Comparison of Designs Towards a Subject-Independent Brain-Computer Interface based on Motor Imagery

Fabien Lotte, Cuntai Guan, and Kai Keng Ang

Abstract—A major limitation of current Brain-Computer Interfaces (BCI) based on Motor Imagery (MI) is that they are subject-specific BCI, which require data recording and system training for each new user. This process is time consuming and inconvenient, especially for casual users or portable BCI with limited computational resources. In this paper, we explore the design of a Subject-Independent (SI) MI-based BCI, i.e., a BCI that can be used immediately by any new user without training the BCI with the user's data. This is achieved by training the BCI on data acquired from several other subjects. In order to assess the possibility to build such a BCI, we compared several designs based on different features and classifiers, on data from 9 subjects. Our results suggested that linear classifiers were the most appropriate for the design of MI-based SI-BCI. We also proposed a filter bank common spatial patterns feature extraction method based on a multi-resolution frequency decomposition which achieved the highest accuracy.

I. INTRODUCTION

Brain-Computer Interfaces (BCI) are communication systems which enable users to send commands to computers by using only their brain activity, this activity being generally measured by ElectroEncephaloGraphy (EEG) [1]. BCI have been revealed as a very promising tool for disabled people [1] as well as for healthy people, e.g., for video games [2]. However, a major limitation of EEG-based BCI systems is that they must be tuned for each new user [1]. This tuning requires a lengthy and tedious recording of EEG signals from the new user, hence preventing immediate use of the BCI. Ideally, a new user should be able to use the BCI from the very first time, without training (neither machine nor human training). Hence, a Subject-Independent (SI) BCI is highly desirable.

Recently, Lu and Guan have proposed a method to design a P300-based SI-BCI [3]. Their results suggested that designing a P300-based SI-BCI was possible as common features of the P300 were shared by different subjects. Whether these findings still hold for other brain signals such as motor imagery (limb movement imagination) is still an open question. Moreover, the features used in P300-based BCI are generally the same for all subjects, more precisely the time course of the EEG signal amplitude. For BCI based on Motor Imagery (MI), the features are generally Subject-Specific (SS) spectral and/or spatial information [4][5]. This

Fabien Lotte, Cuntai Guan and Kai Keng Ang are with A * STAR, Institute for Infocomm Research (I2R), Signal Processing Department, Brain-Computer Interface Laboratory, 1 Fusionopolis way, #21-01 Connexis (South Tower), Singapore 138632 {fprlotte, ctguan, kkang}@i2r.a-star.edu.sg

makes the design of a SI-BCI based on MI an even more challenging task.

This paper explores the design of a SI-BCI based on MI by training the system on data from multiple users. In this study, we compared several designs using different features and classifiers. The remainder of this paper is organized as follows: Section II describes the different BCI designs that were studied and compared. Then, Section III presents the evaluations conducted and Section IV concludes this paper.

II. METHODS

Designing a SI-BCI is rather different from designing a SS-BCI. Indeed, methods that are efficient for SS-BCI design may have poor performances for SI-BCI, as they carry mostly SS information and very few SI information. Thus, we believe that comparing various kind of features and classifiers is necessary in order to gather knowledge about the design of a SI-BCI based on MI. This section describes the features and classifiers studied as well as the design of a BCI using only a-priori knowledge and no machine learning, used as a baseline for comparison.

A. Feature extraction methods

So far, a variety of feature extraction methods have been used to design BCI [6]. In this paper, we study and compare several of them, belonging to the most popular and efficient ones: logarithmic band power, autoregressive coefficients, power spectral densities, common spatial patterns and filter bank common spatial patterns.

Logarithmic Band-Power (BP): Computing a BP feature consists in band-pass filtering the signal in a given frequency band, squaring it, averaging it over a given time window and finally log-transforming it [4]. Here we extracted band power features in the mu (8-13 Hz) and beta (13-30 Hz) frequency bands, as the power in these two bands is known to vary according to the MI task performed [4]. To enhance the results, the EEG signals were re-referenced to Common Average Reference (CAR) [1] before feature extraction.

