Feature extraction for EEG-based brain–computer interfaces by wavelet packet best basis decomposition

2006 J. Neural Eng. 3 251

(http://iopscience.iop.org/1741-2552/3/4/001)

J. Neural Eng. **3** (2006) 251–256 doi:10.1088/1741-2560/3/4/001

Feature extraction for EEG-based brain–computer interfaces by wavelet packet best basis decomposition

Bang-hua Yang, Guo-zheng Yan, Rong-guo Yan and Ting Wu

School of Electronic, Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China

E-mail: ybh@sjtu.edu.cn and yangbanghua@126.com

Received 10 March 2006
Accepted for publication 18 August 2006
Published 5 September 2006
Online at stacks.iop.org/JNE/3/251

Abstract

A method based on wavelet packet best basis decomposition (WPBBD) is investigated for the purpose of extracting features of electroencephalogram signals produced during motor imagery tasks in brain–computer interfaces. The method includes the following three steps. (1) Original signals are decomposed by wavelet packet transform (WPT) and a wavelet packet library can be formed. (2) The best basis for classification is selected from the library. (3) Subband energies included in the best basis are used as effective features. Three different motor imagery tasks are discriminated using the features. The WPBBD produces a 70.3% classification accuracy, which is 4.2% higher than that of the existing wavelet packet method.

1. Introduction

Unilateral limb motor imagery can produce changes in the electroencephalogram (EEG) signal, notably in the sensorimotor cortex contralateral to the limb. These variations can be detected over some specified frequency bands, such as μ rhythm (8–12 Hz) or β rhythm (18–26 Hz) [1]. Different motor imagery tasks can produce different EEG patterns. Recognizing these EEG patterns can help people with severe motor disabilities to acquire alternative methods for communication and control, which is one of the intentions of EEG-based brain–computer interfaces (BCIs).

BCIs that are based on the motor imagery EEG have the advantages that they do not need external stimulus and the production of signals is dependent only on thinking [2]. Therefore, these kinds of BCIs are adopted by many researchers. The accurate recognition of the EEG pattern is a key problem for the realization of the BCIs. The recognition procedure mainly includes the feature extraction and the classification, in which the feature extraction plays an important role for the classification. This paper mainly focuses on the feature extraction.

At present, feature extraction methods for the motor imagery EEG mainly include the following. (1) Fast Fourier transform (FFT): in [3, 4], the Fourier spectral features were computed with the Welch method using windowed Fourier transforms of signal segments. The main disadvantage of this method is that it uses only the frequency information and does not use time domain information. However, the research shows that the combination of frequency information and time domain information can improve the classification performance of the EEG signal [5]. (2) Autoregressive (AR) model: from the AR spectrum, band power was calculated in several frequency bands and the power sum was used as an independent variable [6–8]. In addition, the AR model coefficients or multivariate autoregressive (MVAR) model coefficients were used as features [9–12]. (3) Time–frequency (TF) analysis: Wang *et al* used the TF analysis as a useful tool for oscillatory EEG components during motor imagery [13, 14]. As we all know, oscillatory EEG components produced during motor imagery are both time and frequency related. So the method obtained promising results. However, oscillatory EEG components may cause simultaneous shifts in slow cortical potentials. A combination of two correlated signals might be used to increase the extracted information. The TF method considers only oscillatory EEG components. (4) Wavelet transform and wavelet packet transform (WPT): in [15], raw EEG signals were decomposed using wavelet transform and then event-related (de)synchronization patterns were extracted from symmetric electrode pairs. The weighted energy difference of the electrode pairs were used as features. In [16, 17], raw EEG signals are decomposed using the WPT and subband energies at the last decomposition level were used as features.

Due to the non-stationary property of EEG signals, traditional analysis methods such as the Fourier transform are not very suitable for this work. Wavelet transform and WPT use fast decaying kernel functions, which may better represent and analyze the signals. WPT can be viewed as a generalized version of the wavelet transform and it was also demonstrated to outperform other methods for the feature extraction of EEG signals [16, 17]. However, the wavelet packet decomposition includes multiple bases and different bases will result in different classification performances. The selection of the best basis which can provide the 'best' classification performance for a specified signal is very important.

