■ 2561

Energy extraction method for EEG channel selection

**Hilman Fauzi*[1], M. Abdullah Azzam[2], Mohd. Ibrahim Shapiai[3], Masaki Kyoso[4],
Uswah Khairuddin[5], Tadayasu Komura[6]**
[1,3,5]Malaysia Japan International Institute of Technology, Universiti Teknologi Malaysia,
Kuala Lumpur, Malaysia
[1]Faculty of Electrical Engineering, Telkom University Bandung, Indonesia
[2]Telco Solution Chaosmatic Co. Ltd. Bandung, Indonesia
[4,6]Department of Medical Engineering, Faculty of Engineering, Tokyo City University, Tokyo, Japan
*Corresponding author, e-mail: hilmanfauzitsp@telkomuniversity.ac.id[1], azzam.cyber@gmail.com[2],
md_ibrahim83@utm.my[3], mkyoso@tcu.ac.jp[4], uswah.kl@utm.my[5], g1781408@tcu.ac.jp[6]

Abstract

Channel selection is an improvement technique to optimize EEG-based BCI performance. In previous studies, many channel selection methods—mostly based on spatial information of signals—have been introduced. One of these channel selection techniques is the energy calculation method. In this paper, we introduce an energy optimization calculation method, called the energy extraction method. Energy extraction is an extension of the energy calculation method, and is divided into two steps. The first step is energy calculation and the second is energy selection. In the energy calculation step, l2-norm is used to calculate channel energy, while in the energy selection method we propose three techniques: "high value" (HV), "close to mean" (CM), and "automatic". All proposed framework schemes for energy extraction are applied in two types of datasets. Two classes of datasets i.e. motor movement (hand and foot movement) and motor imagery (imagination of left- and right-hand movement) were used. The system used a Common Spatial Pattern (CSP) method to extract EEG signal features and k-NN as a classification method to classify the signal features with k=3. Based on the test results, all schemes for the proposed energy extraction method yielded improved BCI performance of up to 58%. In summary, the energy extraction approach using the CM energy selection method was found to be the best channel selection technique.

Keywords: channel selection, EEG-based BCI, energy calculation, energy extraction, energy selection

1. Introduction

Electroencephalography (EEG) is the most preferred technology for BCI over other BCI modalities. This is because EEG is more commonly used due to its high temporal resolution, affordability, availability, and portability [1]. However, EEG produces noisy and non-stationary signals [2–4] that cause poor performance in BCI applications. Many techniques have been introduced to solve this problem. One of the techniques to improve brain signal information redundancy is channel selection and reducing or eliminating irrelevant information in channels [5]. There are many reasons why the channel selection method is important in optimizing BCI. Firstly, one drawback of EEG signals is that over fit to noise increases with the number of task-irrelevant features [6, 7]. Secondly, it is difficult to understand which part of the brain generates class-relevant activity [8, 9]. Thirdly, applying a large number of channels is time consuming [10, 11]. Fourthly, it is crucial to reduce the number of EEG channels as well as to maintain good reliability to improve the portability and practicability of BCI systems [12]. Therefore, this study introduced a channel selection method to improve BCI performance, namely the energy extraction method.

In a previous study, the energy extraction method was known as spatial filtering [13] or the energy calculation method [14]. Two previous studies basically used similar methods for energy selection i.e. the highest energy of channels. However, this study proposes energy extraction as a method for optimizing previous studies via the use of different channel selection techniques. In general, the energy extraction method is divided into two processes, which are energy calculation and energy selection. In the energy calculation process, we used l2-norm to generate the energy for all channels, while in energy selection we introduced two techniques to select and define the composition of active channels i.e. *"close to mean"* and automatic selection, and the "*high value*" technique, which is a similar method to that of previous studies.

This paper is organized as follows: section 2 covers previous related works on BCI and channel selection. Section 3 discusses the methodology of the proposed BCI framework via the energy extraction process. Section 4 presents the test result and discussion of BCI performance using the energy extraction method. Finally, concluding remarks and future works are given in section 5.

2. Research Method

EEG is the most widely used device in BCI applications. There are many types of EEG signal actions such as motor action [15, 16], imagination [17–19], emotion [20–22], speech [23–25], visual [26, 27], audio [28–30] etc., all of which simplify the measurement of brain signals. However, the generated EEG signals have less favorable characters that make it easily translatable such as non-linearity [31–33], non-stationary [3, 4, 34] and noisy features [2, 35]. One technique for optimizing EEG signals is channel selection, which can reduce or eliminate irrelevant information in channels [5].

Based on one study [36], the channel selection method can be divided into five methods, which are filtering, wrapper, embedded, hybrid, and human-based. In the filtering method, channel selection is generated via a search algorithm. This method has been used in past studies [9, 37]. The wrapper method uses a classification algorithm to evaluate candidate channels [38, 39]. The embedded method facilitates the interaction between channel selection and classification. The hybrid method is a combination of the filtering and wrapper techniques [40]. The human-based method uses humans in the feedback task for channel selection.

