



OPEN

## Riemannian geometry-based transfer learning for reducing training time in c-VEP BCIs

Jiahui Ying, Qingguo Wei✉ & Xichen Zhou

One of the main problems that a brain-computer interface (BCI) face is that a training stage is required for acquiring training data to calibrate its classification model just before every use. Transfer learning is a promising method for addressing the problem. In this paper, we propose a Riemannian geometry-based transfer learning algorithm for code modulated visual evoked potential (c-VEP)-based BCIs, which can effectively reduce the calibration time without sacrificing the classification accuracy. The algorithm includes the main procedures of log-Euclidean data alignment (LEDA), super-trial construction, covariance matrix estimation, training accuracy-based subject selection (TSS) and minimum distance to mean classification. Among them, the LEDA reduces the difference in data distribution between subjects, whereas the TSS promotes the similarity between a target subject and the source subjects. The resulting performance of transfer learning is improved significantly. Sixteen subjects participated in a c-VEP BCI experiment and the recorded data were used in offline analysis. Leave-one subject-out (LOSO) cross-validation was used to evaluate the proposed algorithm on the data set. The results showed that the algorithm achieved much higher classification accuracy than the subject-specific (baseline) algorithm with the same number of training trials. Equivalently, the algorithm reduces the training time of the BCI at the same performance level and thus facilitates its application in real world.

A brain-computer interface (BCI) is a new communication system that does not rely on the participation of peripheral nerves and muscle tissues. A BCI establishes a direct connection between the brain and a computer or other electronic devices<sup>1,2</sup>. In diverse paradigms of BCIs, visual evoked potential (VEP)-based BCI has recently received increasing attention<sup>3–11</sup> because it is the BCI most likely to first gain widespread applications in real-world<sup>16,8,10</sup>. Electroencephalogram (EEG) recorded on the scalp is widely used in human BCIs due to its non-invasiveness, easiness to acquire and high time resolution.

VEP signals are responses of the brain to visual stimuli and are mainly generated over the occipital lobes. According to the difference in modulation methods of stimulus signals, VEPs can be divided into time modulated VEP (t-VEP), frequency modulated VEP (f-VEP) and pseudorandom code modulated VEP (c-VEP)<sup>12</sup>. Among them, the BCIs based on the latter two are the most potential BCIs that can achieve very high information transfer rate (ITR), the most significant performance metric for a BCI system<sup>1</sup>.

The first c-VEP BCI was proposed in 1984 and tested 8 years later on an amyotrophic lateral sclerosis (ALS) patient by Sutter<sup>13,14</sup>. The result showed that the subject could write 10–12 words/min. In 2001, Bin et al. developed a 32-target c-VEP BCI with the highest ITR (108 bits/min) among all kinds of BCIs at that time<sup>15</sup>. Subsequently, some research groups reported important progresses in c-VEP BCI studies. Thielen et al.<sup>16</sup> developed an approach to fully eliminate the tedious training stage, which is able to systematically reduce the training data in a step fashion and ultimately arrive at a calibration-free method for a c-VEP based BCIs. The results showed that the training-less BCI yielded high communication rates in an online spelling task, proving its feasibility for practical use. Wei et al.<sup>17</sup> proposed a grouping modulation-based c-VEP paradigm that divides all stimulus targets into several groups and targets per group are modulated with a distinct modulation code and its circularly shifting codes. The results indicated that the number of targets and ITR can be increased significantly with the number of groups at the cost of slight sacrifice of classification accuracy. Wittevrongel et al.<sup>18</sup> introduced a novel decoding algorithm based on spatiotemporal beamforming. They showed that this algorithm significantly outperformed an optimized support vector machine (SVM) classifier for a small number of repetitions of the coding sequence. Riechmann et al.<sup>19</sup> developed a c-VEP BCI for fulfilling everyday tasks such as navigation or action selection. The study showed that this work supports the notion of c-VEP BCIs as a particularly fast and robust approach suitable

Department of Electronic Information Engineering, School of Information Engineering, Nanchang University, Nanchang 330031, China. ✉email: wqg07@163.com

for real-world use. Waytowich et al.<sup>20</sup> presented a novel c-VEP BCI paradigm that realizes spatial decoupling of the targets and flashing stimuli via spatial separation and boundary positioning. Results showed classification accuracies for the non-foveal condition comparable with those for direct-foveal condition for longer observation lengths. Spuler et al.<sup>21</sup> made use of one-class support vector machines for creating templates and error-related potentials for target recognition and thus the reliability and accuracy of c-VEP BCIs were improved significantly.

Recently, a new family of approaches based on Riemannian geometry has been presented for EEG signal processing<sup>22–24</sup>. Unlike classical methods based on feature vector classification in Euclidean space, these approaches enable classification of EEG signals by direct manipulation of covariance matrices in Riemannian space. They have been successfully applied to BCIs based on motor imagery (MI)<sup>24–26</sup>, f-VEP<sup>27,28</sup> and P300<sup>29</sup> and achieved promising results. To the best of our knowledge, they have not been employed in c-VEP BCIs so far. Motivated by the above studies, we developed a Riemannian geometry-based classification frame for classifying c-VEP signals.

Due to the non-stationarity of EEG signals and great variability between subjects, it is necessary to acquire a large amount of training data to calibrate the classification model of a BCI system just before every use of it. This time-consuming training process makes users easy to fatigue and thus severely limits the applications of a BCI in real life. Accordingly, it is desired to develop a reliable method to reduce or even suppress the BCI calibration time while keeping the accuracy within an acceptable range. Transfer learning (TL) is a potential solution for achieving the goal. TL is defined as the ability to apply knowledge learnt in previous tasks or fields to new tasks or fields<sup>30,31</sup>. TL has received widespread attention in the field of BCI for improving the generalization performance of classifiers, such as the BCIs based on MI<sup>32–35</sup>, f-VEP<sup>28,36</sup> and c-VEP<sup>37</sup>.

In TL, how to effectively transfer knowledge embedded in EEG data from previous subjects (named source subjects hereinafter) to a new subject (named target subject hereinafter) is a crucial problem that needs to be solved. In terms of BCI systems, there are huge differences in data distribution between subjects due to the inter-subject variability. As a result, not EEG data from all source subjects can be transferred as training data to a target subject to construct a good classifier for classifying his/her testing data. To avoid negative transfer, two methods are usually used for increasing the similarity in data distributions between subjects. One is data alignment and the other is source subject selection. The former aims at reducing the difference in data distributions between subjects by aligning data to a common reference point, whereas the latter aims to select those source subjects whose data distributions are more similar to that of the target subject.

