



Motor imagery task classification using transformation based features



Aida Khorshidtalab*, Momoh J.E. Salami, Rini Akmeliawati

Intelligent Mechatronics System Research Unit, Department of Mechatronics Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:

Received 8 June 2016

Received in revised form 4 December 2016

Accepted 6 December 2016

Keywords:

EEG

Linear prediction coding

QR decomposition

Singular value decomposition

Channel selection

ABSTRACT

This paper proposes a feature extraction method named as LP-QR, based on the decomposition of the LPC filter impulse response matrix of the signal of interest. This feature extraction method is inspired by LP-SVD and is tested in the context of motor imagery electroencephalogram. The extracted features are classified and benchmarked against extracted features of LP-SVD method. The two applied methods are also compared regarding the required execution time, which further highlights their respective merits and demerits. This paper closely examines the contribution of EEG channels of these two information extraction algorithms too. Consequently, a detailed analysis of the role of EEG channels concerning the nature of the extracted information is presented. This study is conducted on the BCI IIIa competition database of four motor imagery movements. The obtained results indicate that the proposed method is the better choice if simplicity is demanded. The investigation into the role of EEG channels reveals that level of contribution each channel can be quite dissimilar for different feature extraction algorithms.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

The purpose of motor imagery EEG signal processing is to create a direct link between the brain and an external electronic device. This link, known as Brain-Computer Interface (BCI), translates brain signals into meaningful commands for the use of external device [1,2]. There are two major criteria, based on which this link is evaluated; first, how fast is the link at translating the commands and second, how accurate the translation is. The main part of the link, which is also the focus of this study, is the processing part; where brain signals are acquired and processed. Thereafter, decisions are made by correlating or assigning the commands to specific signals. In most BCI systems, the decision about which commands the processed signals point to is made by classification algorithms. Evidently, better quality of acquired signals in terms of signal to noise ratio makes the process of decision-making a much easier task. In the same manner, a strong processing method which is capable of highlighting the most informative part of signals can help the decision-making unit just as much. The processing mainly includes feature/information extraction. Among many proposed methods and algorithms, common spatial pattern (CSP) has been proved to have superior performance. CSP [3] produces a transformation with a linear spatial filter that maximizes the variance for one class while

minimizing the variance of the other class. The difference in variance is used as an indication to classify EEG signals into one of two classes. However, CSP is associated with several drawbacks, including the optimality of minimal classification error, the requirement of the method for fine tuning, which is time-consuming, and the fact that CSP is designed for a two-class paradigm. The optimality of CSP for minimal classification error is still an open question

which leaves room for other methods to compete with this method. Moreover, CSP is also known to be very sensitive to noise and prone to overfitting [4]. The first winner group of the BCI competition III for database IIIa provided excellent results using the CSP method [5]. However, it is mentioned in their report that the results were obtained with the help of fine tuning of all parameters. They also applied multi-class CSP in the form of a two-class paradigm with one-against-rest method. There can be many extensions to CSP to make it capable of dealing with multiclass paradigms, which are based on a two-step procedure: 1) computation of a set of potential spatial filters and 2) selection of a subset of these filters [4,6–10]. The selection of spatial filters is mostly evaluated experimentally, which makes it a slow process. In the few cases that the proposed extensions are not heuristic based and are not evaluated experimentally, the computational complexity of the CSP algorithm is considerably increased [11]. To overcome some of the CSP drawbacks, [12] proposed a transformed method referred to as LP_SVD (Linear Prediction in conjunction with Singular Value Decomposition), which needed no parameter tuning and no extension or modification for multiclass paradigms [13]. The only drawback of

* Corresponding author.

E-mail address: ida.khorshidtalab@gmail.com (A. Khorshidtalab).

the proposed method was its computational complexity. Therefore, we came up with the idea of LP_QR (Linear Prediction in conjunction with QR Decomposition), which is inspired by LP_SVD [13]. The merit of the proposed method is that LP_QR is remarkably faster while maintaining the advantages of LP_SVD. In addition to accuracy and time consumption, these two methods are specifically compared with regards to channel selection. The study of channel selection and the role of active EEG channels of these two methods, which are closely related, reveal interesting insights from information extraction point of view.

The rest of paper is organized as follows. Section II.A describes the EEG signals from database IIIa, competition III. Section II.B introduces the LP_QR transform and compares it with LP_SVD. Section III includes the results, performance analysis, and discussion. Finally, section IV concludes the paper.