The existing WPT method in BCIs uses only subband energies at the last decomposition level but does not depend on the best basis and so it cannot ensure the best classification performance. This paper discusses a feature extraction method based on wavelet packet best basis decomposition (WPBBD). The experimental results show that the WPBBD method yields significantly higher classification accuracy than that obtained by previous wavelet packet decomposition without best basis (WPNBBD).

2. Experimental data

Six healthy subjects (three male, three female; age range: 22–34 years; average age = 27 years) who had no experience of BCIs participated in the experiment. All the subjects were right handed. The subjects were seated in a shielded room with dim lighting. A 32-channel elastic electrode cap was used to record EEG. Measurements were made with reference to electrically linked mastoids, A1 and A2. The data were recorded at a sampling rate of 100 Hz with ESI-128, a product of NEUROSCAN Co., USA.

Each subject repeated the experiment for two sessions. Each session comprised 5 runs with 30 trials each resulting in a set of 150 trials. The subjects were asked to imagine performing one of the three motor imagery tasks (playing basketball using left hand, playing basketball using right hand, braking using right foot) in a self-paced mode during each trial. The number of trials for each task was equal. Each trial lasted 5.75–6.25 s (average 6 s) and consisted of three phases (shown in figure 1): (1) a 0.75–1.25 s (random) resting phase during which the computer screen was black; (2) a 1 s preparation phase during which a '+' fixation was displayed at the center of the computer screen; (3) a 4 s motor imagery task phase during which the subjects performed the corresponding motor imagery task according to the direction of the arrow (a left arrow indicating imagining left hand, a right arrow indicating imagining right hand, a down arrow indicating imagining right foot). The arrow was displayed during the first 1 s of the 4 s task phase and the computer screen was black during the other 3 s. The data during the last 4 s of each trial were used to perform off-line analysis.

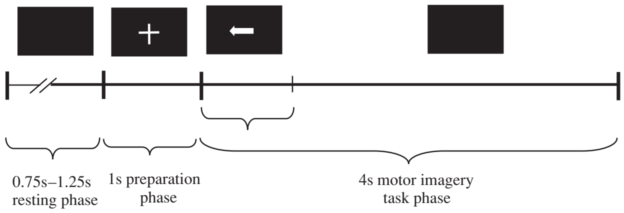

Figure 1. Three phases of a trial (a resting phase, a preparation phase and a task phase).

It should be pointed out that these skillful tasks (playing basketball and braking) may involve more activity in the supplementary motor area (SMA) related with the planning of movements, and these activities can make the classification of EEG signals during motor imagery tasks more difficult. Even so, we still adopt them considering the following reasons. (1) Almost everyone has had the experience of playing basketball and braking, so they are familiar with this motor imagery task and so they can imagine these movements easily. A task familiar to subjects can also shorten their training time. (2) These motor imagery tasks are interesting, so subjects can easily concentrate their attention to fulfill them. (3) Though these mental tasks need skills, they will become easy after subjects master them.

3. Method

3.1. Wavelet packet transform

The wavelet transform splits the original signal into two subspaces, **V** and **W**, which are orthonormally complementary to each other, with **V** being the space that includes the low-frequency information about the original signal and **W** including the high-frequency information. As shown in figure 2(a), we keep repeating the decomposition of the low-frequency subspace **V**. The wavelet transform partitions only the frequency axis finely toward the low frequency, and the WPT is a generalized version, which also decomposes the high-frequency bands that are kept intact in the wavelet transform. WPT leads to a complete wavelet packet tree, which is shown in figure 2(b), where $S(0,0)$ denotes the original signal space, $S(j,k)$ denotes the decomposed subspace, j is the decomposition level and k is the index of the subspace occurring at the jth level.

Let us denote $S(j,0) = V_j$ and $S(j,1) = W_j$; the WPT offers many alternative signal decompositions. Let $G = \{X_i, i = 1, 2, 3, \ldots\}$ be the set of all valid wavelet packet decompositions, such as $S(0,0)$, $S(1,0) \vee S(1,1)$, $S(2,0) \vee S(2,1) \vee S(1,1)$, etc. G constitutes the wavelet packet library and every valid decomposition X_i is called a wavelet packet basis. A valid basis X_i requires that the corresponding subspaces completely cover the entire horizontal 'range' without vertical overlap. The WPBBD selects the best basis X^* for the specified signal based on a certain criterion to provide the 'best' signal classification.

