Single trial EEG Classification

SWISS FEDERAL INSTITUTE OF TECHNOLOGY

AMIT ADLAKHA

Responsable Assistant : **GARY GARCIA MOLINA**

Professer: TOURADJ EBRAHIMI

July 12th, 2002

CONTENTS

INTRODUCTION

IN JUST 50 YEARS, the computer has revolutionized our lives, and our futures. But one problem remains: the interface between computer and mind. This is the Big Barrier, and it is almost impenetrable today as it was in those early days when we used paper tapes.

The race is on to breach the Big Barrier. Researchers in industry and academia are seeking new, more expensive ways to communicate with machines. The idea of using our brains to directly control a machine isn't particularly new. Research into the Brain-Computer Interface, or BCI, began in earnest in the early 70's, when the USA Department of Defense saw the promise of fighter pilots using their minds to directly control their planes.

A successful BCI must be bi-directional. Getting information into the brain is relatively easy: we can use the normal sensory channels, such as sight or hearing. But getting the information out of the brain by studying its electrical signature is a harder problem. However, by focusing on very specific areas of brain activity, such as the motor function, it is possible to analyze EEG data using filters, Fourier transforms, and neural networks, to extract some useful signal the noise.

BCI research is red-hot. Although still at an exploratory stage, the implications of recent research results are phenomenally exciting. Perhaps one day, the Big Barrier will simply disappear altogether.

INTRODUCTION TO EEG SIGNAL ANALYSIS AND CLASSIFICATION:

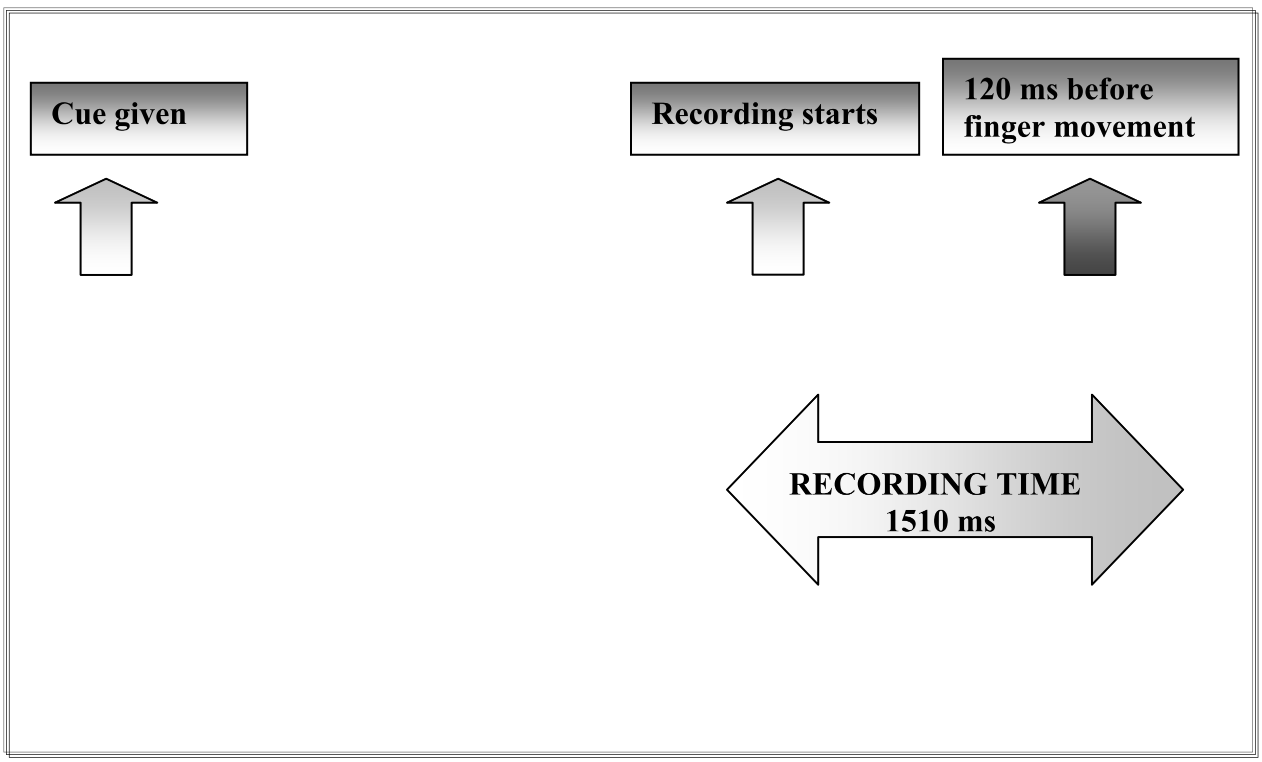
EEG signal analysis and classification is one of the prominent researches in field of Brain Computer Interface. Larger and larger techniques are being devices to analyze the EEG signal and extract useful information out of it. In general, the process of EEG signal analysis and classification consists of the ***following three*** steps:

- **SIGNAL PREPROCESSING**: The process involves filtering that portion of the signal that is more important for signal classification. For example the alpha band is more important as far as the motor actions are concerned. Preprocessing also includes the removing the artifacts that creep into the signal due to various reasons like eye blinking or muscular activity.

- **FEATURE EXTRACTION:** This phase involves extracting those features of the signal that display certain characteristic properties of EEG signal that are unique to the signal and are thus suitable for the classification purpose. Extracting these features also reduces the amount of data that is fed to classifying machine and thus reduce the processing time of BCI system. The features that are generally used for classification include FFTs, the PSDs, The Auto Regressive Coefficients, The Multivariate Autoregressive Coefficients and The Time-Frequency transforms (like the Wigner Ville transform). Feature extraction also includes spatially filtering the multi-channel EEG signals for extracting discriminatory information from the signals. Various techniques like the neural network feature selector, fuzzy entropy based feature ranking and the Signal to noise ratio based technique can be used for identifying the electrodes that provide better discriminatory information.

- **THE CLASSIFICATION PROCESS:** The process involves identifying optimum demarcating boundaries for the various classes of signals in the feature space. This may accomplished using an Analytic Approach, Neural Networks or the Support Vector Machines. The learning machine requires a training data set to identify the demarcation between the various points in the feature space.

DESCRIPTION OF THE DATA USED FOR THE ANALYSIS AND CLASSIFICATION PURPOSE



The subjects were given certain cues to move their left/right index finger some three to four to four minutes before they actually asked to do the same. The EEG signal was recorded from the 27 electrodes during the time span of 1.51 seconds ending just at the instant of the movement the index finger.

POWER SPECTRAL DENSITY APPROACH:

The power spectral density, PSD, describes how the power (or variance) of a time series is distributed with frequency. Mathematically, it is defined as the Fourier Transform of the autocorrelation sequence of the time series. An equivalent definition of PSD is the squared modulus of the Fourier transform of the time series, scaled by a proper constant term.

Power Spectral Density Approach has always been a popular method for classifying EEG signals. The first step in EEG classification is to determine whether the signals have distinguishable features in their power spectrum. After plotting the average PSD spectrum for the left trials and the right trials, it was observed that the two graphs showed some difference at a certain band of frequencies (particularly the **alpha** band) though that difference was not much. This difference can be easily seen in the following graph. This band was collected and was used for training a neural network (An MLP). The network consisted of three layers: the *input layer* consisting of (fill in) nodes, the *hidden layer* 100 neurons and *an output layer* consisting of a single (to indicate the output class). The C3 electrode was used for classification.