2. Material and methods

2.1. Material

Dataset IIIa from the BCI competition III (2005) [14], a widely benchmarked database of multiclass motor imagery movements, is used for evaluation and comparison. This database includes three subjects named as K3b, K6b, and L1b.

The EEG signals were recorded using a 64-channel Neuroscan EEG amplifier. The recording involves only 60 EEG channels. The recording is based on 10–20 standard montage systems. The recorded signals were sampled at 250 Hz and filtered using a 1–50 Hz bandpass filter. Additionally, a notch filter was applied to suppress the interference from the power line.

Each subject was asked to perform the imagery movements associated with the provided visual cues for several trials. Each trial started with an empty black screen at $t = 10$ s; at $t = 2$ s, a short beep tone was presented and a cross '+' appeared on the screen; at $t = 3$ s, an arrow pointing to one of the four main directions was presented for four seconds, disappearing at $t = 7$ s. No feedback was provided to the subject.

The recorded dataset from subject K3b consists of nine runs while for K6b and L1b consist of only six runs. Each run consists of four cues displayed ten times in a random order, resulting in 360 trials for subject K3b and 240 trials for each of the other two subjects.

2.2. Methods

This section is divided into four subsections. The first subsection explains LP_QR transform and compares it with LP_SVD. The second describes the extracted features. The third subsection is about the channel selection procedure, and the last one outlines the applied classifier.

2.2.1. LP_QR transform

QR decomposition is a form of eigenvalue algorithm which is widely used for solving linear least squares problems. The LP_QR transform requires the estimation of LPC filter coefficients plus the QR decomposition of LPC filter impulse response matrix. This computation results in Q_d and R_d matrices, where Q_d is an orthogonal matrix and R_d is an upper triangular matrix.

2.2.2. LPC filter

Linear prediction (LP) [15] is an effective, practical method used in many engineering disciplines, particularly in the field of digital signal processing. LP, in time domain, predicts the current value of

the signal, based on a linear combination of its past samples, which is mathematically expressed as:

$$y(n) = -\sum_{i=1}^P a_i y(n-i) + e(n) \quad (1)$$

where, a_i are the linear prediction coefficients (LPCs), P is the prediction order and $e(n)$ is the prediction error. Eq. (1) can be expressed in a form of matrix notations as follows.

$$\mathbf{y} = \mathbf{H}\mathbf{e} \quad (2)$$

where \mathbf{y} is $N \times 1$ column vectors of the signal samples, \mathbf{e} is the prediction error and \mathbf{H} is a lower triangular and Toeplitz matrix based on the linear prediction coefficients a_i from (1). The \mathbf{H} matrix, the $N \times N$ impulse response matrix of the LPC filter, is in the following form:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ h(1) & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ h(N-1) & h(N-2) & \cdots & 1 \end{bmatrix} \quad (3)$$

where

$$h(n) = \begin{cases} \delta(n) - \sum_{i=1}^P a_i h(n-i), & \text{for } 0 \leq n \leq N-1 \\ 0, & \text{elsewhere} \end{cases}$$

and $\delta(n)$ is the discrete Dirac function.

QR decomposition is not directly applied to the EEG data, instead it is applied to the \mathbf{H} matrix which gives:

$$\mathbf{H} = \mathbf{Q}_d \mathbf{R}_d^T \quad (4)$$

Therefore, in conjunction with Eq. (2), \mathbf{y} can be defined as:

$$\mathbf{y} = \mathbf{Q}_d \mathbf{R}_d^T \mathbf{e} \quad (5)$$

Thus, the transformation that maps the data vector \mathbf{y} to the feature vector $\boldsymbol{\theta}$ is defined as:

$$\boldsymbol{\theta} = \mathbf{Q}_d^T \mathbf{y} \quad (6)$$

This transformation, $(\mathbf{Q}_d^T \mathbf{y})$, compacts a large portion of the signal energy in a few coefficients.

LP_QR transform, algorithm flow:

1. Find the linear prediction coefficient with Eq. (1).
2. Form matrix \mathbf{H} , which is an $N \times N$ impulse response matrix of the LPC filter with Eq. (3).
3. Apply QR decomposition to the \mathbf{H} matrix.
4. The transformed signal, $\boldsymbol{\theta}$, is the result of Eq. (6).