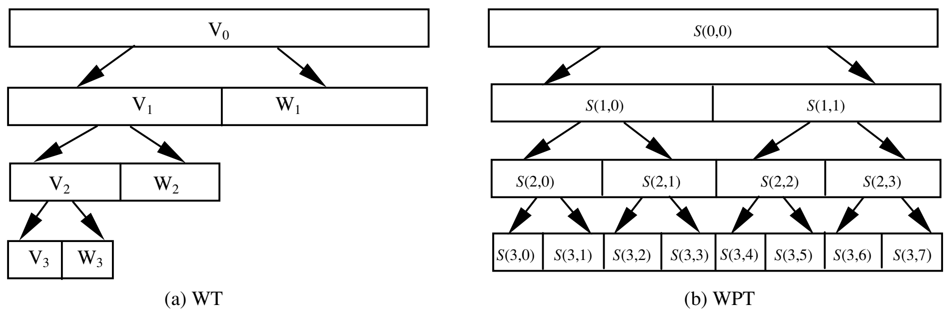

Figure 2. The structures of WT (a) and WPT (b).

3.2. The best basis algorithm

The best basis selection is based on a certain criterion which is used to measure the suitability of a wavelet packet basis. The criterion is purpose dependent. Although entropy is a good measure of information cost, it is suitable for measuring compression effect and may not reflect the classification ability of a wavelet very well [18]. For classification, a basis through which we can maximally separate different classes in the dimensional space is the best one. Therefore, the separability 'distance' among classes should be used to measure the efficiency of a basis. We adopt local discriminant basis (LDB) as a criterion, which is developed by Saito and Coifman [18]. The LDB selects from among the energy distributions of signal classes. The goal is to search for an optimal basis X^* among the wavelet packet library G with the maximum classification distance.

Formally, consider a c class classification problem in which $X = \{(x_k, \omega_k), k = 1, 2, \ldots, N\}$ is a set of N class-labeled training patterns where $x_k \in R^n$ and $\omega_k \in \Omega$, $\Omega = \{1, 2, \ldots, c, \ldots, C\}$. Let N_c be the number of signals belonging to class c, so that we have $N = N_1 + \cdots + N_C$. In the WPD, we denote $W_{j,k,l}\left(x_i^{(c)}\right)$ as the decomposition coefficients of class c signal $x_i^{(c)}$ at subspace $S(j,k)$, where l is the index of the location of decomposition coefficients. Suppose the dimension of signals is $n = 2^{n_0}$, then $l = 0, 1, \ldots, 2^{n_0-j} - 1$. The best basis X^* can be expressed as

$$X^* = \arg\max_{S(j,k)\in G} H(S(j,k)), \tag{1}$$

where

$$H(S(j,k)) \triangleq \sum_{c=1}^{C-1} \sum_{m=c+1}^{C} D(e^{(c)}(j,k), e^{(m)}(j,k)), \tag{2}$$

$$e^{(c)}(j,k) = [e^{(c)}(j,k,1); e^{(c)}(j,k,2) \\ ; \ldots ; e^{(c)}(j,k,2^{n_0-j} - 1)], \tag{3}$$

$$e^{(c)}(j,k,l) \triangleq \sum_{i=1}^{N_C} \left(W_{j,k,l}\left(x_i^c\right)\right)^2 \bigg/ \sum_{i=1}^{N_C} \|x_i^c\|^2. \tag{4}$$

$H(S(j,k))$ is called the discriminant power of the subspace $S(j,k)$, $e^{(c)}(j,k)$ is the normalized energy vector of class c and D is the Euclidean distance. The best subspace can be determined according to the value of $H(S(j,k))$.

3.3. The procedure of feature extraction

The WPBBD feature extraction can be performed according to the following steps.

(1) Select a wavelet function and specify the decomposition level.
(2) Select one of the EEG channels for analysis.
(3) Select a sample from a training set X.
(4) Decompose the sample to the specified level using the selected wavelet.
(5) Repeat steps 2–4 for all training samples.
(6) Calculate the discriminant power for each subband according to (2) and (3).
(7) Select the best basis X^* using the bottom-up search strategy.
(8) Calculate subband energies contained in the best basis X^* and select these energies as features.
(9) Repeat steps 2–8 for all channels.
(10) Combine the features from all EEG channels to form feature vector F.