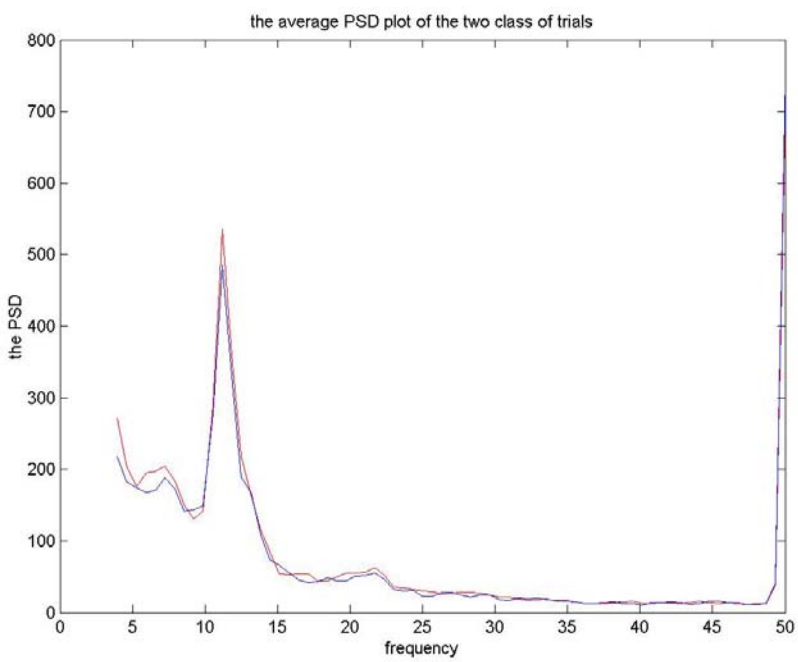


Figure: The above plot shows the average PSD drawn for LEFT trials and the RIGHT trials. Some difference can be seen in the alpha band and the theta band.

The back propagation was algorithm was used to train the network. The network was tested with the cross validation procedure. The test results obtained were poor. This may be because the power spectrum was not able to extract the distinguishing features. This can be easily seen from the graph. The results from the cross validation procedure have been tabulated below.

TRIAL NUMBER	CLASSIFICATION ACCURACY
1-40	40%
61-100	53%
101-140	63%
151-190	51%
AVERAGE	52%

The classification accuracy obtained above is no better than mere chance.

THE AUTO REGRESSIVE COEFFICIENTS APPROACH:

An autoregressive model (AR) is also known in the filter design industry as an infinite impulse response filter (IIR) or an all pole filter, and is sometimes known as a maximum entropy model in physics applications. There is "memory" or feedback and therefore the system can generate internal dynamics.

The definition that will be used here is as follows

$$x_t = \sum_{i=1}^{N} a_i \, x_{t-i} + \varepsilon_t$$

Where ai's are the auto regression coefficients, x (t) is the series under investigation, and N is the order (length) of the filter which is generally very much less than the length of the series. The noise term or residue, epsilon in the above, is almost always assumed to be Gaussian white noise.

Verbally, the current term of the series can be estimated by a linear weighted sum of previous terms in the series. The weights are the auto regression coefficients.

The problem in AR analysis is to derive the "best" values for ai given a series x(t). The majority of methods assume the series x (t) is linear and stationary. By convention the series x (t) is assumed to be zero mean, if not this is simply another term a_0 in front of the summation in the equation above.

Computation methods:

A number of techniques exist for computing AR coefficients. The main two categories are least squares and Burg method. Within each of these there are a few variants, the most common least squares method is based upon the Yule-Walker equations. Mat Lab has a wide range of supported techniques, note that when comparing algorithms from different sources there are two common variations, first is whether or not the mean is removed from the series, the second is the sign of the coefficients returned (this depends on the definition and is fixed by simply inverting the sign of all the coefficients).

The most common method for deriving the coefficients involves multiplying the definition above by x_{t-d}, taking the expectation values and normalizing (see Box and Jenkins, 1976) gives a set of linear equations called the Yule-Walker equations that can be written in matrix form as

$$\begin{pmatrix} 1 & r_1 & r_2 & r_3 & r_4 & \cdots & r_{N-1} \\ r_1 & 1 & r_1 & r_2 & r_3 & \cdots & r_{N-2} \\ r_2 & r_1 & 1 & r_1 & r_2 & \cdots & r_{N-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ r_{N-1} & r_{N-2} & r_{N-3} & r_{N-4} & r_{N-5} & \cdots & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_N \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ \\ r_N \end{pmatrix}$$

Where r_d is the autocorrelation coefficient at delay d. Note: the diagonal is $r_0 = 1$.

There are many other methods for calculating these coefficients like the Burg's method, the least square method, the geometric lattice method and the forward backward method.

AR COEFFICIENTS FOR FEATURE EXTRACTION AND CLASSIFICATION:

The C3 and C4 electrode signals were used for extraction of the AR coefficients. The feature vector consisting of these coefficients was used for classification. An MLP consisting of 20 hidden units was used for the purpose. A cross validation approach was utilized. The training set consisted of 171 trials and the testing set consisted of 20 trials. The classification results have been tabulated below.

TRIAL NUMBER	CLASSIFICATION ACCURACY
21-40	72.5%
41-60	52.5%
61-80	50%
81-100	57.5%
101-120	55%
121-140	62.5%
141-160	72.5%
161-180	62.5%
AVERAGE	59.7%

The results tabulate above reveal that the method is not suitable as features for the data set used. So, for a more robust feature vector we move on to the MVAR coefficients.

THE MVAR COEFFICIENTS APPROACH:

The AR coefficients are somewhat weak in representing the signals because they do not take into account the inter-spectral correlations. To explore the operation of large scale networks in the cerebral cortex, we sought to measure the functional dependence of local field potentials from different areas. To track the transformation of functional dependence accompanying rapid changes, we can use the spectral coherence analysis or the multivariate autoregressive analysis.

A SHORT MATHEMATICAL DEFINITION OF THE MVAR COEFFICIENTS:

A multivariate signal can be modeled by the following equation:

$$V(k,:)' = w' + A1 * v(k-1,:)' + A2 * v(k-2,:)' + eta(k,:)'$$

Where the row vectors eta (k, :) are independent and identically distributed Gaussian noise vectors with zero mean and covariance matrix C. The kth row represented by vector in the left hand side of the above equation of the two column matrix v represents an observation of a process at instant k. The intercept vector w is included to allow for a nonzero mean of the AR (p) process. The matrices A1 and A2 were determined by the Schwarz Bayesian Criterion.

PREPROCESSING AND FEATURE (MVAR COEFFICIENTS) EXTRACTION:

After an estimation of the A matrices using a ***third*** order estimation of the signal using the MVAR technique, the coefficients were grouped to form a column vector which could be used for training a learning machine.

After filtering the signals from 0 to 45 hertz using a 6th order Butterworth filter, the MVAR coefficients were extracted from both the class of recorded signals (*The Cs and some of the CAs and the CPs were used in the process: ten electrodes in all*) using a third order fitting determined by the Schwarz Bayesian Criterion. The coefficients were used to train a support vector machine. It should be noted that a switch over from the neural networks to SVM was needed to enhance the training speed of the learning the features of the training set.

SVMS BETTER THAN NEURAL NETWORKS, WHY?