AutoRegressive (AR) coefficients: AR methods assume that a signal can be modeled as a weighted sum of the values of this signal at previous time steps. The weights in this weighted sum are the AR coefficients used as features for the BCI [4][6]. In this study we used a 4th order AR modelling based on Burg method. Before AR feature extraction, the signals were spatially filtered using a Surface Laplacian [1].

Power Spectral Densities (PSD): PSD features inform on the distribution of the power of a signal between the different frequencies [6]. Here, we computed the PSD values

using the Welch periodogram method and used as features the PSD values in each 2 Hz frequency band within the 8-30 Hz band, without overlap between consecutive bands. Before feature extraction, the signals were re-referenced using CAR.

Common Spatial Patterns (CSP): CSP is based on the decomposition of the EEG signals into spatial patterns [7]. Such patterns are selected in order to maximize the differences between the classes involved once the data have been projected onto these patterns. It is achieved by jointly diagonalizing the covariance matrices of the EEG signals from each class [7]. CSP have proven to be a very efficient feature extraction method [7]. In this study, we classically filtered the EEG signals in 8-30 Hz before using the CSP and we kept only the 3 first pairs of filters [7].

Filter Bank Common Spatial Patterns (FBCSP): The FBCSP is an extension of the classical CSP [5]. FBCSP first decomposes the EEG signals into different frequency bands using a filter bank. Then, a set of CSP filters is created for each of these frequency bands. Finally, a feature selection algorithm is used to determine the most discriminant filters hence created. FBCSP has shown its superiority over other methods during the last BCI competition [8]. In this study we used a filter bank which decomposes the signals into 9 bands of 4Hz (4-8 Hz, 8-12 Hz, ..., 36-40 Hz), created 2 pairs of CSP filter for each frequency band, and selected the 4 most relevant filter pairs using the Mutual Information Best Individual Feature Selection (MIBIFS) algorithm [5]. These settings were reported to be the most optimal ones [5][8].

B. Classifiers

Many classifiers can be used to classify EEG features [9][10]. Among them, we compared the three following algorithms: linear discriminant analysis, quadratic discriminant analysis and Gaussian mixture model.

Linear Discriminant Analysis (LDA): LDA uses a linear hyperplane to separate data from two classes [9]. To compute the best discriminant hyperplane, LDA assumes a Gaussian distribution of the data, with equal covariance matrices for both classes. When using LDA, the class assigned to an unseen feature vector depends on which side of the hyperplane the vector is.

Quadratic Discriminant Analysis (QDA): Like LDA, QDA assumes a Gaussian distribution of the data to identify a decision boundaries between two classes. However, unlike LDA, QDA does not assume equal covariance matrices for both classes, which leads to quadratic decision boundaries [10].

Gaussian Mixture Model (GMM): A GMM classifier assumes that the data of each class are modeled by a mixture of several Gaussians [10]. GMM classifies an unseen feature vector using Bayes' rule. To identify the mixture of Gaussians, a first estimate of the Gaussian location was obtained using K-means clustering, and then optimized using the Expectation-Maximization (EM) algorithm [10]. We used 3 Gaussians to model each class.

The reason for using such classifiers is that they provide various degree of non-linearity. Indeed, LDA is a linear clas-

sifier, QDA is non-linear but provides only quadratic decision boundaries whereas GMM is completely non-linear and can provide arbitraly complex non-linear decision boundaries. These properties will enable us to study the impact of non-linearity on the SI-BCI performances.

C. A-priori-based BCI

From the neurophysiology literature we know that imagining a left-hand movement leads to a decrease of power in the mu and beta bands over the right motor cortex, and vice-versa for an imagined movement of the right hand. As a baseline for comparisons, we also designed a BCI using no machine learning but only a-priori knowledge on MI as described above. To do so, we first computed the logarithmic BP in 8-30 Hz (i.e., mu + beta) for the electrodes CP3, CP4, C5, C3, C1, C2, C4, C6, FP3, FP4 (i.e., the electrodes located over the motor cortices). Then, we averaged these BP values over electrodes CP3, C5, C3, C1 and FP3 (left motor cortex) on one hand, and over CP4, C2, C4, C6 and FP4 (right motor cortex) on the other hand, hence leading to two different features. The sign of the difference between these two features determines the class of the input signal. Indeed, the difference between these two features indicates the motor cortex in which the mu and beta power has decreased.