It is assumed that a small amount of training data be available from a target subject and the EEG data from some source subjects be used as his/her auxiliary training data. By the aid of a suitable TL algorithm, a robust classification model can be built for the target subject. In this paper, we propose a Riemannian geometry-based TL algorithm for c-VEP BCIs, in which EEG data are aligned by a log-Euclidean mean-based data alignment (LEDA) approach, covariance matrices are estimated from super-trials (a particular type of EEG trials), a subset of source subjects is selected by a training accuracy-based subject selection (TSS) algorithm, and finally a minimum distance to mean (MDM) classifier is constructed with the transferred data from source subjects and training data of the target subject. A c-VEP BCI dataset containing 16 subjects was used for evaluating the proposed algorithm. The results show that the algorithm is able to significantly reduce training time while maintaining high accuracy.

The major contributions of the paper are two-fold: (1) A Riemannian geometry-based classification frame is developed for c-VEP BCIs. The frame enables classification of EEG signals by direct manipulation of covariance matrices and thereby avoids the processing procedures of spatial filtering and feature extraction in classical classification models based on Euclidean space; (2) A transfer learning algorithm is proposed for reducing the training time of c-VEP BCIs without the compromise of classification accuracy. By making use of an LEDA approach and a TSS algorithm, the performance of transfer learning is enhanced significantly compared to the baseline algorithm based on subject specific learning.

## Methods

In this section, we introduce the method for classifying c-VEP signals in Riemannian space and the algorithm for transfer learning in c-VEP BCIs. The former includes the concepts such as sample covariance matrix (SCM), super-trial, covariance matrix estimation and Riemannian geometry, whereas the latter contains the concepts such as instance-based transfer learning, data alignment, source subject selection, and transfer learning-based classification. It is noted that in this paper, (stimulus) target and class are two replaceable words because each stimulus target corresponds uniquely to one category. All the methods were performed in accordance with relevant guidelines and regulations.

### Classifying c-VEP signals in Riemannian space.

- (1) SCM: In BCI experiments, a single trial consists of a task period and a resting period, and is denoted by a segment of time-windowed EEG data in the task period. Let  $X_z \in R^{N_c \times N_t}$  be a single-trial bandpass filtered EEG signal with a zero mean, where  $z \in \{1, 2, \dots, Z\}$ ,  $N_c$  and  $N_t$  denote the indices of  $Z$  classes/targets, the number of electrode channels and the number of sampling points respectively. SCM is commonly used for optimizing spatial filters and classifiers in conventional machine learning in Euclidean space. The SCM of an EEG signal  $X_z$  is usually estimated as follows

$$P_z = \frac{1}{N_t - 1} X_z X_z^T \in R^{N_c \times N_c} \quad (1)$$

- Each diagonal element of  $P_z$  holds the variance of the signal at one electrode and each of its off-diagonal elements holds the covariance between one pair of electrodes. Thereby, the SCM contains all spatial information of the EEG signal, particularly its second-order statistics.
- (2) Creating super-trials for c-VEP signals: For the BCI paradigm based on MI, the SCM suffices for discriminating EEG trials of different classes, since MI-based trials for different classes do induce different spatial patterns. For a typical BCI paradigm based on c-VEP, however, all stimulus targets are encoded by the same modulation sequence with different time lags, and thus c-VEP trials can only be differentiated by time lags or phase differences. The SCM of an EEG trial is not useful for classifying c-VEP signals because it does not possess any temporal information at all.

In order to classify the c-VEP signals in Riemannian space, it is necessary to embed the temporal information into a covariance matrix<sup>23</sup>. To this end, we need to first create a super-trial for a training trial

$$X'_z = [\bar{X}_1; \bar{X}_2; \dots; \bar{X}_Z; X_z] \in R^{N_c(Z+1) \times N_t} \quad (2)$$

where  $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_Z$  are the template signals of  $Z$  targets yielded by averaging the EEG trials of the same target, and  $X_z$  is an EEG trial from the target  $z$ . The SCM of the super-trial can be calculated as

$$P'_z = \frac{1}{N_t - 1} X'_z (X'_z)^T \in R^{N_c(Z+1) \times N_c(Z+1)} \quad (3)$$

where the dimension of  $P'_z$  is  $N_c(Z+1) \times N_c(Z+1)$ , increasing significantly compared to that of  $P_z$  in Eq. (1). The class information is embedded into covariance matrix  $P'_z$ , which can be applied for classifying c-VEP signals accordingly. Given an unlabelled testing trial  $X$ , its super-trial and corresponding SCM can be computed using Eq. (2) by replacing  $X_z$  with  $X$  and using Eq. (3) respectively.

- (3) Covariance matrix estimation: Since covariance matrices must be accurate, symmetric positive definite (SPD), and well-conditioned (this property requires that the ratio of the largest singular value to the smallest singular value cannot be too large), the choice of a covariance matrix estimator is very important. Although the SCM estimator formulated in Eq. (1) is computationally efficient, the covariance matrix estimated by it often fails to meet the above requirements, especially when the size of a covariance matrix is large such as  $P'_z$  in Eq. (3). To address the problem, shrinkage estimators<sup>27</sup> were developed as a weighted sum of the SCM  $P'_z$  and a target covariance matrix  $\Gamma$ , which is chosen to be close to the identity matrix. The shrinkage estimators are defined as

$$P_{sh} = \lambda \Gamma + (1 - \lambda) P'_z \quad (4)$$

where  $0 \leq \lambda < 1$  is a weighting coefficient. The shrinkage estimators differ in their definition of target covariance matrix  $\Gamma$ . Ledoit and Wolf<sup>38</sup> proposed  $\Gamma = \nu I_{N_c}$  with  $\nu = \text{tr}(P'_z)$ ; Blankertz<sup>39</sup> also defined  $\Gamma = \nu I_{N_c}$  but with  $\nu = \text{tr}(P'_z)/N_c$ ; Schafer and Strimmer<sup>40</sup> presented several methods for defining  $\Gamma$  relying on the SCM  $P'_z$ . The last shrinkage estimator mentioned above was used in this study because it provides a good tradeoff between performance and operational speed.