The basic ideas behind support vector machines have been explored in the early sixties. The idea of finding robust (generalizing) separating surfaces by placing the separating surface in a position in the input space that is at a maximal distance from all support vectors is not new. The support vector machines learn by finding optimal hyper planes that separate feature points of two different classes in the feature space. This they do in the same feature space or by projecting the features into a higher dimensional plane (where classification is easy). This might not be the case in neural networks because the training is not completely controlled by us. So, determination of an optimum separation hyper plane is not guaranteed. Moreover, the training *and the testing time fro the SVM is generally smaller than the neural networks because of the faster algorithms used for training them.*

RESULTS OBTAINED FROM CROSS VALIDAION:

Testing procedure: A total of 190 trial signals were taken from the data set of each class of signals. The signal set was divided into two parts: the larger part consisting of 170 signals for testing and the smaller part used for testing. This was done for different sets of test signals. A radial basis support vector machine with the variance parameter equal to 10^{-4} (obtain from LIBSVM) was tested for different test data sets using cross validation as mentioned above. The results obtained are tabulated as below:

TRIAL NUMBERS	PERCENTAGE CLASSIFICATION
20-41	50%
40-41	75%
61-80	50%
81-100	60%
101-120	65%
121-140	60%
141-160	65%
161-180	60%
AVERAGE	61%

The average classification rate (61 %) was not good enough. A further analysis of the MVAR coefficients obtained was needed to determine why the method was not showing good results. A list of the wrongly classified signals was also obtained for future use to determine why they were wrongly classified.

FURTHER ANALYSIS OF THE MVAR COEFFICIENTS [INTERSPECTRAL POWER SPECTRAL DENSITY]

The coefficients of the A matrix during the calculation represent interdependence of one electrode signal on the other. For example a (1, 2) represents degree of cross relation between the first electrode and the second electrode. These coefficients when considered separately do show correspondence with the mono variant AR coefficients. So, they (like the AR coefficients) can be used to estimate the inter-spectral power density.

Similar to the previous case, 10 electrodes were used to obtain the MVAR coefficients and the inter-spectral power coefficients were obtained for each of the trials. An average of these spectral densities was obtained for each of two classes of signals and the contrast function was obtained to figure those frequencies and electrode pairs that are important for classification.

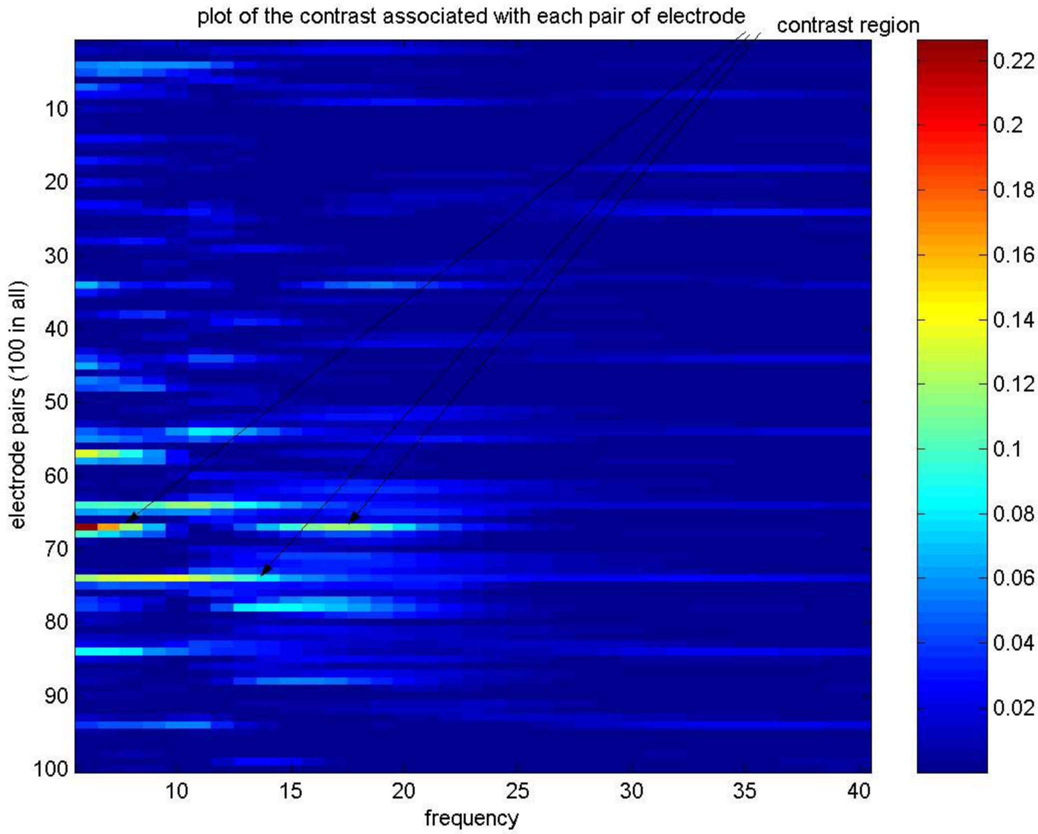


Fig: A plot of the Electrode pairs Vs frequency. The x axis represents frequencies form 5 to 40 hertz . The y axis represents the electrode-pairs (100 electrode pairs in all). The color codes represent the value of the contrast function at that point. The regions of contrast are marked.

It can be easily observed from the graph that there are very few points on the plot that are different, and even if they are the contrast function do not exceed satisfactory limits (generally 0.3-0.4). This is indicative of the fact that the MVAR coefficients were not able to extract much distinguishing features of the signal. It was observed that the electrode pairs did correspond to pairs that had C3 or C4 as one of the electrodes. To further the analysis these points were taken used for classification. Thus the contrast points were used for training a neural network (an MLP). A three layer network was used with 150 hidden units.

To have a flavor of the classification ability of these features, some 150 signals were used for training and the rest were for testing. An average of *67% trials* was correctly classified. This accuracy is not far from that obtained from the previous one.

Unsatisfied with the approach and observing that the signals are not stationary at all time instants of the trial, we move on to the time frequency approach for better feature extraction.

TIME FREQUENCY TRANSFORMS
[FOR BETTER FEATURES EXTRACTION]

WIGNER VILLE TRANSFORMS: All the covariant time-frequency transforms can be obtained from the Cohen class defined as

$$C(t,\omega) = 1/4\pi^2 \iiint s^*(u - \frac{1}{2}\tau) s(u + \frac{1}{2}\tau) \phi(\theta,\tau) e^{-j\theta t - j\tau\omega + j\theta u} du d\tau d\theta$$

Where $\phi(\theta,\omega)$ is a two dimensional function called the kernel. For a WV transform this function is unity for all the input space. The transform gives a good idea of the variation of the frequency content of the signal at different time instants of recording. Thus it has an edge over the Fourier transform and the frequency spectrums which only give an idea of the overall frequency content of the signal. This is important when the signal is not stationary.

FEATURE EXTRACTION AND CLASSIFICATION USING THE WIGNER VILLE TRANSFORM:

The signals from the two data sets were filtered from 1 to 40 hertz using a Butterworth filter. Their WV transforms were manipulated and average for both the classes' was obtained. A contrast between the WV transforms of the two distributions was obtained using the contrast function. The time frequency points showing the maximum frequency were collected and were used to train an SVM. As before, the signal was divided into a larger training set and a smaller testing set. This was done for different testing sets (Cross validation). The training set consisted of 170 trials and the testing set consisted of 20 trials. This was done for different electrodes (the C, CP and CA electrodes). The electrodes giving better were drawn out.

TRIAL NUMBERS	CLASSIFICATION RATE
20-41	62.5%
41-60	72.5%
61-80	85%
81-100	77.5%
101-120	60%
121-140	80%
141-160	75%
161-180	65%

The above table gives the results of cross validation for the 13th electrode (C5 electrode). The C3 electrode provided almost the same classification accuracy.

