Characterizing the EEG Correlates of Exploratory Behavior

Nicolas Bourdaud [a,b], Ricardo Chavarriaga [a], Ferran Galán [a,c] and José del R. Millán [a,b]

IDIAP–RR 08-28

April 2008

to appear in
IEEE Transactions on Neural Systems & Rehabilitation Engineering

[a] IDIAP Research Institute, Martigny
[b] EPFL, Lausanne
[c] University of Barcelona

IDIAP Research Institute **www.idiap.ch**
Av. des Prés–Beudin 20 P.O. Box 592 1920 Martigny – Switzerland
Tel: +41 27 721 77 11 Fax: +41 27 721 77 12 Email: info@idiap.ch

IDIAP Research Report 08-28

Characterizing the EEG Correlates of Exploratory Behavior

Nicolas Bourdaud, Ricardo Chavarriaga,
Ferran Galán and José del R. Millán

April 2008

to appear in
IEEE Transactions on Neural Systems & Rehabilitation Engineering

Abstract. This study aims to characterize the EEG correlates of exploratory behavior. Decision making in an uncertain environment raises a conflict between two opposing needs: gathering information about the environment and exploiting this knowledge in order to optimize the decision. Exploratory behavior has already been studied using fMRI. Based on a usual paradigm in reinforcement learning, this study has shown bilateral activation in the frontal and parietal cortex. To our knowledge, no previous study has been done on it using EEG.

The study of the exploratory behavior using EEG signals raises two difficulties. First, the labels of trial as exploitation or exploration cannot be directly derived from the subject action. In order to access this information, a model of how the subject makes his decision must be built. The exploration related information can be then derived from it. Second, because of the complexity of the task, its EEG correlates are not necessarily time locked with the action. So the EEG processing methods used should be designed in order to handle signals that shift in time across trials.

Using the same experimental protocol as the fMRI study, results show that the bilateral frontal and parietal areas are also the most discriminant. This strongly suggests that the EEG signal also conveys information about the exploratory behavior.

Acknowledgements: This work is supported by the EC-contract number BACS FP6-IST-027140 and by the Swiss National Science Foundation NCCR 'IM2'

1 Introduction

Decision making in an uncertain environment raises a conflict between two opposing needs : gathering information about the environment and exploiting this knowledge in order to optimize the decision. These two needs are opposed because gathering information usually does not lead to the optimal decision, and exploiting the current knowledge precludes from testing other possible options that may lead to optimal long-term performance.

This conflict is an important point of reinforcement learning theory and is classically illustrated by the n-armed bandit problem [1]. In this problem, the subject is faced repeatedly with a choice between different options. After each choice he receives a numerical reward chosen from a probability distribution that depends on the selected action. At each moment the subject may either select the machine he expects to provide the highest payoff (i.e. to exploit) or another machine in order to improve his estimations (i.e. to explore).

Recent brain imaging studies using functional magnetic resonance imaging (fMRI) or positron emission tomography (PET) have focused on the identification of brain signatures of decision making. Usually these studies link neural activity to external variables observed and manipulated [2,3]. However, decision making is often based on internal decision variables not directly observable from the subject behavior. Experiments that aim to study the correlates of these internal variables must build a model of decision based on the observable variables [4–6]. Studying the differences of activation in the brain during exploratory decisions compared to exploitative decisions requires such a model.

Intracranial recordings in primates and fMRI studies in human suggest that the anterior cingulate cortex (ACC) could control the balance between exploitative and exploratory behavior [7]. Recently, Yoshida and Ishii [4] have reported, using fMRI techniques, lateral activation in the prefrontal cortex (PFC) and ACC activation when exploring a virtual maze. Using the same imagery techniques, Daw et al. [8] have shown that activations in the PFC and intraparietal sulcus are correlated with the differences between exploratory decisions and exploitative decisions.

To our knowledge, no electroencephalography (EEG) studies have focused on this issue. Given its time resolution, EEG could give an information that fMRI does not provide, especially in terms of frequency components. However, since both techniques are based on different physical phenomena, the detection using one technique does not necessarily lead to the detection by the other one. Thus, this study aims to determine whether a difference between exploration and exploitation can be detected in scalp EEG.

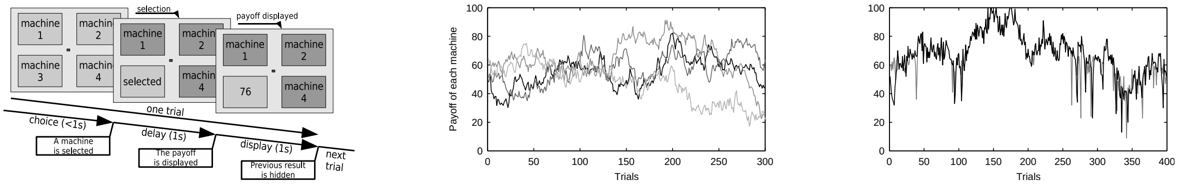
Studying the EEG correlates of exploratory behavior raises the problem of knowing when the decision is made. Because of the complexity of the task, the decision is unlikely to be made at the same instant over all the trials. This makes the search of difference of activity particularly difficult since we cannot synchronize the EEG signals well across trials. The discriminant analysis employed in this study has been developed specifically in order to address this issue.

