Exploring Two Novel Features for EEG-based Brain-Computer Interfaces: Multifractal Cumulants and Predictive Complexity

Nicolas Brodu, Fabien Lotte, Anatole Lécuyer

Exploring Two Novel Features for EEG-based Brain-Computer Interfaces: Multifractal Cumulants and Predictive Complexity

Nicolas Brodu <nicolas.brodu@univ-rennes1.fr>[*],
Fabien Lotte <fabien.lotte@inria.fr>[†],
Anatole Lécuyer <anatole.lecuyer@inria.fr>[‡]

20th May 2011

Abstract

In this paper, we introduce two new features for the design of electroencephalography (EEG) based Brain-Computer Interfaces (BCI): one feature based on multifractal cumulants, and one feature based on the predictive complexity of the EEG time series. The multifractal cumulants feature measures the signal regularity, while the predictive complexity measures the difficulty to predict the future of the signal based on its past, hence a degree of how complex it is. We have conducted an evaluation of the performance of these two novel features on EEG data corresponding to motor-imagery. We also compared them to the gold standard features used in the BCI field, namely the Band-Power features. We evaluated these three kinds of features and their combinations on EEG signals from 13 subjects. Results obtained show that our novel features can lead to BCI designs with improved classification performance, notably when using and combining the three kinds of feature (band-power, multifractal cumulants, predictive complexity) together.

1 Introduction

Brain-Computer Interfaces (BCI) are communication systems that enable users to send commands to a computer by using only their brain activity [53], this activity being generally measured using ElectroEncephaloGraphy (EEG). Most EEG-based BCI are designed around a pattern recognition approach: in a first step features describing the relevant information embedded in the EEG signals are extracted [6]. They are then fed into a classifier which identifies the class of the mental state from these features [32]. Therefore, the efficiency of a BCI, in terms of recognition rate, depends mostly on the choice of appropriate features and classifiers. Moreover, several researchers suggested that the choice of a good feature extraction method has more impact on the final performances than the choice of a good classifier [25, 41].

In this paper, we focus on feature extraction from EEG signals for the design of BCI based on motor imagery (MI) [42]. MI corresponds to the imagination of limb movements (e.g., hand or feet) without actual output. It is known to trigger brain signals variations in specific frequency bands that can be used to drive a BCI [42]. Numerous features have been proposed and explored to design motor imagery-based BCI [35, 6]. Most of these techniques can be described as belonging to one of the two following categories: 1) methods exploiting a temporal information embedded in the signals and 2) methods exploiting a frequential information. Naturally, all these approaches usually exploit a spatial information as well by selecting a relevant subset of EEG channels or by using spatial filtering [42, 33]. In the temporal feature category, researchers have mainly explored methods based on auto-regressive (AR) parameters [3, 22] and their variants such as multivariate AR parameters [4], AR parameters with exogeneous inputs [17] and adaptive auto-regressive parameters (AAR) [44]. They have also explored features such as Hjorth Parameters [39] or time-domain parameters [48]. Since motor imagery triggers amplitude variations in specific oscillations of EEG

[*]Nicolas is with the University of Rennes 1, Campus de Beaulieu, Bât 22, 35042 Rennes Cedex - France.

[†]Fabien was with the Institute for Infocomm Research (I2R), 1 Fusionopolis way, 138632, Singapore. He now is with the national institute for research in computer science and control (INRIA) Bordeaux Sud-Ouest, 351 cours de la libération, 33405, Talence Cedex, France

[‡]Anatole is with the INRIA, at Campus universitaire de Beaulieu, Avenue du Général Leclerc, 35042 Rennes Cedex - France.

signals, the most popular features used in motor imagery-based BCI belong to the category which exploits a frequential information. In this category, the features used, denoted as band-power features, quantify the energy of the EEG signals in various frequency bands [6, 15, 27]. These bands are usually the mu (8-12Hz) and beta (16-24Hz) bands, in which the energy over the motor cortex of the brain is known to be affected by motor imagery [42]. Such band power features can be computed in a number of ways. For instance, they can be computed by band-pass filtering the EEG signals in specific frequency bands and then squaring it [42], by computing the Power Spectral Density of the signals [10, 37], by using various time-frequency representations [27, 18], or by using wavelet decompositions [31, 54, 11]. Naturally time-frequency representations have also been used to exploit both a temporal and a frequential information [21, 50]. In addition to these two main categories of features, some other approaches have been explored in a more marginal way. Among them we can mention methods exploiting a connectivity measure between EEG channels [24, 51] or chaos theory-inspired method such as the fractal dimension [10]. Among all these features, the methods that are based on the band power of the EEG signals are consistently found to be among the most efficient ones [10, 16, 15, 27, 49]. Moreover, band power features are the method of choice used in the design of MI-based BCI from the leading groups worldwide [7, 20, 40]. Therefore, such features are usually considered as the gold standard in MI-based BCI design. Unfortunately, the performances of current BCI are still not satisfactory, and the BCI community has stressed the need to further explore and design alternative features [35].

Consequently, the contribution of this paper is twofold. First, we introduce two new features for MI classification in EEG-based BCI. These two new features are based on: 1) multifractal cumulants [52] and 2) predictive complexity [14]. Second, we perform systematic comparisons and analysis of the performance of these two new features, together with the gold standard feature for MI based-BCI, i.e., band-power feature.

The first new feature, namely, multifractal cumulants, can be seen as a statistic on inter-frequency band relations. This is particularly relevant for BCI as this information is generally ignored in current MI-based BCI designs, mostly based on the sole power in different frequency bands. The multifractal approach adds in information about how the multiple bands relate together at each instant. The hypothesis is that when some specific mental activity occurs, it changes not only the power of different frequency bands of the EEG, but also the repartition of that power between bands. It should be mentioned that a preliminary study of this kind of feature has been presented in a conference paper [13]. The current work expands on this idea and uses a better definition for the multifractal feature (see Section 2).

