This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2015.2398573, IEEE Transactions on Neural Systems and Rehabilitation Engineering

Adaptive Stacked Generalization for Multiclass Motor Imagery-based Brain Computer Interfaces

Luis F. Nicolas-Alonso*, *Student Member, IEEE*, Rebeca Corralejo, *Student Member, IEEE*, Javier Gomez-Pilar, Daniel Álvarez *Member, IEEE,* and Roberto Hornero, *Senior Member, IEEE*

***Abstract*— Practical motor imagery-based brain computer interface (MI-BCI) applications are limited by the difficult to decode brain signals in a reliable way. In this paper, we propose a processing framework to address non-stationarity, as well as handle spectral, temporal, and spatial characteristics associated with execution of motor tasks. Stacked generalization is used to exploit the power of classifier ensembles for combining information coming from multiple sources and reducing the existing uncertainty in EEG signals. The outputs of several regularized linear discriminant analysis (RLDA) models are combined to account for temporal, spatial, and spectral information. The resultant algorithm is called stacked RLDA (SRLDA). Additionally, an adaptive processing stage is introduced before classification to reduce the harmful effect of intersession non-stationarity. The benefits of the proposed method are evaluated on the BCI Competition IV dataset 2a. We demonstrate its effectiveness in binary and multiclass settings with four different motor imagery tasks: left-hand, right-hand, both feet, and tongue movements. The results show that adaptive SRLDA outperforms the winner of the competition and other approaches tested on this multiclass dataset.**

***Index Terms*— Adaptive estimation, Brain Computer Interfaces, Classifier ensembles, Common spatial pattern, Electroencephalography, Linear Discriminant Analysis, Stacked generalization.**

I. Introduction

BRAIN computer interfaces (BCIs) are becoming more popular as a mean to improve the quality of life of severely disabled people and, at the same time, reduce the cost of intensive care [1]. BCIs enable humans to interact with their surroundings, without the involvement of peripheral nerves and muscles, by using control signals generated from brain activity. An alternative non-muscular communication channel is created to translate brain activity directly into sequences of control commands for external devices such as computers, speech synthesizers [2], assistive appliances [3], and neural prostheses [4] amongst many others.

Sensorimotor rhythms have been widely applied for relaying commands in BCI research [5]. Changes in the amplitude of sensorimotor rhythms, known as event-related desynchronization/synchronization (ERD/ERS) [6], can be detected from electroencephalography (EEG) signals when cerebral activity is related to any motor task, even without performing actual movement [7]. However, the applicability of sensorimotor rhythms in real environments is still limited by low transfer rates [8]. Motor imagery based-BCIs (MI-BCIs) or sensorimotor-based BCIs (SMR-BCIs) are usually designed to transmit just two different commands corresponding to two brain states: left and right hand movements. Extending the number of tasks naturally increases misclassification rate because there is a trade-off between number of classes and accuracy [8]. In recent years, temporal, spectral, and spatial features [9-14], as well as combination of feature vectors extracted from independent physiological sources [8] have been researched to enhance MI-BCI performance and speed up communication. With this purpose, a new processing algorithm is presented in this paper.

Non-stationarity of brain activity is a major issue to robust operation of BCIs [15]. Diverse behavioral and mental states continuously change the statistical properties of brain signals. Although there may be multiple types of non-stationarity, we focus on the intersession non-stationarity and temporal variability. The intersession non-stationarity refers to the fact that patterns observed during calibration sessions are different from those recorded during subsequent online sessions [15]. BCI systems are often calibrated by supervised learning during training sessions. They assume that feature vectors extracted from EEG data of different sessions follow similar probability distributions. Hence, non-stationarity can result in degraded performance. Several adaptive techniques such as expectation-maximization-based adaptation [16-19], transductive learning [20], self-training [21-23], bias adaptation [15, 24], covariate shift adaptation [25, 26], data space adaptation (DSA) based on the Kullback-Leibler divergence [27], or dynamic Bayesian classifiers based on the Kalman filter [28-30] were proposed

This work was supported in part by the Project Cero 2011 on Ageing from Fundación General CSIC, Obra Social La Caixa and CSIC and a grant by the Ministerio de Economía y Competitividad and FEDER under project TEC2011-22987. L. F. Nicolas-Alonso was in receipt of a PIF-UVa grant from Universidad de Valladolid.

* L. F. Nicolas-Alonso is with the Biomedical Engineering Group, ETSI Telecomunicación, Universidad de Valladolid, Valladolid, 47011 Spain (e-mail: lnicalo@ribera.tel.uva.es).

R. Corralejo, J. Gomez-Pilar, D. Álvarez, and R. Hornero are with the Biomedical Engineering Group, ETSI Telecomunicación, Universidad de Valladolid, Valladolid, 47011 Spain (e-mail: rebeca.corralejo@gib.tel.uva.es; javier.gomez@gib.tel.uva.es; dalvgon@ribera.tel.uva.es; robhor@tel.uva.es).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2015.2398573, IEEE Transactions on Neural Systems and Rehabilitation Engineering

to address non-stationarity in BCI. Within multiclass MI-BCI context, Xu *et al.* [31] proposed a new adaptive multiclass classifier that enlarged training dataset by adding unlabeled test samples. The increase of performance can be associated with the enhancement of adaptability to non-stationary EEG signals since new incoming data is used to update the classifier. However, the method does not give more importance to the most recent feature vectors. We share the perspective presented by Vidaurre *et al.* [24] that adaptive algorithms should include a forgetting factor. Llera *et al.* [32] and Nicolas-Alonso *et al.* [33] extended the adaptive method proposed by Vidaurre *et al.* [24] to multiclass settings. Continuing this working line, we use an adaptive processing stage before classification stage that estimates and reduces the mismatching running exponentially weighted moving average (EWMA).

Temporal variability in motor imagery tasks arises because of the changing brain dynamics and difficulty of performing long-term mental tasks by users. It results in the detectable ERD/ERS events not occurring in fixed time segments [34]. It can be reduced by constraining experimental conditions. Synchronous MI-BCI provides users with a cue that indicates when the mental task should be executed [35]. However, the task of generating adequate ERD/ERS patterns could be unintentionally delayed. In an effort to address such a problem, methods involving temporal features [36-38] or dynamic classifiers such as hidden Markov models (HMMs) [39], conditional random fields (CRFs) [40] or hidden CRFs (HCRFs) [40] have been used in binary MI-BCIs. Regarding multiclass MI-BCIs, HMMs [41] or CRF [42] have been tested to classify five tasks (four motor tasks and one arithmetic task) and three motor tasks (imagery of the left and right hands and the feet), respectively. More recently, Asensio-Cubero *et al.* [13] developed a multiclass common spatial pattern (CSP)-based algorithm that automatically segmented motor imagery EEG data within a trial to extract spatio-temporal information.

Reliable communication with multiclass MI-BCIs seems to require handling information from different domains. Ensemble learning appears to be particularly suitable for combining information from multiple sources. It has been successfully applied to a wide range of real world problems [43]. The main advantage of such a technique is that the variance and bias can be simultaneously reduced [44]. Voting [45-48], bagging [49], boosting [49, 50], and stacked generalization [49, 51] have been the main types of classifier ensembles in BCI research. They have not been applied to deal with temporal variability yet. In the current study, we investigate whether stacked generalization can increase reliability of multiclass MI-BCI combining temporal, spatial, and spectral information.

The aim of this study is to propose an approach based on adaptive ensemble learning that addresses the intersession non-stationarity and temporal variability in motor imagery EEG data. The novelty of our approach lies in two components. Firstly, adaptive processing with EWMA is introduced in order to ameliorate the intersession non-stationarity effects on classification performance. Secondly, we propose an ensemble classification method that employs stacked generalization to handle multiple information sources. Several regularized linear discriminant analysis (RLDA) models are combined to account for temporal, spatial, and spectral information in EEG data. The resultant algorithm is called stacked RLDA (SRLDA). The benefits of the adaptive stage and stacked generalization are evaluated on the BCI Competition IV dataset 2a [52].

II. BCI Competition IV Dataset 2a Description

BCI Competition IV dataset 2a was provided by Graz University [52] and contains EEG data from 9 healthy subjects performing four motor imagery tasks: movement of the left hand, right hand, feet, and tongue. The dataset provides two sessions for each subject, one for training and the other for evaluation, recorded on different days. Each session includes 288 trials of data (72 for each of the four possible tasks). The electrode montage consisted of 22 EEG channels and 3 monopolar electrooculogram (EOG) channels (with left mastoid serving as reference). In the current study, we use only the EEG channels. The signals were sampled at 250 Hz and band-pass filtered between 0.5 Hz and 100 Hz. A 50 Hz notch filter was also applied to suppress power line noise.

The recording protocol was as follows. At the beginning of each trial, a cross was shown on the black screen and a short warning tone was given. At second 2, a cue in the form of an arrow pointing to the left, right, down, or up (corresponding to one of the four classes: left hand, right hand, foot or tongue) was presented during 1.25 s. At the given signal, the subjects performed the corresponding motor imagery task until the cross disappeared from the screen at second 6. Refer to Tangermann *et al.* [52] for further details on the BCI Competition IV dataset 2a.

