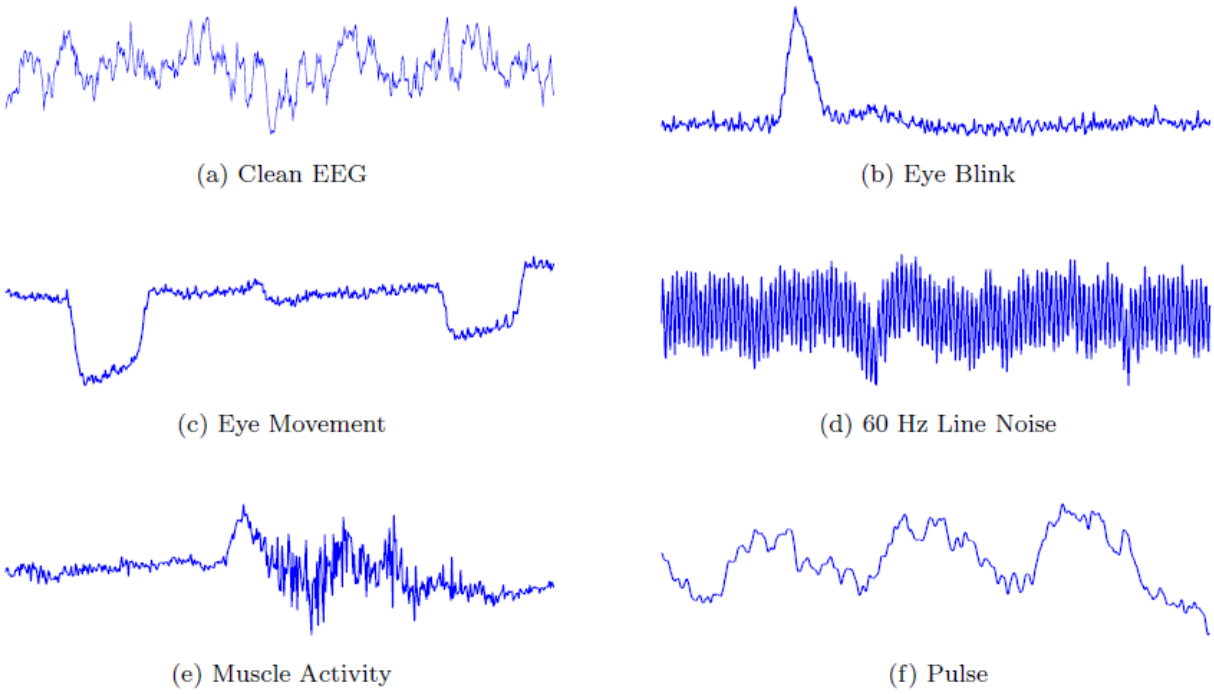


Figure 1.7: Artifacts waveform

Eye Blink: The eye blink artifact is very common in EEG data. It produces a high amplitude signal that can be many times greater than the EEG signals of interest (see Figure 1.2(b)). Because of its high amplitude an eye blink can corrupt data on all electrodes, even those at the back of the head (See Figure 1.1). Eye artifacts are often measured more directly in the electrooculargram (EOG), pairs of electrodes placed above and around the eyes. Unfortunately, these measurements 3 are contaminated with EEG signals of interest and so simple subtraction is not a removal option even if an exact model of EOG diffusion across the scalp is available [31].

Eye Movement: Eye movement artifacts (see Figure 1.2(c)) are caused by the reorientation ofthe retinocorneal dipole [31, 43]. This artifact's diffusion across the scalp is stronger than that of the eye blink artifact. Eye blinks and movements often occur at close intervals producing an effect shown in Figure 1.3

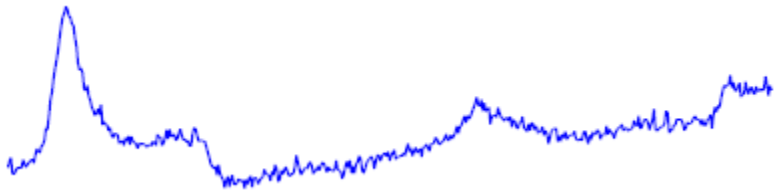


Figure 1.8: Overlap of eye blink and eye movement artifacts.

Line Noise: Strong signals from A/C power supplies (see Figure 1.2(d)) can corrupt EEG data as it is transferred from the scalp electrodes to the recording device. This artifact is often filltered by notch filters, but for lower frequency line noise and harmonics this is often undesirable. If the line noise or harmonics occur in frequency bands of interest they interfere with EEG that occurs in the same band [31]. Notch filtering at these frequencies can remove useful information. Line noise can corrupt the data from some or all of the electrodes depending on the source of the problem.

Muscle Activity: Muscle activity (see Figure 1.2(e)) can be caused by activity in different muscle groups including neck and facial muscles. These signals have a wide frequency range and can be distributed across different sets of electrodes depending on the location of the source muscles.Pulse. The pulse, or heart beat, artifact (see Figure 1.2(f)) occurs when an electrode is placed on or near a blood vessel. The expansion and contraction of the vessel introduce voltage changes into the recordings. The artifact signal has a frequency near 1.2Hz, but can vary with the state of the patient. This artifact can appear as a sharp spike or smooth wave [7].

Respiration Artifact: Two kind of artifacts are produced by respiration. The first kind is in the form of slow and rhythmic activity, that ae synchronous with the body movements of respiration and affects the impedance of one electrode. The other kind can be slow or sharp that occur with inhalation or exhalation

synchronously and involves electrodes on which the patient is lies. To monitor respiration, several commercially available devices can be coupled to the EEG machine. One channel can be dedicated to respiratory movements as with the ECG. To monitor respiration by the EEG technician making notations with a pencil, is the simplest way.

Out of all above artifacts the most obvious and influencing artifact is ocular artifact and it is challenging to remove it because its frequency band overlaps with EEG signal. There are several methods to handle with artifacts, some of them we have discussed in chapter 2 and proposed a method in chapter 3. Chapter 4 will talk about results and in chapter 5 we have given our conclusion

CHAPTER 2

LITERATURE REVIEW

EEG signals must be accurate or free from artifacts for properly diagnosis of the person otherwise the subject is assumed to have malady. As we have discussed in introduction artifacts are obvious in EEG signal, to deal with artifacts very much of research has been done and still going on.

We have gone through various research papers and find various techniques to remove artifacts from the EEG signals.Some methods are as follows:

1) Artifacts removal using Singular Value Decomposition(SVD)
2) Artifacts removal using Independent Component Analysis(ICA)
3) Artifacts removal using Adaptive filtering(AF)
4) Artifacts removal using Artificial Neural Network(ANN)

2.1:Artifacts Removal using SVD

We present here a method to recover EEG signal based on the decomposition of the data space into orthogonal subspaces through singular value decomposition (SVD) . Because of the energy-preserving orthogonal transformation in the SVD, these subspaces correspond to the signal and noise components contained in the data .

To filter the noise and other undesired signal components, we project the data onto the desired subspace by simply setting all the corresponding singular values in the SVD spectrum of the data to zero. The EEG signal estimate is subsequently recovered from its projection.

