Towards a Framework Based on Single Trial Connectivity for Enhancing Knowledge Discovery in BCI

Martin Billinger[1], Clemens Brunner[1], Reinhold Scherer[1], Andreas Holzinger[2], and Gernot R. Müller-Putz[1]

[1] Institute for Knowledge Discovery, BCI Lab, Graz University of Technology, Austria
[2] Research Unit Human-Computer Interaction, Institute for Medical Informatics, Statistics and Documentation, Medical University Graz, Austria
reinhold.scherer@tugraz.at

Abstract. We developed a framework for systematic evaluation of brain-computer interface (BCI) systems. This framework is intended to compare features extracted from a variety of spectral measures related to functional connectivity, effective connectivity, or instantaneous power. Different measures are treated in a consistent manner, allowing fair comparison within a repeated measures design. We applied the framework to BCI data from 14 subjects recorded on two days each, and demonstrated the framework's feasibility by confirming results from the literature. Furthermore, we could show that electrode selection becomes more focal in the second BCI session, but classification accuracy stays unchanged.

Keywords: brain-computer interface, band power, coherence, directed transfer function, connectivity, knowledge discovery.

1 Introduction

Diseases such as amyotrophic lateral sclerosis (ALS) or cerebral palsy (CP) disturb nervous system functions, limiting affected individuals in their abilities to interact with their environment. Voluntary movement or communication can be difficult or even impossible due to loss or impairment of motor functions. A brain-computer interface (BCI) could potentially help to improve quality of life for such individuals by allowing them to communicate or interact with their environment without relying on motor functions [21]. BCIs measure a user's brain signals and translate them into control commands for applications [10, 14].

A commonly used brain signal for BCI control is the electroencephalogram (EEG) [16], which records cortical electrical activity from the scalp. The EEG contains several typical rhythmic activities such as the sensorimotor rhythm (SMR). Individuals can modulate SMR by motor imagery (MI), that is, imagining movement of their extremities, which causes power changes in specific frequency bands. These changes occur in different cortical areas, depending on which movement is imagined [15].

The state-of-the-art method for detecting different MI patterns is classification of band power (BP) features [3]. BP is the instantaneous power in pre-defined

R. Huang et al. (Eds.): AMT 2012, LNCS 7669, pp. 658–667, 2012.

frequency bands of single EEG channels. Although BP features allow a BCI to reliably detect different MI tasks [16], no information about the interaction of different brain areas can be obtained. However, such information could provide more insight in the neurological processes involved, and might improve classification of MI tasks.

The interaction of brain areas is expressed in functional or effective connectivity [6]. While functional connectivity measures correlated activation of brain areas, effective connectivity measures causal information flow. Numerous connectivity measures have been developed and proposed in BCI-related studies [4, 9, 12, 19]. Since these studies vary greatly in procedure, comparing their results is not easily possible. Hence, there is a clear need for a unified procedure to systematically compare feature extraction methods based on different measures of connectivity.

In this article, we present a framework for systematically evaluating BCI systems. This framework supports many different spectral connectivity measures, as well as non-connectivity measures such as BP. This allows fair comparison within a broad range of methods.

Although connectivity measures provide only three dimensional data (source, destination, frequency), they lead to a high dimensional feature space for classifying MI tasks. The number of features usually greatly outnumbers the number of available training samples. Thus, our framework also performs channel and frequency band selection based on statistical significance. This reduces the number of features and helps the researcher to interpret brain interactions relevant for BCI use [7, 8].

2 Methods

2.1 Connectivity Measures

Different measures related to connectivity can be derived from a vector autoregressive model (VAR) representation of the observed signals. The number of free parameters in a VAR model is M^2p, which limits the minimum estimation window length. The number of channels M, model order p, and window length N determine spatial, spectral and temporal resolution of derived measures.

Time resolution of VAR models can be increased by obtaining multiple realizations of the observed time series. This is a common procedure in neurophysiological studies, where repeated trials of a task are performed [13]. Unfortunately, MI BCIs are required to respond to single trials. Thus, a careful trade-off between temporal, spectral and spatial resolutions must be found. While we limit spatial resolution with automatic channel selection, model order and estimation window length are set manually.

