EEG SIGNALS ANALYSIS

SWISS FEDERAL INSTITUTE OF TECHNOLOGY,

By Sameer Kumar,

Indian Institute of Technology,
Kanpur , India.

Project Guide:Gary.N.Garcia

Professor: Touradj Ebrahimi

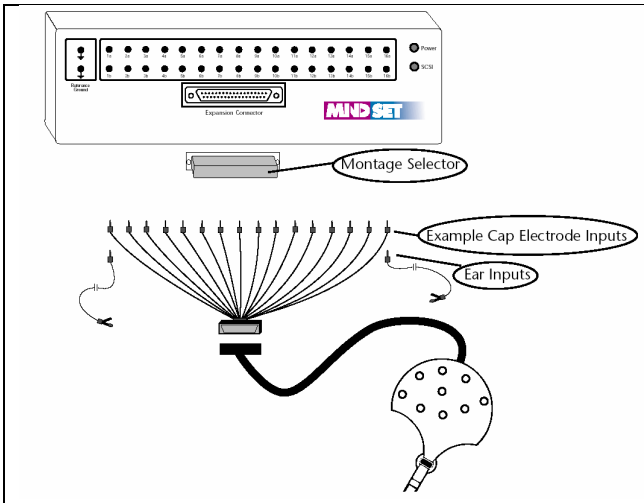
INTRODUCTION

Human Computer Interface has been a growing field of research in recent years. Uptil now, voice and vision have been the key interfaces. Direct *Brain-Computer Interface* is a novel direction in this research. Most exciting BCI's are based on *EEG (Electroencephalogram)*.

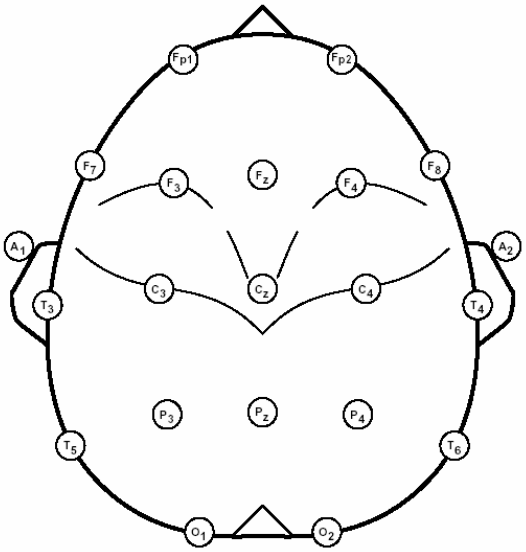
This report deals with the analysis of EEG signals i.e. acquisition, cleaning, feature extraction, and training. Section 1 deals with the *Acquisition of EEG signals* using the *Mindset*. Section 2 contains the principles involved in the *Preprocessing of Signals*. Section 3 presents methods to analyze the signals in order to extract discriminatory *Features*. Section 4 contains the details of the *Experiment*. Section 5 deals with the *Identification of Signals* corresponding to particular mental activities. Section 6 presents the results obtained using various methods for different subjects. Section 7 contains the results with *Alpha Rhythms* for different subjects.

1.Acquisition of EEG signals

The EEG signals are acquired using a **Mindset**. The device is shown in **Figure.1.**Sixteen electrodes namely : 'Fp1' , 'Fp2' , 'F7' , 'F3' , 'F4' , 'F8' , 'T3' , 'C3' , 'C4' , 'T4' , 'T5' , 'P3' , 'P4' , 'T6' , 'O1' , 'O2' are used according to the International Standard as shown in **Figure. 2.**

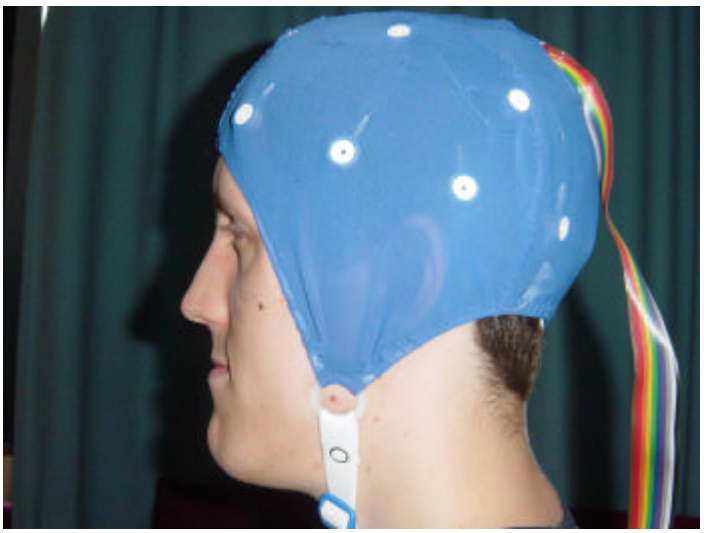


DATA ACQUISITION DEVICE
THE MINDSET
Figure 1



INTERNATIONAL SYSTEM
FOR ELECTRODES
Figure. 2

The subject is asked to wear a *Electrode Cap* through which the signals are acquired. The Cap is shown in **Figure. 3**



A subject wearing an Electrode Cap
Figure. 3

2. PREPROCESSING OF SIGNALS

2.1 Removal of Power Frequency

The power noise (50 Hz) is present in the signals obtained from the *Mindset*. We use a notch filter to remove this noise.

2.2 Removal of Muscular Artifacts

The EEG signals obtained contain perturbations like muscular artifacts. These muscular artifacts are characterized by high frequencies (20 Hz) and high amplitudes. These artifacts are prominent in the electrodes `FP1' and `FP2'. **Figure.4** shows a typical EEG signal containing an artifact, and a binary signal that identifies the artifacts. The artifacts are identified by differentiating the trials in the frequency domain using the K-Means algorithm. These artifacts can also be detected using the amplitudes as the discriminating factor. But this makes the discrimination too sensitive.

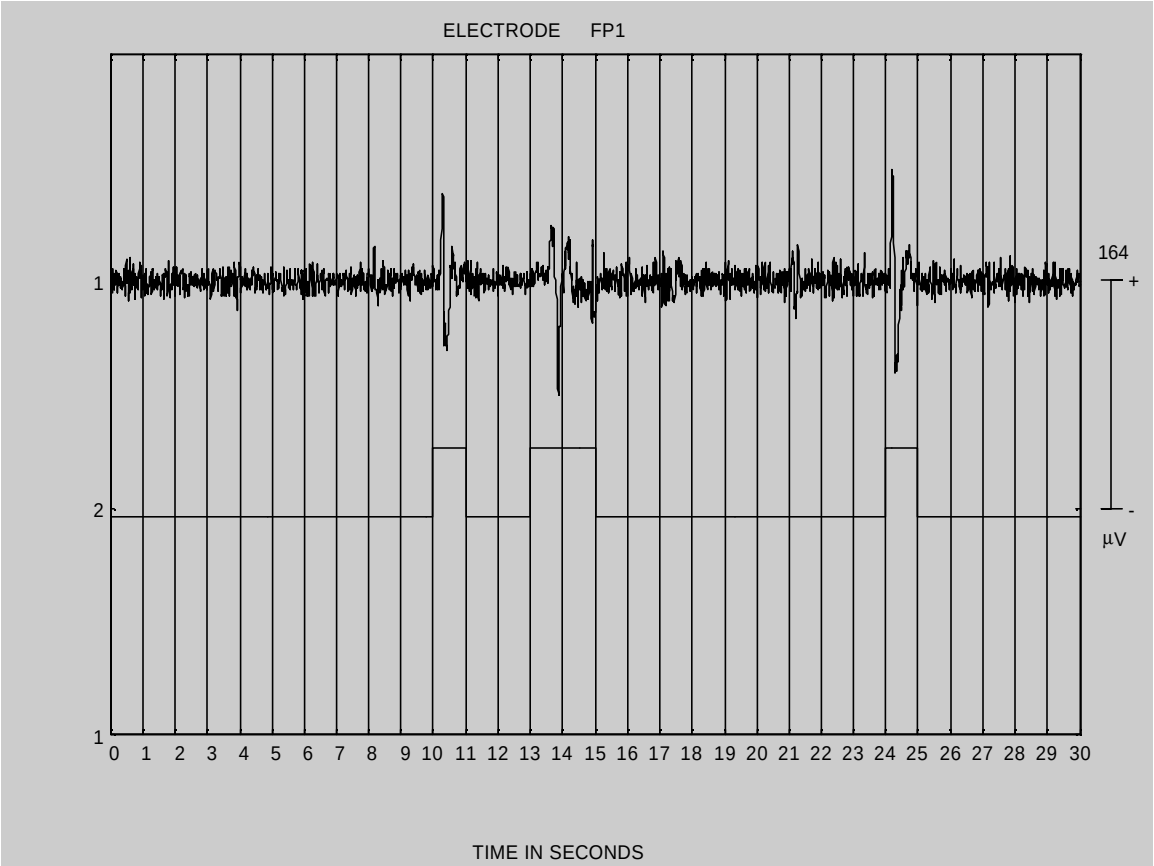


Figure. 4

3. FEATURE EXTRACTION

3.1 Extracting Spatial Features

Our aim is to identify a single mental activity of a subject from his rest state. For this, we use the spatial features of the mental activity. Following method is used to extract the spatial features: Consider a particular mental activity , say , Mental Counting. It is recorded as a matrix **MA1**. Let the rest state signals be recorded as **MA0**.We use 16 electrodes for our readings. Each trial is 4 seconds long. Hence, with a sampling frequency of 256 Hz **MA1** is a [16 × 1024 ×No. of Trials] matrix. The dimensions of **MA0** are chosen to be same as **MA1** .Frequency bands :1-5 Hz, 5-9 Hz ….. 37-41 Hz are used .**MA1** and **MA0** are filtered for each band and the corresponding filtered signals are extracted. For each filtered signal, we take the mean-co-variance of the trials for both **MA1** and **MA0.** Suppose for a particular frequency band, the mean-co-variance is X_1 for **MA1** and X_0 for **MA0** We diagonalise the matrix $(X_1 + X_0)$ as

$$X_1 + X_0 = U^T D_c U$$

Now we form another matrix: $Y = D_c^{-0.5} U \; X_1 U^T D_c^{-0.5}$

and diagonalise it to get $Y = V^T D_v V$

The matrix

$$P_{BAND} = V^T D_c^{-0.5} U$$

gives a projection operator for **MA1** corresponding to a particular frequency band. Similarly , projection operators for other frequency bands are obtained. If the eigen values of all the diagonal matrices are placed in descending oreder, it can be proved mathematically that the first two, and the last two rows of the matrix P will play a dominant role.We will call these four rows of P as the 4 components of the matrix P. Now, given any mental activity, say **MAk** , we can extract its features in the projection space of **MA1** .This is done as

$$F = P_{BANDi} MAk_{BANDi,TRIALj}$$

(where $MAk_{BANDi,TRIALj}$ is the j^{th} trial of **MAk** filtered for the i^{th} band.)

The matrix F obtained gives a feature vector of a trial for a particular frequency band. We can combine these vectors to form a matrix which is the *Spatial Feature* extracted from **MAk** into the projection space of **MA1.** Further , we try to train our algorithm to discriminate between the Spatial feature of **MA1** and the Spatial Feature of **MA0** , both obtained using the Projection P corresponding to **MA1.**

3.2 Extracting Frequency Features

In this approach , we try to identify ***MA1*** from ***MA0*** by using their frequency spectrum. 'FP1' and 'FP2' are just used for detecting the artifacts and are omitted in the analysis. Hence, we just use the last 14 electrodes .In all, 10 frequency bands :1-5 Hz, 5-9 Hz …. 37-41 Hz are used. The mean power in each electrode is calculated for every trial corresponding to each frequency band. Hence we have [14 × 10 × Number of trials] elements to form a *Feature*. This way, we form the feature vectors corresponding to **MA1** and **MA0.**

The average potential of Human brain is zero. So for every sample, we remove the mean potential of the 16 electrodes from each sample. This is known as the Removal of Spatial Mean. But results obtained after the removal of Spatial Mean are not good. So removal of spatial mean is not used in the analysis.

4. THE EXPERIMENT

Subjects were asked to perform a single Mental Activity - *Mental Calculation*. Each subject performed for a session of 15 minutes. The protocol used in a session is shown in **Figure. 5**

Figure. 5

Each session is divided into 3 five-minute slices. These five minute-slices were divided into rest and activity periods. The visual cues used to interact with the user are shown in *Figures*. Each 5 minute slice consists of 30 trials. During each trial, a three digit number is displayed for one second. Then the user performs Mental Calculation **MA1** for 4 seconds in which he subtracts 3 from the number displayed, and continues this with the result obtained. Then the rest state **MA0** starts with a `STOP' displayed for one second. The total duration of **MA0** is randomised between 4 and 6 seconds. The first one-second segment is omitted from **MA1** and **MA0** because of the influence of the visual evoked potentials.

5. IDENTIFYING SIGNALS

Now, our aim is to identify a Mental Activity (Mental Counting) against the rest state of the subject. As mentioned earlier, we first obtain the Spatial features of **MA1** and **MA0** using Projection P corresponding to **MA1** and then try to discriminate these *Features*.

We use a transformation F_H that maps the feature vectors obtained into a space H(of infinite dimension). F_H is defined through a Kernel function that is the internal product in H. The Kernel function used in our approach is the Gaussian Kernel.

If x and y belong to **R ,** the Gaussian Kernel is given by

$$K(x,y) = \exp(-\frac{x.x - 2x.y + y.y}{s^2})$$

The width parameter s is determined through cross-validation. The following procedure is used to train the algorithm. We take two *sets, target-set* and *non-target-set*. The target-set and non-target-set are respectively the `Feature Matrices' corresponding to **MA1** and **MA0** respectively. We take equal number of trials for both the activities for unbiased analysis. We assume that vectors belonging to the target-set are inside a sphere of radius R and centered in Ω (in the space H), whereas the vectors belonging to the non-target-set are outside this sphere. The values of R and Ω can be found by solving an optimization problem which consists in minimizing the radius R [1]. The penalization constants are determined by getting feedback from the subject regarding the estimated percentage of False Positives and False Negatives. The obtained penalization constants are adjusted to get the best results.

In order to apply Cross Validation we divide both the sets(the target-set and the non-target-set) into 10 groups each. Then , we choose 9 groups (at random) from each set and hence form a *training-set* consisting of 18 groups in all. The remaining groups from each set together form a *validation-set.* The parameter s is varied in a wide range logarithmically. For each s ,using the training-set, the optimal R and Ω are obtained. Then each element in the validation-set is identified as either inside or outside the sphere. Hence we obtain the percentage of *False Positives* and *False Negatives* present in the validation-set.

Next, we form another training-set , choosing different groups from target-set as well as non-target-set. The remaining two groups give a new validation-set. The above procedure of finding the sphere , and checking for errors in the validation-set is repeated. In all, we have 10 different training-sets and correspondingly 10 different validation-sets. Using the 10 curves obtained, we take the mean to obtain a final curve. The value of sigma that gives minimum mean validation-error is used to obtain a final sphere which is used to give feedback to the subject in the later sessions.

6. RESULTS

In all, seven male subjects , in the age group of 20-30 years, participated in the experiment. The results were analysed using Spatial Features and Frequency Features. In case of Spatial Features, the projection matrix P was calculated for each subject. Out of the four components of the matrix P(as mentioned earlier) , at first, only the first component was used and the results were analyzed. Then the results were analyzed using two, three and finally with all the four components. The topoplots for all the four components of the matrix P are obtained.

In order to show the behaviour of the components of the matrix P, we compute P for the following simulated data, and observe the topoplots.