It should be noted that the number of features the WPBBD selects depends on the number of subbands that constitute the selected best basis. In this paper, we mainly focus on the feature extraction. As for the classifier, we use the probabilistic neural network (PNN) as our classifier. By virtue of easy training and a solid statistical foundation in Bayesian estimation theory, the PNN has become an effective tool for solving many classification problems [19]. In the PNN, we take the value of the spread of the radial basis function as 1.0. The general flow chart of the WPBBD feature extraction method is shown in figure 3.

4. Results and discussion

4.1. Results

The aim of the analysis is to differentiate among three different motor imagery tasks described in section 2 and so to recognize the human's different intentions. Considering the practicality of BCI systems, we use a few electrodes (C3, C4, P3, P4, O1 and O2 electrodes over the primary sensorimotor cortex). It should be noted that these few electrodes are selected because our cognitive tasks mainly activate local cortical areas below these electrodes and so these few electrodes represent more useful information than other electrodes [20]. In addition,

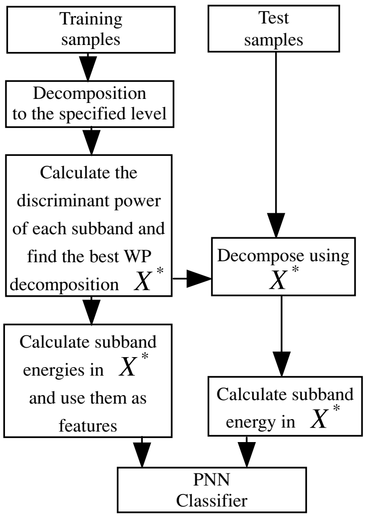

Figure 3. The general flow chart of the WPBBD feature extraction.

all the six subjects use the same set-up of electrodes though the classification accuracy can be increased when optimized electrode selection is applied to each subject. The reasons that the same set-up of electrodes is maintained for each subject in this paper are as follows: (1) the main purpose of this paper is to describe a feature extraction method; (2) we want to verify the generalization of the feature extraction method under the condition that all subjects use the same set-up of electrodes. The optimized electrode selection for each subject will be studied in the following work.

Meanwhile, in order to test the generalization of the established model, we establish a uniform model for all subjects although it may not be very reasonable. In this three-class classification problem, there are 1800 samples for all subjects (each class with 600 samples). We randomly select one third of samples from each class for training and the remaining for test. The following schemes are adopted.

(1) WPBBD: according to the feature extraction procedure described in section 3, we first use the typical 'db4' Daubechies wavelet to decompose the training set into five levels. The 100 Hz sample rate and five-level decomposition result in a 1.5625 Hz frequency resolution.
    Figure 4 shows the best basis X^* from each channel. The subband energies in X^* from all the six channels form the feature vector F. The dimension of F is (the numbers of subbands from C3, C4, P3, P4, O1 and O2 are 17, 9, 12, 12 and 16, separately) 66.
(2) WPNBBD: this method is introduced in [17]. Similar to the WPBBD, the difference between the WPBBD and the WPNBBD is that the latter does not adopt the subband energies contained in X^* but the energies contained in the last level. So, the dimension of the feature vector F is 192 (32 (features per channel) $\times$ 6 (channels)).
(3) AR model: in this paper, we use AR model coefficients as features. The linear AR model for a discrete process or signal $x(n)$ is then

$$x(n) = \sum_{k=1}^{N} a_k x(n-k) + e(n), \tag{5}$$

where N is the model order and $e(n)$ is the noise term. The coefficients a_k can be solved. The most serious disadvantage of the AR method is the problem of selecting the proper model order. There is no good way to determine the model order and so we take $N = 3, 4, 5, 6, 7$. Then, the corresponding number of features can be obtained from each EEG channel. The dimensions of the feature vector F are 18, 24, 30, 36 and 42 (six EEG channels).

The feature vector F is fed into the PNN classifier. The classification accuracies are obtained by ten cross-validations so that the classification accuracies become more stable. A number of well-known wavelets are considered in this study: Haar, Daubechies (db4, db6), Symlet (sym4, sym8). Figure 5 plots the average classification accuracies of the WPBBD and WPNBBD methods for test samples using different wavelet functions. The average values of classification accuracies obtained by all wavelets are 70.3% with the WPBBD and 66.1% with the WPNBBD. The variances of classification accuracies with the WPBBD and WPNBBD methods are 1.25% and 1.31%, respectively. The AR method obtains 58.9%, 57.5%, 61.4%, 62.6% and 60.8% classification accuracies under the above different model orders. The average classification accuracy of all model orders using the AR method is 60.2%.