Algorithm:

- singular value decomposition (SVD) decomposes EEG data space into orthogonal subspaces . These subspaces correspond to the signal and noise components contained in the data.
- Let A be the noisy EEG subspace ,U,V are singular vectors and w is diagonal matrix.

$$\begin{pmatrix} \\ \mathbf{A} \\ \\ \end{pmatrix} = \begin{pmatrix} \\ \mathbf{U} \\ \\ \end{pmatrix} \begin{pmatrix} w_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & w_n \end{pmatrix} \begin{pmatrix} \\ \mathbf{V} \\ \\ \end{pmatrix}^T$$

- To filter out the noise , all the corresponding singular values in the SVD spectrum of the data are set to zero because a small amount of noise in the data changes the zero singular values to small positive nonzero numbers

2.2:Artifact Removal using ICA

In the method suggested for artifact removal is blind source separation using independent component analysis and then zeroing the artifact related components. The basic concept of ICA is in fact that signals can be separated into their actual independent sources. This concept is applicable at places where mixed components can be assumed to be independent of each other and is very useful in denoising of signals.

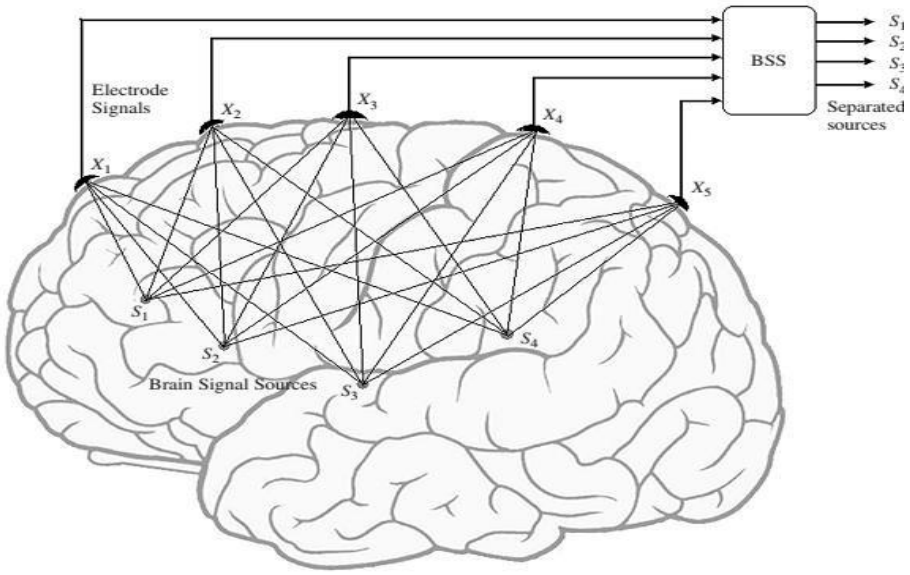


Figure 2.1: BSS concept ,mixing and blind separation of EEG signal

Mathematically we can express the linear ICA model as,

$$X = A \times S$$

Where X is mixed signal recorded through different channels, A is unknown mixing matrix .ICA estimates demixing matrix W such that the estimated independent sources U are very close to actual sources S.

$$u = W \times X \approx S \qquad W = A^{-1}$$

Once the independent components are calculated artifactual sources are identified using some IC marker (some parameter which discriminates artifactual and EEG sources) and the column corresponding to those sources are made zero in estimated mixing matrix. The sources are mixed again by multiplying modified estimated mixing matrix with estimated sources.

Algorithm :

1. Estimate independent components and mixing matrix using ICA.
2. Identify the artifactual sources using some IC marker.
3. Zero the column corresponding to artifactual independent components.
4. Mix the estimated sources by multiplying them with modified estimated mixing matrix.

2.3 Artifacts removal using Adaptive Filtering:

This is the hybrid de-noising approach based on an adaptive model to separate OA and provide a basis for balancing the tracking performance and the computing speed. The hybrid de-noising approach contains two steps:

1. DWT is applied to divide raw EEG signals into different frequency signals (low-frequency components and high- frequency components).
2. An adaptive model is taken to predict the low frequency components with EOG signals and recover true EEG signals.

The key technique is the application of an *Adaptive Predictor Filter* (APF) based on an *Adaptive Autoregressive* (AAR) model to predict low-frequency EEG signals to enable the recovery of the true EEG signal.

Algorithm:

1. ***Signal Decomposition and OAs(Ocular artifacts) Zone Detection Using WT(Wavelet Transformation):*** In this model, DWT is used to decompose the EEG signals and detect OA zones. The frequency range of the EEG signal is 0 to 64 Hz, while OA occur within the range 0 to 16 Hz. Multi-scale DWT decomposition is used to extract the low frequency components and the non-

stationary time series is then decomposed into several approximate stationary time series

2) ***Signal Prediction:*** A variety of studies have been developed with respect to the forecasting of EEG time series. In this model AAR models and an APF are applied to improve prediction. The APF uses an adaptive filter to estimate the future values of a signal based on past values. As the signal is non-stationary, the AAR parameters are allowed to vary in time.There are several AAR estimation algorithms that can be used as predictors, such as the *Least-Mean-Squares* (LMS) approach,the *Recursive-Least-Squares* (RLS) approach.

3) ***Optimization and Selection of Parameters:*** The performance of the AAR model depends heavily upon a good selection of parameters (UC, the model order, ε_1, and ε_2 .The update coefficient UC determines the speed of adaptation, the time-resolution as well as the smoothing of the AR estimates. If UC is too large, the adaptive filter becomes unstable, and if it is too small, adaptation is too slow.

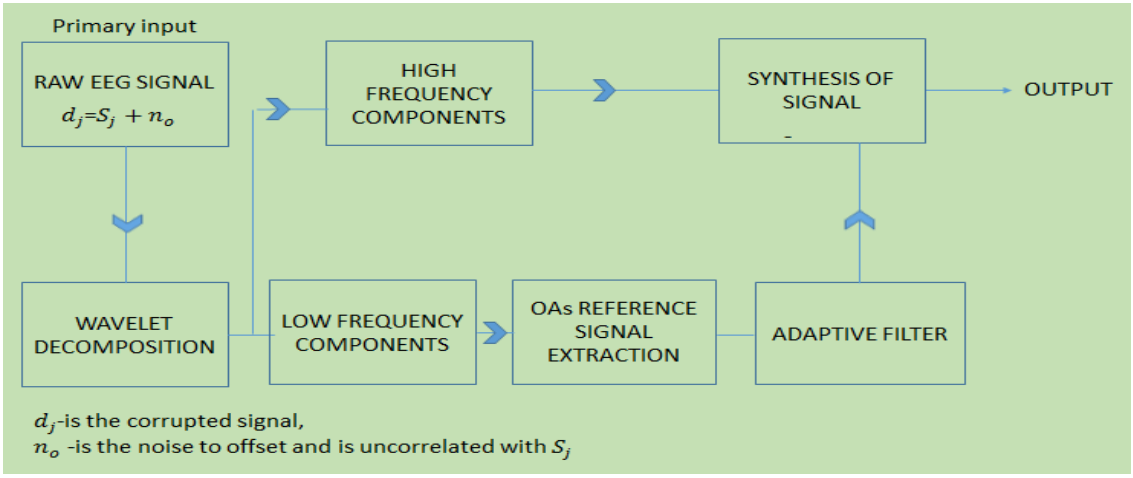


Figure 2.2: EEG denoising using Adaptive Filtering

2.3.1:DWT(Discrete Wavelet Transform):

The discrete wavelet transform (DWT) is a linear transformation that operates on a data vector whose length is an integer power of two, transforming it into a numerically different vector of the same length. It is a tool that separates data into different frequency components, and then studies each component with resolution matched to its scale. DWT is computed with a cascade of filterings followed by a factor 2 subsampling (Fig1).