III. Proposed Method

Fig. 1 illustrates the architecture of the proposed signal processing chain. EEG signals are processed by five stages: multiple band-pass filtering using Finite Impulse Response (FIR) filters, spatial filtering using the CSP algorithm, feature selection, adaptive processing, and SRLDA. All of them except band-pass filtering are optimized using the labelled trials from the training session. The optimized algorithm is then used to predict the motor imagery task for each trial over the evaluation session. Feature vectors are computed using a sliding 2-second window of EEG data. Therefore, classification output is not generated during the first 2 seconds. As the computation and classification of feature vectors at every sample is computationally intensive, the output was only computed on every alternate 10th time sample. That is, every time that features are extracted, the sliding window moves forward 10 samples. Thus, given the sampling frequency of 250 Hz, 25 feature vectors are computed per second. Following sections further explain each stage of the proposed signal processing chain.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2015.2398573, IEEE Transactions on Neural Systems and Rehabilitation Engineering

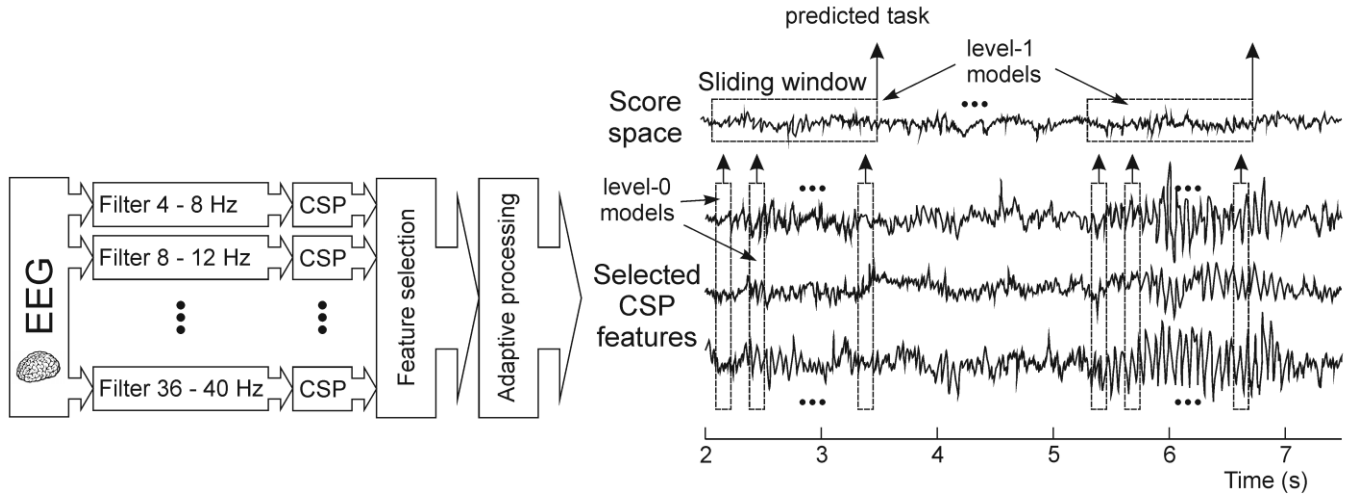


Fig. 1. Architecture of the signal processing chain. Adaptive SRLDA classifier is illustrated in detail. The adaptive processing stage reduces the intersession non-stationarity between training and evaluation session. Level-0 RLDA models classify the selected CSP features over the time producing a score sequence. Then, the scores are combined with a sliding window. Finally, level-1 RLDA models make the final prediction sequence using the combined scores as features.

A. Feature extraction and selection

1) Band-pass filtering

The first stage consists of a filter bank that splits the EEG into 9 spectral pass bands: 4-8 Hz, 8-12 Hz,…, 36-40 Hz [12]. Nine FIR filters designed by means of Kaiser Window are used. FIR filters have linear phase, which does not distort the phase of the filtered signal. It makes them particularly suitable for filter banks. The transition bandwidth is set at 1 Hz. Although other configurations are as effective, this transition bandwidth yields a reasonable order filter and discriminative capacity between spectral bands.

2) Spatial filtering

The second stage spatially filters each band-pass signal using CSP [53]. CSP was originally devised for the analysis of multichannel data belonging to 2-class problems. Then, although other options are feasible in multiclass problems, we adopt the one-versus-rest approach [9]. Spatial filters for multiclass motor imagery classification are computed on the basis of the trials for each class versus the trials for all other classes.

Not all CSP filters are relevant for subsequent classification. The discriminative power of a spatial filter is related to its associated eigenvalue. Therefore, we extract 4 features corresponding to the filter with the 2 highest and the 2 lowest eigenvalues [12]. We obtain 4 features for binary classification and 16 for 4-class classification as a result of repeating the CSP algorithm for each class. Finally, the features of the 9 spectral bands are concatenated to form a single feature vector of 36 and 144 features, respectively.

3) Feature selection

The third stage selects the most discriminative spatio-spectral features using mutual information-based best individual feature (MIBIF) algorithm [12]. Features are sorted according to mutual information in descending order. The first features are then selected. The number of selected features is a free parameter tuned by 10-fold cross validation (CV) on the training session. More details are given below.

B. Classification

1) Adaptive processing

The adaptive processing stage removes the non-stationary bias from each feature. Every incoming feature vector is centered by subtracting the global mean. Firstly, data from all training session are used to estimate the global mean. Across the evaluation session, upon the arrival of a new sample at the time t from the i-th evaluation trial, the global mean $\boldsymbol{\mu}_{\mathrm{G}}(i,t)$ is updated by means of EWMA. EWMA estimates the global mean by an amount that is proportional to the most recent forecast error. The sequential estimates are given by

$$\boldsymbol{\mu}_{\mathrm{G}}(i,t) = (1-\eta)\cdot\boldsymbol{\mu}_{\mathrm{G}}(i-1,t) + \eta\cdot\boldsymbol{x}(i,t) \tag{1}$$

where $\boldsymbol{x}(i,t)$ is the current input feature vector of the i-th evaluation trial at the time t and η is the update coefficient, which has to be configured by the user.

EWMA has been widely used for mean estimation in noisy environments [54] and fits with the idea of adaptation in BCI applications, as proposed by Vidaurre *et al.* [24]. More importance is given to the most recent terms in the time series rather than older data. Simple algebraic manipulation reveals that $\boldsymbol{\mu}_{\mathrm{G}}(i,t)$ can be written as a weighted average of all past observations, in which weights for older samples decay exponentially. Vidaurre *et al.* [24] proposed an adaptive binary LDA based on EWMA. In this study, we extend the work to multiclass problems and other classifiers such as regularized LDA and stacked classifiers. It should be noted that both approaches are quantitatively the same in binary settings when non-regularized LDA is used.

2) Regularized linear discriminant analysis

LDA is the basis of our classifier ensemble SRLDA. We use binary and multiclass versions of LDA [55] for binary and multiclass motor imagery classification. LDA is a standard

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2015.2398573, IEEE Transactions on Neural Systems and Rehabilitation Engineering

tool for classification and dimension reduction that seeks linear projections of features. The distance between class means is maximized while the class variance is minimized [55]. Linear classification has provided accurate results without high computational requirements in numerous BCI applications [56-58]. Although LDA has been proven to be reasonably robust in multiple contexts, if the number of features is comparable to the number of observations, the estimate model can become highly unstable, giving rise to high variance [59]. Applying regularization reduces the variance ensuring good generalization ability for unseen data. In this work, we adopt the Ridge regularization by adding some constant values to the diagonal elements of the within-class scatter matrix [55]. The regularization penalty is controlled by a free parameter $\alpha > 0$ that is empirically tuned by 10-fold CV on the training session.

3) Stacked regularized linear discriminant analysis

Stacked generalization is a way of constructing classifier ensembles combining multiple models to induce a higher-level classifier with improved performance [60]. Stacking introduces the concept of metalearning where the input of the metamodels, also called the level-1 models, consists of the predictions of the base models, or level-0 models. When stacked learning is used for classification, the extracted features are first fed into the level-0 models, and each one produces a score for each possible class. Then, the outputs of level-0 models are fed into the level-1 models, which combine them into the final prediction.

In an attempt to exploit the temporal structure of EEG, we take advantage of stacking to combine spatial, spectral, and temporal information over the trial. The architecture of the SRLDA is illustrated in Fig. 1. In the first stage or level-0, a set of regularized RLDA models processes the CSP features, which provide spatial and spectral characterization of EEG signal. We build a different classifier for each feature vector, that is, 25 RLDA models per second. Each of them produces $C - 1$ scores, where C is the number of classes. In the second stage or level-1, the output of the RLDA classifiers is then fed into another regularized RLDA classifier in order to treat the temporal information throughout the trial. According to Ting and Witten [61], linear projections rather than class predictions are used as input. This second level determines the task that the subject is performing.