2 Methods

2.1 Experimental protocol

We adapt the experimental protocol described in Daw et al. [8]. Nine volunteer healthy human subjects (three females and six males; mean age 26) participated in the experiment. Each subject sits in front of a computer screen where four squares are displayed representing four slot machines (see Figure 1(a)). At each trial the subject chooses one machine by pressing a key using the index or middle finger on both hands (left hand for machines 1 and 3, and right hand for machines 2 and 4). One second after the key-press the payoff of the selected machine is displayed for another second and then, a new trial starts. The subject is asked to select the machines in order to maximize the total gain (i.e. sum of individual payoffs) over a session of 400 trials.

The payoff of each machine –a numerical value between 0 and 100– is drawn from a Gaussian distribution whose mean changes slowly from trial to trial. The use of different, non-stationary



(a) (b) (c)

Figure 1: (a) Experimental protocol: each trial is composed of three phases. *(i) Choice*, the 4 machines are presented. The subject has 1s to select a machine by pressing a key. *(ii) Delay*, once a machine has been chosen, the other machines are deactivated (grayed) and no result is displayed for 1s. *(iii) Display*, the payoff for the selected machine is displayed for 1s. (b) Example of the payoff evolution for the 4 machines represented by different level of gray during one repetition of the experiment. (c) Typical user received payoff during the experiment (black line) and payoff obtained by the user's model (gray line).

distributions for each machine requires the user to regularly update his knowledge about the problem; i.e., he is obliged to explore. Before the experiment, nine examples of the payoff evolution for all machines are graphically shown to the user (one of these examples is shown in Figure 1(b)). The evolution of the payoffs is described in appendix A.

Three sessions were recorded for all subjects, with the exception of one who only did two sessions. During the experiment, subjects fixate a red dot in the center of the screen to reduce ocular artifacts. Moreover, they are specifically instructed not to move the arms during the experiment to reduce EMG artifacts.

2.2 EEG acquisition and processing

EEG activity was recorded using a Biosemi Active II portable system. The signal was acquired with 64 electrodes according to the 10/20 international system at a sampling rate of 2048 Hz; then filtered by an eighth-order low-pass Chebyshev Type I filter with a cutoff frequency of 205 Hz and downsampled to 512 Hz. The data were filtered in both the forward and reverse directions to remove all phase distortion, effectively doubling the filter order. In addition, electrooculogram (EOG) was recorded using two electrodes located below and at the outer canthus of the right eye.

Peripheral scalp electrodes were not taken into account for the study[1]. For the remaining electrodes we extracted windows from 1.0s to 0.1s before the key press. This time window was defined to avoid EMG artifacts associated with the finger movements. We subtracted the continuous component using a fourth-order Chebychev high-pass filter with cutoff frequency of 2 Hz and the common average reference was removed. This referencing removes noise signals that are equally spread over the scalp. The continuous wavelet transform was computed using a Morlet mother wavelet on logarithmically scaled frequencies: 7.5 Hz, 10.0 Hz, 13.2 Hz, 17.6 Hz, 23.3 Hz, 31.0 Hz, 41.1 Hz, 54.6 Hz, 72.6 Hz, 96.4 Hz and 128.0 Hz. This scale regularly covers the full spectrum from 7.5 to 128.0 Hz avoiding redundancy among the different frequency channels. One subject was removed from the study because there were artifacts in his recording.

2.3 Behavioral model

A behavioral model is required to label each trial as corresponding to either an exploratory or exploitative decision. In order to compare our results with those reported using fMRI, we adopt the

[1] Electrodes F1, F3, F5, F7, FC1, FC3, FC5, C1, C3, C5, CP1, CP3, CP5, P1, P3, P5, P7, PO3, PO7, O1, Fz, FCz, Cz, CPz, Pz, POz, Oz, F2, F4, F6, F8, FC2, FC4, FC6, C2, C4, C6, CP2, CP4, CP6, P2, P4, P6, P8, PO4, PO8, O2, AF3, AFz and AF4 are used in the analysis, for a total of 50 electrodes

Table 1: Number of trials

num trial	s1	s2	s3	s4	s5	s6	s7	s8
exploitation	938	852	730	511	686	929	887	828
exploration	57	103	201	96	192	78	106	113
right/exploitation (%)	63.65	65.26	67.26	48.53	48.69	45.43	65.95	64.49
right/exploration (%)	52.63	57.28	51.24	55.21	54.17	52.56	54.72	49.56

same behavioral model proposed in [8]. The model, described in appendix B, assumes that the user estimates the mean payoff of each machine using a Bayesian linear Gaussian rule (i.e. a Kalman filter) and, based on these estimations, selects a machine according to a softmax rule. We assume all the subjects share the same model for tracking the payoff means; thus, we compute these parameters using all the available data. In contrast, independent models of machine selection were built per subject to take into account inter-subject variability. Parameters of the model (for both mean tracking and machine selection) are estimated by maximizing the model likelihood with respect to the subject's choices. A comparison of the choices taken by the user and those given by the model can be seen in Figure 1(c).

At any given trial, the behavioral model provides the mean payoff for all machines taking into account previous observations (i.e. the payoff obtained at previous trials). Comparison between the model's estimated payoff for all machines is used to label that trial as either exploration or exploitation. Those trials where the user selects the machine with the highest estimated mean, are labeled as corresponding to exploitative decisions. In order to increase the reliability of the labeling process, a threshold (with value 4) was set when computing the payoff difference between the machines. Moreover, only exploratory trials (i.e. those trials where the selected machine does not have the highest estimated payoff) corresponding to a change of machine are kept for further analysis. An average of 22% of the trials are discarded at this stage. The total number of trials used in the analysis is shown in Table 1. In order to discard possible bias of the labeling we also show the percentage of samples corresponding to movements of the right hand (machines 2 and 4) for both exploratory and exploitative decisions.