III. EVALUATION

A. EEG data set used

We evaluated the different designs on data set 2a from BCI competition IV [8]. This data set comprises EEG signals from 9 subjects who performed left hand, right hand, foot and tongue MI. For the purpose of this study, only EEG signals corresponding to left and right hand MI were used. Indeed, multiclass classification for BCI is another problem, not in the scope of this paper. This data set comprised a training set and a testing set for each subject. Both sets contained 72 trials for each class. Each trial had a duration of 7 sec. Subjects performed MI from time t = 3 sec to t = 7 sec of each trial.

In this work, we only considered the discrete classification of the trials, i.e., we assigned a class to each trial and not to each trial sample. In order to do so, we extracted the features from the time segment t = 3.5 sec to t = 5.5 sec of each trial, as this was the segment used by the BCI competition winner [8]. Before feature extraction, the data of each channel were centered and divided by the channel standard deviation.

B. BCI considered

Ideally, we would like a SI-BCI to be as good as a SS-BCI. Although this objective might not be attainable, it is desirable to compare a new SI-BCI design with the corresponding SS-BCI in order to assess its efficiency. Thus, in this study, we compared a SS-BCI and a SI-BCI for each kind of feature and classifier. For the SS-BCI, the classifier and CSP filters (if any) were created using the training data of a given subject S and tested on the testing data of this same subject S. This corresponds to the design of current BCI systems. For the SI-BCI, the classifier and CSP filters (if any) were created

Classifier	Type	CSP	FBCSP	BP	AR	PSD	Mean	A-priori-based BCI
LDA	SS-BCI	74.76	**81.1**	73.77	69.52	68.98	**73.63±4.89**	
	SI-BCI	65.12	62.19	**67.59**	63.19	65.05	**64.63±2.07**	
QDA	SS-BCI	73.23	79.94	64.97	63.27	64.97	69.28±7.12	64.2
	SI-BCI	61.81	61.27	63.73	57.64	56.56	60.2±3	
GMM	SS-BCI	71.45	77.24	50.62	61.65	51.31	62.44±11.85	
	SI-BCI	57.56	61.11	62.11	60.73	55.48	59.4±2.77	

using the pooled training data of all subjects expect the data of a given subject S. Then, the resulting BCI was tested on the testing data of subject S. In other words, this BCI was tested on a subject whose training data were not part of the data used to build the BCI.

C. Comparison of different feature extraction methods and classifiers

We evaluated the classification accuracy of each feature extraction method and classifier, for both SS-BCI and SI-BCI. Results obtained are displayed in Table I. Firstly, results in Table I showed that the more linear the classifier, the better the classification accuracy for both SS-BCI and SI-BCI. The best results were obtained with LDA, which outperformed QDA, which outperformed GMM. The difference between the accuracy of LDA and that of QDA and GMM was statistically significant ($p < 0.001$ with a paired t-test). Moreover, the a-piori-based BCI outperformed the QDA and GMM for SI-BCI design, regardless of the features. This suggests that a simple classifier such as a linear one is more robust against subject-to-subject variability than a more complex, non-linear classifier.

Secondly, the results showed that there exists some common information on MI among subjects as the accuracy obtained for the SI-BCI was higher than chance, roughly 60-65 % on average.

Thirdly, results showed that a method which gives very good results for SS-BCI can have very different performances when used for SI-BCI. For instance, when used with LDA, FBCSP clearly gave the best results for SS-BCI whereas it gave the worst results for SI-BCI. Its performance are even lower than that of the a-priori-based BCI that uses no machine learning. This might suggest that there is a tradeoff between good performance for SS-BCI and good performance for SI-BCI. On the other hand, the performances obtained by the classical CSP and BP, using LDA as the classifier, are among the best results for both SS-BCI and SI-BCI. This raises the following question: what made the FBCSP fail when used in a SI-BCI? If we consider the differences between CSP and FBCSP, an answer may be: frequencies. Indeed, the major difference between CSP and FBCSP is in the frequency decomposition of the EEG signals. CSP simply filters the signals in a large frequency band (here 8-30 Hz), whereas FBCSP decomposes the EEG signals into several narrow frequency bands. BP also decomposes the EEG signals into relatively large frequency bands (8-13 Hz

and 13-30 Hz). Therefore, signal decomposition into narrow frequency bands could lead to poor performances for SI-BCI. We may hypothetize that narrow frequency bands are not adapted to deal with the subject-to-subject variability in terms of spectral content. This has motivated the study presented in the next section.