2.2.3. Extracted features

To have a fair comparison between the two proposed methods, the features that contributed to improving the classification accuracy of LP_SVD method are considered for LP_QR too. These features are the LP coefficients (a_i), the prediction error variance (V_r), a subset of significant transform coefficients of $\boldsymbol{\theta}$, Q-statistic and Hotelling's T^2 statistic of the transformed signal. It should be noted that the Q-statistic and Hotelling's T^2 statistic features are modified and adjusted based on the LP_QR transformation.

2.2.4. LPC related features

LP analysis describes the signal in the form of all-poles filter coefficients and prediction error [15]. Once the coefficients are estimated, the prediction error sequence can be computed using Eq.

(1). The prediction error $e(n)$ variance estimation, based on the autocorrelation method is as follows:

$$Vr = \frac{1}{N-1} \sum_{n=1}^N (e(n) - \bar{e})^2, \quad (7)$$

where \bar{e} is the arithmetic mean of the prediction error vector e and N is its length. The autocorrelation method secures the stability of the filter.

2.2.5. Transform coefficients or scores $\hat{\theta}$

The data vector y is presented in the new coordinates q_i by the transform coefficients or scores θ_i . The transform coefficients corresponding to the k largest q values are selected as features. The parameter k has been selected to be four to have a just comparison of these two methods. These four values are the first four values in Fig. 1(b) for LP_SVD and 1(c) for LP_QR marked in black color for a randomly selected signal of subject L1b.

$$\hat{\theta} = \hat{Q}_d y, \text{ The columns of } \hat{Q}_d \text{ are } \{q_1, q_2, \dots, q_k\} \quad (8)$$

2.2.6. Q-statistic

Q-statistic [16], also referred to as the squared prediction error, estimates the error between the vector y and \hat{y} . Where \hat{y} is defined as follow:

$$\hat{y} = \hat{Q}_d \hat{Q}_d^T y, \text{ where } \hat{Q}_d = (q_1, q_2, \dots, q_r) \quad (9)$$

Then the approximation error vector is mathematically expressed as $d = y - \hat{y}$. Therefore, Q-statistic with regards to QR decomposition can be expressed as:

$$Q = y - \hat{y}^2 = y^T \left(I - \hat{Q}_d \hat{Q}_d^T \right) y \quad (10)$$

2.2.7. Hotelling's T^2 statistics

The Hotelling's T^2 statistic measures the overall variability in the transformed data vector. It is defined as the sum of the transform coefficients weighted by σ_i . These values, arranged in descending order, are applied to assign suitable importance to the different elements of the vector θ . To benchmark reasonably with the previous study [13], r is chosen to be 10% of the number of vectors q_i .

$$T^2 = \sum_{i=1}^r \frac{\theta_i^2}{\sigma_i} \quad (11)$$

Eq. (11) can be expressed in a more compact form as:

$$T^2 = y^T \hat{Q}_d \hat{D}^{-1} \hat{Q}_d^T y \quad (12)$$

where \hat{D} is a $r \times r$, diagonal matrix, containing the r largest values of H in descending order.

2.2.8. Channel selection

The applied channel selection in this study, was proposed in [13]. In the channel selection technique used in [13], the classifier is part of the algorithm. The only difference between what is practiced here and in [13] in terms of channel selection, is that instead of extracting features from all the possible channels (stage 2 of the algorithm), features are extracted from only those channels that are more likely to contribute positively to the final result. In other words, features are extracted only from the active channels in the presented map in [5], for each subject. There are advantages and disadvantages in this regard. The benefit of this approach is that it is not as time-consuming as extracting features for all possibilities and testing each one of them with the classifier. While

Table 1
Accuracy of different AR order.

AR model order	Subject			Average
	L1b	K3b	K6b	
	Accuracy (%)			
1	41.25	63.05	34.16	46.15
2	37.08	63.05	31.66	43.93
3	41.25	60.83	32.91	44.99
4	36.25	60.72	30.83	42.60
5	37.50	58.88	31.25	42.54
6	37.50	58.88	32.08	42.82
7	36.66	56.66	36.25	43.19
8	35.41	54.44	34.58	41.47
9	37.50	56.11	33.33	42.31
10	34.16	56.38	35.83	42.12

there is a drawback to this shortcut, there might be some channels that could contribute positively to the final result, but they are not being considered. The argument for this approach is that if the presented map in the [5] is comprehensive, there could be a very narrow chance that existing channels that could contribute significantly are ignored. In short, the convenience of this approach is much greater than its possible drawback.

2.2.9. Classification

The applied classifier in this study is logistic model tree classifier from the Weka package [17]. Evaluation of the classifier is based on 10-fold cross-validation, which means the final results are averaged over 10 times of evaluation in a leave-one-out manner. Therefore, the obtained accuracy is the average accuracy of 10 distinct evaluations on each dataset for each subject.