2.2 BCI Framework

Fig. 1 shows an overview of the BCI framework. The outer cross-validation loop serves to test the framework as a whole. It resembles the flow of typical BCI

operation, where parameters selected from available data are applied to novel data. The optimal time segment for classifier training during MI is determined in the inner cross-validation loop. A detailed description of each functional unit follows below.

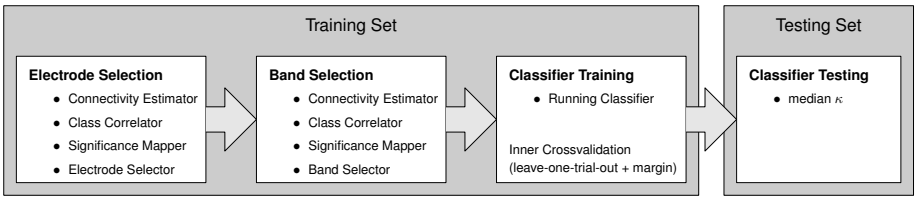

Fig. 1. Framework for the BCI based on functional or effective connectivity. The framework resembles the work-flow of a typical BCI, where the system is optimized on available data and applied to novel data, possibly in an online setting. This work-flow is embedded in a cross-validation procedure, where parts of the pre-recorded data serve as the novel testing set.

Connectivity Estimator. The Connectivity Estimator fits a sliding window VAR model to multiple EEG channels and subsequently extracts any required connectivity measures. For details about retrieving connectivity measures from VAR models see [18].

This results in a four-dimensional data set $c_k(i, j, f, t)$ for each trial k, which describes a measure of connectivity from channel j to i at frequency f and time t relative to the cue. BP can be estimated from the EEG's autospectra obtained by discrete Fourier transform (DFT). In this case $c_k(i, j, f, t) = 0$ for $i \neq j$.

Class Correlator. The correlation coefficient $r(i, j, f, t)$ is calculated between c_k and corresponding class labels y_k. $p(i, j, f, t)$-values are obtained from the asymptotic normal distribution, which are used for testing against the null-hypothesis of no correlation.

Significance Mapper. A modified version of the direct estimation of false discovery rate (FDR) [20] is used to determine which correlations are significant at a specified FDR. Similar to [1], instead of accepting all hypotheses with p-values below a certain threshold, only those p-values are accepted that form clusters of a minimum size in the t/f-plane.

Electrode Selector. Electrode selection is performed in two steps, with a ranking criteria e_l based on the squared correlation coefficient:

$$e_l = \sum_f \sum_t \left(\sum_i r^2(i, l, f, t) + \sum_j r^2(l, j, f, t) \right) \tag{1}$$

Electrodes l with the highest e_l are selected. First, only significant r are considered. If the procedure cannot find a pre-defined number of electrodes, more electrodes are added to the selection based on all r.

Band Selector. Band selection identifies a list of all channel pair frequency bands, in which the measure c_k correlates significantly with the class labels. Bands are identified for each channel pair (i, j) as the frequency ranges in which there is significantly correlated activity of at least two seconds. For functional connectivity measures, where $c_k(i, j, f, t) = c_k(j, i, f, t)$ or $c_k(i, j, f, t) = -c_k(j, i, f, t)$, only one of the channel pairs (i, j) and (j, i) is used.

If no frequency bands are found, the system chooses a set of default bands, which were defined as the frequency range of 5 to 25 Hz for each channel pair.

Classification. First, the optimal time segment for training the linear discriminant analysis (LDA) classifier is determined. This is accomplished by estimating the classification accuracy for each time segment with inner cross–validation. In a leave-one-out procedure, each trial is used as the validation set once and a margin of 10 trials before and after the validation trial is excluded from the training set [11].

Subsequently, the classifier is trained on the full training set in the time segment with the highest accuracy. Testing is performed by applying the classifier to the testing set. This is the only place in the framework where the testing set is used. Thus, the classifier is tested on completely unseen data.

2.3 Data Acquisition

Three electrooculogram (EOG) and 45 EEG channels were recorded. Electrodes were placed according to the international 10–20 System. Fig. 2 (left) shows the exact electrode positions. Signals were recorded at a sampling rate of 300 Hz using three synchronized g.USBamp amplifiers (g.tec, Guger Technologies OEG, Graz, Austria) with passive Ag/AgCl ring electrodes and filtered between 0.5 and 100 Hz. The notch filter was set to 50 Hz to suppress line noise.