A new channel selection technique using the energy calculation method was presented in one study as an optimized signal to improve BCI performance [14]. Basically, the energy calculation method uses a l2-norm to generate channel energy [13, 14]. However, this technique has been used previously in other fields to generate specific information [41]. The EEG signal was generated using l2-norm in a CUR matrix decomposition to prevent 'out of memory' problems because of the large size of the data matrix in non-negative matrix factorization (NMF) [42].

3. Research Method

This study is related to the BCI system processes and is basically structured into three main processes, which are preprocessing, feature extraction, and classification. The energy extraction process was introduced in BCI as a channel selection method to optimize and improve BCI performance. The process of energy extraction occurs in between the preprocessing and feature extraction, as shown in Figure 1.
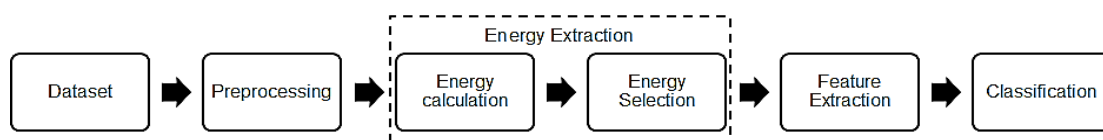

Figure 1. Block diagram of the BCI system

3.1. Dataset

This study used two types of dataset, namely dataset I and dataset II. Dataset I consists of two classes of physical motor movement EEG signals, which were provided by the Biomedical Instrumentation Engineering Laboratory of Tokyo City University, Japan. Motion signals such as grasping of the right hand and raising of the right foot were recorded using seven channels of EEG (C2, Cz, C1, C3, C5, FC3, and FC5) from three healthy adult subjects (sbjA, sbjB, and sbjC) sitting on a chair. Both datasets are explained in Table 1.

Similar to dataset I, dataset II also has two classses of motor imagery EEG signals for left hand imagery (LH) and right hand imagery (RH). This dataset was provided by brain signal researchers at Dr. Cichocki's Lab (Lab. for Advanced Brain Signal Processing), BSI, and via RIKEN collaboration with Shanghai Jiao Tong University. The motor imagery signal was

recorded using 6 EEG channels (C3, Cp3, Cz, Cpz, C4, Cp4) involving five healthy adult subjects (S1, S2, S3, S4, S5).

Table 1. The Dataset Table

Datasets	Subjects	Class	Channel Number	Trial number	Sample rate
Dataset I	SbjA	Hand		40	
	SbjB	Movement/Leg	7	40	200Hz
	SbjC	Movement		40	
Dataset II	S1	Left Hand		170	
	S2	Imagery/Right		158	
	S3	Hand Imagery	6	48	256Hz
	S4			120	
	S5			90	

3.2. Energy Extraction

This study mainly focuses on energy extraction as a channel selection method. The framework starts with a filtered signal in preprocessing using a second-order Butterworth filter at a frequency of 3-14 Hz. The energy extraction method was divided into two processes: the energy calculation process and energy selection process.

3.2.1. Energy Calculation

The purpose of this process is to calculate the energy in the channels using an l2-norm equation. The process begins by calculating the channel energy, as per (1).

$$\mathbf{C} = c_{:,\ pj}\ ,\ \text{where}\quad p_i = \frac{\sum_{j=1}^{n} A(i,j)^2}{\sum_{i=1}^{m}\sum_{j=1}^{n} A(i,j)^2} \tag{1}$$

where, C is a matrix with selected channels, c denotes the columns which later identified as channels, m and n are the number of row and column of matrix C, respectively, and p is the average channel energy. The calculation was applied in all training trials and sessions of subjects, so that every channel is represented by an energy, which later known as channel energy. The next process is energy selection, where the channel energy is selected based on the three techniques to have the best selected channels.

3.2.2. Energy Selection

The energy selection process is the most important process in the energy extraction method for channel selection. This process determines the energy value in the channel to be selected as active channels in regard to the task of class type in a dataset. In this study, we introduced two approaches for the energy selection technique, which are manual and automatic selection. The main difference between the manual and automatic energy selection is the technique used to define the energy combination criteria.

a. Manual selection

The main point of the manual energy selection is to sort channel energy. This paper proposed two techniques for sorting channel energy, where the first is based on the highest energy of selected combination channels called the *"highest value"* method, and the second is based on the adjacency of channel energy, which is called the *"close to mean"* method.

- Highest Value (HV)

This energy selection technique was introduced in previous studies [13, 14]. The combination of active channels is selected sequentially following a number of channel rankings sorted by decreasing energy value. The form of channel combination is further tested for performance accuracy. The combination of channels with the highest accuracy will be the best combination of selected active channels.