3. Result and discussion

3.1. Model order selection

The model order selection process for LP_SVD was based on the highest accuracy that a subset of transform coefficients, the LP coefficients, and the prediction error variance together could obtain for different orders. The order in [13] is varied from one to seven and the obtained results revealed that the best model order was the first order as it could obtain the highest accuracy for all the subjects. The same procedure is followed for finding the best order for LP_QR. The difference is only in the subset of transform coefficients. Subject L1b and K3b have their highest obtained accuracy at first order, but K6b does not follow the same trend. According to Table 1, subject K6b obtained his highest accuracy at order 7. Subsequently, we investigated the best model order up to order 10 for all three subjects. The best order for K6b remains the seventh order. As subject K6b is not in agreement with the other two subjects, the average accuracy is considered for selecting the best model order. The highest average accuracy is obtained for the first order. Therefore, this order has been selected for LP_QR processing too.

3.2. General results

In this section, the obtained results are evaluated and benchmarked in three main categories: 1) accuracy, 2) execution time and 3) channels' contribution.

3.2.1. Accuracy

Regarding accuracy, the two methods are compared when, 1) only the three known central channels, C3–C4–Cz, are considered, 2) all 60 channels are considered, and 3) the selected channels through channel selection are considered.

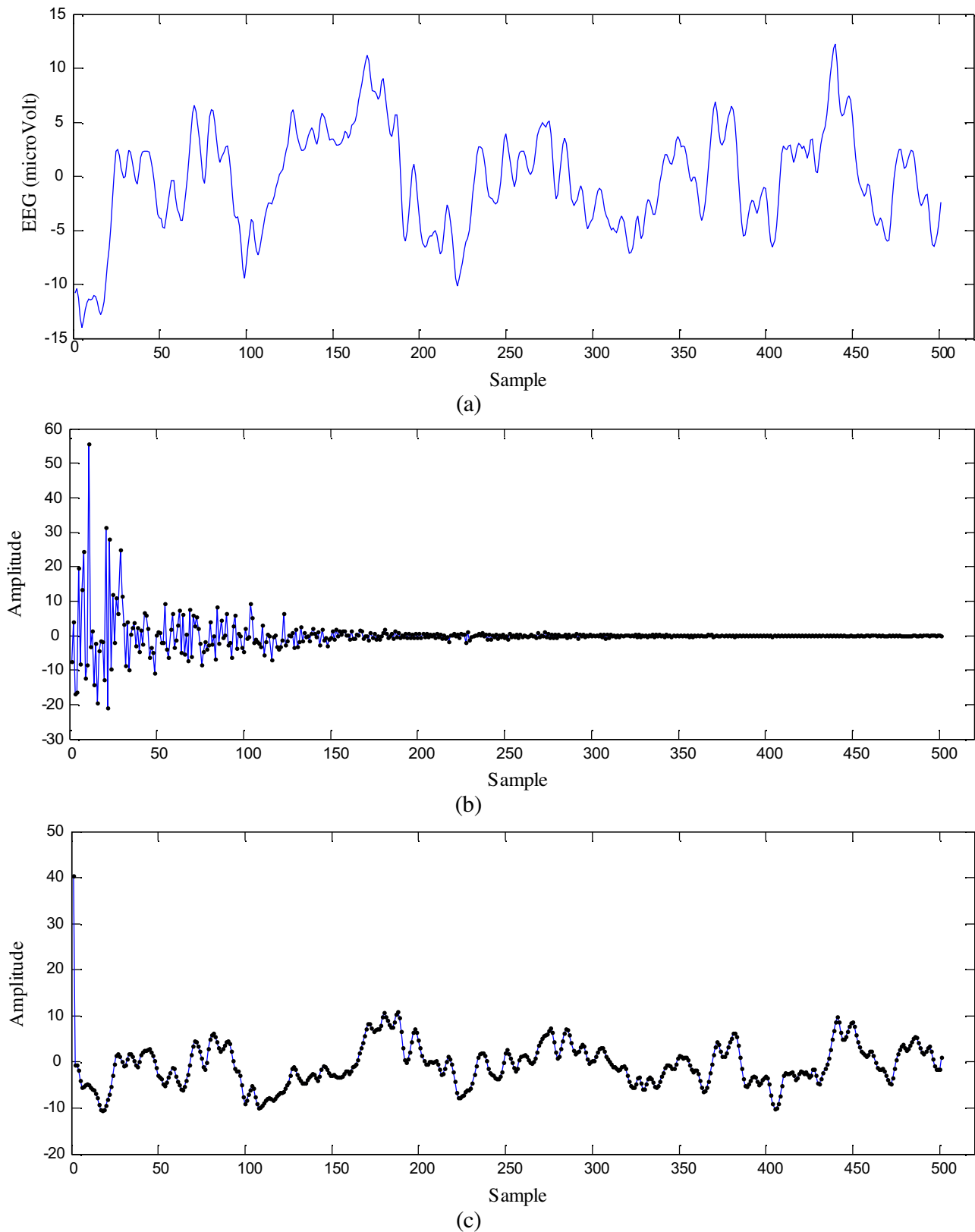


Fig. 1. Signal transformation, (a) Original EEG signal from subject L1b, (b) Transform coefficients with AR(1) and LP_SVD, (c) Transform coefficients with AR(1) and LP_QR.

According to Table 2, for three central EEG channels, subject K3b obtained close accuracies for both methods while subject L1b showed the greatest gap in this regard. The same trend is followed by 60 channels. However, for the selected channels the minimum

difference belongs to subject L1b and the greatest one is for K3b. The closest average accuracy for all three subjects belongs to the category of selected channels. For all subjects and all three categories, the obtained accuracy by LP_SVD is higher than LP_QR, where the

Table 2
Accuracy of LP_SVD and LP_QR for different combination of channels.

	3 central channels		60, all the channels		selected channels	
	LPQR	LPSVD	LPQR	LPSVD	LPQR	LPSVD