2.4 EEG correlates of the exploratory behavior

Given the characteristics of the phenomena we are dealing with, it is unlikely that the correlates of exploratory behavior are well synchronized over all trials, i.e. its neural correlates will not appear always at the same time. Because of this, traditional EEG analysis techniques based on trial averaging in either the time (e.g. event-related potentials) or frequency domain (e.g event-related spectral perturbation) do not seem appropriate for the study of such phenomena [9]. In order to deal with this issue, we propose a method which studies the patterns of the EEG signal over the scalp by considering the distributions of each class (exploration and exploitation) as a "bag of time-samples", meaning that any information about the time location of the sample within the trial is discarded. The rationale of this approach is to detect the most discriminant samples irrespective of appearance time within each trial. In the following subsections, we describe the feature extraction mechanism (2.5), and then the feature selection method (2.6).

2.5 Feature extraction: Canonical Variates Analysis

We compute the power signal for all electrodes per frequency band as described in section 2.2. Features are then extracted using Canonical Variates Analysis (CVA) [10, 11] (also referred to as multidimensional discriminant analysis, MDA). The CVA computes the subspace on which the linear projection of the data maximizes the discriminability of the classes. Such analysis provides the projection of each time sample onto a uni-dimensional space, as well as a measure of the discriminant power (DP) of each electrode. The DP ranks the electrodes according to the correlation between the original signal and the features in the projected space. For our purpose, the discriminant power provides information about which electrodes on the scalp EEG convey more information to distinguish between exploratory

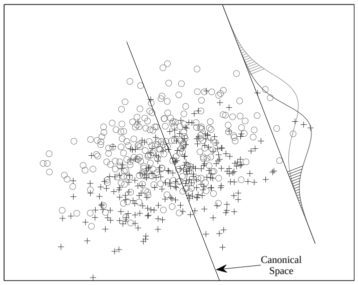


Figure 2: Schematic representation of the processing method. Gray (o) and black (+) symbols correspond to samples of two different classes. The distributions of the projected samples on the canonical space of both class are also presented. The hatched gray and black areas on the distribution function show the frame sets for the gray (o) and black (+) classes.

and exploitative trials.

2.6 Detection of discriminative samples

We propose a method that relies on the detection of the most discriminative samples for each class based on the sample distribution on the canonical projection. Under this approach, we attempt to recognize informative phenomena by identifying samples that lie on non-overlapping regions of the canonical space. We then perform classification based solely on those samples. Following Freeman's theory, we will use the term *frame* to denote these informative samples [12, 13].

Using this approach, a canonical transformation is computed for each subject using a subset of the data (i.e. train set). A sample will be considered as a frame if it lies on the non-overlapping tail of the samples feature distribution in the canonical space (c.f. Figure 2). In this study we use the opposing 5-percentiles of the class distributions, as thresholds to define the frame sets of the corresponding classes.

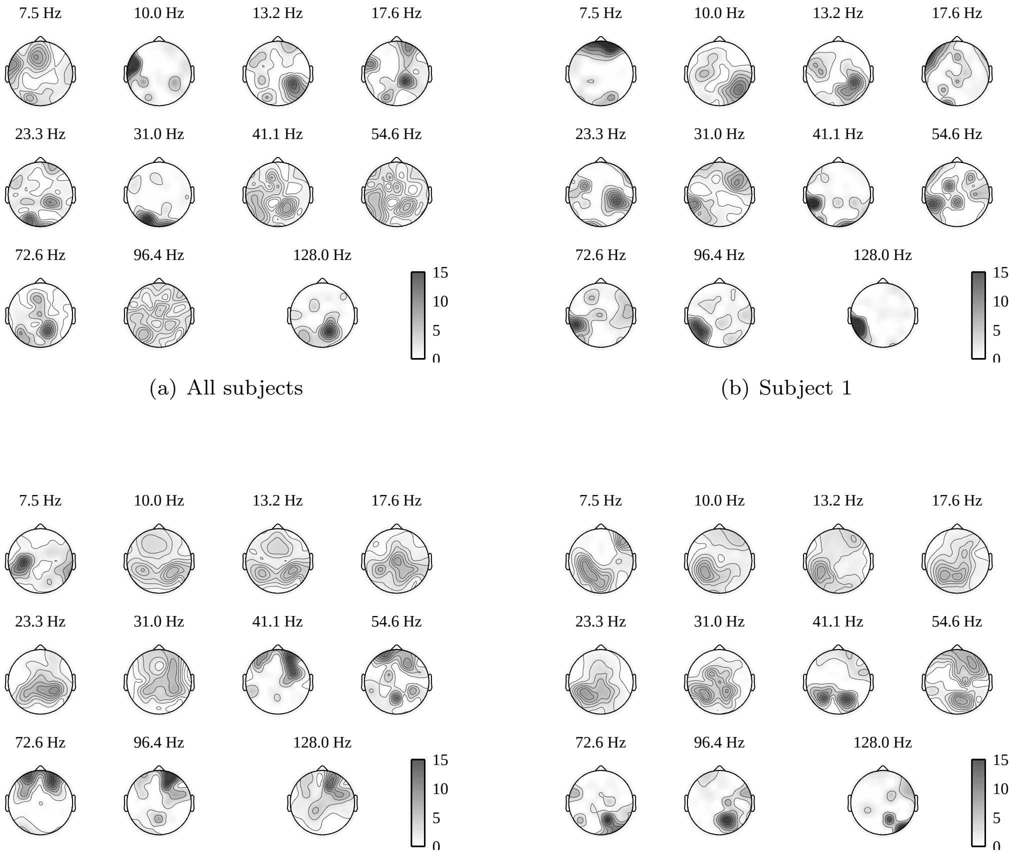
In order to evaluate whether a new test sample corresponds to a frame, its canonical projection is compared to these thresholds. The classification of a trial uses a voting scheme based on the number of identified frames of each class. In case of equality or if no frame has been detected, the trial is marked as unknown.

