Fast Robust Pattern Classification Algorithms for Real Time Neuro-Motor Prosthetic Applications

Hariharan Nalatore[1], *Wilson Truccolo*[2] *AND Govindan Rangarajan*[3]

Abstract | In this paper, we give a brief review of pattern classification algorithms based on discriminant analysis. We then apply these algorithms to classify movement direction based on multivariate local field potentials recorded from a microelectrode array in the primary motor cortex of a monkey performing a reaching task. We obtain prediction accuracies between 55% and 90% using different methods which are significantly above the chance level of 12.5%.

1. Introduction

A significant challenge in applied neuroscience is to build prosthetic devices (e.g. limbs and computer interfaces) controlled by neural signals from the brain[1]. The ultimate goal is to provide paralytic patients and amputees with the means to move and communicate by controlling a prosthetic device using brain activity. Strong relationships between the activity of neurons in the brain's motor cortex and intended movement kinematics in patients with tetraplegia or in able-bodied non-human primates[2–12] have allowed scientists and engineers to get closer and closer to building such devices. Although there has been steady progress in developing both hardware and software needed to directly connect human brains to prosthetic devices[1,13,9,14–17], one of the key remaining issues is the need of fast and robust exploratory algorithms for classification of neural signals into discrete states (e.g. symbols for communication, discrete directional click and point devices) in real time. For example, one might want to train and test a discrete state controller based on few-minute long data segments and test it immediately in a closed-loop application (e.g. [13]). State of the art classification algorithms such as support vector machines are computational expensive and require a reasonable amount of parameter tuning[18,19]. On the other hand, algorithms based on classic discriminant analysis remain off-the-shelf tools for exploratory analyses because of their fast computation and robustness[20].

In this paper, we briefly review various pattern classification algorithms based on discriminant analysis. We then illustrate and compare these algorithms' performances on classification of movement direction based on multivariate local field potentials recorded from a microelectrode array in the primary motor cortex of a monkey performing a reaching task.

2. Pattern classification using discriminant functions

In the previous section, we reduced the problem of decoding movement direction from LFP data to a pattern classification problem. In this section, we briefly describe the theory behind such classification[21].

Given an arbitrary collection of individuals, we need to classify the individuals into various groups so that the individuals belonging to the same group have similar characteristics. The problem that is addressed with discriminant function analysis is how well is it possible to separate two or more

[1]*Applied Research International, New Delhi*
[2]*Department of Neuroscience, Brown University, Providence*
[3]*Department of Mathematics, Indian Institute of Science, Bangalore*

groups of individuals given measurements for these individuals on several variables. In the general case there will be m random samples from different groups of sizes $n_1, n_2, n_3, \ldots, n_m$ and values will be available for p variables $X_1, X_2, X_3, \ldots, X_p$ for each sample member.

2.1. *Discrimination using Mahalanobis distance*

The Mahalanobis distance[21] is frequently used to measure the distance of a single multivariate observation from the center of the population that the observation comes from. The mean vectors for the m samples can be regarded as estimates of the true mean vector for the groups. The Mahalanobis distance of individuals to group centres can then be calculated and each individual can be allocated to the group that is closest to. This may or may not be the group that the individual actually came from. The percentage of correct allocations is clearly an indication of how well groups can be separated using the available variables.

This procedure is defined as follows. Let

$$\boldsymbol{\mu}_i = [\mu_{1i}, \mu_{2i}, \ldots, \mu_{pi}]^T, \quad 1 \le i \le m$$

denote the vector of mean values for the sample from the ith group. Here T refers to the transpose operation. Let C_i denote the covariance matrix for the same sample. The covariance matrix can differ from group to group. However, for using Mahalanobis distance, we require all groups to have the same covariance. To get around this problem, we define $\mathbf{C}$ to be the pooled sample covariance matrix determined by the following equation:

$$\mathbf{C} = \frac{\sum_{i=1}^{m}(n_i - 1)\mathbf{C}_i}{\sum_{i=1}^{m}(n_i - 1)}$$

We use this pooled covariance matrix for all the groups instead of the individual $\mathbf{C}_i$'s. We assume that the classes are described by multidimensional Gaussian probability density function. The probability density function of the class i for an observation

$$\mathbf{x} = [x_1, x_2, \ldots, x_p]^T \tag{1}$$

is assumed to be given by the Gaussian pdf:

$$g_i(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}}|C|^{-1/2} e^{[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T C^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)]}$$

The Mahalanobis distance from an observation $\mathbf{x}$ (equation (1)) to the center of the group i is given by,

$$\begin{aligned} D_i^2 &= (\mathbf{x}-\boldsymbol{\mu}_i)^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) \\ &= \sum_{r=1}^{p}\sum_{s=1}^{p}(x_r - \mu_{ri})c^{rs}(x_s - \mu_{si}) \end{aligned}$$

where c^{rs} is the element in the rth row and sth column of $\mathbf{C}^{-1}$. Clearly,

$$g_i(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}}|\mathbf{C}|^{-1/2} e^{(-D_i^2/2)}$$

The observation $\mathbf{x}$ is allocated to the group for which D_i^2 has the smallest value. In other words, the observation $\mathbf{x}$ is allocated to that group which gives the maximum aposteriori probability.

2.2. *Discrimination using Gaussian quadratic norm*

Here, we calculate the probability density function of the class i for the observation $\mathbf{x}$ by the following relation:

$$g_i(\mathbf{x}) = -\frac{1}{2}ln|\mathbf{C}_i| - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T \mathbf{C}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)$$

In this classifier[21], the classes are not assumed to have the same covariance. The covariance of each class is denoted by $\mathbf{C}_i$ as before. We assign $\mathbf{x}$ to the class j if and only if

$$g_j(\mathbf{x}) \ge g_i(\mathbf{x})$$

for all $i = 1, 2, 3, \ldots, m$. As before, we assign $\mathbf{x}$ to that class (j) which gives maximum probability.

2.3. *Bhattacharya distance*

The data in the original space can be transformed into a feature space by extracting important information (also called significant features) from this data. Therefore, feature selection is an important consideration in any classification scheme, where one needs to choose a smaller subset of features from a complete set of raw measurements, such that the improved subset generates as good or better classification performance compared to original data. This also helps to reduce the dimensionality of the data set.

The Bhattacharya distance[21] is used as a measure of separability of the prior probability distribution of two classes from a given set of features (e.g. channel averaged power at different frequencies). It provides a bound of classification accuracy by taking into account the first and second order statistics. This distance is the sum of two components, one

primarily based on mean differences and the other based on covariance differences.

Under the assumption of multivariate normality (of the pdf's of different classes), the Bhattacharya distance is computed by the equation:

$$B = \frac{1}{8}[\boldsymbol{\mu}_f - \boldsymbol{\mu}_g]^T \left[\frac{\Sigma_f + \Sigma_g}{2}\right]^{-1} [\boldsymbol{\mu}_f - \boldsymbol{\mu}_g] + \frac{1}{2}\ln\left(\frac{|\frac{\Sigma_f+\Sigma_g}{2}|}{\sqrt{|\Sigma_f||\Sigma_g|}}\right).$$

Here $\boldsymbol{\mu}_f$, $\boldsymbol{\mu}_g$ are mean vectors and Σ_f, Σ_g are covariance matrices of the two features f and g. We pick those optimal features (e.g. channel averaged power at some optimal frequency) which maximises the Bhattacharya distance in the feature space.

