Frequency Recognition of Selected Features of EEG Signals with Neuro-Statistical Method

by

Md. Zakir Hossain

A thesis submitted in partial fulfillment of the requirement for the degree of Master of Science in Electrical and Electronic Engineering

Department of Electrical and Electronic Engineering
Khulna University of Engineering & Technology
Khulna 9203, Bangladesh

April 2014

Declaration

This is to certify that the thesis work entitled "*Frequency Recognition of Selected Features of EEG Signals with Neuro-Statistical Method*" has been carried out by *Md. Zakir Hossain* in the Department of *Electrical and Electronic Engineering*, Khulna University of Engineering & Technology, Khulna, Bangladesh. The above thesis work or any part of this work has not been submitted anywhere for the award of any degree or diploma.

20/04/2014
Signature of Supervisor

20.04.14
Signature of Candidate

Approval

This is to certify that the thesis work submitted by *Md. Zakir Hossain* entitled "*Frequency Recognition of Selected Features of EEG Signals with Neuro-Statiscal Method*" has been approved by the board of examiners for the partial fulfillment of the requirements for the degree of *Masters of Science* in the Department of *Electrical and Electronic Engineering*, Khulna University of Engineering & Technology, Khulna, Bangladesh in April 2014.

BOARD OF EXAMINERS

1. 20/04/2014 ______________ Chairman (Supervisor)
   Prof. Dr. Md. Shahjahan
   Department of Electrical and Electronic Engineering,
   Khulna University of Engineering & Technology,
   Khulna, Bangladesh.

2. 20/4/14 ______________ Member
   Head of the Department
   Department of Electrical and Electronic Engineering,
   Khulna University of Engineering & Technology,
   Khulna, Bangladesh.

3. 20/4/14 ______________ Member
   Prof. Dr. Mohiuddin Ahmad
   Department of Electrical and Electronic Engineering,
   Khulna University of Engineering & Technology,
   Khulna, Bangladesh.

4. 20/04/2014 ______________ Member
   Prof. Dr. Mohammad Shaifur Rahman
   Department of Electrical and Electronic Engineering,
   Khulna University of Engineering & Technology,
   Khulna, Bangladesh.

5. M. M. Islam ______________ Member (External)
   Prof. Dr. Md. Monirul Islam
   Department of Computer Science and Engineering,
   Bangladesh University of Engineering & Technology,
   Dhaka, Bangladesh.

Dedicated

To

My Beloved Parents

&

Respected Teachers

Acknowledgement

It is a great pleasure for me to express unbound indebted gratitude to my supervisor Prof. Dr. Md. Shahjahan, Professor, Department of Electrical & Electronic Engineering of Khulna University of Engineering & Technology, for his numerous helpful & constructive suggestions, scholastic guidance, constant inspiration & support, valuable advice and kind co-operation for the successful completion of work "*Frequency Recognition of Selected Features of EEG Signals with Neuro-Statistical Method*". He has always been extremely generous with his time, knowledge and ideas and allowed me great freedom in this research. His enthusiastic approach and endless excitement to research and effervescent personality have made this thesis all the more enjoyable and I am greatly encouraged.

I would also like to thank my family and friends for their continuous encouragement and moral support.

Author

ABSTRACT

The advent of research work to analyze massive multiway oriented electroencephalogram (EEG) signals with low configurable computer is a great challenge. This thesis presents an algorithm for extracting underlying frequency components of such EEG data. Frequency components of these data play a vital role to realize brain-body condition. Usually, a huge amount of time and specially built computers are essential to process these EEG data having different subjects. It also restricts to visualize inherent frequency of EEG for a general practitioner. An algorithm is developed using two-stage cascaded architecture of canonical correlation analysis with neural network named neural canonical correlation analysis (NCCA) to address three major challenges for extracting frequency components from EEG data, such as: (a) It processes massive data which are feed sequentially into neural network, rather than feeding whole data at a time, (b) It uses the conventional personal computer instead of special computer built for such application, (c) It spends very short time for a moderate data set consisting of several ways (time, trials and channels). (d) It considers the nonlinear correlation among the data groups while statistical CCA ignores it. In order to get reliable and robust result, the experimental are carried out with different structures of network such as linear, nonlinear and nonlinear feedback structures. The inherent dominant frequency of 1 Hz having a quite resemblance with EEG landscape has been found. This provides a great opportunity in analyzing brain-body function.

Although it is possible to recognize frequency of massive EEG data at shorter time with NCCA than statistical CCA, but subjects differentiation is still a great challenge. In this view, this paper presents a new feature selection (FS) approach based on NCCA. In order to get robust features subset having maximum correlation and minimum redundancies, NCCA is devised to search highly correlated subsets by maximizing correlation among several subdivisions of raw data and pruning the features of lightly scored weights of CCA network. The result of NCCA is very robust in terms of accuracy. In this sense, frequency recognition is very easy using selected EEG features than original features which are inspected from correlation profiles. The computational complexity is also greatly reduced if selected features are used to recognize frequency which is proved theoretically and experimentally. In this connection, elapsed time is calculated and observed that NCCA is about 2 to 33 times faster to recognize frequencies from selected EEG features than original set.

Contents

		PAGE
	Title Page	I
	Declaration	II
	Certificate of Research	III
	Dedication	IV
	Acknowledgement	V
	Abstract	VI
	Contents	VII
	List of Tables	IX
	List of Algorithms	IX
	List of Figures	X
CHAPTER I	Introduction	1
	1.1 Motivations	1
	1.2 Goals and Approaches	5
	1.2.1 Neuro-Statistical Aspects	5
	1.2.2 Feature Selection	6
	1.2.3 Frequency Recognition	7
	1.3 Organization of the Thesis	8
CHAPTER II	Background and Literature Review	9
	2.1 Electroencephalography (EEG)	9
	2.1.1 Brain Rhythemicities	10
	2.1.2 EEG Potentials	12
	2.2 Data Collection	13
	2.2.1 Experimental Orientations	13
	2.2.2 EEG Data Acquisition – EEG System	14
	2.2.3 Characteristics of Data	15
	2.3 Aspects of Frequency Recognition	16
CHAPTER III	The Proposed Framework: Neuro-Statistical CCA	21
	3.1 Neuro-Statistical Technique	21
	3.2 Traditional CCA	23
	3.3 The Framework	24
	3.4 Feature selection Procedure	25
	3.4.1 Clustering using Wavelet	26
	3.4.2 Data Preparation	27
	3.4.3 Feature Selection using CCA Network	28
	3.5 Procedure for Frequency Recognition	30
	3.5.1 Reference Signal	30
	3.5.2 Implementation of Colin's CCA Network	30
	3.5.3 Extraction of Frequency Components	31

CHAPTER IV	Experimental Studies	35
	4.1 Experimental Setup	35
	4.2 Results of Feature Selection	36
	4.2.1 Class determination	36
	4.2.2 Evaluation with Neural Network (NN)	36
	4.3 Results of Frequency Recognition (FR)	41
	4.3.1 FR from Original EEG	41
	4.3.2 Subjects Variability Realization	45
	4.3.3 FR from Selected Features	48
	4.4 Comparative Study	51
	4.4.1 Correlation Point of View	51
	4.4.2 Computational Complexity	53
	4.4.3 Comparison with Standard Methods	55
CHAPTER V	Conclusions and Future Works	56
	5.1 Conclusions	56
	5.2 Future Works	57
References		57

LIST OF TABLES

Table No.	Description	Page
2.1	Number of features that's are collected SSVEP data	15
4.1	Accuracy (%) for different subsets of EEG data	39
4.2	Confusion Matrix For 15 Features of Subject 4	40
4.3	Confusion Matrix for 15 Features of 28 Hz Stimulation	40

LIST OF ALGORITHMS

Algorithm No.	Description	Page
3.1	Feature selection algorithm using CCA network	29
3.2	Algorithm for recognizing frequencies using two-stage CCA network	34
4.1	NCCA algorithm for inter subject variability test	45

LIST OF FIGURES

Table No.	Description	Page
1.1	Representation of the wrapper and filter approach	3
1.2	General overview of feature selection	6
1.3	General overview to recognize frequency from selected EEG features.	7
2.1	One second of EEG signals	10
2.2	Different types of brain rhythmecities related to frequency band	11
2.3	Gamma wave	11
2.4	Position of 128 electrodes in head to record EEG signal	14
2.5	Single-trial SSVEP response with 128 Bio-semi active electrodes	15
2.6	Schematic diagram to extract frequency from recorded SSVEP	16
3.1	The CCA network	21
3.2	Framework of the two-stage cascaded Neuro-Statistical CCA	25
3.3	FS using CCA network	29
3.4	Illustration of two-stage CCA network for extraction of frequency components	33
4.1	Classification of EEG data using wavelet	36
4.2	Weight updates process of a generalized Standard BP	37
4.3	NN model for accuracy measurement of selected features	37
4.4	Iterations vs. error curve for three sets of selected features	39
4.5	Correlation coefficient of NCCA with (a) linear, (b) nonlinear and (c) nonlinear with feedback networks for H1 settings.	42
4.6	Correlation coefficient of NCCA with (a) linear, (b) nonlinear and (c) nonlinear with feedback networks for H2 settings.	43
4.7	Correlation coefficient of NCCA with (a) linear, (b) nonlinear and (c) nonlinear with feedback networks for H3 settings.	44
4.8	Illustration of NCCA approach for inter-subject variability	45
4.9	Inter subject variability with 8 Hz frequency	46
4.10	Inter subject variability with 14 Hz frequency	46
4.11	Trial to trial variability with sine-cosine reference signal	47
4.12	Trial to trial variability for trial5 of sub2 at 14 Hz	47
4.13	Inter subject variability with sine-cosine signal	48
4.14	Correlation profiles of (a) *S3*, (b) *S4*, (c) *S1* and (d) *S2* with linear CCA network	49
4.15	Correlation profiles of (a) *S4*, (b) *S3*, (c) *S2* and (d) *S2* with nonlinear CCA network	50
4.16	Correlation profiles of (a) *S4*, (b) *S2*, (c) *S2* and (d) *S1* with nonlinear feedback CCA network	51
4.17	Correlation profiles of (a) *S1*, (b) *S2*, (c) *S3* and (d) *S4* for original and selected features of EEG data	53
4.18	Time elapsed to recognize frequency of (a) *S1*, (b) *S4*, (c) *S3* and (d) *S2* with three CCA networks	54

CHAPTER I

Introduction

In this chapter we develop the basic ideas of frequency recognition from the perspective that the standard treatments already contain the motivations for the necessity to know the behavior of human brain-body functions. In this regard, we evaluate electroencephalogram (EEG) signal which is generated from human brain and related to body functions. This signal is roughly less than 100 μV and can be measured with electrodes placed on the scalp, noninvasively. The brain's electrical charge of signals is maintained by billions of neurons. Neurons are electrically charged by membrane transport proteins that pump ions across their membranes. Neurons are constantly exchanging ions with the extracellular milieu, for example to maintain resting potential and to propagate action potentials. Ions of similar charge repel each other and when many ions are pushed out of many neurons at the same time, they can push their neighbors, who push their neighbors, and so on, in a wave. This process is known as volume conduction. When the wave of ions reaches the electrodes on the scalp, they can push or pull electrons on the metal on the electrodes. Since metal conducts the push and pull of electrons easily, the difference in push or pull voltages between any two electrodes can be measured by a voltmeter. Recording these voltages over time gives us the EEG [1].

The frequency recognition from massive multiway [2] EEG data is a great challenge for computational scientists. This requires high capacity machine which are specially built for a particular EEG application and are not available for general users or practitioners. Therefore, a comfortable and user friendly faster method is essential to process multiway data with limited computational resources such as an ordinary personal computer. In this regard, to solve this problem we propose implementation of Canonical Correlation Analysis (CCA) with Neural Network (NN) called Neuro-Statistical CCA.

1.1 Motivations

Frequency components recognition from a signal gives us knowledge about analyzing a waveform to learn something about the source, propagation of waves, simplifies the waveform to understand easily. EEG [3] is one kind of electrical signal which always reflects the summation of the synchronous activity of thousands or millions of neurons that have similar spatial orientation. If the cells do not have similar spatial orientation, their ions do not line up and create waves to be detected. Pyramidal neurons of the cortex are thought to produce the most EEG signal because they are well-aligned and fire together. There are different kinds of potential whose are generated from brain; Steady State Visual Evoked Potential (SSVEP) is one of them. It is evoked over occipital scalp areas, with repetitive external visual stimulation such as flashes, reversing patterns or luminance-modulated images [4]. The strongest responses of SSVEP are occurred in the primary visual (striate) cortex, although other brain areas are also activated in varying degrees.

The EEG is typically described in terms of rhythmic activity. The rhythmic activity is divided into frequency bands. These rhythmic activities within a certain frequency range were noted to have a certain distribution over the scalp or a certain biological significance. Frequency bands are usually extracted using spectral methods (for instance Welch) as implemented for instance in freely available EEG software [5]. A series of operation is repetitively required to get final frequency components if a conventional software or frequency analyzer algorithm is used. Moreover, there has been computational intractability if the data are massive and multiway.

Recent approaches try to find inherent underlying frequency components of these EEG signals. However, EEG signal may be contaminated by noise and it is still a challenge to detect the rhythmic activities of such signals especially at low stimulus frequency [6]. There have been a number of approaches to recognize frequency of EEG signals. A traditional and widely used method for EEG signal recognition is power spectral density analysis (PSDA). PSD is estimated from the EEG signals within a time window typically by Fast Fourier Transform (FFT), and its peak is detected to recognize the target stimulus. It takes longer time window to estimate spectrum with sufficient resolution [7]. Some studies also took the PSDs as features and applied linear discriminate analysis (LDA) or support vector machine (SVM) classifier to classify the desired frequency [8] which may limit the real-time implementation.

Lin has found out the correlations between a set of EEG signals of multiple channels and a set of reference sine-cosine signals with different stimulus frequencies using statistical Canonical Correlation Analysis (CCA) [9]. The desired stimulus is then recognized from conversed correlations by maximization process. It provides better recognition performance than that of the PSDA since it delivers an optimization for the combination of multiple channels and improves the noise tolerance. A comparative analysis between the CCA and PSDA was also discussed in [10]. They also adopted the sine-cosine waves as reference signals used in the CCA for SSVEP recognition. Tensor CCA is an extension of the statistical CCA, which addresses on inspecting the correlation between two multiway data groups, instead of two sets of variables [11]. Multiway CCA (MCCA) [12] has been proposed to address the real time implementation of brain computer interface system. They remove inter subject variability and trial to trial variability in finding the optimized reference signals, although it requires specially built computers which are not available for general purpose. The matlab program for MCCA requires very long time to execute. However, the detail insight realization of correlation profile is still missing.

In this thesis we implement three types of CCA networks such as – (i) linear, (ii) nonlinear and (iii) nonlinear feedback networks to extract underlying frequency components of EEG signals using a two-stage structure of neural CCA which maximizes correlation between a set of sine cosine signals and a set of EEG signals. As a result, an optimized reference signal is obtained in the first stage. In the second stage, a test set of EEG signals and optimized reference signal are applied to the same network to find another optimized signal. Finally, frequency components of EEG data set are determined from above two optimized signals where their correlation becomes maximum. This does not require high capacity machine and it performs better than others since special NN cascade architecture is incorporated.

The SSVEPs of EEG can be processed for different visual stimulation with standard EEG system for different trials of a subject, but there have no significant variability among trials for a subject at same stimulation [13]. If every trail of any subjects is concatenated together for analysis, the data size is so high as well as every feature is not similarly important. The processing of such massive data is a great challenge for computational scientists. In this sense, we search salient features of such high dimensional data with reduced size which will carry important information.

In order to process these kinds of data, huge computational time and resources are often required. In many pattern recognition applications, there have a large number of features those are not equally important for a specific task [14]. Some of the variables may be redundant or even irrelevant. Usually by discarding such variables better performance may be achieved [15]. Moreover, the number of training samples required grows exponentially when the number of features grows [16]. Therefore, dimensionality reduction of the data is most important in many practical applications. Feature Selection (FS) algorithms attempt to reduce insignificant attributes and these are widely applied as a pre-processing tool for pattern recognition, data mining, text categorization, image mining, and frequency recognition etc [17].

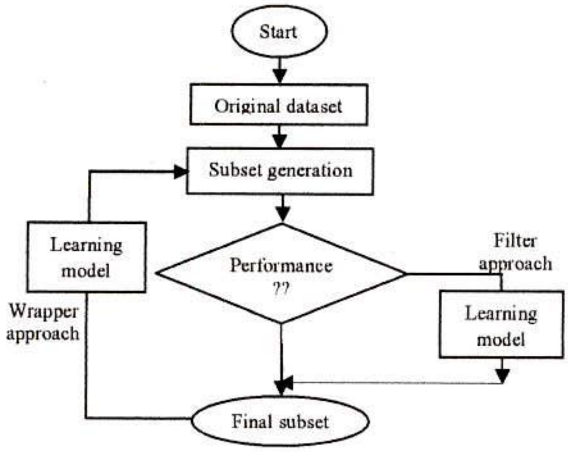


Fig. 1.1: Representation of the wrapper and filter approach

There are three basic types of FS approaches such as wrapper, filter and hybrid [18]. In case of wrapper approach, predetermined learning model is assumed and features are selected that justify the learning performance of the particular learning model. On the other hand, statistical method is utilized without any learning model in case of filter approach. The classification of them is shown in Fig. 1. The hybrid model takes advantage of the complementary strengths of the wrapper and filter approaches by using their different evaluation criteria at different search stages [19]. Different types of search processes are attempted for subsets generation. The search process starts with an empty set and successively adds features in case of sequential forward search (SFS) [20], where sequential backward search (SBS) [21] option starts with a full set and features are successively removed. On the other hand search processes start from both ends and add and remove features simultaneously, for bidirectional selection [22]. A fourth approach [23] is to start a search process with a randomly selected subset using bidirectional or sequential strategy.

There have been several approaches attempted to select important features. Broadly FS techniques can be categories as evolutionary and non-evolutionary. Evolutionary algorithms need very long time to converge due to its very high search spaces [24]. On the other hand, non-evolutionary algorithms require very short time to converge and are useful for online application. Among the non-evolutionary approaches principal component analysis (PCA), canonical correlation analysis (CCA), independent component analysis (ICA), information gain (IG) are worth mentioning.

PCA is a classic tool to seek linear combinations of the data variables that capture a maximum amount of variance [25]. It ranks the variables to explain greatest variation of data and identify the most important variables from amongst large set of possible influences. But there are few restrictions using PCA. One of them is its dealing with only one set of variables. On the other hand, it may consider noisiest variables because they account for most of the variation, where data normalization cannot always be the solution as well as it may not cope with data having class attributes. The most important limitations of PCA however is that each principle component is a linear combination of all variables in the dataset which ends up being very difficult to interpret. There are also different space reductions methods such as SVD (singular value decomposition) and LSA (least square approximation) those are very similar to PCA but only difference is in basis evaluation approach. Although ICA was successful in several attempts for FS [26], it requires non-Gaussian constraints. The performance of ICA depends on the nature of the task and the algorithm used. Moreover there is no statistical difference between ICA and PCA [27].

However it is crucial to identify a subset of features that are informative and significant for classification, easy to interpret and is not sensitive to scale effect. Mutual information based feature selection overcomes all of those challenges. Here, features are selected in a greedy manner [27] by measuring the mutual dependence of two variables. Attribute that gains the most information from a decision tree are chosen [28]. It is very popular due its robustness and excellent speed, although it can sometimes over fit training data, resulting large trees. Important variants can also be identified according to strong interactions of data sets using recursive elimination of features (Relief-F), but it is sensitive to presence of noise in the attributes [29]. This can be overcome by removing variables iteratively with the worst Relief-F scores and update the scores of the remaining variables [30]. A composite score is created from Relief-F and information gain (IG) called evaporative cooling (EC) which demonstrate greater power than iterative Relief-F to detect the continuum of independent and pure interaction effects. But computing IG is sensitive to real valued continuous feature set [31], so discretization of numeric features is required prior to computation IG.

The selection of feature subspaces can be found using a generalized CCA framework using a minimum mean-square-error criterion [32], but it considers only linear combination of two data stream ignoring the nonlinear relationship. Finding the canonical variates is not very difficult while interpretation is cumbersome due to presence of noise in the attributes. Due to the high volume of data available in recent years, it is not easy to analyze such data using classical methods. These problems can be resolved by combining neural network (NN) with statistical standard CCA, so that any linear and nonlinear correlation can be optimized.

We propose a framework for FS based on Colin Fyfe's CCA network [33] that is a simple implementation of statistical CCA with NN called CCA network (NCCA). To search the important features using the framework, firstly entire EEG data of a particular subject are divided into three groups intentionally. Then highly correlated feature subsets are obtained using maximization process through CCA network training and pruning consequently. The joint use of NN and statistics strengthen the FS process and exhibit robust results.

Although expected frequencies can be recognized using whole of the data set, but all features of these high dimensional data are not equally important and may also have noise corrupted data as well. Hence, difference of correlation points between expected frequencies with others is very low. In this sense, firstly features are selected using NCCA approach and then performances of above three networks are analyzed to recognize the frequencies. It is observed that in this case networks can recognize frequencies with higher accuracy and lower computational cost as well as it can differentiate among subjects easily.

1.2 Goals and Approaches

Our goal is to recognize frequencies which is generated from human brain and recorded as SSVEP. Due to the challenges to recognize underlying frequency components of high dimensional EEG signals, it is imperative to take an iterative approach to successfully develop high-performance signal processing loom that extract frequencies with higher accuracy and lower computational cost. In this thesis, we search salient features and underlying frequency components of EEG signals with neuro-statistical method.

1.2.1 Neuro-Statistical Aspects

CCA is a statistical method which finds linear relationship among multivariate data. It search correlation considering whole dataset at once, therefore it degrade the generalization performance of machine and takes very long time to execute a program. In this regard, NN is implemented with CCA which is called Neuro-Statistical CCA. This gives several advantages such as i) it does not require high capacity machine, ii) it uses NN, since it exhibits enhanced correlation than standard statistical methods [35], iii) a set of data out of entire one, is entered sequentially in the NN instead of complete data at a time, iv) remove noisy features easily because it uses training and pruning consequently.

SSVEP is one kind of potential of EEG signal. It is evoked over occipital scalp areas with the same frequency as the visual stimulus and may also include its harmonics when subject focuses on the repetitive flicker of a visual stimulus [3, 8]. Recent approaches try to find inherent underlying frequency components in EEG signals. There have been a number of approaches to recognize frequency of EEG signals. A traditional and widely used method for EEG signal recognition is PSDA. PSD is estimated from the EEG signals within a time window typically by Fast Fourier Transform, and its peak is detected to recognize the target stimulus. It takes longer time window to estimate spectrum with sufficient resolution [21]. CCA was used to find the correlations between the EEG signals of multiple channels and reference signals of sine-cosine with different stimulus frequencies as proposed by Lin et al. [9]. Then, the target stimulus is recognized through maximizing these correlations. We implement CCA with NN in original and selected features of EEG, since NN is well known for their powerful capacity [33, 34].

The EEG data can be configured using different ways such as time, channels and trials. When these data are recorded with higher number of channels for various trials of a subject, then it contains large feature set. The processing of such massive data is a great challenge for computational scientists. This high dimensional EEG data cause learning to be more difficult and also degrade the generalization performance of the learned models. To simplify and improve the quality of dataset, it is needed to select the salient features of high-dimensional datasets. Generally for this purpose FS is used in machine learning. Ordinarily, spurious features are deleted from the original dataset using FS without sacrificing generalization performance [14]. In this sense we employed CCA with NN called neuro-statistical CCA (NCCA) for selecting important features from large EEG dataset.

In this thesis, we try to find frequency components of such EEG signals using neuro-statistical CCA. It is observed that remarkable result is obtained from both original data and selected feature set in this test, but computational cost is greatly reduced when feature selection is performed prior to frequency recognition.

1.2.2 Feature Selection (FS)

In this thesis, we search salient features of massive EEG data with reduced size. SSVEP is one of most important EEG signal which detects the human brain condition at various modes such as reading, writing, watching TV etc that is on the view of opening eyes. But all of these data is not equally important; also handing of such high dimensional data is not an easy task. The correlated data is most important to find the brain conditions. For that motive, we search salient features of EEG data on the basis of NCCA. The correlation coefficient is computed from subdivision of EEG signals. The entire FS scheme is exposed in Fig. 1.2.

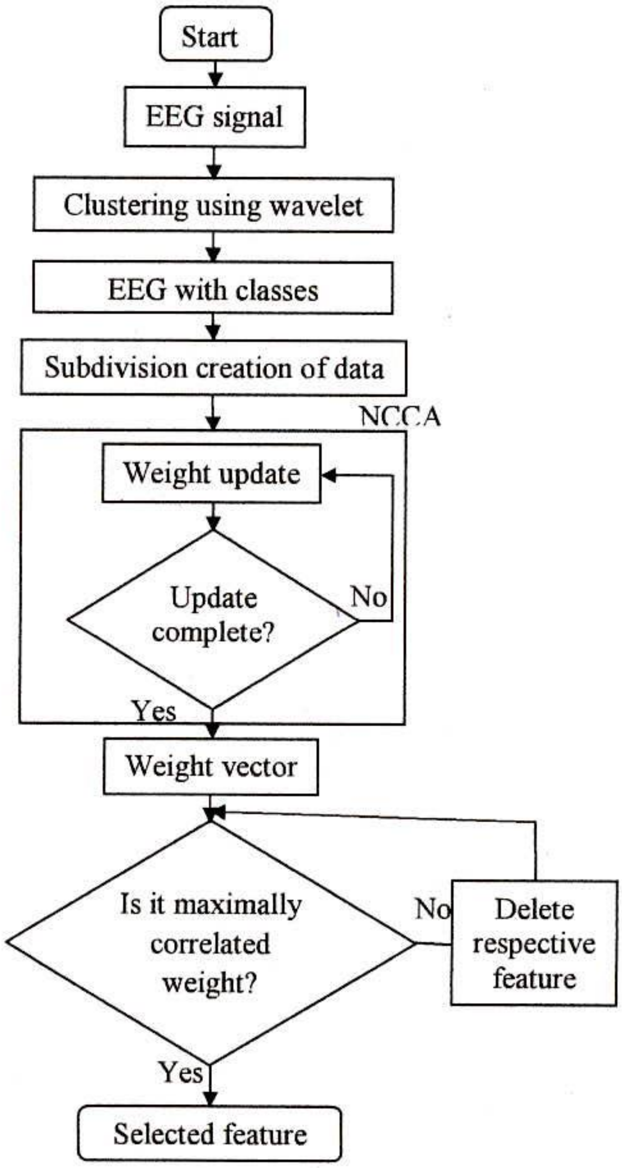


Figure 1.2: General overview of feature selection.