Temporal information is stacked in a sliding fashion as can be noted in Fig. 1. This is because information at the beginning of the trial may have little to do with the information at the end. The size of the sliding window determines the number of level-0 models combined by each level-1 model. This parameter is configured by CV to optimize performance. It should be pointed that, at the beginning of the trial, shorter sliding windows are used. It ensures that classification output is produced even if the number of score samples is less than the length of the window. SRLDA does not require delaying the feedback until having recorded enough EEG samples. Given that CSP in feature extraction stage already delays the feedback 2 seconds, we have designed SRLDA in such a way to avoid delaying the feedback even more.

Finally, we would like to emphasize an issue to train the level-1 classifiers. If all instances from the training session are used to build the level-0 models and the corresponding predictions to train the level-1 model, then too simplistic rules are learned by the level-1 models [62]. This is because the same data are exploited to train and predict the instances needed in the level-1. Then, to imitate the scenario during the evaluation session, we perform leave-one-out CV for every level-0 classifier. Each instance in the training session is predicted using the model trained with the remaining data. Now, the level-0 models have not been trained on the instance to be predicted, therefore their predictions are unbiased. Once the level-1 input data have been generated by this holdout procedure, the level-0 classifiers are retrained on the full training dataset to make slightly better use of the data and leading to better prediction in the test session.

IV. Results

In this section, we describe our experiments on the publicly available BCI Competition IV dataset 2a. Performance analyses are conducted on both binary and multiclass datasets. For binary tests, we split the multiclass dataset into 6 cases: left-hand versus right-hand, left-hand versus foot, left-hand versus tongue, right-hand versus foot, right-hand versus tongue, and foot versus tongue. The effectiveness of the proposed adaptive or non-adaptive SRLDA is analyzed and compared against non-regularized LDA, adaptive or non-adaptive RLDA, as well as some adaptive methods in BCI literature, namely, DSA [27], MPMLDA with equal responsibilities [32], and MPMLDA [32]. All these methods are tested in combination with the filter bank, CSP, and MIBIF. Although DSA was proposed for classifying two motor imagery tasks, it can be directly extended to multiclass classification. We use only the unsupervised version of DSA algorithm with continuous adaptation mode for a fair comparison. Arvaneh *et al.* [27] proposed other versions of DSA that made use of a small number of labelled trials at the beginning of evaluation session to adapt. On the other hand, MPMLDA was designed for multiclass motor imagery classification but MPMLDA with equal responsibilities can be directly applied to binary problems. Finally, the results of the winner of the competition [12] and other studies on the same dataset [10, 13, 33, 63-67] are also reported.

Performance is measured in terms of Cohen's kappa coefficient [68]. According to the evaluation rules of the BCI competition, the maximum kappa throughout the time course of the paradigm is reported. All methods are configured in the same way. Configurable parameters are adjusted running 10-fold CV or chronological validation with partitions containing the same trials. Adaptive SRLDA has five parameters: the number of selected features F, the regularization parameters α_1 and α_2, the size of the sliding window Ω, and the update coefficient η. Adaptive RLDA has three parameters: the number of selected features F, the regularization parameter α, and the update coefficient η. DSA has one parameter the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2015.2398573, IEEE Transactions on Neural Systems and Rehabilitation Engineering

number of selected features F. Finally, MPMLDA has two parameters: the number of selected features F and the update coefficient η. All the parameters but the update coefficients are optimized with 10-fold CV on training session. The update coefficients are optimized with chronological validation on training session. Further information is given below.

A. Design and optimization on the training session

In this section, we give details and results regarding RLDA and SRLDA configuration for multiclass case. The same applies to every binary case. Owing to the large number of parameters, we split SRLDA configuration in three steps. Firstly, the output of the RLDA models in the level-0 is used to determine the optimal F and α_1 by means of 10-fold CV. Secondly, α_2 and Ω in the level-1 are optimized by 10-fold CV as well. Here, we use the previously optimized parameters F and α_1. Although each level-0 model could have been designed with different parameters, we choose a single configuration that corresponds to maximize the kappa value of the model providing the highest performance. The same applies to level-1 models. The two-step optimization of F, α_1, α_2, and Ω is carried out independently of η. The update coefficient η is set to 0.

A wide range of values is defined in order to analyze the effect of the parameters on generalization ability. F is varied from 10 to 144 in steps of 10, α_1 and α_2 take values from 10 to 300 in steps of 5, Ω ranges from 0.40 to 4.40 s in steps of 0.80, and η varies from 0 to 0.2 in steps of 0.01. These ranges are selected taking into account that the maximum number of features is 144 and the classification performance is observed to decay when the parameters fall outside. Fig. 2 illustrates the 10-fold CV classification results for each step according to F, α_1, α_2, and Ω. We use mean kappa value because we select a single set of parameters, which is shared by all the subjects. Mean kappa value is computed obtaining the highest kappa value over the training session for each subject and averaging the results. We choose the maximum as the optimal value (F = 91, α_1 = 85, α_2 = 155, and Ω = 2 s). The amount of features in level-1 is determined by Ω. Given that there are 25 level-0 models per second, each level-1 model combines the C - 1 scores produced by 50 models. That is, the number of features in level-1 is 3 × 50.

We perform chronological validation on the training session [69] because the optimal value of η depends on the natural sequence of trials. The training session is chronologically split into two subsets containing the 60% and 40% of the trials. The first one is used for training and the other for evaluation. This procedure imitates the online learning scenario throughout the evaluation session. Adaptive RLDA and adaptive SRLDA use individually optimized update coefficients. Fig. 3 illustrates the performance in terms of mean kappa as the update coefficient η varies. The improvement of adaptation is notable for a wide range of values. Mean kappa value reaches the maximum at η = 0.02 for RLDA and η = 0.04 for SRLDA. Therefore, η are fixed to these values for the rest of the experiments on the unseen evaluation session.

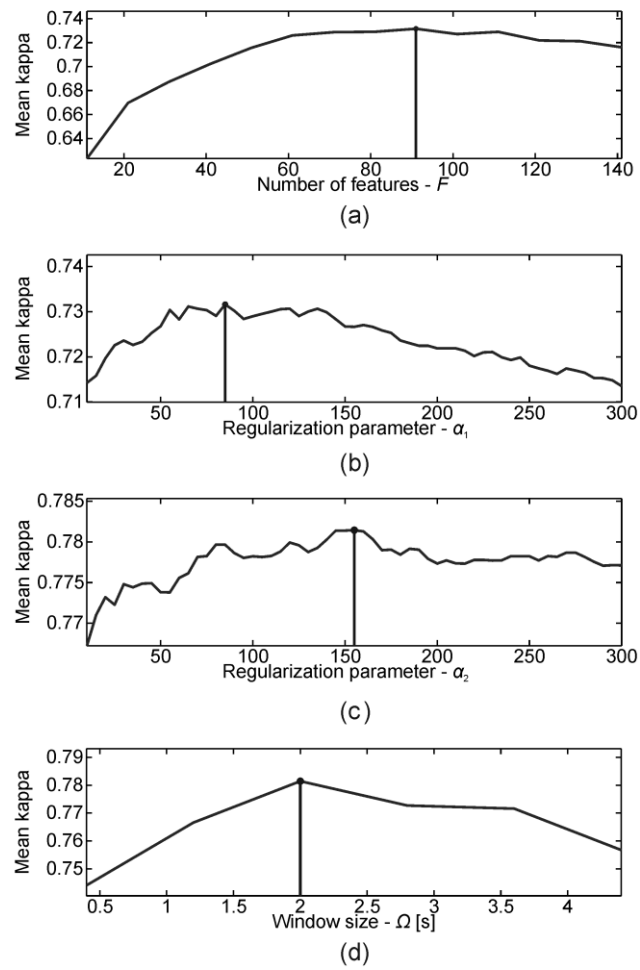


Fig. 2. SRLDA configuration results: mean kappa variation with (a) the number of features F when α_1 is optimal (α_1 = 85), (b) the regularization parameter α_1 when F is optimal (F = 91) (c) the regularization parameter α_2 when F, α_1, and Ω are optimal (F = 91, α_1 = 85, and Ω = 2 s) and (d) the sliding window size Ω when F, α_1, and α_2 are optimal (F = 91, α_1 = 85, and α_2 = 155). Although the parameters are determined in two steps running 10-fold CV, we show the parameter influence in separate plots for clarity.