3. The partitioning of feature space

To classify a point in feature space, we should be able to divide the feature space into an exhaustive set of non overlapping regions; one for each group of interest so that every point in the space is uniquely associated with one of the named classes (groups). The locus of points which separate a pair of groups is called a 'decision boundary'[21]. These are estimated using training samples.

The case of multi-group classification ($m > 2$) is intrinsically harder than binary ($m = 2$) classification because the classification algorithm has to construct more number of decision boundaries for which each group has to be explicitly defined. For m groups in the feature space, the probability of correctly classifying a point by chance is $1/m$.

The modeling of each group is nothing but the modeling of the probability density function for that group in feature space. It must be done in such a way that different groups are as distinct from one another as possible. How well a classifier will work depends on how well the true class density functions can be determined from the training samples.

In the case of the Gaussian Quadratic norm classifier, the decision boundary in the feature space will be a second order hyperspace and its form and location between the class mean values will depend on C_i's. This classifier is a maximum likelihood classifier.

4. The experiment

We consider an experiment in which a well trained monkey performs a Sensorimotor cognitive task as described below. This experiment was performed at John Donoghue's laboratory at Brown University, USA.

The monkey sat on a chair and in front of her, there was a vertical panel with a center button and 8 buttons that were radially placed around the center one. Reaching one of these 8 buttons corresponded to a movement direction of 0°, 45°, 90°, 135°, 180°, 225°, 270° and 315°. The 0° movement corresponded to moving horizontally to the right. The other directions were marked in an anticlockwise direction with respect to this 0° direction.

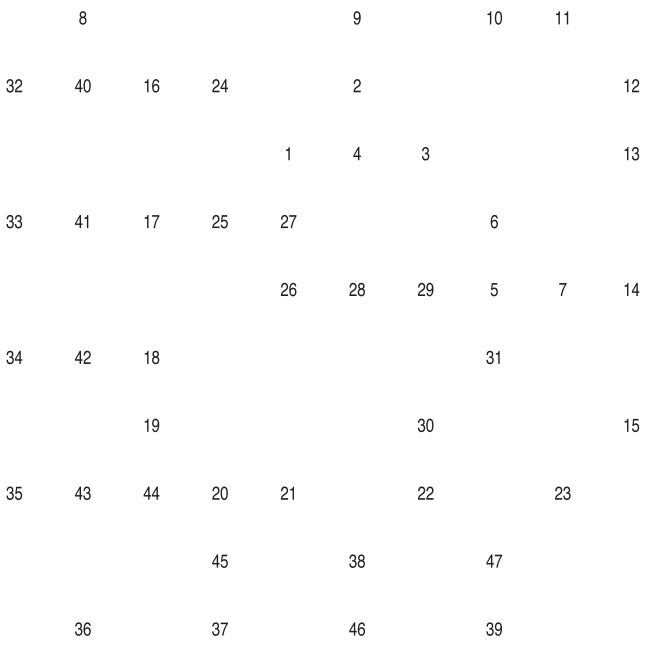
The sequence of events in the tasks was as it follows. First, the center button lit up for a short time. If the monkey reached the center button in less than 2 s, a trial began. The monkey had to hold the button for at least 500 ms, otherwise the trial was aborted. In any given trial, one of the radial buttons glowed randomly for 150 ms (this happened randomly in the time interval [500 ms, 1000 ms]). The start time of this event is here referred to as the "precue time". After this instructed delay, all the radial buttons glowed at once. This time is called the "gocue time". The monkey had to move its arm in that direction where the single radial button had lit up initially. If the monkey reached the correct radial button within 2 s after gocue, the trial was considered to be successful. On the other hand, if the monkey did not reach the correct radial button and held the center button for at least 50 ms, the target was aborted.

In this manner, 411 trials of successful hand reaching movements were performed by the monkey. In each trial, the multichannel Local Field Potentials (LFP's) (which can be thought of as embodying the collective synaptic input of local neuronal clusters[22]) were recorded from a 96-microelectrode array[17] chronically implanted in the primary motor cortex of the monkey which is contralateral to the moving arm. All correct trials were rewarded with juice. Of the 47 channels of data that were recorded, 43 were found to be good for further data analysis.

Figure 1 shows the spatial map of the multichannel electrodes during a recording session.

By considering a single trial multichannel LFP over a brief time interval of about 250 ms prior to gocue (i.e. during the instructed delay period), our goal is to decode the direction of hand movement based on spatiotemporal correlation patterns within the LFP signal. We approach this as a pattern classification problem. First we train a classifier using a training data set (which is a subset of the full LFP data). Once this is done, we assign each individual trial in the remaining data set (called the testing data set) to one of the eight possible directions using various pattern classification algorithms. Moreover, this needs to be accomplished before the hand movement actually starts. Since we know the end result for each trial, we then determine the probability of successful prediction. In our case, since we have to predict the direction of movement in one of eight possible directions, the probability of obtaining the correct result by chance is 0.125.

Figure 1: Spatial locations of the 47 electrodes (out of a total of 100 electrodes in the implanted microelectrode array) used in our study.



5. Results

In our approach, to predict the direction of monkey's movement target, we employ the Bayesian classifiers (Mahalanobis distance and Gaussian quadratic norm as described earlier) with uniform prior probability distribution. We assume the power spectrum to be normally distributed in the entire frequency range. The 8 directions are assumed to be equally likely.

The entire collection of 411 trials is divided into two roughly equal sets, namely the training set and the testing set. The training set is used to train the classifier by determining the parameters (mean vector and covariance matrix) for each direction (group). The testing set is used to evaluate the performance of the classifier.

5.1. *Single frequency feature space approach*

To predict the direction of monkey's target, we proceed as follows. The classifier used here is the Mahalanobis distance. By using the multitrial multichannel Local Field Potential (LFP) recordings between the time interval [250 ms, 500 ms] during the instructed delay period from the training set, we determine the power at a given frequency in the gamma band [31 Hz, 55 Hz] for each channel, each trial and each direction using the multitaper method. The feature used here to obtain the prediction is the multichannel power at a given frequency f_a in the gamma band. The reason for choosing the gamma band is that it has been observed[23,24] to play an important role during the working memory period (i.e. during the time when we wish to predict the movement direction). Figure 2 shows the power spectrum of the monkey during 0–500 ms in the instructed delay period.

By using the multichannel, multitrial power at this frequency from the training data set, we estimate the parameters of the classifier for each direction. We use frequencies in the gamma band since earlier studies have shown that this frequency range hold significant promise for movement prediction.

In order to test the performance of the classifier, we do the actual classification of an arbitrary single trial data from the testing data set as described below. Firstly, we choose an analysis window of 150 ms duration that slides along the time axis of the data (by 2 data points at a time) from 250 ms to 500 ms. This time interval has been chosen such that it is prior to the actual start of the monkey's hand movement. A small moving time window is used for the following reason. As the data is non-stationary, we analyze the single trial data by using highly overlapped time windows of 150 ms duration. Within each time window, the underlying stochastic process generating the data is assumed to be stationary. Hence all the standard techniques from stochastic processes can be used within each time window.

For the data in each time slice, by using the multichannel power at the frequency f_a (computed by multitaper method), we calculate the Mahalanobis distance for all directions and determine that direction which gives the minimum distance. We then predict that the monkey is about to move its arm in the above direction. This procedure when repeated for all trials, enables to calculate the probability of correct prediction since we know the actual direction in which the monkey had moved its arm for each trial. This procedure when repeated for all time windows, gives the temporal evolution of the probability of correct prediction for all directions of analysis as the analysis window slides along the time axis.