High dimensional EEG data are organized into three different groups using wavelet clustering method. This class information is added together with EEG according to pattern. Then NCCA is applied on EEG data to search salient features where statistical analysis of the feature set is utilized without assuming predefined learning model as filter approach [36]. Then spurious features are deleted according to low correlation using pruning. Finally well defined back-propagation (BP) algorithm of NN is used for measuring classification accuracy of selected EEG features.

1.2.3 Frequency Recognition

Frequency recognition from a signal is an important issue that gives us information about the source of signal. Propagation of signal through a medium generally depends on frequency which may say that the waves of different frequencies propagate with different velocities. EEG is one kind of electrical signal which is generated from human brain and related to body functions. To find the behavior of a signal or activities of human brain, we decompose the EEG signal into its different frequency components. It also gives us easy understanding of the signal. The EEG signal is collected from brains with electrodes (channels) placed on the scalp, noninvasively for a sufficient time. Frequency components extraction from this EEG signal play great role for determining brain functions. It is crucial to take an iterative approach to successfully develop high-performance signal processing loom that recognize frequencies with higher accuracy and lower computational cost. This iterative process is divided into the following tasks as shown in Fig. 1.3, most of which are to be pursued simultaneously.

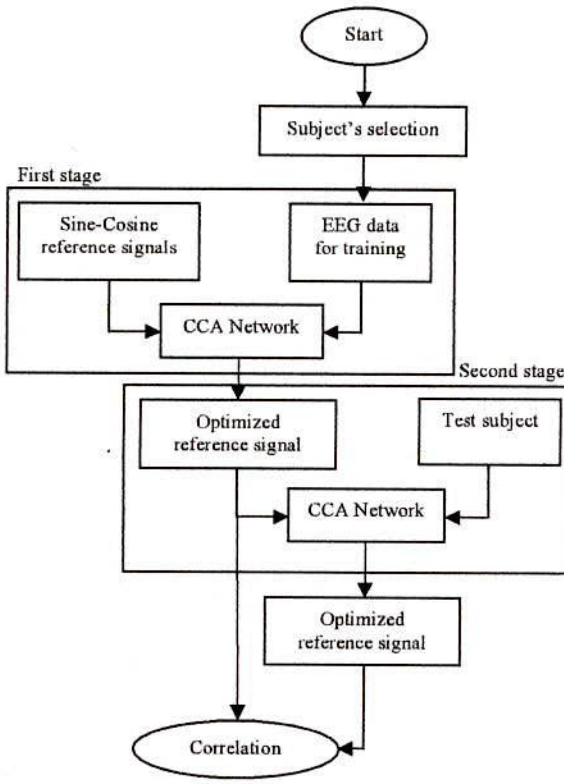


Fig. 1.3: General overview to recognize frequency from selected EEG features.

In this thesis, underlying frequency components of EEG signals are recognized on the basis of correlation maximization with CCA networks. EEG data is collected from EEG database [37]. There are four subjects with five trials of each at three (8 Hz, 14 Hz and 28 Hz) stimulus frequencies. Every trials of a subject are concatenated and about equally subdivided into 'M' subsections prior to present in the network. In this way, expected features are selected from high dimensional EEG data. These selected features are presented into three different CCA networks one by one to recognize frequencies. In this case, two-stage CCA network are utilized where three subjects are concatenated for first stage and rest is used as a test set in the second stage. In the first stage, a reference signal is generated with the sine-cosine set. Correlation maximization is done to find two optimized signals one from first stage and other from second stage. Both of the optimized signals carry the information of subject specific and trial to trial variability because of these are created from EEG signals. Then frequency is recognized from these two optimized signals with maximization process.

Finally, accuracy of selected features is tested with well known BP rule. Comparative analysis of recognizing frequencies from selected and non-selected features are also analyzed with three different CCA networks at three harmonic situations. It is seen that maximum correlations are found at 1 Hz for every network at every situation, because of checkerboard was flickered at this frequency. We can also differentiate among subjects using selected EEG features. The outcome of this work will enable us to design an efficient and user-friendly Brain Computer Interface in future.

1.3 Organization of the thesis

This thesis focuses on an integrated approach to recognize frequencies by optimizing SSVEP of EEG signals. This is accomplished by investigating the brain responses to continuous visual flicker stimulation with very small checkerboards. The motivations and scope of thesis work is described in Chapter I. It also gives an overview of the potentiality of neuro-statistical aspects to select salient features as well as recognize frequencies from noisy high dimensional EEG data.

Background information and literature review on frequency recognition of EEG signals are described in Chapter II. Rhythmic activities of brain signal as well as EEG data collection scheme and various types of EEG potentials are also explained in this chapter.

The proposed framework of neuro-statistical CCA is presented in Chapter III. In this chapter FS and frequency recognition procedures are also addressed with traditional CCA.

Outcome of the thesis is discussed in Chapter IV. Features are selected on the basis of correlation maximization and accuracy is tested with NN. Frequency is recognized with three CCA networks at three harmonic situations from original EEG features as well as from selected features. All of the above are analyzed in this chapter V with their comparative analysis.

Detail summarization of the thesis work is illustrated in chapter V. A comprehensive suggestion for future works is also presented in this chapter.

CHAPTER II

Background and Literature Review

The aim of this chapter is to show the necessity of recognizing frequency from a signal to understand about human-brain behavior. In this perspective we evaluate the behavior of EEG signal, collection of EEG data and critical literature review on frequency recognition from both original EEG signals and selected features.

EEG is a neurological test that uses an electronic monitoring device to measure and record electrical activity in the brain. The human brain is obviously a complex system and exhibits rich spatiotemporal dynamics. Among the noninvasive techniques for probing human brain dynamics, it provides a direct measure of cortical activity with millisecond temporal resolution. It measures voltage fluctuations resulting from ionic current flows within the neurons of the brain [38]. Traditional EEG tracing is now interpreted in much the same way as it was done 50 years ago. More channels are used now and much more is known about the clinical implication of the waves, but the basic EEG display and the quantification of the waves is quite similar to those of their predecessors of a half century ago. There is no taxonomy of EEG patterns that delineates the correspondence between those patterns and brain activity. The clinical interpretation of EEG records is made by a complex process of visual pattern recognition and association on the part of the clinician and significantly more often in the last years (with the introduction of the personal computers) through the use of the Fourier transform. Quantitative EEG analysis as a field includes a wide variety of techniques. These are frequency analysis (Spectral analysis), topographic mapping, compressed spectral arrays, significance probability mapping and other complex analytical techniques [39-41]. Anyway, the most diffused quantitative method in clinical practice is the spectral analysis together with a visual assessment.

2.1 Electroencephalography (EEG)

EEG is the recording of electrical activity along the scalp that measures voltage fluctuations resulting from ionic current flows within the neurons of the brain. It is a tool to image the brain, while it is performing a task. This allows us to detect the location and magnitude of brain activity involved in the various types of functions. In clinical contexts, EEG refers to the recording of the brain's spontaneous electrical activity over a short period of time, usually 20–40 minutes, as recorded from multiple electrodes placed on the scalp. It allows us to view and record the changes in our brain activity during the time we are performing the task.

EEG recordings are achieved by placing electrodes of high conductivity (impedance <5000) in different locations of the head. Measures of the electric potentials can be recorded between pairs of active electrodes (bipolar recordings) or with respect to a supposed passive electrode called reference (monopole recordings). These measures are mainly performed with good mechanical and electrical contact of electrodes on the surface of the head (Scalp EEG) or by using special electrodes placed in the brain after a surgical operation (Intracranial EEG). The changes in the voltage difference between electrodes are sensed and amplified before being transmitted to a computer program to display the tracing of the voltage potential recordings.

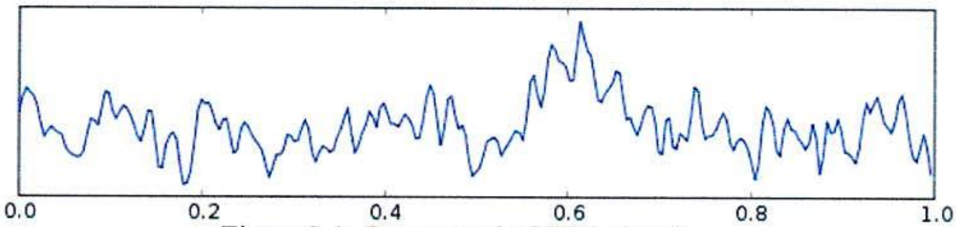


Figure 2.1: One second of EEG signals

2.1.1 Brain rhythemicities

The rhythmic activities of EEG are divided into bands by frequency. To some degree, these frequency bands are a matter of nomenclature (i.e., any rhythmic activity between 8–12 Hz can be described as "alpha"), but these designations arose because rhythmic activity within a certain frequency range was noted to have a certain distribution over the scalp or a certain biological significance. Frequency bands are usually extracted using different spectral methods. Most of the cerebral signal observed in the scalp EEG falls in the range of 1–20 Hz (activity below or above this range is likely to be artifactual, under standard clinical recording techniques).

The study of different types of rhythmicities of the brain and their relation with different pathologies and functions keep the attention of researchers since the beginnings of EEG. Brain osillations were divided in frequency bands that have been related with different brain states, functions or pathologies [42]. Beta activity is closely linked to motor behavior and is generally attenuated during active movements [43]. It is seen usually on both sides in symmetrical distribution and is most evident frontally. Low amplitude beta with multiple and varying frequencies is often associated with active, busy or anxious thinking and active concentration. Rhythmic beta with a dominant set of frequencies is associated with various pathologies and drug effects. It may be absent or reduced in areas of cortical damage. It is the dominant rhythm in patients who are alert or anxious or who have their eyes open. It is best defined in central and frontal locations that have less amplitude than alpha waves and enhanced upon expectancy states or tension. It is traditionally subdivided in β_1 and β_2 oscillations.

Alpha waves are neural oscillations in the frequency range of 8–13 Hz arising from synchronous and coherent (in phase or constructive) electrical activity of thalamic pacemaker cells in humans. They appear spontaneously in normal adults during wakefulness, under relaxation and mental inactivity conditions. They are best seen with eyes closed and most pronounced in occipital locations. They are also called Berger's wave in memory of the founder of EEG. They are reduced with open eyes, drowsiness and sleep. Historically, they were thought to represent the activity of the visual cortex in an idle state. More recent papers have argued that they inhibit areas of the cortex not in use, or alternatively that they play an active role in network coordination and communication [44]. Occipital alpha waves during periods of eyes closed are the strongest EEG brain signals. In addition, there are other normal alpha rhythms such as the mu rhythm (alpha activity in the contralateral sensory and motor cortical areas that emerges when the hands and arms are idle; and the third rhythm (alpha activity in the temporal or frontal lobes) [45]. Alpha can be abnormal; for example, an EEG that has diffuse alpha occurring in coma and is not responsive to external stimuli is referred to as alpha coma.

Theta activity refers to EEG activity within the 4-8 Hz range, prominently seen during sleep and play an important role in infancy and childhood. It is seen normally in young children and drowsiness or arousal in older children and adults; it can also be seen in meditation [46]. In the awake adult, high theta activity is considered abnormal and it is related with different brain disorders. It can be seen as a focal disturbance in focal subcortical lesions, generalized distribution in diffuse disorder or metabolic encephalopathy or deep midline disorders or some instances of hydrocephalus. On the contrary this range has been associated with reports of relaxed, meditative, and creative states. Delta oscillations reflect low-frequency activity upto 4 Hz, typically associated with sleep in healthy humans and neurological pathology. In adults, delta power has been shown to increase in proximity of brain lesions [47] and tumors [48], during anesthesia and during sleep. It may occur focally with subcortical lesions and in general distribution with diffuse lesions, metabolic encephalopathy hydrocephalus or deep midline lesions. It is usually most prominent frontally in adults (e.g. FIRDA - Frontal Intermittent Rhythmic Delta) and posteriorly in children (e.g. OIRDA - Occipital Intermittent Rhythmic Delta).

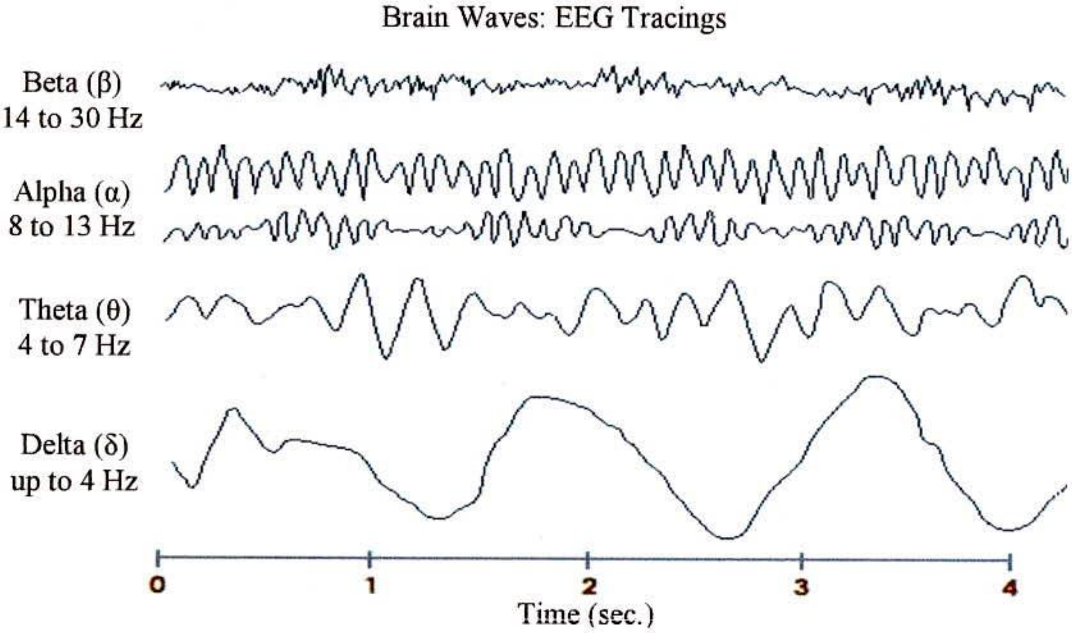


Figure 2.2: Different types of brain rhythmecities related to frequency band

Gamma oscillations have been associated with attention, arousal, object recognition, top-down modulation of sensory processes and in some cases, perceptual binding [49]. They are thought to represent binding of different populations of neurons together into a network for the purpose of carrying out a certain cognitive or motor function.

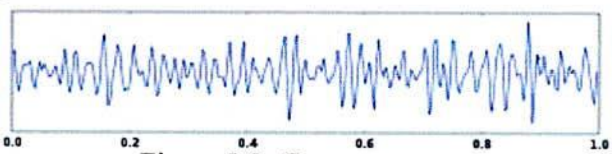


Figure 2.3: Gamma wave

Some features of the EEG are transient rather than rhythmic. Spikes and sharp waves may represent seizure activity or interictal activity in individuals with epilepsy or a predisposition toward epilepsy. Other transient features are normal: vertex waves and sleep spindles are seen in normal sleep. The normal Electroencephalography (EEG) varies by age. The neonatal EEG is quite different from the adult EEG. The EEG in childhood generally has slower frequency oscillations than the adult EEG.

2.1.2 EEG Potentials

EEG potentials are generated typically in microvolt range when electrodes are placed on the scalp, that is present in a spontaneous way or can be generated as a response to an external stimulation (e.g. tone or light flash) or internal stimulation (e.g. omission of an expected stimulus). The alteration of the ongoing EEG due to these stimuli is called event related potential (ERP), in the case of external stimulation also called evoked potential (EP). An ERP is the measured brain response that is the direct result of a specific sensory, cognitive or motor event [50]. More formally, it is any stereotyped electrophysiological response to a stimulus. The study of the brain in this way provides a noninvasive means of evaluating brain functioning in patients with cognitive diseases. Evoked potentials studies measure electrical activity in the brain in response to stimulation of sight, sound or touch. Stimuli delivered to the brain through each of these senses evoke minute electrical signals. These signals travel along the nerves and through the spinal cord to specific regions of the brain and are picked up by electrodes, amplified and displayed for a doctor to interpret.

There are mainly three modalities of stimulation [51]:

a) **Auditory:** These stimuli are single tones of a determined frequency or clicks with a broad band frequency distribution. This test can diagnose hearing ability and can indicate the presence of brain stem tumors and multiple sclerosis. Electrodes are placed on the scalp and earlobes. Auditory stimuli, such as clicking noises and tones are delivered to one ear.

b) **Visual:** This test can diagnose problems with the optic nerves that affect sight. Electrodes are placed along the scalp. The patient is asked to watch a checkerboard pattern flash for several minutes on a screen and the electrical responses in the brain are recorded. These stimuli are produced by a single light or by the reversal of a pattern as for example a checkerboard.

c) **Somatosensory:** These stimuli are elicited by electrical stimulation of peripheral nerves. This test can detect problems with the spinal cord as well as numbness and weakness of the extremities. For this test, electrodes are attached to the wrist, the back of the knee or other locations. A mild electrical stimulus is applied through the electrodes. Electrodes on the scalp then determine the amount of time it takes for the current to travel along the nerve to the brain.

The first two modalities can be combined in what is called bimodal stimulation. Current brain signal processing approaches distinguish between 'spontaneous' and 'evoked' EEG. Spontaneous EEG refers to the measurement of continuous brain waves, including the delta (up to 4Hz), theta (4-8Hz), alpha (8-12Hz), beta (12-30Hz) and gamma (30-100+Hz) waves, while evoked EEG represents brain potentials with limited duration which are recorded in response to specific stimuli, such as visual, auditory, somatosensory or olfactory. But there have also different paradigms that can be based on both spontaneous and evoked brain signals such as motor imagery BCI (brain computer interface) using modulation of spontaneous 'mu' and 'beta' waves or SSVEP-BCI (steady-state visual evoked potentials) using periodically evoked visual responses.

Steady-State Visual Evoked Potentials (SSVEP) is brain responses that are precisely synchronized with fast repetitive external visual stimulation such as flashes, reversing patterns or luminance-modulated images. In neurology, SSVEP are signals that are natural responses to visual stimulation at specific frequencies. When the retina is excited by a visual stimulus ranging from 3.5 Hz to 75 Hz [52], the brain generates electrical activity at the same (or multiples of) frequency of the visual stimulus. This responses can be measured within narrow frequency bands (such as ±0.1 Hz) around the visual stimulation frequency or using other signal processing methods that exploit the specific characteristics of the SSVEP signal such as rhythmicity and synchronization. The strongest responses occur in the primary visual (striate) cortex, although other brain areas are also activated in varying degrees.

While part of the differences in the reported frequency dominance may be due to the diversity of the captured brain processes and areas depending on the recording modality, SSVEP stimulus parameters such as spatial frequency, luminance, contrast and color also play a crucial role. Regan showed that patterned checkerboard stimuli with small checks (such as 0.2° arc side) exhibit low-frequency preferences with response peaks at ~7 Hz, while patterns with larger checks (such as 0.7° arc side) have a higher frequency preference, similarly to un-patterned flicker stimuli [53].

2.2 Data Collection

EEG data is collected from SSVEP database [37], where a single small reversing checkerboard was displayed in the middle of a black screen. Three separate reversal frequencies were used sequentially (8 Hz, 14 Hz and 28 Hz) in order to cover different components in the brain frequency response. Five trial repetitions were used for each frequency. Each trial consisted of 5s baseline rest (black screen) and 15s stimulation [54].

2.2.1 Experimental Orientations

Four healthy subjects participated in both studies. The average age of the group was 38.2 ± 2.4 years. All subjects had normal or corrected-to-normal vision. The participants were fully informed of the procedures in advance. In preparation for the experiments, each subject was screened for history of epilepsy and photosensitivity, and signed an informed consent form including a statement that she/he had no known neurological disorders. In addition, before each experiment the subjects were shown a brief stimulus sequence with increasing frequency in order to test for photosensitive epilepsy and to further decrease the probability of seizure. Subjects were seated 0.9m from a 21" CRT computer display operated at a high vertical refresh rate (setting 170 Hz, measured −168 ± 0.4 Hz). SSVEP stimulation was achieved using small reversing black and white checkerboards with 6 x 6 checks. Each check was 0.3° arc in size so that the diameter of the pattern was 2.5° arc which is slightly larger than the approximate size of the fovea. The stimulus luminance for the white checks and for the black ones (Michelson contrast of 99.2 %) were $12.5 cd/m^2$ and $0.05 cd/m^2$ respectively. Each pattern included a small red fixation point in its center and subjects were instructed to position their gaze on that point. The actual light stimulation which was emitted by the display and reached the eyes was verified using a small photosensitive semiconductor sensor. A chin rest was used by all subjects to prevent excessive contamination of the EEG data with EMG artifacts due to upper body muscle movements.

2.2.2 EEG Data Acquisition – EEG system

Brain signal acquisition was performed using a BIOSEMI EEG system with sintered Ag/AgCl active electrodes. Active electrodes contain miniature electronics to allow substantially higher EEG signal-to-noise ratio and better sensitivity to weak brain signals. Two additional electrodes, the passive Driven Right Leg (DRL) electrode and the active Common Mode Sense (CMS) electrode [55], both located just posterior to the vertex, were used to determine the common mode voltage of the Biosemi EEG system against which all other electrode measurements were recorded. This active electrode arrangement replaced the traditional reference electrode(s) used by previous passive EEG systems.

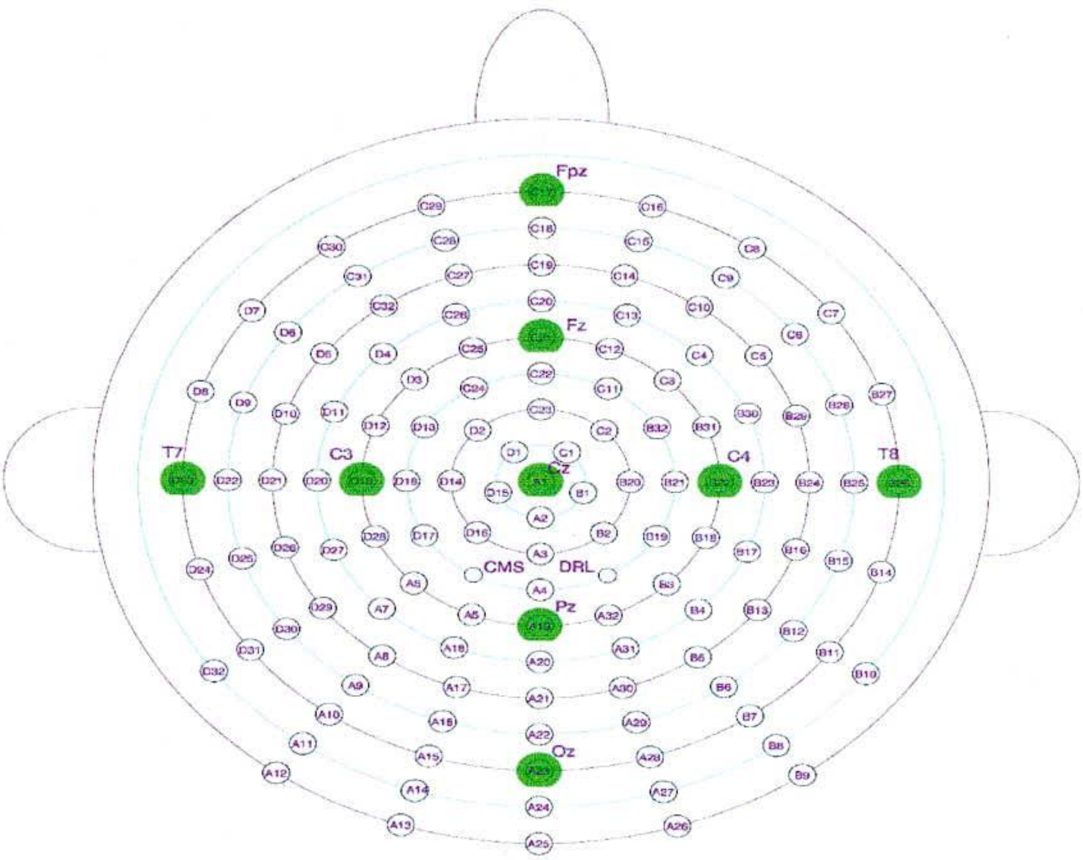


Figure 2.4: Position of 128 electrodes in head to record EEG signal

2.2.3 Characteristics of data

In this study, the experiments were performed with a 128-channel whole-head configuration, using the highest available sampling rate of 2,048 Hz. These 128 channels are used as 128 patterns of collected SSVEP data. When these data are collected for 15s stimulation with 5s baseline rest, more than 6,330 sampling points or attributes are found for a single trial as shown in Table 2.1.