Accuracy						
L1b	39.17	56.25	50	64.16	77.08	77.91
K6b	34.17	46.67	59.58	64.17	74.58	76.25
K3b	64.44	70.83	74.44	77.22	86.38	90
avg	45.93	57.92	61.34	68.52	79.35	81.38

Table 3
Required execution time for feature extraction of different scenarios.

running time (s)						
Subject	L1b		K6b		K3b	
	LPQR	LPSVD	LPQR	LPSVD	LPQR	LPSVD
1 ch, 1 trial	0.19	0.57	0.19	0.55	0.18	0.58
60 chs, 1 trial	11.88	34.55	11.61	33.69	11.89	35.99
Selected chs, 1 trial	4.46	11.46	3.54	7.67	5.33	15.07
All chs, All trials	5.26 m	98.57 m	5.27 m	97.81 m	7.45 m	148.10 m

*m = min.

Table 4
Selected channels and their contribution for subject L1B.

Subject L1b		
in common monopolar channels		
LP-SVD	Channel (#)	LP-QR
Accuracy (%)		Accuracy (%)
3.33	5	5.42
4.58	15	3.33
2.08	24	5.42
2.08	38	5.83
2.08	40	5
Subject L1b		
Not in common monopolar channels		
LP-SVD	Channel (#)	LP-QR
Accuracy (%)		Accuracy (%)
1.25	10	27
0.83	20	
2.91	31	
2.91	42	

difference varies between 0.83% (L1b, selected channel) and 17.08% (L1b, 3 central channel). Although the results of 3 channels and 60 channels are reported to provide different aspects for comparison, the main focus of accuracy is on the selected channel. For the selected channels, LP_SVD could obtain 81.38 while LP_QR obtained 79.35 which is very close and shows that LP_QR can compete with LP_SVD regarding accuracy.

3.2.2. Time consumption

Execution time or time consumption clearly manifests the complexity of applied algorithms. The most accurate and common way of measuring the execution time in MATLAB is using the “tic/toc” function. This function is applied to both LP_SVD and LP_QR, in an identical manner. It is applied after loading the data, and it only includes the part where features are extracted. The reported execution times were obtained from algorithms run under the same conditions. The run time can be less with better processors, but the ratio between the two required execution times is expected to remain the same. To have a complete overview of execution time, the results of four different measurements are reported, which are, 1) the execution time for one channel and one trial of each sub-

Table 5
Selected channels and their contribution for subject K6B.

Subject K6b		
LP-SVD	Channel (#)	LP-QR
Accuracy (%)		Accuracy (%)
2.92	in common monopolar channels 58	9.16
Subject K6b		
LP-SVD	Channel (#)	LP-QR
Accuracy (%)		Accuracy (%)
3.75	Not in common monopolar channels 60	28 33 6.25 2.08

Table 6
Selected channels and their contribution for subject K3B.