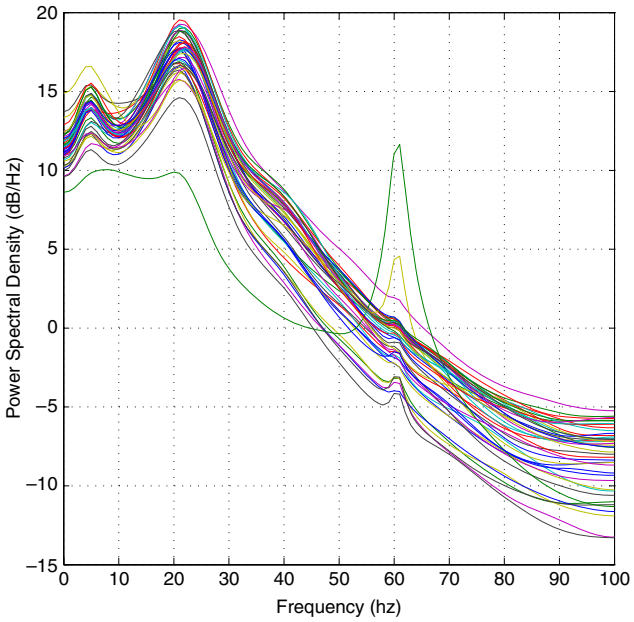
This procedure is repeated for each frequency f_a in the gamma band. We pick that frequency feature f_a^* for which the predictability is maximum. The performance of the classifier is shown in Figures 3 and 4 for various directions.

5.2. *Principal component analysis of single frequency*

5.2.1. *Theory*

In the previous method, we used the power in all the channels at a particular frequency. Since

Figure 2: Power spectrum in all channels for 45 degree direction during instructed delay period.



the corresponding electrodes are all closed spaced (within a distance of approximately 4 mm), the powers in different channels can be quite similar (i.e. not independent) and this can lead to over fitting. To avoid this, principal component analysis can be used to reduce the number of variables.

The objective of Principal component analysis[25] is to take p variables $X_1, X_2, X_3, \ldots, X_p$ and find combinations of these to produce new variables $Z_1, Z_2, Z_3, \ldots, Z_p$ that are uncorrelated. These new variables are ordered so that Z_1 explains the largest amount of variation in the data, Z_2 explains the second largest amount of variation and so on. That is $\mathrm{Var}(Z_1) \geq \mathrm{Var}(Z_2) \geq \ldots \geq \mathrm{Var}(Z_p)$. The new variables Z_i's are called principal components. In performing principal component analysis, there is a possibility that the variances of most Z_i's will be so low as to be negligible. In such a case, the variation in the data set can be adequately described by the first few Z_i variables whose variances are not negligible. Hence, principal component analysis is used to reduce the dimensionality of the problem and to transform interdependent coordinates into significant and independent ones.

The ith principal component Z_i for $(1 \leq i \leq p)$ is given by

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{ip}X_p.$$

The $\mathrm{Var}(Z_i)$ is as large as possible subject to the constraint that

$$a_{i1}^2 + a_{i2}^2 + a_{i3}^2 + \cdots + a_{ip}^2 = 1.$$

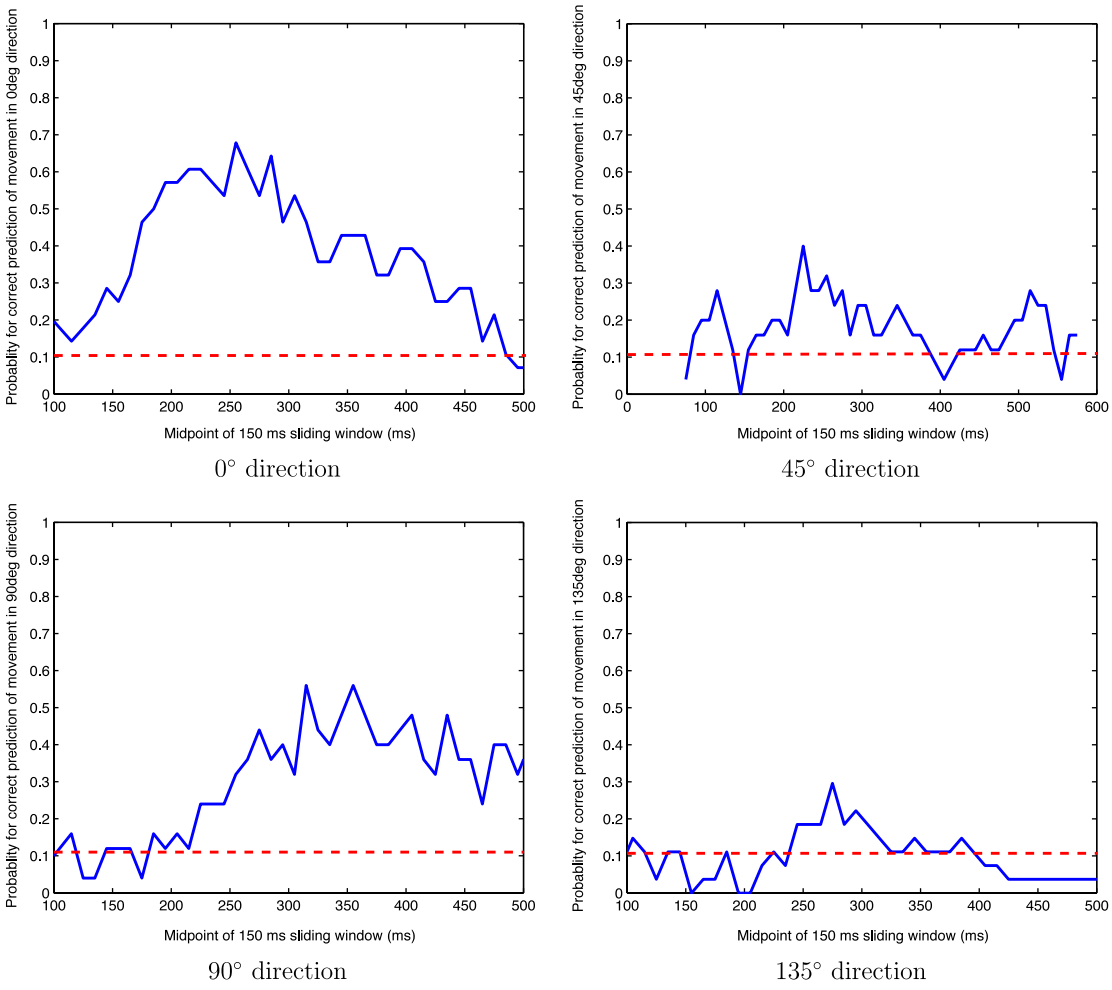
Also, Z_i is uncorrelated with $Z_{i-1}, Z_{i-2}, \ldots Z_2$ and Z_1. Without this constraint on the coefficients a_{ij}, we can see that $\mathrm{Var}(Z_i)$ can be increased by increasing one of the a_{ij} (for a given i).