Simulated Data:

M is a normally distributed random eeg trial with a sine wave of 10 Hz frequency added to the 11th electrode T5. N is another normally distributed random eeg trial. The matrix P is calculated using M as the target-data and N as the non-target-data. $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ The topoplot of all the four components of the matrix P for the frequency band 9-13 Hz is shown in **Figure. 6**

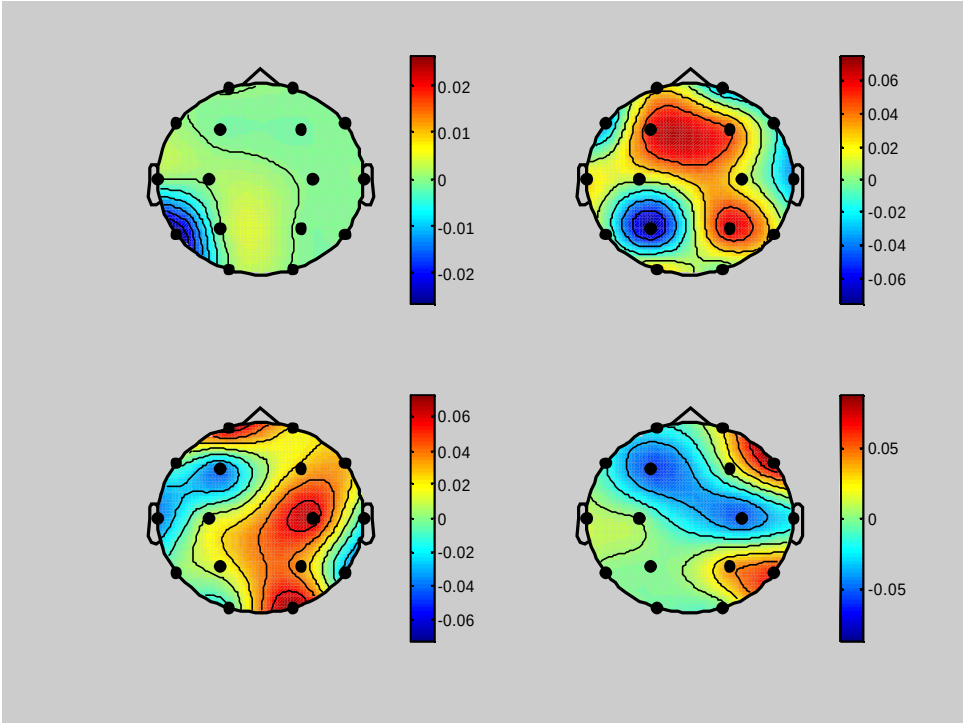


Figure. 6

It can be clearly seen that the first component of P is centred at the 11th Electrode T5.This can be interpreted as :if a subject's data gives best results with a single component of the matrix P, it is quite probable that only one electrode out of the sixteen, is responsible for the discrimination.

Moreover it was observed that for each subject, some electrodes were noisy. These electrodes were unique to a particular subject, and correspondingly each subject's data was analysed after removing the corresponding noisy electrodes.**Figure. 7** shows the signals of Subject 1.It can be seen that the electrodes C3 and C4 are noisy. The signals in these electrodes correspond to muscular perturbations of the brain.

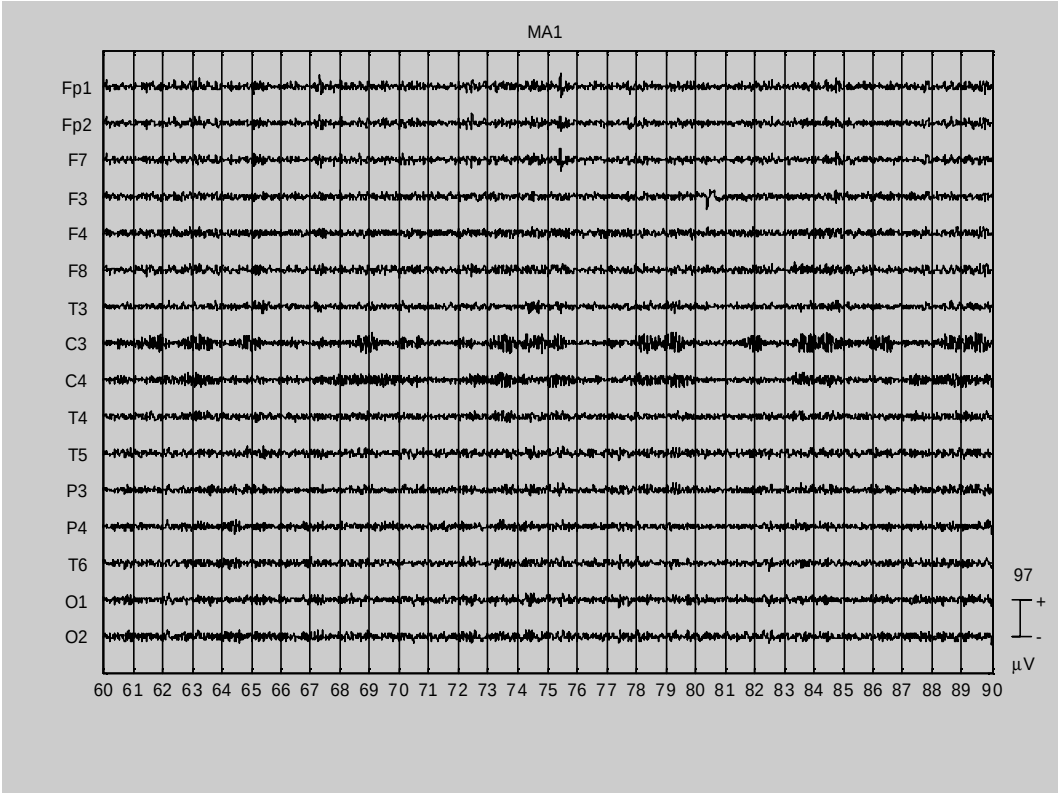
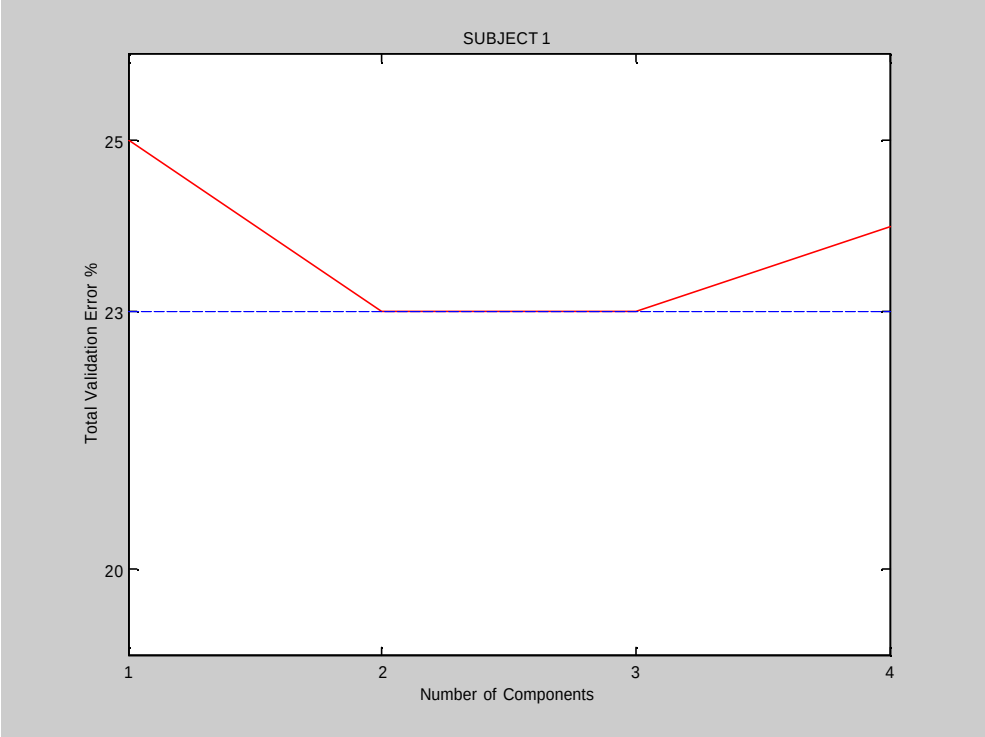


Figure. 7

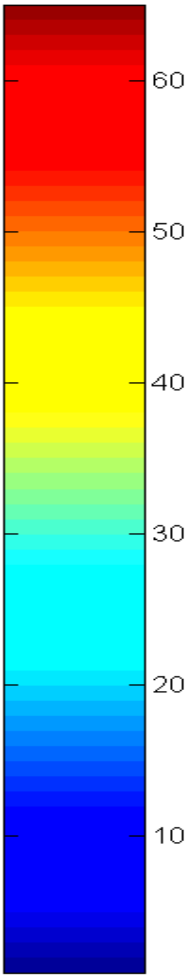
SUBJECT 1

The analysis of the matrix P

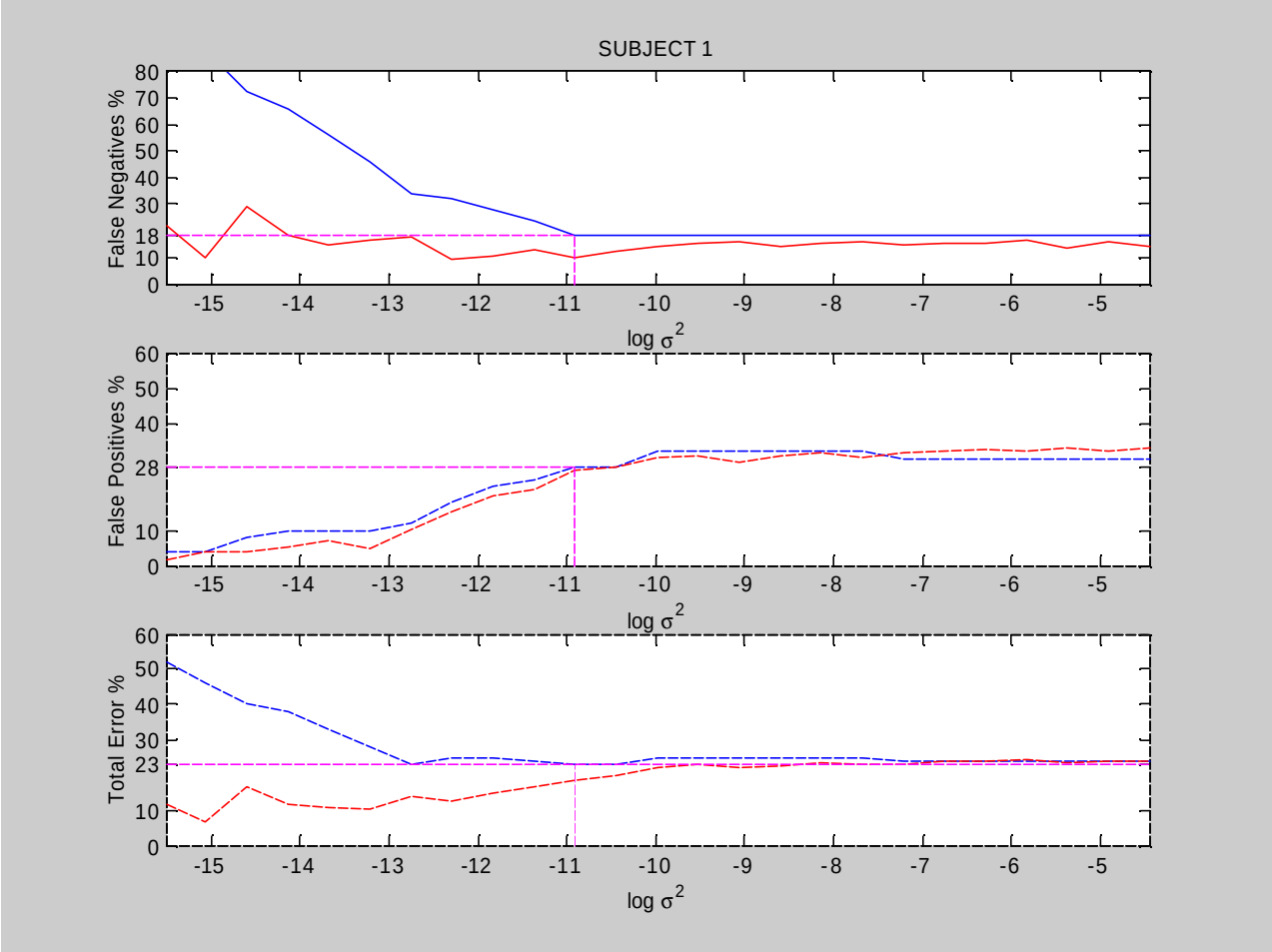


FREQUENCY BAND 1-5 Hz	FREQUENCY BAND 5-9 Hz
FREQUENCY BAND 9-13 Hz	FREQUENCY BAND 13-17 Hz
FREQUENCY BAND 17-21 Hz	FREQUENCY BAND 21-25 Hz
FREQUENCY BAND 25-29 Hz	FREQUENCY BAND 29-33 Hz
FREQUENCY BAND 33-37 Hz	FREQUENCY BAND -37-41 Hz

Topoplots showing the four coefficients of P for each Frequency Band



Result obtained using Spatial Features:



The 'Red' curve depicts the Training Error
The 'Blue' curve depicts the Validation Error

The above curve shows that best results are obtained with log s^2=-10.9 giving the minimum validation error. We show the variation of the distances of the targets and the non-targets in the validation set from the centre of the sphere , as a ratio between the distance and the radius of the sphere. **Figures D1, D2** and **D3** show the respective curves. The Red Curve represents the distances of the Targets from the centre, and the Blue Curve represents the distances of the Non-Targets from the centre.