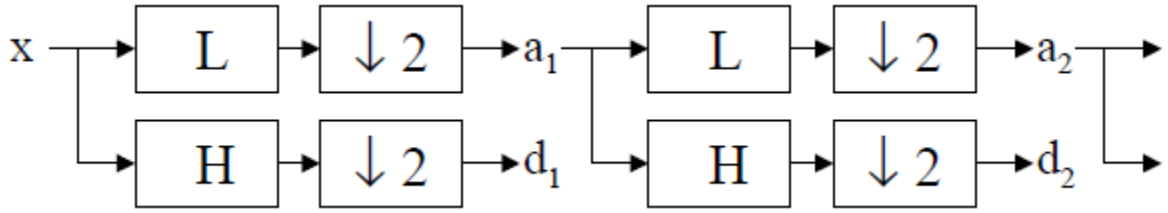


Figure 2.3: DWT Tree

H and *L* denotes high and low-pass filters respectively, ↓ 2 denotes subsampling. Outputs of this filters are given by equations (1) and (2)

Elements *aj* are used for next step (scale) of the transform and elements *dj*, called wavelet coefficients, determine output of the transform. *l[n]* and *h[n]* are coefficients of low and high-pas filters respectively One can assume that on scale *j+1* there is only half from number of *a* and *d* elements on scale *j*. This causes that DWT can be done until only two *aj* elements remain in the analyzed signal These elements are called scaling function coefficients.

DWT algorithm for two-dimensional pictures is similar. The DWT is performed firstly for all image rows and then for all columns (Fig.2).

The main feature of DWT is multiscale representation of function. By using the wavelets, given function can be analyzed at various levels of resolution. The DWT is also invertible and can be orthogonal

2.3.2:LMS(Least mean squares):

The least-mean-square (LMS) algorithm is based on the use of instantaneous estimates of the autocorrelation function $r_{dx}(k)$ and the cross-correlation function $r_x(j,k)$. These estimates are deduced directly from below equations:

$$r_{dx}(k) = E[dx_x] \qquad \text{k = 1, 2, ….. , p}$$

$$r_x(j,k) = E[x_j x_k] \qquad \text{j, k = 1,2,….., p}$$

Where x_k is sensory signal and d is desired response.the estimates are as follows:

$$\hat{r}_k(j,k;n) = x_j(n)x_k(n)$$

and

$$\hat{r_{dx}}(k;n) = x_k(n)d(n)$$

The use of a hat in $\hat{r_x}$ and $\hat{r_{dx}}$ is intended to signify that these quantities are "estimates." The definitions introduced in above equations have been generalized to include a nonstationary environment, in which case all the sensory signals and the desired response assume time-varying forms too. Thus, substituting $\hat{r}_k(j,k;n)$ and $\hat{r_{dx}}(k;n)$ in place of $r_x(j,k)$ and $r_{dx}(k)$ in below equation:

$$w_k(n+1) = w_k(n) + \eta[r_{dx}(k) - \sum_{j=1}^{M} w_j(n) r_x(j,k)]$$

Where k =1,2,….,p , w-weight , n = no. of iteration , η=learning parameter

we get

$$\hat{w}_k(n+1) = \hat{w}_k(n) + \eta\left[x(n)d(n) - \sum_{j=1}^{M} \hat{w}_j\,(n)x_j(n)x_k(n)\right]$$

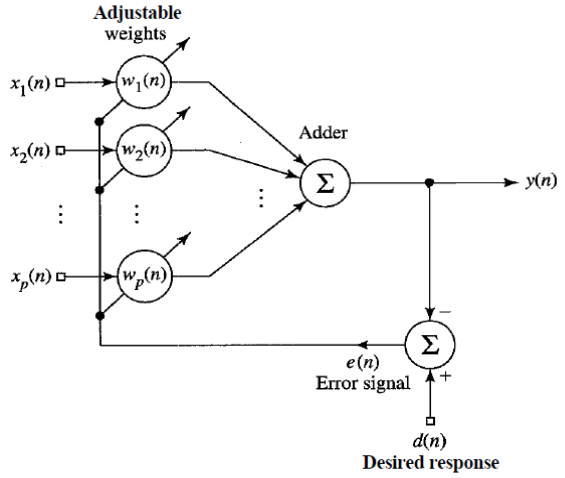
$$\hat{w}_k(n+1) = \hat{w}_k(n) + \eta\left[d(n) - \sum_{j=1}^{M} \hat{w}_j\,(n)x_j(n)\right]x_k(n)$$

$$\hat{w}_k(n+1) = \hat{w}_k(n) + \eta[d(n) - y(n)]x_k(n)$$

where y(n) is the output of the spatial filter computed at iteration n in accordance with the LMS algorithm

$$y(n) = \sum_{j=1}^{M} \hat{w}_j\,(n)x_j(n)$$

Figure 5.3 illustrates the operational environment of the LMS algorithm, which is completely described by Eqs. (5.22) and (5.23). A summary of the LMS algorithm is presented in Table 5.1, which clearly illustrates the simplicity of the algorithm. As indicated in this table, for the initialization of the algorithm, it is customary to set all the initial values of the weights of the filter equal to zero.



1. **Initialization.** Set

$$\hat{w}_k(1) = 0 \qquad \text{for } k = 1, 2, \ldots, p$$

2. **Filtering.** For time $n = 1, 2, \ldots$, compute

$$y(n) = \sum_{j=1}^{p} \hat{w}_j(n) x_j(n)$$

$$e(n) = d(n) - y(n)$$

$$\hat{w}_k(n+1) = w_k(n) + \eta e(n) x_k(n) \qquad \text{for } k = 1, 2, \ldots, p$$

FIGURE 2.4: Adaptive spatial filter

In the method of steepest descent applied to a "known" environment, the weight vector w(n), made up of the weights wl(n), w2(n), . . . , w,(n), starts at some initial value w(O), and then follows a precisely defined trajectory (along the error surface) that eventually terminates on the optimum solution w,, provided that the learning-rate parameter η is chosen properly. In contrast, in the LMS algorithm applied to an "unknown" environment, the weight vector w(n), representing an "estimate" of w(n), follows a random trajectory. For this reason, the LMS algorithm is sometimes referred to as a "stochastic gradient algorithm." As the number of iterations in the LMS algorithm approaches infinity, $\hat{w}$(n) performs a random walk (Brownian motion) about the optimum solution w.

Another way of stating the basic difference between the method of steepest descent and the LMS algorithm is in terms of the error calculations involved. At any iteration n, the method of steepest descent minimizes the mean-squared error J(n). This cost function involves ensemble averaging, the effect of which is to give the method of steepest descent an "exact" gradient vector that improves in pointing accuracy with increasing n.

The LMS algorithm, on the other hand, minimizes an instantaneous estimate of the cost function J(n). Consequently, the gradient vector in the LMS algorithm is "random," and its pointing accuracy improves "on the average" with increasing n. the LMS algorithm has the task of not only seeking the minimum point of the error surface but also trucking it. In this context, the smaller we make the learning-rate parameter η, the better will be the tracking behavior of the algorithm.

2.4:Artifacts removal using Neural Network:

In the given paper Maeda et al uses an Adaline (Adaptive linear Network) for adaptive noise reduction has been selected because of its simple structure and its weights part which improves the convergence rate .An ANN as a filter for EEG recordings is proposed in this paper, developing a novel framework for investigating and comparing the relative performance of an ANN incorporating real EEG recordings.

2.4.1:Artificial Neural Network

Artificial neural networks (ANNs) are effective and powerful tools for removing interference from EEGs. Several methods have been developed, but ANNs appear to be the most effective for reducing muscle and baseline contamination, especially when the contamination is greater in amplitude than the brain signal.