The determination of the coefficients a_{ij} is an eigen value problem of the sample covariance matrix C. The variances of the principal components are the eigen values of the matrix C. Let the eigen values of C be arranged as $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_p \geq 0$. Now, λ_i corresponds to the ith principal component Z_i. The coefficients $a_{i1}, a_{i2}, \ldots, a_{ip}$ are the elements of the corresponding eigen vector (these are scaled so that $\sum_{j=1}^{p} a_{ij}^2 = 1$).

5.2.2. *Implementation*

By considering the multitrial multichannel LFP recordings over the time interval of [250 ms, 500 ms], we compute the power in the gamma frequency band using multitaper method for all the good channels and for all trials belonging to the training set. We choose a frequency f_a in the gamma band. Using the multichannel power at the frequency f_a, we transform this by principal component analysis and pick 5 principal components Z_i's. Each of these principal components is a linear combination of the power in different channels at the frequency f_a. The first 5 components explain between 80 to 90 percent of the total variance across the eight directions. By evaluating these principal components for all trials belonging to various directions, we determine the

Figure 3: Time evolution of probability of correct prediction by single frequency feature space approach for various directions. The horizontal dashed line indicated chance level.



parameters of the classifier for various directions. Here again, we use Mahalanobis distance as the classifier.

To test the performance of the classifier on the testing set, we perform the actual classification by considering each single trial data belonging to the testing set. We proceed as follows. From the single trial multichannel LFP recordings, to get the temporal evolution of the prediction, we divide the data using highly overlapped time windows of 150 ms duration. For the data in each time slice, the multichannel power at the frequency f_a computed by multitaper method is transformed to get 5 principal components from which the Mahalanobis distance is determined for various directions. The given trial is assigned to that direction for which the Mahalanobis distance is minimum. This classification when repeated for all single trial data for all time slice gives the evolution of probability of correct prediction of monkey's movement direction.

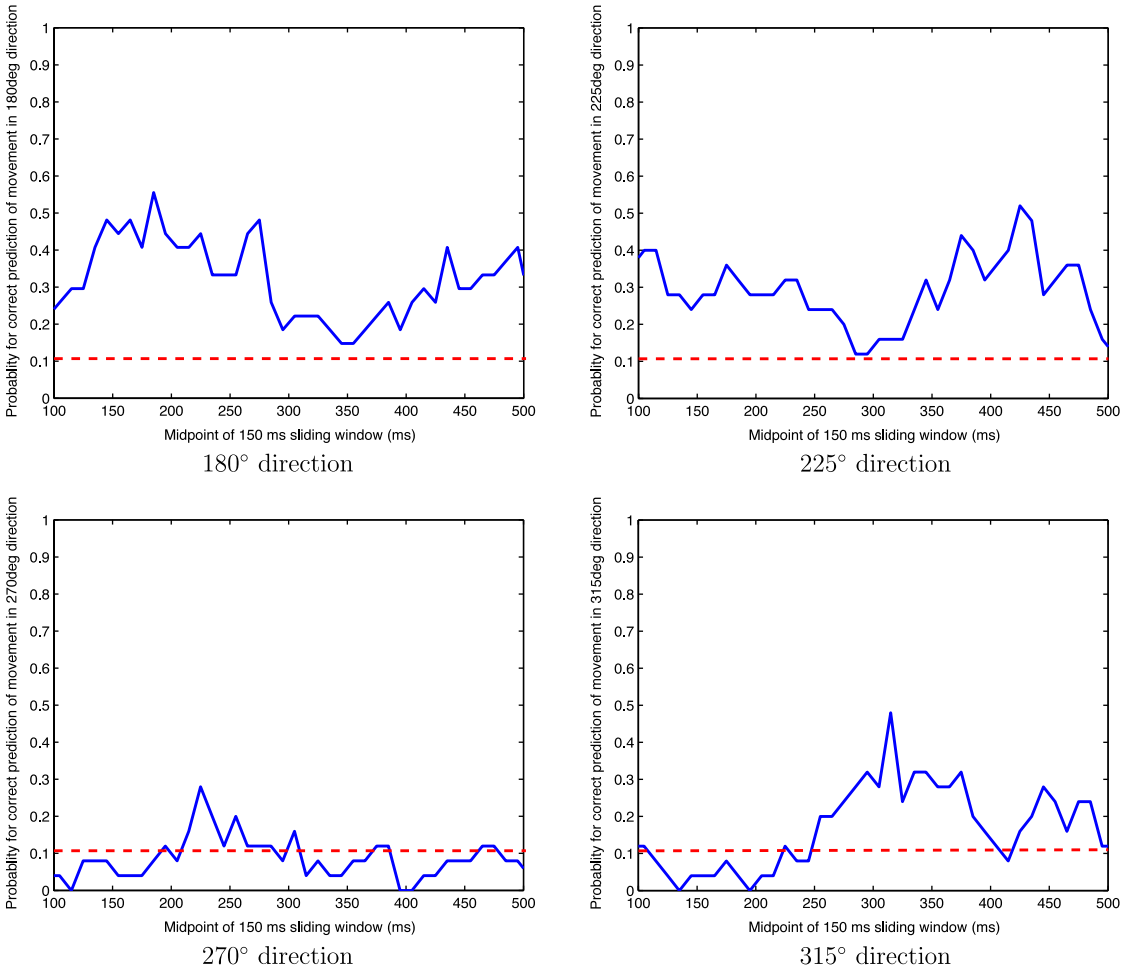
This procedure is repeated for various frequencies f_a in the gamma band, and the optimal frequency that gives the best predictability is chosen. The graphs of probability of correct prediction for various analysis direction are shown in Figures 5 and 6.

5.3. *Prediction by Canonical Discriminant function method*

5.3.1. *Theory*

Next, we consider an alternative method for pattern classification—the canonical discriminant function method[25]. Here, we determine function Z_i's which are linear combination of the variables $X_1, X_2, X_3, \ldots, X_p$ that enable us to separate the m directions (groups) as much as possible. Hence Z

Figure 4: Time evolution of probability of correct prediction by single frequency feature space approach for various directions. The horizontal dashed line indicated chance level.



takes the form

$$Z = a_1X_1 + a_2X_2 + a_3X_3 + \cdots + a_pX_p$$

Groups can be well separated using Z if the mean value changes considerably from group to group,with the values within a group being fairly constant. The coefficients $a_1, a_2, a_3, \ldots, a_p$ are determined so as to maximize a suitable F-ratio for a one way analysis of variance.