Table 2.1: Number of features that's are collected SSVEP data; S1, S2, S3 & S4 indicate subjects 1 to 4 respectively

Stimulus Frequency		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
8 Hz	S1	6330	6361	6361	6330	6361
	S2	6361	6361	6330	6361	6393
	S3	6330	6361	6361	6330	6330
	S4	6361	6330	6330	6330	6361
14 Hz	S1	6343	6361	6379	6379	6398
	S2	6367	6379	6398	6382	6343
	S3	6361	6361	6343	6361	6379
	S4	6361	6379	6379	6389	6388
28 Hz	S1	6370	6379	6379	6389	6388
	S2	6379	6370	6407	6407	6379
	S3	6352	6388	6379	6389	6379
	S4	6389	6379	6379	6379	6388

These high dimensional attributes are found as an EEG wave that's collected with 128 active electrodes. A single trial SSVEP response is shown in Fig. 2.5. It is seen that trial to trial variability of a specific subject is almost negligible. In this regard, five trial of a specific subject are concatenated together to recognize frequency. It is realize that there are more than 30,000 attributes with 128 patterns for a single subject at a specific stimulus frequency. It takes longer time to execute a program and also degrade the generalization performance. In this sense, FS is performed prior to frequency recognition.

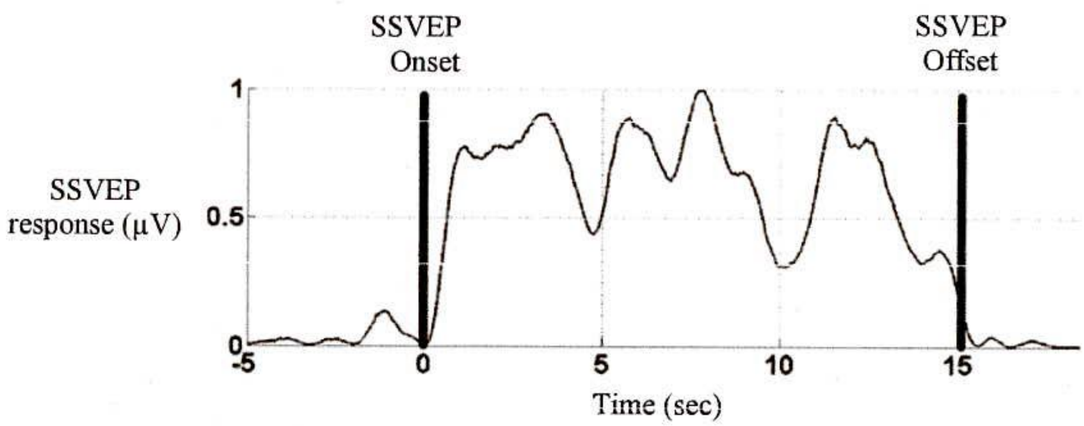


Figure 2.5: Single-trial SSVEP response with 128 Bio-semi active electrodes.

In this thesis frequency components of high dimensional SSVEP data are extracted with low computational cost using neuro-statistical method. Visual stimulations are used to flicker the checkerboard and at the same time EEG signal are recorded from brain with 128 active electrodes. Then frequency components of these EEG signals are extracted with neuro-statistical method. The entire procedures are depicted in Fig. 2.6 schematically.

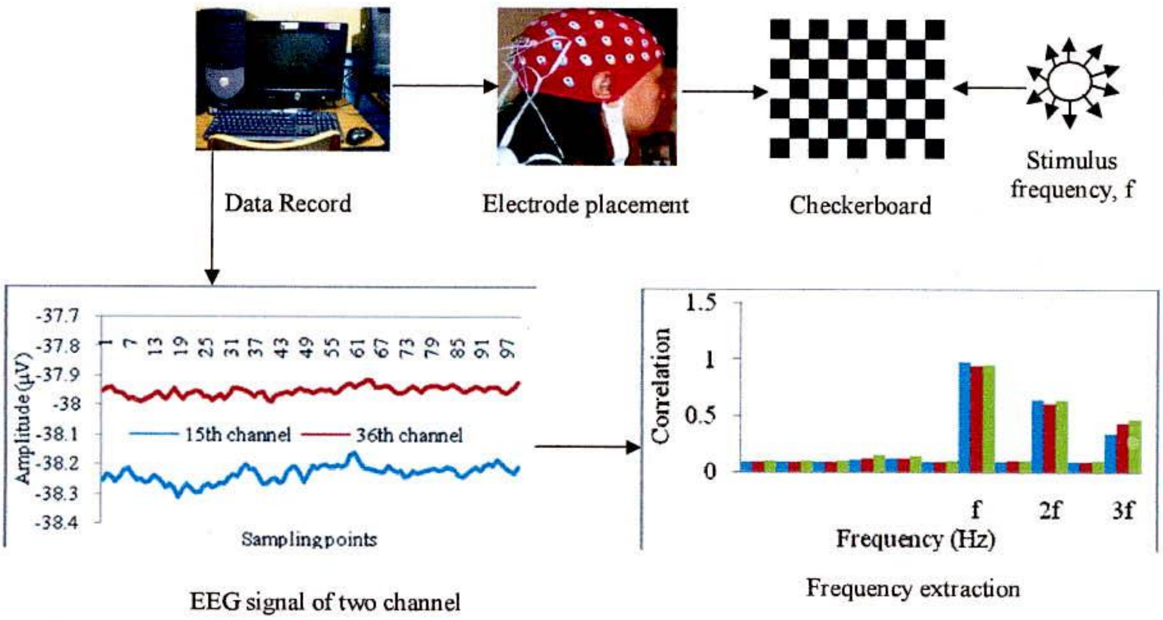


Figure 2.6: Schematic diagram to extract frequency from recorded SSVEP.

2.3 Aspects of Frequency Recognition

It has been well over a century since it was discovered that the mammalian brain generates a small but measurable electrical signal. The EEG of small animals was measured by Caton in 1875 and in man by Berger in 1925. It had been thought by the mathematician Wiener, among others, that generalized harmonic analysis would provide the mathematical tools necessary to penetrate the mysterious relations between the EEG time series and the functioning of the brain. The progress along this path has been slow however and the understanding and interpretation of EEG's remain quite elusive.

A recent approach to the problem of the quantification of EEG series has been presented by nonlinear dynamics [56-62]. The morphology and topography of sharp transients have been correlated with seizure type and therapeutic response to different medications and surgery. An essential component of the traditional visual interpretation of the clinical EEG is the characterization of infrequent, morphologically variable transient events, especially those associated with the epilepsies (spikes, spikes and waves etc.) accordingly, a great deal of energy has been spent over the years in efforts to search automatically long recordings for these phenomena and epileptiform transient detection, but with different results [39-40]. Anyway, the most diffused quantitative method in clinical practice is the spectral analysis together with a visual assessment [41].

According to frequency domain the EEG activities can be divided into three different categories: (i) spontaneous non-paroxysmal or background, (ii) spontaneous paroxysmal activity and (iii) activity evoked by external sensory stimulation. Consequently, it is quite obvious that in the frequency domain representation, rhythmic components are relatively enhanced at corresponding frequencies, whereas transients (epileptic spikes, isolated paroxysm, etc.) are smeared over the spectrum and therefore are no longer recognizable. From this, it follows that the principle field of spectral analysis is the background activity, which means the first category mentioned above, whereas in the other two categories there exist only special cases to which standard spectral analysis can be successfully applied [39].

The methods mentioned above are applied to the activity analysis in a single channel, independently of the activity of the other channels. The most common methods of studying interactions between two channels are cross correlation and the cross spectral analysis [63]. The average amount of mutual information and nonlinear correlation are recently developed methods [64-66]. These methods all try to determine whether two channels have a common activity and often whether one channel contains activity induced by the activity in the other channel. Clearly providing causality is extremely difficult, but it can sometimes be inferred by measuring time differences.

From another point of view, an EEG may be considered as a time series measured on a dynamical system that represents brain activity. This subject has caught the attention of several researchers in this field, having found the important feature that an attractor [60]. The treatment of EEG series using the approach of nonlinear dynamic systems has opened new possibilities for knowledge of the brain dynamic. However, the aims are not limited to this, but also to obtaining new forms to quantify differences in the EEG series that have some kind of clinical application. The metric approach usually employed in nonlinear dynamical analysis is based on distances and assumes the stationary of the data sets. Distances between points in appropriate embeddings of the data are used to compute a set of metric properties. These quantities are difficult to compute, require large data sets and degrade rapidly with additive noise [67].

Mayer-Kress and Layne [57] used the reconstruction techniques in the time series of the EEG to obtain their phase portrait. These diagrams suggest chaotic attractors with divergent trajectories and therefore EEG series seem non stationary. This means that the average position of a series defined over some interval changes in another. Layne, Mayer-Kress and Holzfuss [56] conclude that the EEG series are non stationary and present high dimensionality, in which case the concepts of attractor and fractal dimension would not be applied because these are asymptotic or stationary properties of dynamical system. However, Babloyantz and Destexhe [60] focused their attention on the fact that this non stationary is strictly true for awaking states but could be different for states of the sleep cycles or for patients with certain pathologies.

This problem has not been well studied and it has brought about a great variety of results exposed by different authors [58, 60]. Due to great extension of EEG series that is necessary for nonlinear metric treatment (satisfying the entire mathematical hypothesis) a criterion is almost impossible to satisfy in practice. Consequently, if the time series are non stationary, the metric algorithms must not be used. Statistical tests of stationary EEG have revealed a variety of results depending on conditions, with estimates of the amount of time during which the EEG is stationary ranging from several seconds to several minutes [68-70]. However as a practical matter, whether or not the same data segment is considered stationary depends on the problem being studied, the type of analysis being performed and the measured (features) used to characterize the data.

The time evolution of the frequency rhythm of an EEG signal and visualized the frequency engagement during epileptic activity as well as paroxysm activities are analyzed in [70]. In these case the correaltion between the obtained frequency evolution series for the differnt channels and bands are used to obtain some knowledge about the interaction and consequently causality between channels and bands. Recently EEG based Brain-Computer Interface (BCI) uses electrical signals from the cortex to control external devices like a computer or other systems and is aimed to facilitate communication for subjects with severe motor impairments. As also reported by numerous authors [71], we use the principle of SSVEP. The SSVEP is a periodic response to a visual stimulus which has the same fundamental frequency as that of the visual stimulus as well as its harmonics. The SSVEP can be recorded from the surface of the scalp over the visual cortex.

Many characteristics of SSVEP, such as the amplitude, distribution and available frequency range, show great user variation [72]. So in many of the previous research, parameter optimization and channel selection for each subject to improve the performance of BCI have been widely adopted [71, 73]. These optimizations limit the practical applicability of the SSVEP. Moreover, the SSVEP has the same fundamental frequency as the visual stimulus as well as its harmonics. The traditional SSVEP detection techniques cannot identify the targets flickering at harmonic frequencies. Thus, stimuli with harmonic frequencies cannot be used in the previous system [74-76]. This limits the number of targets. This disadvantage looms large in a system with a monitor as the stimulus. In a monitor, the number of stimulus frequencies is limited due to the small variability of the screen refresh rate and many of the obtainable stimulus frequencies are a whole-number multiple of some others (i.e. harmonics). Thus, the previous BCI system that used PC monitors as a visual stimulator usually had only two to four targets [75]. A method which can recognize the frequency with a harmonic relationship can greatly improve the performance of the SSVEP recognition.

A multiple-channel SSVEP recognition may be able to improve these disadvantages. Recently, many methods were proposed for frequency recognition from multiple-channel EEG signals. Friman et al [76] proposed a minimum energy method (MEC) which shows many advantages such as high detection accuracy and no calibration data. A traditional and widely used method for SSVEP recognition is PSDA. PSD is estimated from the EEG signals within a TW typically by FFT and its peak is detected to recognize the target stimulus [73]. Lin et al [9] proposed the use of CCA method for multi-channel SSVEP detection and also showed highly increased detection accuracy. Linet al's method was tested in offline data and channel selection was required which indicates that CCA is a very promising method for the multi-channel SSVEP recognition. The matlab program for this statistical CCA requires very long time to execute.

In this regard, we implement CCA networks to recognize underlying frequency components of EEG signals which does not require high capacity machine and performs better than others since special NN cascade architecture is incorporated. Though SSVEPs of EEG can be processed for different visual stimulation for different trials of a subject, it comprises with a high dimensional feature set. The processing of such massive data is a great challenge for computational scientists. In order to process these kinds of data, huge computational time and resources are often required. In many pattern recognition applications, there have a large number of features those are not equally important for a specific task. In this sense, FS is performed prior to frequency recognition.

FS performance is greatly dependent on the search technique in finding the salient features from a given dataset [77]. Among different FS algorithms, most are involved with either sequential search [21] or global search technique [31]. On the contrary, the existing FS algorithms can be categorized into three ways on the basis of search strategies and evaluating the generated subset, such as, wrapper [20], filter [78] and hybrid [79]. In addition, there are several works, where various FS techniques can also be found in [18].

In solving FS, filter approaches estimate the performance of features without any learning model assumed between outputs and inputs of the data for that reason they are faster to implement. Using predefined criteria, such as, mutual information [80], feature weighing [81], principal component analysis [25], independent component analysis [82], class separability measure [83], or variable ranking [84] are used for feature selection or removal. Though it is computationally efficient, but the saliency of the selected features is insufficient, because biases of classification models are not considered. There have a significant number of filter approaches in the literature. Among them ranking (or score), distance, statistics based attempts are worth mentioning. 'Laplacian Score' selects feature by calculating its local preserving power [85]. On the other hand, 'Fisher Score' seeks a feature on the basis of highest discriminative distance between patterns of different classes [86]. They however are sensitive to noises [87]. Diffusion score is another approach which preserves the diffusion distance computed with error minimization and Markov matrix [88]. The authors applied this diffusion distance technique to select features from cartoon images only. Although it improves the recognition rate, it needs to compute Markov Matrix which is computationally very expensive.

A number of algorithms have been proposed [21] for wrapper model that use the sequential search strategy for FS. By following the SFS strategy, significant features are added sequentially to the NN during training in [20], where progresses of the process were depended on the improvement of NN performance. Kabir et al. [89] proposed an idea on the basis of SFS-based FS, where these approaches have provided the correlation information of input features to the NN classifiers during training. The least salient features are deleted in stepwise fashion during the training of NN in [90]; those are SBS based FS process. Different algorithms use different heuristic strategies for finding saliency of the individual features. For example, at a time only one feature is used in the input layer for NN training in. Also full feature set is used in the NN training scheme on different weight analysis-based heuristic techniques [91]. Each feature is temporarily deleted in training, with a cross check of NN performance. In addition, for selecting salient features Guyon et al. [92] uses a recursive feature elimination process that is a variant of

SBS. The least ranking features are deleted in each step during the training of SVMs for expected feature.

In this case, neuro-statistical (CCA network) method [33] is proposed to address several problems of previous methods mentioned above. This system mainly consists of two phases. In the first phase, entire feature set of a particular problem are divided into several groups of features to train the CCA network. Highly and lightly correlated feature subsets are then obtained by pruning a large number of features from corresponding lightly and heavily weights of CCA network respectively. Since the information is lost due to pruning in the first phase, highly and lightly correlated subset are undergone to CCA network training in the second phase. In this way informative features which carry maximum correlation and minimum redundancies are found. The joint use of NN and statistics strengthen the FS process and exhibit robust results. This approach is different from previous methods on the following aspects. (i) A number of feature groups are used instead of the traditional use of individual features. (ii) The system is noise tolerant due to presence of training. (iii) No additional search algorithms are required except CCA network. (iv) It is faster than other approaches due to the use of less number of instructions and there was no huge computational burden in the program. (v) CCA network searches global correlation among feature groups instead of traditional computation of mutual correlation between two variables.

In this thesis, neuro-statistical CCA is utilized for both FS and frequency recognition. It removes computational complexity and extra burden. It is observed that frequency can be recognized from selected features as well as from original EEG data. When FS is performed prior to frequency recognition, it reduces computational complexity, time and cost. Also harmonics frequencies and differentiations among subjects can easily be identified if frequencies are recognized from selected features.

CHAPTER III

The Proposed Framework: Neuro-Statistical CCA

CCA finds two bases, one for each variable and corresponding correlation that is a way of measuring the linear relationship between two multidimensional variables [93]. The dimensionality of these new bases is equal to or less than the smallest dimensionality of the two variables. The standard statistical CCA is an extension of multiple regressions, where the second set contains multiple response variables. The variables in each set are linearly combined in a low dimensional space such that the linear combinations have a maximal correlation. Where, it avoids nonlinear relationship between datasets. Incorporating NN with standard CCA (called neuro-statistical CCA) can consider nonlinearity to find correlation among two or more sets of variables [33] and remove noisy features easily because it uses training and pruning consequently. This chapter describes this framework to search important features and recognize frequencies.

3.1 Neuro-Statistical Technique

Artificial Neural Networks (ANNs) is well known for their capacity as implementations of powerful statistical transformations. Some of the first demonstrations of this came from the family of networks [94] which extract the Principal Components of the input data. These give the best (in the sense of least mean square error) linear compression of a data set. Nonlinear extensions of PCA networks have been shown to be capable of more sophisticated statistical techniques such as Exploratory Projection Pursuit [95], Factor Analysis [96] etc. On this basis, Colin [33] proposed NN with CCA called Colin's CCA network. Canonical Correlation Analysis [97] is used when there have two or more data sets which we believe have some underling correlation. Consider two sets of input data, from which we draw two input matrices as x_1 and x_2. Then we attempt to find the linear combination of the variables that gives us maximum correlation between the combinations as described in Fig. 3.1. Let

$$\mathbf{y_1} = \mathbf{w_1 x_1} = \sum_j w_{1j} x_{1j} \qquad (3.1)$$

$$\mathbf{y_2} = \mathbf{w_2 x_2} = \sum_j w_{2j} x_{2j} \qquad (3.2)$$

Where, j is the number of attributes in every pattern. Then the values of w_1 and w_2 are searching that maximize the correlation between y_1 and y_2.

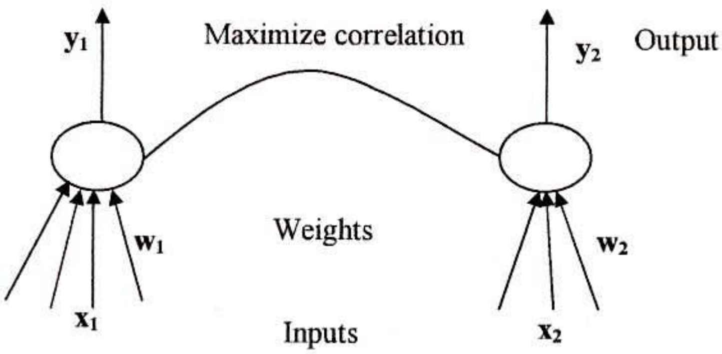


Figure 3.1: The CCA network, by adjusting weights, $\mathbf{w_1}$ and $\mathbf{w_2}$, the correlation between y_1 and y_2 is maximized.

Whereas Principal Components Analysis and Factor Analysis deal with the interrelationships within a set of variables, CCA deals with relationships between two sets of variables. If the relation between y_1 and y_2 is believed to be casual, then it may be viewed that the process as one of finding the best predictor of the set x_2 by the set x_1, and similarly of finding the most predictable criterion in the set x_2 from the x_1 data set.

The input data comprises two matrices x_1 and x_2. A complete column for a row of a particular subject is entered in the CCA network at a time as input (x_1 and x_2). In this way, every row is presented in the network sequentially. Activation is fed forward from each input to the corresponding output through the respective weights w_1 and w_2.

For input matrices x_1 and x_2, correlation maximization is done as $E(y_1y_2)$ where $E()$ denotes the expectation which will be taken over the joint distribution of x_1 and x_2. It may be regarded that this problem maximize the function as $g_1\left(\frac{w_1}{w_2}\right) = E(y_1y_2)$ which is defined to be a function of the weights, w_1 given the other set of parameters, w_2. This is an unconstrained maximization problem that has no finite solution and so we must constrain the maximization. Typically in CCA, constraint $E(y_1^2 = 1)$ is added and similarly when maximize $g_2\left(\frac{w_2}{w_1}\right)$, constraint $E(y_2^2 = 1)$ is added with y_2. Using the method of Lagrange multipliers, this yields the constrained optimization functions,

$$J_1 = E(y_1y_2) + (1/2)\lambda_1(1 - y_1^2) \tag{3.3}$$

$$J_2 = E(y_1y_2) + (1/2)\lambda_2(1 - y_2^2) \tag{3.4}$$

In this way following equation can be equivalently use

$$J = E(y_1y_2) + (1/2)\lambda_1(1 - y_1^2) + (1/2)\lambda_2(1 - y_2^2) \tag{3.5}$$

But it will be more convenient to regard these as separate criteria which can be optimized independently by implicitly assuming w_1 is constant when w_2 is changing and vice versa. Colin's wish to find the optimal solution using gradient ascent and so to find the derivative of the instantaneous version of each of these functions with respect to both the weights w_1 and w_2, and the Lagrange multipliers λ_1 and λ_2. The relative strength of the constraint, compared to the optimizing function is changed by varying the Lagrange multipliers in proportion to the derivatives of J. This allows to smoothly maximizing the function in the region, where the constraint is satisfied.

Noting that

$$\partial g_1(w_1 / w_2) / \partial w_1 = \partial(y_1y_2) / \partial w_1 = \partial(w_1x_1y_2) / \partial w_1 = x_1y_2 \tag{3.6}$$

These yields, respectively

$$\partial J_1 / \partial w_1 = x_1y_2 - \lambda_1 y_1 x_1 = x_1(y_2 - \lambda_1 y_1) \tag{3.7}$$

$$\partial J_1 / \partial w_1 \infty (1 - y_1^{\,2}) \tag{3.8}$$

Similarly, with the J_2 function $\mathbf{w_2}$ and λ_2. This gives a method of changing the weights and Lagrange multipliers on an on line basis. Finally the following joint learning rules are used for linear correlation.

$$\Delta w_{1j} = \eta x_{1j}(y_2 - \lambda_1 y_1) \tag{3.9}$$
$$\Delta\lambda_1 = \eta_0(1 - y_1^2) \tag{3.10}$$
$$\Delta w_{2j} = \eta x_{2j}(y_3 - \lambda_2 y_2) \tag{3.11}$$
$$\Delta\lambda_2 = \eta_0(1 - y_2^2) \tag{3.12}$$

Where λ_1 and λ_2 are Lagrange multipliers, w_{1j} is the j^{th} element of weight vector, $\mathbf{w}_1$, etc. It has been found empirically that best results are achieved when $\eta_0 \gg \eta$.

3.2 Traditional CCA

Canonical analysis is one of a family of correlation techniques. Generally there are three basic types of correlation those measures relationship among data, such as

a) Product Moment Correlation (PMC) measures the relationship between the observed values of two variables.

b) Multiple Regression Analysis (MRA) measures the relationships between the observed values of one variable, and the observed values of a set of variables.

c) Canonical Correlation Analysis (CCA) measures the relationships between the observed values of two sets of variables.

CCA is a way of measuring the linear relationship between two multidimensional variables [98]. It finds two bases, one for each variable, that are optimal with respect to correlations and, at the same time, it finds the corresponding correlations. In other words, it finds the two bases in which the correlation matrix between the variables is diagonal and the correlations on the diagonal are maximized. The dimensionality of these new bases is equal to or less than the smallest dimensionality of the two variables.

CCA can also be defined as the problem of finding two sets of basis vectors, one for $\boldsymbol{x}$ and the other for $\boldsymbol{y}$, such that the correlations between the projections of the variables onto these basis vectors are mutually maximized. Consider the linear combinations $\boldsymbol{x} = \boldsymbol{x}^T\widehat{\boldsymbol{w}}_x$ and $\mathbf{y} = \boldsymbol{y}^T\widehat{\boldsymbol{w}}_y$ of the two variables respectively. This means that the function to be maximized is

$$\rho = \frac{E[xy]}{\sqrt{E[x^2]E[y^2]}} = \frac{E[\widehat{\boldsymbol{w}}_x^T\boldsymbol{x}\boldsymbol{y}^T\widehat{\boldsymbol{w}}_y]}{\sqrt{E[\widehat{\boldsymbol{w}}_x^T\boldsymbol{x}\boldsymbol{x}^T\widehat{\boldsymbol{w}}_x]E[\widehat{\boldsymbol{w}}_y^T\boldsymbol{y}\boldsymbol{y}^T\widehat{\boldsymbol{w}}_y]}} = \frac{\boldsymbol{w}_x^T\boldsymbol{C}_{xy}\boldsymbol{w}_y}{\sqrt{\boldsymbol{w}_x^T\boldsymbol{C}_{xx}\boldsymbol{w}_x\boldsymbol{w}_y^T\boldsymbol{C}_{yy}\boldsymbol{w}_y}} \tag{3.13}$$

The maximum of ρ with respect to $\boldsymbol{w}_x$ and $\boldsymbol{w}_y$ is the maximum canonical correlation. The subsequent canonical correlations are uncorrelated for different solutions, i.e.

$$\begin{cases} E[x_i x_j] = E[\boldsymbol{w}_{xi}^T \boldsymbol{x}\boldsymbol{x}^T w_{xj}] = \boldsymbol{w}_{xi}^T \boldsymbol{C}_{xx} \boldsymbol{w}_{xj} = 0 \\ E[y_i y_j] = E[\boldsymbol{w}_{yi}^T \boldsymbol{x}\boldsymbol{x}^T w_{yj}] = \boldsymbol{w}_{yi}^T \boldsymbol{C}_{yy} \boldsymbol{w}_{yj} = 0 \\ E[x_i y_j] = E[\boldsymbol{w}_{xi}^T \boldsymbol{x}\boldsymbol{y}^T w_{yj}] = \boldsymbol{w}_{xi}^T \boldsymbol{C}_{xy} \boldsymbol{w}_{yj} = 0 \end{cases} \quad \text{for} \quad i \neq j \tag{3.14}$$

The projections onto $\boldsymbol{w_x}$ and $\boldsymbol{w_y}$ i.e. x and y are called canonical variates.