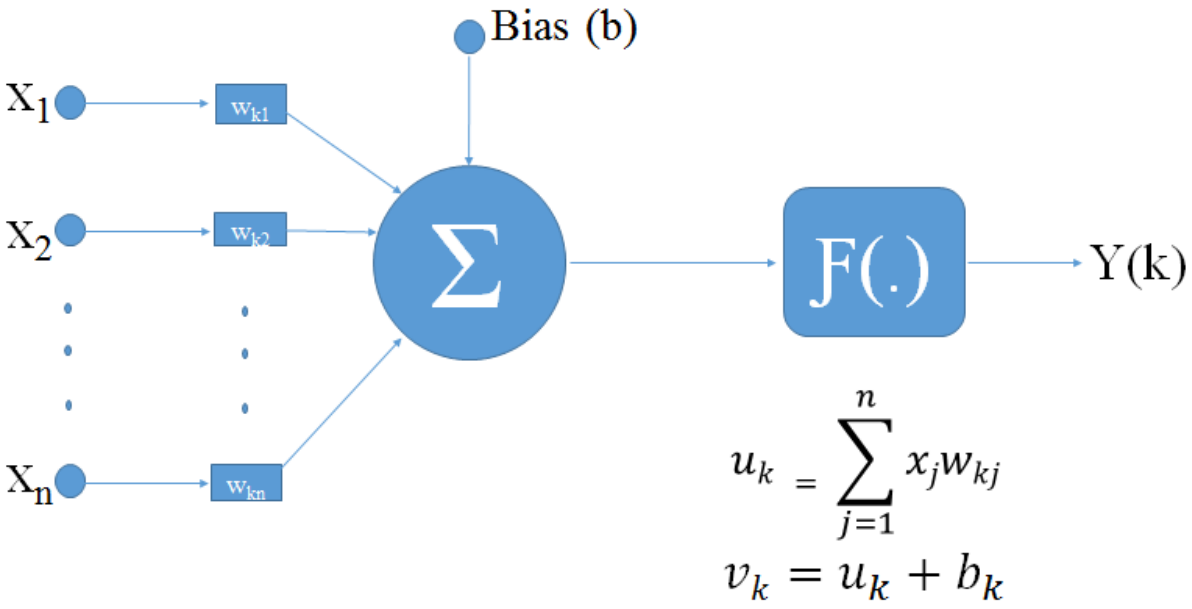


Figure 2.5 Neuron model

This method is based on a growing ANN that optimized the number of nodes in the hidden layer and the coefficient matrices, which are optimized by the simultaneous perturbation method. The ANN improves the results obtained with the conventional EEG filtering techniques: wavelet, singular value decomposition, principal component analysis, adaptive filtering and independent components analysis. The system has been evaluated within a wide range of EEG signals. The present study introduces a new method of reducing all EEG interference signals in one step with low EEG distortion and high noise reduction.

Algorithm:

1) Take noisy EEG signal samples as input vector to NN.
2) Set weights and bias of the network initially and activation function is taken
3) Output is taken and error is calculated from the desired output
4) Weights and biased are changed using learning process until we get minimum error.

Procedure for applying Artificial Neural Network on EEG:

1) Samples of the EEG signal must be taken as input vector.In Fig, $x = (x_i)$ is the input vector,$y = (y_i)$ is the output vector (clean signal),$w = (y_{ij})$ is the matrix of weights between the input layer and the hidden layer,$v = (v_{ij})$is the matrix of weights between the hidden layer and the output layer, and $d_i(k)$ represents the expected output of the pth set of the input vector.

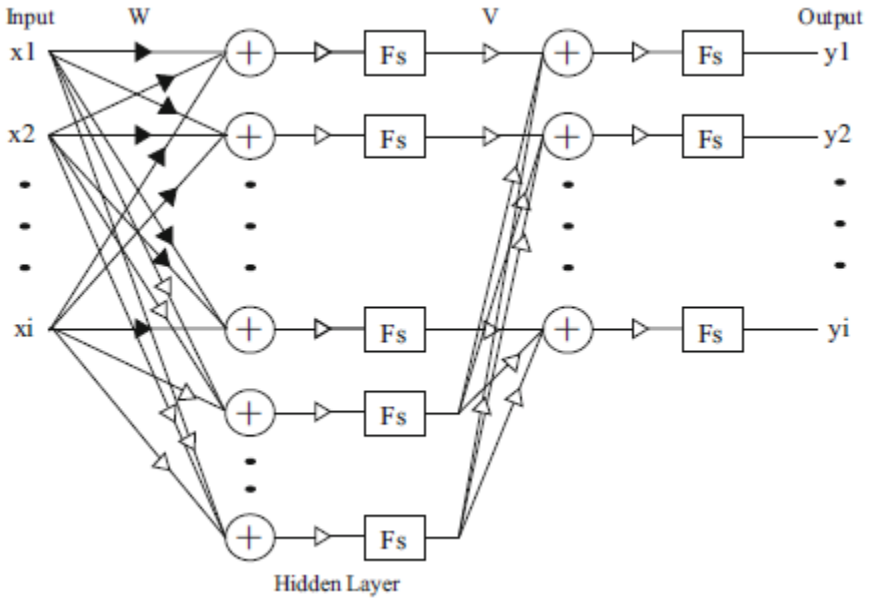


Figure 2.6 : ANN with a neuron in the hidden layer where X is the input vector, Y is the output vector and W and V are the weights.

2) Initially weights and bias are assigned to the network.
3) Using simultaneous perturbation learning process network is trained.
4) Error is calculated and weights and bias are changed until we get minimum error through number of iterations performed.

The proposed ANN was based on supervised learning, which required knowing the associated values in each input. These pairs of input/output are

$\{x_1, d_1\}, \ldots\ldots\ldots\ldots \{x_n, d_n\}, \ldots\ldots\ldots \{x_q, d_q\}$ n =1,2,3……q

where x_n is the network input and d_n is its corresponding expected output (target). When x input was presented to the proposed network, the network output was compared to d value, which was associated with it. The SP method was implemented by Spall, who reported different kinds of applications, including optimization problems, stochastic approach, adaptive control and ANNs.

2.4.2Mathematical Interpretation:

a) Inputs taken be $x_1(n), x_2(n), x_3(n), \ldots\ldots\ldots\ldots, x_p(n)$ and weights and bias are assigned to the model as $w_0(n), w_1(n), w_2(n), \ldots\ldots\ldots\ldots, w_p(n)$ where w_0 is the bias provided.

b) Output $y(n)$ is calculated after passing through the activation function (linear) .Here output is calculated as

$$y(n) = w_0(n) + \sum_{i=1}^{p} x_p(n) w_p(n)$$

c) Error is calculated

$$e(n) = d(n) - y(n)$$

d) Using this error values weights can be changed

$$w_k(n+1) = w_k(n) + \eta e(n) x_k(n)$$

e) Therefore, in Adaline network weights are adjusted until we get the minimum error.

CHAPTER 3

PROPOSED METHOD

We have discussed four methods for artifacts removal in previous chapter, Maeda et al proposed a method of Neural Network which gives good results in artifact suppression and uses simultaneous perturbation learning algorithm to find out the optimized neural network which gives better results with appropriate activation function. Here ,we have proposed different learning algorithms and different activation function and find out the best possible network to get minimum error .

3.1 Method:

A) Firstly , we have taken some EEG signals of subjects from site www.physionet.org .The signals comprises of 2560 samples and we have taken only 35 samples(taking (35:35:35) as the optimized network as proposed in paper).