By this approach, one can determine several linear combinations for separating groups. In general, the number is s which is given by $\min(p, m-1)$. These s functions are called the canonical discriminant functions. The first function

$$Z_1 = \sum_{j=1}^{p} a_{1j}X_j$$

gives the maximum possible F-ratio on a one-way analysis of variance for the variation within and between groups. If there is more than one function, then the second one

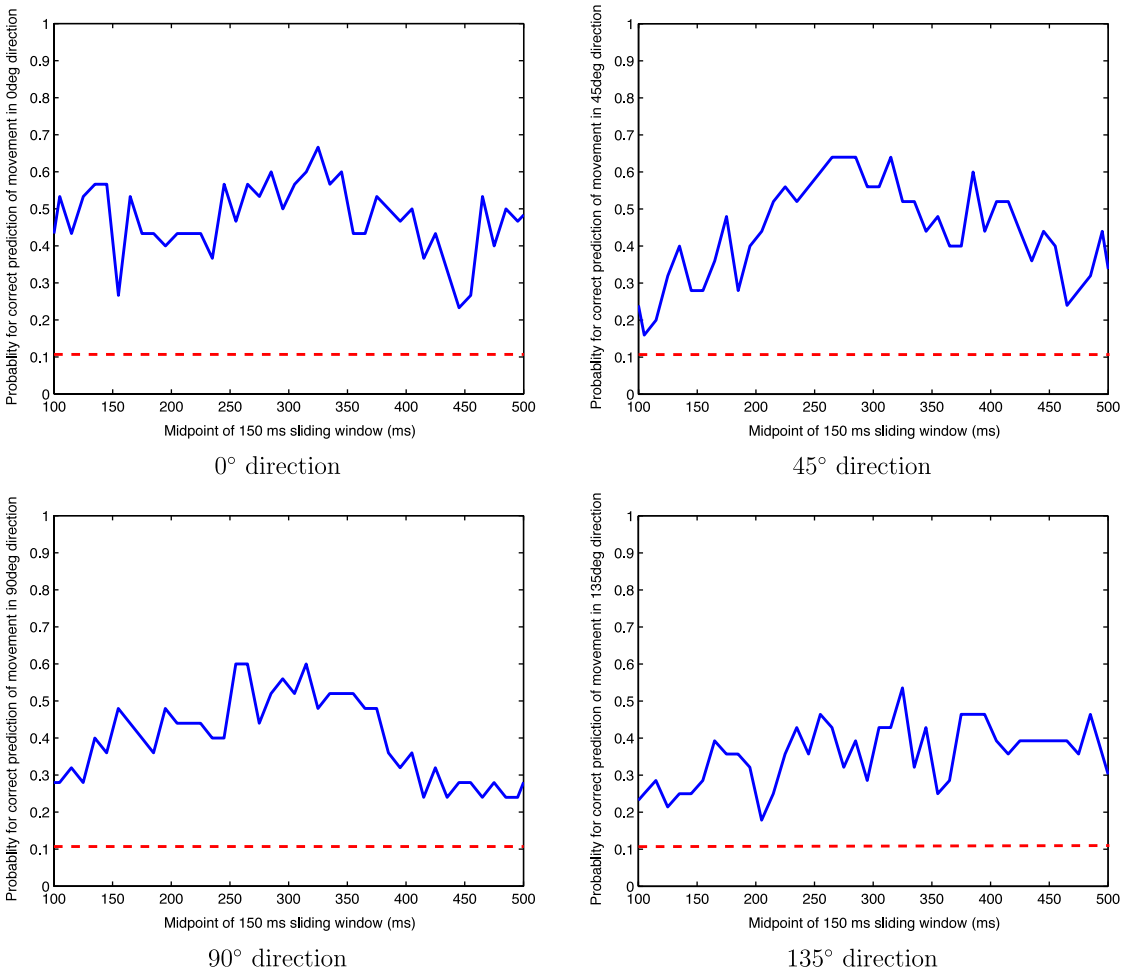
$$Z_2 = \sum_{j=1}^{p} a_{2j}X_j$$

gives the maximum possible F-ratio on a one-way analysis of variance subject to the condition that there is no correlation between Z_1 and Z_2 within groups. The other functions are defined in the same way. Thus the ith canonical discriminant function

$$Z_i = \sum_{j=1}^{p} a_{ij}X_j$$

is the linear combination for which the F-ratio on an analysis of variance is maximised, subject to

Figure 5: Time evolution of probability of correct prediction by principal component analysis approach for various directions. The horizontal dashed line indicated chance level.



Z_i being uncorrelated with $Z_1, Z_2, Z_3, \ldots$ and Z_{i-1} within groups. Thus the functions $Z_1, Z_2, Z_3, \ldots, Z_s$ are chosen in such a way that Z_1 reflects group difference as much as possible; Z_2 captures as much as possible of the group differences not displayed by Z_1; Z_3 captures as much as possible of the group differences not displayed by Z_1 and Z_2; and so on.

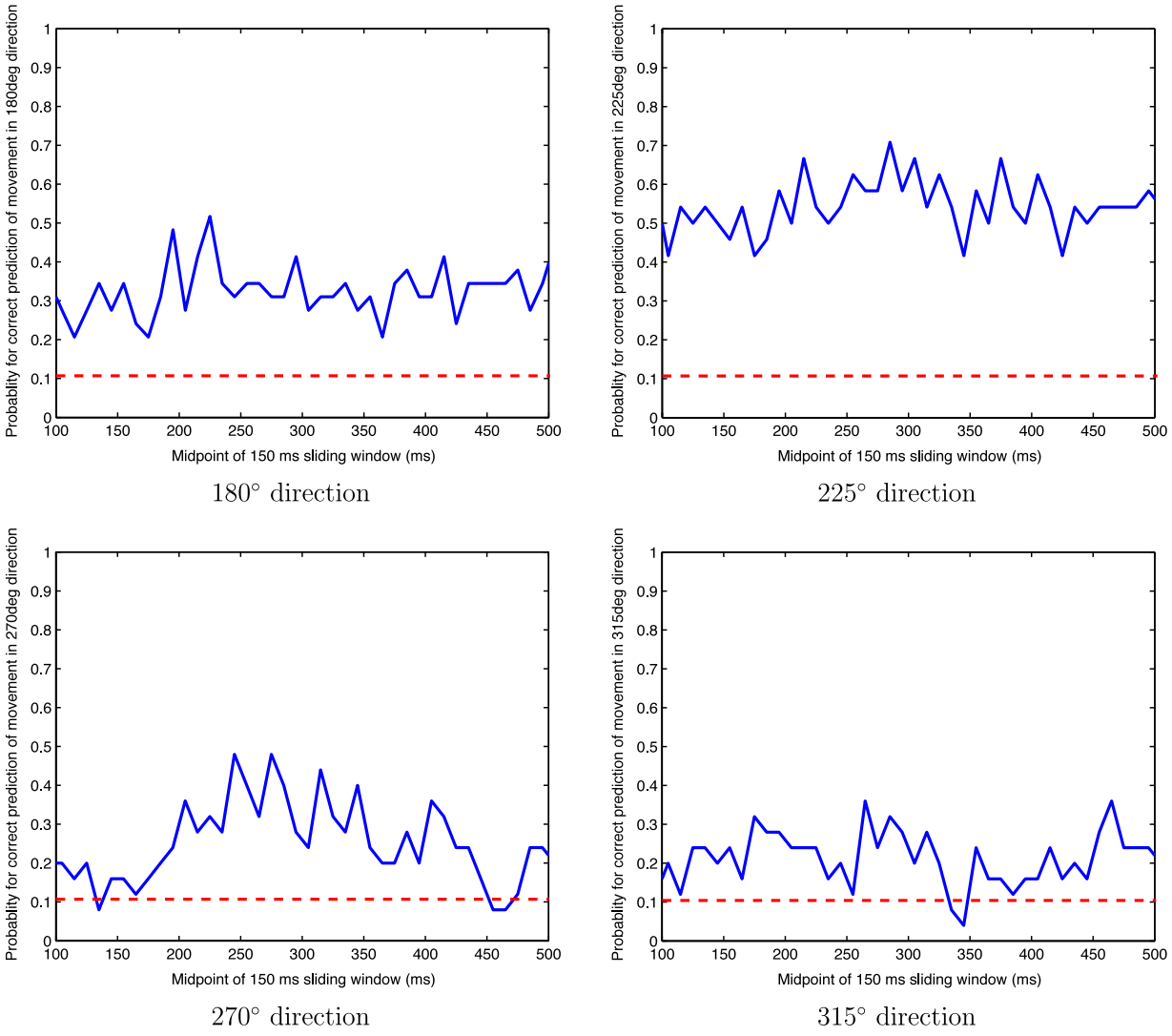
5.3.2. *Implementation*

Given the frequency f_a in the gamma band, by using the multichannel power at this frequency, we determine the coefficients of the canonical discriminant function. Since, **Z** is a vector CDF, we use Mahalanobis distance to classify the single trial LFP data from this. By determining the parameters for all directions from the training set, we do the actual classification for a single trial LFP data from the testing set. By considering the data in various time slices of 150 ms duration, we get the evolution of the probability of the correct prediction. This procedure is repeated for each frequency f_a in the gamma band and we pick the optimal frequency feature which gives best predicability of the monkey's movement target. It was found that this procedure does not give good predictions. Hence we have not displayed the figures.