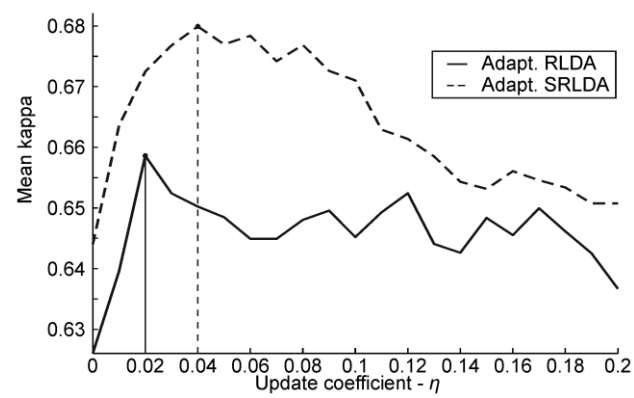


Fig. 3. Mean kappa of adaptive RLDA and adaptive SRLDA over update coefficient η. The other configurable parameters of RLDA and SRLDA were fixed to optimal values: RLDA (F = 91, α_1 = 85) and SRLDA (F = 91, α_1 = 85, α_2 = 155, and Ω = 2 s).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2015.2398573, IEEE Transactions on Neural Systems and Rehabilitation Engineering

B. Binary datasets: performance assessment on the evaluation session

The comparative analyses of the six cases of motor imagery binary classification are presented in Table I. The highest classification performance for each subject is in boldface. MPMLDA is not applicable to binary problems. The binary version of MPMLDA with equal responsibilities is used instead. Actually, it corresponds to the pooled mean update proposed by Vidaurre *et al.* [24].

The results evidence the superiority of adaptive methods, namely, DSA, MPMLDA, and our EWMA-based methods, with respect to LDA or RLDA ($\eta = 0$). However, there does not seem to be clear reasons for recommending a particular method based on kappa values.

Stacked generalization is shown to increase the performance further. Table II contains the *p*-values of Wilcoxon's test to evaluate the statistical significance of the difference between the kappa values for SRLDA ($\eta > 0$) and other methods.

C. Multiclass dataset: performance assessment on the evaluation session

The classification performances on the multiclass BCI Competition IV dataset 2a are presented in Table III. We also report the results of other studies on the same dataset [10, 12, 13, 33, 63-67]. The highest classification performance for each subject is in boldface.

Adaptive classification and stacking generalization bring a substantial improvement. SRLDA ($\eta > 0$) yields the highest mean kappa value (0.74). Table IV presents the results of Wilcoxon's test to evaluate the statistical significance of the difference between the performance of RLDA and SRLDA with or without adaptation, and the other methods.

TABLE I
SUMMARY OF KAPPA VALUES OVER DIFFERENT TYPES OF BINARY-CLASS MOTOR IMAGERY.

LEFT – RIGHT

METHOD	SUBJECTS									AVG.
	A1	A2	A3	A4	A5	A6	A7	A8	A9	
LDA	**0.85**	0.10	0.83	**0.46**	0.75	0.36	0.81	0.93	0.79	0.65
DSA	0.84	0.04	0.91	0.52	0.76	0.37	0.77	0.94	0.80	0.66
MPMLDA (EQUAL RESP.)	**0.85**	0.18	**0.93**	**0.46**	0.75	0.36	0.81	0.93	0.79	0.67
RLDA ($\eta = 0$)	0.78	0.11	0.83	0.33	0.67	0.29	0.83	0.92	**0.83**	0.62
SRLDA ($\eta = 0$)	0.78	0.28	0.83	0.43	**0.89**	0.43	**0.96**	0.85	0.78	0.69
RLDA ($\eta > 0$)	0.81	0.03	**0.93**	0.44	0.74	0.32	0.88	0.93	0.82	0.65
SRLDA ($\eta > 0$)	0.82	**0.39**	0.92	0.51	**0.89**	**0.49**	**0.96**	**0.96**	0.81	**0.75**

LEFT – FOOT

METHOD	SUBJECTS									AVG.
	A1	A2	A3	A4	A5	A6	A7	A8	A9	
LDA	**0.97**	0.57	0.92	0.76	0.24	0.35	0.99	0.82	0.94	0.73
DSA	**0.97**	0.58	0.91	0.80	0.66	0.25	**1**	0.81	0.94	0.77
MPMLDA (EQUAL RESP.)	**0.97**	0.56	0.93	**0.81**	0.57	0.38	**1**	0.83	0.93	0.77
RLDA ($\eta = 0$)	0.92	0.47	0.89	0.68	0.15	0.33	**1**	0.82	0.94	0.69
SRLDA ($\eta = 0$)	0.94	0.78	0.94	0.71	0.40	0.29	0.99	0.88	0.86	0.75
RLDA ($\eta > 0$)	0.93	0.53	0.92	0.72	0.50	0.31	**1**	0.86	**0.96**	0.75
SRLDA ($\eta > 0$)	0.96	**0.82**	**0.96**	0.75	**0.71**	**0.61**	**1**	**0.88**	**0.96**	**0.85**

LEFT – TONGUE

METHOD	SUBJECTS									AVG.
	A1	A2	A3	A4	A5	A6	A7	A8	A9	
LDA	0.94	0.35	0.90	0.81	0.46	0.35	0.88	0.83	0.94	0.72
DSA	**0.96**	0.38	0.94	**0.84**	0.61	0.32	0.85	0.90	0.94	0.75
MPMLDA (EQUAL RESP.)	0.93	0.44	**0.96**	0.75	0.56	**0.44**	0.92	0.89	0.94	0.76
RLDA ($\eta = 0$)	0.94	0.33	0.92	0.67	0.46	0.43	0.88	0.78	0.94	0.71
SRLDA ($\eta = 0$)	0.93	0.60	0.92	0.78	0.40	0.36	0.90	0.68	**0.96**	0.73
RLDA ($\eta > 0$)	0.93	0.32	0.94	0.72	0.53	0.50	**0.96**	0.90	0.93	0.75
SRLDA ($\eta > 0$)	0.93	**0.63**	0.90	0.81	**0.68**	0.33	**0.96**	**0.92**	0.94	**0.79**

RIGHT – FOOT

METHOD	SUBJECTS									AVG.
	A1	A2	A3	A4	A5	A6	A7	A8	A9	
LDA	**0.97**	0.67	0.85	0.75	0.54	0.28	**1**	0.82	0.49	0.71
DSA	0.96	0.64	**0.94**	0.79	0.71	0.34	**1**	0.85	0.52	0.75
MPMLDA (EQUAL RESP.)	**0.97**	0.65	0.93	0.74	0.57	0.32	**1**	0.82	0.47	0.72
RLDA ($\eta = 0$)	0.94	0.57	0.85	0.72	0.32	0.31	**1**	0.81	0.46	0.66
SRLDA ($\eta = 0$)	0.96	**0.93**	0.83	**0.90**	0.51	0.32	0.99	0.86	0.38	0.74
RLDA ($\eta > 0$)	0.96	0.54	0.93	0.78	0.68	0.46	**1**	0.82	0.58	0.75
SRLDA ($\eta > 0$)	**0.97**	0.92	**0.94**	**0.90**	**0.76**	**0.56**	0.99	**0.88**	**0.64**	**0.84**

RIGHT – TONGUE

METHOD	SUBJECTS									AVG.
	A1	A2	A3	A4	A5	A6	A7	A8	A9	
LDA	0.97	0.39	0.89	0.74	0.67	0.29	0.82	0.65	0.71	0.68
DSA	0.96	0.45	0.93	0.76	0.72	**0.41**	0.88	0.78	0.72	0.73
MPMLDA (EQUAL RESP.)	0.97	0.51	0.92	0.75	0.68	0.40	0.92	0.76	0.74	0.74
RLDA ($\eta = 0$)	**1**	0.44	0.86	0.74	0.39	0.22	0.89	068	0.36	0.62
SRLDA ($\eta = 0$)	0.97	0.46	0.89	0.78	0.67	0.26	0.92	0.68	0.74	0.71
RLDA ($\eta > 0$)	0.99	0.43	0.97	0.75	0.60	0.24	0.97	**0.79**	0.78	0.72
SRLDA ($\eta > 0$)	0.99	**0.53**	**0.94**	**0.86**	**0.86**	0.36	**1**	0.75	**0.83**	**0.79**

FOOT – TONGUE

METHOD	SUBJECTS									AVG.
	A1	A2	A3	A4	A5	A6	A7	A8	A9	
LDA	0.57	0.54	0.50	0.58	**0.44**	0.33	0.64	0.69	0.75	0.56
DSA	0.62	0.60	0.63	0.61	0.42	0.44	0.65	0.75	0.87	0.62
MPMLDA (EQUAL RESP.)	0.58	0.53	0.64	0.56	0.50	0.43	0.68	0.74	**0.89**	0.62
RLDA ($\eta = 0$)	0.69	0.67	0.53	**0.68**	0.26	0.39	0.63	0.64	0.51	0.56
SRLDA ($\eta = 0$)	0.63	0.78	0.69	0.65	0.40	0.54	**0.76**	0.68	0.67	0.65
RLDA ($\eta > 0$)	0.68	0.65	0.69	0.65	0.31	0.43	0.65	0.74	0.76	0.62
SRLDA ($\eta > 0$)	**0.72**	**0.81**	**0.86**	0.61	0.43	**0.68**	0.75	**0.81**	0.83	**0.72**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2015.2398573, IEEE Transactions on Neural Systems and Rehabilitation Engineering

TABLE II
WILCOXON'S TEST RESULTS (*P*-VALUES) EVALUATING THE STATISTICAL SIGNIFICANCE OF THE DIFFERENCE BETWEEN THE BINARY PERFORMANCE OF SRLDA ($\eta > 0$) AND OTHER METHODS. SIGNIFICANT DIFFERENCES ARE MARKED WITH A SYMBOL (* < 0.05 AND ** < 0.01).