To take into account the different sizes of the data sets for both classes (c.f. Table 1) we use the normalized Mathews correlation coefficient (MCC) [14] that takes into account the rate of correctly classified samples for each class :

$$MCC = \frac{t_1 t_2 - f_1 f_2}{\sqrt{(t_1 + f_1 t)(t_1 + f_2)(t_2 + f_1)(t_2 + f_2)}} \tag{1}$$

where t_1 and f_1 denote correctly and incorrectly classified *exploratory* trials respectively and the same for t_2 and f_2 for the *exploitative* trials. A MCC coefficient of 1 corresponds to perfect classification for both exploration and exploitation samples, while a value of zero corresponds to random performance. If all samples are misclassified, the resulting coefficient is equal to -1.

Separate classifiers are built for each frequency band. In addition, a combined classifier is built using the data from several selected frequency bands. A frequency band is selected for this combined classifier if the classifier based on the single band has a MCC higher than 0.16 on training set (equivalent to a classification accuracy of 58% with equally sized sets). The choice of this threshold is based on a trade-off between selecting more than one band and a reasonable confidence that the classifier based on this single band performs better than random. To classify a trial, this combined classifier attributes a label to a trial according to the total number of frames detected in *all selected* bands of each class.



(a) All subjects

(b) Subject 1

(c) Subject 2

(d) Subject 9

Figure 3: Discriminant power (DP) of the electrodes activity using: (a) data of all subjects, (b)–(d) data of three different subjects.

3 Results

3.1 Discriminant analysis

We computed the canonical space projection and the electrodes discriminant power (DP) using data for all subjects altogether (representing 3.8×10^6 time samples) as described in subsection 2.5. Figure 3(a) shows the electrode discriminant power for the different frequency bands. The figure shows that the most informative scalp areas (high DP) correspond to left frontal and bilateral parietal: 11 of the 15 most discriminant electrodes are located in these areas.

The same analysis was done independently per subject (using 4.1×10^5 time samples per subject on average, standard deviation: 5.4×10^4). See Figures 3(b), 3(c) and 3(d) for examples. Comparing among subjects, we report a high inter-subject variability in the precise location of the source of discrimination. But consistently with the global analysis, left frontal and bilateral parietal areas are often discriminant: left frontal electrodes were found to be in the 15 most discriminant electrodes for 7 subjects, right parietal area for 7 subjects and left parietal area for 6 subjects. In addition, right frontal electrodes were found among the 15 most discriminant for all subjects. Finally, bilateral frontal and parietal activities seem to be, in spite of the inter-subject variability, the most discriminant activities of the exploratory behavior.

From the analysis made on all EEG data of all subjects, we can observe that frontal and parietal electrodes are not discriminant in the same frequency bands. Discriminant frontal activity is mainly

Table 2: Classification accuracies per subjects

(a) training set

freq.	s1	s2	s3	s4	s5	s6	s7	s8
7.5	0.31±0.07(34)	0.26±0.02(28)	0.24±0.02(21)	0.29±0.04(31)	0.31±0.02(32)	0.18±0.02(27)	0.21±0.01(19)	0.31±0.02(47)
10.0	0.22±0.01(26)	0.41±0.03(59)	0.26±0.02(25)	0.24±0.03(24)	0.22±0.02(20)	0.15±0.02(21)	0.27±0.02(23)	0.32±0.02(45)
13.2	0.20±0.02(21)	0.26±0.03(40)	0.21±0.02(10)	0.26±0.02(18)	0.29±0.02(17)	0.19±0.02(21)	0.21±0.03(13)	0.25±0.02(29)
17.6	0.17±0.03(6)	0.25±0.02(18)	0.23±0.01(4)	0.19±0.03(8)	0.22±0.02(8)	0.17±0.01(12)	0.20±0.02(6)	0.36±0.02(28)
23.3	0.14±0.03(5)	0.36±0.01(21)	0.21±0.03(3)	0.24±0.02(4)	0.17±0.03(3)	0.16±0.02(9)	0.11±0.02(3)	0.31±0.02(19)
31.0	0.16±0.02(4)	0.17±0.02(4)	0.17±0.03(2)	0.19±0.02(3)	0.23±0.03(2)	0.14±0.02(6)	0.21±0.02(4)	0.26±0.02(5)
41.1	0.13±0.02(1)	0.12±0.01(2)	0.13±0.02(7)	0.18±0.02(4)	0.03±0.07(3)	0.14±0.02(3)	0.18±0.03(4)	0.08±0.02(1)
54.6	0.13±0.03(4)	0.16±0.02(4)	0.17±0.01(9)	0.18±0.03(18)	0.09±0.07(8)	0.12±0.03(4)	0.18±0.02(11)	0.11±0.03(5)
72.6	0.10±0.02(1)	0.09±0.01(2)	0.11±0.02(3)	0.24±0.02(2)	0.12±0.03(2)	0.11±0.01(3)	0.19±0.02(2)	0.12±0.02(1)
96.4	0.15±0.02(2)	0.15±0.01(2)	0.12±0.03(10)	0.22±0.02(2)	0.07±0.07(5)	0.11±0.02(4)	0.20±0.02(2)	0.08±0.03(2)
128	0.07±0.02(1)	0.09±0.02(1)	0.01±0.10(4)	0.14±0.02(1)	0.02±0.08(1)	0.10±0.01(1)	0.19±0.02(1)	0.11±0.01(2)
CC	0.30±0.04(1)	0.36±0.03(0)	0.36±0.02(0)	0.39±0.02(0)	0.43±0.02(0)	0.24±0.01(1)	0.31±0.02(0)	0.38±0.01(0)