5.4. *Histogram approach*

This method is also a single frequency feature space approach. Here we do not use any classifier. Instead we determine the prior probability density function for each directions as follows. We choose a frequency f_a in the gamma band. By using the multitrial multichannel power at this frequency, we average the power across all channels for each trial in the training set. By considering the collection of trials in a given direction, we get an estimate of the pdf's for each direction by using the channel averaged power

Figure 6: Time evolution of probability of correct prediction by principal component analysis approach for various directions. The horizontal dashed line indicated chance level.



(at frequency f_a). Hence, we have an estimate of the pdf's for each direction (group) from the training set.

Given a single trial LFP recording, we choose the multichannel recording in a given time window (slice). By evaluating the channel averaged power at the frequency f_a, we assign this trial to that direction that gives the maximum aposterior probability. In the same manner, the direction allocation of this data is repeated for all trials in the testing set.

Repeating this procedure for all time windows enables us to get the temporal evolution of the probability of correct prediction. This procedure is repeated for all frequency in the gamma band and the frequency that gives best predictability is chosen. The results obtained are shown in the Figures 7 and 8.

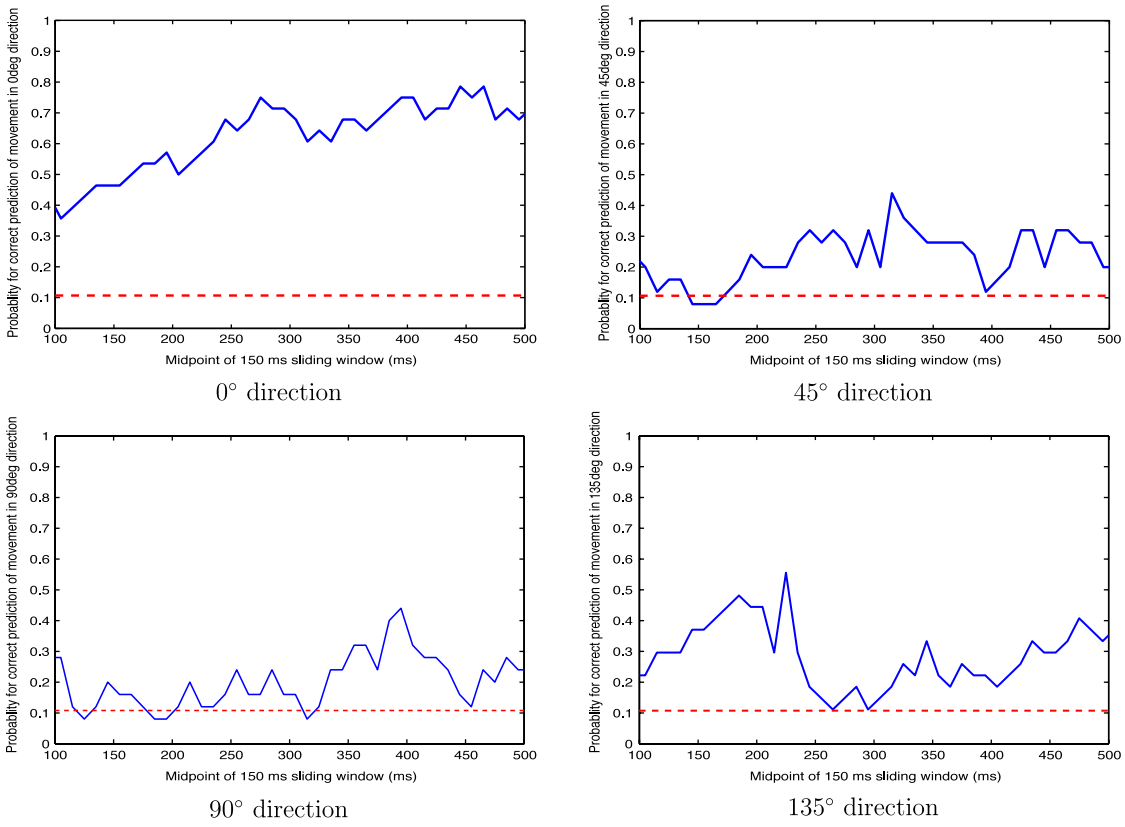
5.5. *General remarks on subsequent methods*

We now describe other methods that we used to classify the testing set data and obtain predictions of movement directions. However, none of these methods gave results better than the first few methods. Hence only a brief description of the method will be given for the sake of completeness. No figures will be included.

The following key steps are common in each of the subsequent methods for decoding monkey's movement direction:

(1) By using the multitrial multichannel LFP recordings between the time intervals [250 ms, 500 ms] during instructed delay, we compute the power spectrum by multitaper method for all channels and trials. The power in the gamma

Figure 7: Time evolution of probability of correct prediction for various directions by histogram approach. The horizontal dashed line indicated chance level.



frequency band [31 Hz, 55 Hz] is used during the classification of single trial data in various directions.

(2) While doing actual classification on training set, we divide the single trial multichannel data by using highly overlapped time windows of 150 ms duration as the signal is non-stationary. We classify the trial using some algorithm in each time slice. As we slide the time window across the data, we obtain the temporal evolution of the probability of correct prediction for each of the methods.