- (4) Riemannian geometry: In conventional BCI systems, many methods for feature extraction and classification developed in Euclidean space such as common spatial pattern (CSP), canonical correlation analysis (CCA) and linear discriminant analysis (LDA), are based on the SCMs of EEG signals. The SCM of an EEG trial, however, is actually an SPD matrix and thus lies in a Riemannian manifold, a subset of Euclidean space. Because the space of SPD matrices endowed with Riemannian distance is a differentiable Riemannian manifold<sup>41,42</sup>, the notions from Riemannian geometry such as Riemannian distance and Riemannian mean, can be used to analyze and classify SCMs.

Denote by  $\mathcal{S}(n) = \{S \in M(n), S^T = S\}$  the space of all  $n \times n$  symmetric matrices in the space of square real matrices  $M(n)$  and  $\mathcal{P}(n) = \{P \in \mathcal{S}(n), u^T P u > 0, \forall u \in R^n\}$  the set of all  $n \times n$  SPD matrices. As for the Riemannian distance of two SPD matrices, there are several different definitions in literature<sup>43</sup>, the most commonly used two of which are affine-invariant distance and log-Euclidean distance. The expression of the former is as follows

$$d_{AI}(P_1, P_2) = \|\log(P_1^{-1} P_2)\|_F = \sqrt{\sum_{e=1}^{N_e} \log^2 \lambda_e} \quad (5)$$

where  $\lambda_e$ ,  $e = 1, 2, \dots, N_e$  is the  $e$ th eigenvalue of matrix  $P_1^{-1} P_2$ ; whereas that of the latter is as follows

$$d_{LE}(P_1, P_2) = \|\log(P_1) - \log(P_2)\|_F \quad (6)$$

Since  $P_1$  and  $P_2$  are SPD matrices, the two types of Riemannian distances have an important property termed as congruence invariance

$$d_B(P_1, P_2) = d_B(UP_1U^T, UP_2U^T), \quad B = AI \text{ or } LE \quad (7)$$

where  $U$  denotes an invertible and orthogonal matrix. Because of the excellent property, the linear transformation of two SPD matrices will not change their relative distance in the Riemannian space.

The Riemannian mean of multiple SPD matrices is defined by their Riemannian distance with a variational approach<sup>44</sup>. Since  $P_n, n = 1, 2, \dots, N$  lies in a Riemannian manifold of non-positive curvature, the Riemannian mean is defined as the matrix that minimizes the sum of the squared Riemannian distances

$$\bar{P}_R(P_1, \dots, P_N) = \arg \min_{P \in P(n)} \sum_{i=1}^N d_R^2(P_i, P) \quad (8)$$

The existence and unicity of the Riemannian mean were proved in Ref.<sup>45</sup>. When the affine-invariant distance metric is used as Riemannian distance, an explicit solution exists only for  $N = 2$ , where it coincides with the middle point of the geodesic connecting the two SPD matrices. For  $N > 2$ , a solution can be found iteratively, and several different algorithms have been developed in Ref.<sup>46</sup>; When the log-Euclidean distance is used as Riemannian distance, an explicit solution exists and can be formulated as

$$\bar{P}_{LE}(P_1, \dots, P_N) = \arg \min_{P \in P(n)} \sum_{i=1}^N d_{LE}^2(P_i, P) = \exp \left( \sum_{i=1}^N \log(P_i) \right) \quad (9)$$

In Riemannian space, SPD matrices function as feature signals and the decoding of EEG signals is able to directly operate on them. There are two methods for achieving the task: One is the minimum distance to mean (MDM) classifier proposed by Congedo et al.<sup>23</sup> and the other is tangent space LDA (TSLDA). The former classifies an EEG trial by comparing the Riemannian distances of its covariance matrix to the Riemannian means of all classes, whereas the latter first maps covariance matrices onto tangent space, then the matrices are vectorized, and finally the feature vectors are classified with conventional algorithms from the Euclidean space. This study adopted the first method due to the simplicity of its concept.

With the concepts of distance and mean in Riemannian space, an MDM classifier can be used to classify EEG trials by directly manipulating their covariance matrices. Thereby, the classification frame avoids spatial filtering and feature extraction in conventional classification algorithm based on Euclidean geometry. In a c-VEP BCI, the training data from the reference target can be shifted to obtain the training data of other targets according to their time lags in modulation codes. Using labelled training trials, we can calculate the mean covariance matrix of each class  $z$ , denoted as  $M_z, z = 1, 2, \dots, Z$ . Then an unlabelled testing trial is assigned to the class with the closest mean, which means that the distance between the covariance matrix of the testing trial and the  $M_z$  of the assigned class  $z$  is the smallest. The detail of this classification algorithm can be referred to Ref.<sup>23</sup>.

### Transfer learning for c-VEP BCIs.

- (1) Instance-based transfer learning: Transfer learning (TL) is an important branch of machine learning<sup>30,31</sup>. In TL, a domain  $\mathcal{D}$  consists of a feature space  $\mathcal{X}$  and its marginal probability distribution  $P(X)$ , i.e.,  $\mathcal{D} = \{\mathcal{X}, P(X)\}$ , where  $X \in \mathcal{X}$ . Two domains  $\mathcal{D}_S$  and  $\mathcal{D}_T$  are different if  $\mathcal{X}_S \neq \mathcal{X}_T$  and/or  $P_S(X) \neq P_T(X)$ . A task  $\mathcal{T}$  consists of a label space  $\mathcal{Y}$  and a conditional probability distribution  $Q(\mathcal{Y}|X)$ . Two tasks  $\mathcal{T}_S$  and  $\mathcal{T}_T$  are different if  $\mathcal{Y}_S \neq \mathcal{Y}_T$  and/or  $Q_S(\mathcal{Y}|X) \neq Q_T(\mathcal{Y}|X)$ . TL uses the similarity among data, tasks, or models to transfer the data or knowledge from one domain (termed source domain) to help solve the task in another domain (termed target domain).

In most BCI applications, the feature space and label space are usually assumed the same for source and target domains because the electrode channels used for recording EEG data and the mental tasks the subjects fulfilled are the same, but their marginal and conditional probability distributions are different due to the non-stationarity of EEG signals. There are three approaches for transfer learning in the BCI field<sup>31</sup>, named feature representation transfer, instance transfer and classifier transfer. Based on the second approach, this study made use of the labelled EEG trials of source subjects to help the target subject with a small number of labelled training data calibrate his/her classification model.