Calculating canonical correlations

Consider two random variables $\boldsymbol{x}$ and $\boldsymbol{y}$ with zero mean. The total covariance matrix is a block matrix where $\boldsymbol{C}_{xx}$ and $\boldsymbol{C}_{yy}$ are the within-sets covariance matrices of $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively and $\boldsymbol{C}_{xy} = C_{yx}^T$ is the between-sets covariance matrix.

$$C = \begin{bmatrix} \boldsymbol{C}_{xx} & \boldsymbol{C}xy \\ \boldsymbol{C}_{yx} & \boldsymbol{C}_{yy} \end{bmatrix} = E\left[\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix}^T\right] \tag{3.15}$$

The canonical correlations between $\boldsymbol{x}$ and $\boldsymbol{y}$ can be found by solving the Eigen value equations

$$\begin{cases} \boldsymbol{C}_{xx}^{-1} \boldsymbol{C}_{xy} \boldsymbol{C}_{yy}^{-1} \boldsymbol{C}_{yx} \widehat{\boldsymbol{w}}_x = \rho^2 \widehat{\boldsymbol{w}}_x \\ \boldsymbol{C}_{yy}^{-1} \boldsymbol{C}_{yx} \boldsymbol{C}_{xx}^{-1} \boldsymbol{C}_{xy} \widehat{\boldsymbol{w}}_y = \rho^2 \widehat{\boldsymbol{w}}_y \end{cases} \tag{3.16}$$

Where the Eigen values ρ^2 are the squared canonical correlations and the eigenvectors $\widehat{\boldsymbol{w}}_x$ and $\widehat{\boldsymbol{w}}_y$ are the normalized canonical correlation basis vectors. The numbers of non-zero solutions to these equations are limited to the smallest dimensionality of $\boldsymbol{x}$ and $\boldsymbol{y}$. As an example, if the dimensionality of $\boldsymbol{x}$ and $\boldsymbol{y}$ is 8 and 5 respectively, the maximum number of canonical correlations is 5.

Only one of the Eigen value equations needs to be solved since the solutions are related by

$$\begin{cases} \boldsymbol{C}_{xy} \widehat{\boldsymbol{w}}_y = \rho \lambda_x \boldsymbol{C}_{xx} \widehat{\boldsymbol{w}}_x \\ \boldsymbol{C}_{yx} \widehat{\boldsymbol{w}}_x = \rho \lambda_y \boldsymbol{C}_{yy} \widehat{\boldsymbol{w}}_y \end{cases} \tag{3.17}$$

Where

$$\lambda_x = \lambda_y^{-1} = \sqrt{\frac{\widehat{\boldsymbol{w}}_y^T \boldsymbol{C}_{yy} \widehat{\boldsymbol{w}}_y}{\widehat{\boldsymbol{w}}_x^T \boldsymbol{C}_{xx} \widehat{\boldsymbol{w}}_x}} \tag{31.8}$$

3.3 The Framework

The general frequency recognition framework is exposed in Fig. 3.2. Our aim is to extract the frequency components from large EEG data. In order to make it possible we devise two stage cascade CCA. First stage determines optimized reference signal from a set of sine-cosine reference signals and a set of EEG data. Second stage determines another optimized signals from the first stage optimized signals and a new set of EEG data. Then correlation coefficient is computed from two optimized signals. This process is repeated at different stimulus frequency or time window. We determine the target frequency by maximizing correlation.

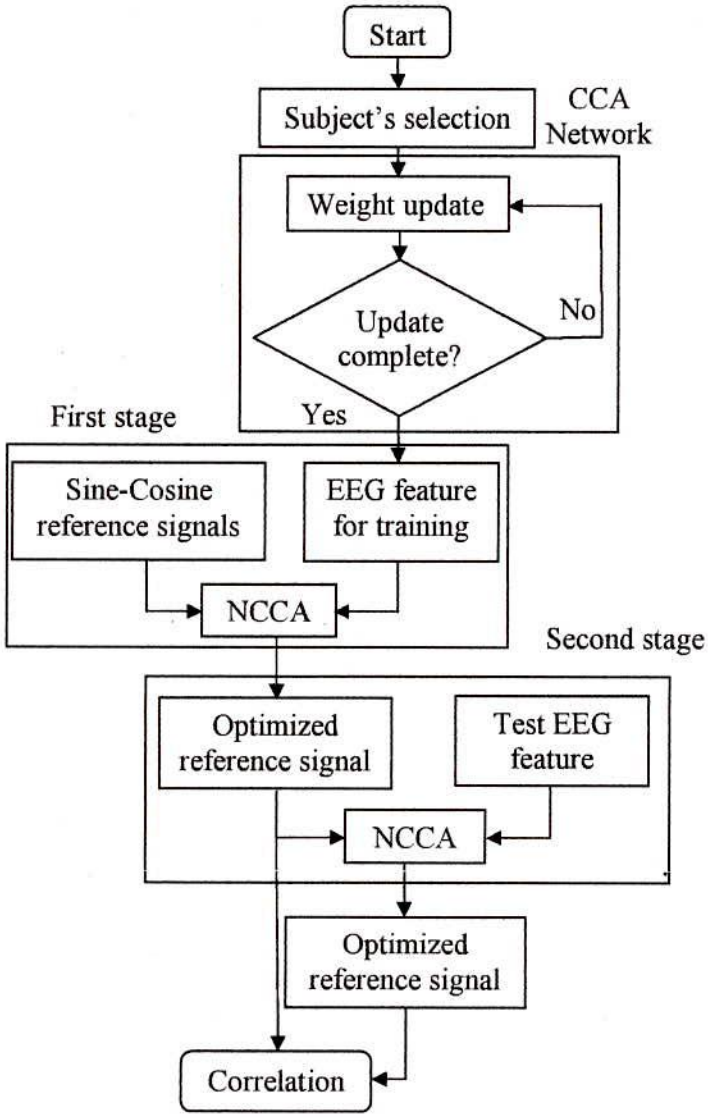


Figure 3.2: Framework of the two-stage cascaded Neuro-Statistical CCA.

3.4 Feature Selection Procedure

A major problem of much data processing is its high dimensionality. In order to process these kinds of data, huge computational time and resources are often required. In many pattern recognition applications, there have a large number of features those are not equally important for a specific task. Some of the variables may be redundant or even irrelevant. Usually by discarding such variables better performance may be achieved. In order to get informative feature subset having maximum correlation and minimum redundancies, CCA network is devised here. To explain the complete work, three subsections are made as clustering using wavelet, data preparation and FS with CCA network.

3.4.1 Clustering using wavelet

Clustering [99] is the process of organizing data or objects into groups whose members are similar in some way. It is the most important unsupervised learning problem that deals with finding a structure in a collection of unlabeled data. Cluster analysis is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups. It is a main task of exploratory data mining and a common technique for data analysis, used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics.

Cluster analysis itself is not one specific algorithm, but the general task to be solved. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Popular notions of clusters include groups with small distances among the cluster members, dense areas of the data space, intervals or particular statistical distributions. Clustering can therefore be formulated as a multi-objective optimization problem. The appropriate clustering algorithm and parameter settings depend on the individual data set and intended use of the results. Cluster analysis as such is not an automatic task, but an iterative process of knowledge discovery or interactive multi-objective optimization that involves trial and failure. It will often be necessary to modify data preprocessing and model parameters until the result achieves the desired properties. Cluster analysis was originated in anthropology by Driver and Kroeber in 1932 and introduced to psychology by Zubin in 1938 and Robert Tryon in 1939 [100] and famously used by Cattell beginning in 1943 for trait theory classification in personality psychology.

Two or more objects belong to the same cluster if they are close according to a given geometrical distance called distance-based clustering. Another kind of clustering is conceptual clustering where two or more objects belong to the same cluster if this one defines a concept common to all that objects. In other words, objects are grouped according to their fit to descriptive concepts, not according to simple similarity measures.

Though there have so many clustering algorithms, we use wavelet for clustering EEG signals. When a small oscillatory wave concentrates its energy in time is called a wavelet. It is a suitable tool for transient, non-stationary or time-varying signals analysis, which has ability to allow simultaneous time and frequency analysis. Clustering signals using wavelet is proposed in [101], where a classical clustering strategy is applied to a suitably chosen set of wavelet coefficients that offers a useful tool to carry out both significant noise reduction and efficient compression.

Wavelets are short wavelike functions that can be scaled and translated. These are mathematical functions that cut up data into different frequency components and then study each component with are solution matched to its scale. They have advantages over traditional Fourier methods in analyzing physical situations where the signal contains discontinuities and sharp spikes. Wavelets were developed independently in the fields of mathematics, quantum physics, electrical engineering and seismic geology. Interchanges between these fields during the last ten years have led to many new wavelet applications such as image compression, turbulence, human vision, earthquake prediction and radar.

Wavelet transforms take any signal and express it in terms of scaled and translated wavelets. The resulting wavelet transform is a representation of the signal at different scales. Many time series in geophysics exhibit non-stationary in their statistics. While the series may contain dominant periodic signals, these signals can vary in both amplitude and frequency over long periods of time.

A 1-D multi-signal is a set of 1-D signals of same length stored as a matrix organized row-wise or column-wise. In this process we decompose EEG signal using wavelet decomposition (WD) for clustering according to row. WD includes multiple bases and different basis will result in different classification performance and covers the shortage of fixed time–frequency decomposition in discrete wavelet transform [102]. The WD splits the original signal into two sub spaces as V and W. Complementary to each other, with V being the space that includes the low frequency information about the original signal and W includes the high frequency information. The decomposition of the low frequency subspace V was repeated and WD only partitions the frequency axis finely toward low frequency.

It is well known to any scientist and engineer who work with a real world data that signals do not exist without noise, which may be negligible (i.e. high SNR) under certain conditions. However, there are many cases in which the noise corrupts the signals in a significant manner and it must be removed from the data in order to proceed with further data analysis. The process of noise removal is generally referred to as signal de-noising or simply de-noising.

De-noising and compressing are two of the main applications of wavelets, often used as a preprocessing step before clustering. Threshold is a technique used for signal de-noising. Wavelet Transform has emerged as a powerful mathematical tool for signal compression. When we decompose a signal using the wavelet transform, we are left with a set of wavelet coefficients that correlates to the high frequency sub bands. These high frequency sub bands consist of the details in the data set. If these details are small enough, they might be omitted without substantially affecting the main features of the data set. Additionally, these small details are often those associated with noise; therefore, by setting these coefficients to zero, we are essentially killing the noise. This becomes the basic concept behind threshold set all frequency sub band coefficients that are less than a particular threshold to zero and use these coefficients in an inverse wavelet transformation to reconstruct the data set.

Multi-signal 1-D wavelet decomposition is used for EEG signal analysis in matlab environment. We exploit 7 levels near symmetric wavelet according to rows (128 channels). Universal threshold is utilized for signals de-noising, where level dependent estimation of level noise is used for rescaling. Signals are compressed using energy ratio with threshold '99'. Then three clusters are made using cell array that contains the list of EEG data to classify, also signals with coefficients of approximation at level '7' is used.

3.4.2 Data Preparation

There were four subjects for SSVEP acquisition. EEG data were collected for every subject with three different stimulus frequencies (8Hz, 14Hz and 28Hz). For that reason, there are total three datasets for one subject and total of 12 datasets for four subjects. For every subject, there have more that 31500 attributes for 128 patterns. Class information was fixed for every patterns found by wavelet. The class information of subject *i* are concatenated with EEG data of subject *i*, where *i=1, 2, 3* and *4* respectively. Then for one set of EEG data with size of $\boldsymbol{S} \times \boldsymbol{F}$, three subsections with sizes of $\boldsymbol{S} \times \boldsymbol{k}$, $\boldsymbol{S} \times \boldsymbol{m}$ and $\boldsymbol{S} \times \boldsymbol{n}$ are created manually. For first subsection $\boldsymbol{k}$ =1 to 10,500 columns, $\boldsymbol{m}$ =10,501 to 21,000 columns for second subsection and $\boldsymbol{n}$ =21,001 columns to rest are used for third subsection.

3.4.3 FS with CCA Network

The Colin's CCA network is implemented for selecting features according to weight maximization. It takes the advantages of correlation maximization and pruning process. Firstly whole dataset is subdivided into three subsections and they are fed sample by sample into the CCA network. Then highly correlated features are selected by discarding minimum weight respective attributes from original dataset. For the sake of simplicity a brief description of feature selection process is presented here.

Step 1. Consider entire dataset of size $\boldsymbol{S} \times \boldsymbol{F}$, where $\boldsymbol{S}$ is the number of samples and $\boldsymbol{F}$ is the number of features.

Step 2. Three multivariate datasets $\boldsymbol{x_1}$, $\boldsymbol{x_2}$ and $\boldsymbol{x_3}$ with dimensions of $\boldsymbol{S} \times \boldsymbol{k}$, $\boldsymbol{S} \times \boldsymbol{m}$ and $\boldsymbol{S} \times \boldsymbol{n}$ respectively are generated from $\boldsymbol{S} \times \boldsymbol{F}$, where $F = k + m + n$.

Step 3. Initialize Lagrange multipliers $\boldsymbol{\lambda_1}$, $\boldsymbol{\lambda_2}$, $\boldsymbol{\lambda_3}$ and learning constant η_0, η. Generate random weight $\boldsymbol{w_1}$, $\boldsymbol{w_2}$ and $\boldsymbol{w_3}$ according to the dimensions of $\boldsymbol{x_1}$, $\boldsymbol{x_2}$ and $\boldsymbol{x_3}$.

Step 4. Update weights and Lagrange multipliers according to following joint learning rules. Where, j is the number of features for every sample.

$$\mathbf{y_1} = \mathbf{w_1 x_1} = \sum_j w_{1j} x_{1j} \tag{3.1}$$
$$\mathbf{y_2} = \mathbf{w_2 x_2} = \sum_j w_{2j} x_{2j} \tag{3.2}$$
$$\mathbf{y_3} = \mathbf{w_3 x_3} = \sum_j w_{3j} x_{3j} \tag{3.19}$$
$$\Delta \mathrm{w_{1j}} = \eta \mathrm{x_{1j}} (\mathrm{y_2} - \lambda_1 \mathrm{y_1}) \tag{3.9}$$
$$\Delta \lambda_1 = \eta_0 (1 - \mathrm{y_1^2}) \tag{3.10}$$
$$\Delta \mathrm{w_{2j}} = \eta \mathrm{x_{2j}} (\mathrm{y_3} - \lambda_2 \mathrm{y_2}) \tag{3.11}$$
$$\Delta \lambda_2 = \eta_0 (1 - \mathrm{y_2^2}) \tag{3.12}$$
$$\Delta \mathrm{w_{3j}} = \eta \mathrm{x_{3j}} (\mathrm{y_1} - \lambda_3 \mathrm{y_3}) \tag{3.20}$$
$$\Delta \lambda_3 = \eta_0 (1 - \mathrm{y_3^2}) \tag{3.21}$$

Step 5. Select a $\boldsymbol{S} \times \boldsymbol{p}$ matrix from $\boldsymbol{S} \times \boldsymbol{k}$, $\boldsymbol{S} \times \boldsymbol{m}$ and $\boldsymbol{S} \times \boldsymbol{n}$ matrices by discarding $(k + m + n) - p$ features according to maximizing correlated features, where $\mathrm{p} \ll (\mathrm{k} + \mathrm{m} + \mathrm{n})$.

The CCA network is applied for EEG dataset. Though it starts with different initial random weight for every run, but it did not show any significance difference for weight maximization as well as subset selection. Also it did not show any variation with change in the value of learning constant and Lagrange multipliers. Weights were normalized for finding representative result. The FS process using CCA network is explored in Fig. 3.3.

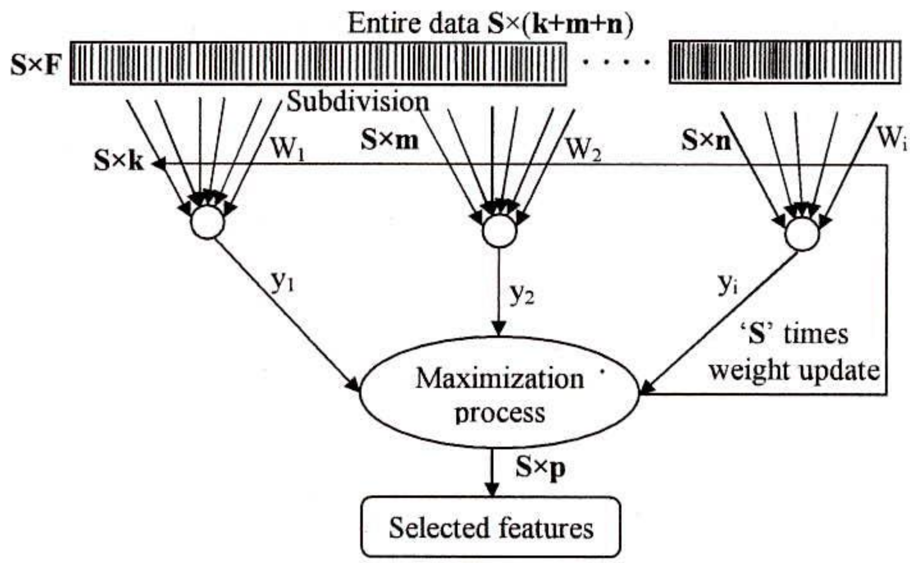


Figure 3.3: FS using CCA network

For SSVEP recognition, CCA works quite well and correlation is very important in order to assess the relationship between two time series. The joint weight update rules of eq. 3.7 to eq. 3.8 and eq. 3.15 to eq. 3.21 are applied to find the canonical correlations with different subjects. The correlations and weight maximization are made among most important features of a signal. Features are deleted those show lightly correlated weight. We want to keep only 15, 30 and 60 features from more than 31,500 features. For that reason, only 5, 10 and 20 features exist accordingly for every subset of a subject. It is done using pruning process i.e: whose correlation is lowest it is deleted firstly. In this way every samples are deleted without our expected features. Then outputs of every subset are added sequentially for select final feature of a subject. The algorithm of CCA network for selecting features is outlined in A3.1 below.

A3.1: Feature selection algorithm using CCA network

***Input**: $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$* are three *subsections of EEG data of one subject corresponding to S sampling points.*

***Output**: Correlation S_m*

***for** m=1 to S do*

Random initialization for $\mathbf{w}_1$, $\mathbf{w}_2$ and $\mathbf{w}_3$

repeat

Find $\mathbf{w}_1$, $\mathbf{w}_2$ and $\mathbf{w}_3$ which maximize correlation between y_1 and y_2, y_2 and y_3 & y_3 and y_1 by the CCA

***until** the maximum number of iteration is reached*

Compute the optimized signals y_1, y_2 and y_3

end

Compute update weight $\mathbf{w}_1$, $\mathbf{w}_2$ and $\mathbf{w}_3$ for correlation S_m

Select the features for weight maximization

3.5 Procedure for Frequency Recognition

Humen brain generated EEG signals have a certain frequency range that was noted to have a certain distribution over the scalp or a certain biological significance. To recognize these frequencies from collected SSVEP of EEG, we devise three types of CCA networks. The entire methodology for frequency recognition is examined in three subsections as reference signal, Colin's CCA network and extraction of frequency components respectively.

3.5.1 Reference Signal

Usually a signal can be represented in terms of sine-cosine waveform according to theory of Fourier transform. Therefore we consider sine-cosine as reference signals in order to determine underlying frequency components in the EEG data. The reference signals are constructed by sine-cosine waves at the m-th stimulus frequency f_m $(m = 1, 2 \ldots M)$ as follows [12]:

$$Y_m = \begin{pmatrix} \sin(2\pi f_m 1/f_s) \ldots\ldots\ldots \sin(2\pi f_m J/f_s) \\ \cos(2\pi f_m 1/f_s) \ldots\ldots\ldots \cos(2\pi f_m J/f_s) \\ \vdots \\ \sin(2\pi H f_m 1/f_s) \ldots\ldots\ldots \sin(2\pi H f_m J/f_s) \\ \cos(2\pi H f_m 1/f_s) \ldots\ldots\ldots \cos(2\pi H f_m J/f_s) \end{pmatrix} \tag{3.22}$$

Where H denotes the number of used harmonics, J is the number of sampling points and f_s represents the sampling rate. These signals are used in first stage of Neural CCA to find an optimized reference signal. Since pure sine-cosine reference signals do not contain any information on EEG data, we generate this kind of optimized reference signals with CCA network.

3.5.2 Implementation of Colin's CCA Network

CCA is a multivariable statistical method to reveal the underlying correlation between two sets of data [98]. It finds two bases, one for each variable, that are optimal with respect to correlations and at the same time, it finds the corresponding correlations. The standard statistical CCA avoids nonlinear relationship between datasets. Extension of standard CCA with NN can overcome this problem [33]. A brief description of neural CCA is presented here for the sake of simplicity. In this paper we use three kinds of neural networks - linear, nonlinear and nonlinear with feedback and they are described briefly below. Consider one set of reference signals $\mathbf{x_1}$ and one set of EEG signals or selected feature sets of EEG signals $\mathbf{x_2}$, and then we attempt to find the linear combination of the signals that gives us maximum correlation between the combinations. Let

$$\mathbf{y_1} = \mathbf{w_1}\mathbf{x_1} = \textstyle\sum_j w_{1j} x_{1j} \tag{3.1}$$
$$\mathbf{y_2} = \mathbf{w_2}\mathbf{x_2} = \textstyle\sum_j w_{2j} x_{2j} \tag{3.2}$$

Where j is the number of column in every row, there were total 128 rows for 128 electrodes. Then we wish to find those values of $\mathbf{w_1}$ and $\mathbf{w_2}$ that maximize the correlation between y_1 and y_2.

The input data comprises two vectors $\mathbf{x}_1$ and $\mathbf{x}_2$. A complete column for a row of a particular subjects/reference signals are entered in the CCA network at a time as input ($\mathbf{x}_1$, $\mathbf{x}_2$). In this way, every row is presented in the network sequentially. Activation is fed forward from each input to the corresponding output through the respective weights, $\mathbf{w}_1$ and $\mathbf{w}_2$. We use the joint learning rules for linear correlation.

$$\Delta w_{1j} = \eta x_{1j}(y_2 - \lambda_1 y_1) \tag{3.9}$$
$$\Delta\lambda_1 = \eta_0(1 - y_1^2) \tag{3.10}$$
$$\Delta w_{2j} = \eta x_{2j}(y_1 - \lambda_2 y_2) \tag{3.11}$$
$$\Delta\lambda_2 = \eta_0(1 - y_2^2) \tag{3.12}$$

Where λ_1 and λ_2 are Lagrange multipliers, w_{1j} is the j^{th} element of weight vector, $\mathbf{w}_1$, etc.

To find the nonlinear and nonlinear feedback combination of the signals that gives us maximum correlation between the reference signals set $\mathbf{x}_1$ and EEG data set $\mathbf{x}_2$, let

$$\mathbf{y}_1 = \mathbf{w}_1\mathbf{f}_1 = \sum_j w_{1j} \tanh(v_{1j}x_{1j}) \tag{3.23}$$
$$\mathbf{y}_2 = \mathbf{w}_2\mathbf{f}_2 = \sum_j w_{2j} \tanh(v_{2j}\, x_{2j}) \tag{3.24}$$

The joint learning rules for nonlinear correlation are

$$\Delta w_{1j} = \eta \mathbf{f}_1(y_2 - \lambda_1 y_1) \tag{3.25}$$
$$\Delta v_{1j} = \eta \mathbf{x}_{1j}\mathbf{w}_{1j}(y_2 - \lambda_1 y_1)(1 - \mathbf{f}_1^2) \tag{3.26}$$
$$\Delta\lambda_1 = \eta_0(1 - y_1^2) \tag{3.27}$$
$$\Delta w_{2j} = \eta \mathbf{f}_2(y_1 - \lambda_2 y_2) \tag{3.28}$$
$$\Delta v_{2j} = \eta \mathbf{x}_{2j}\mathbf{w}_{2j}(y_2 - \lambda_1 y_1)(1 - \mathbf{f}_2^2) \tag{3.29}$$
$$\Delta\lambda_2 = \eta_0(1 - y_2^2) \tag{3.30}$$

The joint learning rules for nonlinear feedback network are

$$\Delta w_{1j}(t) = \eta \mathbf{f}_1(y_2(t) - \lambda_1 y_1(t)) + 0.5y_1(t-1) \tag{3.31}$$
$$\Delta v_{1j}(t) = \eta \mathbf{x}_{1j}\mathbf{w}_{1j}(y_2(t) - \lambda_1 y_1(t))(1 - \mathbf{f}_1^2) + 0.5y_1(t-1) \tag{3.32}$$
$$\Delta w_{2j}(t) = \eta \mathbf{f}_2\big(y_1(t) - \lambda_2 y_2(t)\big) + 0.5y_2(t-1) \tag{3.33}$$
$$\Delta v_{2j}(t) = \eta \mathbf{x}_{2j}\mathbf{w}_{2j}\big(y_2(t) - \lambda_1 y_1(t)\big)(1 - \mathbf{f}_2^2) + 0.5y_2(t-1) \tag{3.34}$$

3.5.3 Extraction of Frequency Components

We here visualize how the frequency components of EEG signals are extracted using CCA network. Since there is no information of EEG signals on sine-cosine signals, we optimize reference signals from sine-cosine and EEG signals. We use four-fold cross validation i.e. three subjects consisting of 45 trials are taken for training in the first stage and remaining subject (15 trials) is used for finding final optimize reference signal in the second stage. We concatenated 45 trials for first stage and 15 trials for second stage before presenting in the neural network. In this way channel wise EEG patterns are presented in the network and updated as visualize in Fig. 3.4. We use three different types of networks such as linear, nonlinear and nonlinear with feedback. The correlation coefficient S_m which reflects the relationship between y_2 and y_4 is calculated as:

$$S_m = \sqrt{1 - \frac{||y_2 - y_4||^2}{||y_2 - E[y_4]||^2}} \tag{3.35}$$

where ||.|| denotes norm. Larger S_m implies more significant relationship between y_2 and y_4 [31]. From correlation $S_{m,}$ we find target frequency f_t by using

$$f_t = {}^{max}_{f_m} S_m \tag{3.36}$$

Where, f_m is the stimulus frequency of sine-cosine reference signal. We consider time window$(TW) = \frac{1}{f_m}$. At different f_m we have a correlation profile. The maximum correlation converges at one or more f_m value(s) which is (are) our desired frequency component(s).