B) Four different noises of various amplitude and frequencies are taken and added to the original EEG signal and out of the four three noisy signals are taken as input pattern to the neural network with corresponding target value.

C) Now, network is trained with different learning algorithms ,activation functions and for different network which we get by regularly updating the number of neurons in the hidden layer with an increment of 5 neurons per cycle.

D) Then, combinedly we simulate the particular network for fourth noisy signal which is taken as test signal and identify various parameters on the

basis of which we finally conclude which network is best with which learning algorithm and activation

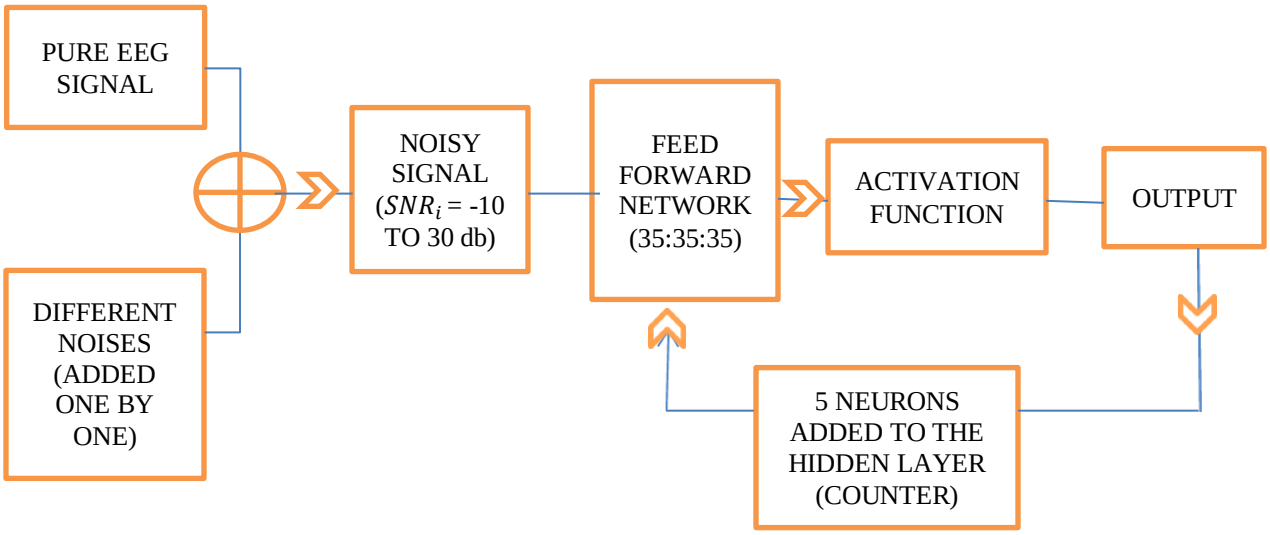


Figure 3.1:Denoising EEG signal using ANN

3.2 Mathematical Interpretation:

a) Here We have considered a feedforward network having 35 input neurons,35 hidden neurons and 35 output neuron .

b) Now,the network is trained with the training patterns $\{(x_1, t_1), (x_2, t_2), (x_3, t_3)\}$ and weights, bias are already assigned to the model where n represents no. of samples taken.

c) After getting the output (O_i) Error is calculated

$$E = \frac{1}{2}\sum_{i=1}^{N}(O_i - t_i)^2$$

d) After minimizing this function for the training set, new unknown input patterns are presented to the network and we expect it to interpolate. The backpropagation algorithm we used here is used to find a local minimum of the error function. The network is initialized with randomly chosen weights.

The gradient of the error function is computed and used to correct the initial weights. Mathematically as follows:

$$\nabla E = (\frac{\partial E}{\partial w_1}, \frac{\partial E}{\partial w_2}, \ldots\ldots \frac{\partial E}{\partial w_i})$$

e) Each weight is updated using the increment

$$\nabla w_i = -\gamma \frac{\partial E}{\partial w_i}$$

where γ represents a learning constant, i.e., a proportionality parameter which defines the step length of each iteration in the negative gradient direction.

3.3 Learning mechanism:

The steps followed during learning are :

a) Each input pattern in a training set is applied to the input units and then propagated forward.
b) The pattern of activation arriving at the output layer is compared with the correct (associated) output pattern to calculate an error signal.
c) The error signal for each such target output pattern is then back-propagated from the outputs to the inputs in order to appropriately adjust the weights in each layer of the network.
d) After a BackProp network has learned the correct classification for a set of inputs, it can be tested on a second set of inputs to see how well it classifies untrained patterns.

3.4 Learning algorithms:

Three different artificial neural network (ANN) training algorithms, Levenberg-Marquardt, conjugate gradient and resilient back-propagation, are used in the present study. This was done with a view to see which algorithm produces better results and has faster training for the application under consideration.The objective of training is to reduce the global error *E* defined as

$$E = \frac{1}{P}\sum_{P=1}^{P} E_p$$

where *P* is the total number of training patterns; and E_p is the error for training pattern *p. EP* is calculated by the following formula:

$$E_p = \frac{1}{2}\sum_{i=1}^{N}(O_i - t_i)^2$$

where *N* is the total number of output nodes, o_i is the network output at the i_{th} output node, and t_i is the target output at the i_{th} output node. In every training algorithm, an attempt is made to reduce this global error by adjusting the weights and biases.

3.4.1:Levenberg marquardet learning algorithm:

The Levenberg-Marquardt (LM) algorithm is an iterative technique that locates the minimum of a multivariate function that is expressed as the sum of squares of non-linear real-valued functions . It has become a standard technique for non-linear least-squares problems , widely adopted in a broad spectrum of disciplines. Levenberg-Marquardt algorithm was designed to approach second-order training speed without having to compute the Hessian matrix24. When the performance function has the form of a sum of squares (as is typical in training feed forward networks), then the Hessian matrix can be approximated as

$$H = J^T J$$

and the gradient can be computed as

$$g = J^T e$$

where *J* is the Jacobian matrix, which contains first derivatives of the network errors with respect to the weights and biases, and *e* is a vector of network errors. The Jacobian matrix can be computed through a standard back-propagation technique that is much less complex than computing the Hessian matrix. The Levenberg-Marquardt algorithm uses this approximation to the Hessian matrix in the following Newton-like update:

$$x_{k+1} = x_k - [J^T J + \mu I]^{-1} J^T e$$

When the scalar μ is zero, this is just Newton's method, using the approximate Hessian matrix. When μ is large, this becomes gradient descent with a small step-size. Newton's method is faster and more accurate near a minimum error, so the aim is to shift towards Newton's method as quickly as possible.

Thus, μ is decreased after each successful step (reduction in performance function) and is increased only when a tentative step would increase the performance function. In this way, the performance function will always be reduced at each iteration of the algorithm. The Levenberg-Marquardt optimization technique is more powerful than the conventional gradient descent techniques.