- (2) Data alignment: Due to the randomness and non-stationarity of EEG signals, there are big differences in data distribution between subjects. Thereby, to transfer data from a subject to another, it is necessary to reduce the difference in data distribution between them prior to transferring data. Data alignment (DA) is an efficient method for achieving the goal<sup>47–49</sup>. DA aligns either the EEG trials or their covariance matrices from different domains (i.e., subjects or sessions) to a common reference and thus reduces the marginal probability distribution shift of different domains. DA can be used as a preprocessing step to enhance similarity between EEG trials of different subjects. Several DA approaches exist in literature such as Riemannian alignment (RA)<sup>47</sup>, Euclidean alignment (EA)<sup>48</sup> and online pre-alignment (OPS)<sup>49</sup>, depending upon how the reference matrix is defined. LEDA, whose reference matrix is derived from log-Euclidean mean, is applied in the TL framework because it outperforms other approaches for c-VEP data.

LEDA can be used in either supervised or unsupervised manner. For testing data from the target subject, LEDA is unsupervised since they do not have any labelled information. LEDA calculates the log-Euclidean mean  $M_{LE}$  of the covariance matrices from all classes using Eq. (9), and the reference matrix is calculated as  $M_{ref} = P_{LE}^{-1/2}$ . A single-trial EEG signal (or rather a super-trial)  $X_i$  is aligned as

$$X'_i = M_{ref} X_i \quad (10)$$

For the EEG data from source subjects and a few labelled training data from the target subject, LEDA can be used in a supervised manner. The log-Euclidean mean  $M_{LE,z}$  of covariance matrices can be calculated using all

EEG trials belonging to the same class  $z \in \{1, 2, \dots, Z\}$ , and the reference matrix is  $M_{ref,z} = M_{LE,z}^{-1/2}$ . A single-trial EEG signal  $X_{i,z}$  is aligned as follows

$$X'_{i,z} = M_{ref,z} X_{i,z} \quad (11)$$

DA does not change the relative distance between any two covariance matrices in a data set to be aligned because of the property of congruence invariance in Eq. (7). After LEDA, marginal probability distribution shift of EEG data between subjects is reduced, and resulting EEG trials are whitened accordingly. The distribution of EEG trials from different subjects tends to be consistent, which is conducive to construction of the classifier of a target subject.

- (3) Source subject selection: Subject selection (SS) aims to determine an optimal subset of source subjects for a target subject to ensure the highest classification accuracy of his/her testing data with the help of EEG data from the chosen source subjects. Lotte et al.<sup>50</sup> used the sequential forward floating search algorithm to select the subjects to be added to or removed from the current subject subset in order to maximize the accuracy for the target subject. Qi et al.<sup>51</sup> proposed to use Riemannian distance to select samples from the existing subject database to help construction of the classifier. Giles et al.<sup>52</sup> proposed a new similarity measurement algorithm, called Jensen Shannon ratio (JSR), which is used to compare the data of the target subject with the existing subject database, and to select the previously calibrated BCI model with the highest similarity with the target subject as the target model of BCI. Azab et al.<sup>32</sup> proposed a similarity measurement method based on Kullback–Leibler divergence (KL), which is used to measure the similarity between two feature spaces. Despite these methods, SS remains an open and tough problem.

---

Algorithm 1 Training accuracy-based subject selection (TSS)

---

Input:  $P_T$ : Covariance matrix of the target subject;

$P_S^N$ : Covariance matrix of  $N$  source subjects;

Output:  $P_S^C$ : Covariance matrix of  $C$  chosen source subjects with high similarity to the target subject;

$C$ : The number of chosen source subjects.

---

1: Use  $acc_1(i) = \text{trainThenTest}(P_S^i, P_T)$  to obtain  $\{acc_1(i), i = 1 : N\}$

2: Sort  $\{acc_1(i), i = 1 : N\}$  in descending order, then obtain the index

3: Use  $acc_2(n) = \text{trainThenTest}(P_S^{\text{index}(1:n)}, P_T)$  to obtain  $\{acc_2(n), n = 1 : N\}$

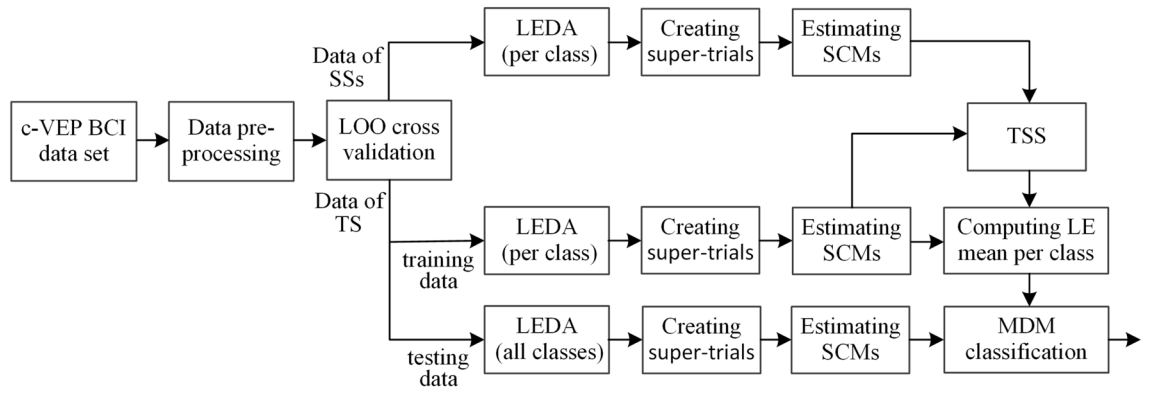
4: Obtain the  $\text{maxIndex}$  by  $\max(acc_2)$

5: Then  $P_S^C = P_S^{\text{index}(1:\text{maxIndex})}$ ,  $C = \text{index}(1:\text{maxIndex})$