- Close to mean (CM)

The CM energy selection technique sorts energy in increasing order, which has the smallest difference between channel energy and average channel energy in the first rank. Using this technique, the entire channel energy is first calculated on average and used as a reference threshold to create channel energy order. This is done by calculating the difference in energy between the reference and each channel energy. The *"close to mean"* selection method is not that different from the *"highest value"* method, as it selects a number of channel

combinations following the rankings in the order that has been established. The selected channel is generated by the highest performance test outlined by (2):

$$\varepsilon_{sc} = \min|\sum_{i=1}^{n} \bar{\varepsilon}_n - \varepsilon_i| \tag{2}$$

where, ε is the channel energy, which is the "*sc*" selected channel energy that represents the selected channels, and "*n*" is the number of channels, while $\bar{\varepsilon}_n$ is the average of all channel energy.

b. Automatic Selection

The main difference between the process of automatic and manual energy selection is the technique used for determining energy threshold value. The automatic selection process defines the combination of selected channels by the boundary of threshold energy value based on a defined region. This method determines the combination of active channels based on channel energy calculation, which is higher than the average channel energy (energy threshold) as denoted by (3). Based on this method, the selected channels will be the channels with a higher energy than the threshold.

$$\varepsilon_{sc} \geq \bar{\varepsilon}_n \tag{3}$$

3.3. Common Spatial Pattern (CSP)

CSP is known as a popular and powerful method for discriminating two classes of EEG [43]. It generates maximum variances in one-multi channel signal and at the same time minimizes other variances [44]. The main idea of the Common Spatial Pattern (CSP) is to use a linear transformation to project multi-channel EEG data into a low-dimensional matrix. The normalized covariance matrix can be calculated using (4):

$$C_y = \frac{1}{n_y} \sum \frac{E_j(y)E_j^T(y)}{\text{trace}(E_j(y)E_j^T(y))} \tag{4}$$

where, E_j denotes the EEG signal for the j-th trial, n_y represents a number of trials for the y class, and y is a class (e.g. left and right). T is the transpose operator and trace is the diagonal sum operation of matrix elements. The covariance matrix is computed over trials in each class of EEG data. The combination of spatial covariance can be computed using (5), where C is the average spatial covariance and 1 and 2 are the name of classes (in this case, 1 is for the left class and 2 is for the right class, while t is the combination of spatial covariance).

$$Ct = C1 + C2 \tag{5}$$

The combination of spatial covariance is used as a factorized formula in (6) to whiten the transformation process in (7):

$$Ct = \hat{U}\hat{A}\hat{U}T \tag{6}$$

$$P = \hat{A}\text{-}1/2\hat{U}T \tag{7}$$

the notation $\hat{U}$ represents the eigenvector matrix and $\hat{A}$ is the diagonal matrix of the corresponding eigenvalues. Transformations of the covariance matrices as whitened signals for both classes are carried out by using (8) and (9):

$$S1 = PC1\,PT \quad S2 = PC2\,PT \tag{8}$$

$$S1 = \bar{B}\lambda1\,\bar{B}^T \quad S2 = \bar{B}\lambda2\,\bar{B}^T \quad \lambda1 + \lambda2 = I \tag{9}$$

where S_1 and S_2 share common eigenvectors and the sum of corresponding two matrices will always be one. The eigenvectors with the highest eigenvalues for S_1 have the smallest eigenvalues for S_2 and vice versa. The projection matrix *W* is denoted by (10):

$$W = \bar{B}^T\,P \tag{10}$$

using the projection matrix of *W*, the original EEG can be transformed into uncorrelated components in the *X* matrix, as per (11):

$$X = W\text{-}1Z \tag{11}$$

Z is known as an EEG source that contains common and specific components in different tasks. W^{-1} is the inverse matrix of W. The columns of W^{-1} can be considered as EEG source distribution vectors. The first and last columns of W^{-1} are the most important spatial patterns that explain the highest and smallest variance of a task.

3.4. K-Nearest Neighbor (k-NN)

The k-Nearest Neighbor (k-NN) is a method that uses a supervised algorithm in which the results of a new instance query are classified according to the majority categories in k-NN. The aim of the k-NN technique is to classify new objects based on attributes and training samples. In BCI, the k-NN is usually obtained using a distance metric [45]. This method can approximate several functions in an adequate number of training samples which includes suitable value for k. Therefore, k-NN enables the reduction of nonlinear decision boundaries. k-NN may be efficient for BCI implementation with low-dimensional feature vectors.

The k-NN method algorithm is simple, and only calculates the shortest distance from query instance to training sample in order to determine its k-NN. Training samples are projected onto a many-dimensional space, where each dimension represents a feature of the data. This space is divided into sections based on training sample classification. The neighbor distances are usually calculated based on Euclidean Distance, which is represented by (12):

$$D(a, b) = \sqrt{\sum_{k=1}^{d}(a_k - b_k)^2} \tag{12}$$

where, the *D(a,b)* matrix is the scalar distance between the *a* vector and *b* vector from the matrix with *d* dimension. In the training phase, this algorithm only retains feature vectors and sample training data classification. In the classification phase, the same features are calculated for the testing data (unknown classification) which is in this study, we use use k = 3 for k-NN.