5.6. *Two frequency feature space approach*

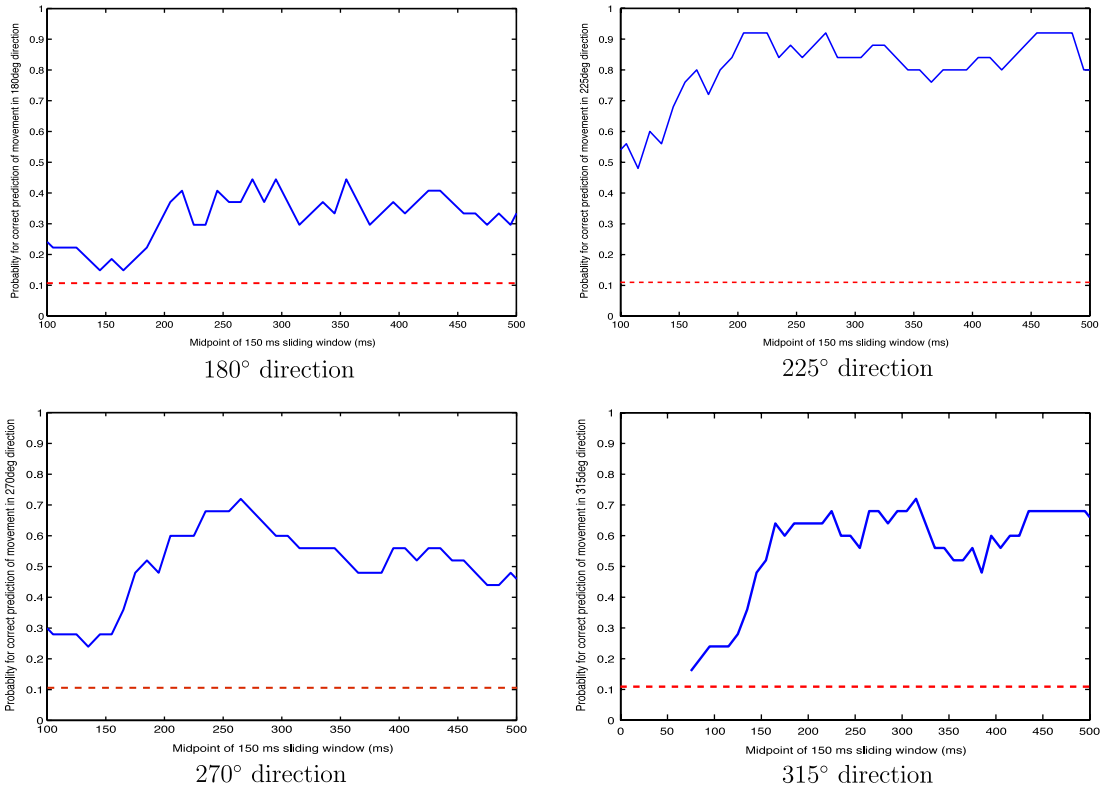
In this method, to choose a set of optimal frequencies, (similar to the single frequency f_a^* in the previous methods) we make use of Bhattacharya distance. First, the power in the gamma band is averaged over channels for all trials and frequencies. We want to pick a pair of frequencies (f_1^*, f_2^*) so that the channel averaged power at the above optimal pair of frequencies separates the pdf's of each direction the most. To do this, we evaluate the Bhattacharya distance using channel averaged power at all possible pairs of frequencies (f_1, f_2). The optimal pair of frequencies (f_1^*, f_2^*) corresponds to those frequencies for which the Bhattacharya distance is maximum. The classifier used here is the Gaussian quadratic norm. The parameters of the classifier are evaluated using channel averaged power at this pair of optimal frequencies for each direction. The time evolution of the probability of correct prediction is obtained in the usual manner.

5.7. *Four channel feature space approach*

The best feature subset is obtained as follows. We sum the power over frequencies in the gamma band for each channel and trial from the training set. Here, the frequency summed power of the ith channel $(1 \leq i \leq 43)$ is considered as the ith feature. We form all possible 4-tuple channel combinations of these frequency summed powers. For each such 4-tuple channel combination, we obtain the pairwise separability using Bhattacharya distance between all possible pairs of directions.

We consider that direction pair for which the average Bhattacharya distance across all channel combinations is the least. Then for this optimal direction pair, we choose that channel combination

Figure 8: Time evolution of probability of correct prediction for various directions by histogram approach. The horizontal dashed line indicated chance level.



which gives the largest distance. Here, we are trying to maximize the successful classification for that pair of direction which are hardest to classify. By using the frequency summed power of this optimal 4-channel combination, we determine the parameters of the Gaussian quadratic norm for each direction. The time evolution of the probability of correct prediction is obtained in the usual manner.

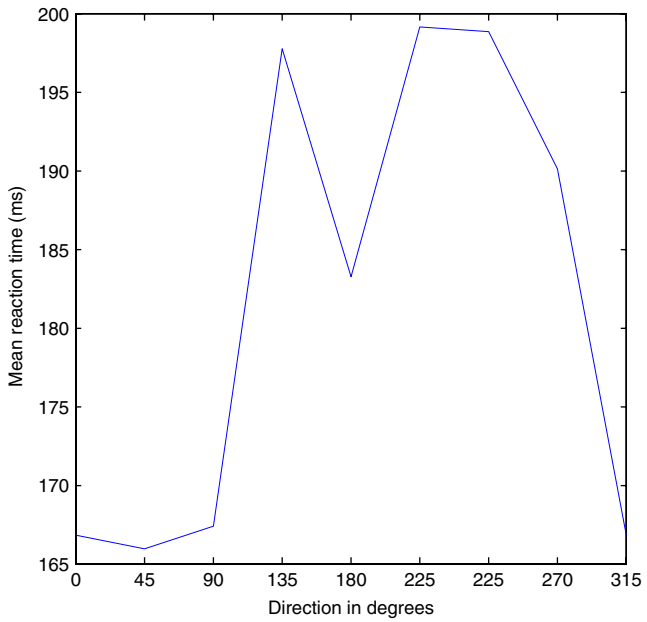
5.8. *3 Channel 1 frequency feature space approach*
Here, the power at a given frequency f_1 of the ith channel is considered to be the ith feature to be used. We form a 3-tuple channel combination of power at the frequency f_1 and obtain pairwise separability using Bhattacharya distance between all pairs of direction. This pairwise separability distance is averaged across the pair of directions for all 3-tuple channel power (at frequency f_1). We pick that 3-tuple channel combination of power at the frequency (f_1^*) for which the across direction average distance is maximum. Using this optimal 3-tuple channel combination of power at the frequency f_1^*, we determine the parameters of the Gaussian quadratic norm for all directions using the training set.

5.9. *3-channel 2-frequency feature space approach*
Here also, the ith channel power at a given frequency is considered to be the ith feature to be used. We form a 3-tuple channel combination of power at a pair of frequencies (f_1, f_2) and obtain pairwise separability using Bhattacharya distance between all pairs of directions. This pairwise separability distance is averaged across the pair of directions for all possible 3-tuple combinations (of power at al pairs of frequencies). We pick the optimal 3-tuple channel combination of power at the pairs of frequencies (f_1^*, f_2^*) for which the across direction average distance is a maximum. Using this 3-tuple channel combination of power at the pair of frequencies (f_1^*, f_2^*), the parameters of various directions are determined. We again make use of the Gaussian quadratic norm as the classifier. The time evolution of the probability of correct prediction is obtained in the usual manner.

6. Reaction times

The reaction time is defined as the time difference between the go cue and the start of hand movement. It represents the total time taken by the visual signal to reach the brain and for the appropriate

Figure 9: Plot of mean reaction times for the eight directions.



command to reach the hand. We wish to correlate this time with the preferred and anti-preferred directions for monkey's movement as obtained through LFP's in the gamma band. The preferred (anti-preferred) direction is the direction for which the relative power in the gamma band is the maximum (minimum). From the data, the preferred (anti-preferred) direction is found to be 45° (270°). The plot of the mean reaction times for the eight directions is shown in figure 9. We see that 45° direction has the minimum mean reaction time. This corresponds to the preferred direction for the monkey. On the other hand, 225°/270° directions have the maximum mean reaction time. This corresponds to the anti-preferred direction for the monkey. Thus the intuitive expectation that the preferred (anti-preferred) direction should have the minimum (maximum) reaction time is borne out by the data.