4.2. Discussion

It can be seen from section 4.1 that the WPBBD, the WPNBBD and the AR model obtain 70.3%, 66.1% and 60.2% classification accuracies, respectively. The proposed WPBBD yields the best performance among the three methods. We can also see that the WPBBD obtains significantly higher classification accuracy than the WPNBBD method although the variances obtained by them are very close. In addition, the WPBBD uses fewer features than the WPNBBD.

In the present study, we have demonstrated that the WPBBD is an efficient way for the feature extraction of the motor imagery EEG signal. First, the WPT is an excellent signal analysis tool, especially for non-stationary signals. Due to the non-stationary property of EEG signals, the WPT is very appropriate to analyze the EEG signal, which has also been demonstrated in [16, 17]. Second, the WPT offers many alternative signal decompositions and includes multiple bases. Different bases can result in different performances. The WPBBD uses the best basis algorithm to search for the best basis which can produce the best performance. The best basis selection provides the basis for obtaining excellent classification results. The classification results comparing the WPBBD with the WPNBBD also prove the effectiveness of the best basis method.

Compared with the WPT, the FFT and the AR are used to analyze the stationary signals and so cannot obtain good performance when they are used to analyze the non-stationary signals, such as EEG. Our experimental results show that the WPT is superior to the AR model. In addition, the TF analysis is used to analyze the EEG signals during motor imagery in [13, 14]. In the TF method, raw EEG signals are filtered by the Laplacian method and then the filtered EEG

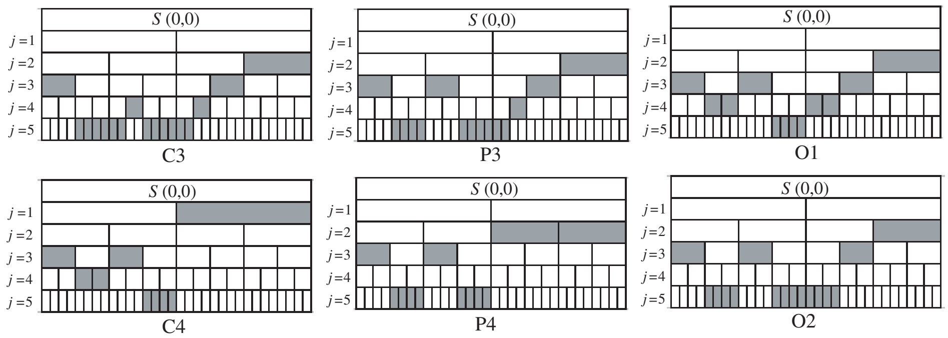

Figure 4. The best basis of each channel (the shadowed subbands represent the best basis).

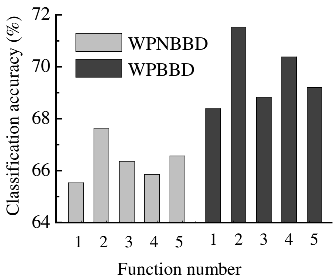

Figure 5. Classification accuracies obtained by different wavelet functions (1: Haar; 2: db4; 3: db6; 4: sym4; 5: sym8).

signals are decomposed into some band bins with band-pass filters. The corresponding envelopes of EEG in each band are down-sampled to form features. The TF method obtains a promising result due to the following reasons. (1) Spatial filters are used as a means of accentuating localized activity and reducing diffused activity, which is favorable for the feature extraction. (2) The combination of time domain information and frequency information can provide better classification performance than using any one of them. The WPT is a good time–frequency analysis tool and in our research we will continue to use the WPT as an analysis tool. In addition, we will adopt some advantages of other methods, such as the Laplacian filter technique used in the TF method. It should also be pointed out that the wavelet packet best basis varies between subjects and so the specific best basis for each subject can further improve the classification performance.