3.4.2:Scaled Conjugate Gradient learning algorithm:

SCG is a supervised learning algorithm for feedforward neural networks, and is a member of the class of conjugate gradient methods.The basic back-propagation algorithm adjusts the weights in the steepest descent direction (the most negative

of the gradients). This is the direction in which the performance function is decreasing most rapidly. It turns out that, although the function decreases most rapidly along the negative of the gradient, this does not necessarily produce the fastest convergence. In the conjugate gradient algorithms a search is performed along conjugate directions, which produces generally faster convergence than steepest descent directions27. In most of the conjugate gradient algorithms the step-size is adjusted at each iteration. A search is made along the conjugate gradient direction to determine the step-size, which will minimize the performance function along that line. All of the conjugate gradient algorithms start out by searching in the steepest descent direction on the first iteration.

$$P_o = -g_o$$

A line search is then performed to determine the optimal distance to move along the current search direction:

$$x_{k+1} = x_k + \alpha_k g_k$$

Then the next search direction is determined so that it is conjugate to previous search directions. The general procedure for determining the new search direction is to combine the new steepest descent direction with the previous search direction:

$$p_k = -g_k + \beta_k p_{k-1}$$

The various versions of conjugate gradient algorithms are distinguished by the manner in which the constant β_k is computed. For the Fletcher-Reeves update the procedure is

$$\beta_k = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}$$

This is the ratio of the norm squared of the current gradient to the norm squared of the previous gradient. This update procedure was used in the study and denoted as CGF (conjugate gradient with Fletcher-Reeves).

3.4.3Resilient backpropagation learning algorithm:

Resilient backpropagation is a learning heuristic for supervised learning in feedforward artificial neural networks.this is a first order optimization algorithm .Multilayer networks typically use sigmoidal transfer function in the hidden layer These functions are often called "squashing" functions, since they compress an infinite input range into a finite output range. Sigmoid functions are characterized by the fact that their slope must approach zero as the input gets large.

This causes a problem when using steepestdescent to train a multi-layer network with sigmoid functions, since the gradient can have a very small magnitude; and therefore, cause small changes in the weights and biases, even though the weights and biases are far from their optimal values. The purpose of the resilient back-propagation (RB) training algorithm is to eliminate these harmful effects of the magnitudes of the partial derivatives. Only the sign of the derivative is used to determine the direction of the weight update; the magnitude of the derivative has no effect on the weight update.

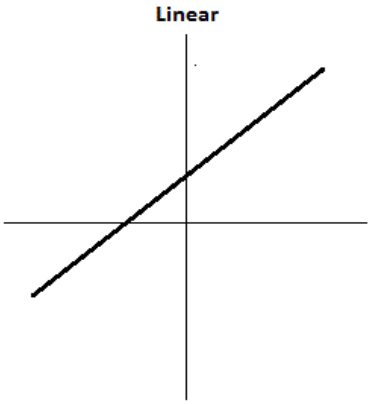
The size of the weight change is determined by a separate update value. The update value for each weight and bias is increased by a factor whenever the derivative of the performance function with respect to that weight has the same sign for two successive iterations. The update value is decreased by a factor whenever the derivative with respect to weight changes sign from the previous iteration. If the derivative is zero, then the update value remains the same.

Whenever the weights are oscillating the weight change will be reduced. If the weight continues to change in the same direction for several iterations, then the magnitude of the weight change will be increased.

3.5 Activation Function:

3.5.1:Linear activation function:

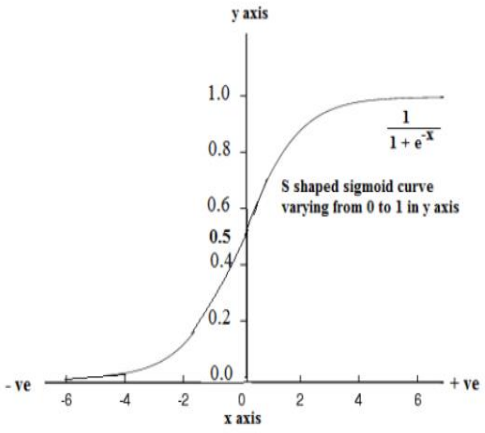
A Linear activation function is a mathematical function which divides plane into two parts(straight line). Like a linear regression, a linear activation function transforms the weighted sum inputs of the neuron to an output using a linear function.



3.5.2:Sigmoidal activation function:

A sigmoid function is a mathematical function having an "S" shape (sigmoid curve). Sigmoid function is given by the following formula:

$$S(t) = \frac{1}{1 + e^{-t}}$$



Sigmoid functions are very similar to the input-output relationships of biological neurons, although not exactly the same. Sigmoid function exhibits smoothness and has the desired asymptotic properties. The sigmoid curve is shown.As t goes to minus infinity, S(t) goes to 0. As t goes to infinity, S(t) goes to 1.As t =0, S(t) =0.5

CHAPTER 4

RESULTS AND DISCUSSION

We have taken single channel EEG signal from www.physionet.org and is plotted in matlab as shown below:

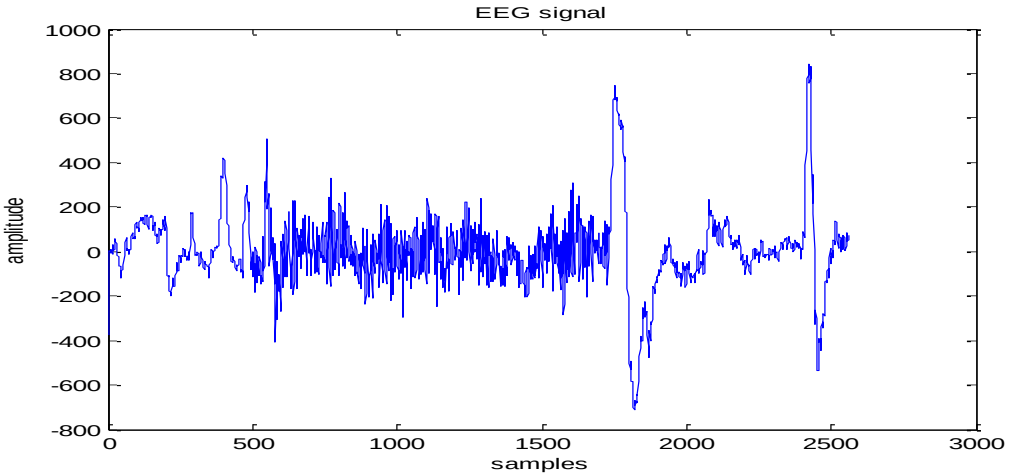


Figure 4.1:Original EEG signal

From above signal we have taken only first 35 samples(originally have 2560 samples) and the plot of 35 samples is shown below:

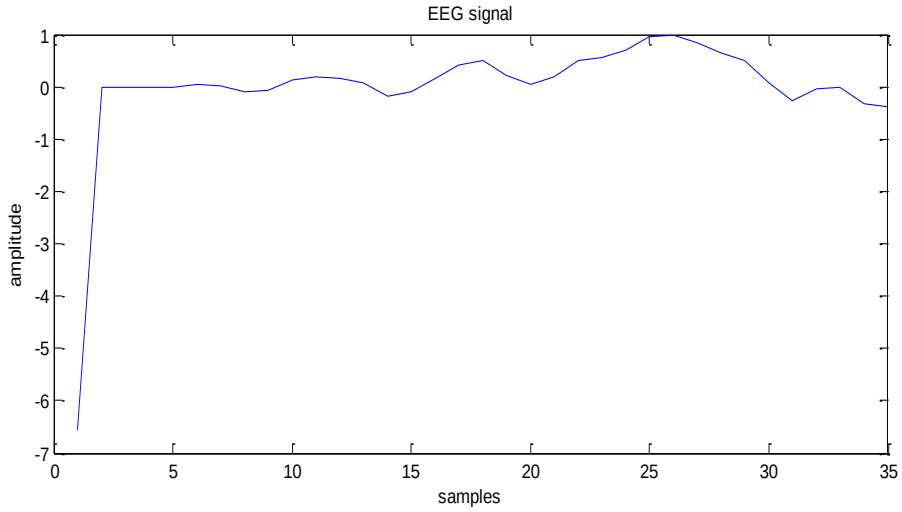


Figure 4.2:35 samples of EEG signal

Now the diferent noises are added to the signal and the signal got corrupt and this is shown below:

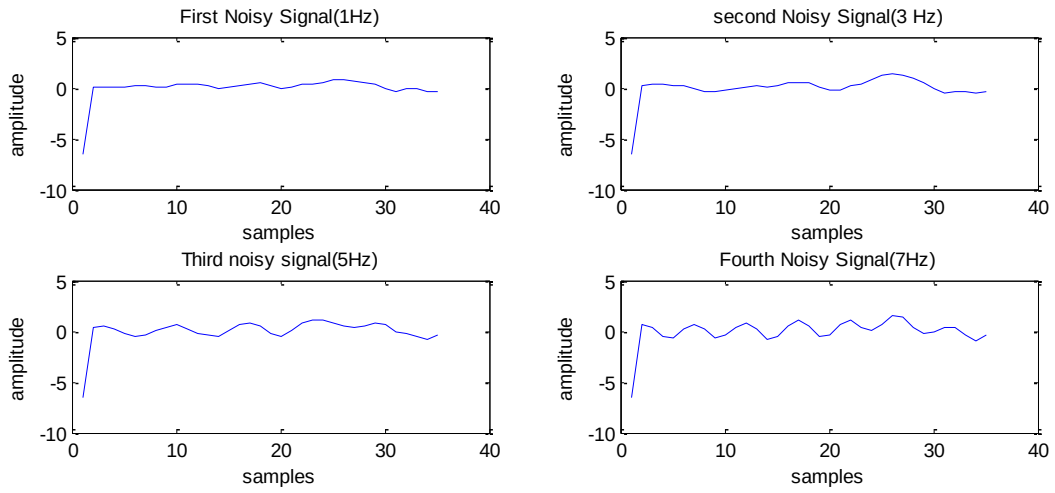


Figure 4.3: different noisy signal

Then we trained our network one by one using the above starting three noisy signals and then tested by using the fourth noisy signal and calculate signal to noise ratio(SNR),mean square error(mse),correlation to find which network provide better results using three learning mechanisms and two activation function as discussed above.

4.1:SIGNAL TO NOISE RATIO(SNR):

Signal to Noise Ratio or SNR is defined as the ratio of the transmitted power from the access point to the ambient (noise floor) energy present. To calculate the SNR value, we add the Signal Value to the Noise Value to get the SNR ratio. A positive value of the SNR ratio is always better. The SNR is a measure of signal strength relative to background noise after denoising. Theoretically, the SNR of a filtered EEG signal should be large in amplitude in order to recover the useful signal.

$$SNR = 10log\frac{\sum_{i=1}^{N} y^2(n)}{\sum_{i=1}^{N}|x(n)-y(n)|^2}$$

4.2:MEAN SQUARE ERROR:

In statistics, the mean squared error (MSE) or mean squared deviation (MSD) of an estimator (of a procedure for estimating an unobserved quantity) measures the average of the squares of the errors or deviations—that is, the difference between the estimator and what is estimated. MSE is a risk function, corresponding to the expected value of the squared error loss or quadratic loss. The difference occurs because of randomness or because the estimator doesn't account for information that could produce a more accurate estimate. The MSE is a measure of the quality of an estimator—it is always non-negative, and values closer to zero are better.The MSE is the second moment (about the origin) of the error, and thus incorporates both the variance of the estimator and its bias. Mathematically,given as:

$$MSE = \sqrt{(d(n) - y(n))^2}$$

4.3:CORRELATION:

Correlation is used to measure the linear relationship between to random variables, it used second order statistics to find the measure of similarity, the maximum value of correlation between two signal can be '1'. It can be mathematically given as,

$$r_{xs} = \frac{cov(x,s)}{\sigma_x \sigma_s}$$

Here x is the raw EEG signal and s is the noise free EEG signal, σ is the standard deviation and cov is the covariance of two random variable x and s.

For different networks using first learning mechanism i.e. Levenberg Marquardt and purelinear activation function and sigmoidal activation function above parameters are calculated (SNR, correlation, MSE, test error)and values are given in table below:

LM-BACKPROPAGATION:

- PURELINEAR

NETWORK	SNR	CORRELATION	ERROR	MSE
35:40:35	49.2225	0.9828	0.1847	0.2318
35:45:35	45.0989	0.9571	0.2928	0.3726
35:50:35	46.2047	0.9660	0.2684	0.3281
35:55:35	47.3468	0.9704	0.2395	0.2877
35:60:35	46.7938	0.9664	0.2311	0.3066
35:65:35	44.9497	0.9524	0.2825	0.3791
35:70:35	45.8416	0.9589	0.2457	0.3421

- SIGMOIDAL

NETWORK	SNR	CORRELATION	ERROR	MSE
35:40:35	79.6654	0.9489	0.058	0.0070
35:45:35	82.8564	0.9572	0.034	0.0048
35:50:35	79.8406	0.9425	0.055	0.0068
35:55:35	85.2182	0.9693	0.021	0.0036
35:60:35	85.2215	0.9744	0.021	0.0037
35:65:35	94.9321	0.9955	0.004	0.0012
35:70:35	80.2563	0.9583	0.055	0.0062

For different networks using second learning mechanism i.e. scaled conjugate gradient backpropagation and purelinear activation function and sigmoidal activation function above parameters are calculated (SNR, correlation, MSE, test error)and values are given in table below:

SCALED CONJUGATE GRADIENT BACKPROPAGATION

- PURELINEAR

NETWORK	SNR	CORRELATION	ERROR	MSE
35:40:35	49.2156	0.9828	0.1835	0.2310
35:45:35	45.1128	0.9572	0.2922	0.3720
35:50:35	46.2144	0.9661	0.2679	0.3277
35:55:35	47.3791	0.9706	0.2383	0.2866
35:60:35	46.8039	0.9666	0.2309	0.3062
35:65:35	44.9720	0.9526	0.2815	0.3781
35:70:35	45.8780	0.9592	0.2451	0.3407

- SIGMOIDAL

NETWORK	SNR	CORRELATION	ERROR	MSE
35:40:35	80.6057	0.9441	0.048	0.0063
35:45:35	83.2970	0.9507	0.027	0.0046
35:50:35	78.6207	0.9536	0.070	0.0079
35:55:35	89.2082	0.9819	0.006	0.0023
35:60:35	84.7750	0.9667	0.021	0.0039
35:65:35	84.6054	0.9633	0.026	0.0039
35:70:35	80.6668	0.9513	0.050	0.0062

For different networks using second learning mechanism i.e. resilient backpropagation and purelinear activation function and sigmoidal activation function above parameters are calculated (SNR, correlation, MSE, test error)and values are given in table below:

RESILIENT BACKPROPAGATION

- PURELINEAR

NETWORK	SNR	CORRELATION	ERROR	MSE
35:40:35	49.3440	0.9821	0.1711	0.2286
35:45:35	45.7762	0.9549	0.2967	0.3867
35:50:35	47.5674	0.9584	0.2886	0.3531
35:55:35	46.2235	0.9691	0.2394	0.2918
35:60:35	46.0280	0.9615	0.2663	0.3348
35:65:35	45.2138	0.9553	0.2803	0.3677
35:70:35	45.4863	0.9553	0.2658	0.3564

- SIGMOIDAL

NETWORK	SNR	CORRELATION	ERROR	MSE
35:40:35	81.0209	0.9539	0.047	0.0060
35:45:35	81.8482	0.9593	0.042	0.0054
35:50:35	79.0078	0.9574	0.065	0.0075
35:55:35	84.1873	0.9666	0.028	0.0041
35:60:35	84.8409	0.9668	0.023	0.0038
35:65:35	82.4735	0.9605	0.037	0.0050
35:70:35	79.6640	0.9504	0.059	0.0070

CHAPTER 5

CONCLUSION

In this study, a new ANN model to reduce different types of noise in EEG signals has been proposed using different learning mechanism with appropriate activation function. Noise reduction and cancellation are important for obtaining a clear and useful signal. Some signals, such as EEG, are nonstationary, and the noise statistical property is complicated because of the complexity of the signal.