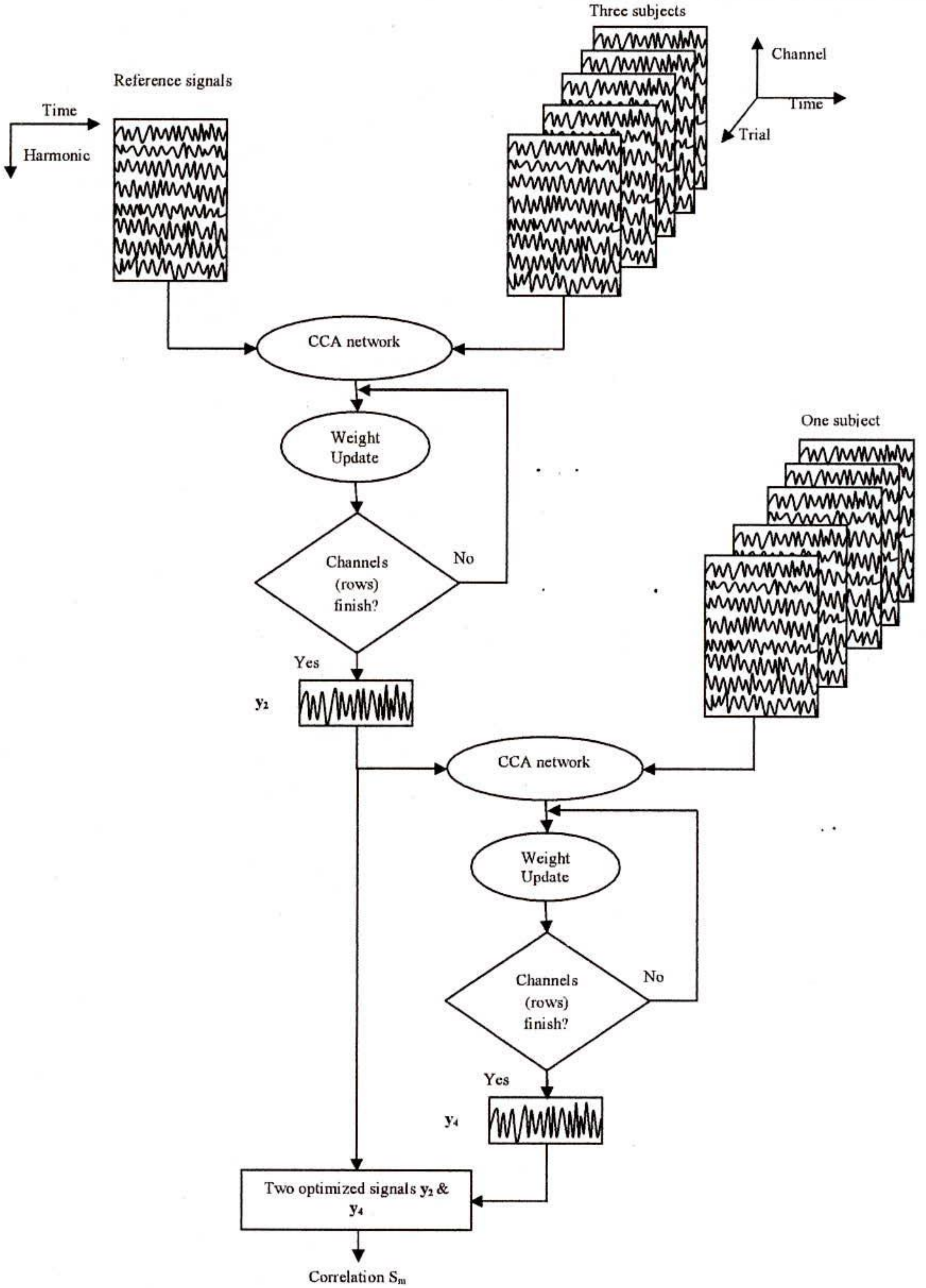


Figure 3.4: Illustration of two-stage CCA network for extraction of frequency components

The correlation is very important in order to assess the relationship between two time series. Experiments were done using time series of EEG signals of different subjects and sine-cosine reference signals with different harmonics. Lin et al. [9] introduced the CCA to recognize SSVEP for the first time. Tensor CCA is an extension of the standard CCA, which focuses on inspecting the correlation between two multiway data arrays, instead of two sets of vector-based variables [11], where a first-order tensor is a vector, a second-order tensor is a matrix and higher order is a tensor. Borrowing the idea of Multiway CCA which maximizes the correlation between multiway data (tensor) and two-way data (matrix) to optimize the reference signals used in correlation analysis for SSVEP recognition, we introduce two-stage multiway neural CCA to find frequency components of EEG data set. We consider that EEG data from the trials with a specific stimulus frequency form a third-order (three-way) data tensor, (channel × time × trial) and an original reference signal matrix (harmonic × time) is constructed by the sine-cosine signals with frequencies as the stimulus frequency and its higher harmonics. Our aim is to find more efficient reference signals for SSVEP recognition from different domains and then to find underling frequency components of EEG data set, based on the optimized reference signals of sine-cosine and multiway data arrays. A brief description of two-stage CCA network algorithm is outlined in A3.2 below.

A3.2: Algorithm for recognizing frequencies using two-stage CCA network

Input: EEG data $x_1, x_2, x_3, .x_m \in R^{IxJxK}$ and reference sine-cosine signals Y_m $(m=1,2,...M) \in R^{2HxJ}$ corresponding to M stimulus frequencies. Here I, J, K and H indicate the total number of channels, sampling points, trials and harmonics respectively.
Output: Frequency f_t

Begin loop
for *m=1 to M do*
  begin stage 1
    Input: *Training subjects (EEG data set) and reference sine-cosine signals* Y_m
    Output: *Optimized signals* y_1 *and* y_2
        Random initialization of $\mathbf{w_1}$ *and* $\mathbf{w_2}$
    for *i=1 to I do*
        repeat
            Update $\mathbf{w_1}$ *and* $\mathbf{w_2}$
        until *the maximum number of iteration is reached*
    end
        Compute the optimized signals y_1 *and* y_2
  end stage 1
  begin stage 2
    Input: *Testing subject (EEG data set) and optimized reference signal* y_2
    Output: *Optimized signals* y_3 *and* y_4
        Random initialization for weight $\mathbf{w_3}$ *and* $\mathbf{w_4}$
    for *i=1 to I do*
        repeat
            Update $\mathbf{w_3}$ *and* $\mathbf{w_4}$
        until *the maximum number of iteration is reached*
    end
        Compute the optimized signals y_3 *and* y_4
  end stage 2
    Compute correlation S_m ***from*** y_2 ***and*** y_4
    for *i= 1 to I do*
        repeat
            Compute S_m
        until *the maximum number of iteration is reached*
    end
    Recognize frequencies f_t *from maximum* S_m
End loop

CHAPTER IV

Experimental Studies

The basic objective of this chapter is to evaluate the experimental results of our proposed CCA network. Firstly evaluations of selected features are done using NN where data are subdivided into three subsections prior to FS. Then performance of CCA network to recognize frequencies is evaluated. In this case, we implemented two-stage CCA network which gives the advantage of EEG signal optimization perfectly by reference signal. Finally it is found that frequency recognition is perfect and advantageous when FS is performed prior to recognize frequency. A comparative study between these two cases and among other standard methods with our proposed method is discussed at the end of this chapter.

4.1 Experimental Setup

SSVEP are one kind of potentials collected from brain by standard EEG systems. It is related to the visual stimulation. If a checkerboard is flickered with a definite frequency and seen by any subjects, then this frequency related potentials will be adopted to the brains of corresponding subjects. It is possible to determine the corresponding frequency by analyzing SSVEP of the subjects. In this situation, CCA network is applied to the SSVEP data to extract the corresponding frequencies. Here, SSVEP data were collected for 5s with 128 standard Biosemi active electrodes. There were four subjects with five trials of each one. For this reason, there were 128 patterns with more than 30,000 features of a subject. All of the features of this high dimensional data are not similarly important. It also takes longer time to execute as well as frequency detection is not sharp. In this sense, firstly effective features are selected using CCA network and then applied to the three types of CCA network to extract the corresponding frequency. In this case, it is seen that FS is made with higher accuracy as well as frequency extraction is sharp and quick where subjects can be differentiated easily.

The performance of proposed FS algorithm and frequency recognition with CCA network has been evaluated on SSVEP based EEG datasets. Classification accuracy as well as Sum Square Errors (SSE) of selected features was measured using NN classifier. We performed the program by an Intel(R) Core(TM) i5-2450M CPU with 2.5GHz, 4.00 GB of RAM, Operating system 64-bit computer using matlab 2012. In case of FS, firstly whole attribute of a dataset is subdivided into possible equal three subdivisions with including classes of dataset. There are more than $\boldsymbol{F}$=31,500 attributes of a subject for definite stimulas frequency. Then they are subdivided as $\boldsymbol{k}$=10,500, $\boldsymbol{m}$=10,500 and $\boldsymbol{n}$=21,001 columns to rest attributes for $\boldsymbol{x_1}$, $\boldsymbol{x_2}$ and $\boldsymbol{x_3}$ respectively. This is applied for other subjects as well, where all attributes are subdivided into three subdivisions almost equally. Then random weight is generated within a certain small value those have dimensions as original subdivisions. In case of frequency recognition, two sets are made to provide to the CCA network – one set is created from reference signal and other set is obtained from original EEG subjects or from selected features.

4.2 Results of Feature Selection

Prior to FS, class information is determined for SSVEP data of every subject using wavelet. Features are selected using CCA network and results are evaluated with NN classifier. These results are evaluated into following subsections.

4.2.1 Class Determination

Wavelet is used for class determination of EEG data. For a subject, there have total of 128 attributes (rows) and more than 31500 sampling points (columns). We find three class cluster using wavelet. Seven levels, near symmetric wavelet are used for clustering and row is the direction of decomposition. There were four subjects and EEG data is collected for every subject at three different stimulus frequencies. So, there are total of twelve EEG dataset and wavelet is used for clustering every dataset differently. For understanding conveniently, clustering result of second subject whose is stimulated at 8 Hz are given below in Fig. 4.1. For further analysis we use binary 001, 010, and 100 for 1, 2, and 3 respectively. Here Sub denotes subject.

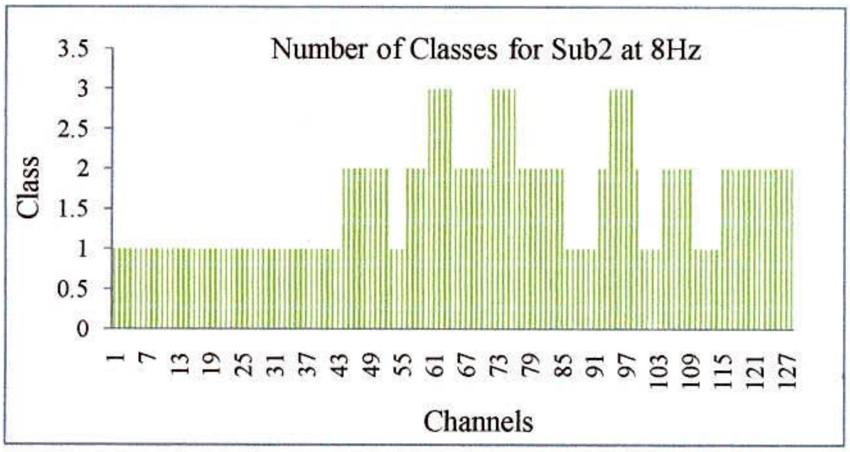


Figure 4.1: Classes of EEG data using wavelet.

4.2.2 Evaluation with Neural Network (NN)

NN is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements called neurons, working in unison to solve specific problems. NNs, like people, learn by example. A NN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. This is true of NNs as well.

In this case, NN is used for accuracy measurement of classified selected features of EEG data. The Pattern recognition tools can be implemented by using a feed-forward NN that has been trained accordingly. During training, the network is trained to associate outputs with input patterns. When the network is used, it identifies the input pattern and tries to output the associated output pattern. The power of neural networks comes to life when a pattern that has no output associated with it, is given as an input. In this case, the network gives the output that corresponds to a taught input pattern that is least different from the given pattern.

The Back-propagation (BP) algorithm is most effective learning method that use gradient descent to tune network parameters to best fit a training set of input-output pairs. The activation functions in BP, denoted by φ(x), defines the output of a neuron in terms of the net input x. Sometimes it referred as transfer function since it limits and rectifies the output signal according to its characteristics. The sigmoid function is the most common form of activation function used in construction of artificial neural network, defined by

$$\varphi(x) = \frac{1}{1+\exp(-x)} \tag{4.1}$$

The input from unit i into unit j is denoted x_{ji}, and the weight from unit i to unit j is denoted w_{ji}. For each ($\underline{x}$, $\underline{t}$), in training examples, propagate the input forward through the network [103].

Standard BP (training_examples, ή)

Each traing example is a pair of the form ($\underline{x}$, $\underline{t}$), where $\underline{x}$ is the vector of input values, and t is the target output value, ή is the learning rate.

- Initialize each w_i to some small random value
- Until the termination condition is met, Do
- Initialize each Δw_i to zero
- For each ($\underline{x}$, $\underline{t}$), in training_examples, Do
- Input the instance $\underline{x}$ to the unit and compute the output o
- For each linear unit weight w_i, Do
  $\Delta w_i \leftarrow \Delta w_i + ή(t - o)x_i$
- For each linear unit weight w_i, Do
  $w_i \leftarrow w_i + \Delta w_i$

Figure 4.2: Weight updates process of a generalized Standard BP

The BP with feed forward NN model is shown in Fig. 4.3. BP is a common method of training artificial neural networks so as to minimize the objective function.

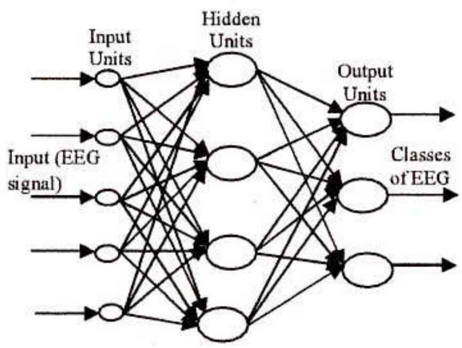


Figure 4.3: NN model for accuracy measurement of selected features.

Arthur E. Bryson and Yu-Chi Ho described it as a multi-stage dynamic system optimization method in 1969 [104]. It wasn't until 1974 and later, when applied in the context of neural networks and through the work of Paul Werbos [105]. NN generally refers to the interconnected groups of nodes in the different layers of each system, akin to the vast network of neurons in a brain. Three layers are used in this system. EEG signals are fed into the input neurons of input layer. There are 128 attributes for every (15, 30 and 60) EEG feature set. So these are fed into 128 input neurons of one input layer. Firstly 128 attributes of first feature is fed and accordingly all feature are fed to input layer one by one. These input neurons send data via synapses to the hidden layer of neurons. There are 7, 15 and 30 hidden neurons in one hidden layer for 15, 30 and 60 samples respectively and they send data via synapses to the output layer of output neurons. There are three output neurons in one output layer for three classes of EEG data. The synapses store weights to manipulate the data in the calculations. The sigmoid activation function is used here to convert a neuron's weighted input to its output activation. Each circular node in Fig. 4.3 represents an artificial neuron and an arrow represents a connection from the output of one neuron to the input of another.

NN are also used for finding sum squares errors (SSE), as shown in eq. 4.2, where 'P' denotes the numbers of patterns, 'O' denotes the numbers of output, 't' for target output and 'a' for actual output.

$$SSE = \frac{1}{2}\sum_{i=1}^{P}\sum_{j=1}^{O}\left(t_{i,j} - a_{i,j}\right)^2 \tag{4.2}$$

Subdivided EEG data is entered into the CCA network sequentially. Weight is updated for correlation maximization according to CCA networks weight update rules shown in Eq3.1 to Eq3.2 and Eq3.9 to Eq3.12 of chapter III. Then features are selected according to pruning process. According to minimum correlated weight respective features deletion maximum correlated features are selected. We accept three subsets for a subject at one stimulus frequency. We choose 15, 30 and 60 attributes for three subsets of each subject. So there are nine subsets for a subject and total of 36 subsets. These subsets are analyzed using BP rule of NN. In this way, the accuracy is measured for the selected features. In BP network 7, 15 and 30 hidden nodes are used for 15, 30 and 60 features accordingly. TABLE I shows the accuracy for every subject at different stimulus frequency for various sizes of subsets. We choose 75% (96 rows) for training set and 25% (32 rows) for testing set. We take average accuracy of five runs for every feature set. It is shown from Table 5.1 that for subject 4 at 8 Hz, classifier cannot classify nine attributes. Also classifier cannot detect one attributes for subject 1 and 2 at 14 Hz. It is for noise contamination of EEG signals at that time when data is collected from brain. But for every other sector classifier shows 100 percent accuracy. For that reason, we can say that NCCA is a great tool for feature selection. This method is suitable for converting very high dimensional data into very low feature set which may contain almost same information as original set.

Table 4.1- Accuracy (%) for different subsets of selected features. 'Acc' indicate Accuracy.

Stimulus Freuency		*No. of selected features*	*Acc (%)*	*No. of selected features*	*Acc (%)*	*No. of selected features*	*Acc (%)*
8Hz	S1	15	100	30	100	60	100
	S2		100		100		100
	S3		100		100		100
	S4		72		72		72
14Hz	S1	15	97	30	97	60	97
	S2		97		97		97
	S3		100		100		100
	S4		100		100		100
28Hz	S1	15	100	30	100	60	100
	S2		100		100		100
	S3		100		100		100
	S4		100		100		100

To find SSE, 1000 iterations are made for a selected feature as shown in eq. 4.2 of chapter IV. The error curve with 100 iterations is shown in Fig. 4.4. We explore 8Hz_Sub2_30 for 30 features of second subject at 8 Hz stimulation. In this way, other features are kept in the figure. For 1000th iteration, errors 0.001869, 0.002549 and 0.001869 are found for 8Hz_Sub2_30, 14Hz_Sub1_30 and 28Hz_Sub4_30 respectively. Thus errors are less than 0.3% which shows performance of CCA network is better.

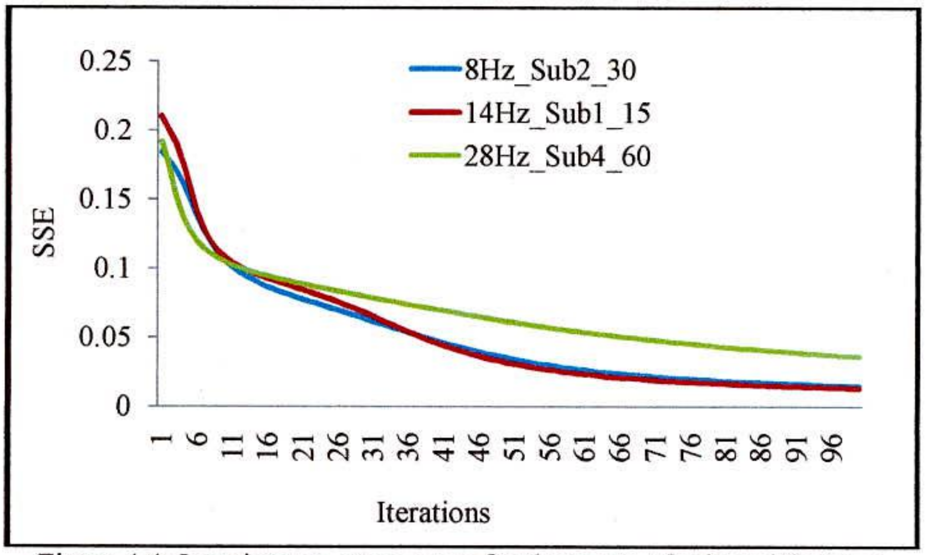


Figure 4.4: Iterations vs. error curve for three sets of selected features.

This EEG data contains brain conditions of human being. It may vary with subject's variations as well as with different stimulations. But though these are the time varying signals, they show coherency with one another on the basis of subjects and stimulations. Also, since NCCA finds out the features among the subdivision of entire EEG data via correlation maximization, there must have some kind of dependency among selected attributes. We observe this dependency with NN by varying the testing set with different subjects and stimulations. Table 4.2 shows dependency with different stimulus frequency as a confusion matrix for 15 attributes of subject 4. We have total 128 patterns, 75% (96 patterns) of them are used as training set and rest (32 patterns) are used as a testing set. It is shown from Table 4.2 that there has no coincidence when test by 14 Hz stimulation but trained by 8 Hz stimulation; because all test data are misclassified with training set. But 56.25% test dataset of 28 Hz are coinciding with trained set of 8 Hz.

Table 4.2: Confusion Matrix for 15 Features of Subject 4

		Training dataset		
		8 Hz	*14 Hz*	*28 Hz*
Test dataset	*8 Hz*	23	4	8
	14 Hz	0	32	15
	28 Hz	18	13	32

We also find recognition rate by considering different subjects as testing set. Table 4.3 shows these rates for 15 attributes of 28 Hz stimulation. In Table 4.3, we denote subject1 as S1 and in this way for others. Firstly, we train the network using 96 patterns of S1, and then we test using 32 patterns of S1, S2, S3 and S4 accordingly, which shows that 32 patterns are classified for S1, 21 for S2, 13 for S3, but all are misclassified for S4. That's why we can say that 100% of S1, 64.73% of S2, 40.63% of S3 data are coincide with S1, but no coincidence data between S4 and S1. So, we can say that though brain signals are time varying quantity, it is possible to test the coincidence between signals from selected features of CCA network. It is compatible and reliable, because highest classified are found for own dataset. Also it shows low computational time.

Table 4.3: Confusion Matrix for 15 Features of 28 Hz Stimulation

		Training dataset			
		S1	***S2***	***S3***	***S4***
Test dataset	***S1***	32	16	21	09
	S2	21	32	23	01
	S3	13	10	32	17
	S4	0	09	14	32

In our EEG datasets, there have almost above 31,500 sampling points for a subject. When such a high dimensional data are feed into statistical CCA or NN separately, it takes very long time to find features. On the other hand, we divide the whole dataset into three subdivisions and they are feed into NCCA network sequentially. From these subdivisions of data one can find out feature subsets of any desired length. It also shows computation only within a few seconds for find a feature subset.

4.3 Results of Frequency Recognition (FR)

Human-body conditions can be known by recognizing the frequencies of brain signal. Here, we utilized CCA network to recognize frequency from SSVEP of EEG. The SSVEP is collected for various trials of a specific subject, but there have no significant difference among trials for a specific subject. When frequency is recognized from high dimensional EEG data of a specific subject, it takes longer time to execute and degrade the generalization performance. Thus informative features are selected using training and pruning which remove noisy features as well as improves the recognition quality. These results are evaluated into following subsections.

4.3.1 FR from Original EEG data

The CCA algorithm is used to extract underlying frequency components of multiway EEG data. The idea is to find maximum correlation points between reference sine-cosine signals and EEG signals of different subjects. Initially random weights are generated for both stages of the networks. Weights are updated according to update rules and normalized after presenting each data into the network. Three harmonic settings are utilized for each types of network. The harmonic settings of reference sine-cosine signals are regarded as $H1 \in$ Fundamental and multiple of fundamental, $H2 \in 2^{nd}$ Harmonic and multiple of 2^{nd} Harmonic, $H3 \in 3^{rd}$ Harmonic and multiple of 3^{rd} Harmonic. Subjects 1 to 4 are denoted as S1, S2, S3 and S4 respectively

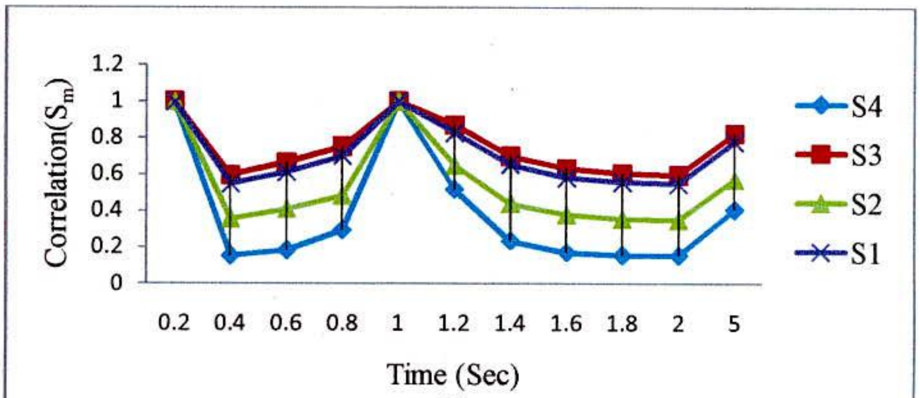
Figures 4.5, 4.6 and 4.7 describe correlation profiles against time ($1/f_m$) in seconds. At $H1$, maximum correlation occurs at 1 Hz stimulus frequency for different subjects as observed in Figure 4. It is seen that maximum correlation occurs at 5 Hz and 1 Hz with linear network for S1, S2, S3 and S4 as shown in Figure 4.5 (a). Similar results are also observed with nonlinear feedback network as shown in Figure 4.5 (c). However, maximum correlation becomes 1 Hz only with feedback free nonlinear network across all subjects as shown in Figure 4.5 (b).