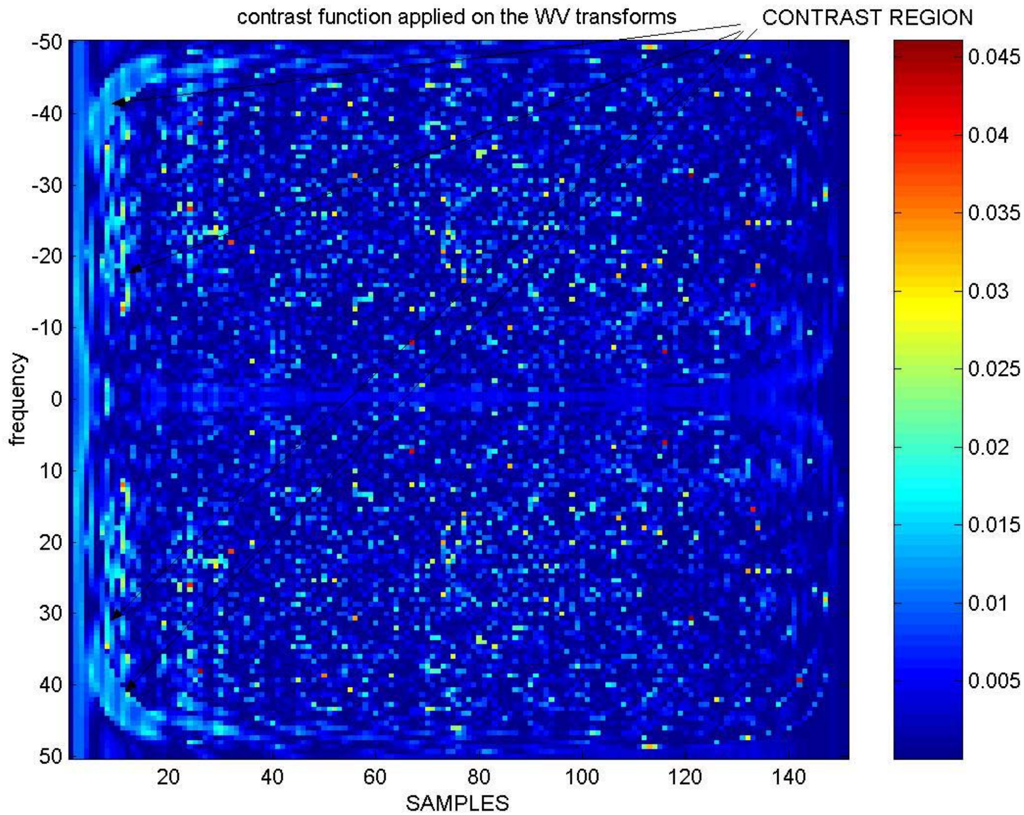
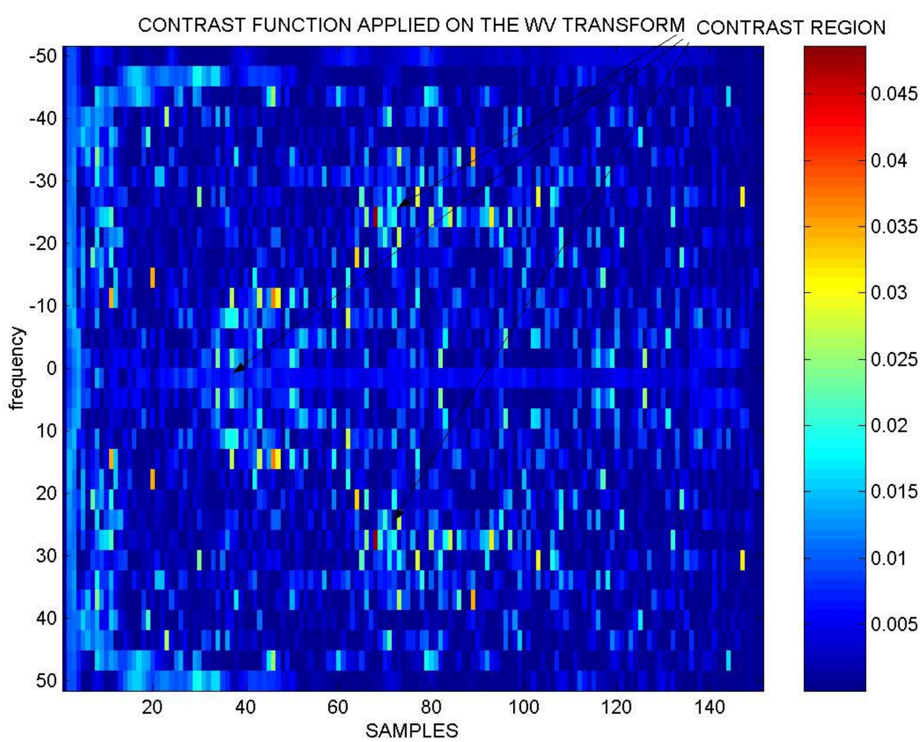


Figure: Contrast function applied on the WV transforms (high and low resolution). <u>The color code represents the contrast value at that time-frequency point.</u>



OPTIMAL SPATIAL FILTERING: The purpose of spatial filters is to obtain a new time series which has better variance and discriminatory features than the EEG signal from the individual electrodes. The method used to design such spatial filters is based on the simultaneous diagonalization of two covariance matrices.

Since, at this stage it becomes important to search for the electrode that shows the most distinguishing features we adopt the method called spatial filtering to obtain a new time series which has better discriminatory features. So you save yourselves the pain of looking for electrodes with better results.

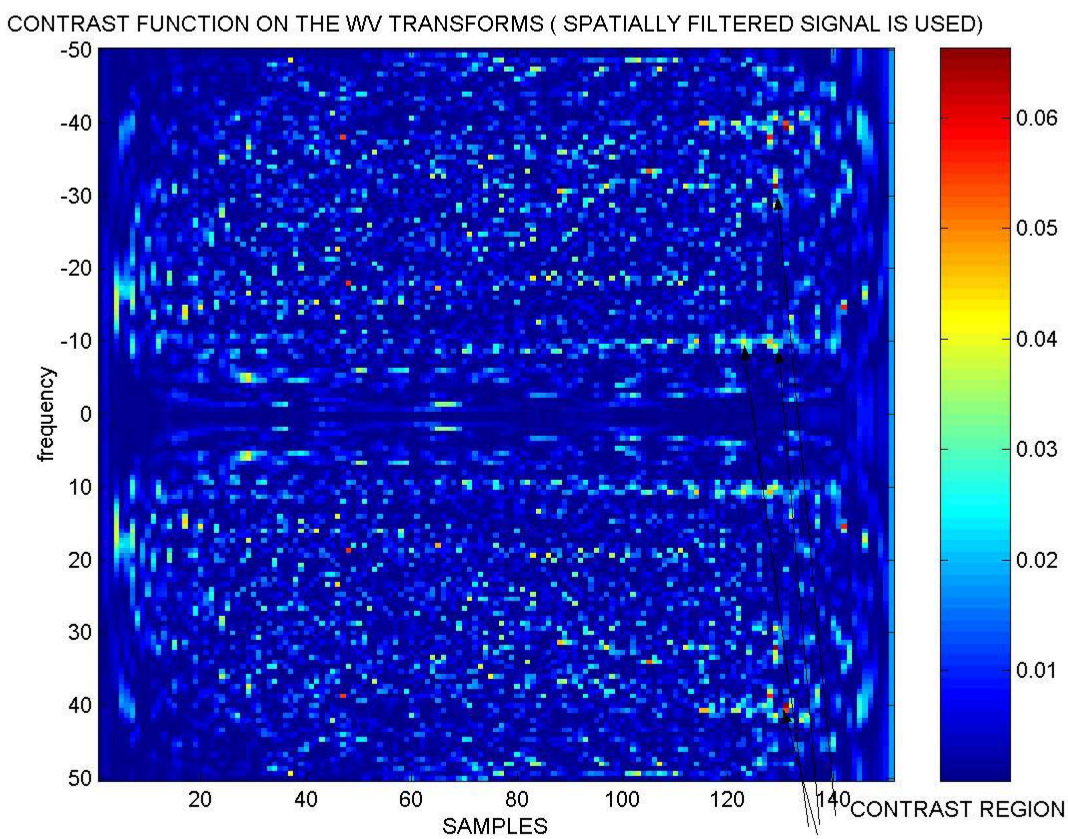


Figure: the figure shows the contrast between the WIGNER VILLE transforms for the signal showing the largest difference in the Eigen values (the largest difference the variance) after spatial filtering. <u>The color codes represent the value of the contrast function at time-frequency</u>. The regions with good contrast have been marked.