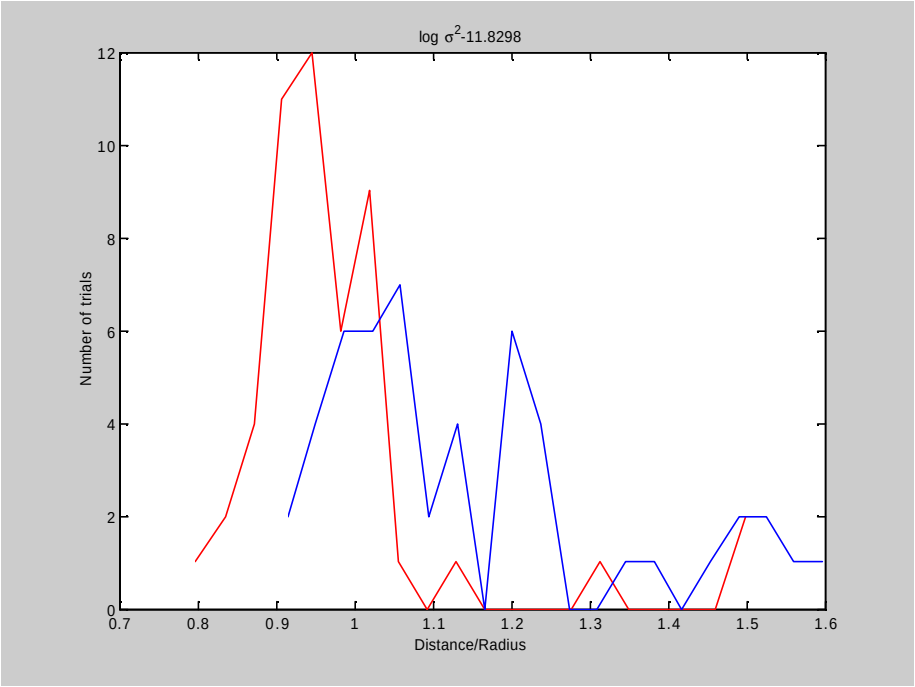


Figure D1

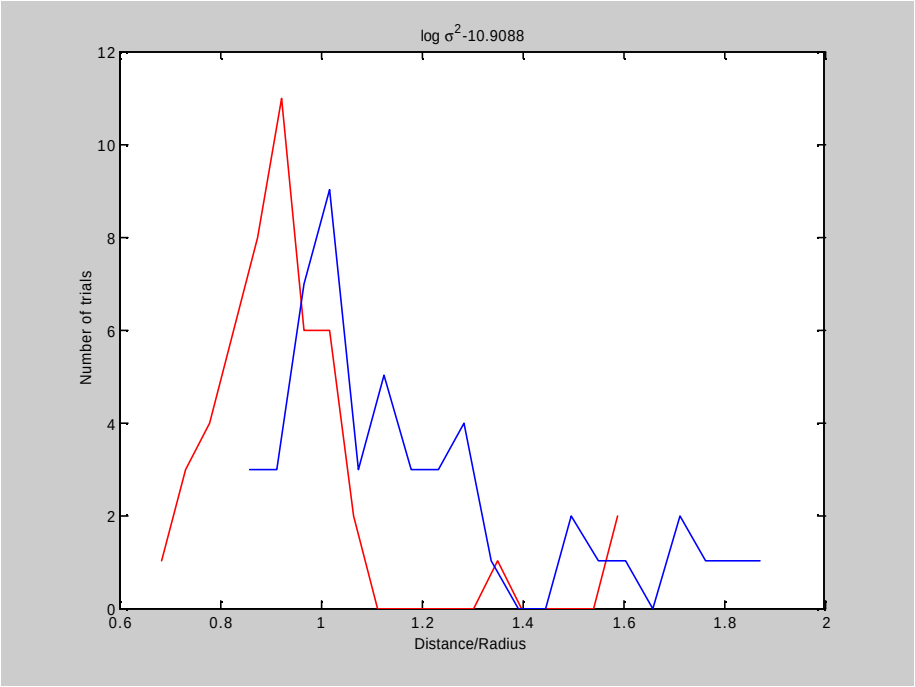


Figure D2

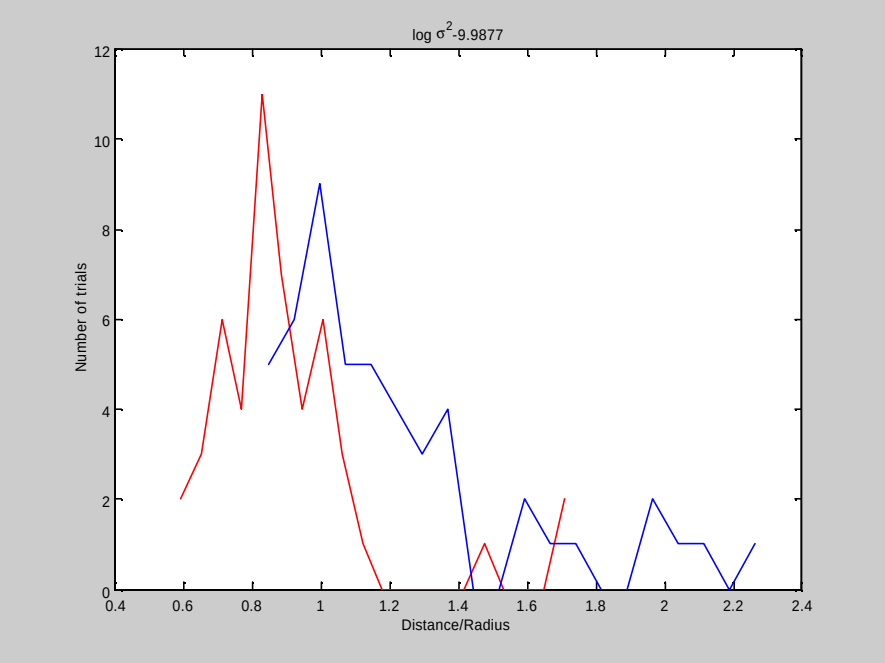
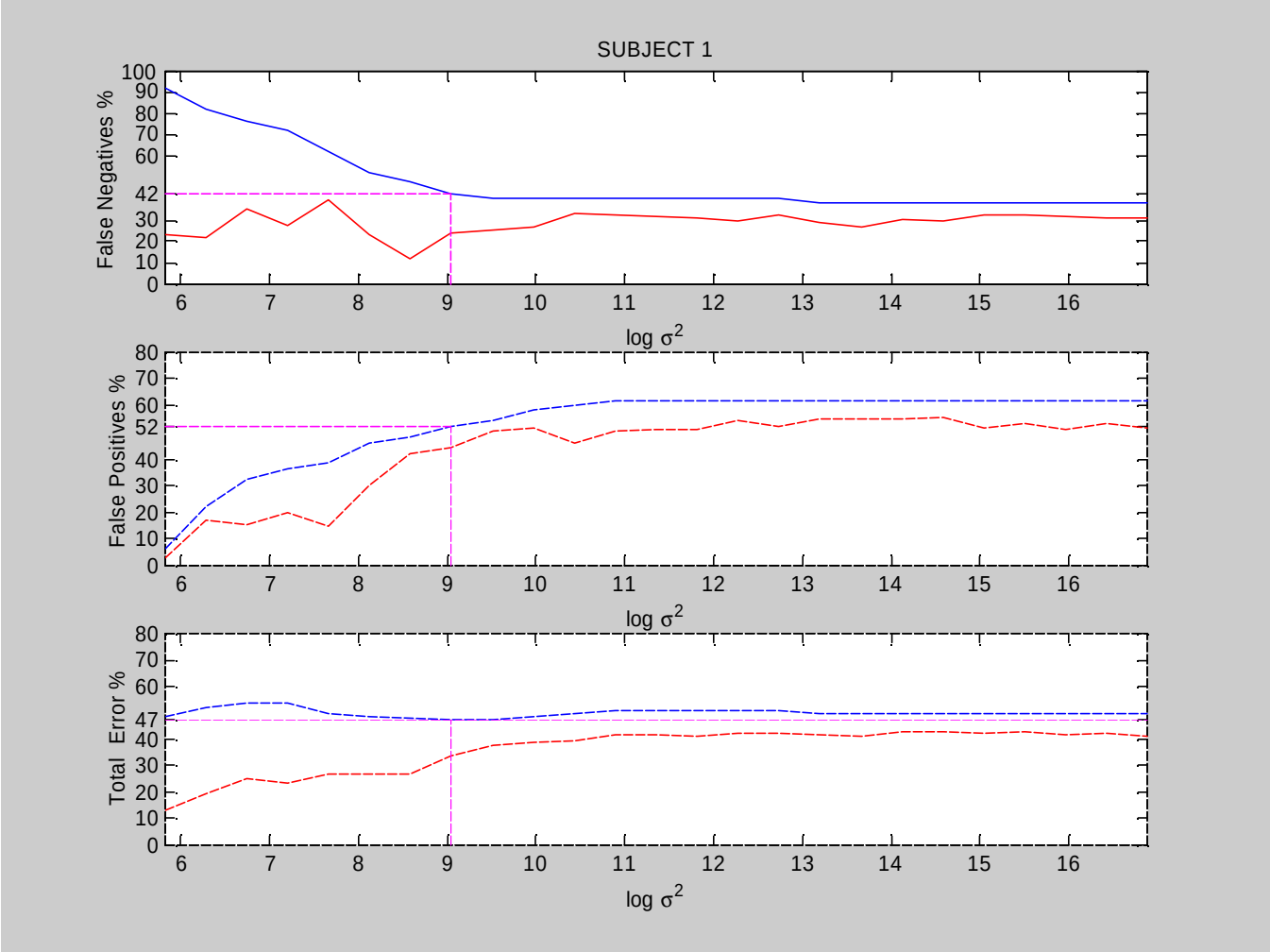

log σ2-9.9877
Number of trials
Distance/Radius
0
2
4
6
8
10
12
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
2.2
2.4

Figure D3

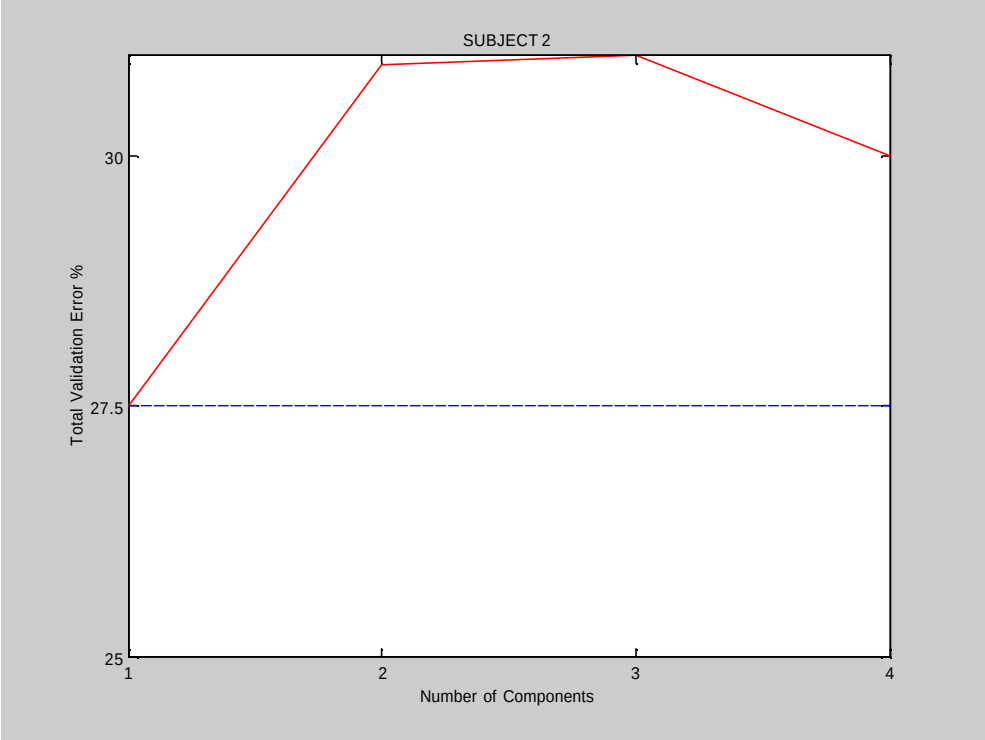
Results obtained using Frequency Features

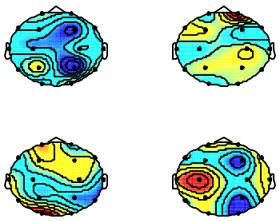
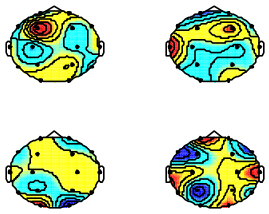
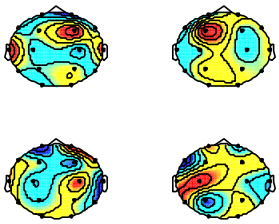
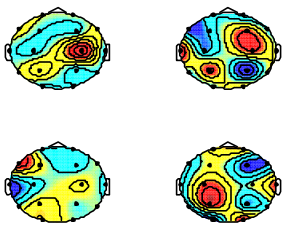
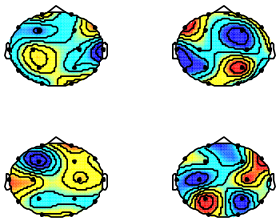
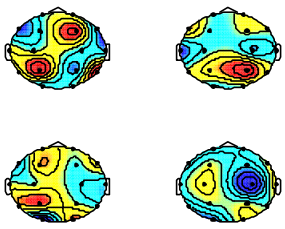
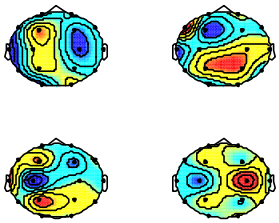
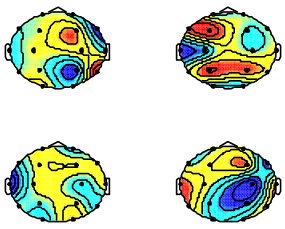
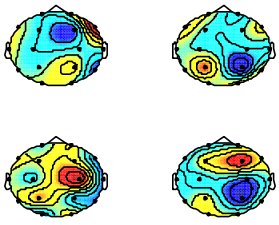
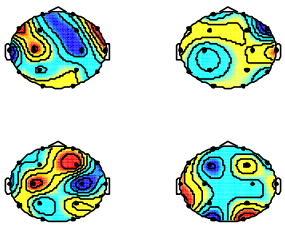


The ‘Red’ curve depicts the Training Error
The ‘Blue’ curve depicts the Validation Error

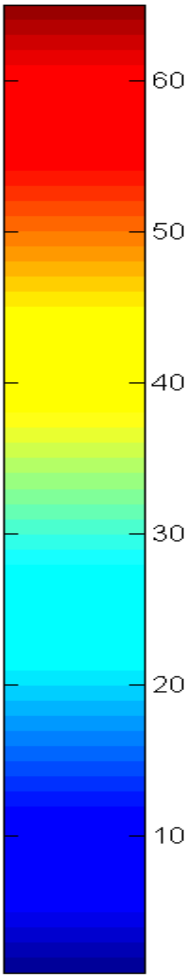
SUBJECT 2

The analysis of the matrix P

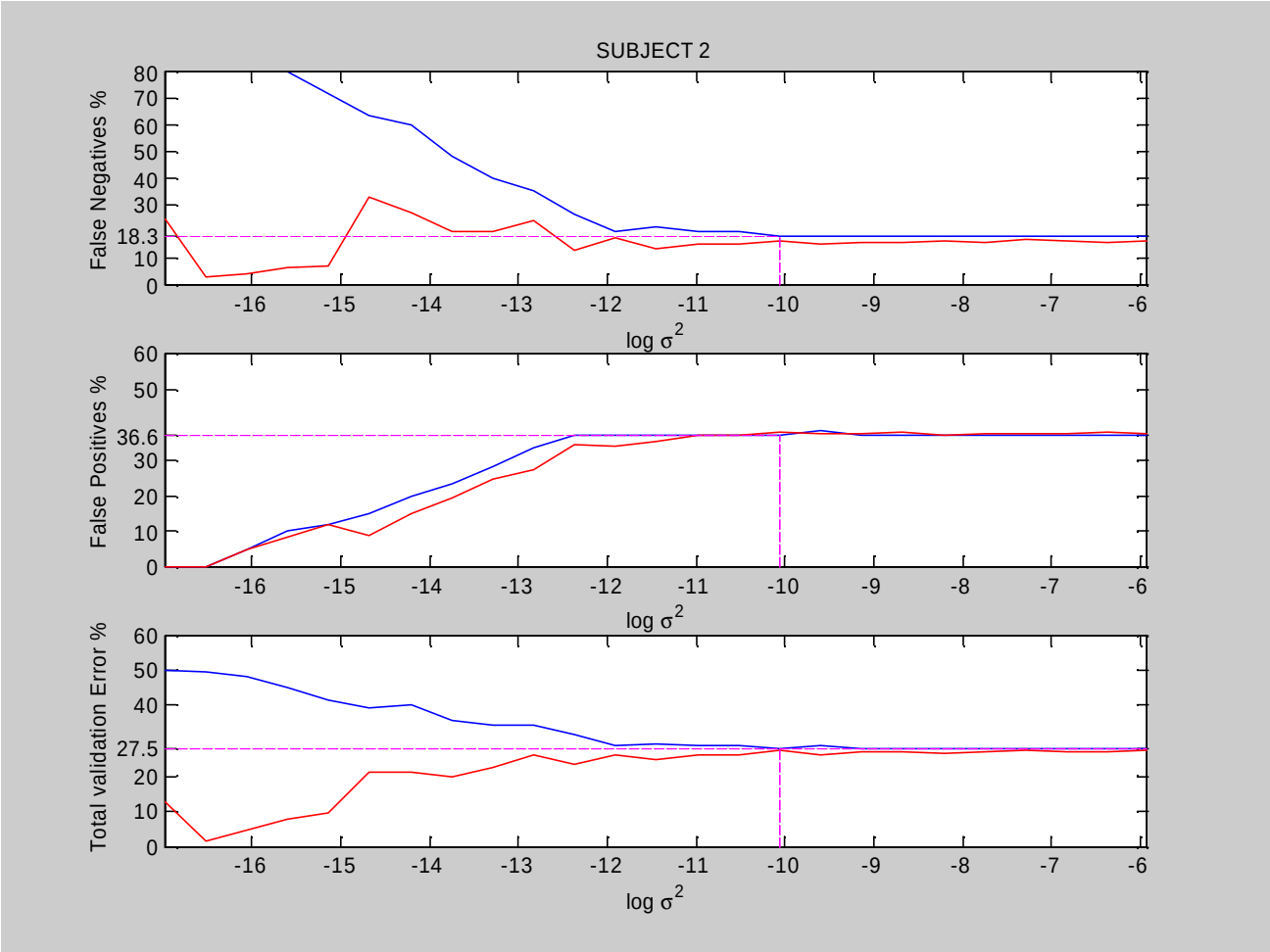


FREQUENCY BAND 1-5 Hz 	FREQUENCY BAND 5-9 Hz 
FREQUENCY BAND 9-13 Hz 	FREQUENCY BAND 13-17 Hz 
FREQUENCY BAND 17-21 Hz 	FREQUENCY BAND 21-25 Hz 
FREQUENCY BAND 25-29 Hz 	FREQUENCY BAND 29-33 Hz 
FREQUENCY BAND 33-37 Hz 	FREQUENCY BAND -37-41 Hz 