Subject K3b		
in common monopolar channels		
LP-SVD	Channel (#)	LP-QR
Accuracy (%)		Accuracy (%)
1.95	11	2.22
2.22	16	4.72
0.56	38	1.66
0.84	52	2.77
Subject K3b		
Not in common monopolar channels		
LP-SVD	Channel (#)	LP-QR
Accuracy (%)		Accuracy (%)
0.84	17	13
1.39	18	31
0.56	23	39
	44	2.77
	8	5
	53	4.16

ject, 2) the execution time for all the available channels and one trial, 3) the execution time of the selected channels for one trial and 4) the execution time of all the channels and all the available trials. From the obtained results, reported in Table 3, it is evident that LP_QR needs considerably lesser execution time compared to LP_SVD. 4.46 s compared to 11.46 s for subject L1b, 11.61 s compared to 33.69 for subject K6b and 11.89 compared to 35.99 for subject K3b are reported for LP_QR and LP_SVD, respectively. For all the subjects, LP_QR is almost three times faster than LP_SVD at delivering the expected results.

The gap in required execution time is clearest when all channels and all trials go through this process. The scan of all channels is necessary for most of the channel selection algorithms like wrapping methods. For instance, subject K3b, which has a total of 360 trials and the number of channels is 60, a time of 7.45 min is required for LP_QR while 148.10 min are needed for LP_SVD.

3.2.3. Channels' contribution

This section introduces channels that contributed to achieving better accuracies. The number of channels is an important factor because as the number of channels increases, so do the required execution time.

Tables 4–6 show the contribution of monopolar channels and the number of the monopolar channels that contributed positively. It is reported only for monopolar channels because in bipolar

Table 7
Contributed accuracy for channels of different montage (%).

	in common monopolar channels		Not in common monopolar channels		bipolar channels	
	LPQR	LPSVD	LPQR	LPSVD	LPQR	LPSVD
L1b	25	14.15	6.67	7.9	45.41	55.86
K6b	9.16	2.92	8.33	3.75	56.67	69.58
K3b	11.37	5.57	18.87	2.79	56.14	81.64