RESULTS OBTAINED USING THE WV TRANSFORMS OF THE SPATIALLY FILTERED SIGNALS: As usual, the signals were filtered between 1 to 40 hertz. The filtered signals were then spatially filtered and the signal with the largest difference in variance is obtained. This WIGNER VILLE transforms of this signal is obtained and for 171 trials from each of the 2 classes. In a similar fashion, the contrast function was obtained and the distinguishing points were extracted t form a feature set. These were then used for training and testing using the method of cross validation. The training set was used to train an SVM (radial basis) with a Gaussian coefficient 10 to the power (-13).The results are tabulated below:

Tabulated results:

TRIAL NUMBERS	CLASSIFICATION
21-40	78%
41-60	73%
61-80	78%
81-100	83%
101-120	73%
121-140	78%
141-160	68%

TESTING THE NEW APPROACH ON NEW DATA SET

After obtaining good results on one set of data, it becomes necessary to test technique on new data set to check its validity. A new data set is obtained with 6 types of trials. Unfortunately, the sampling frequency is unknown. Data trials from two of the classes is obtained is used to validate the method. The same preprocessing was done and the contrast points were obtained. ***The second set comes from a contest where in principle the exact protocol was not detailed, only the task indices were provided. The only information provided is that in all, 64 electrodes were used for recording.***

The following figure shows value of the contrast function for the time frequency points. We observe that there is a lot of difference in the signals in the later time span of the recordings of the trials.

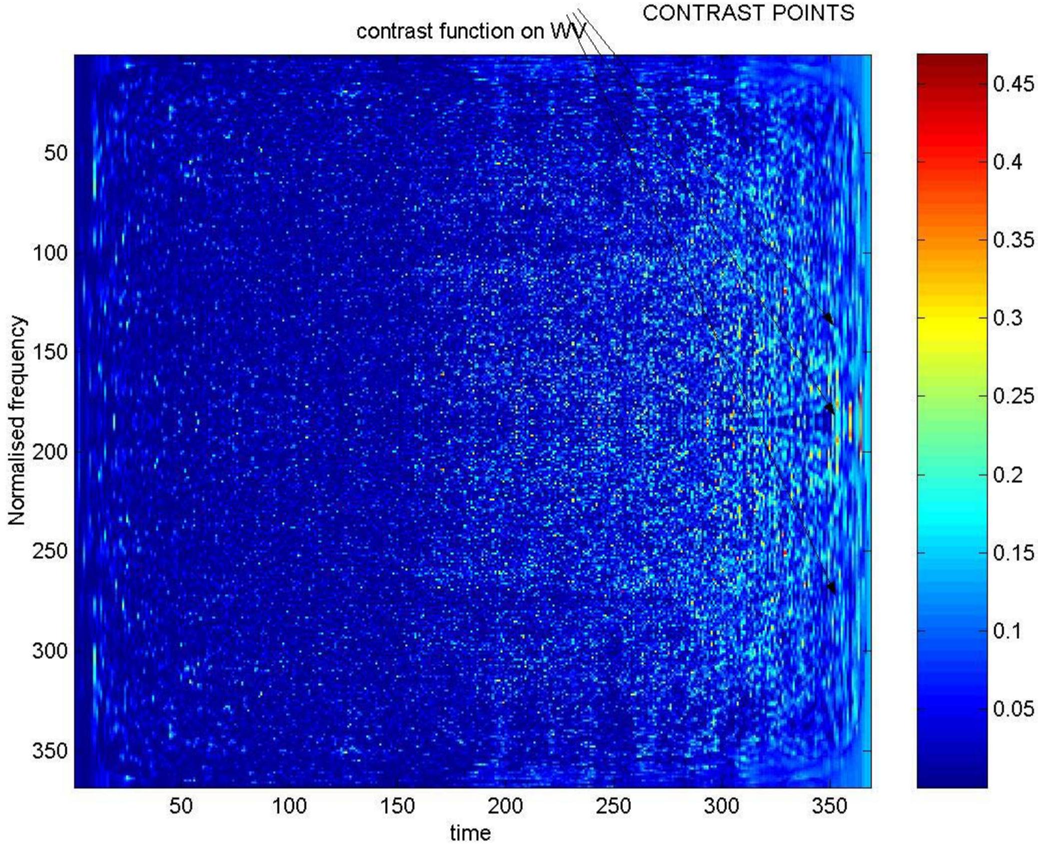


Figure: It shows the contrast between the 2 classes of signal in the new data set. The normalized frequency is used because unfortunately the sampling frequency is not known. <u>The color codes represent the values of the contrast function at that time-frequency point. The regions with good contrast have been marked (the later period of trials recording)</u>

TABULATED TESULTS: The data set of each class consisted of 47 trials. To test validity of the technique, the data set was divided into two sets: the first consisting of 30 training trials and the second consisting of 17 testing trials. This was done for two sets of testing data. The first time the last trials were used for training and for the second the first 17 trials were used for testing. As usual a Gaussian SVM was used with g parameter equal to 10^{-20} .The tabulated results are as follows:

TRIAL NUMBER	CLASSIFICATION	NO. OF SUPPORT VECTORS
1-17	82%	52
31-47	80%	51

EEG SIGNAL ANALYSIS USING MORPHOLOGICAL/FRACTAL ANALYSIS

A fractal description of EEG signals can be a useful tool for feature extraction. Fractals are objects which possess a form of self-scaling: Parts of the whole can be made to fit the whole in some way or the other in some way by shifting and stretching. Fractals features represent the morphology of the signals in some way or the other. These morphological differences can be picked up and used for several applications. They have already found applications in the field of traffic control, image analysis and compression. Such utility of fractal/morphological analysis is a source of motivation to consider it as a useful tool for feature extraction in EEG signal analysis. They have already been used to monitor patients during anesthetic procedures. There have been only one or two papers that have actually used the chaotic behavior of brain signals for EEG classification. One of them used the box counting dimension for EEG signal classification.

The plausibility of signal classification using fractal classification depends on the nature of the signal. “Brain Signals are chaotic” is still a question to be answered?

There are several features based on fractal theory/morphological analysis that can be extracted from the normal signals.

FRACTAL FEATURES FOR EEG SIGNAL CLASSIFICATION:

EEG signal classification is all about extracting features and classifying the signals based on these features. Some of the features based on fractal analysis are as follows:

1. **FRACTAL DIMENSION**: Fractal Dimensions are a measure of the self similarity of the signals. A lot of dimensions have been defined in this field. Some of these are as follows :

- **Regularization Dimension**: This dimension is defined in the following way: One first computes smoother and smoother versions of the original signal, obtained through convolutions with a kernel. When the original signal is fractal, its graph has infinite length, while all regularized versions have finite length. When the smoothing parameter tends to 0, the smoothened version tends to the original signal, and its length will to tend to infinity. The regularization dimension measures the speed with which the convergence takes place.
- **Box Dimension**:
  We define the *box-counting dimension* (or just ``box dimension") of a set S contained in ^{n} as follows: For any $\epsilon > 0$, let N_{ϵ}(S) be the minimum number of *n*-dimensional cubes of side-length ϵ needed to cover S. If there is a number *d* so that

$$N_{\varepsilon}(S) \sim 1/\varepsilon^{d} \quad \text{as} \quad \varepsilon \to 0$$

We say that the box-counting dimension of S is *d*. We will denote this by S = *d*.