The second new feature is based on the statistical complexity and predictive properties of the time series [14]. The information (quantified in bits) that is extracted this way measures how difficult it is to make an optimal prediction based on past information. It is null both for totally ordered and totally random systems, and increases in between. It has already been applied to single simulated neurons [12] and a related form was applied to measure synchronisation in the brain [30]. The assumption is that performing a mental task makes the EEG signal either more or less predictable, which can be detected by a classifier when quantified by this second new feature. For example, a motor activity generates rythmic patterns, which are distinctive compared to the neuron resting state. Irrespectively of the implied frequencies of both motor activity and resting state, we assume that the presence of some kind of activity will change the predictability of the EEG signal.

The remainder of this paper is organised as follows: Sections 2 and 3 present the two new features that we propose. Section 4 presents the experimental evaluation, including the data sets that were used, and the results. Section 5 concludes this study.

The code used for all the experiments in this paper is provided as Free/Libre software. The data used is available online (link given in appendix) and all the presented results are reproducible independently.

2 Multifractal cumulants

The multifractal formalism is described in details in [52, 38]. This section presents a short overview for the needs of this document.

Intuitively, the multifractal cumulants of the signal capture a signature of inter-band relations (see below). This contrasts to the power in each frequency band that is generally used. As shown in [13] the multifractal spectrum can in itself be used for EEG classification. When considering multifractal in addition to power band feature vectors, the resulting combination may improve the classification accuracy.

The method we chose for extracting the multifractal spectrum is a discrete wavelet transform of the signal, out of which we extract the wavelet leader coefficients [1]. Following the directions of [52] we then use the cumulants of the leaders as the features for classification, unlike what we previously did in [13].

Let $x(t)$ be the signal to analyse. One view on multifractal analysis [34] is to relate the statistical properties of $x(t)$ and of a scaled version of it $x(at)$. In terms of frequency analysis, that scaling in time corresponds to a frequency shift. Hence, another view of the multifractal cumulants feature is that they characterize some form of inter-frequency information, as mentioned in the introduction of this section.

More precisely:

- The signal $x(t)$ is decomposed using a Discrete Wavelet Transform to get the wavelet coefficients $w(s, t_s)$ at each dyadic scale s and time interval t_s.
- The wavelet leaders at each scale s are then extracted by computing the maxima of the wavelet coefficients over all samples involved for computing $w(s, t_s - 1)$, $w(s, t_s)$ and $w(s, t_s + 1)$ (including lower scales). Using three coefficients instead of one is justified in [1] and achieves a bias/variance tradeoff.
- Typically the partition functions are then computed: $F(s, q) = \frac{1}{N_s} \sum_{t_s=1}^{N_s} |w(s, t_s)|^q$ for an adequate range of exponents q, then
- a Legendre transform or a direct Holder exponent density estimation as in [13] is performed to obtain the multifractal spectrum. Here we use instead the recent technique introduced by [52] and compute the wavelet leader cumulants of orders 1 to 5 (see that reference). As noted in [52] the first few cumulants already contain most of the information useful in practice for characterising the distribution of the Holder exponents. For a classification task this information can now be exploited in a more condensed form. Compared to what we previously did in [13] we thus have a feature that better characterizes the multifractal properties of the signal.
- The 5 first cumulants are computed for the leaders at each scale s. Considering there is at most L levels of wavelet transform in a signal of size between 2^L and 2^{L+1}, we get a total of $5 \times L$ cumulants for the signal, that progressively encompass more and more frequency bands as the scale increases. These $5 \times L$ cumulants per channel are used as the feature vector.

This method can be quite sensitive to the presence of electromagnetic interferences at 50 Hz. We thus pre-filter the signals as described in section 4.1 before proceeding to the multifractal cumulants estimation.

3 Predictive complexity measure

This paper introduces for the first time the predictive complexity measure of [14] in the context of EEG classification.

The intuitive idea behind this feature is to quantify the amount of information that is necessary to retain from the past of the series in order to be able to predict optimally the future of the series ([46], and see below for the optimality criterion).

We had indication from related previous works that the feature could be relevant for EEGs:

- At the level of neurons: statistical complexity was used to describe the computational structure of spike trains [26]. It was also shown to decrease while the neurons are learning in an artificial spiking neural network [12].
- At the level of the brain: information coherence and synchronisation mechanism between communities of neurons are presented in [30], relying on related techniques.

3.1 Decisional states and the corresponding complexity measure

Informally, the idea behind the decisional states is to construct a Markovian automaton [19, 46] whose states correspond to taking the same decisions [14], according to a user-defined utility function. These decisions are those that one can take based on predictions of the future and their expected utility.

The complexity of the series is then computed as the mutual information between the internal states of the Markovian automaton, and the series itself. The complexity is null for a very regular series, for example a constant series or a series where we always take the same decision: there is only a single state in the automaton. Similarly the complexity of a completely random series is also null: it can be modelled by successive independent draws from a fixed probability distribution, whose expected utility we take to make our decision. This leads again to a single Markovian automaton state, hence a null complexity. The complexity measure increases only for more complicated series with many internal states (i.e. many distinct probability distributions of what happens next, depending on what previously happened, leading to different decisions).

Presumably when the EEG corresponds or not to some functional activity, the complexity of the series should change. The idea is to plug machine learning techniques for monitoring that change. For example, when there is only a background activity in the left and right motor areas of the brain we shall observe a given complexity level of the corresponding EEG signals. If there is an activity in only one of these two brain areas then the complexity should probably change in that area and not in the other. Even if very unpredictable and very predictable signals have in theory about the same complexity, in practice we never expect EEG signals to be either totally unpredictable (i.e. they are only background patterns) or very predictable (i.e. the EEGs have inherent fluctuations unrelated to the task we monitor) during a motor imagery task. Thus we expect that a machine learning algorithm trained to distinguish the differences in complexities of the EEG signals recorded over the left and right motor cortices would successfully identify some change in mental state. Results in section 4 show that for some subjects this criterion alone allows for more than 90% recognition rate.

Formal description

Formally, let (s_t) be a time series, with t the time index. Let $s_t^{-\infty} = (s_u)_{u \leq t}$ and $s_t^{+\infty} = (s_u)_{t<u}$ be the past and future histories at time t. In practice and for real measures, the time range is finite: $0 \leq t \leq T$. Similarly we measure the past and future histories with finite horizons: $s_t^{-h} = (s_u)_{t-h \leq u \leq t}$ and $s_t^{+k} = (s_u)_{t<u \leq t+k}$.