3.5.   Performance Evaluation
3.5.1. Accuracy Test

The accuracy test is used to measure the performance of the system. It was calculated using Mean Squared Error (MSE) in (13) which calculates the initial accuracy parameter for the created system.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2 \tag{13}$$

where $\hat{Y}$ is the vector of *n* predictions, and *Y* is the vector of observed values corresponding to the inputs to the function, which generated the predictions. The accuracy test was applied in the classification process, and output a value between 0 and 1.

3.5.2. Power Spectrum Distribution

This test finds the distribution of energy information in a channel. Energy signals take the form of a contour map in the range of 0 (black color) to 1 (white color), where a lighter color means a higher and more positive energy signal, and a darker color indicates an irrelevant signal. This test is equipped with the composition of selected active channels, which yielded the best performance in every energy extraction technique for foot movement and right hand imagery in dataset I and dataset II, respectively.

4. Results and Analysis

The result of this study is presented in terms of accuracy performance, power spectrum distribution, and selected active channels by energy selection method. Using these evaluation criteria, the proposed energy selection method can be evaluated as a part of energy extraction in the best channel selection method to improve BCI performance. All evaluation is then compared with the original BCI scheme (without the energy extraction scheme).

4.1. Accuracy Test
4.1.1. Manual Selection
Firstly, the energy extraction method is applied in datasets without channel selection. In dataset I, the average accuracy of the HV method is the same to the original scheme, but the CM method has better accuracy of up to 58.7% with a range of subject accuracy enhancement from 36% to 81%. The highest improvement in accuracy was observed for sbjA. In the HV method, the arrangement of energy channel is similar to that of the original, while in the CM method, the channel energy arrangement is different for hand and foot movements; thus, the accuracy improved because it is easier to define the classes of the dataset in the energy channel arrangement of CSP.

There is not much difference between the accuracy results of dataset II and dataset I. The average accuracy of the HV method in dataset II is similar to the original in six channels, but an improved accuracy of 6.3% was observed for subjects S2 and S3. Meanwhile, the CM method accuracy improved to 55.5% on average.

In terms of channel selection, dataset I yielded the best improvement for four active channels using the HV method with an accuracy of 52.3% better than the original. Meanwhile, the CM method reached maximum accuracy because it did not apply channel selection and yielded an average accuracy value of 1.00. The accuracy of the CM method was still at its highest until the channel selection was applied in five active channels. Based on the test results for the manual channel selection of dataset I, the HV and CM energy selection methods showed better improvement for all numbers of channel combinations, except for 1 channel in which the accuracy was the same. The details of dataset I test result can be seen in Table 2.

Meanwhile, in dataset II, the HV method generated the best accuracy for five active channels, where the average accuracy improved by 4.7%. Actually, the HV method did not show improved accuracy for all subjects in the dataset. Significant average improvement was observed in subjects S3 and S5, for which the improvement went up to 19.6%. In the CM energy selection method, the selected active channels, which yielded the best-improved accuracy was observed in the five channels. The accuracy improvement is similar to that of the CM method in six channels compared to the original. Overall, for the test results of dataset II as shown Table 3, improvement in accuracy performance with the CM method could be observed for all number of channel selections, except for 1 channel. The average accuracy improvement ranged from 3% to 55.56%.

Table 2. Classification Accuracy of Dataset I in Manual Energy Selection Method

Subj.		Number of Channels												
		7		6		5		4		3		2		1
	Ori.	HV	CM	HV	CM	HV	CM	HV	CM	HV	CM	HV	CM	HV	CM
SbjA	0.55	0.55	1.00	0.75	1.00	0.80	1.00	1.00	1.00	0.93	0.77	0.85	0.50	0.50	0.50
SbjB	0.73	0.73	1.00	0.88	1.00	1.00	1.00	1.00	1.00	0.90	0.82	0.75	0.47	0.50	0.50
SbjC	0.60	0.60	1.00	1.00	1.00	0.90	1.00	0.88	0.92	0.85	0.92	0.50	0.35	0.50	0.50
Avg	0.63	0.63	1.00	0.88	1.00	0.90	1.00	0.96	0.97	0.89	0.84	0.70	0.44	0.50	0.50

Table 3. Classification Accuracy of Dataset II in Manual Energy Selection Method

| Subj.		Number of Channels											
		6		5		4		3		2		1
	Ori.	HV	CM	HV	CM	HV	CM	HV	CM	HV	CM	HV	CM
S1	0.65	0.64	0.98	0.61	0.98	0.55	0.98	0.58	0.98	0.48	0.64	0.50	0.50
S2	0.56	0.60	0.98	0.55	0.98	0.53	0.98	0.44	0.91	0.61	0.47	0.50	0.50
S3	0.56	0.60	0.95	0.67	0.95	0.33	0.95	0.50	0.95	0.58	0.75	0.50	0.50
S4	0.67	0.60	0.98	0.65	0.98	0.60	0.98	0.64	0.89	0.62	0.69	0.52	0.52
S5	0.73	0.72	1.00	0.80	1.00	0.57	0.97	0.57	0.84	0.57	0.72	0.50	0.50
Avg	0.63	0.63	0.98	0.66	0.98	0.52	0.97	0.55	0.91	0.57	0.65	0.50	0.50

In general, the accuracy performance of the channel selection methods relatively decreased and was linearly proportionate to the number of selected channels in dataset I and dataset II, especially for the CM method as shown in Figure 2. A high accuracy was achieved with a high number of selected channels, and vice versa. The same could not be said for the HV method, however, which yielded the highest accuracy for only one of the selected channel combinations with no dependence on the number of selected channels. For example, in dataset

I and dataset II, the highest accuracy was observed in four selected channels, and five selected channels, respectively.