Topoplots showing the four coefficients of P for each Frequency Band

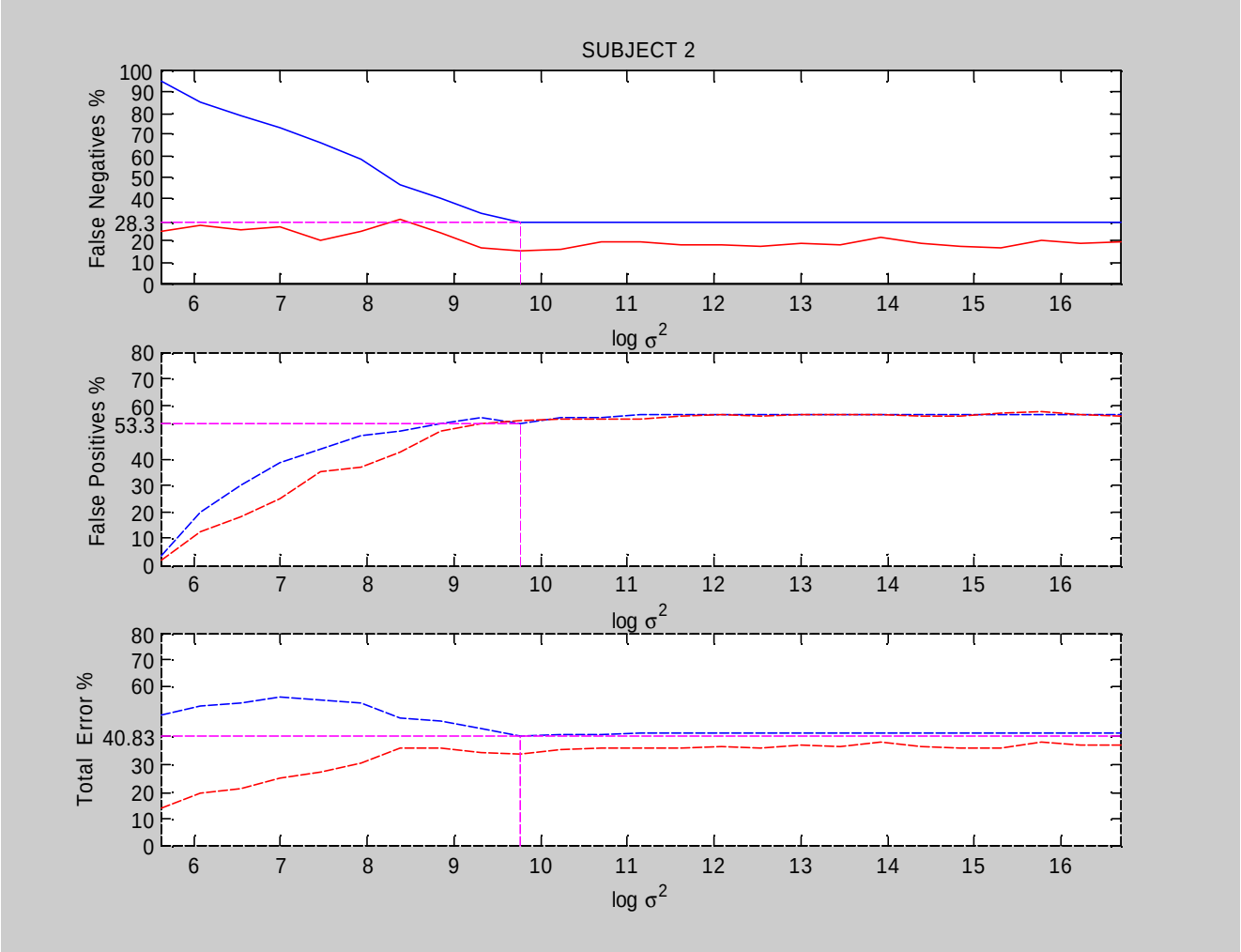


Result obtained using Spatial Features:



The ‘Red’ curve depicts the Training Error
The ‘Blue’ curve depicts the Validation Error

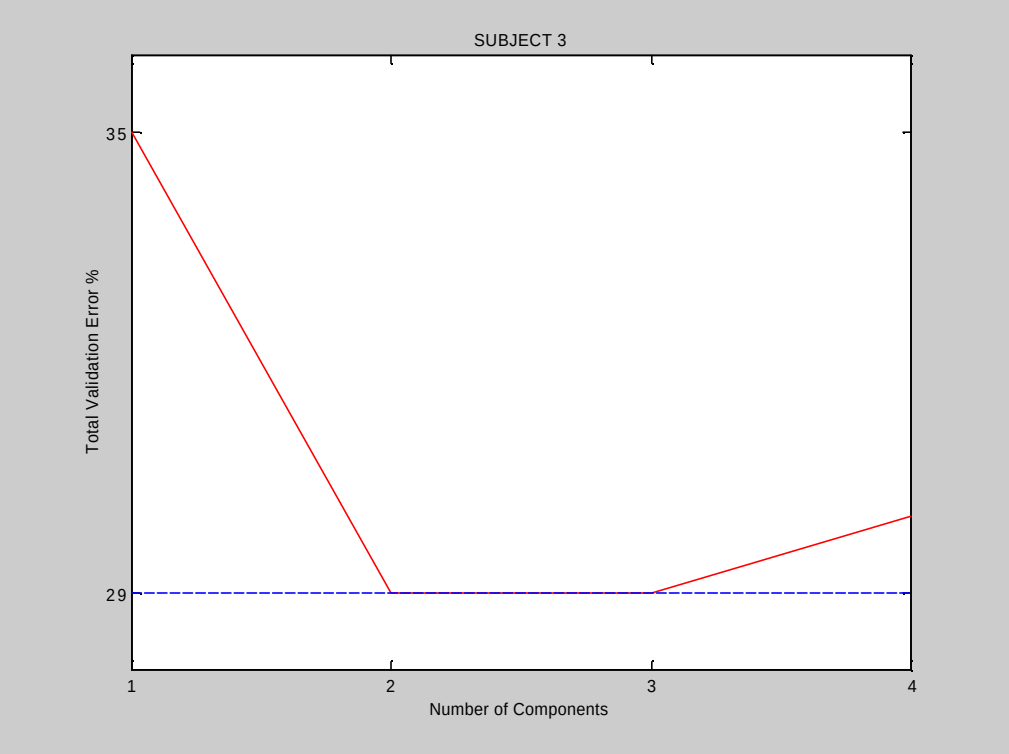
Results obtained using Frequency Features

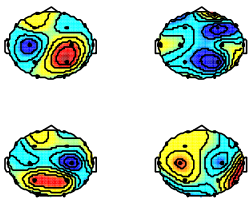
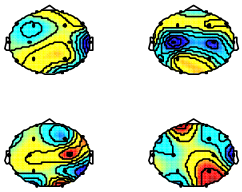
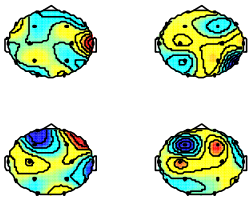
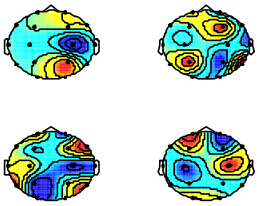
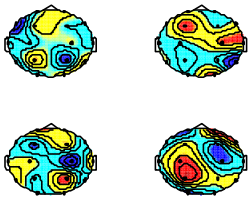
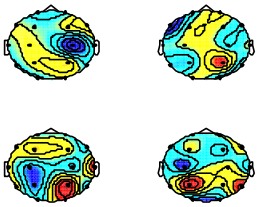
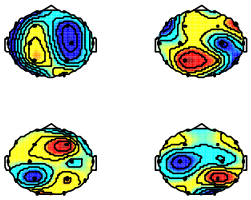
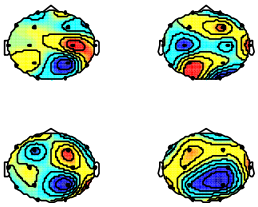
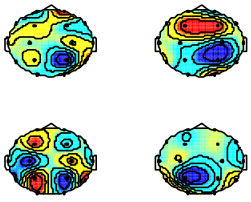
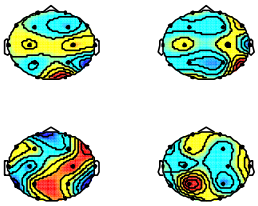


The ‘Red’ curve depicts the Training Error
The ‘Blue’ curve depicts the Validation Error

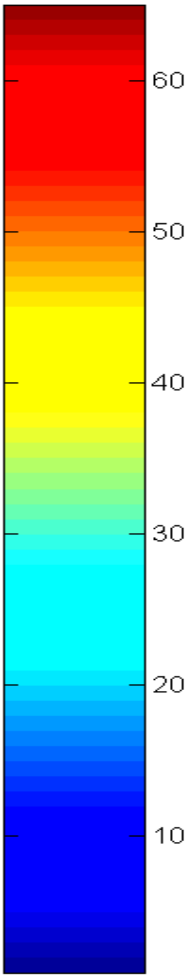
SUBJECT 3

The analysis of the matrix P

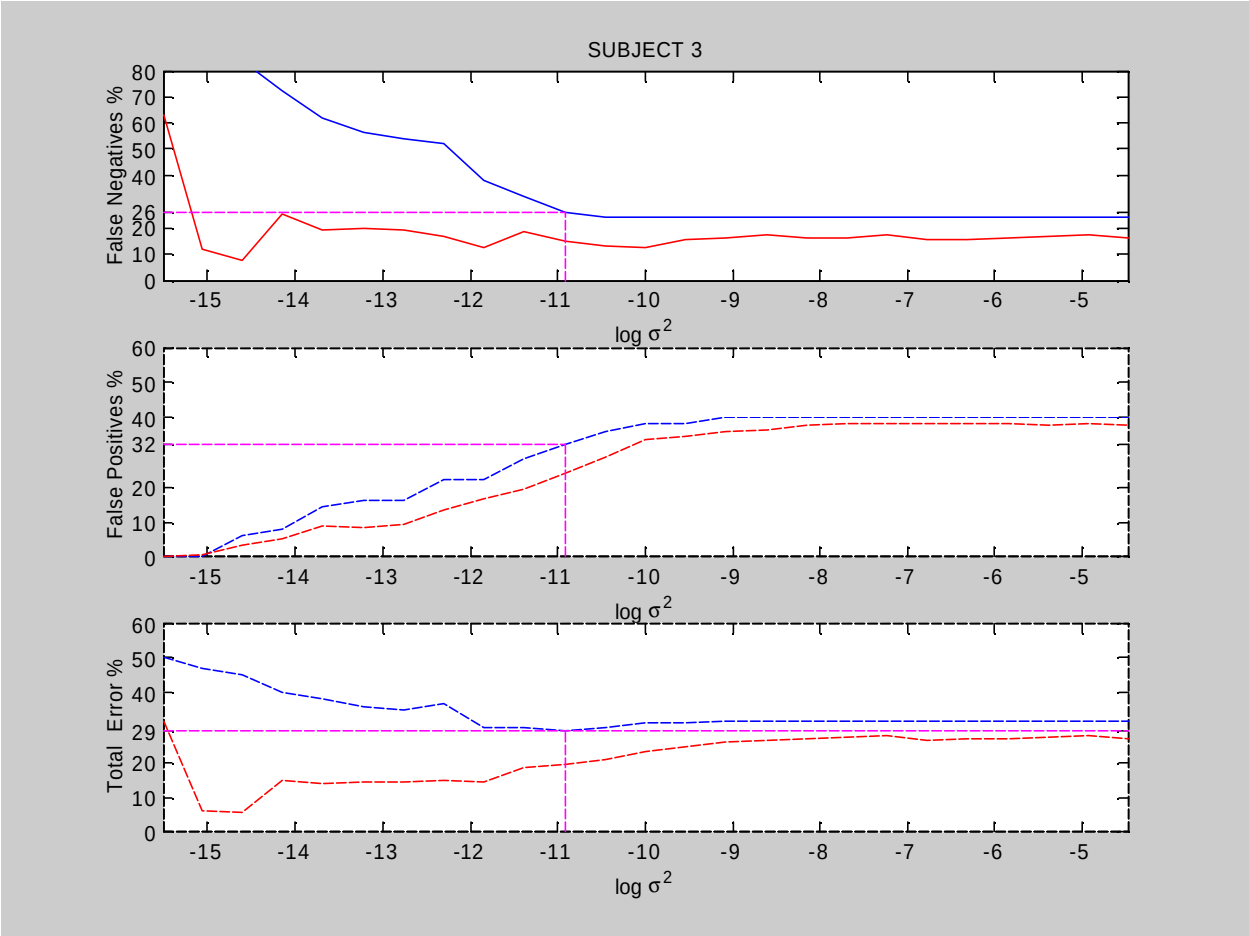


FREQUENCY BAND 1-5 Hz 	FREQUENCY BAND 5-9 Hz 
FREQUENCY BAND 9-13 Hz 	FREQUENCY BAND 13-17 Hz 
FREQUENCY BAND 17-21 Hz 	FREQUENCY BAND 21-25 Hz 
FREQUENCY BAND 25-29 Hz 	FREQUENCY BAND 29-33 Hz 
FREQUENCY BAND 33-37 Hz 	FREQUENCY BAND -37-41 Hz 