Let us now consider predicting the future from the past statistically: we seek to determine $P\left(s_t^{+k} | s_t^{-h}\right)$ at each time t (possibly with h or k infinite). Assuming the system is conditionally stationary, that distribution does not change through time: the same causes produce the same consequences. In the context of BCI we assume that the brain patterns monitored over a few seconds during motor imagery tasks respect this criteria. This is indeed a reasonable hypothesis: the signal itself may not be stationnary, but over that short period the cause of the detected patterns does not change.

Let then $S^{-h} = \left\{s_t^{-h}\right\}_{\forall t}$ and $S^{+k} = \left\{s_t^{+k}\right\}_{\forall t}$ the sets of all past and future histories. We will drop the time indices from now on to indicate the time shift invariance.

The causal states ζ are defined as the equivalence classes of past histories with the same conditional distribution of futures: $\zeta\left(s^{-h}\right) = \left\{x \in S^{-h} : P\left(s^{+k}|x\right) = P\left(s^{+k}|s^{-h}\right)\right\} = \left\{x : x \overset{c}{\equiv} s^{-h}\right\}$. Knowing the causal state at the current time is the minimal information needed for making optimal predictions [46] using the full conditional probability distribution.

In the discrete case the series are strings of symbols drawn from an alphabet $\mathcal{A}$. Each time step implies a symbol transition, which possibly leads to a different causal state. The corresponding automaton is called the ϵ-machine [19].

Let us now introduce a utility function $u : \left(S^{+k}\right)^2 \mapsto \mathbb{R}$, such that $u\left(r^{+k}, s^{+k}\right)$ quantifies the gain (positive) or loss (negative) when the user relied on the prediction r^{+k} while s^{+k} actually happened. We can now define an expected utility: $\mathbb{U}\left[r^{+k}|s^{-h}\right] = \mathbb{E}_{s^{+k} \in S^{+k}}\left[u\left(r^{+k}, s^{+k}\right)|s^{-h}\right]$, quantifying what utility can be expected on average when choosing the prediction r^{+k} for the current system state s^{-h}. The set of optimal predictions, realising the maximal expected utility, can now be defined as $Y\left(s^{-h}\right) = \text{argmax}_{r^{+k} \in S^{+k}} \mathbb{U}\left[r^{+k}|s^{-h}\right]$.

The following equivalence relations $\overset{p}{\equiv}$, $\overset{u}{\equiv}$, and $\overset{d}{\equiv}$ naturally extend the causal state equivalence relation $\overset{c}{\equiv}$, taking into account the utility function:

- $r^{-h} \overset{p}{\equiv} s^{-h}$ when $Y\left(r^{-h}\right) = Y\left(s^{-h}\right)$, with the corresponding *iso-prediction sets* as equivalence classes. All past histories within the same class lead to choosing the same predictions, even though the expected utility may change from one past history to the other.

- $r^{-h} \overset{u}{\equiv} s^{-h}$ when $\max_{r^{+k} \in S^{+k}} \mathbb{U}\left[r^{+k} | r^{-h}\right] = \max_{r^{+k} \in S^{+k}} \mathbb{U}\left[r^{+k} | s^{-h}\right]$, with the corresponding *iso-utility sets* as equivalence classes. All past histories within the same class lead to the same maximum expected utility, even though the optimal predictions to choose for reaching this utility are not specified.

- $r^{-h} \overset{d}{\equiv} s^{-h}$ when $r^{-h} \overset{p}{\equiv} s^{-h}$ and $r^{-h} \overset{u}{\equiv} s^{-h}$. The intersection of the above iso-utility and iso-prediction sets define the *decisional states*: $\Psi(s^{-k}) = \left\{x \in S^{-h} : x \overset{p}{\equiv} s^{-h}, x \overset{u}{\equiv} s^{-h}\right\}$. We assume that when both the maximal expected utility and the optimal predictions are the same, the user will reach the same decisions. In other words, the utility function encodes all the user needs to know to reach a decision.

It can be easily shown [14] that the causal states sub-partition the decisional states. That is to say, the causal states have lost their minimality property due to the fact that we are not interested in the full conditional distribution of futures but only in the optimal decisions with respect to a user-defined utility function.

The causal states are an intrinsic property of the data set. The mutual information $M_c = I(s^{-h}; \zeta\left(s^{-h}\right))$ between the causal $\zeta\left(s^{-h}\right)$ state and the series of s^{-h} defines an intrinsic measure, the statistical complexity [19], quantifying how hard it is to get the conditional distribution of futures from the current observed past.

The decisional states correspond to the structure implied by the user utility function on top of the causal states. As for the causal states, knowing the decisional state at the current time is the information needed for making optimal predictions maximising the user-defined utility function. The mutual information $M_d = I(s^{-h}; \Psi\left(s^{-h}\right))$ similarly defines a complexity measure for how hard it is to make these predictions, called decisional complexity by analogy with the statistical complexity.

3.2 Application to EEG data

In the present study the chosen utility function is the negative sum of square error: $u\left(r^{+k}, s^{+k}\right) = -\left\|r^{+k} - s^{+k}\right\|^2$.