14 healthy volunteers without prior experience in BCI control participated in a BCI experiment. On each of two separate days (sessions) six training runs and three feedback runs were performed. During feedback, the participants were instructed to control a virtual plane along a path with two different MI tasks (for details, see [2]). The training paradigm was based on the synchronous Graz BCI training paradigm [16], modified to visually resemble the continuous feedback paradigm. At the beginning of each session, an artifact run was performed to estimate the influence of artifacts such as eye movement and eye blinks on the EEG. Subsequently, four training runs were followed by one feedback run, two further training runs, and two final feedback runs. In this article, we only use the data recorded during training.

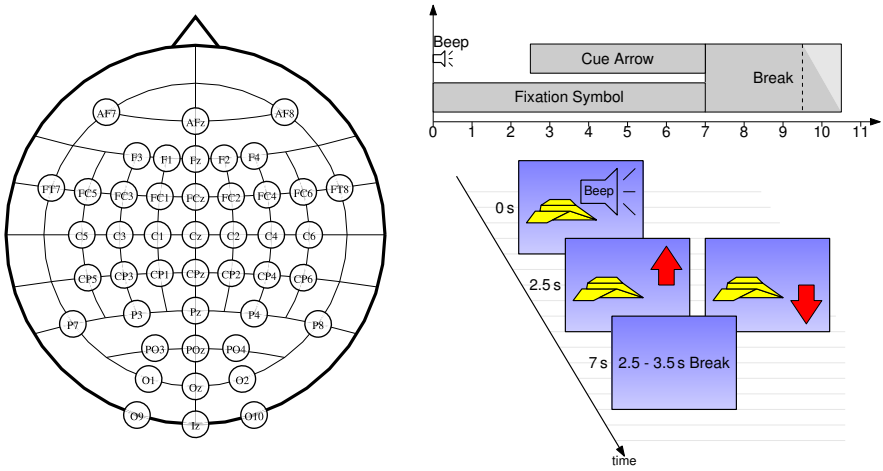

Fig. 2. Left: EEG electrode placement. Right: experimental paradigm.

From each session a total of 90 trials of right hand MI and 90 trials of feet MI are available. Fig. 2 (right) shows the timing of a trial. Trial duration is 7 s, followed by a break of 3 ± 0.5 s. Trial start was indicated by an acoustic beep and appearance of the fixation object. At $t = 2.5$ s an arrow appeared, pointing up or down, to cue the participant to perform hand or feet MI, respectively.

2.4 Data Analysis

Prior to application of the framework to the data, some pre-processing steps were performed. The EOG was removed from the EEG using a regression based approach [17]. Subsequently, the EEG data was down-sampled from 300 to 100 Hz.

The VAR model order was chosen as $p = 9$. This allows reasonable frequency resolution with reasonable window lengths. Window length was set to 3 s prior to electrode selection, and to 1.5 s after electrode selection. The number of electrodes selected was 10.

The framework was applied to the comparison of BP, coherence (COH), and directed transfer function (DTF) features. BP is known to work well with this kind of task and serves as the reference method. COH is a measure for functional connectivity, and DTF is a measure for effective connectivity. Each measure was independently applied to the electrode selection and the classification step.

The electrode selection step is evaluated using the concept of entropy. Entropy is a measure of uncertainty in a probability distribution, defined in (2), where p_l is the probability that each electrode l is selected from the set of all electrodes L.

$$E = -\sum_{l \in L} p_l \log_2(p_l) \tag{2}$$

The probabilities p_l are estimated from the electrode selection step in the outer cross-validation. An entropy of $E = 0$ corresponds to perfectly consistent selection, where always the same set of electrodes is selected. Increasing values of E indicate increasing uncertainty.

Classification performance is measured with Cohen's Kappa κ. To obtain a robust measure, κ is estimated for each time segment between cue and end of trial, and the median value of κ is reported as classification performance.

The results of electrode selection were analyzed using a repeated measures analysis of variance (ANOVA), with the dependent variable *entropy*, and the factors *selection* (measure used for selection) and *session*. Similiarly, classification performance was analyzed with the dependent variable κ, and the factors *selection*, *method* (measure used for classification) and *session*. Sphericity corrections were applied when required. Factors found to be significant by the ANOVAs were subject to paired t-tests with Holm-Bonferroni correction.