Method	SRLDA ($\eta > 0$)					
	LEFT - RIGHT	LEFT - FOOT	LEFT - TONGUE	RIGHT - FOOT	RIGHT - TONGUE	FOOT - TONGUE
LDA	0.0156*	0.0469*	0.1250	0.0156*	0.0039**	0.0078**
DSA	0.0547	0.0781	0.4258	0.0117*	0.0742	0.0195*
MPMLDA (EQUAL RESP.)	0.0430*	0.1016	0.3906	0.0234*	0.0703	0.0508
RLDA ($\eta = 0$)	0.0078**	0.0078**	0.1563	0.0078**	0.0078**	0.0117*
SRLDA ($\eta = 0$)	0.0156*	0.0078**	0.1406	0.0625	0.0039**	0.0391*
RLDA ($\eta > 0$)	0.0313*	0.0156*	0.4531	0.0156*	0.0625	0.0117*

TABLE III
SUMMARY OF KAPPA VALUES ON MULTICLASS BCI COMPETITION IV DATASET 2A.

METHOD	SUBJECTS									AVG.
	A1	A2	A3	A4	A5	A6	A7	A8	A9	
ANG *ET AL.* [12] (WINNER)	0.68	0.42	0.75	0.48	0.40	0.27	0.77	0.75	0.61	0.57
GOUY-PAILLER *ET AL.* [10]	0.66	0.42	0.77	0.51	0.50	0.21	0.30	0.69	0.46	0.50
WANG [63]	0.67	0.49	0.77	0.59	0.52	0.31	0.48	0.75	0.65	0.58
BARACHANT *ET AL.* [64]	0.74	0.38	0.72	0.50	0.26	0.34	0.69	0.71	0.76	0.57
WANG *ET AL.* [65]	0.56	0.41	0.43	0.41	**0.68**	**0.48**	0.80	0.72	0.63	0.57
KAM *ET AL.* [66]	0.74	0.35	0.76	0.53	0.38	0.31	0.84	0.74	0.74	0.60
ASENSIO-CUBERO *ET AL.* [13]	0.75	0.50	0.74	0.40	0.19	0.41	0.78	0.72	**0.78**	0.59
ASENSIO-CUBERO *ET AL.* [67]	0.76	0.32	0.76	0.47	0.31	0.34	0.59	0.76	0.74	0.56
NICOLAS-ALONSO *ET AL.* [33]	0.77	0.39	0.87	0.55	0.47	0.32	0.74	0.79	0.72	0.62
LDA	0.76	0.41	0.83	0.56	0.35	0.26	0.79	0.73	0.53	0.58
DSA	0.78	0.50	0.83	0.63	0.59	0.38	0.85	0.81	0.67	0.67
MPMLDA (EQUAL RESP.)	0.65	0.51	0.80	0.63	0.52	0.33	0.85	0.75	0.62	0.63
MPMLDA	0.77	0.53	0.85	0.65	0.55	0.32	0.86	0.79	0.68	0.67
RLDA ($\eta = 0$)	0.79	0.48	0.75	0.63	0.40	0.34	0.84	0.79	0.59	0.62
SRLDA ($\eta = 0$)	0.83	**0.55**	0.79	0.62	0.47	0.34	0.92	0.80	0.60	0.66
RLDA ($\eta > 0$)	0.77	0.50	0.88	0.67	0.51	0.36	0.88	0.83	0.73	0.68
SRLDA ($\eta > 0$)	**0.84**	**0.55**	**0.90**	**0.71**	0.66	0.44	**0.94**	**0.85**	0.76	**0.74**

TABLE IV
WILCOXON'S TEST RESULTS (*P*-VALUES) EVALUATING THE STATISTICAL SIGNIFICANCE OF THE DIFFERENCE BETWEEN THE MULTICLASS PERFORMANCES OF RLDA AND SRLDA, AND OTHER METHODS. SIGNIFICANT DIFFERENCES ARE MARKED WITH A SYMBOL (* < 0.05 AND ** < 0.01).

Method	RLDA ($\eta = 0$)	SRLDA ($\eta = 0$)	RLDA ($\eta > 0$)	SRLDA ($\eta > 0$)
ANG *ET AL.* [12] (WINNER)	0.1250	0.0391*	0.0039**	0.0039**
GOUY-PAILLER *ET AL.* [10]	0.1797	0.0391*	0.0039**	0.0039**
WANG [63]	0.5078	0.1797	0.0391*	0.0039**
BARACHANT *ET AL.* [64]	0.0703	0.0703	0.0391*	0.0078**
WANG *ET AL.* [65]	0.5078	0.5078	0.1797	0.1796
KAM *ET AL.* [66]	0.2891	0.0391*	0.0391*	0.0039**
ASENSIO-CUBERO *ET AL.* [13]	0.5078	0.1797	0.2891	0.0391*
ASENSIO-CUBERO *ET AL.* [67]	0.2891	0.0703	0.0391*	0.0039**
NICOLAS-ALONSO *ET AL.* [33]	0.7266	0.2891	0.0078**	0.0039**
LDA	0.0391*	0.0391*	0.0039**	0.0039**
DSA	0.0703	0.5078	0.7266	0.0039**
MPMLDA (EQUAL RESP.)	0.0703	0.5078	0.5078	0.0039**
MPMLDA	0.2891	0.5078	0.2891	0.0039**
RLDA ($\eta = 0$)	-	0.0703	0.0391*	0.0039**
SRLDA ($\eta = 0$)	-	-	0.5078	0.0078**
RLDA ($\eta > 0$)	-	-	-	0.0039**

The time courses of the mean kappa value throughout evaluation session by RLDA and SRLDA with or without adaptive processing stage are shown in Fig. 4.

V. DISCUSSION AND CONCLUSION

In this paper, we addressed the problem of motor imagery classification. We studied the performance of an adaptive classifier ensemble that integrates temporal, spectral, and spatial information. The feature extraction stage employed a filter bank and the CSP algorithm. Performance analyses were conducted on the publicly available BCI Competition IV dataset 2a. Our findings suggest that stacked generalization and adaptive processing with EWMA are able to increase the classification performance of binary and multiclass MI-BCIs.

The power of SRLDA stems from its ability to model temporal discriminant features. It combines information from different time points to obtain additional information and deal with temporal variability in ERD/ERS patterns. Other studies evaluated dynamic classifiers such as HMM, CRF, or HCRF [40, 42] to exploit temporal information in experiments with synchronous BCI as well. While there is no complete consensus, HMMs were reported to provide comparable or even lower performance than static classifiers such as LDA [40, 42]. HMM is a generative approach that estimates the joint probability density function between class labels and the observed sequence of data samples. Although HMM is one of the most widely used classifier in the task of labeling multivariate time series [70], it requires making assumptions about independence of the data at each time point conditioned on the states, which are violated in practical BCI scenarios [40]. Discriminant approaches such as CRFs and HCRF,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2015.2398573, IEEE Transactions on Neural Systems and Rehabilitation Engineering

which directly model the class conditional distributions, were proposed to overcome this issue [40, 42]. Both approaches, particularly HCRF, perform better than HMM. However, their applicability in real-time BCI applications is limited by the high computational cost. SRLDA does not present such an inconvenient, since linear models are not as computationally demanding as computing the likelihoods of sequences.

Another important point related to temporal variability is the selection of the time point of good performance. The BCI Competition IV dataset 2a organizers did not take into account this issue. The performance was measured obtaining the highest kappa value across the evaluation session for each subject and averaging the results. Although we follow this criterion for comparison purposes, it leads to an optimistic measure. It does not imitate the real scenario in discrete motor imagery applications where the algorithm has to choose the time point of decision in advance. Then, we configure the time points of decision for RLDA ($\eta = 0$) and SRLDA ($\eta = 0$) based on 10-fold CV performance over the training session. The other parameters F, α_1, α_2, and Ω are set to the previously optimized values. The time point of decision is chosen based on the maximum mean kappa over the training session ($t_{\mathrm{RLDA}} = 4.52$ s and $t_{\mathrm{SRLDA}} = 5.68$ s). The same values are used for the adaptive counterparts. We assume that the optimal time point of decision is not largely affected by the update coefficient. RLDA ($\eta = 0$), SRLDA ($\eta = 0$), RLDA ($\eta > 0$), and SRLDA ($\eta > 0$) with a single time point of decision are tested on the evaluation session yielding mean kappa values of 0.57, 0.63, 0.64, and 0.72, respectively. They are lower than in Table III as expected, but the losses are slightly smaller with SRLDA. This suggests that SRLDA is less sensitive to the time point of decision. Interestingly, the time window of good performance of SRLDA is longer compared to RLDA (Fig. 4).