5. Conclusion

In this paper, the WPBBD feature extraction method for classifying EEG signals during the motor imagery tasks is investigated. The WPT yields a redundant representation of the signal and its over-complete structure provides the flexibility for features' representation to achieve better accuracy. We adopt the distance criterion to select the best basis for EEG signals. A signal can be better represented with the best basis than without the best basis. The experimental results show that the WPBBD outperforms the previous WPNBBD method. The WPBBD provides a more suitable feature extraction method for EEG-based BCIs. We will continue to study the best basis fitted for each subject separately, which is expected to further improve the classification accuracy.

References

[1] Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 Brain–computer interfaces for communication and control *Clin. Neurophysiol.* **113** 767–91

[2] Millán J d R 2003 Adaptive brain interfaces *Commun. ACM* **46** 74–80

[3] Varsta M, Heikkonen J, Millán JdR and Mouriño J 2000 Evaluating the performance of three feature sets for brain–computer interfaces with an early stopping MLP committee *15th Int. Conf. on Pattern Recognition* **2** 907–10

[4] Mark P and Aleksandar K 1998 Feature extraction in development of brain–computer interface: a case study *Proc. 20th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society* **20** 2058–61

[5] Mensh B D, Werfel J and Seung S H 2004 BCI Competition 2003—Data Set Ia: combining gamma-band power with slow cortical potentials to improve single-trial classification of electroencephalographic signals *IEEE Trans. Biomed. Eng.* **51** 1052–6

[6] McFarland D J, Lefkowicz A T and Wolpaw J R 1997 Design and operation of an EEG-based brain–computer interface with digital signal processing technology *Behav. Res. Methods Instrum. Comput.* **29** 337–45

[7] McFarland D J, McCane L M, David S V and Wolpaw J R 1997 Spatial filter selection for EEG-based communication *Electroencephalogr. Clin. Neurophysiol.* **103** 386–94

[8] McFarland D J, Miner L A, Vaughan T M and Wolpaw J R 2000 Mu and beta rhythm topographies during motor imagery and actual movements *Brain Topogr.* **12** 177–86

[9] Burke D P, Kelly S P, de Chazal P, Reilly R B and Finucane C 2005 A parametric feature extraction and classification strategy for brain–computer interfacing *IEEE Trans. Neural Syst. Rehabil. Eng.* **13** 12–7

[10] Pei X M and Zheng C X 2004 Feature extraction and classification of brain motor imagery task based on MVAR model *Proc. 3rd Int. Conf. on Machine Learning and Cybernetics (Shanghai, China)* **6** 3726–30

[11] Keirn Z A and Aunon J I 1990 A new mode of communication between man and his surroundings *IEEE Trans. Biomed. Eng.* **37** 1209–14

[12] Anderson C W, Stolz E A and Shamsunder S 1998 Multivariate autoregressive models for classification of spontaneous electroencephalographic signals *IEEE Trans. Biomed. Eng.* **45** 277–86

[13] Wang T and He B 2004 An efficient rhythmic component expression and weighting synthesis strategy for classifying

motor imagery EEG in a brain–computer interface *J. Neural Eng.* **1** 1–7

[14] Wang T, Deng J and He B 2004 Classifying EEG-based motor imagery tasks by means of time–frequency synthesized spatial patterns *Clin. Neurophysiol.* **115** 2744–53

[15] Qin L and He B 2005 A wavelet-based time–frequency analysis approach for classification of motor imagery for brain–computer interface applications *J. Neural Eng.* **2** 65–72

[16] Graimann B, Huggins J E and Levine S P 2004 Toward a direct brain interface based on human subdural recordings and wavelet-packet analysis *IEEE Trans. Biomed. Eng.* **51** 954–62

[17] Xue J Z, Zhang H, Zheng C X and Yan X G 2003 Wavelet packet transform for feature extraction of EEG during mental tasks *Proc. 2nd Int. Conf. on Machine Learning and Cybernetics (Xi'an, China)* **1** 360–3

[18] Saito N and Coifman R R 1995 Local discriminant bases and their applications *J. Math Vis. Imaging* **5** 337–58

[19] Mao K Z, Tan K C and Ser W 2000 Probabilistic neural-network structure determination for pattern classification *IEEE Trans. Neural Netw.* **11** 1010–7

[20] Millán JdR, Mouriño J, Franzé M, Cincotti F and Varsta M 2002 A local neural classifier for the recognition of EEG patterns associated to mental tasks *IEEE Trans. Neural Netw.* **13** 678–86