At $H2$, when 2^{nd} Harmonic and multiple of 2^{nd} Harmonic are used as sine-cosine reference signals set, it is found that maximum correlation is found at 0.5, 1, 2.5 and 5 Hz with linear network, as shown in Figure 4.6 (a). We also use nonlinear and nonlinear feedback networks, as observe in Figures 4.6 (b) and 4.6 (c) respectively. In the last two cases, maximum correlation is found at 1 Hz stimulus frequency only.

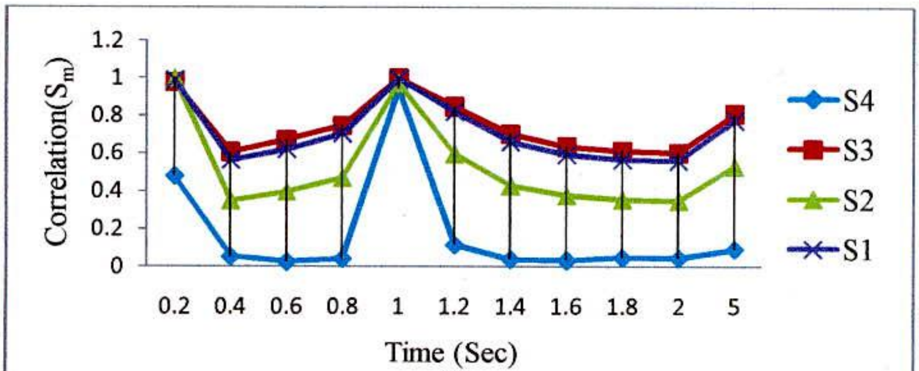
At $H3$, when 3^{rd} Harmonic and multiple of 3^{rd} Harmonic are used as sine-cosine reference signals set, it is found that maximum correlation is found at 0.625, 0.71, 1, 1.672 and 5 Hz with linear network, these are shown in Figure 4.7 (a). Nonlinear and nonlinear feedback networks are also studied. The results are plotted in Figures 4.7 (b) and 4.7 (c). Maximum correlation is found at 1 Hz stimulus frequency for both cases. In this sense, we claim that a frequency of 1 Hz is dominant for these experimental EEG data of different subjects. The appearance of maximum correlations at frequencies other than 1 Hz may be due to the harmonic and subjects variations.

CCA networks involve a number of used specified parameters such as learning rate (η, η_0) and Lagrange multipliers (λ_1, λ_2). The values of them are selected with few initial trial runs. It has been found empirically that best results are achieved when, $\eta_0 \gg \eta$. We choose $\lambda_1 = 0.0005$, $\lambda_2 = 0.000002$, $\eta_0 = 0.0005$ and $\eta = 0.00000015$ for representative result. Correlation profile does not change significantly if these values of constants are increased or decreased. Iteration is one of important factors for the convergence of NN. If iteration is increased the correlation profile is improved, resulting no change in frequency characteristic.

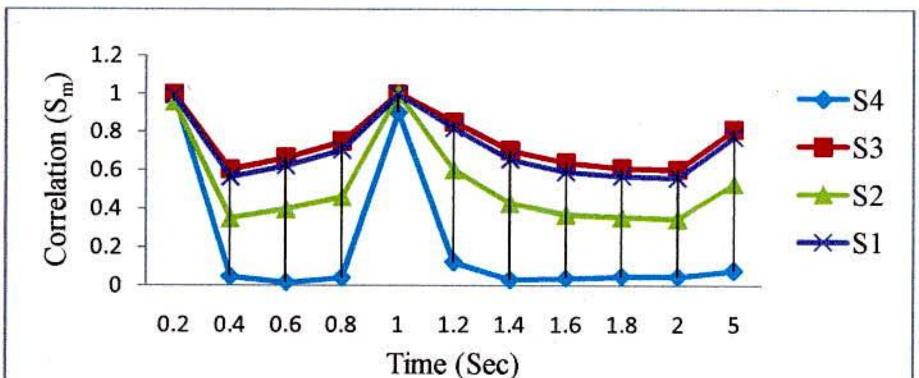
One can determine brain condition of a particular subject using this approach which is easy to program in an ordinary machine. Usually massive parallel EEG data requires a number of days to observe the final result. Our approach is simple to execute within a minute and does not require special machine.



(a)



(b)



(c)

Figure 4.5: Correlation coefficient of NCCA with (a) linear network, (b) nonlinear network, and (c) nonlinear network with feedback for H1 settings.

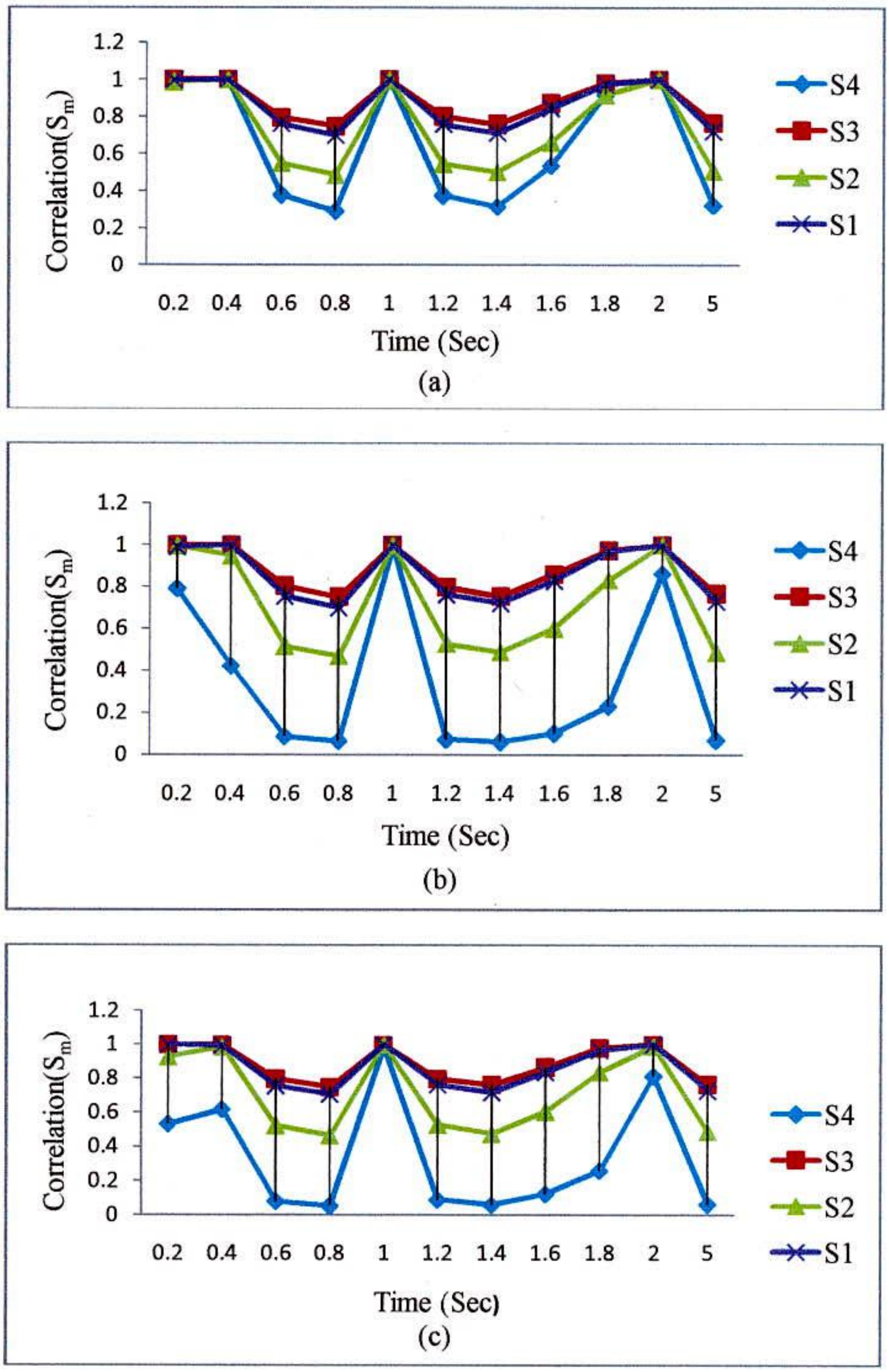


Figure 4.6: Correlation coefficient of NCCA with (a) linear network, (b) nonlinear network, and (c) nonlinear network with feedback for H2 settings.

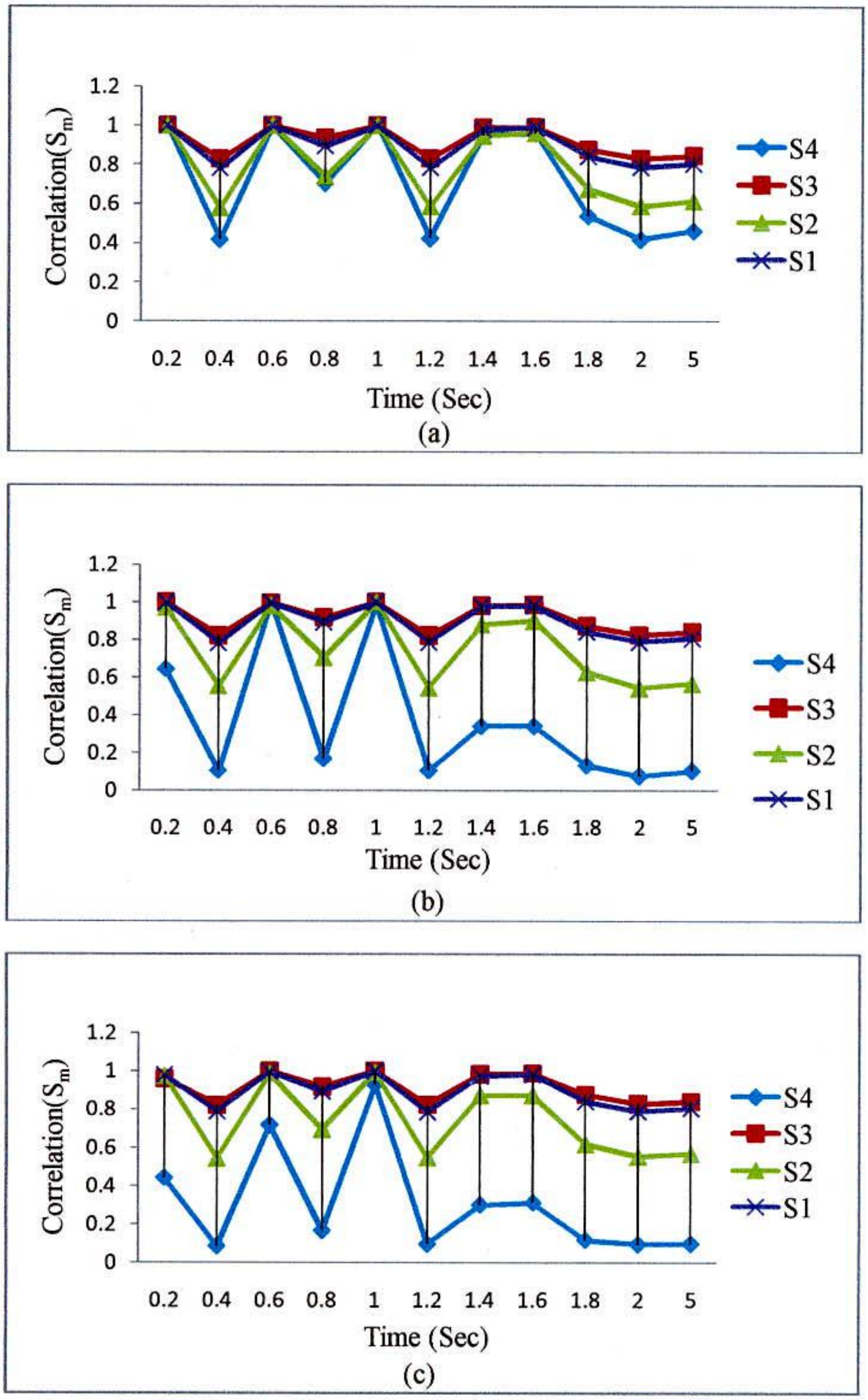


Figure 4.7: Correlation coefficient of NCCA with (a) linear network, (b) nonlinear network, and (c) nonlinear network with feedback for H3 settings.

4.3.2 Subjects Variability Realization

The CCA network is now used to compare inter subject variability and variability with reference sine-cosine signal of different subjects using MATLAB. Also trial to trial variability is studied. Initially random weight is generated and used to update for final output. In this situation, eq. 3.22 is utilized as reference signals set and eq. 3.35 is to compute correlation. The entire process can be understood from Fig. 4.8. The CCA algorithm to realize variability is outlined in A4.1 below.

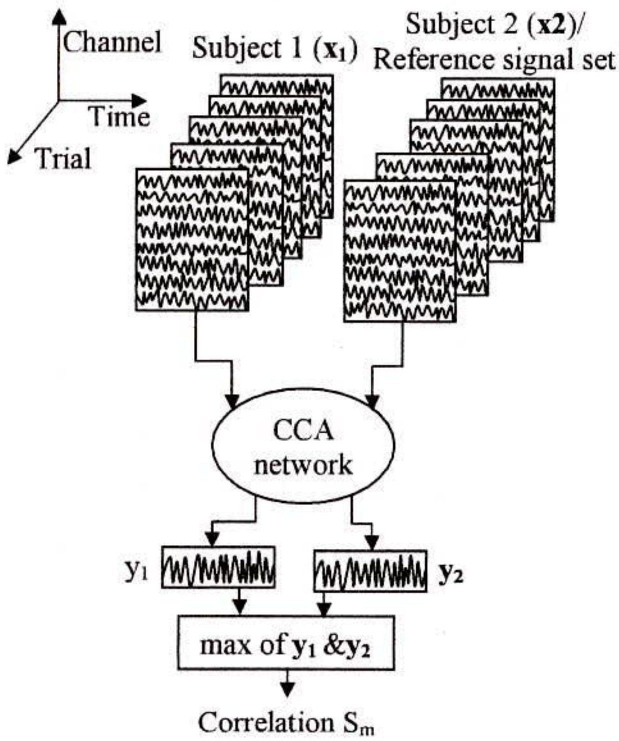


Figure 4.8: Illustration of NCCA approach for inter-subject variability

A4.1: NCCA algorithm for inter subject variability test

***Input**: EEG data $x_1, x_2, x_3, \ldots \ldots x_m \in R^{IxJxK}$ and sine-cosine signals y_m $(m=1,2,\ldots M) \in R^{2HxJ}$ corresponding to M stimulus frequencies, respectively.*

***Output**: Correlation S_m*

***for** m=1 to M do*

Random initialization for $\mathbf{w}_1$ and $\mathbf{w}_2$

repeat

Find $\mathbf{w}_1$ and $\mathbf{w}_2$ which maximize correlation between $\mathbf{y}_1$ and $\mathbf{y}_2$ by the CCA

***until** the maximum number of iteration is reached*

Compute the optimized signals y_1 and y_2

end

Compute correlation S_m

It is seen that 100% correlation for a subject of the same frequency, but it differs with different frequency for same subject. Like Sub1 (8Hz), it shows 100% correlation with Sub1 (8Hz), but 98% with Sub1 (14 Hz) and 97.6% with Sub1 (28Hz) which is shown in Fig. 4.9. In this figure, the horizontal numbers 1 to 4 for 8Hz, 5 to 8 for 14Hz and 9 to 12 for 28 Hz are used for four subjects accordingly. There is very low correlation with Sub3 (8Hz) and Sub4 (14Hz), which is almost 23%.

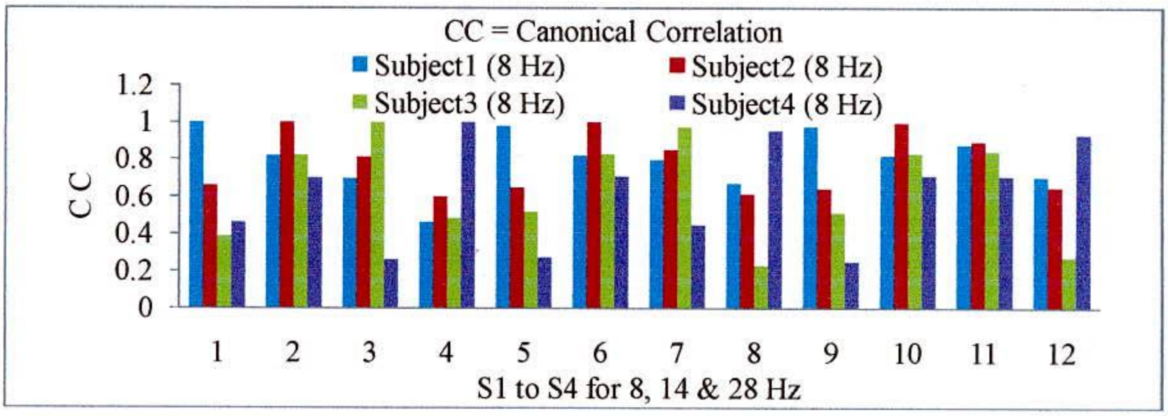


Figure 4.9: Inter subject variability with 8 Hz frequency

There is 100% correlation for Sub1 (14Hz) with Sub1 (14Hz) or Sub3 (14Hz) with Sub3 (14Hz), but for Sub3 (14Hz) with Sub3 (8 Hz) and Sub3 (28Hz), they are 97.8%and 95% respectively, as shown in Fig. 4.10. There is very low correlation with Sub3 (14Hz) and Sub4 (14Hz), that is about 15%.

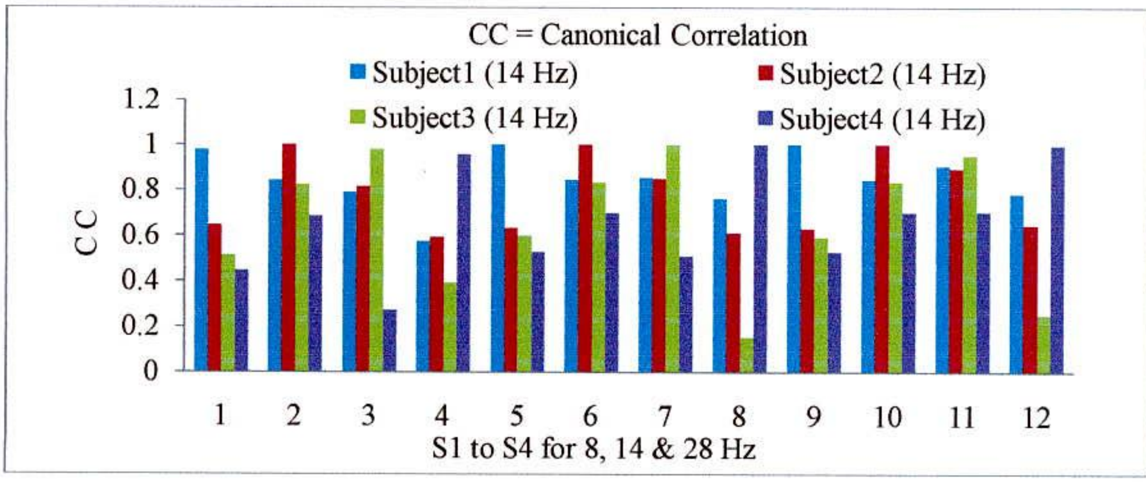


Figure 4.10: Inter subject variability with 14 Hz frequency

In this way, we have seen 100% correlation for Sub2 (28Hz) with Sub2 (28Hz) or Sub3 (28Hz) with Sub3 (28Hz), but for Sub3 (28Hz) with Sub3 (8 Hz) and Sub3 (14Hz) correlations are 90%

and 96% accordingly. Very low correlation (about 18%) is found between Sub3 (8Hz) and Sub4 (28Hz).

Sine-cosine reference signals are used to test the correlation of different trials of different subjects. There are total 20 trials for every stimulus frequency. We use respective stimulus frequency for sine-cosine signal. It is seen that almost same correlation for every trial of a subject with sine-cosine signals as shown in Fig. 4.11. Highest correlation is found with Sub3 of 28Hz (about 94%) for trial5 of sub3 at 28Hz stimulus frequency. But lowest correlation is found with Sub1 of 14Hz (about 26%) for trial1 of sub1 at 14Hz stimulus frequency.

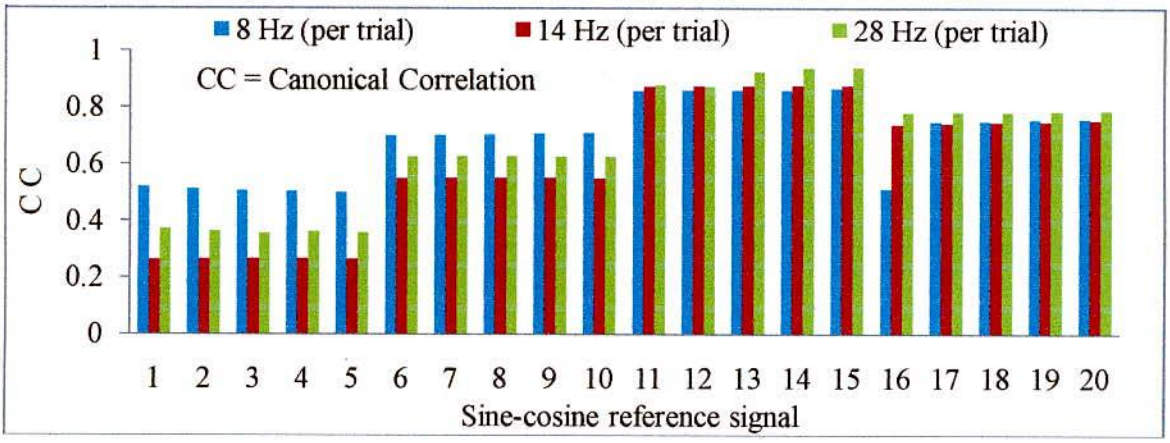


Figure 4.11: Trial to trial variability with sine-cosine reference signal

There was no big difference among trail to trail correlations for same subject. For example, correlations of trial1 to trial5 of subject1 at 28 Hz stimulus frequency with trial5 of subject2 at 14Hz stimulus frequency are 62.39% to 62.90%, which is shown in Fig. 4.12.

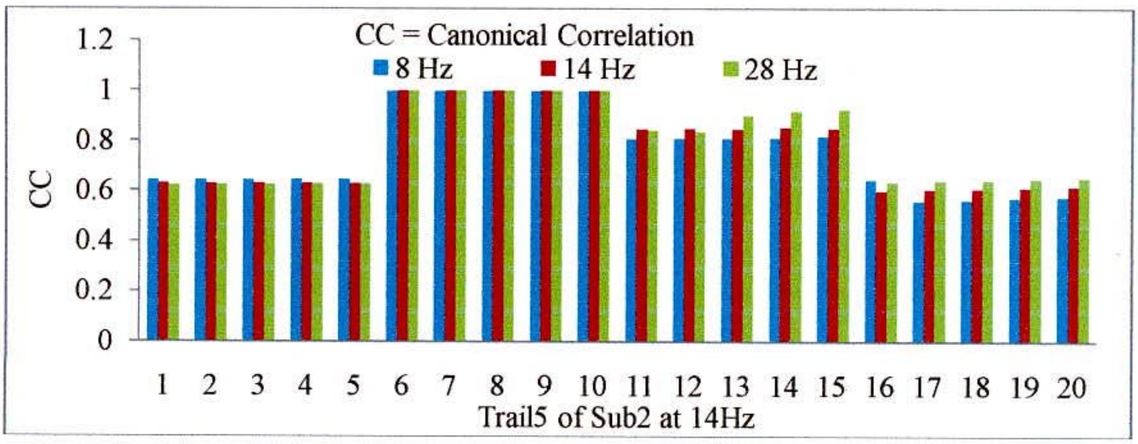


Figure 4.12: Trial to trial variability for trial5 of sub2 at 14 Hz

The trial to trial variability is checked for almost every subject and it is almost negligible. Finally, we apply sine-cosine reference signal to test the correlation of different subjects. We use respective stimulus frequency for sine-cosine signal. Highest correlation is found with Sub3 of 28Hz and it is about 91.91%. But lowest correlation is found with Sub1 of 14Hz and it is about 26.77% as shown in Fig. 4.13.

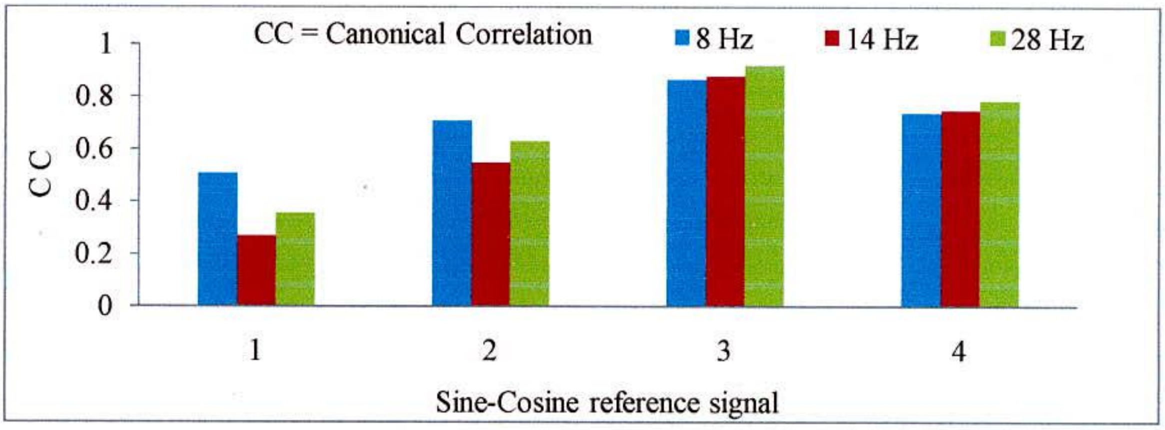


Figure 4.13: Inter subject variability with sine-cosine signal

From this study, it is realize that inter subject variability is most considerable candidate because correlation varies significantly with different subjects whether trial to trial variability of correlation is almost unchanged for a particular subject. When sine-cosine reference signal is used in place of a subject, higher correlation is found with subject 3 for 28Hz stimulus frequency and lower correlation with subject 1 for 14Hz stimulus frequency.