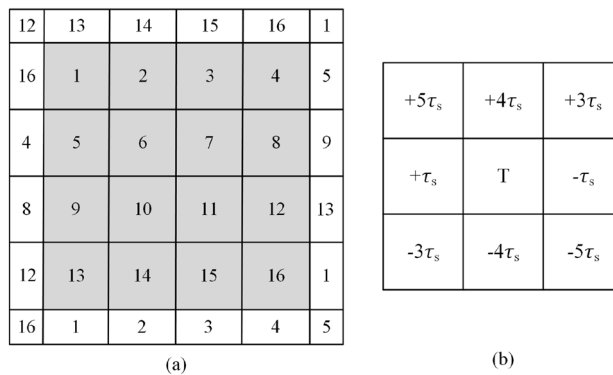
---

We propose a training accuracy-based subject selection (TSS) algorithm shown in Algorithm 1, where the function  $acc = \text{train Then Test}(P_S^N, P_T)$  returns the accuracy obtained when the MDM is trained on the data set  $P_S^N$  from  $N$  source subjects and then is used to classify training data set  $P_T$  from the target subject. All source subjects are sorted by their classification accuracies. The first  $C$  source subjects are selected as transferred subjects, if their EEG data are pooled together to re-train an MDM classifier and it achieves the highest accuracy on data set  $P_T$  among all subsets of source subjects. The processing procedures of the algorithm are summarized as follows: All source subjects for a target subject are first sorted according to their classification accuracies by using the EEG data from a single source subject as training data and classifying the training data from the target subject. Then  $c$  (from 1 to 15) source subjects are sequentially extracted from sorting list and their EEG data are collectively used for creating a classification model, which is used to classify the training data of the target subject. Finally, the  $C$  is decided as the  $c$  value that yields the highest classification accuracy.

- (4) Transfer learning-based classification: Given a c-VEP BCI data set, the EEG data from each subject are first preprocessed by bandpass filtering and data segmenting into single-trial signals. The bandpass filtering is done between 2 and 30 Hz via a Butterworth IIR filter of order 8. A leave-one subject-out (LOO) cross validation strategy is used to divide all subjects in the data set into the target subject and source subjects, i.e., each subject in the data set acts as the target subject once, and the remaining subjects act as the source subjects. Then the preprocessed EEG trials from a subject are aligned with LEDA in either supervised or unsupervised manner depending on the types of the trials. For each source subject, the EEG trials function as training data and are aligned by class; for the target subject, the training trials are also aligned by class, whereas the testing trials from all classes are aligned as a whole. Next, for an aligned trial, a super-trial is created with template signals of all targets and an SCM is estimated using the specified shrinkage estimator. Next, the transferred subjects are selected from source subjects according to the TSS algorithm. Finally, the EEG trials of the transferred subjects and the training trials of the target subject are pooled together as overall training data and are employed to computing the log-Euclidean means of all classes, which are



**Figure 1.** Processing flowchart of the Riemannian geometry-based transfer learning in c-VEP BCIs. The meanings of the acronyms in the figure are as follows. LOO: leave-one subject-out; SSs: source subjects; TS: target subject; LE(DA): log-Euclidean (data alignment); SCMs: sample covariance matrices; TSS: training accuracy-based subject selection; MDM: minimum distance to mean.



**Figure 2.** (a) The visual stimulator used in the study. It consists of 16 target stimuli in the gray area and 20 complementary non-target stimuli in the white area. The 16 target stimuli are modulated by a 63-bit M sequence and its circularly shifting sequences. The target and non-target stimuli tagged with the same numbers are modulated by the same modulation code. (b) The principle of equivalent neighbors is used in design of the stimulator. Each target T and the eight adjacent targets around it maintain a fixed time delay relationship.  $\tau_s$  (4 bits in the study) is the time lag of modulation codes between two adjacent target stimuli.

utilized to classify the SCMs of testing trials from the target subject with an MDM classifier. The algorithmic flowchart of the transfer learning-based c-VEP BCI classification in Riemannian space is shown in Fig. 1.

### BCI experiment

**Subjects.** An offline experiment was designed for the c-VEP BCI. Sixteen healthy subjects (7 females, aged 21 ~ 26 years, mean age: 23 years) with normal or corrected to normal vision participated in the experiment. None of them had a previous history of epilepsy or seizures, which can be induced by flashing stimuli. Six subjects participated in a c-VEP BCI experiment before and were familiar with the experimental setting. Each subject was asked to read and sign an informed consent form before the experiment. The study was approved by the Human Research and Ethics Committee, Nanchang University. After the experiment, these subjects were paid for their contribution to the study.

**Target modulation.** A visual stimulator comprising 16 target stimuli is shown in Fig. 2a, the design principle of which is similar to that of Bin et al.<sup>12</sup>. The stimulator was presented on an LCD monitor. Principle of equivalent neighbors was adopted for designing the stimulator. 16 target stimuli and 20 complementary non-target stimuli are arranged in grey and white area respectively. All these stimuli are tightly placed without intervals between them. The 16 target stimuli are modulated by a 63-bit pseudorandom M sequence and its circularly shifting sequences. The 63-bit M sequence corresponds to a stimulus period of  $P = 63/60 = 1.05$  s for the screen refresh rate of 60 Hz, and is circularly shifted integral multiples of 4 bits to generate 16 sequences of different phases used for modulating the 16 target stimuli. The target and non-target stimuli tagged with the same numbers are modulated by the same sequence. The purpose of these non-target stimuli is to ensure that the target stimuli have equivalent neighbors in the directions of left, right, up, down and diagonals, i.e. the time lags

between modulation codes of target stimuli and their neighbors in same directions are the same, as shown in Fig. 2b.

**Experimental setup.** The experimental system consists of a personal computer (Lenovo China) operated under a Windows 7 system and an EEG amplifier (a Synamps2 system, Neuroscan Inc.) with 64 EEG channels. The computer has a 24-in. liquid crystal display (LCD) monitor and a parallel port linked to the amplifier. The LCD monitor has a refresh rate of 60 Hz and a resolution of  $1920 \times 1080$  pixels. Stimulus presentation was operated in the PC and controlled by the stimulus program developed under Visual C++ 9.0 platform. DirectX (Microsoft Inc.) was employed to ensure the stability of frame-based rendering of stimuli. Event trigger signals yielded by the stimulus program were sent from the parallel port of the computer to the EEG amplifier and recorded on an event channel for synchronizing stimulus presentation and EEG data recordings. Each target stimulus was rendered within a square of  $140 \times 140$  pixels.