(b) test set

freq.	s1	s2	s3	s4	s5	s6	s7	s8
7.5	0.08±0.14(36)	0.15±0.10(26)	0.13±0.08(19)	0.08±0.12(30)	0.18±0.10(30)	-0.06±0.11(28)	0.05±0.14(22)	0.20±0.15(45)
10.0	0.07±0.14(28)	0.30±0.22(59)	0.16±0.08(24)	0.12±0.09(23)	0.10±0.10(18)	-0.00±0.10(19)	0.15±0.13(21)	0.27±0.09(45)
13.2	0.01±0.14(19)	0.18±0.13(38)	0.12±0.11(10)	0.15±0.16(17)	0.19±0.12(18)	0.03±0.12(21)	0.04±0.09(14)	0.16±0.11(26)
17.6	0.01±0.13(6)	0.19±0.08(18)	0.13±0.11(5)	0.07±0.10(8)	0.14±0.12(9)	0.07±0.08(12)	0.00±0.17(8)	0.31±0.11(28)
23.3	-0.02±0.09(4)	0.30±0.10(21)	0.11±0.15(4)	0.05±0.13(3)	0.02±0.15(3)	0.01±0.10(9)	-0.09±0.12(4)	0.26±0.15(19)
31.0	0.06±0.12(3)	0.07±0.13(5)	-0.01±0.13(2)	0.02±0.13(3)	0.08±0.10(3)	0.05±0.09(5)	0.06±0.15(7)	0.16±0.14(5)
41.1	0.06±0.11(1)	0.04±0.08(1)	0.02±0.06(5)	-0.04±0.11(4)	-0.06±0.17(1)	0.06±0.10(4)	-0.02±0.15(4)	-0.02±0.13(1)
54.6	0.08±0.13(3)	0.08±0.11(4)	-0.01±0.06(8)	-0.01±0.10(20)	-0.08±0.19(6)	0.06±0.08(4)	-0.00±0.17(5)	0.02±0.09(3)
72.6	0.02±0.10(1)	0.05±0.13(2)	0.04±0.08(3)	0.12±0.14(3)	0.04±0.08(1)	0.03±0.14(2)	0.02±0.12(2)	-0.01±0.13(1)
96.4	0.09±0.15(2)	0.10±0.12(1)	-0.04±0.05(12)	0.05±0.17(2)	0.01±0.08(2)	0.08±0.10(4)	0.05±0.16(2)	-0.00±0.07(1)
128	0.03±0.13(1)	0.06±0.10(2)	0.01±0.03(3)	0.04±0.12(1)	-0.02±0.04(0)	0.05±0.08(1)	0.01±0.16(2)	0.00±0.09(2)
CC	0.05±0.14(1)	0.29±0.09(0)	0.16±0.13(0)	0.11±0.19(0)	0.21±0.14(0)	0.03±0.08(1)	0.04±0.15(0)	0.29±0.13(0)

The performances are reported as "mean MCC±SD(UNK)" of the 10-fold cross validation where UNK is the percentage of trials labelled as unknown by the classifier. CC refers to the combined classifiers. See text for details.

found between 7.5 and 41 Hz while parietal activity is found in the full spectrum of analysis.

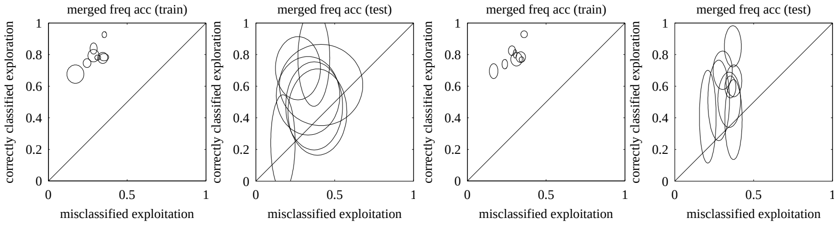
3.2 Classification

A 10-fold cross-validation procedure was used to assess the performance of the classification: trials are sorted chronologically and divided into 10 consecutive and approximately equal subsets of trials. Each fold in the cross validation procedure is formed by one of these subsets which is used as the testing set and by the remaining nine subsets used as the training set. This procedure differs from the usual cross-validation procedure in the sense that the trials in test set do not come from time epochs that have been used to train the classifiers. The classification performances reported in the study are the average and standard deviation of the MCC value of all folds.