4.3.3 FR from Selected Features

In this case, underlying frequency components of EEG data are extracted from selected features. Firstly, update rule CCA network is applied to select the 15, 30 and 60 features from every EEG subjects. Two-stage CCA network is further applied in the selected features, to extract the underlying frequency components of each subject. In this sense, we use three different networks as – (i) linear CCA, (ii) nonlinear CCA and (iii) nonlinear CCA with feedback. Initially random weights are generated for every networks and correlation is maximized by different updates rule that is shown in chapter III. There have analyzed three harmonic conditions of sine-cosine reference signals for every selected features set. These harmonic conditions are denoted as *H1*, *H2* and *H3*. Here, *H1* denotes the fundamental and multiple of fundamental frequencies, *H2* denote the 2nd Harmonic and multiple of 2nd Harmonic as well as *H3* denote the 3rd Harmonic and multiple of 3rd Harmonic. Subjects 1 to 4 are denoted as *S1*, *S2*, *S3* and *S4* respectively. We use fourfold cross validation that is when *S4* is in the second stage as a test set, other three subjects are cascaded in the first stage as a training set and this is rounded for every subject. The test set subjects are indicated in the figures.

Firstly, we apply the linear CCA network on various selected features of different subjects at three harmonic conditions as shown in Fig. 4.14. Highest correlation is found at 1 Hz for every selected features of *S3* at *H1* condition as shown in Fig. 4.14(a). In this case, *S3* is the test set. Though correlation variation is also found at 5 Hz, but negligible correlation is originated for every other frequencies. In this sense, the govern frequency is 1 Hz for *S3*. The dominating frequency for *S4* is 1 Hz and 5 Hz that is shown from Fig. 4.14(b). It is also found that dominating frequency is 1 Hz for *S1* and *S2* at *H1* condition. At *H2* condition, the dictated frequency is found at 1 Hz for every subject as shown in Fig. 4.14(c) for *S1*, although variation is analyzed at 0.5, 2.5 and 5 Hz which is the effect of the second harmonic. Dominating frequency of 1 Hz is also found when test with *H3* condition as shown in Fig. 4.14(d) for *S2*, though variation is explored at 1.67 and 5 Hz for the effect of third harmonic. While the flickering of checkerboard is done at 1 Hz stimulus frequency, the result is acceptable. Although variation is found for subject and harmonic variation but dominating frequency is 1 Hz.

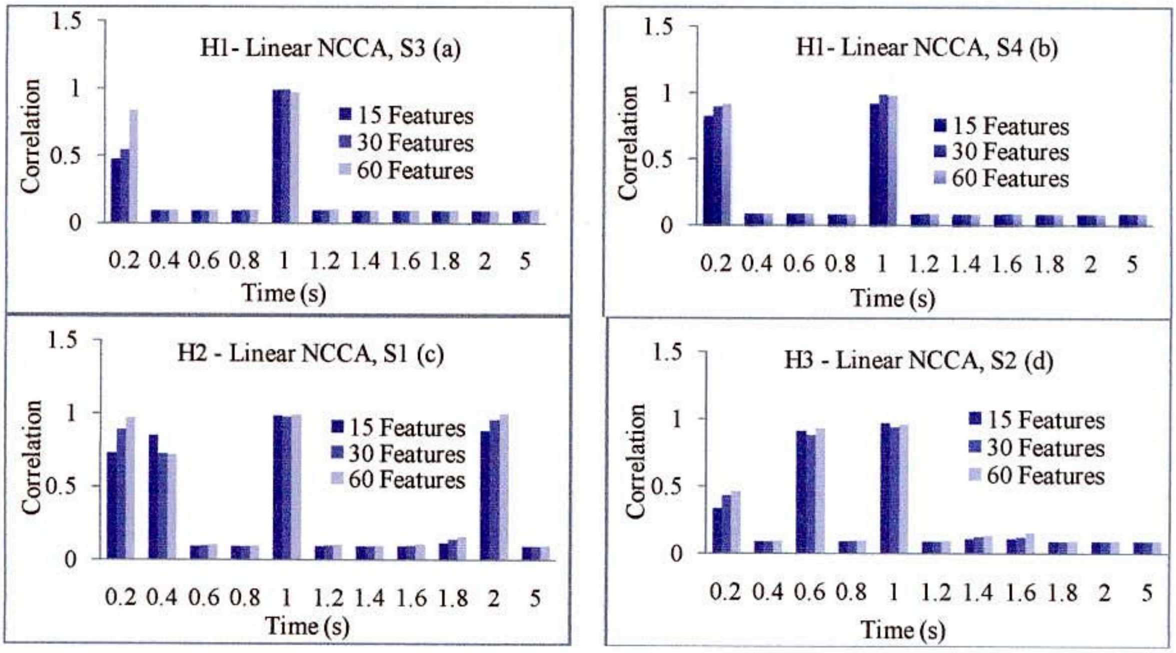


Figure 4.14: Correlation profiles of (a) *S3*, (b) *S4*, (c) *S1* and (d) *S2* with linear CCA network.

The correlation profiles are also shown in Fig. 4.15 those are found by nonlinear CCA network. The maximum correlation is obtained at 1 Hz with *H1* condition for every subject as an example it is seen from Figs. 4.15(a) and 4.15(b) for *S4* and *S3* respectively. Though effects of harmonics are different for different subjects, it is also explored that maximum correlation is occurred at 1 Hz stimulus frequency for every subject. This is proven by Figs. 4.15(c) and 4.15(d) for *S2* at *H2* and *H3* conditions respectively.

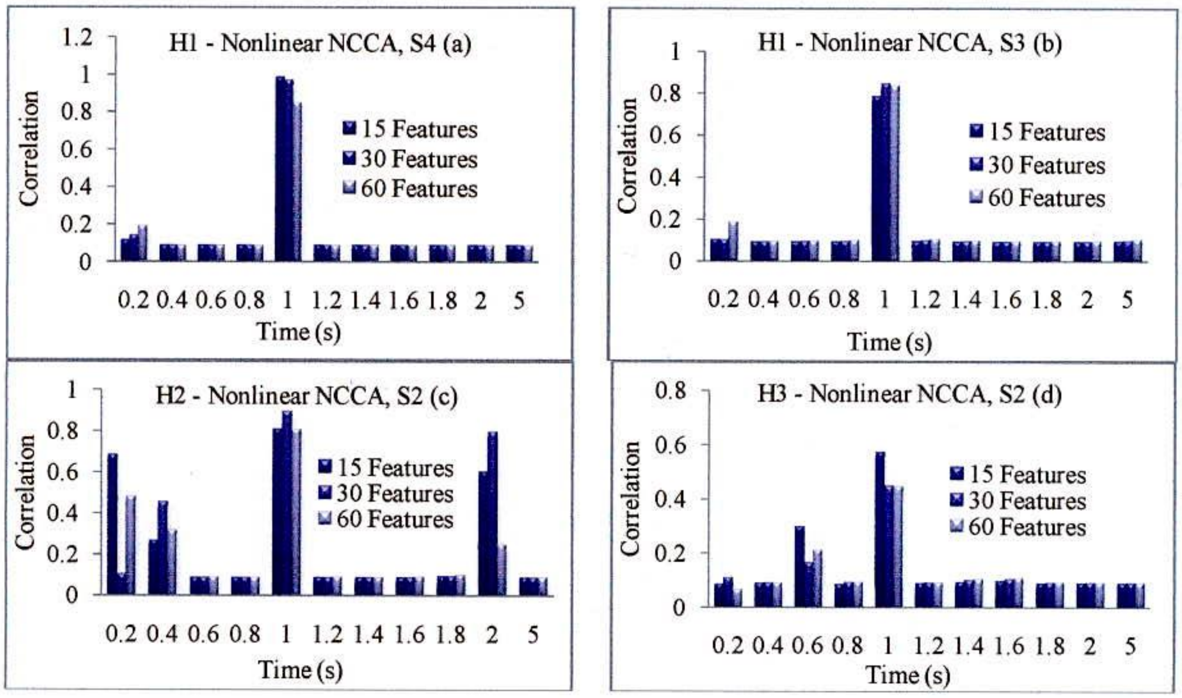
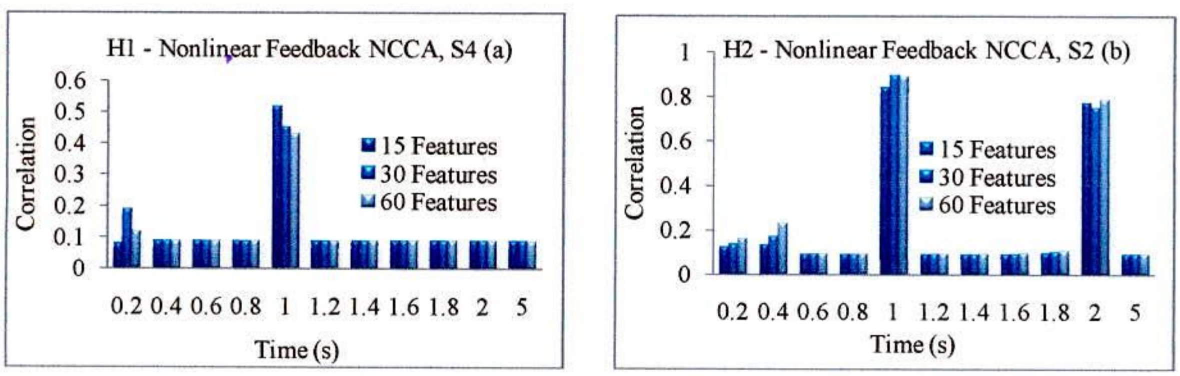


Figure 4.15: Correlation profiles of (a) *S4*, (b) *S3*, (c) *S2* and (d) *S2* with nonlinear CCA network.

The correlation profiles of different selected features with nonlinear feedback CCA network are shown in Fig. 4.16. The maximum correlations of every subject are shown at 1 Hz with *H1* condition as an example for *S4*, is explored at Fig. 4.16 (a). The correlations are also maximized at 1 Hz for *H2* and *H3* conditions, but due to harmonic variations maximize correlation points are also found at 0.5 Hz for *H2* and at 1.67 Hz for *H3* that is explored from Figs. 4.16 (b) and 4.16 (c). In Fig. 4.16 (d), maximum correlation is found at 1 Hz for *S1*, but variations are also found at 1.67 and 5 Hz due to effect of 3rd harmonics and multiple of 3rd harmonics. Finally we may say that though variations of maximum correlation points are found at different frequencies for the effect of harmonics and subjects variations, but exact dominant flickering frequency is 1 Hz.



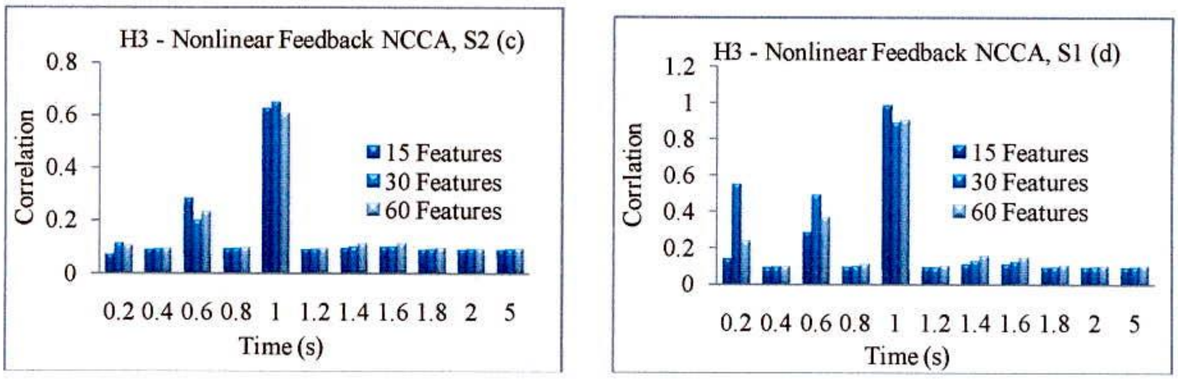


Figure 4.16: Correlation profiles of (a) *S4*, (b) *S2*, (c) *S2* and (d) *S1* with nonlinear feedback CCA network.

4.4 Comparative Study

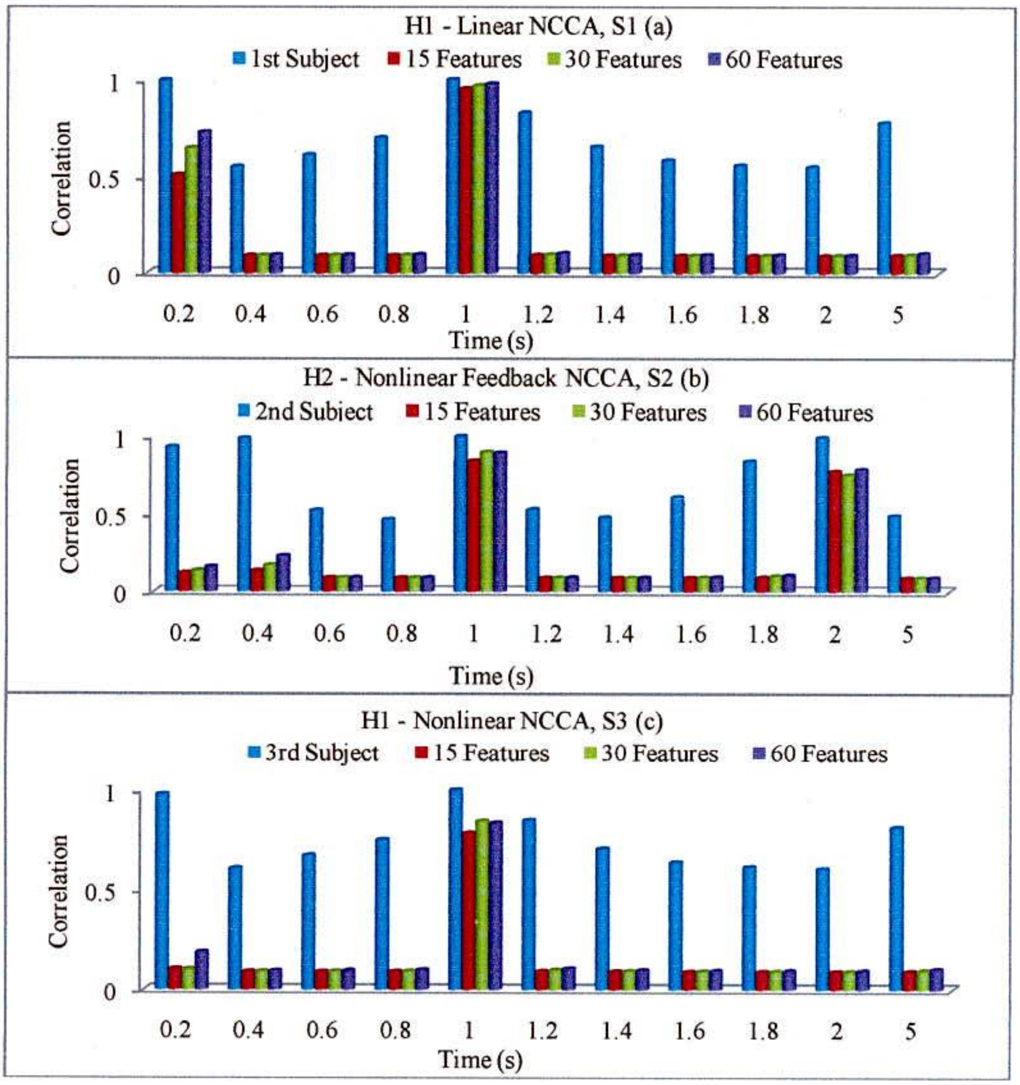
CCA networks need lower computational complexity and time than other methods though it utilizes training and pruning. It removes noisy features which increases generalization performance. When features are selected prior to recognize frequency, it greatly increases the performance of networks. In this way, subject's differentiation and stimuli at harmonic frequencies can easily be identified.

4.4.1 Correlation Point of View

Frequencies of EEG data are recognized by observing the maximum correlation points. In this sense firstly EEG data are concatenated for five trials of a subject and applied to the two-stage CCA network to find out exact frequencies. On the other hand, features are selected for a concatenated subject and obtained the maximum correlation points from two-stage CCA network at a define frequency. Fig. 4.17 shows the comparative outline for original and selected EEG data set at different harmonic conditions.

The correlation profiles for *S1* at *H1* condition is shown in Fig. 4.17 (a). Maximum correlations are found at 1 Hz for original and selected EEG data set. In the case of 1.25 Hz, it is seen that correlation is 0.701 for *S1* but 0.0898, 0.0924 and 0.0957 for three types of selected features of *S1*. The correlation point's variation is not high for original EEG, but selected features show high correlation differences between expected frequency and others that are examined from Fig. 4.11. There have little bit correlation variations at 5 Hz for original and selected EG features at *H1* condition that is verified from Fig. 4.17 (a). It is the effect of fundamental and multiple of fundamental frequencies of a definite subject. The effect of 2nd harmonic and multiple of 2nd harmonic is shown at 0.5 Hz for S2 that is examined from Fig. 4.17 (b). The effect of *H1* at different frequencies is negligible for *S3* that is examined from Fig. 4.17 (c). We may also select a subject or differentiate among subjects to see the correlation profiles at different frequencies. It is seen from Figs. 4.17 (a) and 4.17 (c) that correlation variations are found at 5 Hz for *S1* but not for *S3* where both are in *H1* condition. The effect of 3rd harmonic and multiple of 3rd harmonic is explored at 5 Hz and 1.67 Hz for *S4* which is examined from Fig. 4.17 (d).

In all of the above cases we see that maximum correlations are found at 1 Hz due to the fact that, the checkerboard is flickered at this frequency. In this sense, the network is suitable for finding the recognized frequency as well as it can be differentiated among subjects by observing the correlation profiles. Although highest correlations are found at 1 Hz for all of the above cases, but negligible correlations are found at other frequencies for selected features but not for original EEG data. In this sense, expected frequency can easy be obtained from correlation profiles of selected EEG features than original one.



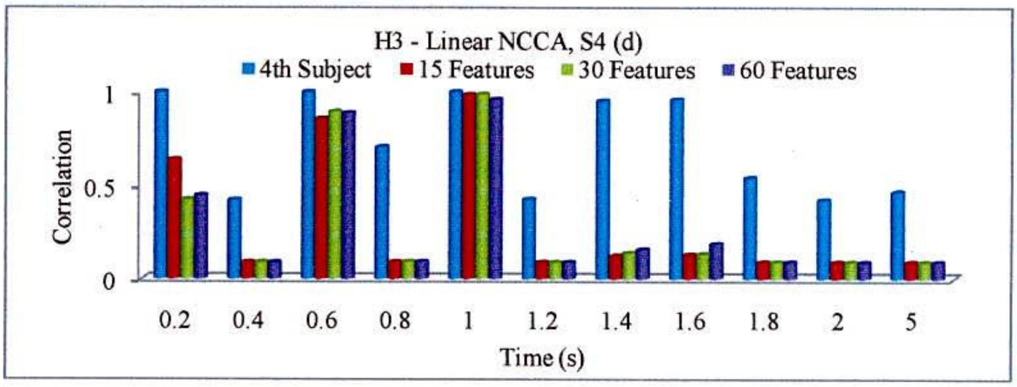


Figure 4.17: Correlation profiles of (a) *S1*, (b) *S2*, (c) *S3* and (d) *S4* for original and selected features of EEG data.

4.4.2 Computational Complexity

Original EEG: Two-stage CCA networks are utilized for frequency recognition. Let, there are '*F*' numbers of feature for each subject. Three subjects are concatenated in the first stage, so computational complexity for these concatenated subject is $O(3F)$. The number of sampling points of reference sine-cosine signals depends on stimulus frequency f_m. When, $f_m = 1$ the sampling points are same as total number of features '*3F*' that's why computational complexity for reference signals can be denoted as $f_m \times O(3F)$. The number of sampling points is increased with increasing stimulus frequency. In the first stage total computational complexity is found as $f_m \times O(3F) + O(3F)$. The size of the optimized reference signal of first stage is '*F/128*', where total number of patterns is 128. The computational complexity for first stage optimized reference signal is $O(F/128) \ll f_m \times O(3F) + O(3F)$. There is a subject as test set in the second stage which has '*F*' number of features. The computational complexity for the test EEG features is $O(F)$. In this sense, total computational complexity for two-stage NCCA network is $f_m \times O(3F) + O(3F) + O(F)$. Further another optimized reference signal is generated from second stage that's has a complexity of $O(F/128) \ll f_m \times O(3F) + O(3F) + O(F)$. The total computational complexity for frequency recognition from original EEG data is $f_m \times O(3F) + O(3F) + O(F)$.

Selected EEG features: In this case, EEG data is subdivided into three subsections before selecting features denoted as *m*, *n* and *k* respectively. The total number of features for a given dataset is $F = m + n + k$, and then the cost of measuring correlation is $O(m) + O(n) + O(k)$. In addition there require $O(mlog(m)) + O(nlog(n)) + O(klog(k))$ for pruning process. In this case, *m*, *n*, and *k* are reduced by one in each course of pruning. It is obtained that $O(m) + O(n) + O(k) \gg O(mlog(m)) + O(nlog(n)) + O(klog(k))$. The number of selected features of a subject is $p \ll (m + n + k)$ and computational cost for selecting this features is $O(m) + O(n) + O(k) + O(p) \approx O(m) + O(n) + O(k)$.

Here frequency is recognized from 'p' number of features not from '*F*' number of features where $p \ll F$. In this sense computational cost for frequency recognition from selected EEG features is $f_m \times O(3p) + O(3p) + O(p) \ll f_m \times O(3F) + O(3F) + O(F)$. the total computational complexity for frequency recognition from selected EEG features is $f_m \times O(3p) + O(3p) + O(p) + O(m) + O(n) + O(k)$.

Feature selection is a sub-procedure for recognizing frequency from selected EEG features where linear NCCA network takes about 140 to 150s for searching expected features. The elapsed time for recognizing frequency is analyzed in Fig. 4.18. The linear NCCA network spend about 57.678s at 5 Hz for *S1*, but around 27s for selected features of *S1* where at 0.2 Hz, *S1* takes 31.642s but selected features take around 1.5s as shown in Fig. 4.18 (a). These results are almost same for other subjects with applying linear CCA network that is examined from Fig. 4.18 (b) for *S4*. The effects of harmonics are negligible in these cases study. When nonlinear CCA network is utilized, 138.72s is spent at 5 Hz for *S3* but around 51s for selected EEG features of *S3* where at 0.2 Hz *S3* takes 67.673s and selected features take only around 2s as examined from Fig. 4.18 (c). We also analyzed this case by utilizing the nonlinear feedback NCCA network. It is seen from Fig. 4.18 (d) that 137.27s is required for computing correlation at 5 Hz of *S2* but selected features of *S2* is required around 52s where 67.768s is required at 0.2 Hz of *S2* but around 2.5s for selected features of *S2*. In this case, variation of recognition time is only varied by network variation but do not depend on the subjects or harmonic variations. Finally it may be say that higher recognition time is required for analyzing original EEG data, but very low recognition time is required to analyze selected features which also give remarkable results.

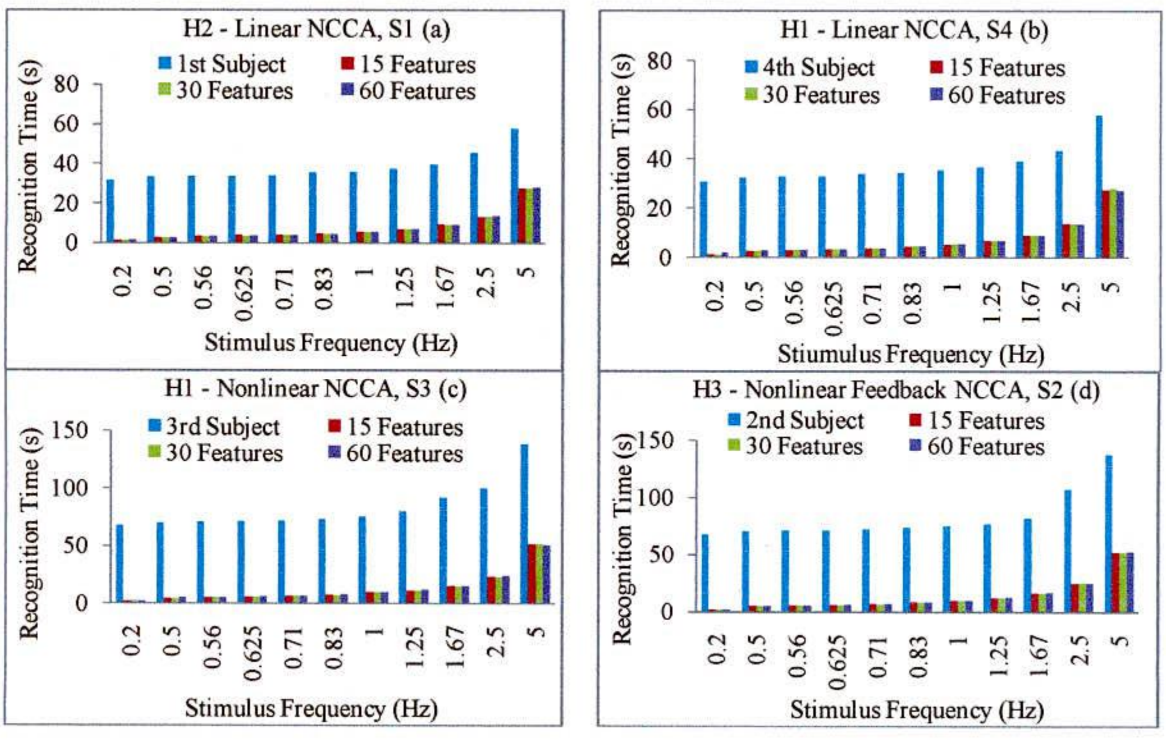


Figure 4.18: Time elapsed to recognize frequency of (a) *S1*, (b) *S4*, (c) *S3* and (d) *S2* with three CCA networks.