During last years, numerous techniques have been proposed to reject different kinds of noise in EEG signals, including wavelet, SVD, AF and ICA. However, optimized results have not yet been obtained. In addition, these conventional filtering techniques can contain ripples that do not correspond to the original EEG. We have found that ANN improves all results obtained by the previous methods significantly reducing the interference. Moreover, SNR evolution has been more constant with ANN and more linear because the proposed method has introduced the lowest distortion to the signal.

With the different learning mechanism and activation function ,we found that LM-backpropagation provides the best results (n/w-35:65:35,SNR-94.9321,corr-0.995,test error-0.004,MSE-0.0012) with sigmoidal activation function.The present study has demonstrated how the proposed ANN network(35:65:35)can be used to reduce muscle and baseline noise in EEG recordings in one step. The above network has obtained greater corelation and SNR values than the other methods. Besides, this method has revealed to be an effective enhancement tool in all practical cases which have been studied, keeping clinical information in the brain signal.

REFERENCE

[1] Graupe D (2007) Principles of artificial neural networks, 2nd edn. World Scientific, Singapore.

[2] Kathirvalavakumar T (2010) Neural networks: FNN training algorithms: simultaneous perturbation, backpropagation and tunneling methods. VDM Verlag, Saarbru¨cken.

[3]. Haykin S (2008) Neural networks and learning machines, 3rd edn. Prentice Hall, Englewood Cliffs, NJ S.A. Hillyard and R. Galambos. Eye-movement artifact in the CNV. Electroencephalography and Clinical Neurophysiology, 28:173{182, 1970.

[4] R.A. Horn and C.R. Johnson. Matrix Analysis. Cambridge University Press, Cambridge, UK,1985.

[5] D.R. Hundley, M.J. Kirby, and M. Anderle. A solution procedure for blind signal separation using the maximum noise fraction approach: algorithms and examples. In Proceedings of the Conference on Independent Component Analysis, pages 337{342, December 2001.

[6] D.R. Hundley, M.J. Kirby, and M. Anderle. Blind source separation using the maximum signal fraction approach. Signal Processing, 82:1505{1508, 2002}.

[7] A. HyvÄarinen, J. Karhunen, and E. Oja. Independent Component Analysis. John Wiley & Sons, New York, 2001.

[8] A. HyvÄarinen and E. Oja. Independent component analysis: Algorithms and applications. Neural Networks, 13(4-5):411{430, 2000}.

[9] H.H. Jasper. The ten-twenty electrode system of the international federation. Electroen-cephalography and Clinical Neurophysiology, 10:371{373, 1958}.

[10] T-P. Jung, C. Humphries, M. Lee, V. Iragui, S. Makeig, and T. Sejnowski. Removing electroencephalographic artifacts: Comparison between ICA and PCA. In IEEE International Workshop on Neural Networks for Signal Processing, pages 63{72, 1998.

[11] O'Leary, J. L., and Goldring, S.,Science and Epilepsy, Raven Press, New York, 1976, pp.19–152.

[12] Gotman, J., Ives, J. R., and Gloor, R., 'Automaticrecognition of interictal epileptic activity in prolonged EEG recordings',Electroencephalogr. Clin. Neurophysiol., 46, 1979, 510–520.

[13] Estrada E, Nazeran H, Sierra G, Ebrahimi F, Setarehdan SK: Wavelet-based EEG denoising for automatic sleep stage classification, in 21st International Conference on Electrical Communications and Computers (CONIELECOMP). , San Andres Cholula; 2011:295–298.

[14] Daubechies, I, "Ten Lectures on Wavelets" (CBMS-NSF Regional Conference Series in applied Mathematics), SIAM, 1992.

[15] Coifman R R and Donoho D L 1995 Translation-invariant de-noisingTechnical Report475 (Department of Statistics, Stanford University)

[16] Percival D B and Walden A 2000Wavelet Methods for Time Series Analysis (Cambridge: Cambridge University Press).W. O. Tatum, B. A. Dworetzky, and D. L. Schomer, "Artifact and recording concepts in EEG," *J. Clinical Neurophysiol.*, vol. 28, no. 3, pp. 252–263, Jun. 2011.

[17] J. L. Kenemans, P.C.M.Molenaar, M. N. Verbaten, and J. L. Slangen, "Removal of the ocular artifact from the EEG: A comparison of time and frequency domain methods with simulated and real data," *Psychophysiology*, vol. 28, pp. 114–121, Jan. 1991.

[18] V. Krishnaveni, S. Jayaraman, S. Aravind, V. Hariharasudhan, and K. Ramadoss, "Automatic identification and removal of ocular artifacts from EEG using wavelet transform,"*Meas. Sci. Rev.*, vol. 6, no. 4, pp. 45–57, 2006.

[19] S. Hu, M. Stead, and G. A. Worrell, "Automatic identification and removal of scalp reference signal for intracranial EEGs based on independent component analysis," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 9, pp. 1560–1572, Sep. 2007.

[20] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, Eds. New York: Springer-Verlag, 1995, vol. 103, pp. 125–150.

[21] H. Ocak, “Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy,” *Expert Syst. With Appl.*, vol. 36, no. 2, pp. 2027–2036, 2009.

[22] R. E. J. Yohanes, S. Wee, and H. Guang-Bin, “Discrete wavelet transform coefficients for emotion recognition from EEG signals,” in *Proc. 2012 Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, pp. 2251–2254.

[23] M. S. Reddy, M. R. Kumar, K. P. Chander, and K. S. Rao, “Complex wavelet transform driven de-noising method for an EOG signals,” in *Proc. 2011 Annu. IEEE India Conf. (INDICON)*, pp. 1–5.

[24] C. A. Joyce, I. F. Gorodnitsky, and M. Kutas, “Automatic removal of eye movement and blink artifacts from EEG data using blind component separation,” *Psychophysiology*, vol. 41, no. 2, pp. 313–325, Mar. 2004.

[25] Y. Okad, J. Jung, and T. Kobayashi, “An automatic identification and removal method for eye-blink artifacts in event-related magnetoencephalographic measurements,” *Physiol. Meas.*, vol. 28, no. 12, p. 1523, Oct. 2007.

[26]Yu-Quan Z, Ji-Shun O, Geng C, Hai-Ping Y (2011) Dynamic weighting ensemble classifiers based on cross-validation. Neural Comput Appl 20(3):309–317.

[27] Mateo J, Torres AM, Garcı´a MA (2014) Dynamic fuzzy neural network based learning algorithms for ocular artefact reduction in EEG recordings. Neural Process Lett 39(1):45–67.

[28] Zhang L, He C (2012) Quantitative methods for detecting cerebral infarction from multiple channel EEG recordings. Neural Comput Appl 21(6):1159–1166.

[29] Sokouti B, Haghipour S, Tabrizi A (2014) A framework for diagnosing cervical cancer disease based on feedforward MLP neural network and ThinPrep histopathological cell image features.Neural Comput Appl 24(1):221–232