Another benefit of stacked generalization is that more stable correct output is produced. It is of great interest for real MI-BCI applications because it reduces the amount of spurious transitions between correct and incorrect outputs. Fig. 5 depicts the temporal evolution of RLDA ($\eta = 0$) and SRLDA ($\eta = 0$) accuracy for one trial. Label 1 means that the algorithm predicted correctly the task and label 0 implies the opposite. In view of the figure, RLDA ($\eta = 0$) output presents several transitions between correct and incorrect states. In contrast, the number of transitions is clearly reduced by SRLDA ($\eta = 0$). Although we show just one trial, similar results are found for the remainder of the trials. Classifiers that generate stable correct output avoid highly changing feedback, which can mislead the user.

Regarding the adaptive processing stage, our results show that it increases the classification performance in binary and multiclass settings. EWMA is able to estimate small fluctuations in global mean ameliorating intersession non-stationarity effects. Our findings agree with other previous studies on two-class [15, 24, 27] or multiclass [31, 32] BCIs. Our adaptive method is an extension of the pooled mean update proposed by Vidaurre *et al.* [24] to multiclass classification. The novelty is that our architecture can be used in multiclass settings and with other classifiers. Both would be

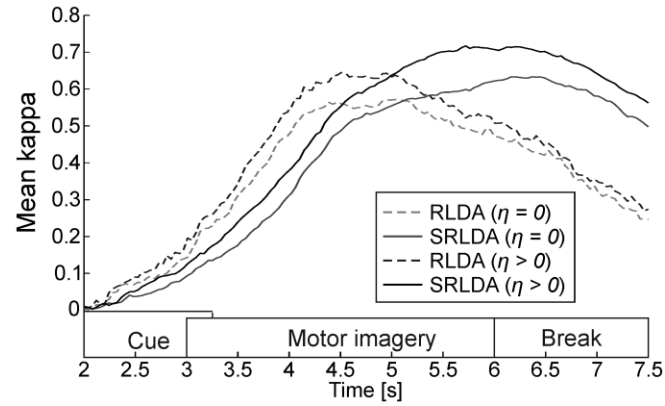


Fig. 4. Time course of the mean kappa value for RLDA and SRLDA with or without adaptive processing stage throughout evaluation session. Mean kappa value is computed obtaining the kappa time course for each subject and averaging these values across the subjects.

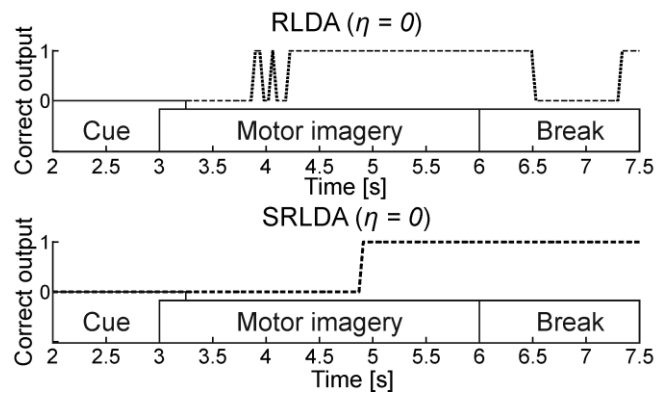


Fig. 5. Comparison between RLDA ($\eta = 0$) (upper plot) and SRLDA ($\eta = 0$) (bottom plot) accuracy over the time for one trial. Label 1 means the predicted task is correct whereas label 0 means an error is made.

quantitatively the same in the case of binary classification and non-regularized LDA.

The binary and multiclass comparative analyses amongst DSA, MPMLDA, and our adaptive approach show that none of them provides clearly superior performance. However, Llera *et al.* [32] stated that MPMLDA outperformed DSA in multiclass settings. The explanation for this discrepancy can be that Llera *et al.* tested MPMLDA on a different set of features extracted by tangent space mapping or the way in which we configure update coefficient by means of chronological validation on the training session. Leaving aside the quantitative results, DSA and our adaptive approach have the advantage that they can be used with any classifier. The computational cost needed to re-adjust other more costly classifiers sequentially throughout the evaluation sessions would be avoided.

EWMA presents the difficulty of the proper choice of the update coefficient η. With a large η, the estimated mean follows the features too truly presenting peaks, whereas, with a small η, peaks are suppressed but the changes in the real mean are followed too slowly by the estimation. Unlike Vidaurre *et al.* [24], who configured η with several datasets, we use chronological validation on training data to estimate η. It has the advantage that there is no need for recording EEG data in different days but performance might be affected by a

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2015.2398573, IEEE Transactions on Neural Systems and Rehabilitation Engineering

large difference between evaluation and train data. Thus, the selected η value might not be optimal anymore. We verify this issue configuring η with the training session and the first 15 trials of the evaluation session. The optimal update coefficients for RLDA ($\eta > 0$) and SRLDA ($\eta > 0$) are η = 0.08 and η = 0.11 yielding mean kappa values of 0.68 and 0.73, respectively. Wilcoxon's test reveals no significant differences for RLDA ($\eta > 0$) (p-value = 0.5078) or SRLDA ($\eta > 0$) (p-value = 0.8981).

The comparative analysis with other previous studies on the multiclass BCI Competition IV dataset 2a (Table III) shows that, although several studies have been conducted, there has been little improvement since the competition in 2008. Whereas Kam *et al.* [66] obtained a mean kappa of 0.60 in 2013, adaptive SRLDA increases mean kappa to 0.74. It can be also observed that LDA slightly outperforms the winner of the competition despite of both being very similar. The use of the LDA instead of Naïve Bayesian Parzen Window classifier or the different number of selected features are the reasons for such a difference. A Wilcoxon's test reveals no significant difference (p-value = 0.8448).

SRLDA requires adjusting five parameters by CV. It can raise questions concerning stability, overfitting, resilience to outliers, and the actual usability of the system in real life applications. Fig. 2 shows the performance is little sensitive to the adjustable parameters. Optimal values of F, α_1, α_2, and Ω are in the flattest part of the optimization curves. It is true that update coefficient seems to affect the performance more than other parameters. Nevertheless, as can be seen in Fig. 3, the improvement is notable with respect to the static classification ($\eta = 0$) for a wide range of values.

The risk of overfitting is assessed comparing the performances in training and evaluation sessions. RLDA (η = 0) and SRLDA (η = 0) yield mean kappa values of 0.72 and 0.78 on 10-fold CV training session, respectively. These values are considerably higher than the respective mean kappa values of 0.62 and 0.68 on evaluation session, which would suggest overfitting. The number of configurable parameters do not seem to be an important reason as both RLDA ($\eta = 0$) and SRLDA (η = 0) are equally affected. The performances of RLDA ($\eta > 0$) and SRLDA ($\eta > 0$), which yield mean kappa values of 0.68 and 0.74 on evaluation session, point to the intersession non-stationarity as the main cause of such performance reduction. Furthermore, RLDA and SRLDA incorporate regularization, which should protect against overfitting to some extent [55]. On the other hand, the performance reduction may be also caused by CSP, which is prone to overfitting [71].

In order to evaluate the resilience to outliers, we repeat multiclass experiments training RLDA and SRLDA models without discarding the training trials marked as artifacts during the BCI Competition. RLDA ($\eta = 0$), SRLDA ($\eta = 0$), RLDA ($\eta > 0$), and SRLDA ($\eta > 0$) produce mean kappa values of 0.63, 0.67, 0.69, and 0.75, respectively. Wilcoxon's test reveals no significant differences with respect to rejecting invalid trials (p-values = 0.5078, 0.2891, 0.2891, and 0.7266).

Regarding real-life usage, the time needed for configuring adaptive or static SRLDA is about 45 min for each subject using a PC with an Intel Core i7-2600 @ 3.40 GHz processor and 16 GB RAM. The training time for each subject, including feature extraction and selection, is less than 3 minutes. Although SRLDA cannot be configured in few minutes, given the stability of the configurable parameters and the adaptive nature of our method, there could be no need for reconfiguration before every session.

Some limitations of this research have to be considered. Firstly, it is important to note that binary and multiclass results present large variations amongst BCI users. Therefore, the results should be cautiously interpreted. Secondly, although stacking generalization is able to increase the classification performance, there may be even more room for improvement. Our approach selects a single time point of decision for task prediction. Performance could be further improved by methods that adaptively compute this point for each trial. Further analyses should be carried out with new and larger datasets to test the usefulness of our methodology prospectively. In addition, future work should test methods that in an unsupervised way evaluate classification output reliability to compute the optimal time of decision adaptively throughout the evaluation session.

In summary, the current study presents a new methodology that addresses intersession non-stationarity and temporal variability. Spatial and spectral information are also considered to characterize EEG signals. Our findings suggest that both binary and multiclass MI-BCI performance can be significantly increased with either combined or separated use of adaptive processing and stacking generalization. Adaptive processing stage reduces the mismatch between sessions before classification whereas stacked generalization is able to handle temporal, spatial, and spectral information, and cope with temporal variability. In binary and multiclass motor imagery experiments, adaptive SRLDA significantly outperforms the state-of-the-art methods. Adaptive SRLDA can serve to design more reliable and robust MI-BCIs, which are of great interest for speeding up communication in real-life applications.

REFERENCES

[1] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors,* vol. 12, pp. 1211-1279, Jan. 2012.