4.4.3 Comparison with Standard Methods

The most widely used frequency detection method in SSVEP is PSDA but it might still be sensitive to noise, shows higher inter-subject variability, and cannot easily identify stimuli at harmonic frequencies. The traditional CCA method has shown significantly better recognition performance than PSDA. A potential problem of CCA is that all parameters for recognition are estimated from test data since the reference signals of sine-cosine waves do not include features from training data. Hence, the CCA method often does not result in the optimal recognition accuracy of SSVEP frequency due to possible over fitting.

The multi-way CCA method [12] (use tensor data) has shown improved recognition performance of SSVEP frequency compared to the CCA method. On the other hand, a phase constrained CCA method (PCCA) [106] has also been proposed for SSVEP frequency recognition. The PCCA method achieved significant accuracy improvement in comparison with the CCA method, by embedding the phase information estimated from training data into the reference signals. However, the procedures of reference signal optimization in both the multi-way CCA and the PCCA methods are not completely based on training data but still need to resort to the pre-constructed sine-cosine waves. Multi-set CCA was developed as an extension of CCA to find multiple linear transforms that maximize the overall correlation among canonical variates from multiple sets of random variables [107]. Although multi-set CCA method is completely based on training data and outperforms on different CCA methods for SSVEP recognition, but it is only suitable for a small number of channels.

In this thesis, we use high dimensional SSVEP data which have 128 patterns with more 30,000 features of a single subject. When we use above traditional CCA methods to recognize frequency from these data, it takes very long time (about 24 to 25 hours) to execute a matlab program with a traditional computer. In this regard, we provide training and pruning by introducing NN with CCA that's we called CCA network. This method may be solved above problems clearly because it is completely based on training data and reduces computational time and cost.

CHAPTER V

Conclusions and Future Works

5.1 Conclusions

Frequency recognition from SSVEP plays a vital role to know the brain conditions due to visual stimulations. In this study, two-stage CCA approach is proposed to recognize the stimulus frequency. CCA network is implemented between the EEG data and reference sine-cosine signals to get optimized reference signals in the first stage. It is also applied in the second stage to inspect the correlation between the test EEG data and optimized reference signals of the first stage. Both optimized signals contain information of subject-specific and trial-to-trial variability meaning that NCCA converges to underlying frequency components. Finally frequency components are extracted from two optimized signals. Three different networks such as – (i) linear CCA, (ii) nonlinear CCA and (iii) nonlinear feedback CCA are used to recognize frequencies at three different harmonic situations. It is seen that though maximum correlations are found at different frequencies for different subject but 1 Hz frequency is dominant for every subject. The proposed CCA method can executes a program within a single minute with user friendly ordinary computer, whereas the statistical CCA take several hours for execution.

Although it is possible to find the expected frequencies by using whole of the data, but all features of these high dimensional data are not equally important and may also have noise corrupted data as well. In this sense, the CCA approach is employed for feature selection (FS) of high dimensional EEG dataset. Firstly, every attribute of a subject is classified using wavelet clustering method. Then class information is added to the original dataset. Expected subsections are created from whole database of a subject. CCA network is applied for finding correlation by weight maximization. Then features are deleted from original dataset according to weight minimization process. Thus only maximum weights respective features are remained. These features are tested according to Back-propagation (BP) rule, where training is done by 75% attributes and rests are used for testing. Finally, we get 100% accuracy mostly for every dataset except subject 4 for 8 Hz stimulation due to higher level of noise. So, CCA network is a suitable choice for FS with a low computational cost.

It is also possible to recognize frequency of SSVEP to know the brain condition from selected EEG features with higher accuracy and lower computation cost. In this view, correlations are computed using above three networks with three harmonic situations. It is seen that maximum correlations are obtained at 1 Hz because of the checkerboard was flickered at this frequency. It is observed from correlation profiles that difference between correlated and non-correlated points is higher for selected EEG features than original EEG features due to the removal of noisy features. It can also be differentiated among subjects that are observed from correlation profiles as well. CCA networks are also very much computationally inexpensive to recognize frequencies from selected EEG features than original one. The linear CCA network is shown about 2 to 21 times faster as well as nonlinear CCA and nonlinear feedback CCA networks are shown about 2.5 to 33 times faster to recognize frequencies from selected EEG features due to different stimulus frequencies. Therefore CCA network is quite suitable choice for FS as well as to recognize frequencies from EEG signals with higher accuracy.

5.2 Future Works

The brain technology has limitless possibilities to aid the disabled users and assist the healthy people in society. The emergences of successful Brain-Computer Interfaces (BCI) based on noninvasive scalp EEG have become an increasingly active research area. The SSVEP are precisely synchronized brain responses with the fast repetitive external visual stimulation. These brain responses are found from subjects with the use of EEG data acquisition system. CCA network is a suitable choice to detect the existence of the SSVEP and determine its frequency. The command will be created by determining these frequencies to control the device which will be our further study.

This research will be helpful to patients with motor disabilities to improve their quality of life. This technology can also be used to improve the performance of normal healthy users. Overall, it will be advantageous to employ emotional-face video stimuli in SSVEP-based practical applications such as BCI.

References

[1] Tatum, W. O., Husain, A. M., Benbadis, S. R., 2008, "Handbook of EEG Interpretation". Demos Medical Publishing.

[2] Cichocki, A., Zdunek, R., Phan, A. & Amari, S., 2009, "Nonnegative matrix and tensor factorizations: Applications to exploratory multi-way data analysis and blind source separation". New York, NY: Wiley.

[3] Nunez PL, Srinivasan R., 1981, "Electric fields of the brain: The neurophysics of EEG". Oxford University Press.

[4] D. H. Zhu, J. Bieger, G. G. Molina and R. M. Aarts, 2010, "A survey of stimulation methods used in SSVEP-based BCIs," Comput. Intell. And Neurosci, vol. 46, No. 15 pp. 7023-57

[5] Website: http://sccn.ucsd.edu/eeglab/

[6] Zhang Y., Jin J., Qing X., Wang B. & Wang X., 2011, "LASSO based stimulus frequency recognition model for SSVEP BCIs". Biomed. Signal Process. Control. Vol. 49, no. 2 pp. 181-191.

[7] Cheng, M., Gao, X., Gao S. & Xu, D., 2002, "Design and implementation of a brain-computer interface with high transfer rates". IEEE Trans. Biomed. Eng., vol. 49, pp. 1181-1186.

[8] MÄuller-Putz, G.R., Scherer, R., Brauneis, C., and Pfurtscheller, G., 2005, "Steady-state visual evoked potential (SSVEP)-based communication: impact of harmonic frequency components". J. Neural. Eng., vol. 2, pp. 123-130.

[9] Lin Z., Zhang, C., Wu, W., & Gao X., 2006, "Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs". IEEE Trans. Biomed. Eng., vol. 53, pp. 2610-2614.

[10] Hakvoort G., Reuderink B. & Obbink, M., 2011, "Comparison of PSDA and CCA detection methods in a SSVEP-based BCI-system". Technical Report, TR-CTIT-11-03, EEMCS, ISSN: 1381-3625.

[11] Kim, T.K. & Cipolla, R., 2009, "Canonical correlation analysis of video volume tensor for action categorization and detection". IEEE Trans. PAMI. 31, pp. 1415-1428.

[12] Zhang, Yu. Zhou, G., Zhao, Q., Onishi, A., Wang, J. J., & Cichocki, A., 2011, "Multiway Canonical Correlation Analysis for Frequency Components Recognition in SSVEP-based BCIs". Lect. Notes in Comp. Sci., vol. 70, no. 62, pp. 287-295.

[13] M. Z. Hossain, M. J. A. Rabin, A. F. M. N. Uddin, M. Shahjahan, 2013, "Canonical correlation analysis with neural network for inter subject variability realization of EEG data," in proc. IEEE of the 2nd Int. Conf. ICIEV, Dhaka, 17-18, pp. 1-5.

[14] Md. Monirul Kabir., Md. Shahjahan., & Kazuyuki Murase, 2012, "A new hydrid ant colony optimization algorithm for feature selection". Expert Systems with Applications, vol. 39, pp. 3747-3763

[15] Steppe, J. M., Bauer, K. W., 1996, "Improved feature screening in feed forward neural networks". Neurocomputing, vol. 13, pp. 47–58.

[16] Duda, R. O., Hart, P.E., 1973, "Classification and Scene Analysis". Wiley, New York.

[17] Aghdam, M.H., Aghaee, N. G., & Basiri, M. E. 2009, "Text feature selection using ant colony optimization". Expert Systems with Applications, vol. 36, pp. 6843-6853.

[18] Liu, H., & Tu, L. 2005, "Toward integrating feature selection algorithms for classification and clustering". IEEE Transactions on Knowledge and Data Engineering, vol. 17 no. 4, pp. 491-502.

[19] Huang, J., Cai, Y., & Xu, X. 2007, "A hybrid genetic algorithm for feature selection wrapper based on mutual information". Pattern Recognition Letters, vol. 28, pp. 1825-1844.

[20] Guan, S., Liu, J., & Qi, Y. 2004, "An incremental approach to contribution-based feature selection". Journal of Intelligence Systems, vol. 13 no. 1.

[21] Hsu, C., Huang, H., & Schuschel, D. 2002, "The ANNIGMA-wrapper approach to fast feature selection for neural nets". IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, vol. 32 no. 2, pp. 207-212.

[22] Caruana, R., & Freitag, D. 1994, "Greedy attribute selection". In Proceedings of the 11th international conference of machine learning. USA: Margan Kaufman.

[23] Lai, C., Reinders, M. J. T., & Wessels, L. 2000, "Random subspace method for multivariate feature selection". Pattern Recognition Letters, vol. 27, pp. 1067-1076.

[24] Yang, J. H., & Honavar, V. 1998, "Feature subset selection using a genetic algorithm". IEEE intelligent Systems, vol. 13 no. 2, pp.44-49.

[25] Bishop, C. M., 1995, "Neural Networks for Pattern Recognition". Clarendon Press, Oxford.

[26] K. H. Liu, J. Zhang, B. Li and J. X. Du., 2009, "A GA-based approach to ICA feature selection: An efficient method to classify microarray datasets". Advances in neural networks, Lecture notes in computer science. Vol. 5552, pp 432-441.

[27] B. A. Draper, K. Baek, M. S. Bartlett and J. R. Beveridge., 2003, "Recognizing faces with PCA and ICA". Computer Vision and Image Understanding, vol. 91 pp. 115–137.

[28] JR. Quinlan., 1986, "Induction of decision trees". Mach Learn, pp. 81–106.

[29] B. Draper, C. Kaito and J. Bins., 2003, "Iterative Relief". Madison. Workshop on Learning in Computer Vision and Pattern Recognition, vol. 58, pp. 63-69.

[30] JH. Moore and BC. White, 2007, "Tuning Relief-F for Genome-Wide Genetic Analysis". Lecture Notes in Computer Science, pp. 166–175.

[31] Z. Zhu, Y. S. Ong, M. Dash, 2007, "Markov blanket-embedded genetic algorithm for gene selection." Pattern Recognition, vol. 49 no. 11 pp. 3236–3248.

[32] Biliana Paskaleva, Majeed M. Hayat, Zhipeng Wang, J. ScottTyo, and Sanjay Krishna, 2008, "Canonical Correlation Feature Selection for Sensors with Overlapping Bands: Theory and Application". IEEE transactions on geo science and remote sensing, vol. 46, no. 10

[33] P.L. Lai, C. Fyfe., 1999, "A neural implementation of canonical correlation analysis". Neural Networks, vol. 12, pp. 1391-1397.

[34] Oja, E. 1982, "A simplified neuron model as a principle component analyzer". Journal of Mathematical Biology, vol. 16, pp. 267-273.

[35] Sanger, T., 1990, "Analysis of the two-dimensional receptive fields learned by the generalized hebbian algorithm in response to random dot input". Biological Cybernetics, vol. 63, pp. 221-228.

[36] L. Ke, Z. Feng, Z. Ren, 2008, "An efficient ant colony optimization approach to attribute reduction in rough set theory", Pattern Recognition Letters, vol. 29 pp. 1351–1357.

[37] Website: http://www.bakardjian.com/work/ssvep_data_Bakardjian.html

[38] Niedermeyer E. and da Silva F.L., 2004, "Electroencephalography: Basic Principles, Clinical Applications, and Related Fields". Lippincot Williams & Wilkins.

[39] A. Gevins and A. Remond., 1987, "Handbook of Electroencephalography and clinical Neuro-physiology, Methods of analysis of brain electrical and magnetic signals". Elsevier, Amsterdam, vol. 1.

[40] F. H. Lopes da Silva, Storm van Leeuwen and A. Remond., 1986, "Handbook of Electroencephalography and clinical Neuro-physiology, Clinical applications of computer analysis of EEG and other neuro physiological signals". Elsevier, Amsterdam, vol. 2.

[41] D. D. Daly and T. A. Pedley., 1990, "Current practice of Clinical Electroencephalography". 2nd ed. Reven, New York.

[42] Steriade M., 1993, "Cellular substrates of brain rhythms. In: E. Niedermeyer and F. H. Lopes da Silva (Eds.), Electroencephalography: Basic Principles, Clinical Applications and Related Fields". Baltimore: Williams and Wilkins 3rd ed., pp. 27-62.

[43] Pfurtscheller, G.; Lopes Da Silva, F.H., 1999, "Event-related EEG/MEG synchronization and desynchronization: Basic principles". Clinical Neurophysiology, vol. 110 no. 11 pp. 1842–57.

[44] Palva, S. and Palva, J.M., 2007, "New vistas for a-frequency band oscillations". Trends Neurosci.

[45] Feshchenko, Vladimir A.; Reinsel, Ruth A.; Veselis, Robert A., 2001, "Multiplicity of the alpha Rhythm in Normal Humans". Journal of Clinical Neurophysiology, vol. 18 no. 4 pp. 331–44.

[46] Cahn, B. Rael; Polich, and John, 2006, "Meditation states and traits: EEG, ERP, and neuroimaging studies". Psychological Bulletin, vol. 132, no. 2, pp. 180–211.

[47] Gilmore, P. C. & Brenner, R. P., 1981, "Correlation of EEG, computerized tomography, and clinical findings. Study of 100 patients with focal delta activity". Archives of Neurology, vol. 38, pp. 371-372.

[48] Fernandez-Bouzas, A., Harmony, T., Bosch, J., Aubert, E., Fernandez, T., Valdes, P., 1999, "Sources of abnormal EEG activity in the presence of brain lesions". Clinical Electroencephalography, vol. 30, pp. 46-52.

[49] Engel, A. K., Fries, P., & Singer, W., 2001, "Dynamic predictions: oscillations and synchrony in top-down processing". Nature Reviews Neuroscience, vol. 2, pp. 704-716.

[50] Luck, Steven J., 2005, "An Introduction to the Event-Related Potential Technique". The MIT Press.

[51] Regan D., 1989, "Human brain electrophysiology. Evoked potentials and evoked magnetic fields in science and medicine". Elsevier, Amsterdam.
[52] Beverina F, Palmas G, Silvoni S, Piccione F and Giove S., 2003, "User adaptive BCIs: SSVEP and P300 based interfaces". Psych Nol. J. vol. 1, pp. 331–54
[53] Regan D, 1977, "Steady-state evoked potentials". J Opt Soc Am, vol. 67 no. 11 pp. 1475-89.
[54] Astolfi L., Bakardjian H., Cincotti F., Mattia D., Marciani M.G., De Vico Fallani F., Colosimo A., Salinari S., Miwakeichi F., Yamaguchi Y., Martinez P., Cichocki A., Tocci A., Babiloni F., 2007, "Estimate of causality between independent cortical spatial patterns during movement volition in spinal cord injured patients". Brain Topogr, vol. 19 no. 3 pp. 107-23.
[55] Metting van Rijn AC, Peper A, Grimbergen CA, 1994, "Amplifiers for bioelectric events: a design with a minimal number of parts". Med & Biol Eng & Comput, vol. 32 pp. 305-310.
[56] S. C. Layne. G. Mayer-Kress and J. Holzfuss., 1986, "Dimensions and entropies in chaotic systems". Springer-Verlag, Berlin, pp. 246-256.
[57] G. Mayer-Kress and S. C. Layne, 1987, Ann. N. Y. Acad. Sci. vol. 504, no. 11
[58] E. Basar, 1990, "Chaos in Brain function". Springer-Verlag, Berlin.
[59] A. Babloyantz and A. Destexhe, 1987, "Electroencephalogram." in proceedings of the IEEE first international conferences on Neural Networks, edited by M. Caudill and C. Butler vol. 4, pp. 31-40.
[60] J. P. N. Pijn, J. Van Neerven, A. Noest and F. Lopes da Silva, 1991, "Electroencephalogram". Clin. Neuro physiol. Vol. 79, pp. 371.
[61] B. J. West, 1990, "Fractal Physiology and Chaos in Medicine, Nonlinear Phenomena in Life Science". World Scientific, Singapore, vol. 1.
[62] B. J. West, 1990, "Patterns, Information and Chaos in Neuronal Systems, Nonlinear Phenomena in Life Science". World Scientific, Singapore, vol. 2.
[63] J. Gotman, D. D. Daly and T. A. Pedley., 1990, "In current practice of clinical encephalography". Reven, New York, pp. 51-83.
[64] J. P. M. Pijn, 1990, "Ph.D. thesis", Universities van Amsterdam.
[65] H. J. A. Beldhis, T. Suzuki, J. P. M. Pijn, P. Teisman and F. Lopes da Silva, 1993, "Practice of clinical encephalography". Eur. J. Neuroscience, vol. 5, pp. 944.
[66] J. P. M. Pijn and F. Lopes da Silva, 1993, "Basic Mechanism of the EEG", Bishauser, Boston.
[67] J. G. Caputo, N. B. Abraham, A. M. Albano, A. Passamante and P. Rapp, 1989, "Measures of complexity and chaos." NATO advanced study institute series B: Physics, Plenum, New York, vol. 208, pp. 99-110.
[68] J. A. McEwen and G. B. Anderson, 1975, "Measures and Mechanism of the EEG", IEEE Trans. Biomed. Eng., vol. 22, pp. 361.
[69] T. Gasser, A. Remond., 1977, "EEG informatics: a Didactic review of methods and application of EEG data processing". Elseveir, New York, pp. 37-55.
[70] H. Sugimoto, N. Ishii and N. Suzumura, 1977, "EEG informatics." Comput. Programs Biomed, vol. 7, pp. 293.
[71] S. Blanco, R. Quian, O. A. Rosso and S. Kochen, 1995, "Time frequency analysis of electroencephalogram series". Physical Review E, vol. 51, no. 3, pp. 2624-2631.
[72] Wolpaw J, Birbaumer N, McFarland D, Pfurtscheller G, VaughanT. 2002, "Brain-Computer Interfaces for Communication and Control", Clinical Neurophysiology, Elsevier, Ireland, vol. 113, pp. 767-791.

[73] Cheng M., Gao X., Gao S., Xu D. 2002, "Design and implementation of a brain-computer interface with high transfer rates". IEEE Trans. Biomed. Eng. vol. 49, pp. 1181-1186.

[74] Middendorf M, McMillan G, Calhoun G and Jones K., 2000, "Brain–computer interfaces based on the steady-state visual-evoked response," IEEE Trans. Rehabil. Eng., vol. 8, pp. 211–214

[75] Kelly S, Lalor E, Finucane C, McDarby G and Reilly R., 2005, "Visual spatial attention control in an independent brain–computer interface." IEEE Trans. Biomed. Eng, vol. 52, pp. 1588–1596.

[76] Friman O, Volosyak I and Graser A., 2007, "Multiple channel detection of steady-state visual evoked potentials for brain–computer interfaces." IEEE Trans. Biomed. Eng, vol. 54, pp. 742–750.

[77] D. P. Muni, N. R. Pal, J. Das, 2006, "Genetic programming for simultaneous feature selection and classifier design". IEEE Transactions on Systems, Man, and Cybernetics—PartB: Cybernetics, vol. 36 no. 1 pp. 106–117.

[78] X. Wang, J. Yang, X. Teng, W. Xia, R. Jensen, 2007, "Feature selection based on rough sets and particle swarm optimization". Pattern Recognition Letters, vol. 28 no. 4 pp. 459–471.

[79] D. Chakraborty, N. R. Pal., 2004, "A neuro-fuzzy scheme for simultaneous feature selection and fuzzy rule-based classification" IEEE Transactions on Neural Networks, vol. 15 no. 1 pp. 110–123.

[80] H. Peng, F. Long, C. Ding, 2005, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy". IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 8, pp. 1226–1238.

[81] Y. Sun, 2007, "Iterative RELIEF for feature weighting: algorithms, theories, and applications". IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 6, pp. 1035–1051.

[82] A. D. Back, T. P. Trappenberg, 2001, "Selecting inputs for modeling using normalized higher order statistics and independent component analysis". IEEE Transactions on Neural Networks, vol. 12, no. 3, pp. 612–617.

[83] Mao, K. Z. 2002, "Fast orthogonal forward selection algorithm for feature subset selection". IEEE Transactions on Neural Network, vol. 13 no. 5, pp. 1218–1224.

[84] Caruana, R., & Sa, V. D. 2003, "Benefitting from the variables that variable selection discards". Journal of Machine Learning Research, vol. 3, pp. 1245–1264.

[85] X. He, D. Cai, P. Niyogi, 2005, "Laplacian score for feature selection", in: Advances in Neural Information Processing Systems.

[86] K. Tsuda, M. Kawanabe, K. Muller, 2003, "Clustering with the Fisher Score", in: Advances in Neural Information Processing System, pp. 729–736.

[87] X. Tian, D. Tao, Y. Rui, 2012, "Sparse transfer learning for interactive video search reranking", in: Proceedings of the ACM Transactions on Multimedia Computing, Communications and Applications.

[88] Jun Yu, 2013, "Cartoon features selection using Diffusion Score", Elsevier, Signal Processing, vol. 93 pp. 1510–1520.

[89] Md. Monirul Kabir, Md. Monirul Islam, Kazuyuki Murase, 2010, "A new wrapper feature selection approach using neural network". Neurocomputing, vol. 73, pp. 3273–3283

[90] S. Abe, 2005, "Modified backward feature selection by cross validation". Proceedings of the European Symposium on Artificial Neural Networks, pp. 163–168.

[91] E. Gasca, J. S. Sanchez, R. Alonso., 2006, "Eliminating redundancy and irrelevance using a new MLP-based feature selection method". Pattern Recognition, vol. 39, pp. 313–315.

[92] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, 2002, "Gene selection for cancer classification using support vector machines". Machine Learning, vol. 46, pp. 389–422.

[93] Härdle, Wolfgang; Simar, Léopold, 2007, "Canonical Correlation Analysis". Applied Multivariate Statistical Analysis. pp. 321–330.

[94] Sanger, T. 1990, "Analysis of the two-dimensional receptive fields learned by the generalized hebbian algorithm in response to random input". Biological Cybernetics, vol. 63, pp. 221-228.

[95] Fyfe, C., & Baddeley, R., 1995, "Nonlinear data structure extraction using simple hebbin networks". Biological Cybernetics, vol. 72, no. 6, pp. 533-541.

[96] Charles, D., & Fyfe, C. 1998, "Modelling multiple cause structure using rectification constraints". Network: Computation in neural systems, vol. 9, pp. 167-182.

[97] Mardia, K. V., Kent, J., & Bibby, J. 1979, "Multivariate analysis". New York: Academic Press

[98] H. Hotelling., 1936, "Relations between two sets of variates". Biometrika, vol. 28, pp. 321–377.

[99] R. B. Cattell, 1943, "The description of personality: Basic traits resolved into clusters," Journal of Abnormal and Social Psychology, vol. 38, pp. 476-506.

[100] Bailey, Ken., 1994, "Numerical Taxonomy and Cluster Analysis". Typologies and Taxonomies. pp. 34.

[101] M. Misiti, Y. Misiti, G. Oppenheim, J. M. Poggi., 2007, "Clustering Signals Using Wavelets". Computational and Ambient Intelligence, Lecture Notes in Computer Science, vol. 4507, pp. 514-521.

[102] J.Z. Xue, H. Zhang, C.X. Zheng, 2003, "Wavelet packet transform for feature extraction of EEG during mental tasks". Presented at Proceedings of the Second International Conference on Machine Learning and Cybernetics, Xi'an.

[103] Tom M. Mitchell, 1997, "Machine Learning". McGraw-Hill Science/Engineering/Math, pp. 93

[104] Arthur Earl Bryson, Yu-Chi Ho, 1969, "Applied optimal control: optimization, estimation, and control". Blaisdell Publishing Company or Xerox College Publishing, pp. 481.

[105] Paul J. Werbos, 1974, "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences". PhD thesis, Harvard University.

[106] J. Pan, X. Gao, F. Duan, Z. Yan and S.Gao, 2011, "Enhancing the classification accuracy of steady-state visual evoked potential-based brain-computer interfaces using phase constrained canonical correlation analysis". J. Neural Eng. vol. 8, no. 3, pp. 036027.

[107] J. Kettenring, 1971, "Canonical analysis of several sets of variables". Biometrika vol. 53, no. 3, pp. 433–451.