The CM energy selection method defines active channels based on the energy that appears in most active channels. Therefore, every channel has an energy coverage that will give the best information to CSP to extract the best feature of the signals. When the channel is not selected (in manual selection), the energy coverage will be smaller and will reduce the accuracy. Hence, in the CM method, accuracy is affected by the number of selected channels. In addition, the HV method finds the best combination of selected channels by highest energy without calculating the difference in energy value between the closest energy values in the selected channels. Therefore, the HV method will yield the highest accuracy without being affected by the number of channels.
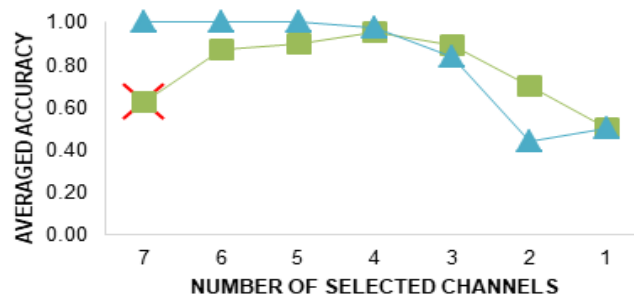
(a)
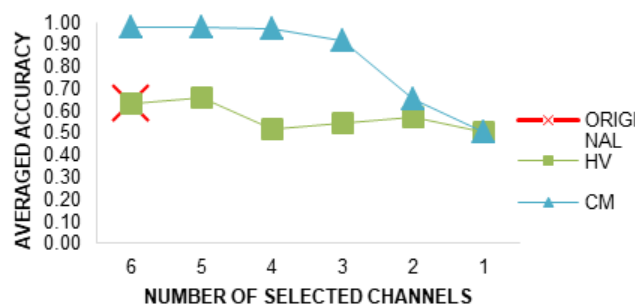
(b)

Figure 2. Performance comparison of the selected channels number for
the manual channel selection method: (a) Dataset I (b) Dataset II

4.1.2. Automatic Selection

The automatic energy selection method selects the best energy and combination of channels in one process selection. This method yields faster computation time compared to manual methods. Based on the accuracy performance test result, both datasets I and II showed better accuracy than the original method. An accuracy of 0.94 for dataset I with an average accuracy improvement of 49.2% was observed. Dataset II yielded an accuracy of 0.889 or 33.3% better accuracy compared to the original. Although dataset I and dataset II experienced a significant improvement in accuracy, the accuracy is no better than the manual method, especially for the HV and CM methods in dataset I and the CM method in dataset II, when comparing the best accuracy of channel selections for all methods. This automatic method gives better performance than the same number of selected channels (3 channels) in the manual method, where on average, the performance of the automatic method is better than two of the manual methods in dataset I. Meanwhile, in dataset II, the accuracy of the automatic method is better than the HV manual method with the highest accuracy being observed for subject S3 at a value of 0.93. The accuracy result of automatic energy selection method is shown in Table 4.

Table 4. Classification Accuracy in Automatic Energy Selection Method

Subjects	Dataset I			Dataset II				
	SbjA	sbjB	sbjC	S1	S2	S3	S4	S5
Accuracy	1	1	0.825	0.98	0.73	0.96	0.98	0.53
Acc. Avg.		0.94				0.89		
Num. of Selected Channels	3	3	3	3	3	3	3	3

4.2. Selected Channel and Power Spectrum Distribution

The selected channels are mapped onto the scalp in Figure 3 to determine the channels to be selected by the energy selection methods (HV, CM, and automatic). This mapping of selected channels as described in Table 5 comes from the active channels selected by HV, CM, and automatic methods for the foot movement of dataset I and the right-hand imagery of dataset II.

In dataset I, two channels were selected of all the energy selection methods for foot movement, which are Cz and C1. Meanwhile, in dataset II, CP3 and CP4 were selected as the active channels for right hand imagery in all the energy selection methods. The result shows that the foot movement in dataset I and right-hand imagery in dataset II are defined by these selected channels.

The power spectrum distribution mapping was generated from the foot movement of dataset I and the right-hand imagery of dataset II. The CM method was the best-performing energy selection method in dataset I and dataset II. One of the factors that made CM so powerful is its power distribution, which has a wider area of high energy than other methods. These areas or distribution energies were generated by the selected channels, where more selected channels would generate higher areas of energy and yield better performance, as explained before in test result of CM section. The power spectrum distribution maps are illustrated in Table 6.