Note that the box-counting dimension is *d* if and only if there is some positive constant *k* so that

$$\lim_{\epsilon \to 0} \frac{N_\epsilon(S)}{1/\epsilon^d} = k.$$

HOLDER FUNCTIONS/EXPONENTS: Holder functions are a measure of the measure of the degree of regularity of the signals/functions. Some of the holder exponents are as follows.

- **Point-wise Holder Functions:** The point-wise Holder exponents, which characterizes the regularity of the measure/function under consideration.
- **The local Holder exponent:** It characterizes the regularity of the function around any given point.
- **The long range dependence parameter:** This one describes power law behavior of the Fourier power spectrum near the zero frequencies.

ONE DIMENSIONAL MULTIFRACTAL SPECTRA: Basically, any of these spectra provides an information as to which singularities occur in your signal, and which are the dominant; a spectrum is a one dimensional curve, where abscissa represents the Holder exponents present in your signal, and the ordinates are related to amount of points where you'll encounter a given singularity . We have worked on two kinds of these spectra:

- **The Legendre spectrum:** This is based on the Legendre transform a signal. It may be the Discrete Wavelet Transform based Legendre spectrum or a CWT based Legendre spectrum.
- **The large deviation spectrum:** this spectrum yield statistical information related to the probability of finding a point with a given holder exponent in the signal. More precisely, it measures how this probability behaves with the change in resolution.

ANALYSIS OF OUR EEG DATA USING FRACTAL ANALYSIS

Fractal Dimension (Box Dimension): An Analysis of the left trials and the right trials with the box dimension (N_MIN =2 and N_MAX=5) using the least square algorithm, we observe that both signals have an average box dimension of 1.3. Classifying them on their dimension is difficult.

GIFS based point-wise holder functions: It calculates the holder functions at each point based on Gifs technique. The left and the right trials were used to extract the GIFS holder exponents and an average was performed. The C3 electrode was used for the purpose. The following graph shows the average of both the left trials.

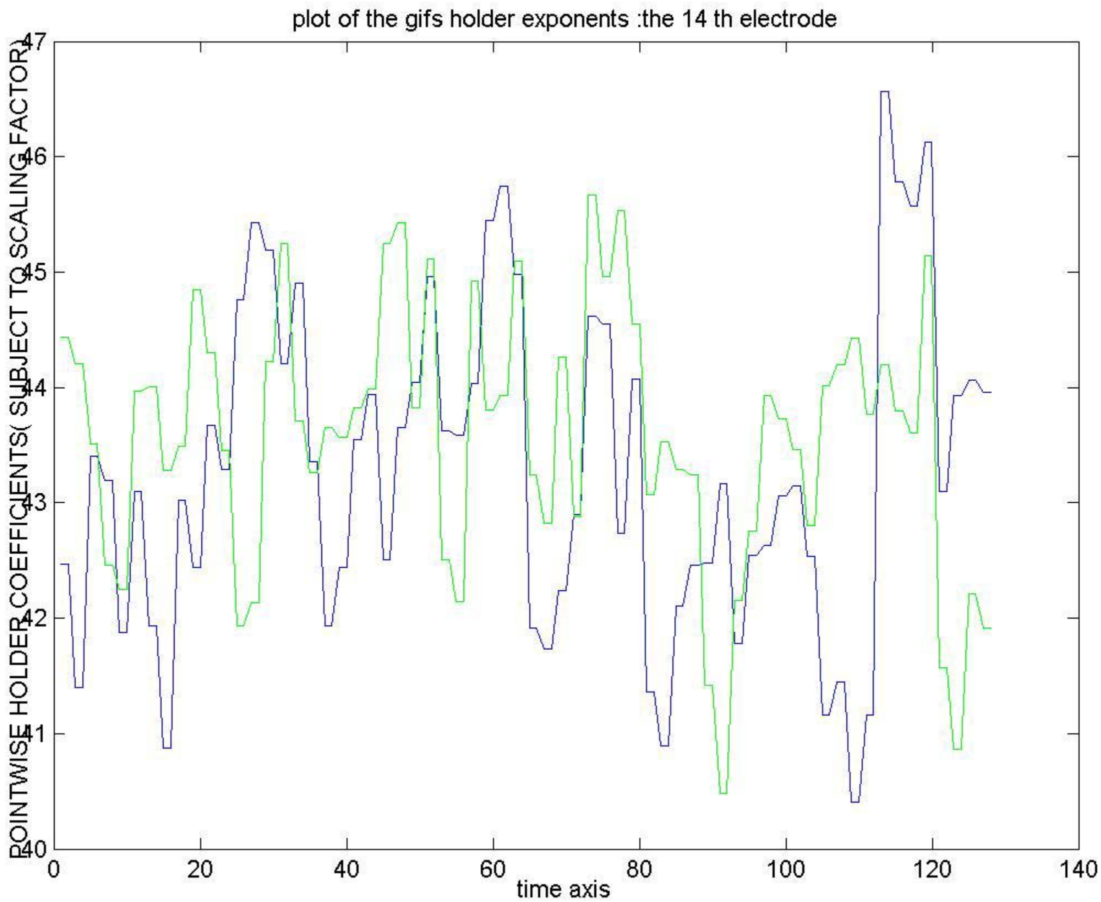


Figure: An average of the GIFS holder function plotted against time samples (Only 128 samples are considered). <u>The average is obtained over 170 trials of the LEFT TRIALS as well on the RIGHT TRIALS</u>. The C3 electrode has been used for the purpose.

Since the holder functions are very sensitive to noise, we try considering an average of the holder coefficients.

A similar technique was used on the other data set and an average was performed on intervals to minimize the effect of noise. The size of the interval was chosen to be 32 samples. Only 256 samples were considered.

LARGE DEVIATION BASED SPECTRUM: The large deviation spectrum yields a statistical information, related to the probability of finding a point with a given holder exponent in the signal. It measures how the probability behaves under change of resolution.

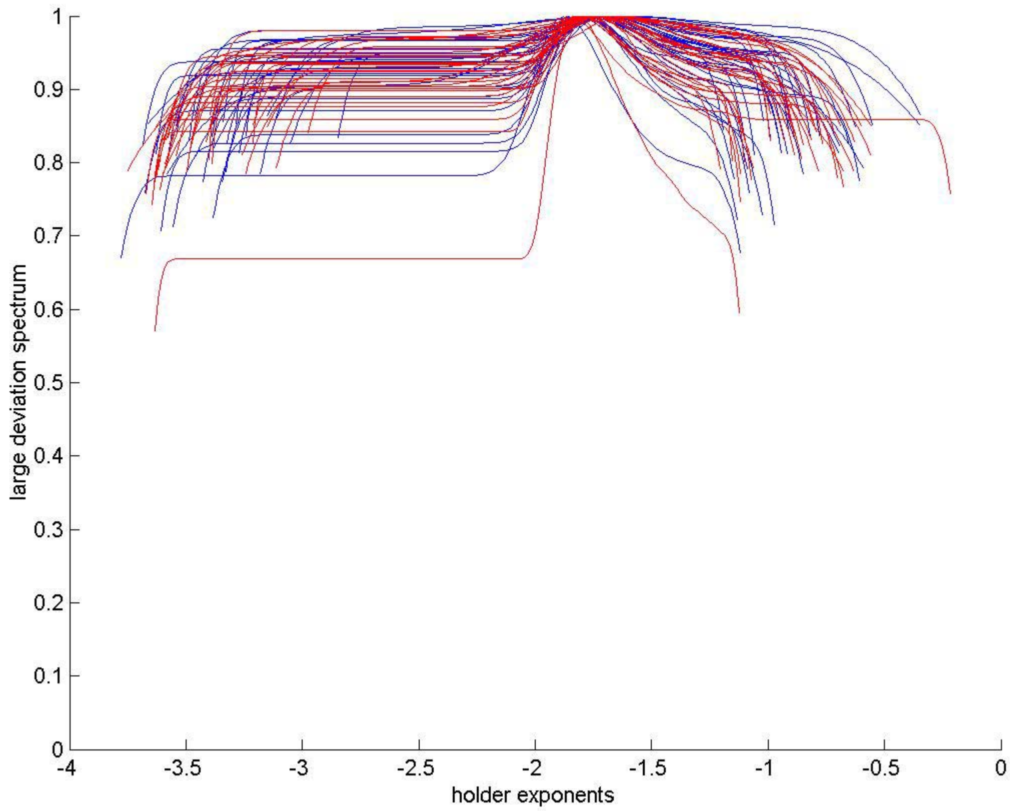


Figure: A superimposed graph of the large deviation spectrum for the new data set. <u>The red colored graphs correspond to the 2nd class trials and the blue graphs correspond to the class 1 trails.</u>