[2] F. H. Guenther, J. S. Brumberg, E. J. Wright, A. Nieto-Castanon, J. A. Tourville, M. Panko, *et al.*, "A wireless brain-machine interface for real-time speech synthesis," *PloS one,* vol. 4, p. e8218, 2009.

[3] R. Corralejo, L. F. Nicolás-Alonso, D. Álvarez, and R. Hornero, "A P300-based brain–computer interface aimed at operating electronic devices at home for severely disabled people," *Med. Biol. Eng. Comput.,* vol. 52, pp. 861-872, 2014.

[4] G. R. Muller-Putz and G. Pfurtscheller, "Control of an electrical prosthesis with an SSVEP-based BCI," *IEEE Trans. Biomed. Eng.,* vol. 55, pp. 361-364, 2008.

[5] H. Yuan and B. He, "Brain-Computer Interfaces Using Sensorimotor Rhythms: Current State and Future Perspectives," *IEEE Trans. Biomed. Eng.,* 2014.

[6] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clin. Neurophysiol.,* vol. 110, pp. 1842-1857, Nov. 1999.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2015.2398573, IEEE Transactions on Neural Systems and Rehabilitation Engineering

[7] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer, "EEG-based discrimination between imagination of right and left hand movement," *Electroencephalogr. Clin. Neurophysiol.,* vol. 103, pp. 642-651, Dec. 1997.

[8] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Muller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," *IEEE Trans. Biomed. Eng.,* vol. 51, pp. 993-1002, Jun. 2004.

[9] G. Townsend, B. Graimann, and G. Pfurtscheller, "A comparison of common spatial patterns with complex band power features in a four-class BCI experiment," *IEEE Trans. Biomed. Eng.,* vol. 53, pp. 642-651, Apr. 2006.

[10] C. Gouy-Pailler, M. Congedo, C. Brunner, C. Jutten, and G. Pfurtscheller, "Nonstationary brain source separation for multiclass motor imagery," *IEEE Trans. Biomed. Eng.,* vol. 57, pp. 469-478, Feb. 2010.

[11] H. I. Suk and S. W. Lee, "Subject and class specific frequency bands selection for multiclass motor imagery classification," *Int. J. Imag. Syst. Tech.,* vol. 21, pp. 123-130, Feb. 2011.

[12] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter Bank Common Spatial Pattern algorithm on BCI Competition IV Datasets 2a and 2b," *Front. Neurosci.,* vol. 6, Mar. 2012.

[13] J. Asensio-Cubero, J. Q. Gan, and R. Palaniappan, "Extracting optimal tempo-spatial features using local discriminant bases and common spatial patterns for brain computer interfacing," *Biomed. Signal. Proces.,* vol. 8, pp. 772-778, Nov. 2013.

[14] N. Robinson, C. Guan, A. Vinod, K. K. Ang, and K. P. Tee, "Multi-class EEG classification of voluntary hand movement directions," *J. Neural Eng.,* vol. 10, p. 056018, May. 2013.

[15] P. Shenoy, M. Krauledat, B. Blankertz, R. P. N. Rao, and K. R. Müller, "Towards adaptive classification for BCI," *J. Neural Eng.,* vol. 3, p. R13, Mar. 2006.

[16] G. Liu, G. Huang, J. Meng, D. Zhang, and X. Zhu, "Improved GMM with parameter initialization for unsupervised adaptation of brain–computer interface," *Int. J. Numer. Meth. Bio-Med. Eng.,* vol. 26, pp. 681-691, Jun. 2010.

[17] J. Blumberg, J. Rickert, S. Waldert, A. Schulze-Bonhage, A. Aertsen, and C. Mehring, "Adaptive classification for brain computer interfaces," in *Proc. IEEE Eng. Med. Biol. Soc.*, Lyon, France, 2007, pp. 2536-2539.

[18] Y. Li and C. Guan, "An extended EM algorithm for joint feature extraction and classification in brain-computer interfaces," *Neural Comput.,* vol. 18, pp. 2730-2761, Nov. 2006.

[19] A. Llera, V. Gomez, and H. J. Kappen, "Adaptive Classification on Brain-Computer Interfaces Using Reinforcement Signals," *Neural Comput.,* vol. 24, pp. 2900-2923, Nov 2012.

[20] X. Liao, D. Yao, and C. Li, "Transductive SVM for reducing the training effort in BCI," *J. Neural Eng.,* vol. 4, p. 246, Sep. 2007.

[21] J. Qin, Y. Li, and W. Sun, "A semisupervised support vector machines algorithm for BCI systems," *Comput. Intell. Neurosci.,* vol. 2007, Jul. 2007.

[22] Y. Li, C. Guan, H. Li, and Z. Chin, "A self-training semi-supervised SVM algorithm and its application in an EEG-based brain computer interface speller system," *Pattern Recogn. Lett.,* vol. 29, pp. 1285-1294, Jul. 2008.

[23] G. Zhenghui, Y. Zhuliang, S. Zhifang, and L. Yuanqing, "An Online Semi-supervised Brain-Computer Interface," *IEEE Trans. Biomed. Eng.,* vol. 60, pp. 2614-2623, Sep. 2013.

[24] C. Vidaurre, M. Kawanabe, P. von Bünau, B. Blankertz, and K. Muller, "Toward unsupervised adaptation of LDA for brain–computer interfaces," *IEEE Trans. Biomed. Eng.,* vol. 58, pp. 587-597, Mar. 2011.

[25] M. Sugiyama, M. Krauledat, and K.-R. Müller, "Covariate shift adaptation by importance weighted cross validation," *J. Mach. Learn. Res.,* vol. 8, pp. 985-1005, Jul. 2007.

[26] Y. Li, H. Kambara, Y. Koike, and M. Sugiyama, "Application of covariate shift adaptation techniques in brain–computer interfaces," *IEEE Trans. Biomed. Eng.,* vol. 57, pp. 1318-1324, Jun. 2010.

[27] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "EEG Data Space Adaptation to Reduce Intersession Nonstationarity in Brain-Computer Interface," *Neural Comput.,* vol. 25, pp. 2146-2171, Aug. 2013.

[28] P. Sykacek, S. J. Roberts, and M. Stokes, "Adaptive BCI based on variational Bayesian Kalman filtering: an empirical evaluation," *IEEE Trans. Biomed. Eng.,* vol. 51, pp. 719-727, May. 2004.

[29] J. W. Yoon, S. J. Roberts, M. Dyson, and J. Q. Gan, "Adaptive classification for brain computer interface systems using sequential Monte Carlo sampling," *Neural Networks,* vol. 22, pp. 1286-1294, Jun. 2009.

[30] J. W. Yoon, S. J. Roberts, M. Dyson, and J. Q. Gan, "Bayesian inference for an adaptive Ordered Probit model: An application to Brain Computer Interfacing," *Neural Networks,* vol. 24, pp. 726-734, Sep. 2011.

[31] P. Xu, P. Yang, X. Lei, and D. Yao, "An enhanced probabilistic LDA for multi-class brain computer interface," *PloS One,* vol. 6, p. e14634, Jan. 2011.

[32] A. Llera, V. Gómez, and H. Kappen, "Adaptive Multiclass Classification for Brain Computer Interfaces," *Neural Comput.,* vol. 26, pp. 1108-1127, 2014.

[33] L. Nicolas-Alonso, R. Corralejo, D. Álvarez, and R. Hornero, "Adaptive Classification Framework for Multiclass Motor Imagery-Based BCI," in *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013*, ed: Springer International Publishing, 2014, pp. 762-765.

[34] E. V. C. Friedrich, R. Scherer, and C. Neuper, "Stability of event-related (de-) synchronization during brain–computer interface-relevant mental tasks," *Clin. Neurophysiol.,* vol. 124, pp. 61-69, Jan. 2013.

[35] G. Townsend, B. Graimann, and G. Pfurtscheller, "Continuous EEG classification during motor imagery-simulation of an asynchronous BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.,* vol. 12, pp. 258-265, Feb. 2004.

[36] N. F. Ince, A. H. Tewfik, and S. Arica, "Extraction subject-specific motor imagery time–frequency patterns for single trial EEG classification," *Comput. Biol. Med.,* vol. 37, pp. 499-508, Apr. 2007.

[37] B. Wu, F. Yang, J. Zhang, Y. Wang, X. Zheng, and W. Chen, "A frequency-temporal–spatial method for motor-related electroencephalography pattern recognition by comprehensive feature optimization," *Comput. Biol. Med.,* vol. 42, pp. 353-363, Apr. 2012.

[38] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Mutual information-based selection of optimal spatial–temporal patterns for single-trial EEG-based BCIs," *Pattern Recognition,* vol. 45, pp. 2137-2144, Jun. 2012.

[39] B. Obermaier, C. Guger, C. Neuper, and G. Pfurtscheller, "Hidden Markov models for online classification of single trial EEG data," *Pattern Recogn. Lett.,* vol. 22, pp. 1299-1309, Oct. 2001.