3 Results

Table 1 lists the ANOVA results for electrode selection. Factors *selection* and *session* are both highly significant ($p < 0.01$), and there are no significant interaction effects between factors. Fig. 3 shows the results of post tests on both significant factors. BP has the lowest selection entropy of all methods. Also, entropy is significantly lower in the second session. Fig. 4 shows an example of electrode selection from both sessions for one subject. In the first session, the distribution of selection probabilities is more spread out than in the second session, which is indicated by a higher entropy value.

Table 1. ANOVA results for electrode selection. d_1 and d_2 are the between-group and within-group degrees of freedom of the F-statistic F.

Effect	d_1	d_2	F	p	sig
selection	2	26	13.956	0.00008	**
session	1	13	9.292	0.00933	**
selection:session	2	26	2.169	0.13458	

Table 2 lists the ANOVA results for classification. Factor selection is significant ($p < 0.05$), and factor method is highly significant ($p < 0.01$). All other main and interaction effects are not significant. Fig. 5 shows the results of post tests on both significant factors. Electrode selection with DTF leads to the highest classification performance, while classification with DTF is significantly worst. Differences between BP and COH are not significant.

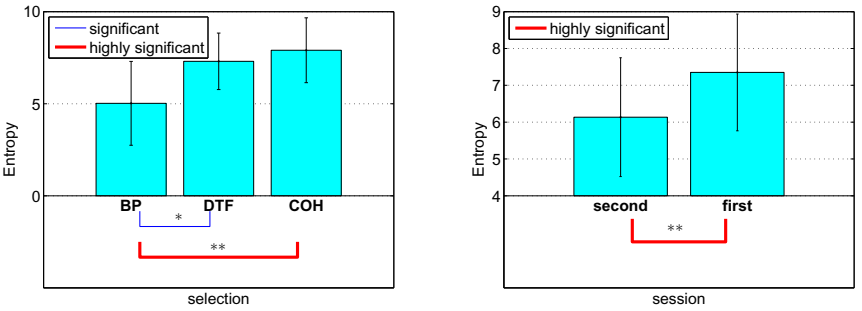

Fig. 3. Entropy of electrode selection. Left: comparison of connectivity measures (averaged over both sessions). Right: comparison of first and second session (averaged over all selection methods). The brackets below the bar charts indicate if differences are significant ($p < 0.05$, *) or highly significant ($p < 0.01$, **).

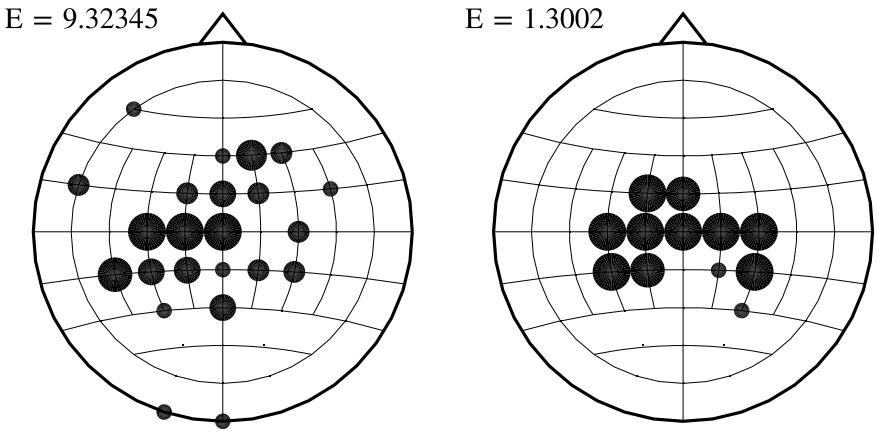

Fig. 4. Exemplary electrode selection probability for subject BV3, using COH. Left: first session. Right: second session. Bigger circles indicate higher probability of an electrode to be selected.

Table 2. ANOVA results for classification performance. The column p[GG] contains Greenhouse-Geisser corrected p-values for factors that violate sphericity assumptions.