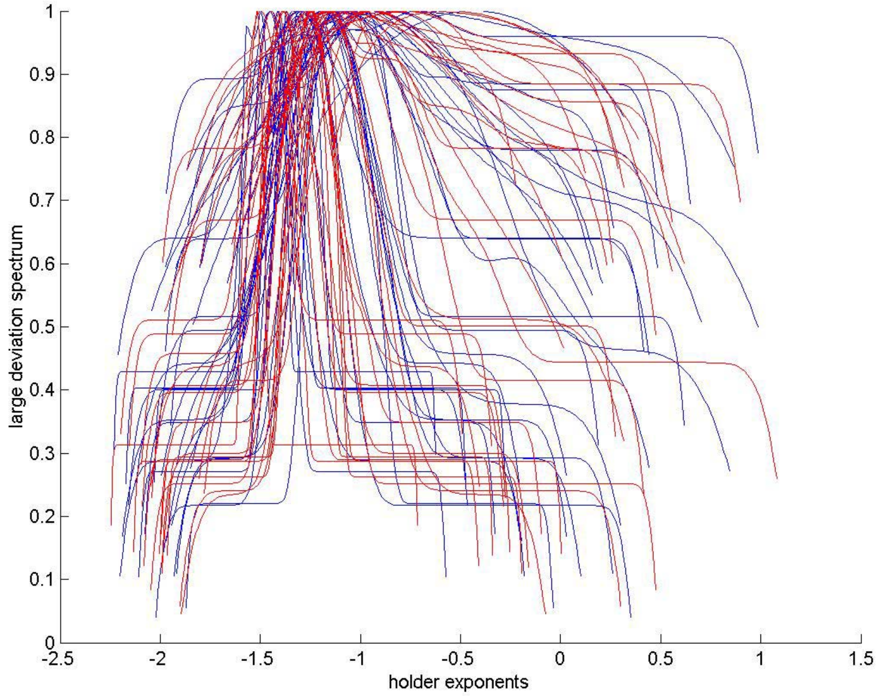


Figure: A superimposed version of the large deviation spectrum for the Left trials and the Right trials. The red graphs denote the Right finger movement trials. Large overlapping is observed in this case.

LEGENDRE SPECTRUM: It is a concave approximation to the large deviation spectrum.

The new data was analyzed with the Legendre spectrum to find out whether that could be a useful feature for classification. The results were negative. The following graphs show the following.

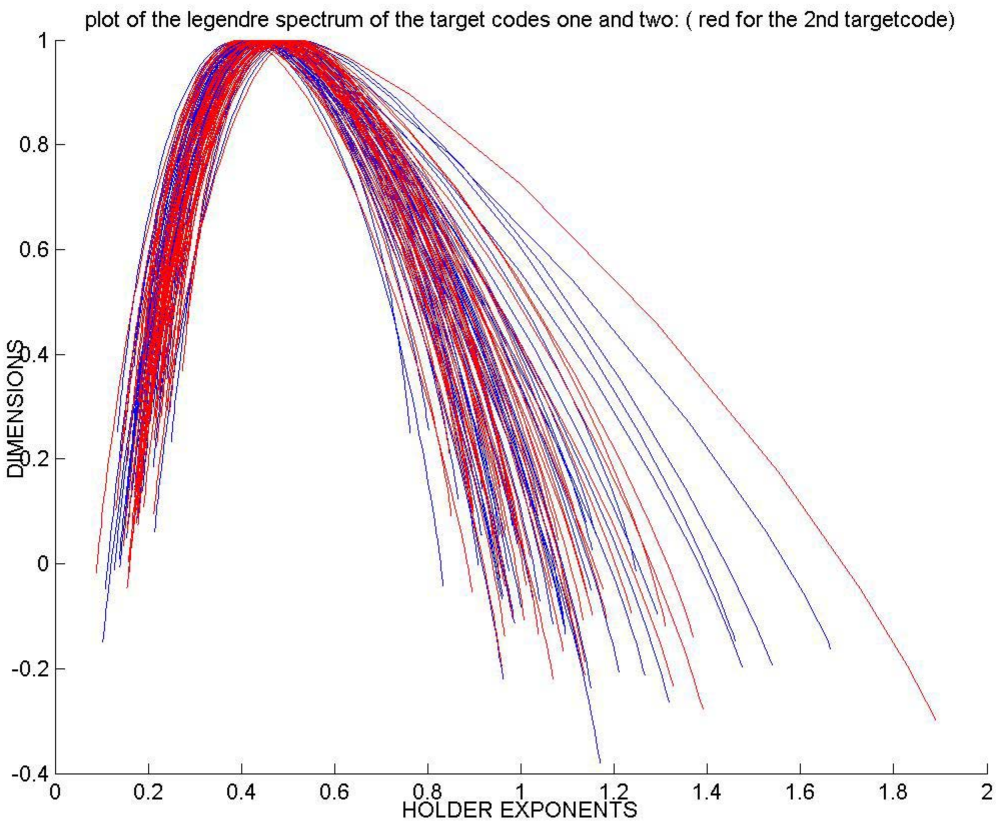


Figure: Superimposed graphs of the Legendre spectrum for the trials from the new data set. The above graph shows that the holder exponents do not work well with the new data set. <u>The red graphs correspond to the class 2 trials and vice-versa</u>. Too much overlapping is observed.

CLASSIFICATION USING THE GIFS COEFFICIENTS:

The new data set used for classifying the new data set (The first class and the second class). The gifs coefficients were used for as features. The gifs coefficients were averaged out for an interval of 32 samples and the new feature vector obtained was used for classification. The results obtained were poor. The training set consisted of 30 trials and the testing consisted of 17 trials. A radial basis support vector machine was used for classification. The percentage of correctly classified signals was 62%. This shows that the feature extraction using gifs coefficients does not provide robust features for classification.

ARTIFACT REMOVAL USING THE HOLDER EXPONENTS

Artifacts have been one of the major problems of EEG signal processing. Their removal is a must in order to use the EEG signal for processing and classification. Some potent methods like ICA and JADE have been suggested by some researchers for EEG data cleaning. The artifacts can also be removed using the fractal coefficients like the point wise holder function.