Each EEG series is split in time windows of $h + k$ samples, defining each a $\left(s^{-h}, s^{+k}\right)$ pair. These observations are fed in the reconstruction algorithm presented in [14]. That algorithm estimates the joint probabilities $p\left(s^{-h}, s^{+k}\right)$ using a kernel density estimation with Gaussian kernels. The conditional probabilities are then computed by integration over S^{+k}: $p\left(s^{+k} | s^{-h}\right) = p\left(s^{-h}, s^{+k}\right) / \int_{\sigma^{+k} \in S^{+k}} p\left(s^{-h}, \sigma^{+k}\right)$. These are then clustered into causal state estimates $\zeta\left(s^{-h}\right) = \left\{x \in S^{-h} : p\left(s^{+k} | x\right) = p\left(s^{+k} | s^{-h}\right)\right\}$ using

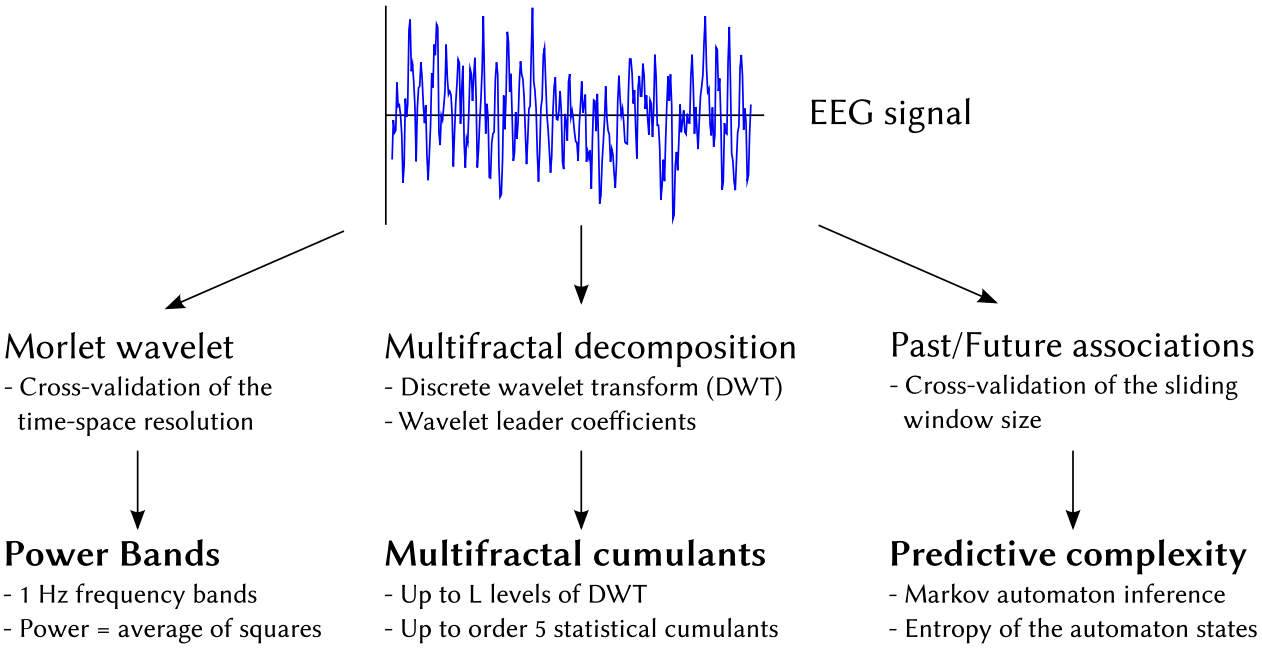

Figure 1: Extraction steps for each feature

connected components up to a fixed Bhattacharyya distance threshold (chosen to be 0.05) between the conditional distributions. The utility function is applied on top of the aggregated causal states distributions $p\left(s^{+k}|\zeta\right) = \text{avg}_{s^{-h}\in\zeta} p\left(s^{+k}|s^{-h}\right)$ in order to get the expected utilities $\mathbb{U}\left[r^{+k}|\zeta\right]$. Finally, the causal states are themselves clustered into iso-utility and iso-prediction sets, which are intersected to get the decisional states ω.

The decisional complexity is the feature that is extracted from the EEG series. We thus get one feature per channel.

Figure 1 displays the extraction steps for each feature.

4 Evaluation

We evaluated the performances of our new features and their combinations on classical motor-imagery benchmarks, so they can be compared with the litterature [45]. We also worked on raw data captured by home-made software, that allows us to avoid any interference with the pre-filtering performed in the BCI competition data sets [8][9]. That last raw data is made available to the community, a link is provided at the end of the document.

4.1 Data sets

Four data sets for a total of 13 subjects were used for evaluation in this study. Data for 12 subjects come from the international BCI competitions[1] II, III, and IV [8][9] and was captured at the Department of Medical Informatics, Institute for Biomedical Engineering, University of Technology Graz [45]. The data for the last subject was acquired at INRIA (French National Research Institute on Computer Science and Control) Rennes-Bretagne Atlantique, using the OpenViBE software platform [43][2]. In all these data sets, EEG signals were collected while subjects were performing left hand and right motor imagery, according to the Graz BCI protocol [42]. With this protocol, During each trial, the subject is presented with a visual cue indicating either left or right at random, and shall then imagine a movement of the corresponding hand during 6s. For each data set, we used all the available EEG data over the motor imagery period in order to extract the features. Differences between the data sets include: the number of subjects from which EEG were recorded, the EEG electrodes used, the sampling rate, the number of trials collected per class and the preprocessing method used. These differences are highlighted in the following.

4.1.1 BCI competition II, data set III (1 subject)

This data set contains 280 trials sampled at 128 Hz. There was 140 trials of left class (left hand motor imagery) and 140 trials of the right class (right hand motor imagery). EEG were recorded using the C3, C4 and Cz electrodes, however, for the purpose of this evaluation, we used only the C3 and C4 electrodes as recommended in [31]. More details about this data set can be found in [8].

The data was already preprocessed by a band-pass filter between 0.5 and 30 Hz. Unfortunately, the nature of the filter is not specified, and the DC component was not well removed. Since this interferes in particular with the algorithm for estimating the series complexity, we designed a new FIR filter using METEOR [47] in order to get: less than 1dB change in 4-30Hz with a linear phase response, at least -50dB at 0 Hz (and above 40 Hz, though in the present case the signal is already filtered) and 1/4 sec delay.