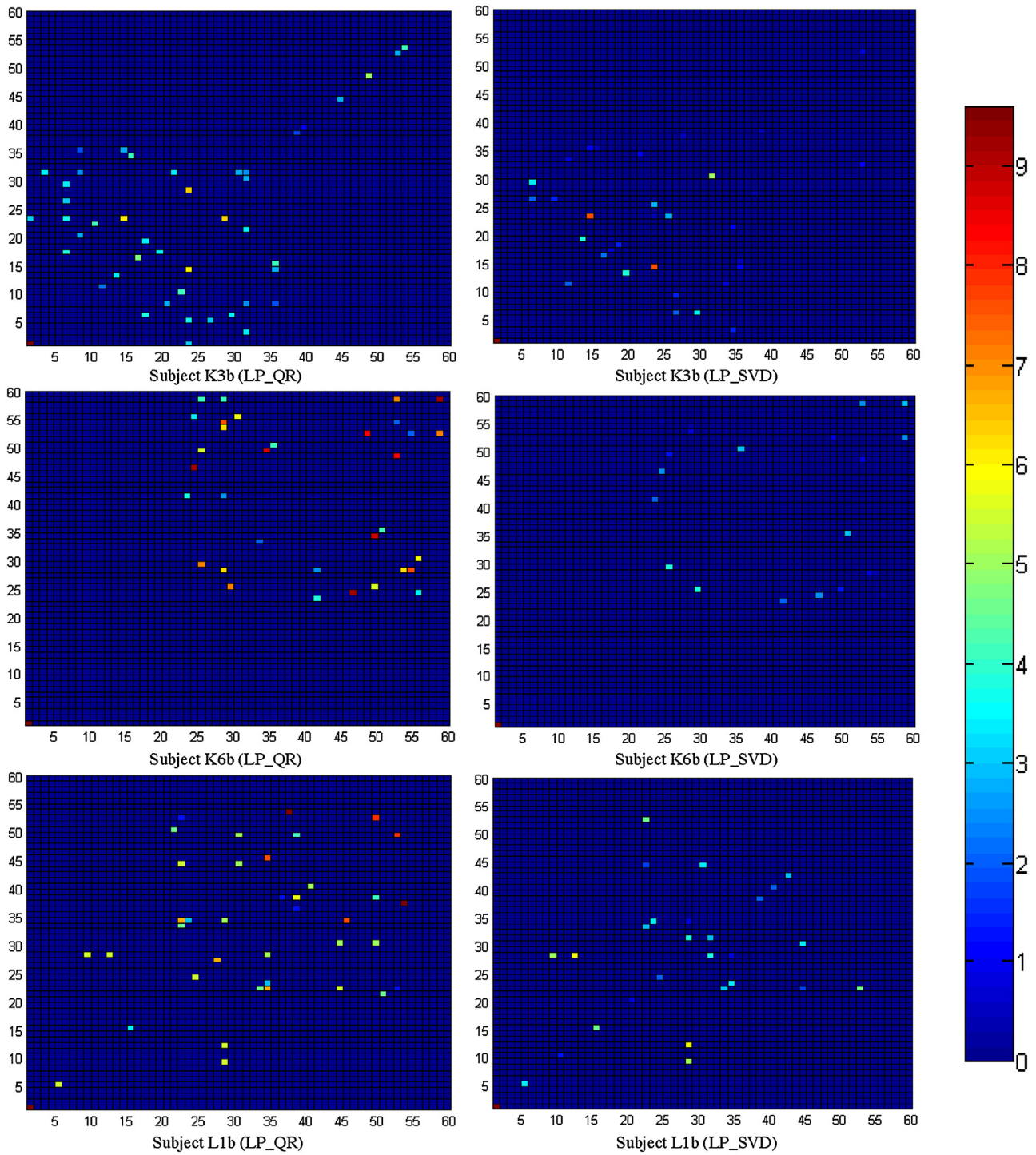


Fig. 2. Map of active channels and their contribution.

Table 8
Benchmarking with the state-of-art methods.

Method	Avg	Requirement for parameter tuning
Guan et al. [14]	84.44	Fine, intensive tuning
Proposed (LP-SVD)	81.38	No
Wentrup et al. [11]	80.60	Yes
Proposed (LP-QR)	79.35	No
Koprinska [19]	78.42	Yes
Wentrup et al. [18]	78.40	Yes
Wentrup et al. [18]	78.23	Yes
Gao et al. [14]	76.20	Yes
Hill and Schröder [14]	72.03	Yes
Schlögl et al. [5]	61.16	No

montage it is not clear which channel is the cause of the contribution. In case of subject L1b, Table 4, nine monopolar channels for LP_SVD and six monopolar channels for LP_QR contributed positively, among them five are identical; however, the amount of contribution of these identical channels is different.

For LP_SVD, these five channels contributed 14.15% to the total obtained accuracy while the same channels could obtain 25% of the total accuracy for LP_QR. A similar trend is observed for the other two subjects. For all the three subjects, the number of selected monopolar channel is the same; however, the obtained accuracy through LP_QR is much higher than LP_SVD. In other words, the extracted information through LP_QR is more informative for the classifier as compared to the information from LP_SVD. Table 7 outlines the contribution of monopolar channels which are in common between the two methods, channels which are not in common, and also the contribution of bipolar channels for LP_QR and LP_SVD separately.

For ease of comparison, the active monopolar and bipolar channels and the amount they contributed are presented in the form of maps for each subject in Fig. 2. The amount each active channel contributed is indicated by its color.

3.2.4. Benchmarking

The results of winners of BCI competition IIIa and the remarkable and competing works in this field are benchmarked against the two proposed methods. As mentioned before, the highest result regarding average accuracy belongs to the first winner of competition. They clarified that what they obtained is the result of fine tuning of all parameters. Tuning of parameters makes a method a subject specific method, which means for each subject the parameter should be tuned again. Another drawback of tuning is the time that should be spent on a parameter to find its optimum value, which makes the process more time-consuming. With regards to recent demands for real-time brain-computer-interface systems, methods that do not need tuning are preferred over others that need it. Table 8, compares the achieved accuracy with available methods for BCI systems. Considering the requirements for parameter tuning, the proposed methods are among the successful results.

4. Conclusion

This paper presents a feature extraction method based on autoregressive modeling and QR decomposition. The proposed method is closely related to LP_SVD; therefore, they are compared against each other. LP_QR could obtain accuracies very close to that of LP_SVD and is capable of competing with it while requiring far less execution time. This very low execution time made it possible for the channel selection algorithm to scan all possibilities and find all the channels that could contribute positively, active channels.