D. Comparison of different frequency decompositions

In order to assess the impact of the frequency bands used on BCI performance, we evaluated the FBCSP approach using the following frequency decompositions:

Default decomposition: The default decomposition used in FBCSP decomposes the signals into 9 bands of 4 Hz (4-8 Hz, 8-12 Hz, ..., 36-40 Hz).

6 bands of 5 Hz: The signals are decomposed into larger bands than the default decomposition, using 6 bands of 5 Hz (8-13 Hz,13-18 Hz, ...,33-38 Hz).

5 bands of 6 Hz: The signals are decomposed into even larger bands than the previous decomposition, using 5 bands of 6 Hz (7-13 Hz,13-19 Hz, ..., 31-37 Hz).

4 bands of 7 Hz: The signals are decomposed into the largest frequency bands, using 4 bands of 7 Hz (6-13 Hz, 13-20 Hz, 20-27 Hz and 27-34 Hz).

Multi-resolution decomposition: The signals are decomposed using three sub-decompositions with different resolutions. The first decomposition corresponds to a coarse resolution and uses only the 8-30 Hz frequency band, as in the classical CSP method. The second decomposition uses a medium resolution which decomposes the signals into the four standard rhythms: theta (4-7 Hz), mu (8-13 Hz), beta (13-30 Hz) and gamma (30-40 Hz). Finally, the last decomposition corresponds to a finer resolution as it decomposes the signals into 5 bands of 6 Hz as described above. By using several resolutions, we hope to capture the subject-independent information with low resolutions and the subject-specific ones with finer resolutions, hence leading to improved performances. Since this decomposition leads to a larger number of features, we also compared the performances of this method while including different number of features in the feature selection step: 6, 10 and 20 pairs of features.

Performances obtained by the different frequency decompositions are displayed in Table II. Results showed that using larger frequency bands for the filter bank led to better performances for SI-BCI, while leaving the SS-BCI performances almost unchanged. Overall, the best performances were obtained with the Multi-Resolution (MR) decomposition for

4545

Classifier	Type	Default	4 bands of 7 Hz	5 bands of 6 Hz	6 bands of 5 Hz	Multi resolution 6 feature pairs	Multi resolution 10 feature pairs	Multi resolution 20 feature pairs
LDA	SS-BCI	81.1	79.94	81.4	79.94	80.17	80.48	79.94
	SI-BCI	62.19	65.35	67.67	66.59	68.52	70.07	**70.99**
QDA	SS-BCI	79.94	79.32	**81.56**	78.47	80.4	80.94	78.94
	SI-BCI	61.27	64.04	67.05	68.52	69.29	66.74	68.06
GMM	SS-BCI	77.16	77.24	80.09	77.7	79.63	73.38	60.57
	SI-BCI	61.11	64.04	64.27	62.11	63.58	66.13	65.28

each classifier. For the SI-BCI, whatever the number of features (6, 10 or 20 pairs), the MR FBCSP was significantly better than the standard FBCSP ($p < 0.005$ with a paired t-test). Still for the SI-BCI, using 10 or 20 feature pairs, the MR FBCSP was also significantly better ($p < 0.05$) than BP features which reached the best results in the previous study (see Section III-C). Using LDA with 20 pairs of features, MR FBCSP reached an average accuracy of about 71 %, which was the best score among all feature extraction methods and classifiers. Interestingly enough, when using BP features extracted from the same frequency bands as those used in the MR decomposition, rather than from mu and beta, the accuracy of the resulting SI-BCI remained roughly unchanged. More precisely, when using all features and the LDA classifier, the average accuracy was 67.82 % whereas when using only 100 features (selected using MIBIFS) this accuracy was 68.82 %. Finally, when using 40 features, i.e., as few features as with the FBCSP approach using 20 feature pairs, the accuracy was 68.6 %.