During the experiment, all stimuli were flashed simultaneously and periodically according to their modulation codes. Each subject was seated in a comfortable chair in an unshielded, dimly lit room, approximately 60 cm away from the monitor. They were instructed to focus attention to an intended target, gaze at its center and not to blink as much as possible. Nine Ag/AgCl electrodes over the occipital lobes (P3, Pz, P4, PO7, POz, PO8, O1, Oz, O2) in line with international 10/20 system were used for recording EEG data. The reference electrode was positioned at the vertex. Electrode impedances were kept below  $10 \text{ k}\Omega$  during the experiment. The EEG signals were digitized at a sampling rate of 1000 Hz.

**Data acquisition.** This experiment was divided into two phases: training and testing. The purpose of the training stage was to acquire training data of a reference target. Any one of the 16 targets could be specified as the reference target, which was the 11th target in the study. The subject should fixate continuously at the reference target for 100 stimulus cycles (i.e. 100 trials of one stimulus cycle). In testing stage, the subject was asked to complete the testing task of all 16 targets, which appeared in a random order. The subject was required to fixate continuously on the current target for five stimulus cycles (i.e., five trials), and subsequently, the test of next target began immediately. Each test started with a prompting duration of 1 s, in which the target turned red and the subject should shift his/her gaze to the target as soon as possible. After the cue duration was over, the red target cue was switched to a small cycle at bottom right corner of the target so that the subject found it easily.

## Results

The difference in data distribution between the raw data and the aligned data is visualized in 2-dimensional space to examine the effect of LEDA. The following four Riemannian geometry-based algorithms are compared using the first algorithm as baseline: (1) subject-specific learning (SSL); (2) transfer learning using EEG data of all source subjects without data alignment (TL-ASS); (3) transfer learning using EEG data of all source subjects with LEDA (TL-LEDA-ASS); and (4) transfer learning using EEG data of chosen source subjects by TSS with LEDA (TL-LEDA-TSS). It is noted that for each source subject, only testing data (5 trials per target) are used for transfer learning because each subject had only training data from one stimulus target. The proposed classification algorithm is evaluated by classification accuracy and ITR based on varying lengths of data, numbers of channels and numbers of training trials. The ITR in bits/minute defined by Wolpaw et al.<sup>1</sup> is calculated as follows

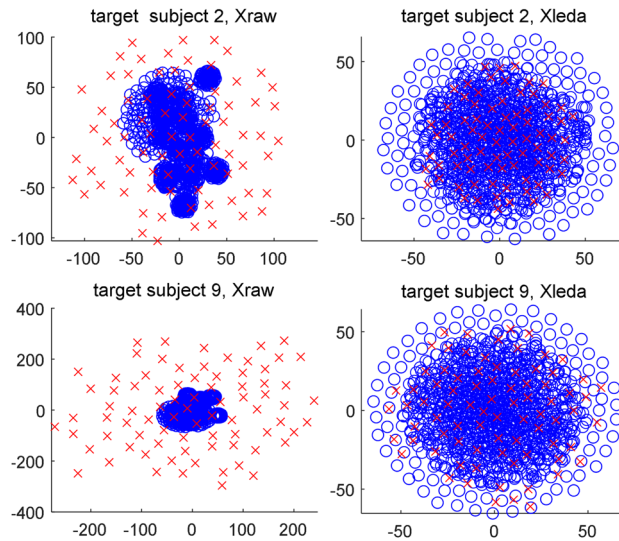
$$ITR = \left( \log_2 M + P \log_2 P + (1 - P) \log_2 \left[ \frac{1 - P}{M - 1} \right] \right) * \left( \frac{60}{T} \right) \quad (12)$$

where  $M$  is the number of targets,  $P$  is the detection accuracy of targets, and  $T$  in seconds/selection is the average time for a selection. For the estimation of ITRs, the time of 1 s for gaze-shifting was included in target selection.

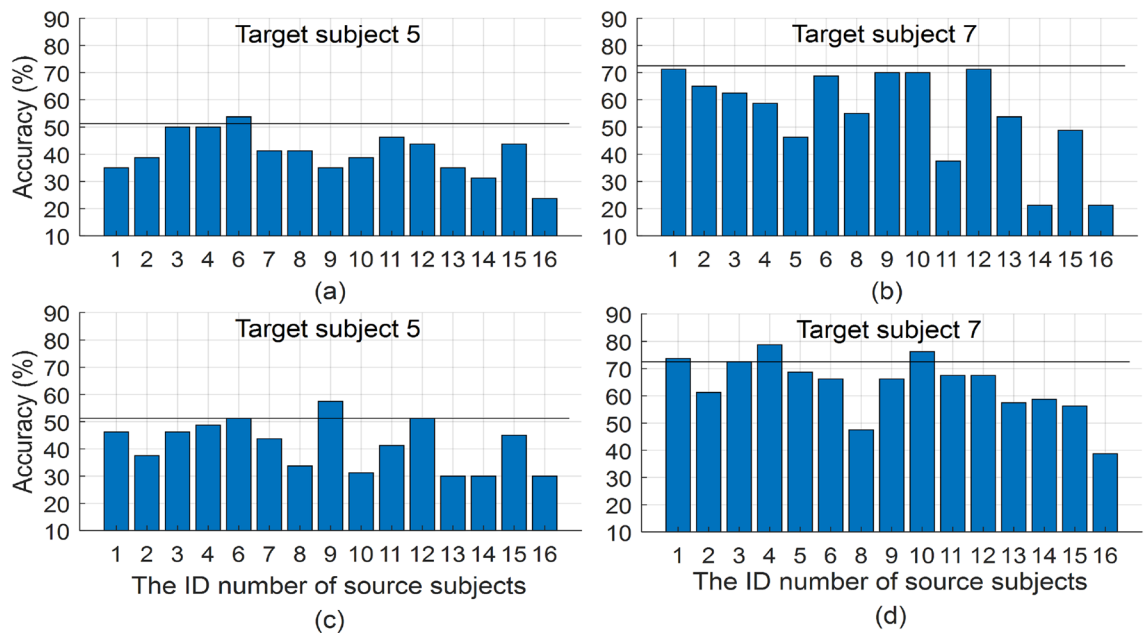
This definition for ITR is a simplified computational model based on Shannon channel theory under several assumptions and the most widely used metric to assess the overall performance of a BCI system. Although several scholars, e.g., Yuan et al.<sup>53</sup>, pointed out its shortcomings and proposed some improvements, the vast majority of studies still use this definition to measure the performance of BCI systems so far.