[40] J. F. D. Saa and M. Çetin, "A latent discriminative model-based approach for classification of imaginary motor tasks from EEG data," *J. Neural Eng.,* vol. 9, p. 026020, Apr. 2012.

[41] B. Obermaier, C. Neuper, C. Guger, and G. Pfurtscheller, "Information transfer rate in a five-classes brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.,* vol. 9, pp. 283-288, Mar. 2001.

[42] B. A. S. Hasan and J. Q. Gan, "Conditional random fields as classifiers for three-class motor-imagery brain–computer interfaces," *J. Neural Eng.,* vol. 8, p. 025013, Mar. 2011.

[43] N. C. Oza and K. Tumer, "Classifier ensembles: Select real-world applications," *Information Fusion,* vol. 9, pp. 4-20, Jan. 2008.

[44] L. Kuncheva, "Combining pattern classifiers: methods and algorithms," ed: John Wiley & Sons, 2004.

[45] G. Pfurtscheller, D. Flotzinger, and J. Kalcher, "Brain-Computer Interface—a new communication device for handicapped persons," *J. Microcomput. Appl.,* vol. 16, pp. 293-299, Jul. 1993.

[46] A. Rakotomamonjy and V. Guigue, "BCI competition III: dataset II-ensemble of SVMs for BCI P300 speller," *IEEE Trans. Biomed. Eng.,* vol. 55, pp. 1147-1154, Mar. 2008.

[47] Z. S. Shirehjini, S. B. Shouraki, and M. Esmailee, "Variant Combination of Multiple Classifiers Methods for Classifying the EEG Signals in Brain-Computer Interface," in *Advances in Computer Science and Engineering*, ed: Springer, 2009, pp. 477-484.

[48] C. Chun-Hsiang, K. Li-Wei, L. Yuan-Pin, J. Tzyy-Ping, and L. Chin-Teng, "Independent Component Ensemble of EEG for Brain Computer Interface," *IEEE Trans. Neural Syst. Rehabil. Eng.,* vol. 22, pp. 230-238, Feb. 2014.

[49] O. AlZoubi, I. Koprinska, and R. A. Calvo, "Classification of brain-computer interface data," in *Proc. of the 7th Australasian Data Min. Conf.*, Glenelg, South Australia, 2008, pp. 123-131.

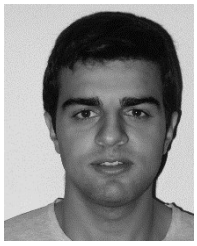
[50] B. Reza and M. Mohammad Hassan, "A new approach in the BCI research based on fractal dimension as feature and Adaboost as classifier," *J. Neural Eng.,* vol. 1, p. 212, Dec. 2004.

[51] J. R. White, T. Levy, W. Bishop, and J. D. Beaty, "Real-Time Decision Fusion for Multimodal Neural Prosthetic Devices," *PLoS One,* vol. 5, p. e9493, Mar. 2010.

[52] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, *et al.*, "Review of the BCI Competition IV," *Front. Neurosci.,* vol. 6, p. 55, Jul. 2012.

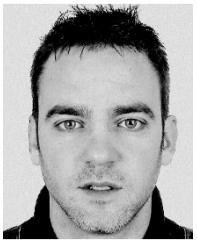
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TNSRE.2015.2398573, IEEE Transactions on Neural Systems and Rehabilitation Engineering

[53] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Proc. Mag.,* vol. 25, pp. 41-56, Jan. 2008.
[54] J. M. Lucas and M. S. Saccucci, "Exponentially weighted moving average control schemes: properties and enhancements," *Technometrics,* vol. 32, pp. 1-12, Feb. 1990.
[55] D. Cai, X. He, and J. Han, "SRDA: An efficient algorithm for large-scale discriminant analysis," *IEEE Trans. Knowl. Data En.,* vol. 20, pp. 1-12, Jan. 2008.
[56] D. P. Burke, S. P. Kelly, P. De Chazal, R. B. Reilly, and C. Finucane, "A parametric feature extraction and classification strategy for brain-computer interfacing," *IEEE Trans. Neural Syst. Rehabil. Eng.,* vol. 13, pp. 12-17, Jan. 2005.
[57] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, "Single-trial analysis and classification of ERP components — a tutorial," *NeuroImage,* vol. 56, pp. 814-825, May. 2011.
[58] Z. Jinyin, G. Sudre, L. Xin, W. Wei, D. J. Weber, and A. Bagic, "Clustering Linear Discriminant Analysis for MEG-Based Brain Computer Interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.,* vol. 19, pp. 221-231, Mar. 2011.
[59] J. H. Friedman, "Regularized discriminant analysis," *Journal of the American statistical association,* vol. 84, pp. 165-175, Oct. 1989.
[60] D. H. Wolpert, "Stacked generalization," *Neural Networks,* vol. 5, pp. 241-259, Jul. 1992.
[61] K. M. Ting and I. H. Witten, "Issues in stacked generalization," *J. Artif. Int. Res.,* vol. 10, pp. 271-289, Oct. 1999.
[62] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann, 2005.
[63] H. Wang, "Multiclass filters by a weighted pairwise criterion for EEG single-trial classification," *IEEE Trans. Biomed. Eng.,* vol. 58, pp. 1412-1420, May. 2011.
[64] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass Brain–Computer Interface Classification by Riemannian Geometry," *IEEE Trans. Biomed. Eng.,* vol. 59, pp. 920-928, Apr. 2012.
[65] D. Wang, D. Miao, and G. Blohm, "Multi-class motor imagery EEG decoding for brain-computer interfaces," *Front. Neurosci.,* vol. 6, Oct. 2012.
[66] T.-E. Kam, H.-I. Suk, and S.-W. Lee, "Non-homogeneous spatial filter optimization for ElectroEncephaloGram (EEG)-based motor imagery classification," *Neurocomputing,* Jan. 2013.
[67] J. Asensio-Cubero, J. Gan, and R. Palaniappan, "Multiresolution analysis over simple graphs for brain computer interfaces," *J. Neural Eng.,* vol. 10, p. 046014, Jul. 2013.
[68] G. Dornhege, J. d. R. Millán, T. Hinterberger, D. McFarland, and K.-R. Müller, *Toward brain-computer interfacing* vol. 74. Boston, MA: MIT press Cambridge, 2007.
[69] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K.-R. Muller, "Combined optimization of spatial and temporal filters for improving brain-computer interfacing," *IEEE Trans. Biomed. Eng.,* vol. 53, pp. 2274-2281, Nov. 2006.
[70] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE,* vol. 77, pp. 257-286, Feb. 1989.
[71] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms," *IEEE Trans. Biomed. Eng.,* vol. 58, pp. 355-362, 2011.



Luis F. Nicolas-Alonso (S'13) was born in Valladolid, Spain in 1988. He received the M.S. degree in telecommunication engineering, in 2011, from the University of Valladolid (Spain), where he is currently working toward the Ph.D. degree in the Department of Signal Theory and Communications and is a scholarship holder. He is a member of the Biomedical Engineering Group. His current research interests include signal processing and pattern recognition techniques applied to brain computer interfaces.

Rebeca Corralejo (S'10) was born in Burgos, Spain, in 1983. He received the M.S. degree in telecommunication engineering, in 2008, from the University of Valladolid (Spain), where she is currently working toward the Ph.D. degree in the Department of Signal Theory and Communications and is a scholarship holder. She is a member of the Biomedical Engineering Group. Her current research interests include biomedical signal processing applied to brain computer interfaces and development of assistive applications.



Javier Gomez-Pilar was born in Alicante, Spain in 1983. He received the M.S. degree in telecommunication engineering, in 2012, from the University of Valladolid, Valladolid, Spain, where he is currently working toward the Ph.D. degree in the Department of Signal Theory and Communications. He is a Researcher at the Biomedical Engineering Group. His current research interests include signal processing, complex network theory, and rehabilitation techniques using brain computer interfaces.

Daniel Alvarez (M'12) was born in Bembibre, Spain, in 1978. He received the M.S. degree in telecommunication engineering and the Ph.D. degree from the University of Valladolid (Spain) in 2005 and 2011, respectively. Since 2005, he is a member of the Biomedical Engineering Group of the University of Valladolid. He is currently a researcher in the Department of Signal Theory and Communications of the University of Valladolid. His research interests include multivariate analysis and pattern recognition of biomedical signals. His work focuses on the development of novel methodologies aimed at assisting in the diagnosis of sleep disorders, as well as on the design of new assistive tools based on brain computer interfaces for disabled/dependent people.

Roberto Hornero (M'04–SM'11) was born in Plasencia, Spain, in 1972. He received the M.S. degree in telecommunication engineering and the Ph.D. degree from the University of Valladolid (Spain), in 1995 and 1998, respectively. He is currently Professor in the Department of Signal Theory and Communications at the University of Valladolid. His main research interest is spectral and nonlinear analysis of biomedical signals to help physicians in the clinical diagnosis. He founded the Biomedical Engineering Group in 2004. The research interests of this group are connected with the field of nonlinear dynamics, chaotic theory, and wavelet transform with applications in biomedical signal and image processing.