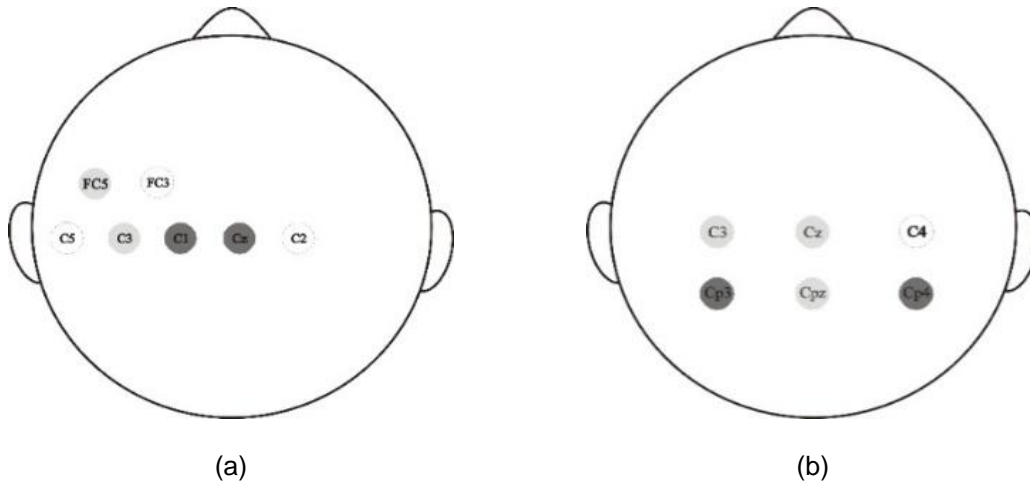(a)   (b)

Figure 3. The intersection of selected active channel maps between HV, CM, and automatic energy selection methods; the dark gray color indicates the channels that were selected via all the energy selection methods, the light gray color indicates the selected channels selected by two methods, and the white color indicates channels selected by one method, where:
(a) the active channel map for foot movement in dataset I and
(b) the active channel map for right hand movement in dataset II

Table 5. The Composition of Selected Channels of Foot Movement and
Right Hand Imagery in All Energy Extraction Methods

Datasets	Selected Channels		
	HV	CM	Auto
Dataset I	Cz-C1-C3-FC5	FC3-FC5-C2-Cz-C1	Cz-C1-C3
Dataset II	C3-Cp3-C4-Cp4-Cz	C3-Cp3-C4-Cz-Cpz	Cp3-Cp4-Cpz

Table 6. Power Spectrum Distribution
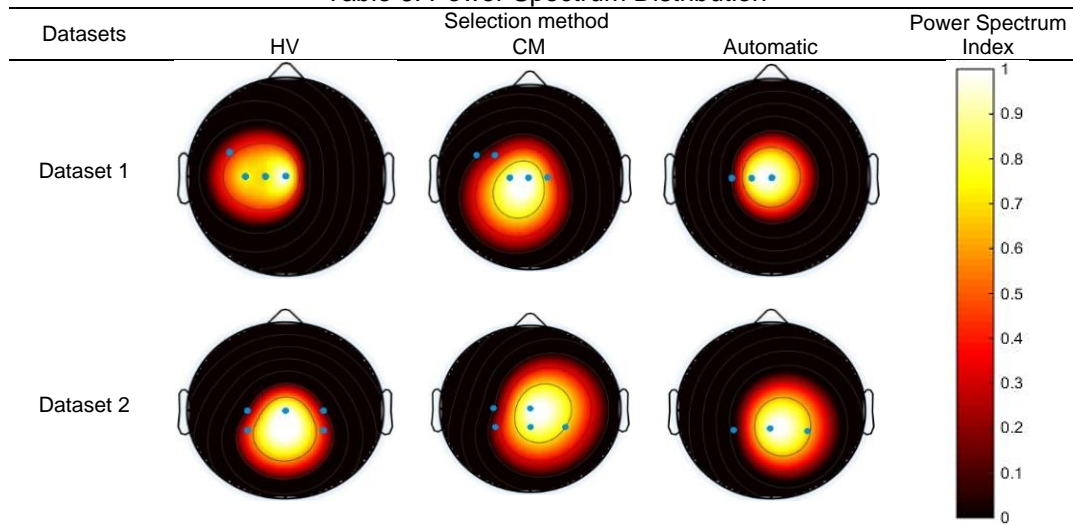
5. Conclusion

In general, like what has been shown by previous studies, an energy extraction method with improved BCI performance has been achieved. Energy extraction via all energy selection methods in this study showed an enhanced performance of up to 50% compared to the original scheme for dataset I and dataset II. In line with the better performance, the CM energy selection method gives the best performance compared to the other energy selection methods. Besides that, CM technique also offer a better performance of BCI even though in complete number of channels. It can be assumed that BCI performance using CM technique in energy extraction method will be dependant on the number of channels. When less active channels are selected, it will reduce the BCI performance and vice versa. The framework in this study was applied on small number of channels and a specific location for motor and motor imagery location. Therefore, it is proven that the framework offered better BCI performance for EEG with a small number of channels. In the future, this framework could use datasets with a larger number of channels or electrodes. Besides that, because of the results of the test that yielded better and improved performance, the energy extraction method could also define the common channels in different subjects or sessions in a dataset to improve upon the results.