Topoplots showing the four coefficients of P for each Frequency Band

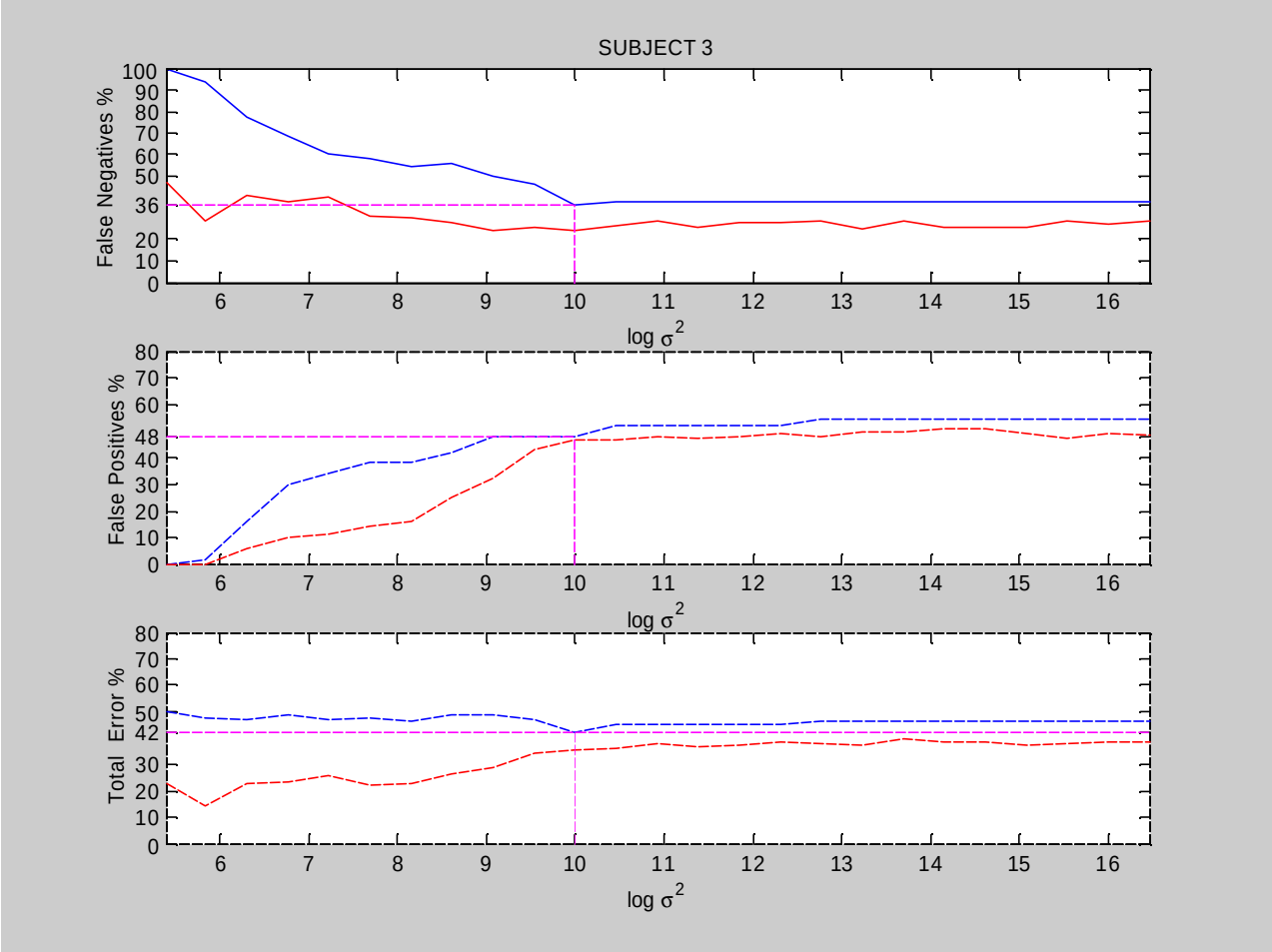


Result obtained using Spatial Features:



The ‘Red’ curve depicts the Training Error
The ‘Blue’ curve depicts the Validation Error

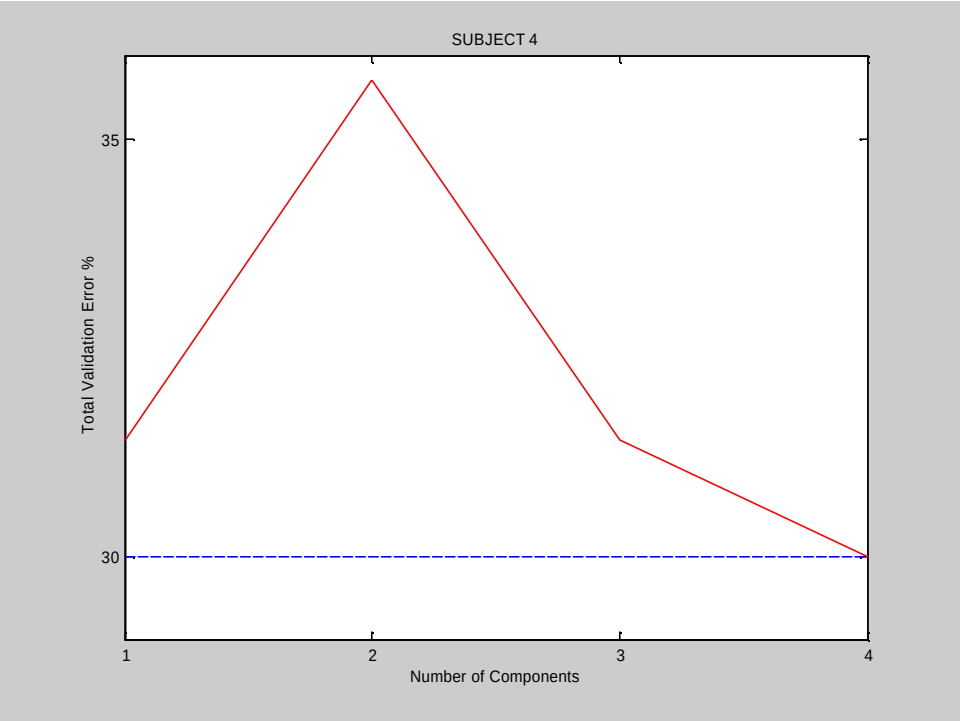
Results obtained using Frequency Features

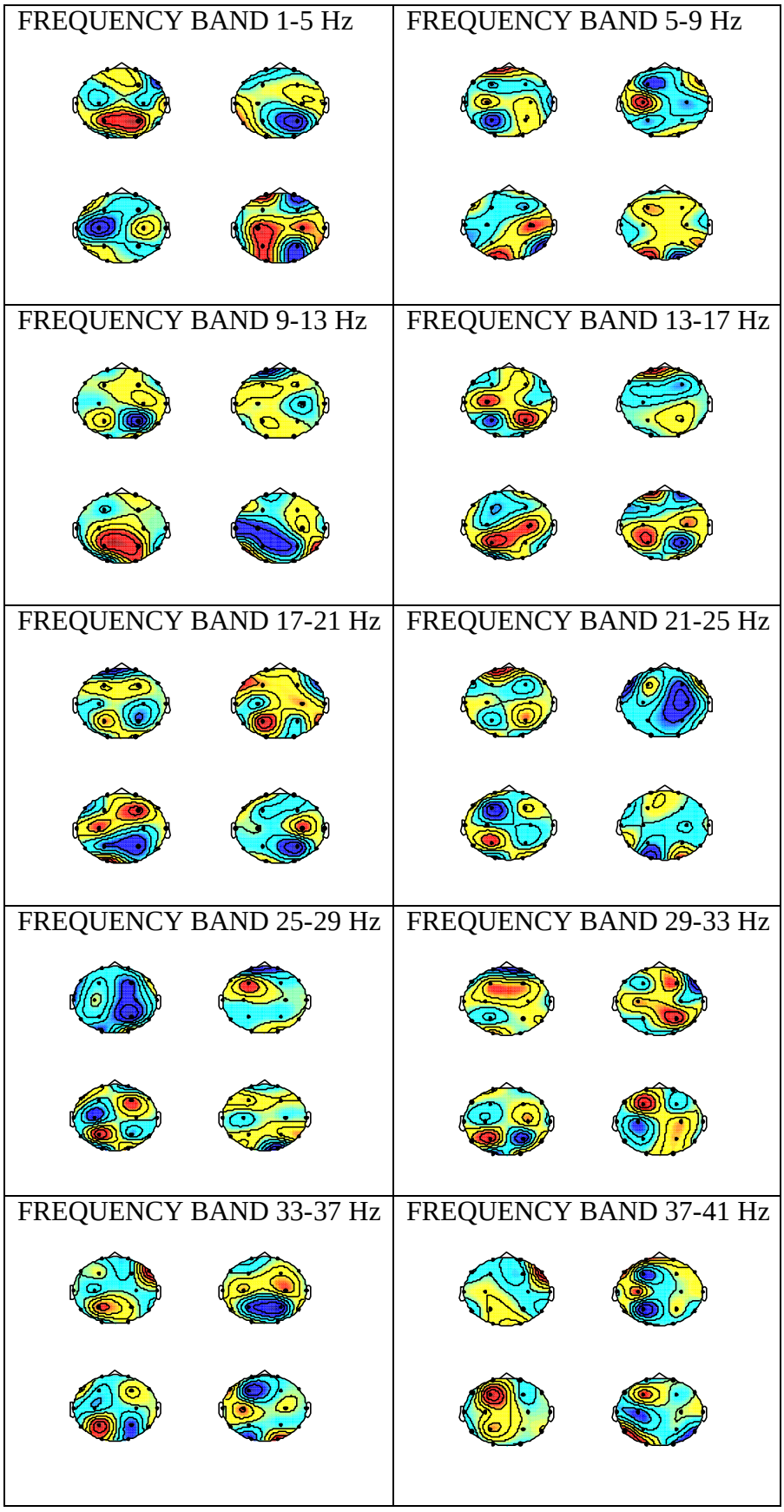


The ‘Red’ curve depicts the Training Error
The ‘Blue’ curve depicts the Validation Error

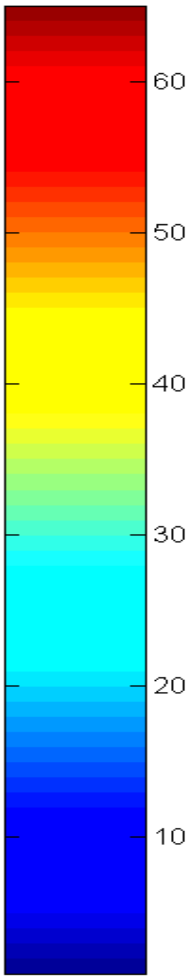
SUBJECT 4

The analysis of the matrix P

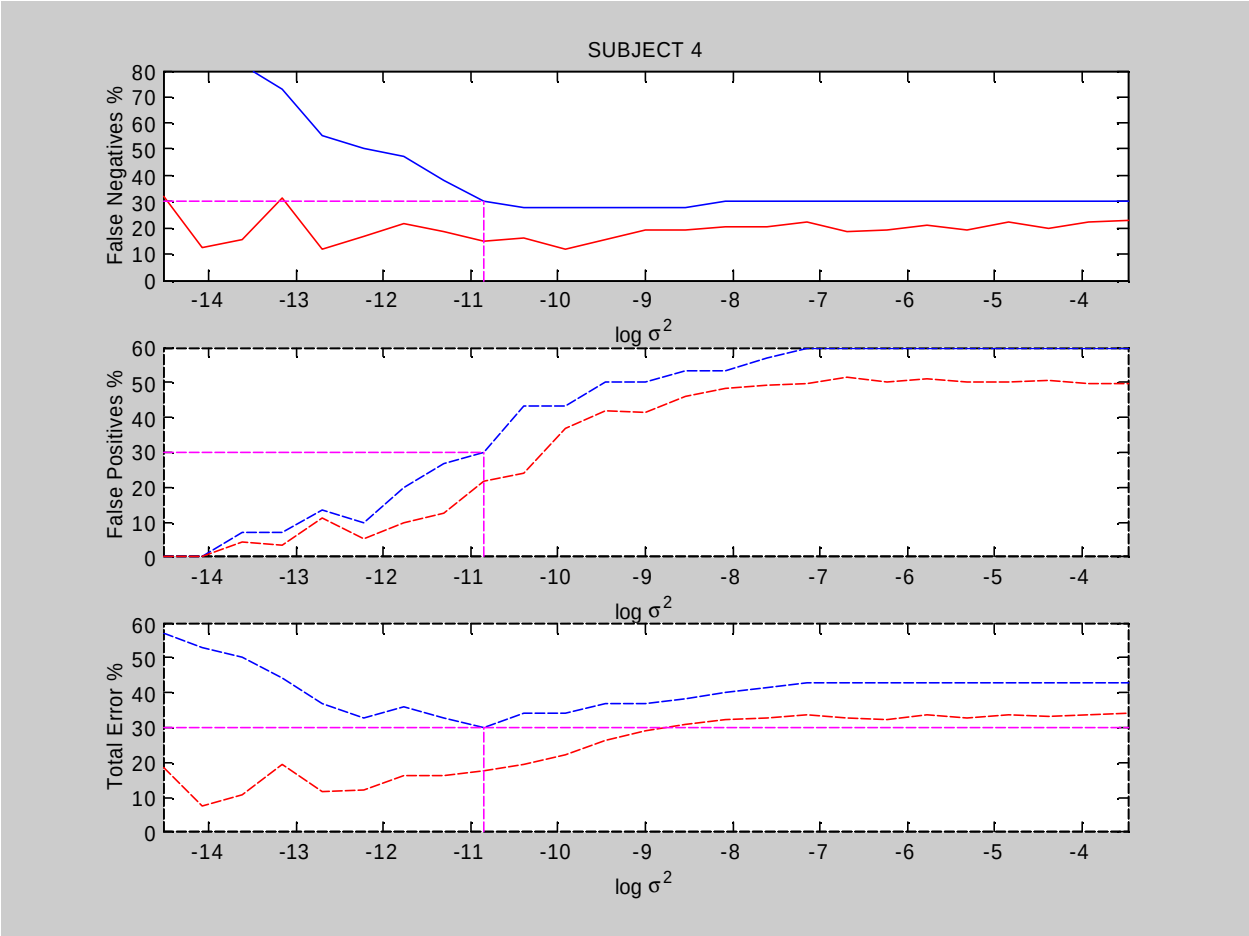




Topoplots showing the four coefficients of P for each Frequency Band

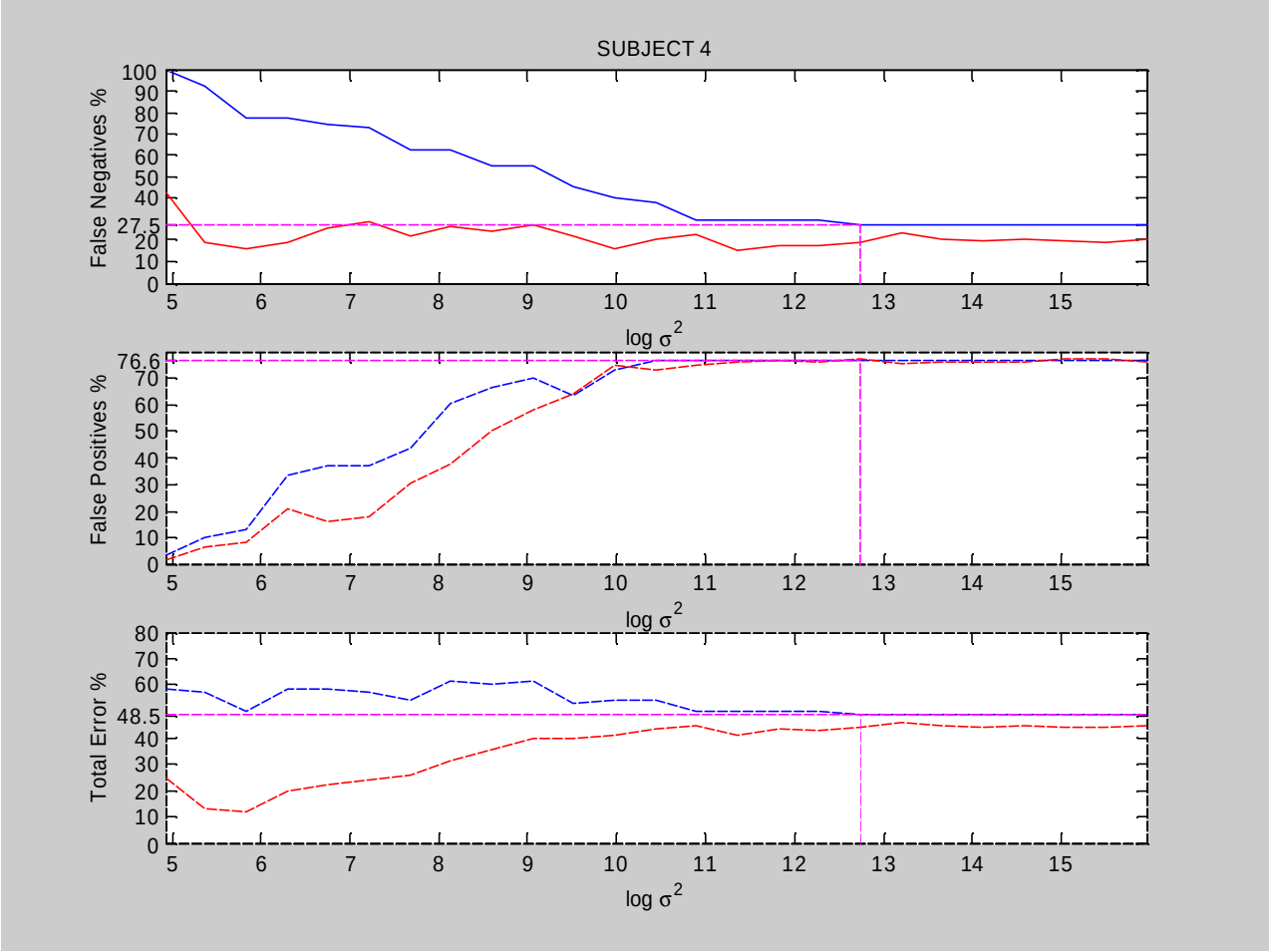


Result obtained using Spatial Features:



The ‘Red’ curve depicts the Training Error
The ‘Blue’ curve depicts the Validation Error

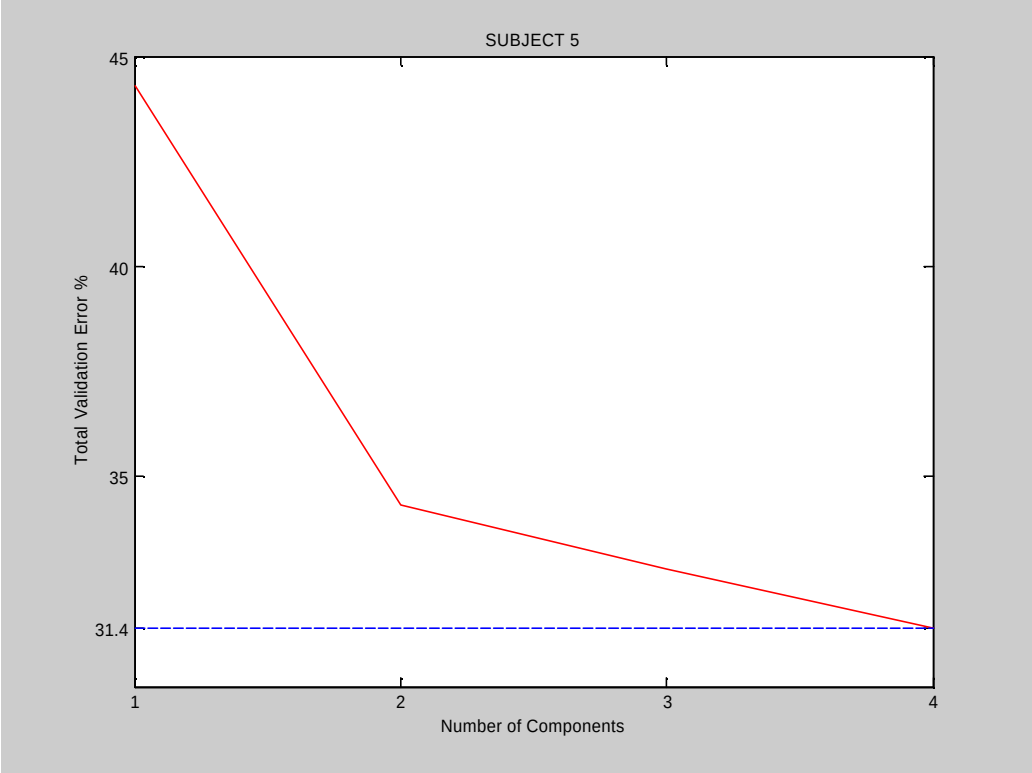
Results obtained using Frequency Features

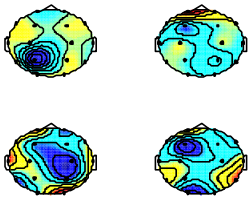
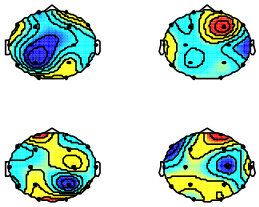
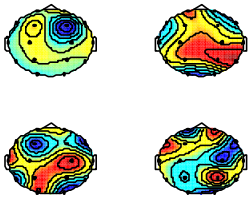
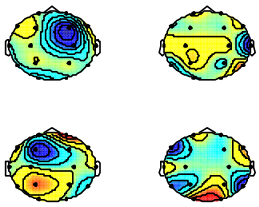
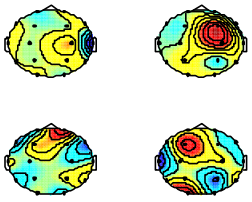
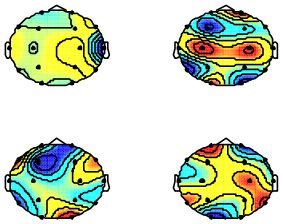
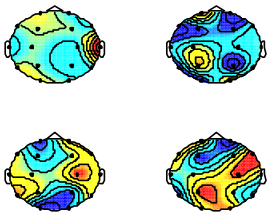
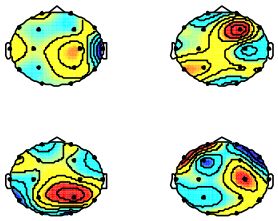
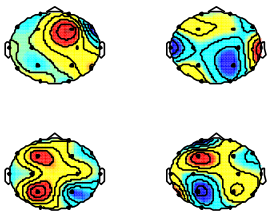
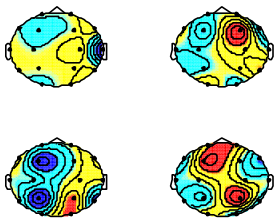


The ‘Red’ curve depicts the Training Error
The ‘Blue’ curve depicts the Validation Error

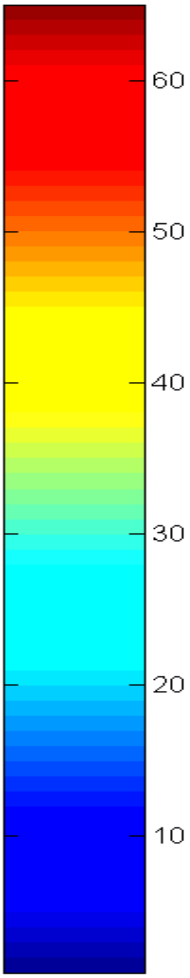
SUBJECT 5

The analysis of the matrix P

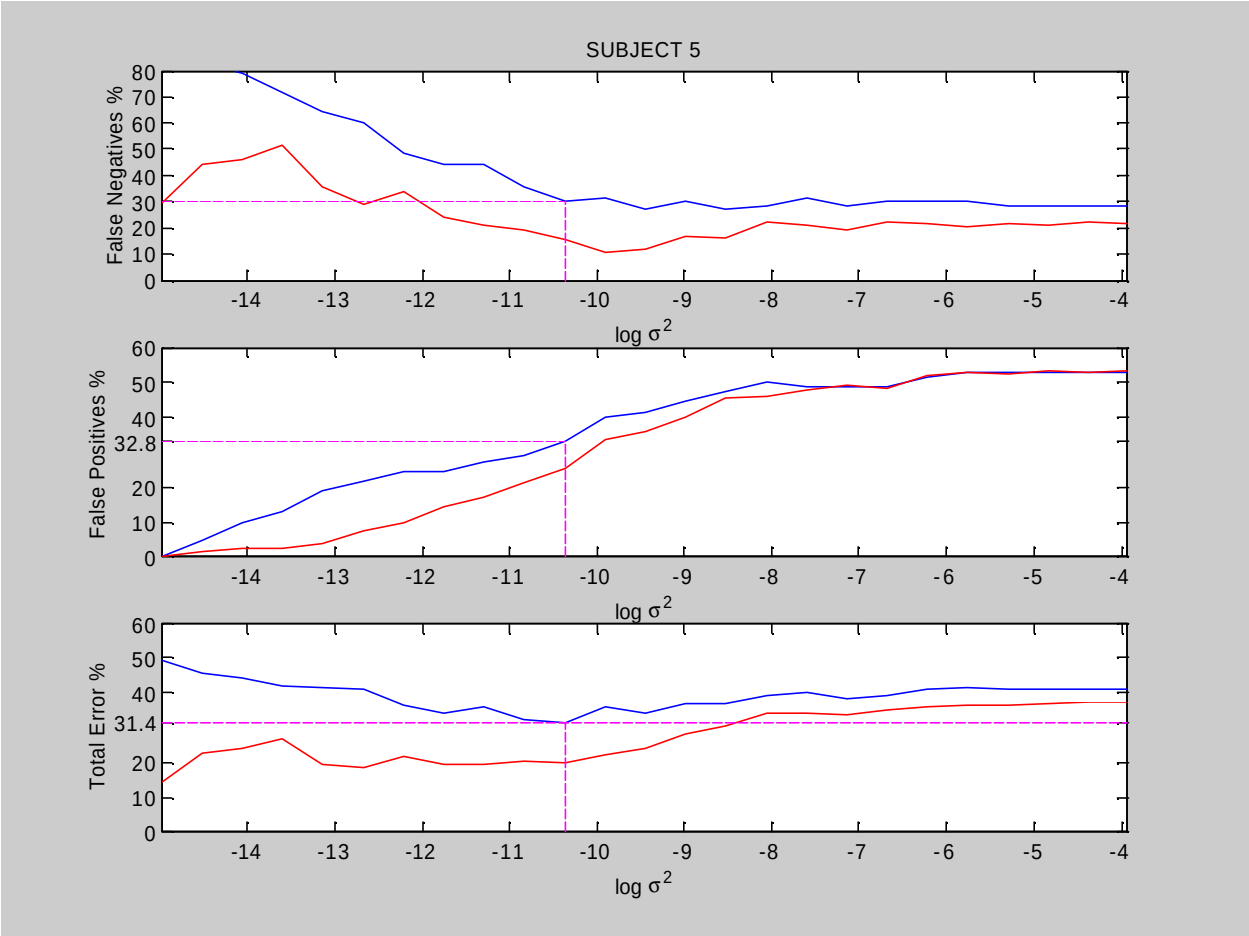


FREQUENCY BAND 1-5 Hz 	FREQUENCY BAND 5-9 Hz 
FREQUENCY BAND 9-13 Hz 	FREQUENCY BAND 13-17 Hz 
FREQUENCY BAND 17-21 Hz 	FREQUENCY BAND 21-25 Hz 
FREQUENCY BAND 25-29 Hz 	FREQUENCY BAND 29-33 Hz 
FREQUENCY BAND 33-37 Hz 	FREQUENCY BAND -37-41 Hz 

Topoplots showing the four coefficients of P for each Frequency Band

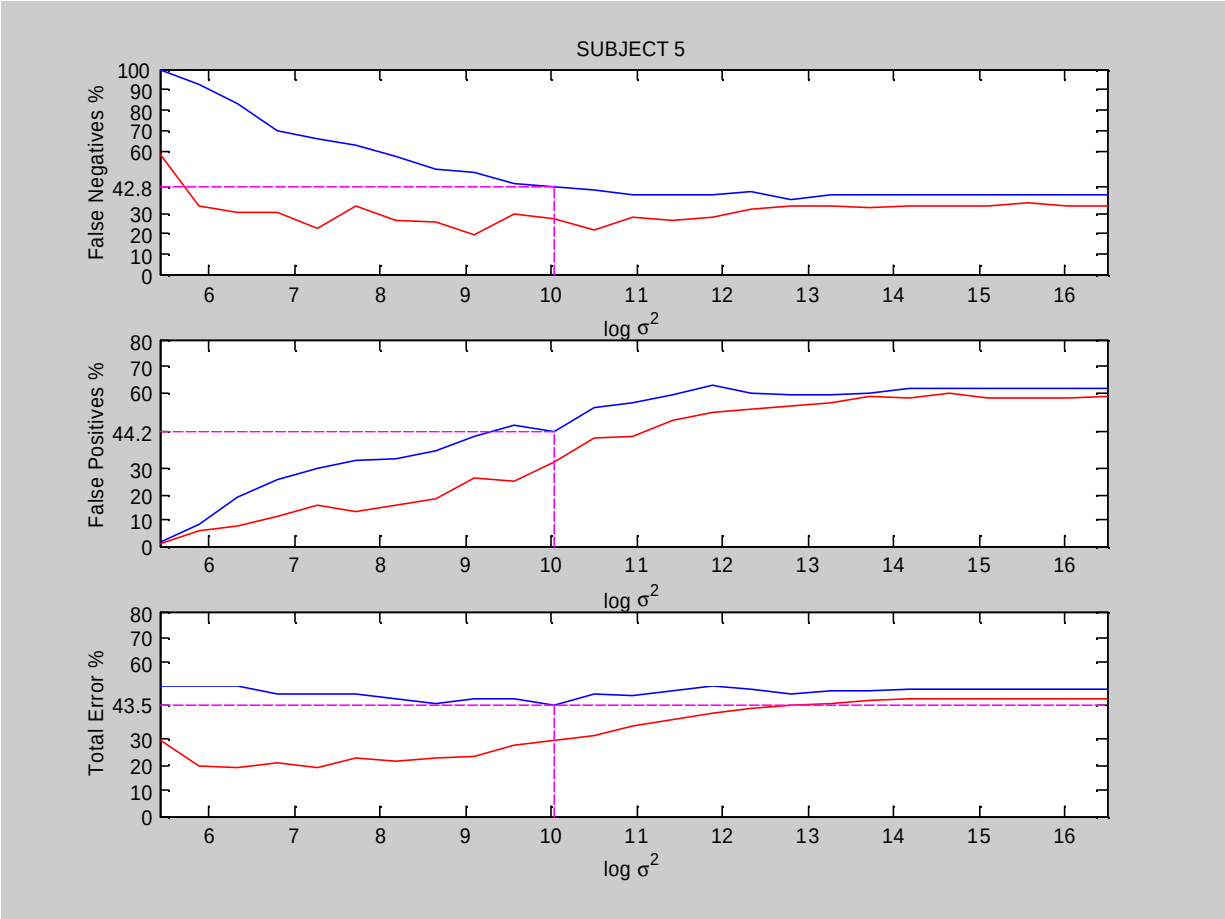


Result obtained using Spatial Features:



The ‘Red’ curve depicts the Training Error
The ‘Blue’ curve depicts the Validation Error