4.1.2 BCI competition III, data set IIIb (2 subjects)

This data set originally consists of EEG signals from 3 subjects. However, for the purpose of this study, only subjects labeled S4 and X11 were used. Indeed, EEG signals for subject O3VR were recorded using a different protocol and the data file provided online contained erroneously duplicate signals[3]. The data was captured at a 125 Hz sampling rate using electrodes C3 and C4. More details about this data set can be found in [9]. EEG signals in this data set were already band-pass filtered in 0.5-30 Hz. As for the previous

[1] http://www.bbci.de/competition/

[2] http://openvibe.inria.fr/?q=datasets

[3] See http://www.bbci.de/competition/iii/desc_IIIb_ps.html for details

data set, an additional FIR filter with the same characteristics as the previous one was applied, for the same reasons.

4.1.3 BCI competition IV, data set IIb (9 subjects)

This third data set contains occular artifacts that interfere with the brain signals. Although this is less convenient for the purpose of testing our two new features, it is also more realistic. We thus choosed to ignore the artifacts and process as if the signals were clean EEGs, in order to assert the robustness of our features to occular artifacts.

Data coming from nine subjets is sampled at 250 Hz, and pre-processed by a band-pass filter between 0.5 and 100 Hz with a notch at 50 Hz. For the aforementioned reasons we had to re-filter the signals. We choosed a FIR design with a band-pass between 6 and 30 Hz, with a null at 0Hz to suppress the DC component and -50dB above 40 Hz.

4.1.4 OpenViBE / INRIA data (1 subject)

. This data set comprises EEG signals from a single subject for which 560 trials of motor imagery (280 trials per class) were recorded over a 2 week period. Half of the trials (randomly selected from all experiments over time) is used for training, the remaining half for testing.

EEG data was sampled at 512 Hz and recorded using the following electrodes: C3, C4, FC3, FC4, C5, C1, C2, C6, CP3, CP4, with a nose reference electrode. The use of such electrodes enables us to apply a discrete Laplacian spatial filter [29] over C3 and C4 in order to obtain better signals, as recommended in [36].

The data was preprocessed by a FIR filter we designed with METEOR [47] in order to get: less than 1dB change in 4-30Hz and a linear phase response in this range, at least -50dB at 0 Hz and above 50 Hz, null responses at 50 Hz and all harmonics, and a 1/4 sec delay.

4.2 Results obtained with all features and their combinations

The goal of this section is to show what results can be obtained with each feature considered in isolation, and the effect of combining them. The hypothesis we wish to test is that each feature extracts a different information from the signal, and thus that combining them can improve the classification accuracy.

4.2.1 Features

The following methodology was applied:

- The power in frequency bands was extracted for each subject. As was shown by [27] and in our own study [15] estimating this information from data is sensitive to choice of the time-frequency decomposition algorithm. We extracted the power information in each frequency band using a Morlet Wavelet which support is determined by cross-validation. The Morlet Wavelet is indeed adapted for BCI tasks, as shown for example by the winners of the BCI competition 2003 [31]. Moreover, according to [28]: "Using sinusoidal wavelets like the Morlet wavelet is ideally suited for detecting sinusoidal EEG activity since the wavelet transform is similar to detecting whether the used wavelet is contained in the signal". Finally, as we showed in [15], cross-validating the wavelet support performs better than other techniques for extracting the power-information. All bands were then kept between 4 and 30 hertz, for the C3 and C4 channels: this leads to a 52-dimensional feature vector.

- The multifractal cumulants feature was extracted according to the method described in Section 2. Cross-validation on the training set was used to determine the wavelet support and the number of decomposition levels to retain. Since we have 5 cumulants per level per electrode, this leads to a $10 * N$-dimensional vector with N the number of retained levels.

- Parameters for estimating the predictive complexity feature were also determined by maximising the cross-validation performance on the training set: the number of points to retain from the past and the future, the sub-sampling factor for the series, the kernel size used for estimating the conditional probability distributions, and whether to work on the raw series or the first differences. Cross-validation

selected 1 point in the future, 5 points from the past with a sub-sampling factor of 8 for the OpenViBE subject, and 6 points from the past with a sub-sampling factor of 2 for the BCI II and III subjets. These sub-sampling parameters correspond to having a sampling frequency of 62.5 Hz for the S4 and X11 subjects and 64 Hz for the others, which thanks to Nyquist theorem matches the filtering operation that was performed on the signals: cross-validation selected the most economical sub-sampling parameter that still captures the remaining frequencies in the signal between 4 and 30 Hz. We thus saved computational time for the BCI IV subjets by fixing the subsampling parameters and cross-validating only the kernel width. All sub-sampled signals (ex: all 8 possible series for a sub-sampling by factor 8) were kept for building the statistics in the complexity computations.

Therefore the feature vectors include:

- 52 power features for frequencies between 4 and 30 Hz. (26 for each C3 / C4 channel)
- 30, 40, or 50 features for multifractal cumulants.
- 2 features for the predictive complexity (one for each channel).

As a side remark we shall point out that cross-validating all the above parameters required a quite consequent computational power, which was obtained using the BOINC distributed computing infrastructure [5]. Nevertheless, once the optimal parameters were determined off-line by cross-validation, the computational requirements for extracting the features are nothing exceptional: the features can be computed in real-time on a standard PC.

4.2.2 Classifier and combination rule

In order to classify the extracted features, we used a Linear Discriminant Analysis (LDA), one of the most popular and efficient classifier for EEG-based BCI [32].

Each series was classified independently into either the "left hand" or "right hand" class of the motor imagery task using each feature independently. The percentage of correct classifications on the test set is reported for each subjet in the first 3 colums of Table 1.