The detailed comparison between the contributions of the selected channels, shows that although these two methods are closely related and have similar nature, the active channels for each method are quite different. Interestingly, even the amount of contribution of channels is different for them. The results of both methods are obtained without the help of any filtering in pre-processing stage, any feature selection algorithm, or any fine tuning of parameters. Thus, future studies can address these issues to further improve the results.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.bspc.2016.12.006>.

References

- [1] T.O. Zander, C. Kothe, S. Jatzew, M. Gaertner, Enhancing human-computer interaction with input from active and passive brain-computer interfaces, in: *Brain-Computer Interfaces*, Springer, London, 2010, pp. 181–199.
- [2] Jonathan R. Wolpaw, J. McFarland, Dennis, W. Neat Gregory, A. Catherine Forneris, An EEG-based brain-computer interface for cursor control, *Electroencephalogr. Clin. Neurophysiol.* 78 (3) (1991) 252–259.
- [3] H. Ramoser, J. Müller-Gerking, G. Pfurtscheller, Optimal spatial filtering of single trial EEG during imagined hand movement, *Rehabil. Eng. IEEE Trans.* 8 (4) (2000) 441–446.
- [4] F. Lotte, C. Guan, Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms, *Biomed. Eng. IEEE Trans.* 58 (2) (2011) 355–362.
- [5] A. Schlögl, F. Lee, H. Bischof, G. Pfurtscheller, Characterization of four-class motor imagery EEG data for the BCI-competition 2005, *J. Neural Eng.* 2 (4) (2005) L14.
- [6] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K.R. Müller, Optimizing spatial filters for robust EEG single-trial analysis, *Signal Process. Mag. IEEE* 25 (1) (2008) 41–56.
- [7] F. Lotte, C. Guan, Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms, *Biomed. Eng. IEEE Trans.* 58 (2) (2011) 355–362.
- [8] J. Müller-Gerking, G. Pfurtscheller, H. Flyvbjerg, Designing optimal spatial filters for single-trial EEG classification in a movement task, *Clin. Neurophysiol.* 110 (5) (1999) 787–798.
- [9] W. Wu, Z. Chen, X. Gao, Y. Li, E.N. Brown, S. Gao, Probabilistic common spatial patterns for multichannel EEG analysis, *Pattern Anal. Mach. Intell. IEEE Trans.* 37 (3) (2015) 639–653.
- [10] X. Li, X. Lu, H. Wang, Robust common spatial patterns with sparsity, *Biomed. Signal Process. Control* 26 (2016) 52–57.
- [11] M. Grosse-Wentrup, M. Buss, Multiclass common spatial patterns and information theoretic feature extraction, *Biomed. Eng. IEEE Trans.* 55 (8) (2008) 1991–2000.
- [12] H. Baali, R. Akmeliawati, M.J.E. Salami, A. Khorshidtalab, E.G. Lim, ECG parametric modeling based on signal dependent orthogonal transform, *Signal Process. Lett. IEEE* 21 (10) (2014) 1293–1297.
- [13] H. Baali, A. Khorshidtalab, M. Mesbah, M.J.E. Salami, A transform-Based feature extraction approach for motor imagery tasks classification, *Trans. Eng. Health Med. IEEE J.* 3 (2015) 1–8.
- [14] B. Blankertz, K.R. Müller, D.J. Krusienski, G. Schalk, J.R. Wolpaw, A. Schlögl, N. Birbaumer, The BCI competition III: Validating alternative approaches to actual BCI problems, *Neural Syst. Rehabil. Eng. IEEE Trans.* 14 (2) (2006) 153–159.
- [15] P. Strobach, *Linear Prediction Theory: a Mathematical Basis for Adaptive Systems*, vol. 21, Springer Science & Business Media, 2012.
- [16] Y. Leal, M. Ruiz, C. Lorencio, J. Bondia, L. Mujica, J. Vehi, Principal component analysis in combination with case-based reasoning for detecting therapeutically correct and incorrect measurements in continuous glucose monitoring systems, *Biomed. Signal Process. Control* 8 (6) (2013) 603–614.
- [17] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, *The WEKA data mining software: an update*, *ACM SIGKDD Explor. Newslett.* 11 (1) (2009) 10–18.
- [18] M. Grosse-Wentrup, *Feature Extraction in Non-Invasive Brain-Computer Interfaces*, Universitätsbibliothek der TU München, 2008 (Doctoral dissertation).
- [19] I. Koprinska, Feature selection for brain-computer interfaces, in: *New Frontiers in Applied Data Mining*, Springer, Berlin Heidelberg, 2009, pp. 106–117.