Effect	DFn	DFd	F	p	p[GG]	sig
selection	2	26	5.386	0.01104		*
method	2	26	17.980	0.00001	0.00015	**
session	1	13	0.758	0.39964		
selection:*method*	4	52	0.534	0.71136		
selection:*session*	2	26	0.458	0.63760		
method:*session*	2	26	1.130	0.33849		
selection:*method*:*session*	4	52	0.549	0.70077		

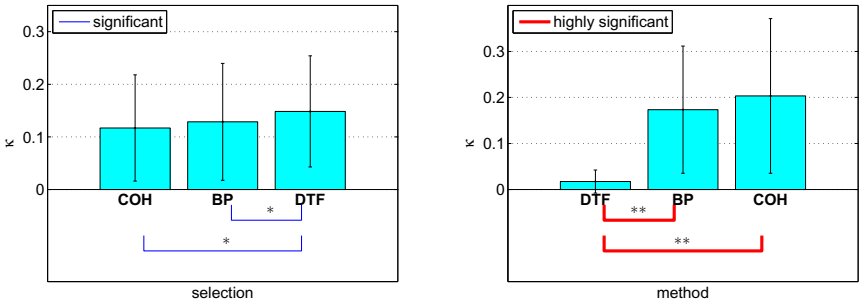

Fig. 5. Classification performance (Cohen's Kappa). Left: comparison of connectivity measures used for electrode selection. Right comparison of connectivity measures used for classification. The brackets below are explained in Fig. 3.

4 Discussion

We proposed a framework for simulating offline BCIs based on spectral features. Features are selected from connectivity or band power measures. To avoid overfitting, the whole work flow is embedded in nested cross-validation procedures. The framework is designed to be modular. Thus, components can easily be replaced or added. For example, the simple LDA could be replaced by more sophisticated classifiers. Additional connectivity measures or pre-processing steps can be added with little effort.

Our simulation results show that BP and COH work equally well for classification. This finding is in line with [9], who argue that this is caused by a bias towards zero-phase between electrodes. Although DTF provides the worst features for classification, it gives the best electrode selection in terms of classification performance. This is an interesting finding that will require further investigation.

The issue of zero-phase bias is inherent to autoregressive models, since instantaneous terms are not directly modeled. This is addressed in [5], who propose an extension to VAR models to include an instantaneous term. An alternative could be to use independent component analysis (ICA) for pre-processing, which maximizes independence between components, and effectively removes zero-phase components.

Electrode selection entropy is lower in the second session. This could indicate a training effect induced by the feedback. Activation becomes more focused as the subjects get accustomed to the task. However, this effect has no influence on classification accuracy.

The framework helps researchers to manage their knowledge at various levels. On the methodological level, any number of measures, each providing a high dimensional feature space, are reduced to their respective classification performance. Statistical comparison allows to easily comprehend their suitability for BCI use (Figs. 3, 5). On the level of individual measures, relevant electrodes are

identified from the multitude of possible connectivities between all channel pairs in all frequency bins (Fig. 4). Furthermore, the subset of actual connectivities can be further analysed, deepening the understanding of brain connectivity in respect to specific tasks.

5 Conclusions

We could demonstrate that the proposed framework for BCI evaluation works correctly by confirming some of the results of [9]. Thus, the framework could be used to determine the best set of methods and electrodes to be used in individuals or patient groups.

The DTF, a measure of effective connectivity, appears to be useful for electrode selection. This indicates that connectivity measures can provide useful information for BCIs. However, to realize a BCI based on connectivity features, more work is required to address the issue of zero-phase bias between EEG channels.

Acknowledgments. This work was supported by the FP7 Framework EU Research Project ABC (No. 287774) and the FWF Project Coupling Measures for BCIs (P20848-N15). This paper only reflects the authors' views and funding agencies are not liable for any use that may be made of the information contained herein.

References

1. Billinger, M., Kaiser, V., Neuper, C., Brunner, C.: Automatic frequency band selection for BCIs with ERDS difference maps. In: Proceedings of the 5th International Brain–Computer Interface Conference (2011)
2. Billinger, M., Neuper, C., Müller-Putz, G.R., Brunner, C.: User-centric performance estimation in a continuous online BCI. In: Proceedings of the 3rd TOBI Workshop (2012)
3. Brunner, C., Billinger, M., Vidaurre, C., Neuper, C.: A comparison of univariate, vector, bilinear autoregressive, and band power features for brain computer interfaces. Medical and Biological Engineering and Computing 49, 1337–1346 (2011), http://dx.doi.org/10.1007/s11517-011-0828-x, doi:10.1007/s11517-011-0828-x
4. Brunner, C., Scherer, R., Graimann, B., Supp, G., Pfurtscheller, G.: Online control of a brain-computer interface using phase synchronization. IEEE Transactions on Biomedical Engineering 53, 2501–2506 (2006)
5. Erla, S., Faes, L., Tranquillini, E., Orrico, D., Nollo, G.: Multivariate autoregressive model with instantaneous effects to improve brain connectivity estimation. International Journal of Bioelectromagnetism 11(2), 74–79 (2009)
6. Friston, K.J.: Functional and effective connectivity in neuroimaging: A synthesis. Hum. Brain Mapping 2, 56–78 (1994)