We also evaluated the accuracy obtained when combining these three features together. To do so, we first estimate the global accuracy of each feature alone by using the Fisher discriminant ratio F on the training data set, hence we get F^{pow}, F^{mfa}, F^{cpx} for the three features Band-Power, Multifractal cumulants, and Complexity, respectively. In order to combine the results we tried different methods (results given in Table 1):

- Method 1: We gather the outputs of the 3 linear classifiers trained independently on each feature by performing a simple arithmetic average for each instance $P_i^{\mathrm{combi}} = F^{\mathrm{pow}} * P_i^{\mathrm{pow}} + F^{\mathrm{mfa}} * P_i^{\mathrm{mfa}} + F^{\mathrm{cpx}} * P_i^{\mathrm{cpx}}$ where each P_i is a prediction (classifier output) for the instance number i.
- Method 2: We use the median prediction for each instance $P_i^{\mathrm{combi}} = Median[F^{\mathrm{pow}} * P_i^{\mathrm{pow}}, F^{\mathrm{mfa}} * P_i^{\mathrm{mfa}}, F^{\mathrm{cpx}} * P_i^{\mathrm{cpx}}]$. The idea is to be less sensitive to badly performing features than when using the first combination rule.
- Method 3: We use the maximum prediction for each instance $P_i^{\mathrm{combi}} = Max_{\mathrm{m=pow,mfa,cpx}}[P_i^{\mathrm{m}}, right;\ 1 - P_i^{\mathrm{m}}, left]$ for classifying in the right of left class. This is motivated by previous results in [13].
- Unique Vector: For comparison purpose only. We include all features in a unique vector. This is the naive approach, but it is expected to fail due to the curse of dimensionality and due to giving all features the same weight irrespectively of the number of dimension they take.

4.2.3 Results

The multifractal cumulants and the predictive complexity features allow by themselves to classify the BCI data sets with a good accuracy. This, in itself, is a contribution of this paper. Although the accuracy obtained using either the Multifractal Cumulants or the Predictive Complexity is slightly lower on average that when using only the Band Power feature (sometimes higher), we have extracted some information from the signal

Subjet	**Band-Power**	**Multifractal**	**Complexity**	Average method	Gain (%)	Median method	Gain (%)	Maximum method	Gain (%)	Unique vector	Gain
bci2	77.1	80.7	77.9	80.7	3.57	80.7	3.57	81.4	4.29	75.7	-1.43
s4	81.5	74.4	65.0	81.7	0.19	81.7	0.19	80.7	-0.74	78.1	-3.33
x11	80.4	68.7	70.2	80.9	0.56	80.9	0.56	80.2	-0.19	79.4	-0.93
OpenVibe	92.9	85.7	76.1	93.2	0.36	93.2	0.36	92.1	-0.71	92.1	-0.71
bci4s1	77.5	65.9	55.9	74.4	-3.12	75.6	-1.88	72.8	-4.69	73.4	-4.06
bci4s2	56.4	57.5	56.1	56.1	-0.36	56.4	0.00	57.1	0.71	52.1	-4.29
bci4s3	51.9	51.2	54.1	52.2	0.31	51.9	0.00	54.7	2.81	53.8	1.88
bci4s4	93.4	91.2	90.3	95.6	2.19	95.0	1.56	95.0	1.56	91.6	-1.88
bci4s5	96.9	88.8	74.1	95.9	-0.94	96.2	-0.62	94.1	-2.81	93.8	-3.12
bci4s6	87.8	82.2	68.4	89.1	1.25	89.1	1.25	88.4	0.62	87.8	0.00
bci4s7	70.6	68.1	70.3	72.2	1.56	72.2	1.56	72.8	2.19	68.4	-2.19
bci4s8	80.0	88.4	90.6	89.1	9.06	89.1	9.06	89.1	9.06	85.9	5.94
bci4s9	78.8	83.1	78.8	82.8	4.06	80.9	2.19	80.6	1.88	80.9	2.19
Average	78.9	75.8	71.4	80.3	1.44	80.2	1.37	79.9	1.08	77.9	-0.92

Legend: Green when the classification accuracy of the combination is higher by at least 1% than when using only the usual Band-Power feature, blue when the change in performance with the combination is between -1% and 1%, orange when the combination is worse.

Table 1: Classification accuracies over the test sets for each subject and feature

that is adequate for classification. Moreover for some subjects the new features perform better than the band power feature. This effect is variable from subject to subject, but this is a well known issue in the BCI litterature [23, 2] so it is not suprising that our new features also exhibit inter-subject variability. Possible reasons include differences in activity localization between subjects, thus the electrodes used work no better for our new features than they do for the band power (ex: bci4s2, bci4s3). However we obtained higher results with the new features for some subjects with low results using classical band power feature (bci4s8, bci2): a to-be-invented better combination technique might possibly exploit these or other new features for compensating the inter-subject variability.

Results also show that combining all features is better than using the Band Power alone for 6 subjects, on the same order for 6 subjects, and with a negative impact for only one subject, for both the average and the median combination rules. As expected the median rule gives slightly more consistent results across individuals than the average combination rule: the expected gain is about the same at 1.4%, but with a lower standard deviation. The maximum decision rule on the other hand leads to more contrasted results (see Table 1). The combination approach based on a unique vector led to the lowest performances among the difference combination approaches. Again, this was expected, as concatenating all features together makes the classifier even more sensitive to the curse-of-dimensionality (which is a well known issue in BCI [32]), since the feature vector dimension greatly increased but not the number of training data.

Adding features has a net benefit on global performances, even if for some subjects the combination is occasionnaly worse than using the band power alone. But this is also true for our new features: occasionally adding the band power is worse than using the predictive complexity alone. In other words, there is no a priori reason not to consider only one feature in isolation of the other, as the combination leads to better results on average. The inter-subject variability manifests itself again: in real situations with a training phase for the subject, a quick test with each feature alone may determine if one or the other works best in isolation.

In any case these results not only confirm that some different information was extracted from the signal with the new features, but also that our new features can improve the performance of BCI classification tasks. Moreover, depending on the subject, this gain can go up to 9%. Finally, the increase in computational

complexity is not a serious issue due to the rapidly growing computational performances of computers.

5 Conclusion

This study has introduced two new features for Brain-Computer Interface design: multifractal cumulants and predictive complexity. The information contained in the multifractal cumulants feature corresponds to a relation between frequency bands, rather than the power extracted in each band. The complexity feature measures the difficulty to predict the future of the EEG signals based on their past.

Interestingly enough, our results showed that the two new features, i.e., multifractal cumulants and predictive complexity measure, could be used by themselves to discriminate between different motor imagery mental states as measured by EEG. This is especially interesting as the complexity feature only adds one scalar value per electrode, compared to the other features we considered (ex: one dimension per frequency band in the band power case).