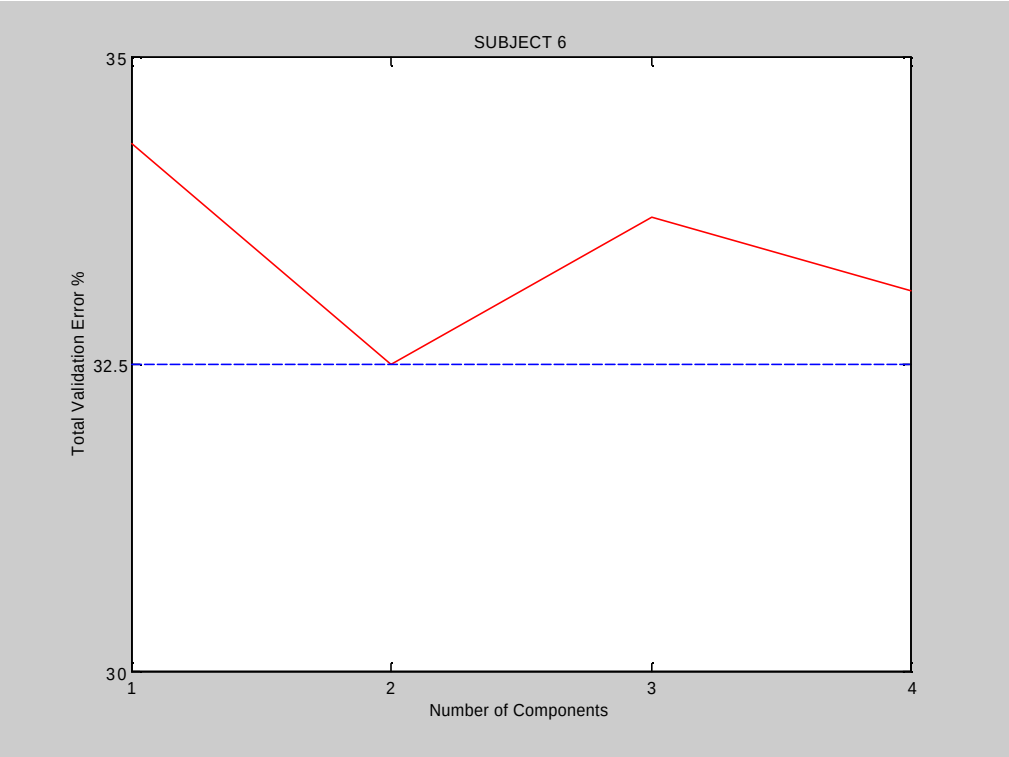
Results obtained using Frequency Features

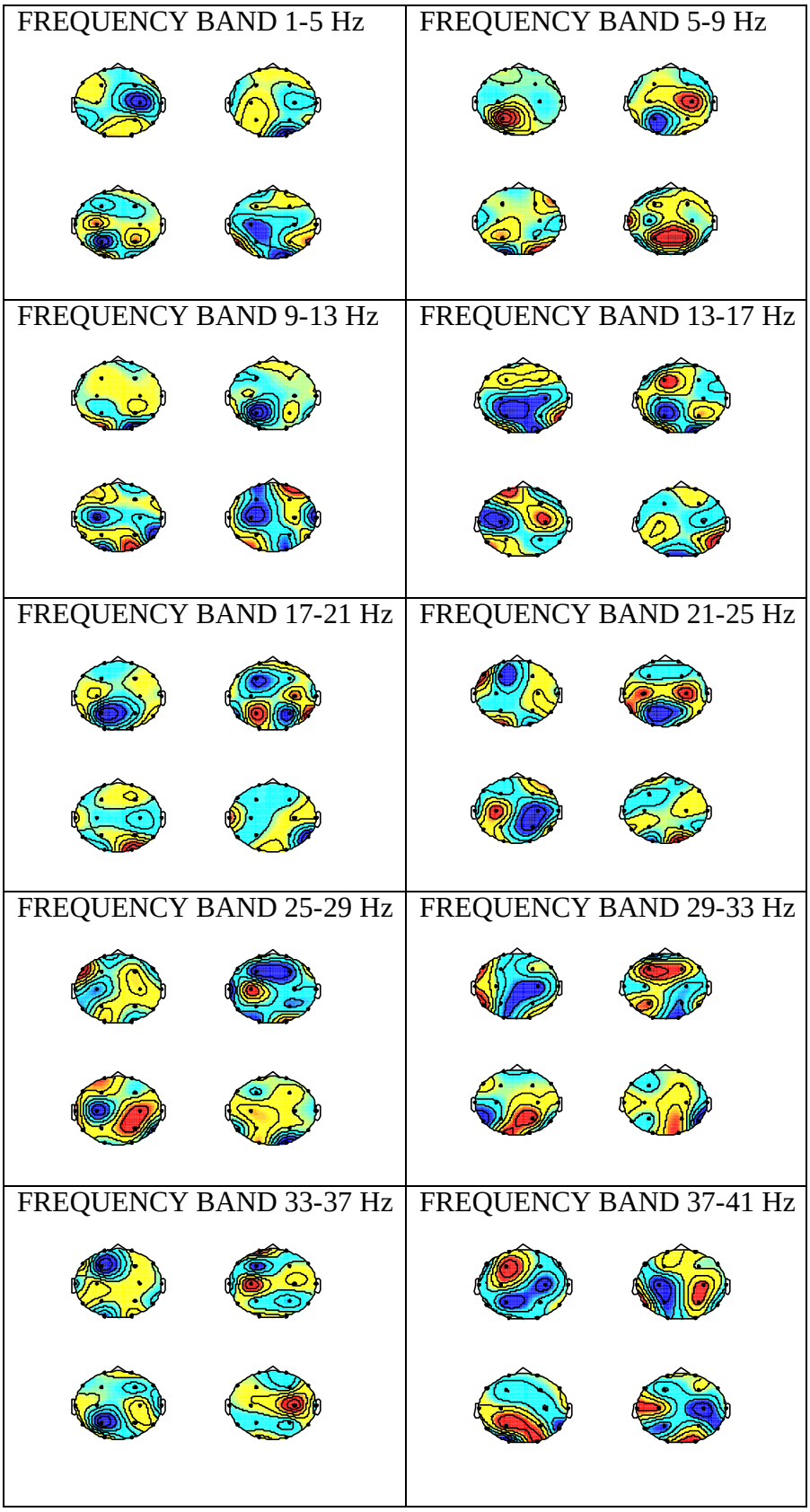


The ‘Red’ curve depicts the Training Error
The ‘Blue’ curve depicts the Validation Error

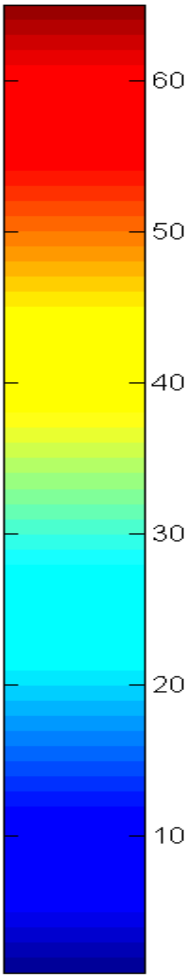
SUBJECT 6

The analysis of the matrix P

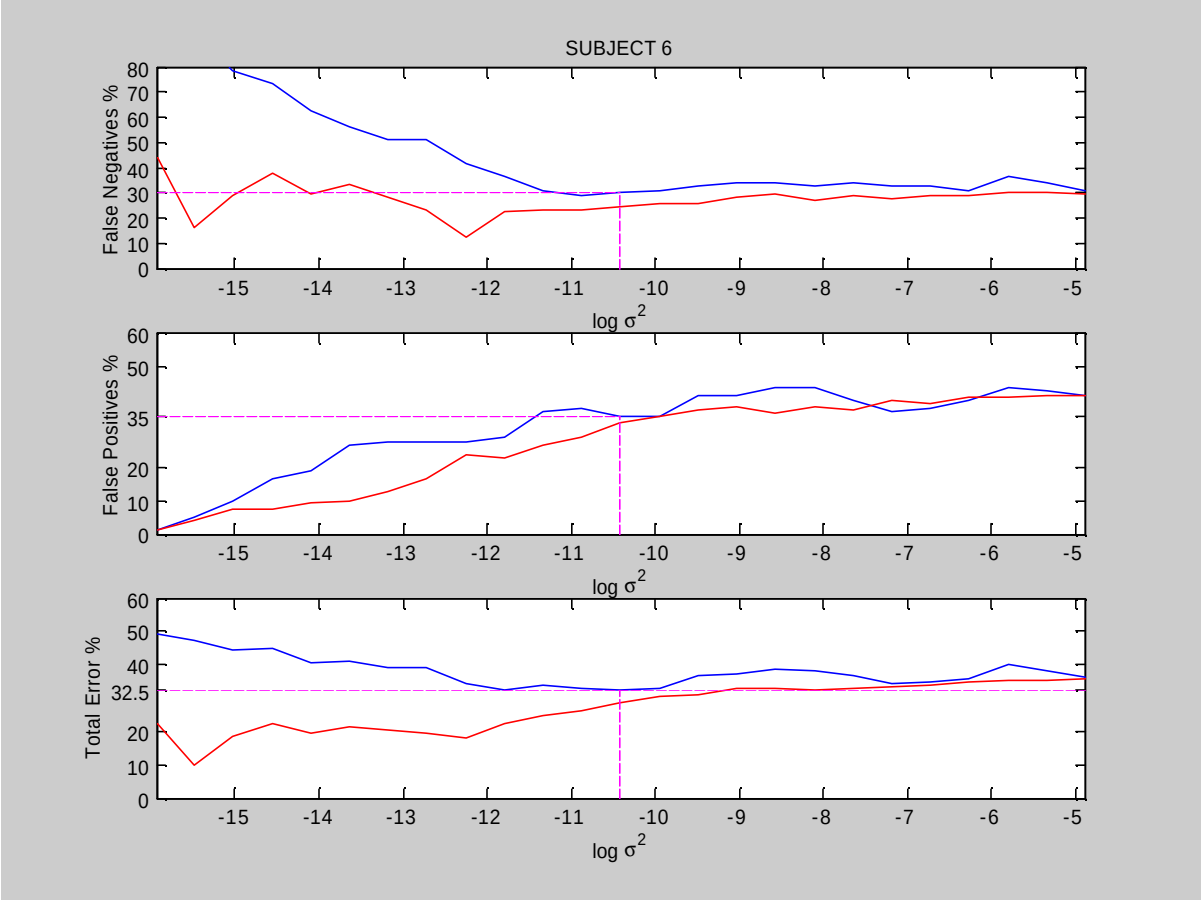




Topoplots showing the four coefficients of P for each Frequency Band

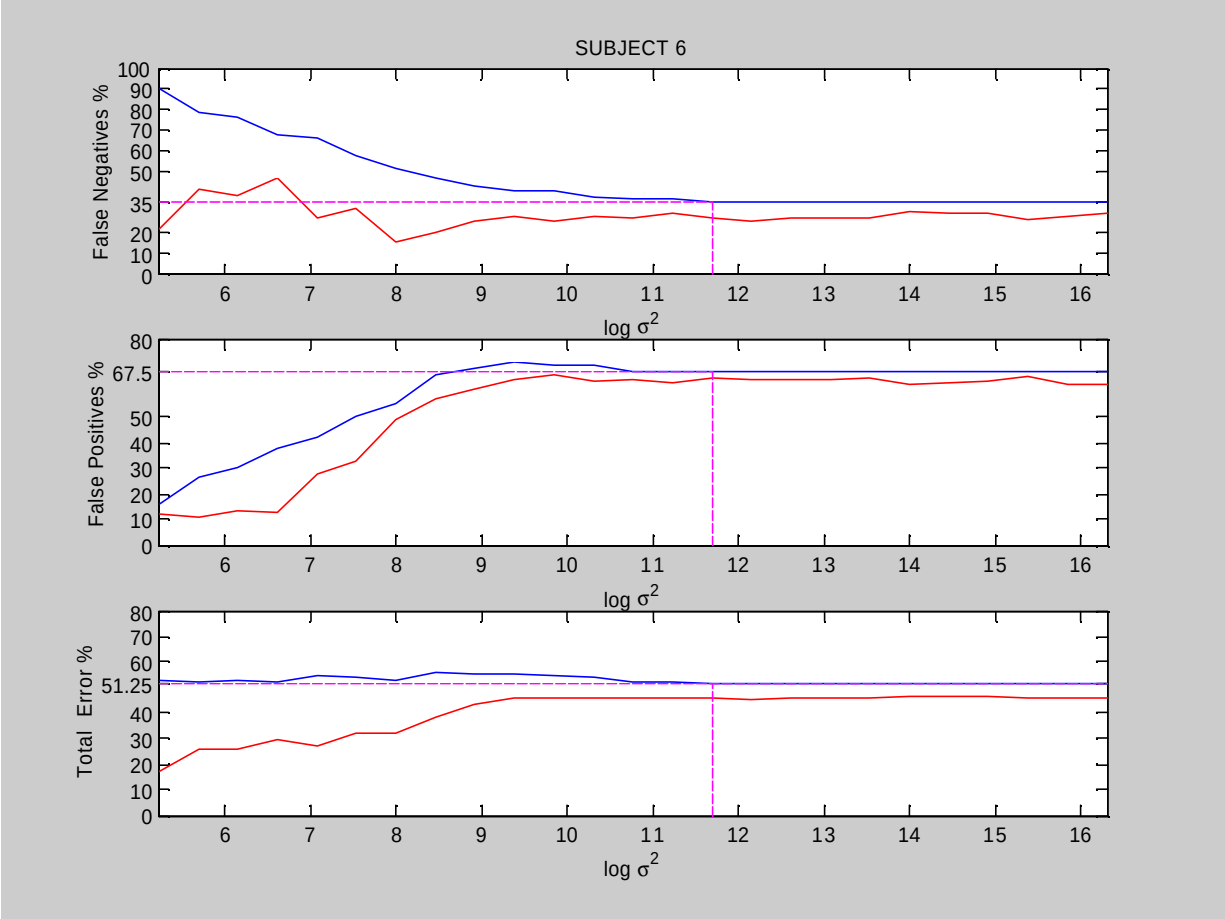


Result obtained using Spatial Features:



The ‘Red’ curve depicts the Training Error
The ‘Blue’ curve depicts the Validation Error

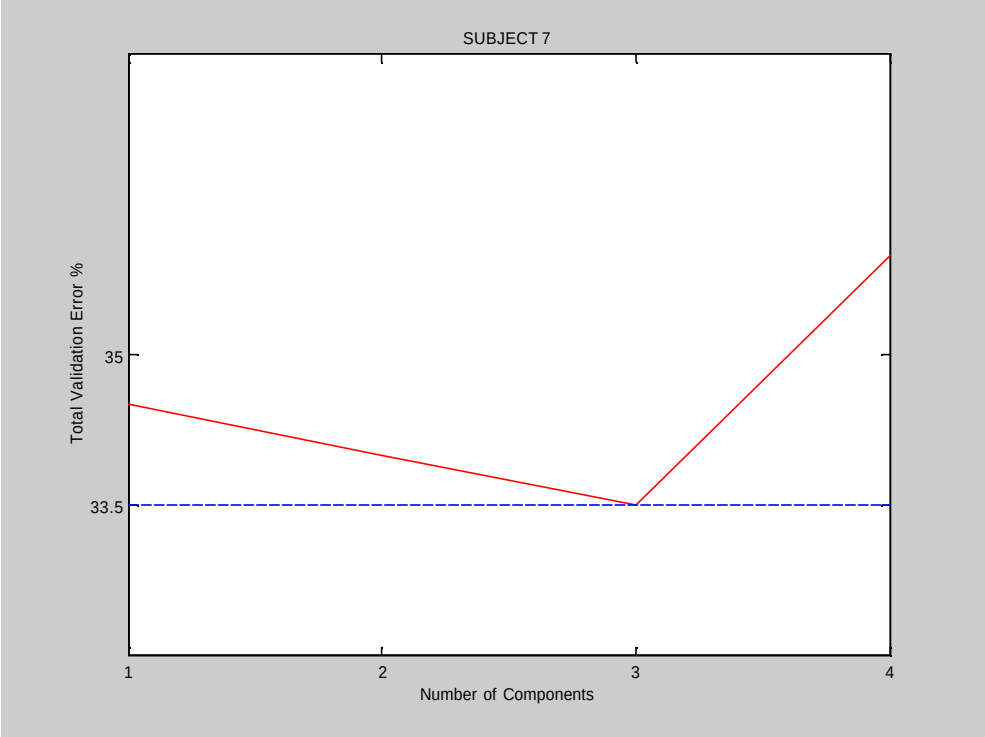
Results obtained using Frequency Features