7. Holzinger, A.: On knowledge discovery and interactive intelligent visualization of biomedical data - challenges in human-computer interaction & biomedical informatics. In: Proceedings of the 9th International Joint Conference on e-Business and Telecommunications (ICETE 2012), Rome, Italy, pp. IS9–IS20 (2012)
8. Holzinger, A., Scherer, R., Seeber, M., Wagner, J., Müller-Putz, G.: Computational Sensemaking on Examples of Knowledge Discovery from Neuroscience Data: Towards Enhancing Stroke Rehabilitation. In: Böhm, C., Khuri, S., Lhotská, L., Renda, M.E. (eds.) ITBAM 2012. LNCS, vol. 7451, pp. 166–168. Springer, Heidelberg (2012)
9. Krusienski, D.J., McFarland, D.J., Wolpaw, J.R.: Value of amplitude, phase, and coherence features for a sensorimotor rhythm-based brain computer interface. Brain Research Bulletin 87(1), 130–134 (2012)
10. Kübler, A., Furdea, A., Halder, S., Hammer, E.M., Nijboer, F., Kotchoubey, B.: A brain-computer interface controlled auditory event-related potential (P300) spelling system for locked-in patients. Annals of the New York Acadamy of Sciences 1157, 90–100 (2009)
11. Lemm, S., Blankertz, B., Dickhaus, T., Müller, K.R.: Introduction to machine learning for brain imaging. NeuroImage 56(2), 387–399 (2011), http://www.sciencedirect.com/science/article/pii/S1053811910014163
12. Lim, J.H., Hwang, H.J., Jung, Y.J., Im, C.H.: Feature extraction for brain–computer interface (BCI) based on the functional causality analysis of brain signals. In: Proceedings of the 5th International Brain–Computer Interface Conference (2011)
13. Möller, E., Schack, B., Vath, N., Witte, H.: Fitting of one ARMA model to multiple trials increases the time resolution of instantaneous coherence. Biological Cybernetics 89, 303–312 (2003)
14. Müller-Putz, G.R., Scherer, R., Pfurtscheller, G., Rupp, R.: Brain-computer interfaces for control of neuroprostheses: from synchronous to asynchronous mode of operation. Biomedizinische Technik 51, 57–63 (2006)
15. Neuper, C., Pfurtscheller, G.: Event-related dynamics of cortical rhythms: frequency-specific features and functional correlates. International Journal of Psychophysiology 43, 41–58 (2001)
16. Pfurtscheller, G., Neuper, C.: Motor imagery and direct brain-computer communication. Proceedings of the IEEE 89, 1123–1134 (2001)
17. Schlögl, A., Keinrath, C., Zimmermann, D., Scherer, R., Leeb, R., Pfurtscheller, G.: A fully automated correction method of EOG artifacts in EEG recordings. Clinical Neurophysiology 118, 98–104 (2007)
18. Schlögl, A., Supp, G.: Analyzing event-related EEG data with multivariate autoregressive parameters. In: Neuper, C., Klimesch, W. (eds.) Event-related Dynamics of Brain Oscillations, pp. 135–147. Elsevier (2006)
19. Shoker, L., Sanei, S., Sumich, A.: Distinguishing between left and right finger movement from eeg using svm. In: 27th Annual International Conference on Engineering in Medicine and Biology Society, IEEE-EMBS 2005, pp. 5420–5423 (January 2005)
20. Storey, J.D.: A direct approach to false discovery rates. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64(3), 479–498 (2002), http://dx.doi.org/10.1111/1467-9868.00346
21. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain-computer interfaces for communication and control. Clinical Neurophysiology 113, 767–791 (2002)