Acknowledgement

References

[1] Mohanchandra K, Saha S. Optimal channel selection for robust EEG single-trial analysis. *AASRI Procedia.* 2014; 9: 64-71.
[2] Sosa OP, Quijano Y, Doniz M, Chong-Quero JE. *BCI: A historical analysis and technology comparison.* 2011 Pan American Health Care Exchanges. 2011: 205–209.
[3] Samek W, Müller KR, Kawanabe M, Vidaurre C. *Brain-computer interfacing in discriminative and stationary subspaces.* 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2012: 2873–2876.
[4] Li KG, Shapiai MI, Adam A, Ibrahim Z. *Feature scaling for EEG human concentration using particle swarm optimization.* Proceedings of 2016 8th International Conference on Information Technology and Electrical Engineering: Empowering Technology for Better Future, ICITEE 2016. 2017: 1-6.
[5] Qiu Z, Jin J, Lam HK, Zhang Y, Wang X, Cichocki A. Improved SFFS method for channel selection in motor imagery based BCI. *Neurocomputing.* 2016; 207: 519–527.

[6]   Duque-Muñoz L, Guerrero-Mosquera C, Castellanos-Dominguez G. *Stochastic relevance analysis of epileptic EEG signals for channel selection and classification*. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS. 2013: 2104–2107.

[7]   Arvaneh M, Guan C, Ang KK, Quek C. Optimizing the channel selection and classification accuracy in EEG-based BCI. *IEEE Transactions on Biomedical Engineering*. 2011; 58(6): 1865–1873.

[8]   Schröder M, *et al.* Robust EEG channel selection across subjects for brain-computer interfaces. *EURASIP J. Appl. Signal Processing*. 2005; 2005: 3103–3112.

[9]   He L, Hu Y, Li Y, Li D. Channel selection by Rayleigh coefficient maximization based genetic algorithm for classifying single-trial motor imagery EEG. *Neurocomputing*. 2013; 121: 423–433.

[10] Astigarraga A, Arruti A, Muguerza J, Santana R, Martin JI, Sierra B. User adapted motor-imaginary brain-computer interface by means of EEG channel selection based on estimation of distributed algorithms. *Mathematical Problems in Engineering*. 2016; 2016: 1-12.

[11] Naeem M, Brunner C, Pfurtscheller G. Dimensionality reduction and channel selection of motor imagery electroencephalographic data. *Computational intelligence and neuroscience*. 2009; 2009: 1-8.

[12] Arvaneh M, Guan C, Ang KK, Quek C. *Robust EEG channel selection across sessions in brain-computer interface involving stroke patients*. The 2012 International Joint Conference on Neural Networks (IJCNN). 2012: 1-6.

[13] Fauzi H, Shapiai MI, Yusof R, Remijn GB, Setiawan NA, Ibrahim Z. *The design of spatial selection using CUR decomposition to improve common spatial pattern for multi-trial EEG classification*. Asian Simulation Conference. Singapore. 2017: 428-442.

[14] Fauzi H, Shapiai MI, Setiawan NA, Jaafar J, Mustafa M. *Channel selection for common spatial pattern Based on energy calculation of motor imagery EEG signal*. 2017 International Conference on Control, Electronics, Renewable Energy and Communications (ICCREC). 2017: 33–39.

[15] Gwin JT, Ferris D. *High-density EEG and independent component analysis mixture models distinguish knee contractions from ankle contractions*. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS. 2011: 4195–4198.

[16] Park W, Jeong W, Kwon GH, Kim YH, Kim L. *A rehabilitation device to improve the hand grasp function of stroke patients using a patient-driven approach*. IEEE International Conference on Rehabilitation Robotics. 2013.

[17] Schaefer RS, Desain P, Farquhar J. Shared processing of perception and imagery of music in decomposed EEG. *Neuroimage*. 2013; 70: 317–326.

[18] Liu Y, *et al.* A tensor-based scheme for stroke patients' motor imagery EEG analysis in BCI-FES rehabilitation training. *J. Neurosci. Methods*. 2014; 222: 238–249.

[19] Gürkök H, Poel M, Zwiers J. *Classifying motor imagery in presence of speech*. Proceedings of the International Joint Conference on Neural Networks. 2010.

[20] Zheng WL, Zhang YQ, Zhu JY, Lu BL. T*ransfer components between subjects for EEG-based emotion recognition*. 2015 International Conference on Affective Computing and Intelligent Interaction, ACII 2015. 2015: 917–922.