The classification performances of train and test sets are reported in Table 2. For all subjects, the best performances are mainly obtained below 23.3 Hz (low beta and alpha band). Considering the performance on the test set, we have observed that the classifiers based on the low frequency bands also perform the best. For 6 subjects, these classifiers produce results significantly better than random. Despite having a high performance, classifiers centered around 7.5 and 10 Hz have a high rejection rate (more than 20% of the trials are labelled as unknown).

For all subjects, performance of the combined classifier is comparable to the one of the best classifiers based on a single frequency band. However, its rejection rate is very low on the two sets and for four subjects, its performances are significantly better than random while it rejects only 0.45% of the trials on average. The frequency bands selected to be combined are mostly below 31 Hz. Since the rejection rate of the combined classifier is significantly lower than that of any other classifier, this shows that it does not rely on the data from only one band but instead, its performance is effectively based on the combination of several bands.

The same analysis was done using EOG data instead of EEG data to rule out signal contamination due to eye or facial movements. No classifier based on any single frequency band nor any combined classifier performed better than random for every subject even on the training set. This shows that no EOG artifacts had contaminated the EEG analysis.



(a) cross-validation used the study (b) traditional cross-validation

Figure 4: Performance of the combined classifier using: (a) the cross validation procedure used in the study (preserving data time order), (b) the traditional cross validation. Each plot shows the exploration classification accuracy versus the exploitation error rate. A perfect classifier would get a point in the upper-left corner of the plot and random classification results would be lie on the diagonal line. In addition, the standard deviations of the estimation of the accuracy and the error rate in cross-validation procedure are reported, thus each classification performance is represented by an ellipse.

To compare the cross-validation procedure used in this study with the traditional cross validation (that is, by mixing in the training and test set trials from different time instants), we report in Figure 4 the performance –estimated by the 2 methods– of the combined classifier for each subject. The traditional cross-validation presents slightly better classification accuracies with a smaller variance in the set of exploitation trials.

4 Discussion

In this study, we were able to find scalp EEG activity discriminant between exploratory and exploitative behavior. This activity was mainly located in bilateral frontal and parietal areas, which is consistent with the intracranial activity reported by previous studies using fMRI. Daw et al. [8], using the same protocol, have reported activity in prefrontal cortex (PFC) and in parietal areas as discriminative of exploratory behavior. Similarly, Yoshida and Ishii [4] found lateral PFC and parietal activity, although they suggest the latter activity to be related to the maintenance of the spatial information of the task. As it was done in previous studies [15], we plan to apply inverse methods [16, 17] to estimate the intracranial sources of the EEG activity and perform a further comparison of our results with those reported in fMRI studies.

Our results show that the discriminant frontal activity is mainly found in the alpha and beta band, whereas the parietal activity is not constraint to a particular sub-band. Moreover, the performance of classifiers based on single frequencies confirms the importance of the alpha and beta band for recognition of exploratory behavior from EEG. However, the increased performance of the combined classifiers suggests that discriminant EEG correlates are unlikely to be restricted to a single frequency band.

We achieved test classification performance above random levels in four of the subjects. In particular, classifiers using single frequency bands equal or below 23.3 Hz have the highest performance, although they have a high rejection rate. The use of combined classifiers increase the classification performance while dramatically reducing the number of trials labelled as unknown. It should also be noticed that we measure the classification performance using a cross-validation method that preserves the temporal order of the data –as opposed to traditional data partitioning for cross-validation– thus giving a better approximation of the method's ability to generalize non-stationary data (c.f. Figure 4). In addition, classifier generalization capabilities may also be affected by the difference in the number of trials corresponding to each class (c.f. Table 1).

It has been argued that EEG signals above 20 Hz are highly affected by EMG artifacts [18, 19]. However, distant EMG activity (i.e. generated by limb movements) is unlikely to have biased the results since the time window used in the study is located before the actual movement. Moreover, the number of trials corresponding to movements done by right and left hand are roughly equal in both classes (see Table 1). EMG artifacts from eye, face or neck movements could have affected the EEG signal, but the classification results using EOG data and visual inspection of the signal energy distribution across different frequencies contradict this hypothesis.

As explained previously, a behavioral model of the subject's decisions was used to label trials as exploration or exploitation. The fitting of the parameters of the model has been shown to be consistent with the actual statistical parameters of the machines. More refined models have been proposed to explicitly include the subject's need for exploring [20]. We plan to study these models in order to improve the discriminant analysis of exploratory decisions.

In this study, we assume that discriminant features are not synchronized to any observable event (e.g. key press to select a machine). The use of a detection approach allowed us to overcome the issue of the non-time-locked signal, as suggested by the classification results. However, the applied technique is not able to capture temporal relationships between several discriminative patterns of activity (potentially in different frequency bands). These might be even more relevant since our results show that taking into account combined information from different bands helps to better discriminate between the two types of decision.

To address this issue, several techniques can be applied. For instance, trials can be synchronized according to the most discriminative time sample in the canonical space (i.e. frame) and then average the EEG pattern in the time-frequency domain. Alternatively, Vialatte et al. [21] have proposed a method that consists of detecting some elementary patterns in the time-frequency domain for each trial and then determining a relevant structure between these patterns. Furthermore, analysing the phase synchrony between frontal and parietal areas, both involved in the discrimination of exploratory behavior, can give some insights about information transfer between these two areas [22].