**Data visualization.** t-distributed stochastic neighbor embedding (t-SNE)<sup>54</sup> is a commonly used method for data dimensionality reduction and visualization. High-dimensional data can be projected to a 2- or 3-dimensional space by t-SNE and their characteristics in distribution are observed in the space. We visualize whether LEDA can decrease the difference in data distributions between the target subject and source subjects. Figure 3 shows two examples of 2-dimensional distributions of the raw data and the aligned data by LEDA. In each subplot, the red symbols "x" indicate the EEG trials of the target subject, whereas the blue symbols "o" denote the EEG trials of all source subjects. From the figure, it is clearly observed that for the raw data, the data distributions of the target subject and the source subjects are totally different, whereas for the aligned data, they tend to be consistent. Thereby, the data alignment by LEDA is very conducive to decreasing the difference in data distribution between subjects and subsequent transfer learning for c-VEP BCIs.

**Transferring data from a single source subject.** Figure 4 depicts the classification accuracies of two target subjects (#5 and #7) yielded by the two transfer learning algorithms TL (the first row) and TL-LEDA (the second row) with the EEG trials (5 trials per class) of only one source subject transferred to the target subject. The accuracy achieved by SSL algorithm is also shown with a solid line as a baseline in each subplot. The same 5 training trials per stimulus target (obtained by shifting the 5 trials of the reference target) from the target subject were used for building classification model in both transfer algorithms and SSL algorithm. From the figure, it is



**Figure 3.** t-SNE visualization of the distribution of raw data (Xraw, the first column) and aligned data by LEDA (Xleda, the second column). The red symbols ‘x’ denote the EEG trials from the target subject (16 × 5 = 80 trials), whereas the blue symbols ‘o’ stand for the EEG trials from all source subjects (80 × 15 = 1200 trials).



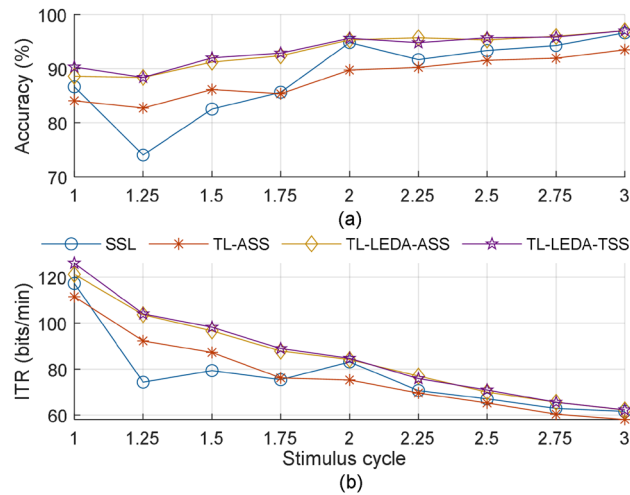
**Figure 4.** Classification accuracies of two target subjects (#5 and #7) yielded by the two transfer learning algorithms TL (a,b) and TL-LEDA (c,d) with the EEG trials of only one source subject transferred to the target subject. The solid line in each subplot stands for the accuracy yielded by SSL (baseline) algorithm.

easily observed that whether the EEG trials are aligned or not, transferring EEG data between different pairs of source-target subjects, classification accuracies differ substantially. In most cases, such a transfer was negative because it deteriorated the accuracy. However, data distributions of some sources are complementary and their overall distribution may be more similar to that of the target subject. To avoid negative transfers, it is necessary to select those relevant source subjects whose total data distribution is more similar to that of the target subject.

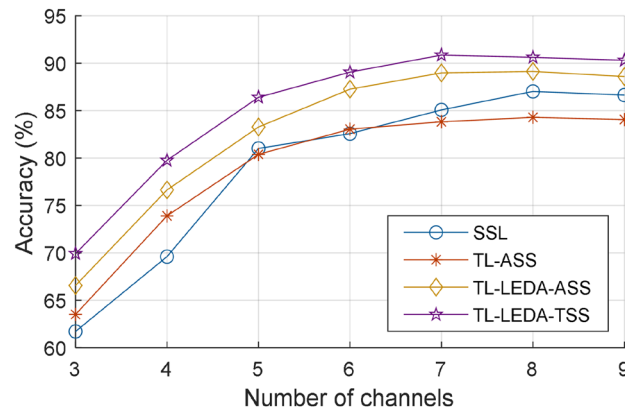
**Transferring data from multiple source subjects.**

- (1) Data length: Fig. 5 illustrates the relationship between the averaged accuracy and ITR of the four algorithms (SSL, TL-ASS, TL-LEDA-ASS and TL-LEDA-TSS) across subjects and the data length used for target recognition. The data length is denoted as multiple of the stimulus cycle, and 9 different multiples



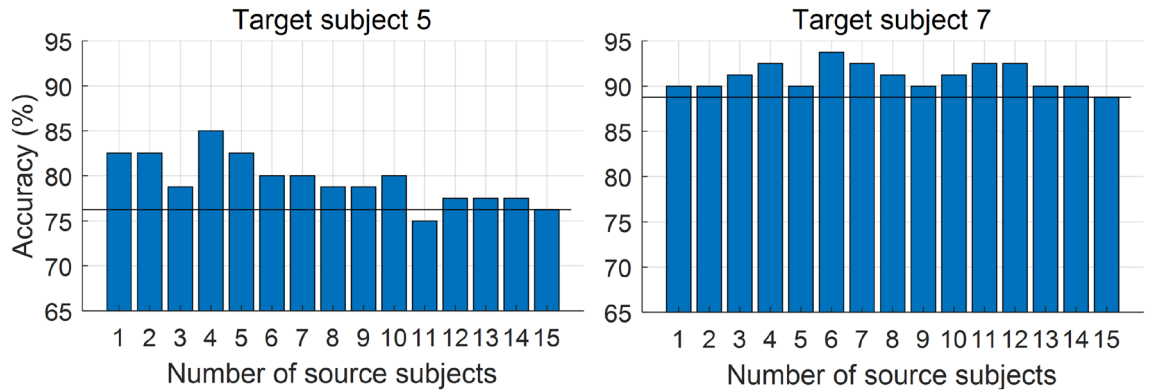


**Figure 5.** Relationship between the averaged accuracy (a) and ITR (b) of the four algorithms (SSL, TL-ASS, TL-LEDA-ASS and TL-LEDA-TSS) across subjects and the data length used for target recognition. The data length is denoted as the multiple of the stimulus cycle.