[21] Y Lin YP, Jung TP. Improving EEG-based emotion classification using conditional transfer learning. *Frontiers in human neuroscience*. 2017; 11: 334.

[22] Ding H, Lu Q, Gao H, Peng Z. Non-invasive prediction of hemoglobin levels by principal component and back propagation artificial neural network. *Biomed. Opt. Express*. 2014; 5(4): 1145–1152.

[23] Brigham K, Kumar BVKV. *Subject identification from Electroencephalogram (EEG) signals during imagined speech*. IEEE 4th International Conference on Biometrics: Theory, Applications and Systems, BTAS 2010. 2010.

[24] Mirkovic B, Debener S, Jaeger M, De Vos M. Decoding the attended speech stream with multi-channel EEG: implications for online, daily-life applications. *Journal of neural engineering*. 2015; 12(4): 046007.

[25] Gow DW, Segawa JA, Ahlfors SP, Lin FH. Lexical influences on speech perception: A Granger causality analysis of MEG and EEG source estimates. *Neuroimage*. 2008; 43(3): 614–623.

[26] Busch NA, Dubois J, VanRullen R. The phase of ongoing EEG oscillations predicts visual perception. *Journal of Neuroscience*. 2009; 29(24): 7869–7876.

[27] Kamienkowski JE, Ison MJ, Quiroga RQ, Sigman M. Fixation-related potentials in visual search: a combined EEG and eye tracking study. *Journal of vision*. 2012; 12(7): 4.

[28] Vialatte FB, Dauwels J, Musha T, Cichocki A. Audio representations of multi-channel EEG: a new tool for diagnosis of brain disorders. *American journal of neurodegenerative disease*. 2012; 1(3): 292–304.

[29] Lee YC, Lin WC, King JT, Ko LW, Huang YT, Cherng FY. *An EEG-based approach for evaluating audio notifications under ambient sounds*. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014: 3817–3826.

[30] Vialatte FB, Dauwels J, Musha T, Cichocki A. Audio representations of multi-channel EEG: a new tool for diagnosis of brain disorders. *Am. J. Neurodegener. Dis.* 2012; 1(3): 292–304.

[31] Sitges C, Bornas X, Llabrés J, Noguera M, Montoya P. Linear and nonlinear analyses of EEG dynamics during non-painful somatosensory processing in chronic pain patients. *Int. J. Psychophysiol.* 2010; 77(2): 176–183.

[32] Arns M, Cerquera A, Gutiérrez RM, Hasselman F, Freund JA. Non-linear EEG analyses predict non-response to rTMS treatment in major depressive disorder. *Clin. Neurophysiol.* 2014; 125(7): 1392–1399.

[33] Geng S, Zhou W, Yuan Q, Cai D, Zeng Y. EEG non-linear feature extraction using correlation dimension and Hurst exponent. *Neurological research.* 2011; 33(9): 908–912.

[34] Von Bünau P, Meinecke FC, Scholler S, Müller KR. *Finding stationary brain sources in EEG data.* 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. 2010: 2810–2813.

[35] Rao R, Scherer R. Brain-Computer Interfacing. *IEEE Signal Processing Magazine.* 2010; 27(4): 152–150.

[36] Alotaiby T, El-Samie FE, Alshebeili SA, Ahmad I. A review of channel selection algorithms for EEG signal processing. *EURASIP Journal on Advances in Signal Processing.* 2015; 2015(1): 66.

[37] He L, Yu Z, Gu Z, Li Y. *Bhattacharyya bound based channel selection for classification of motor imageries in EEG signals.* 2009 Chinese Control and Decision Conference, CCDC 2009. 2009: 2353–2356.

[38] Yang Y, Kyrgyzov O, Wiart J, Bloch I. *Subject-specific channel selection for classification of motor imagery electroencephalographic data.* 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. 2013: 1277–1280.

[39] Kamrunnahar M, Dias NS, Schiff SJ. *Optimization of electrode channels in brain computer interfaces.* Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009. 2009: 6477–6480.

[40] Li M, Ma J, Jia S. *Optimal combination of channels selection based on common spatial pattern algorithm.* 2011 IEEE International Conference on Mechatronics and Automation. 2011: 295–300.

[41] Mitrovic N, Asif MT, Rasheed U, Dauwels J, Jaillet P. *CUR decomposition for compression and compressed sensing of large-scale traffic data.* 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013). 2013: 1475–1480.

[42] Lee H, Choi S. *CUR+ NMF for learning spectral features from large data matrix.* 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). 2008: 1592–1597.

[43] Asyraf H, Shapiai MI, Setiawan NA, Musa WW. Masking covariance for common spatial pattern as feature extraction. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC).* 2016; 8(11): 81–85.

[44] Blankertz B, Tomioka R, Lemm S, Kawanabe M, Muller KR. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal processing magazine.* 2008; 25(1): 41–56.

[45] Lotte F, Congedo M, Lécuyer A, Lamarche F, Arnaldi B. A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of neural engineering.* 2007; 4(2): R1–R13.