To sum up, this study has shown that scalp EEG conveys discriminative information between exploratory and exploitative decisions. The spatial pattern of these signals (i.e. most discriminative electrodes) was found to be consistent with previous fMRI studies. Moreover, using a feature detection approach we achieve classification performance above random levels in four out of eight subjects. Further studies will be performed to better characterize EEG correlates with the two-fold goal of performing a further comparison with other imaging studies, as well as to increase the classification performance.

A Distribution of the payoff

The payoff $r_{i,k}$ associated with the ith machine at trial k is drawn from a Gaussian distribution (mean $\mu_{i,k}$, SD: σ_0) and rounded to the nearest integer between [0 , 100]. At each time step, the mean $\mu_{i,k}$ is diffused in a decaying Gaussian random walk,

$$\begin{aligned} \mu_{i,k+1} &= \lambda\mu_{i,k} + (1-\lambda)\theta + e \qquad (2)\\ &\qquad \text{where } e \sim \mathcal{N}\left(0, \sigma_d^2\right) \\ r'_{i,k} &\sim \mathcal{N}\left(\mu_{i,k}, \sigma_0^2\right) \qquad (3)\\ r_{i,k} &= round\left(r'_{i,k}\right) \qquad (4) \end{aligned}$$

where λ, θ controls the random walk of the mean $\mu_{i,k}$ and e corresponds to Gaussian noise with zero mean and SD σ_d. The values of these parameters are reported in Table 3. The mean payoff $\mu_{i,0}$ at the beginning of the experiment was set to the result of computing 30 diffusion steps (equation (2)) with an initial value of 50.

Table 3: Estimation of the parameters of the behavioral model

	λ	θ	σ_d	σ_0
Real values	0.9836	50	2.8	4
Est. values	0.92	51.37	8.12	N/A

subj.	1	2	3	4	5	6	7	8
β	0.37	0.28	0.19	0.21	0.19	0.29	0.29	0.23

B Modelling user estimation and selection

We model the subject strategy for tracking the payoff of each machine by a Kalman filter and assume that parameters do not change over trials. *After* a machine j has been selected at trial k, the estimated payoff distribution $(\hat{\mu}_{j,k}^{post},\ \hat{\sigma}_{j,k}^{post\ 2})$ can be updated given the received payoff r_k and the estimation *before* the observation $(\hat{\mu}_{j,k}^{pre},\ \hat{\sigma}_{j,k}^{pre\ 2})$,

$$\hat{\mu}_{j,k}^{post} = \hat{\mu}_{j,k}^{pre} + \kappa_k \left(r_k - \hat{\mu}_{j,k}^{pre} \right) \tag{5}$$

$$\hat{\sigma}_{j,k}^{post\ 2} = (1 - \kappa_k)\ \hat{\sigma}_{j,k}^{pre\ 2} \tag{6}$$

$$\text{where} \quad \kappa_k = \frac{\hat{\sigma}_{j,k}^{pre\ 2}}{\hat{\sigma}_{j,k}^{pre\ 2} + \hat{\sigma}_0^2} \tag{7}$$

Since the user cannot observe the payoff of the remaining machines, the mean estimation for these machines does not change as a result of the choice. That is,

$$\forall i \neq j \quad \begin{cases} \hat{\mu}_{i,k}^{post} = \hat{\mu}_{i,k}^{pre} \\ \hat{\sigma}_{i,k}^{post} = \hat{\sigma}_{i,k}^{pre} \end{cases} \tag{8}$$

Then, the estimations are updated in time according to the diffusion rule seen in (2),

$$\hat{\mu}_{i,k+1}^{pre} = \hat{\lambda}\ \hat{\mu}_{i,k}^{post} + \left(1 - \hat{\lambda}\right) \hat{\theta} \tag{9}$$

$$\hat{\sigma}_{i,k+1}^{pre\ 2} = \hat{\lambda}^2\ \hat{\sigma}_{i,k}^{post\ 2} + \hat{\sigma}_d^2 \tag{10}$$

We model the choice of the subjects by a softmax rule, i.e. at each trial k the probability of choosing the machine i is :

$$P_{i,k} = \frac{\exp\left(\beta \hat{\mu}_{i,k}^{pre}\right)}{\sum_j \exp\left(\beta \hat{\mu}_{j,k}^{pre}\right)} \tag{11}$$

The parameters of the behavioral model are estimated by maximizing the log-likelihood under constraints ($\hat{\lambda} \in [0,1]$, $\hat{\theta} \in [0,100]$, $\hat{\sigma}_d \in [0,100]$, $\beta \geq 0$). The parameters of the mean payoff tracking ($\hat{\sigma}_0$, $\hat{\lambda}$, $\hat{\theta}$ and $\hat{\sigma}_d$) are shared by all subjects, while the parameter of the selection (β) is specific to each subject. To speed up convergence, parameters $\hat{\sigma}_0$, $\hat{\mu}_{i,0}^{pre}$ and $\hat{\sigma}_{i,0}^{pre}$ are fixed to the real values σ_0, $\mu_{i,0}$ and $\sigma_{i,0}$. Fixing these last two parameters does not significantly affect the estimation of the others because their influence vanishes quickly within a few trials. Table 3 shows the estimated values of the model, which are consistent with the real values of the machines.