IV. CONCLUSION AND DISCUSSION

In this paper, we have explored the design of a SI MI-based BCI, i.e., a BCI that can be used immediately by a new user without the need of training the BCI with data from this new user. We compared several designs based on different feature extraction methods and different classifiers on EEG data from 9 subjects, available from data set 2a of BCI competition IV. Results revealed that linear classifiers were the most appropriate for MI-based SI-BCI design. We also proposed a FBCSP feature extraction method based on a MR frequency decomposition which achieved the highest accuracy. Overall, combining MR FBCSP with LDA achieved a promising average accuracy of about 71 % for a SI-BCI design whereas the best SS-BCI design obtained an average accuracy of about 82 %. This score is encouraging as an accuracy of 70 % is sufficient to control a BCI, as mentioned by Birbaumer [11]. Thus, these results suggested the possibility of designing SI-BCI based on MI.

However, there is still room for improvements in terms of classification accuracy. Future works could study more advanced classifiers [9][12][13] or features [6][14] to improve the performance. A more promising direction could be to incrementally transform the SI-BCI into a SS-BCI by using EEG data from the current user with unsupervised adaptation [15][16]. It could also prove rewarding to study normalization strategies in order to reduce subject-to-subject variability.

REFERENCES

[1] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M. Vaughan. Brain-computer interfaces for communication and control. *Clin. Neurophysiol.*, 113(6):767–791, 2002.
[2] A. Lécuyer, F. Lotte, R.B. Reilly, R. Leeb, M. Hirose, and M. Slater. Brain-computer interfaces, virtual reality and videogames: Current applications and future trends. *IEEE Comput.*, 41(10):66–72, 2008.
[3] S. Lu and C. Guan. Boosting-based subject-independent brain computer interface. In *ICPR*, 2008.
[4] G. Pfurtscheller and C. Neuper. Motor imagery and direct brain-computer communication. *proc. of the IEEE*, 89(7):1123–1134, 2001.
[5] K.K. Ang, Z.Y. Chin, H. Zhang, and C. Guan. Filter bank common spatial pattern (FBCSP) in brain-computer interface. In *IJCNN*, pages 2391–2398, 2008.
[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. *J. of Neur. Eng.*, 4(2):R35–57, 2007.
[7] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans. on Rehab.*, 8(4):441–446, 2000.
[8] BCI competition IV website: http://ida.first.fhg.de/ projects/bci/competition_iv/.
[9] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms for EEG-based brain-computer interfaces. *J. of Neur. Eng.*, 4:R1–R13, 2007.
[10] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Recognition, second edition*. WILEY-INTERSCIENCE, 2001.
[11] N. Birbaumer. Breaking the silence: Brain-computer interfaces (BCI) for communication and motor control. *Psychophysiology*, 43(6):517–532, 2006.
[12] F. Lotte, A. Lécuyer, F. Lamarche, and B. Arnaldi. Studying the use of fuzzy inference systems for motor imagery classification. *IEEE Trans. on Neur. Syst. and Rehab.*, 15(2):322–324, 2007.
[13] M. Zhong, F. Lotte, M. Girolami, and A. Lécuyer. Classifying EEG for brain computer interfaces using gaussian processes. *Pattern Recog. Lett.*, 29:354–359, 2008.
[14] F. Lotte, A. Lécuyer, and B. Arnaldi. FuRIA: An inverse solution based feature extraction algorithm using fuzzy set theory for brain-computer interfaces. *IEEE Trans. on Sig. Proc.*, 2009.
[15] S. Lu, C. Guan, and H. Zhang. Unsupervised brain computer interface based on inter-subject information and online adaptation. *IEEE Trans. on Neur. Syst. and Rehab.*, 2009.
[16] C. Vidaurre, A. Schlogl, R. Cabeza, R. Scherer, and G. Pfurtscheller. A fully on-line adaptive BCI. *IEEE Trans. on Biomed. Eng.*, 53(6):1214–1219, 2006.
[17] M. Naeem, C. Brunner, R. Leeb, B. Graimann, and G. Pfurtscheller. Seperability of four-class motor imagery data using independent components analysis. *J. of Neur. Eng.*, 3:208–216, 2006.