Artifacts are always accompanied by a morphological change in the EEG signal. This fact can be exploited to indicate when an artifact has crept into the signal. Point wise holder functions have a larger value when the signal becomes more regular. This is generally the case when artifacts are introduced. So, the value of the holder function can be used as an indication of an artifact. To expedite on the fact a model signal was analyzed with the point wise holder function. A model signal very close in nature to an EEG signal corrupted by artifact noise is used for the purpose. This can be easily seen in the following graphs:

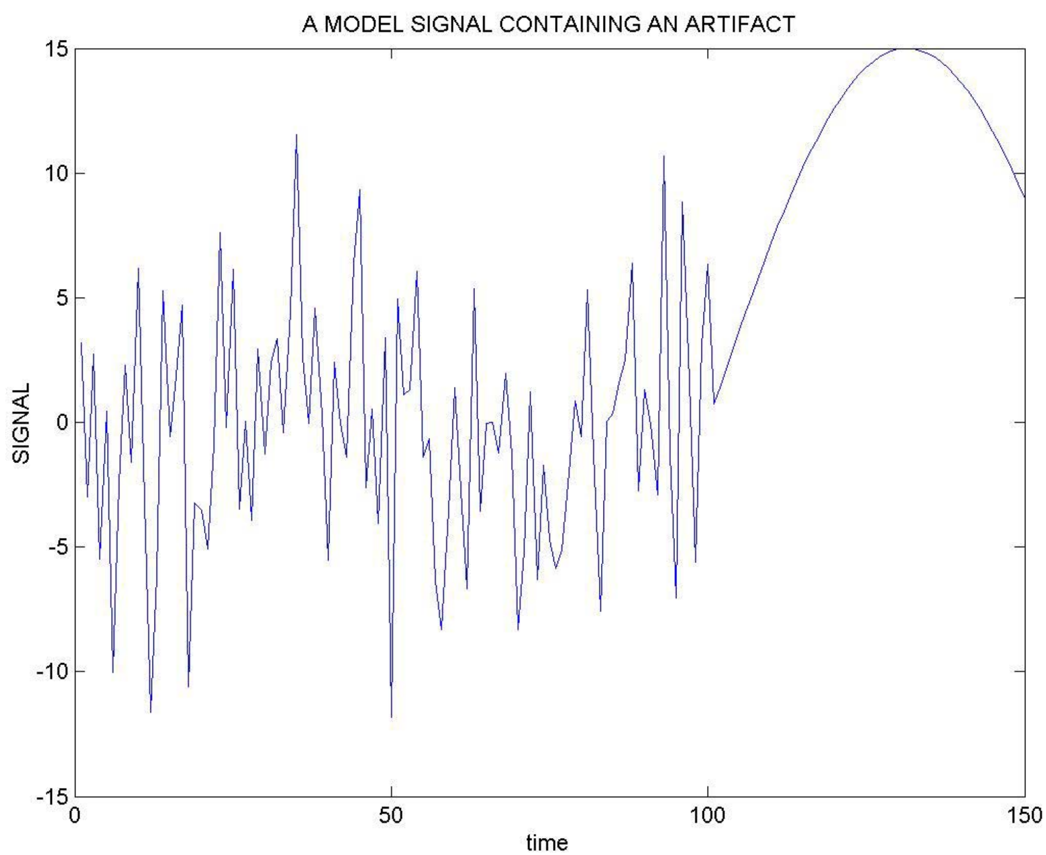


Figure: A model signal generated to test the utility of point wise holder function for artifact detection. <u>The later part of the signal shows the appearance of an ocular artifact</u>.

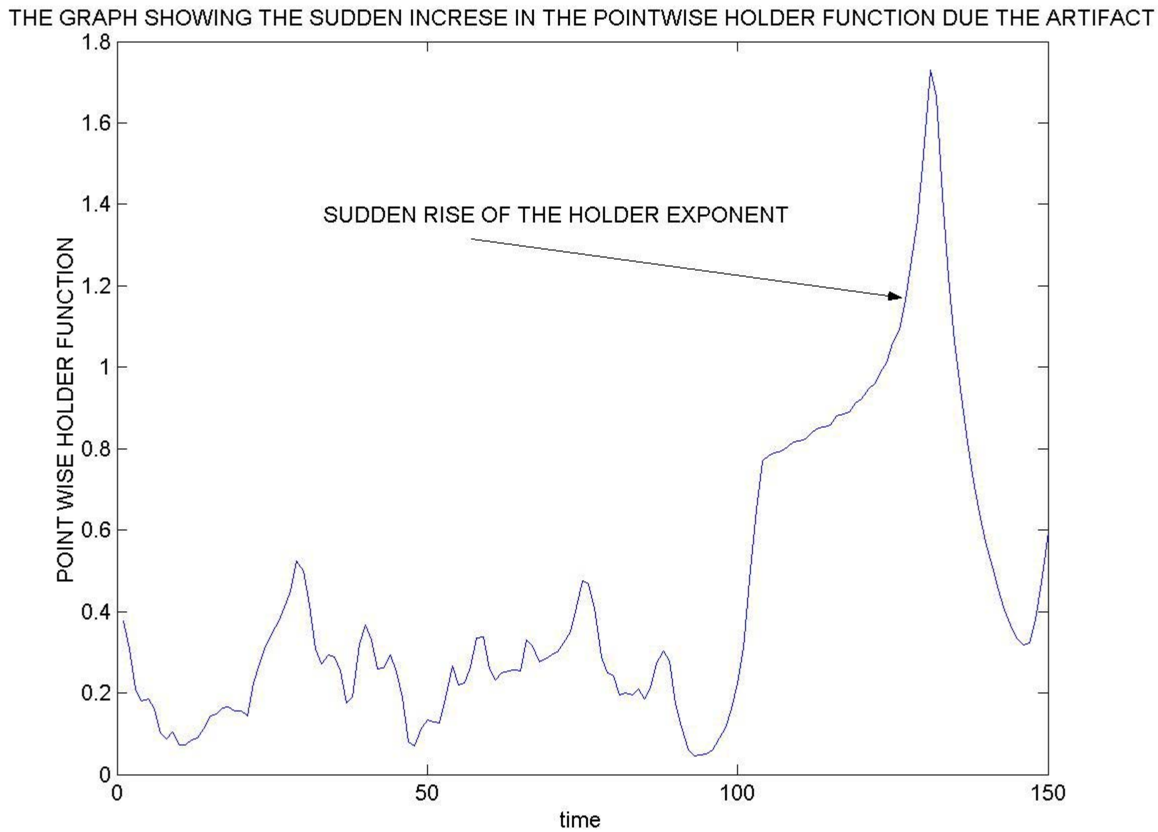


Figure: The above figure shows a sudden rise in the point wise holder function at the arrival of the artifact in the previous signal used to model an artifact.

CONCLUSIONS

A rigorous analysis of the EEG data, which I used, shows that simple frequency transforms like the FFT and the PSD are not always helpful in detecting major differences between the signals. They only bring out the overall frequency content of the EEG signal.

Although a parametric model using the autoregressive coefficients is helpful in extracting modeling the data well, they treat signals as decoupled from the other signal. Thus, an analysis of the signals using inter-electrode correlations is always better than using decoupled mono variant analysis of the signals. Thus, a multi variant analysis of signals always provide better distinguishing features than the mono variant analysis of the EEG signal. The better performance of the MVAR coefficients bears a strong testimony this fact.

A time frequency transform provides better analysis of the signals than a direct frequency transforms. They also provide a good view of the variation of the frequency content of the signal with time, which a direct frequency transform cannot. Thus, they are able to provide better contrasting features. An extraction of these features using the contrasting function was helpful in improving the classification results.

A run time analysis of the SVM machine shows that by choosing the right parameters we can better results than the neural networks. Also, they provide faster learning because of faster learning algorithms than the normal back propagation algorithm used to train an MLP.

A morphological analysis of the signal suggests that a regularity variation measure could be a useful tool for classification. This is only the case when we have good morphological difference between the signals. For signal with little or no morphological difference, the method provides little help.

The fractal/morphological analysis can be very helpful in detecting an ocular artifact because these are always accompanied by some changes in the morphological properties of the signal.