References

[1] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction.* Cambridge, MA: MIT Press, 1998.

[2] J. Tanabe, L. Thompson, E. Claus, M. Dalwani, K. Hutchison, and M. T. Banich, “Prefrontal cortex activity is reduced in gambling and nongambling substance users during decision-making.” *Hum Brain Mapp*, vol. 28, pp. 1276–1286, 2007.

[3] M. G. Philiastides and P. Sajda, "EEG-informed fMRI reveals spatiotemporal characteristics of perceptual decision making." *J Neurosci*, vol. 27, pp. 13 082–13 091, 2007.

[4] W. Yoshida and S. Ishii, "Resolution of uncertainty in prefrontal cortex." *Neuron*, vol. 50, pp. 781–9, 2006.

[5] ——, "Model-based reinforcement learning: A computational model and an fMRI study," *Neurocomputing*, vol. 63, pp. 253–269, 2005.

[6] G. Corrado and K. Doya, "Understanding neural coding through the model-based analysis of decision making." *J Neurosci*, vol. 27, pp. 8178–8180, 2007.

[7] M. F. S. Rushworth, "Intention, choice, and the medial frontal cortex," *Ann. N. Y. Acad. Sci.*, vol. 1124, pp. 181–207, 2008.

[8] N. D. Daw, J. P. O'Doherty, P. Dayan, B. Seymour, and R. J. Dolan, "Cortical substrates for exploratory decisions in humans." *Nature*, vol. 441, pp. 876–9, 2006.

[9] C. Tallon-Baudry and O. Bertrand, "Oscillatory gamma activity in humans and its role in object representation." *Trends in Cognitive Sciences*, vol. 3, no. 4, pp. 151–162, Apr. 1999.

[10] J. Krzanowski, *Principles of Multivariate Analysis.* Oxford: Oxford University Press, 1998.

[11] F. Galán, P. W. Ferrez, F. Oliva, J. Guardia, and J. d. R. Millán, "Feature extraction for multi-class BCI using canonical variates analysis," in *Proc. IEEE International Symposium on Intelligent Signal Processing WISP 2007*, 2007.

[12] W. J. Freeman, "Origin, structure, and role of background EEG activity. Part 4: Neural frame simulation." *Clin Neurophysiol*, vol. 117, pp. 572–89, 2006.

[13] F. Galán, J. Palix, R. Chavarriaga, P. Ferrez, E. Lew, C. Hauert, and J. Millán, "Visuo-spatial attention frame recognition for brain-computer interfaces," in *Proceedings of the 1st International Conference on Cognitive Neurodynamics*, 2007.

[14] P. Baldi, S. Brunak, Y. Chauvin, C. A. Andersen, and H. Nielsen, "Assessing the accuracy of prediction algorithms for classification: an overview." *Bioinformatics*, vol. 16, pp. 412–424, 2000.

[15] P. W. Ferrez and J. del R. Millán, "Error-related EEG potentials generated during simulated brain-computer interaction," *IEEE Trans. Biomed. Eng.*, vol. 55, pp. 923–929, 2008.

[16] F. Babiloni, F. Cincotti, C. Babiloni, F. Carducci, D. Mattia, L. Astolfi, A. Basilisco, P. M. Rossini, L. Ding, Y. Ni, J. Cheng, K. Christine, J. Sweeney, and B. He, "Estimation of the cortical functional connectivity with the multimodal integration of high-resolution EEG and fMRI data by directed transfer function." *NeuroImage*, vol. 24, pp. 118–31, 2005.

[17] R. D. Pascual-Marqui, "Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details." *Methods Find Exp Clin Pharmacol*, vol. 24 Suppl D, pp. 5–12, 2002.

[18] E. M. Whitham, K. J. Pope, S. P. Fitzgibbon, T. Lewis, C. R. Clark, S. Loveless, M. Broberg, A. Wallace, D. DeLosAngeles, P. Lillie, A. Hardy, R. Fronsko, A. Pulbrook, and J. O. Willoughby, "Scalp electrical recording during paralysis: Quantitative evidence that EEG frequencies above 20 Hz are contaminated by EMG." *Clin Neurophysiol*, vol. 118, pp. 1877–1888, 2007.

[19] E. M. Whitham, T. Lewis, K. J. Pope, S. P. Fitzgibbon, C. R. Clark, S. Loveless, D. Delosangeles, A. K. Wallace, M. Broberg, and J. O. Willoughby, "Thinking activates EMG in scalp electrical recordings." *Clin Neurophysiol*, 2008.

[20] S. Ishii, W. Yoshida, and J. Yoshimoto, "Control of exploitation-exploration meta-parameter in reinforcement learning." *Neural networks*, vol. 15, pp. 665–87, 2002.

[21] F. B. Vialatte, C. Martin, R. Dubois, J. Haddad, B. Quenet, R. Gervais, and G. Dreyfus, "A machine learning approach to the analysis of time-frequency maps, and its application to neural dynamics." *Neural Netw*, vol. 20, pp. 194–209, 2007.

[22] F. J. Varela, J.-P. Lachaux, E. Rodriguez, and J. Martinerie, "The brainweb: Phase synchronization and large-scale integration." *Nature Reviews Neuroscience*, vol. 2, pp. 229–39, 2001.