Moreover, our results showed that when combining these two features together with band power features, the resulting BCI could reach a better classification accuracy than when using the usual band power feature alone. Thus this suggests that these new features are a good complement to currently used features for BCI design and that they can lead to improved BCI design. Therefore we would recommend BCI designers to consider these two features as additional features in the conception of a BCI system, in order to obtain better performance.

Future work might involve the exploration of novel ways to combine the features and feature selection, as well as the application of these features to BCI tasks other than motor imagery. Work is also needed on the design of novel algorithms including physiologically relevant error functions for EEG signal predictions for the complexity feature.

Acknowlegdements

Part of this work was developped by Nicolas Brodu while being at the Laboratoire du Traitement du Signal et de l'Image (LTSI) at University of Rennes 1

Part of this work was financed by the French National Research Agency (ANR) in the form of the OpenViBE2 project.

Appendix: Source code, data, and web information

All the results in this study are reproducible independently. The code for the experiments is provided as Free/Libre software on the main author web site:

- http://nicolas.brodu.numerimoire.net/en/recherche/publications/

The data that was used can be downloaded on the BCI competitions website and on the OpenViBE project page:

- http://www.bbci.de/competition/
- http://openvibe.inria.fr/?q=datasets

References

[1] P. Abry, S. Jaffard, and B. Lashermes. Revisiting scaling, multifractal, and multiplicative cascades with the wavelet leader lens. In *SPIE*, pages 103–117, 2004.

[2] B. Allison and C. Neuper. *Could Anyone Use a BCI?*, chapter 3, pages 35–54. Human-Computer Interaction Series. Springer London, 2010.

[3] C. W. Anderson and Z. Sijercic. Classication of EEG signals from four subjects during five mental tasks. In *Solving Engineering Problems with Neural Networks: Proceedings of the International Conference on Engineering Applications of Neural Networks (EANN'96)*, 1996.

[4] C. W. Anderson, E. A. Stolz, and S. Shamsunder. Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks. *IEEE Transactions on Biomedical Engineering*, 45:277–286, 1998.

[5] D.P. Anderson. BOINC: A system for public-resource computing and storage. In *5th IEEE/ACM International Workshop on Grid Computing.*, November 2004.

[6] A. Bashashati, M. Fatourechi, R. K. Ward, and G. E. Birch. A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of Neural engineering*, 4(2):R35–57, 2007.

[7] B. Blankertz, F. Losch, M. Krauledat, G. Dornhege, G. Curio, and K.-R. Müller. The berlin brain-computer interface: Accurate performance from first-session in BCI-naive subjects. *IEEE Trans Biomed Eng*, 55(10):2452–2462, 2008.

[8] B. Blankertz, K. R. Müller, G. Curio, T. M. Vaughan, G. Schalk, J. R. Wolpaw, A. Schlögl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schröder, and N. Birbaumer. The BCI competition 2003: Progress and perspectives in detection and discrimination of EEG single trials. *IEEE Transactions on Biomedical Engineering*, 51(6):1044–1051, 2004.

[9] B. Blankertz, K. R. Muller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlogl, G. Pfurtscheller, J. D. R. Millan, M. Schroder, and N. Birbaumer. The BCI competition III: Validating alternative approaches to actual BCI problems. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):153–159, 2006.

[10] R. Boostani, B. Graimann, M.H. Moradi, and G. Pfurtscheller. A comparison approach toward finding the best feature and classifier in cue-based BCI. *Medical and Biological Engineering and Computing*, 45(4):403–412, 2007.

[11] V. Bostanov. BCI competition 2003–data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram. *IEEE Transactions on Biomedical Engineering*, 51(6):1057–1061, 2004.

[12] N. Brodu. Quantifying the effect of learning on recurrent spikin neurons. In *IJCNN*, pages 512–517, 2007.

[13] N. Brodu. Multifractal feature vectors for brain-computer interfaces. In *IJCNN*, pages 2883–2890, 2008.

[14] N. Brodu. Reconstruction of epsilon-machines in predictive frameworks and decisional states. *Advances in Complex Systems*, 2011. accepted for publication, to appear.

[15] N. Brodu, F. Lotte, and A. Lécuyer. Comparative study of band-power extraction techniques for EEG-based brain-computer interfaces. In *IEEE Symposium Series on Computational Intelligence, Cognitive Algorithms, Mind and Brain (IEEE SSCI CCMB 2011)*, 2011.

[16] C. Brunner, M. Billinger, and C. Neuper. A comparison of univariate, multivariate, bilinear autoregressive, and bandpower features for brain-computer interfaces. In *Fourth International BCI Meeting*, 2010.

[17] D. P. Burke, S. P. Kelly, P. deChazal deChazal, R. B. Reilly, and C. Finucane. A parametric feature extraction and classification strategy for brain-computer interfacing. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13(1), 2005.

[18] D. Coyle, G. Prasad, and T. M. McGinnity. A time-frequency approach to feature extraction for a brain-computer interface with a comparative analysis of performance measures. *EURASIP J. Appl. Signal Process.*, 2005(1):3141–3151, 2005.

[19] J. Crutchfield and K. Young. Inferring statistical complexity. *Physical Review Letters*, 62(2):105–108, 1989.

[20] J. del R. Millán, P.W. Ferrez, and A. Buttfield. *Towards brain-computer interfacing*, chapter The IDIAP brain-computer interface: An Asynchronous Multiclass Approach. MIT Press, G. Dornhege, R. Millan Jdel, T. Hinterberger, D. J. McFarland & K. R. Müller edition, 2007.

[21] Gary N. Garcia, T. Ebrahimi, and J.-M. Vesin. Joint time-frequency-space classification of EEG in a brain-computer interface application. *EURASIP J. Appl. Signal Process.*, 2003(1):713–729, 2003.

[22] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut. Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. *IEEE Transactions on Neural System and Rehabilitation Engineering*, 11:141–144, 2003.

[23] C. Guger, G. Edlinger, W. Harkam, I. Niedermayer, and G. Pfurtscheller. How many people are able to operate an EEG-based brain-computer interface (BCI). *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, (11):145–147, 2003.