7. Conclusions

In this paper, we considered the problem of predicting the direction of a monkey's hand movement using a single trial of multichannel LFP's recorded from its motor cortex. We approached this problem as a pattern classification problem and studied fast and robust algorithms for decoding the movement direction based on discriminant analysis. We observed that the histogram method was able to decode the direction better than other methods for most directions. For directions where it performed well, the percentage of correct prediction ranged between 55% and 90%. However, it did poorly for 90° and 180° directions and not so well for 45° direction. For these directions, the PCA method performed better. For these directions, the percentage of correct predictions by PCA ranged from 50% to 65%. In all cases, we were able to obtain predictions significantly above the chance level (which gives the probability of correct prediction to be only 0.125). Further, we observe a distinct time evolution pattern of the probability for correct prediction in almost all the cases. The probability becomes maximum at some time prior to the start of the actual hand movement as expected in the case of recordings from the motor cortex. We hope to study in the future the application and performance of these algorithms in the context of human closed-loop neuro-motor prosthetic devices.

8. Acknowledgements

GR's work was supported in part by grants from DRDO, DST (SR/S4/MS:419/07) and UGC (under SAP-DSA Phase IV). He is also associated with the Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore as a Honorary Faculty Member. WT's work was supported in part by NIH-NINDS 1K01NS057389-01A1.

Received 30 November 2007; revised 11 December 2007.

References

1. Hochberg, L. R. Serruya, M. D. Friehs, G. M. Mukand, J. A. Saleh, M. Caplan, A. H., Branner, A. Chen, D. Penn, R. D. Donoghue, J. P. (2006) Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature, 442, 7099, 164–171.
2. Truccolo, W., Friehs, G. M., Donoghue, J. P., Hochberg, L. R. (2008, in press) Primary motor cortex tuning to intended movement kinematics in humans with tetraplegia. J of Neuroscience.
3. Paninski, L., Fellows, M. R., Hatsopoulos, N. G., Donoghue, J. P. (2004) Spatiotemporal tuning of motor neurons for hand position and velocity. J Neurophysiol 91: 515–532.
4. A. P. Georgopoulos, *Ann. Rev. Neurosci*, **14**, (1991), 361–377.
5. Moran D. W. and Schwartz A. B. (1999) Motor cortical representation of speed and direction during reaching. J Neurophysiol 82: 2676–2692.
6. Mehring, C., Rickert, J., Vaadia, E., de Oliveira, S. C., Aertsen, A. and Rotter, S. (2003) Inference of hand movements from local field potentials in monkey motor cortex. Nature Neuroscience, 6 (12): 1253–1254.
7. Serruya, M. D., Hatsopoulos, N. G., Paninski, L., Fellows, M. R., Donoghue, J. P. (2002) Instant neural control of a movement signal. Nature 416, 141–142.
8. Truccolo, W. A., Saleh, M., Hatsopoulos, N., Donoghue, J. P. (2002). Relationships between LFPs, Spike Activity and Movement Direction in Primary Motor Cortex. Society for Neuroscience Annual Meeting, Abstract Number 357.3.
9. Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. Journal of Neurophysiology 93(2): 1074–1089.

10. Scherberger H., Jarvis M. R., Andersen R. A. (2005) Cortical local field potential encodes movement intentions in the posterior parietal cortex. Neuron 46, 347–354.
11. Mehring C., Nawrot M. P., Cardoso de Oliveira S., Vaadia E., Schulze-Bonhage A., Aertsen A., Ball T. (2005) Comparing information about arm movement direction in single channels of local and epicortical field potentials from monkey and human motor cortex. J. Physiol. 98, 498–506.
12. Rickert J., Cardoso de Oliveira S., Vaadia E., Aertsen A., Rotter S., Mehring C. (2005) Encoding of movement direction in different frequency ranges of motor cortical local field potentials. J. Neurosci. 25, 8815–24.
13. Kim, S. P., Simeral, J. D., Hochberg, L. R., Donoghue, J. P., Friehs, G. M., Black, M. J. (2007) Multi-state decoding of point-and-click control signals from motor cortical activity in a human with tetraplegia. Proceedings of the 3rd International IEEE EMBS Conference on Neural Engineering, pp. 486–489.
14. Truccolo, W. and Donoghue, J. P. (2007) Nonparametric modeling of Neural Point Processes via Stochastic Gradient Boosting Regression. Neural Computation, 19, 672–705.
15. Wu, W., Gao, Y., Bienenstock, E., Donoghue, J. P. and Black, M. J. (2005) Bayesian population coding of motor cortical activity using a Kalman Filter. Neural Computation, vol.18, pp. 80–1118.
16. Srinivasan, L., Eden, U.T., Mitter, S.K., Brown, E. N. (2007) General-purpose filter design for neural prosthetic devices. J Neurophys 98, 2456–2475.
17. Suner, S., Fellows, M. R., Vargas-Irwin, C., Nakata, K. and Donoghue, J. P. (2005) Reliability of signals from chronically implanted, silicon-based electrode array in non-human primate primary motor cortex. IEEE Trans. Neural Syst. Rehabil. Eng. 13, 524–541.
18. Scholkopf, B., Smola, A.J. (2002) Learning with Kernels - Support vector machines, optimization and beyond. MIT Press, MS.
19. Vapnik, V. (1998). Statistical Learning Theory. Wiley, New York.
20. Hastie, T., Tibshirani, T., and Friedman, J.H. (2001). Elements of Statistical Learning. Springer-Verlag, New York.
21. D. A. Landgrebe, Signal Theory Methods in Multispectral Remote Sensing, John Wiley, Hoboken, 2003.
22. C. M. Gray, *J. Computat. Neurosci.* **1**, 11–38 (1994).
23. B. Pesaran et. al, *Nature Neurosci*, **5**, (2002), 805–811.
24. R. A. Andersen, J. W. Burdick, S. Musallam, B. Pesaran and J. Cham, *Trends in Cognitive Sciences*, **8**, (2004), 486–493.
25. R. A. Johnson and D. W. Wichern, Applied Multivariate Statistical Analysis, Pearson, New Delhi, 2002.

Govindan Rangarajan completed his Ph.D. from University of Maryland, College Park, USA. Later he was employed as a Staff Scientist at the Lawrence Berkeley Laboratory, University of California, Berkeley. He returned to India in 1992 as an Assistant Professor in the Indian Institute of Science, Bangalore. He is currently Professor and Chairman, Department of Mathematics. He is also Convener, Digital Information Services Centre and Convener, IISc Mathematics Initiative. His research interests include nonlinear dynamics and chaos, applications of time series analysis to neuroscience, and applications of stochastic processes. He has over 60 papers to his credit. He was recently made a Chevalier (Knight) of the Order of Palms by the Government of France. He has been awarded the Homi Bhabha Fellowship and is a Fellow of the National Academy of Sciences.

Wilson Truccolo received the Ph. D. degree in Complex Systems and Brain Sciences from the Florida Atlantic University, Boca Raton, FL. He is currently a research Assistant Professor in the Neuroscience Department, Brown University, and a Research Fellow in the Neurology Department, Massachusetts General Hospital. Mr. Truccolo is a principal investigator having recently received a NIH-NINDS K01 career award. His research focuses on stochastic modeling of cortical dynamics & computation, sensorimotor neuroprostheses, neural modulation and control.

Hariharan Nalatore is currently a Project Scientist at Applied Research International, New Delhi. He obtained his Ph.D. from the Indian Institute of Science in 2005 and worked as a Postdoctoral Associate in the Department of Biomedical Engineering, University of Florida, USA. His primary research interest is in time series analysis applied to data generated from neuroscience experiments.