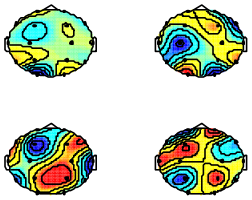
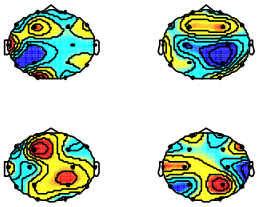
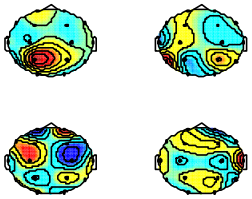
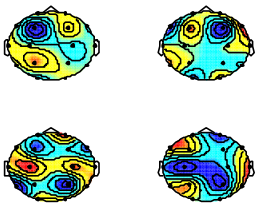
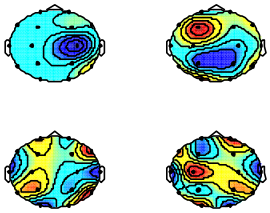
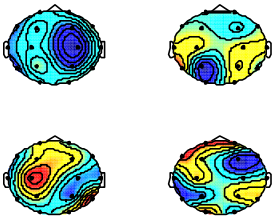
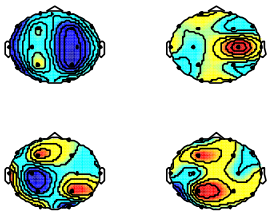
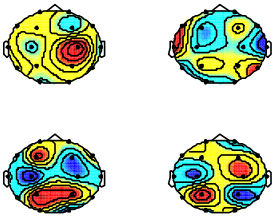
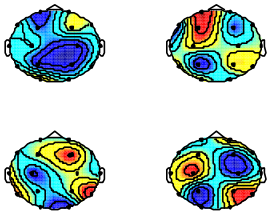
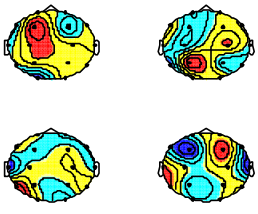


The ‘Red’ curve depicts the Training Error
The ‘Blue’ curve depicts the Validation Error

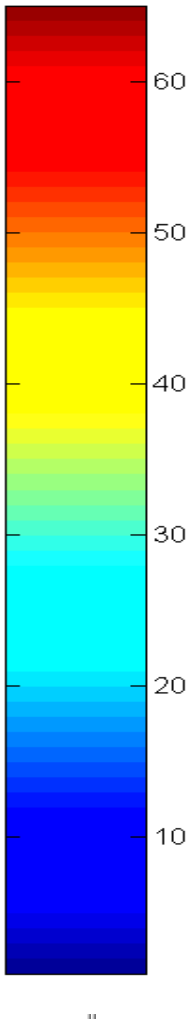
SUBJECT 7

The analysis of the matrix P

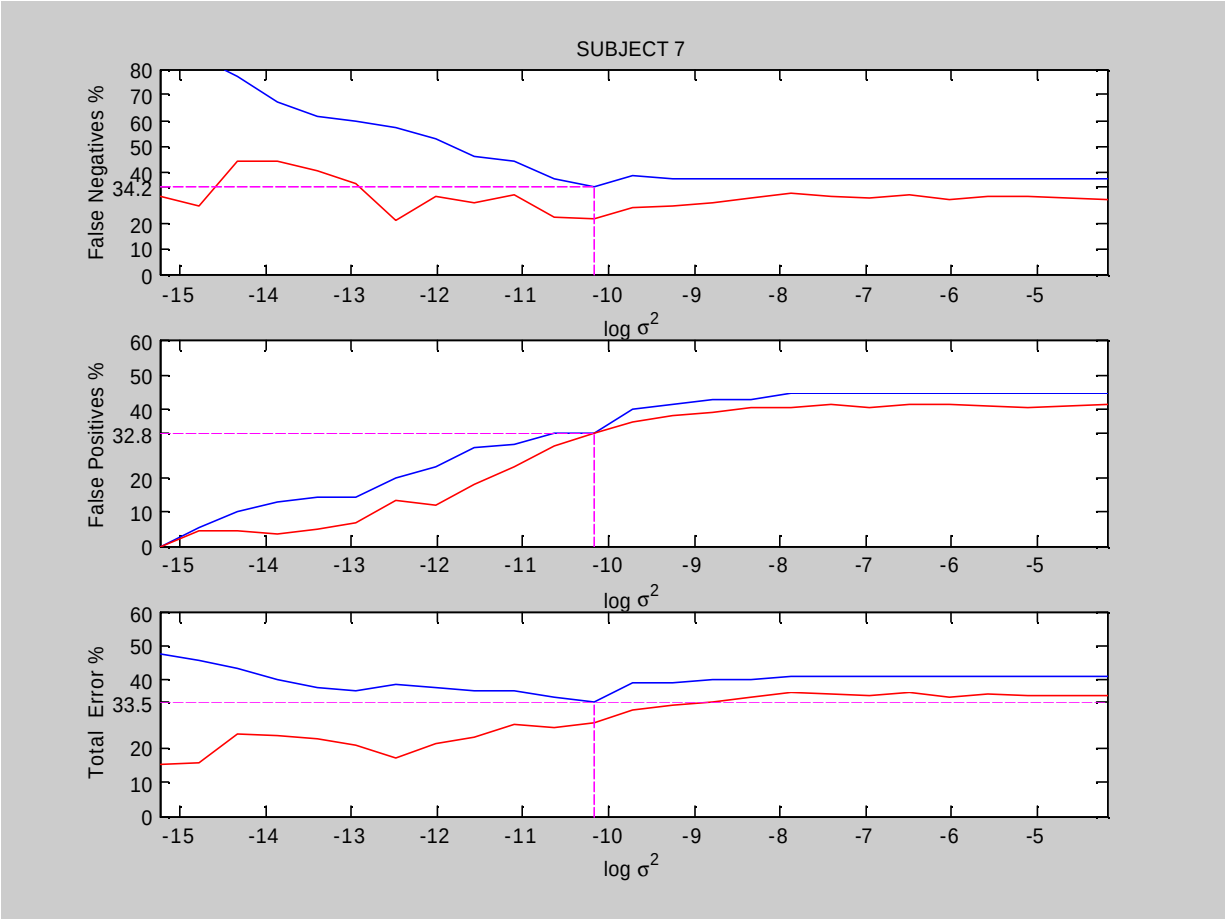


FREQUENCY BAND 1-5 Hz 	FREQUENCY BAND 5-9 Hz 
FREQUENCY BAND 9-13 Hz 	FREQUENCY BAND 13-17 Hz 
FREQUENCY BAND 17-21 Hz 	FREQUENCY BAND 21-25 Hz 
FREQUENCY BAND 25-29 Hz 	FREQUENCY BAND 29-33 Hz 
FREQUENCY BAND 33-37 Hz 	FREQUENCY BAND -37-41 Hz 

Topoplots showing the four coefficients of P for each Frequency Band

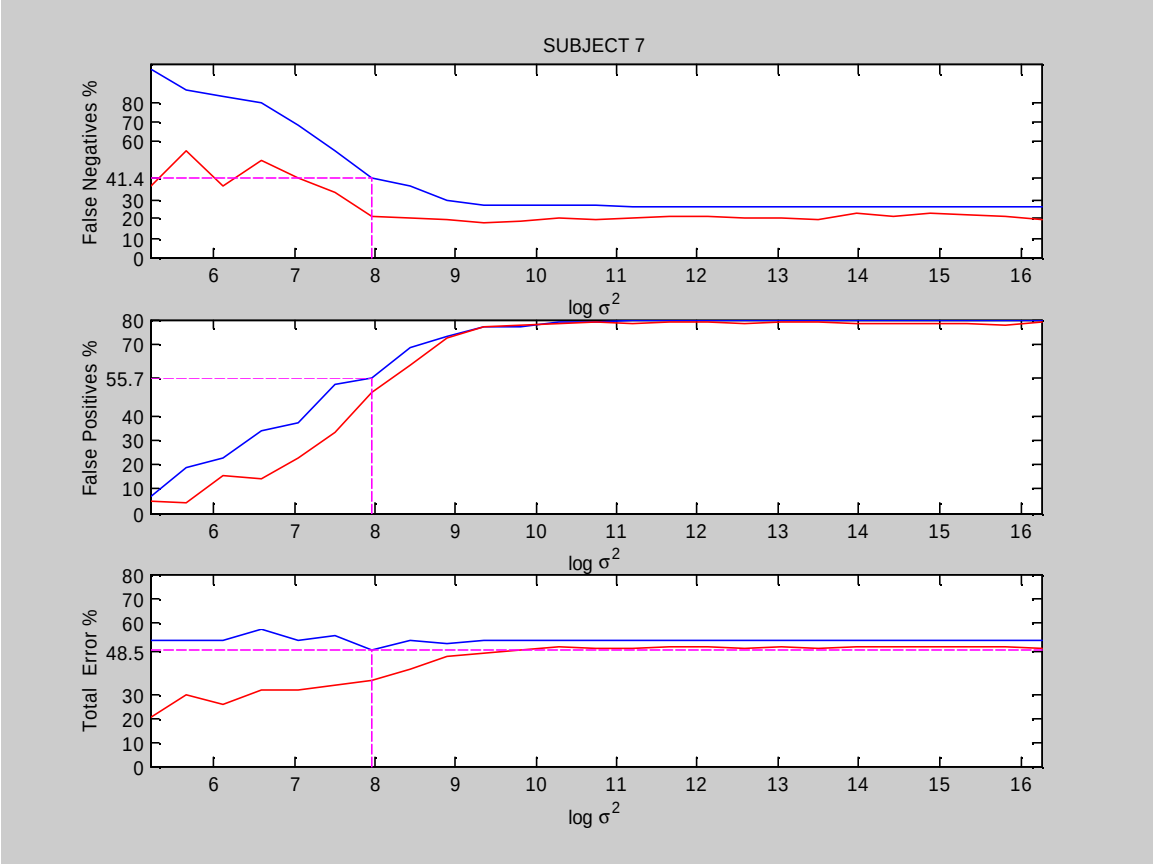


Result obtained using Spatial Features:



The ‘Red’ curve depicts the Training Error
The ‘Blue’ curve depicts the Validation Error

Results obtained using Frequency Features



The ‘Red’ curve depicts the Training Error
The ‘Blue’ curve depicts the Validation Error

7. ALPHA RHYTHM

The alpha rhythm are present in a specific frequency band-(8-13Hz).The alpha waves, of moderate amplitude , are typical of relaxed wakefulness(idling).We performed the following experiment to record the alpha rhythm of a person.

The subject was asked to alternately close and open his eyes,every 2 seconds. Closed eyes produce alpha rhythm which is depicted by large amplitude in the last six electrodes namely 'T5' , 'P3' , 'P4' , 'T6' , 'O1' , 'O2'.The power spectrum in the range 8-13 Hz is calculated for both the closed eyes and the open eyes.

The following figure depicts the power spectrum for the nine subjects, eight males and one female, in the age-group of 20-30 years , who participated in the experiment.

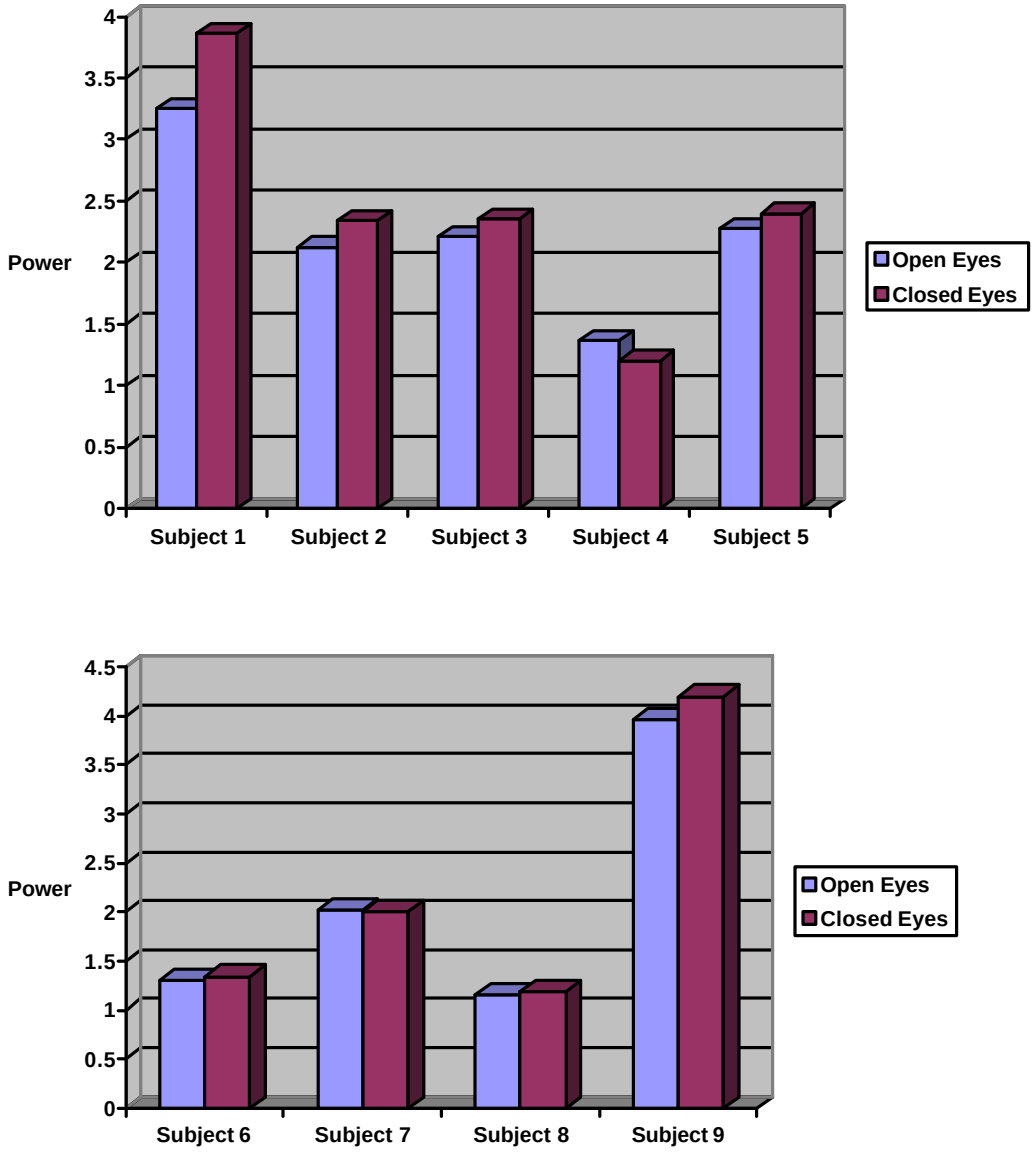


Figure.9

From **Figure. 9,** we can interpret that the closed eye spectra is higher than the open eye spectra for most of the subjects. Some subjects do have the capability to produce large amplitude alpha rhythm even with open eyes. But some readings in **Figure. 9** are not exactly correct as 2-3 subjects reported blinking of eyes during the trial, which leads to wrong data.

8. CONCLUSION

The results obtained from different subjects show that the Spatial Features are more discriminatory than the Frequency Features.On an average, the error with the Spatial Filters is around 30% as compared to nearly 50% for the Frequency Features. The analysis of different components of the matrix P depict that not all components are necessary for every subject to get the best results. Infact , this shows that its possible for a subject to have just one electrode as the discriminating electrode for the two mental activities.

The analysis of the alpha rhythm shows that some subjects have the capability of producing high power alpha rhythm. These subjects can use this power to control an operation willingly.

References:

[1] G.Garcia, T.Ebrahimi and J.-M. Vesin , 'Human-Computer Adaptation for EEG based Communication'

[2]T.Ebrahimi, J.-M Vesin and G.Garcia , 'Brain-Computer Interface in Multimedia Communication' IEEE Signal Processing Magazine January 2003.

[3] Gary N. Garcia, T. Ebrahimi and J.-M. Vesin, 'Joint Time Frequency-Space Classification of EEG in a Brain Computer Interface Application' , Eurasip Journal on Applied Signal Processing, special issue on Neuromorphical Signal Processing.