[24] E. Gysels and P. Celka. Phase synchronization for the recognition of mental tasks in a brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 12(4):406–415, 2004.

[25] P.S. Hammon and V.R. de Sa. Preprocessing and meta-classification for brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*, 54(3):518–525, 2007.

[26] R. Haslinger, K. Klinkner, and C.R. Shalizi. The computational structure of spike trains. *Neural Computation*, 22:121–157, 2010.

[27] P. Herman, G. Prasad, T.M. McGinnity, and D. Coyle. Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16(4), August 2008.

[28] C.S. Herrmann. *EEG oscillations and wavelet analysis.* Maren Grigutsch & Niko A. Busch, MIT Press, 2005.

[29] B. Hjorth. An on-line transformation of EEG scalp potentials into orthogonal source derivations. *Electroencephalography and Clinical Neurophysiology*, 39:526–530, 1975.

[30] K. Klinkner, C.R. Shalizi, and M. Camperi. Measuring shared information and coordinated activity in neuronal networks. *Advances in Neural Information Processing Systems*, 18:667–674, 2006.

[31] S. Lemm, C. Schafer, and G. Curio. BCI competition 2003–data set III: probabilistic modeling of sensorimotor mu rhythms for classification of imaginary hand movements. *IEEE Transactions on Biomedical Engineering*, 51(6):1077–1080, 2004.

[32] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms for EEG-based brain-computer interfaces. *Journal of Neural Engineering*, 4:R1–R13, 2007.

[33] F. Lotte and C.T. Guan. Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms. *IEEE Transactions on Biomedical Engineering*, 58(2):355–362, 2011.

[34] B. Mandelbrot, A. Fisher, and L. Calvet. A multifractal model of asset returns. Technical Report 1164, Cowles Fundation, 1997.

[35] D. J. McFarland, C. W. Anderson, K.-R. Muller, A. Schlogl, and D. J. Krusienski. BCI meeting 2005-workshop on BCI signal processing: feature extraction and translation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):135–138, 2006.

[36] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw. Spatial filter selection for EEG-based communication. *Electroencephalographic Clinical Neurophysiology*, 103(3):386–394, 1997.

[37] J.R. Millán, M. Franzé, J. Mouriño, F. Cincotti, and F. Babiloni. Relevant EEG features for the classification of spontaneous motor-related tasks. *Biological Cybernetics*, 86(2):89–95, 2002.

[38] J.-F. Muzy, E. Bacry, and A. Arneodo. Multifractal formalism for fractal signals: The structure-function approach versus the wavelet-transform modulus-maxima method. *Physical Review E.*, 47:875–884, 1993.

[39] B. Obermeier, C. Guger, C. Neuper, and G. Pfurtscheller. Hidden markov models for online classification of single trial EEG. *Pattern recognition letters*, pages 1299–1309, 2001.

[40] G. Pfurtscheller, C. Brunner, R. Leeb, R. Scherer, G. Müller-Putz, and C. Neuper. The graz brain-computer interface. In Avshalom C. Elitzur, Mark P. Silverman, Jack Tuszynski, Rüdiger Vaas, H. Dieter Zeh, Bernhard Graimann, Gert Pfurtscheller, and Brendan Allison, editors, *Brain-Computer Interfaces*, The Frontiers Collection, pages 79–96. Springer Berlin Heidelberg, 2010.

[41] G. Pfurtscheller, D. Flotzinger, and J. Kalcher. Brain-computer interface-a new communication device for handicapped persons. *journal of microcomputer application*, 16:293–299, 1993.

[42] G. Pfurtscheller and C. Neuper. Motor imagery and direct brain-computer communication. *proceedings of the IEEE*, 89(7):1123–1134, 2001.

[43] Y. Renard, F. Lotte, G. Gibert, M. Congedo, E. Maby, V. Delannoy, O. Bertrand, and A. Lécuyer. OpenViBE: An open-source software platform to design, test and use brain-computer interfaces in real and virtual environments. *Presence: teleoperators and virtual environments*, 19(1), 2010.

[44] A. Schlögl, K. Lugger, and G. Pfurtscheller. Using adaptive autoregressive parameters for a brain-computer-interface experiment. In *Proceedings 19th International Conference IEEE/EMBS*, pages 1533–1535, 1997.

[45] A. Schlögl, C. Neuper, and G. Pfurtscheller. Estimating the mutual information of an EEG-based brain-computer-interface. *Biomedizinische Technik*, 47(1-2):3–8, 2002.

[46] C.R. Shalizi. *Causal Architecture, Complexity, and Self-Organization in Time Series and Cellular Automata*. PhD thesis, University of Wisconsin at Madison, 2001.

[47] K. Steiglitz, T. W. Parks, and J. F. Kaiser. METEOR: A constraint-based FIR filter design program. *IEEE Transactions on Signal Processing*, 40(8):1901–1909, 1992.

[48] C. Vidaurre, N. Krämer, B. Blankertz, and A. Schlögl. Time domain parameters as a feature for EEG-based brain computer interfaces. *Neural Networks*, 22:1313–1319, 2009.

[49] C. Vidaurre, R. Scherer, R. Cabeza, A. Schlögl, and G. Pfurtscheller. Study of discriminant analysis applied to motor imagery bipolar data. *Med. Biol. Engineering and Computing*, 45(1):61–68, 2007.

[50] T. Wang, J. Deng, and B. He. Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns. *Clinical Neurophysiology*, 115(12):2744–2753, 2004.

[51] Q. Wei, Y. Wang, X. Gao, and S. Gao. Amplitude and phase coupling measures for feature extraction in an EEG-based brain–computer interface. *Journal of Neural Engineering*, 4(2):120, 2007.

[52] H. Wendt, P. Abry, and S. Jaffard. Bootstrap for empirical multifractal analysis. *IEEE Signal Processing Magazine*, (38), 2007.

[53] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M. Vaughan. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767–791, 2002.

[54] Y.P.A. Yong, N.J. Hurley, and G.C.M. Silvestre. Single-trial EEG classification for brain-computer interfaces using wavelet decomposition. In *European Signal Processing Conference, EUSIPCO 2005*, 2005.