**Figure 6.** Relationship between the averaged accuracy across subjects and the number of channels used for target recognition for the four algorithms SSL, TL-ASS, TL-LEDA-ASS and TL-LEDA-TSS.

- are taken from 1 to 3 with an interval of 0.25, i.e.  $k = 1, 1.25, \dots, 3$ . The number of training trials and the number of channels used for classification are 20 and 9 respectively. It is easily seen from the figure that although the overall trend of the accuracy yielded by each algorithm is to increase with data length, the ITR indeed decreases monotonically. Statistical analysis based on paired t-tests showed that the ITRs of the four algorithms at one stimulus cycle are significantly higher than those at all other data lengths with  $p < 0.001$ . Therefore, the data length of one stimulus cycle was used in the following analysis.
- (2) The number of channels: Fig. 6 shows the relationship between the averaged classification accuracy across subjects and the number of channels for each of the four algorithms SSL, TL-ASS, TL-LEDA-ASS and TL-LEDA-TSS. In the study, the nine channels (P3, Pz, P4, PO7, POz, PO8, O1, Oz, O2) were used to record EEG data and kept sequentially in the dataset. The number of channels used for classification was taken from 3 to 9 with the step length of 1. To simplify the analysis, the channels were selected sequentially according to their order in the dataset. The number of training trials used for classification is 20. As shown in the figure, accuracies of all the four algorithms first rise rapidly, then slow down and fall after reaching the maxima. For SSL, TL-ASS and TL-LEDA-ASS, the maximal accuracies were achieved at 8 channels, whereas for TL-LEDA-TSS, that was derived at 7 channels. Paired t-tests showed that TL-LEDA-TSS has no significant differences in accuracy between 7 and 8 channels with  $p = 0.10$ . Thus, 8 channels were used for following analysis.
  - (3) The number of chosen source subjects: In this study, we screened the source subjects with Algorithm 1. It is noted that the testing data set  $P_T$  of the target subject was derived from training data of all targets obtained by shifting those of the reference target, whereas the training data set  $P_S^N$  of  $N$  source subjects was derived from the testing data, i.e., 5 trials per target. Figure 7 shows the classification accuracies of the two target subjects (#5 and #7 as examples) varying with the number of source subjects used for transferring data.



**Figure 7.** The classification accuracies of the two target subjects (#5 and #7) change with the number of chosen source subjects. The solid line in each subplot denotes the accuracy of SSL (baseline) algorithm for comparison.

TS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	M
Acc1	100 (2)	100 (7)	100 (15)	97.5 (1)	82.5 (4)	100 (1)	90 (15)	90 (9)	100 (6)	97.5 (15)	88.75 (7)	100 (8)	75 (5)	86.25 (15)	91.25 (10)	62.5 (1)	91.33 (7.56)
Acc2	98.75 (1)	98.75 (1)	95 (3)	92.5 (10)	75 (1)	95 (15)	75 (2)	70 (2)	97.5 (8)	92.5 (14)	83.75 (2)	95 (1)	65 (12)	65 (2)	87.5 (1)	62.5 (1)	84.3 (4.75)
Acc3	100 (13)	100 (9)	98.75 (9)	97.5 (4)	81.25 (6)	98.75 (11)	91.25 (8)	87.5 (9)	100 (7)	97.5 (8)	90 (8)	100 (8)	72.5 (5)	82.5 (10)	86.25 (8)	62.5 (7)	90.39 (8.13)

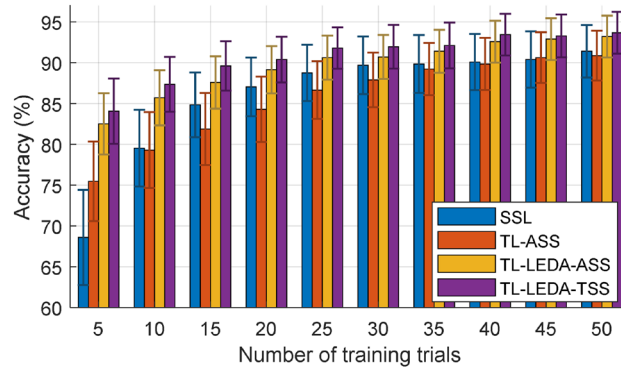
**Table 1.** The classification accuracies (%) of the 16 target subjects (TS) and their mean (M) under three conditions: Acc1—the maximum accuracy achieved actually; Acc2—the minimum accuracy achieved actually; Acc3—the accuracy yielded by TL-LEDA-TSS algorithm. The digits in parentheses denote the corresponding number of source subjects.

The number of training trials is taken as 20. The solid line in each subplot denotes the accuracy of SSL (baseline) algorithm for comparison. Obviously, the accuracy does not increase monotonically with the number of source subjects, but fluctuates up and down near the baseline accuracy. Based on the TL-LEDA-TSS algorithm, the number of chosen source subjects yielding the highest accuracy is 4 and 6 for the two target subjects respectively.

Table 1 lists the classification accuracies of the 16 target subjects under three conditions: Acc1—the maximum accuracy achieved actually; Acc2—the minimum accuracy achieved actually; Acc3—the accuracy yielded by TL-LEDA-TSS algorithm. The corresponding number of transferred source subjects is shown in parentheses. As shown in the table, the accuracy of each target subject yielded by TL-LEDA-TSS algorithm (Acc3) is not identical to the maximum accuracy achieved actually (Acc1), so is their corresponding numbers of chosen source subjects. The mean accuracy of all target subjects achieved by the algorithm is lower than Acc1, meaning that this TSS algorithm is not optimal. However, the TSS algorithm achieved a near-optimal mean accuracy much higher than the minimum accuracy achieved actually. It is better to show which source subjects were selected for a target subject. We ignored this problem for sparing space.

- (4) The number of training trials: Fig. 8 illustrates the averaged classification accuracies across all subjects of the three transfer learning algorithms varying with the number of training trials, which ranges from 5 to 50 with the interval of 5 trials, i.e.,  $N_{tr} = 5, 10, \dots, 50$ . The classification accuracy of SSL algorithm is shown as baseline for comparison. Clearly, the accuracy of each algorithm increases consistently with the number of training trials, but the increase slows down as it gets bigger. The accuracies of TL-ASS are always lower than those of SSL except for  $N_{tr} = 5$  trials, meaning that without additional processing of c-VEP data, directly transferring them from source domains to a target domain is generally not feasible.

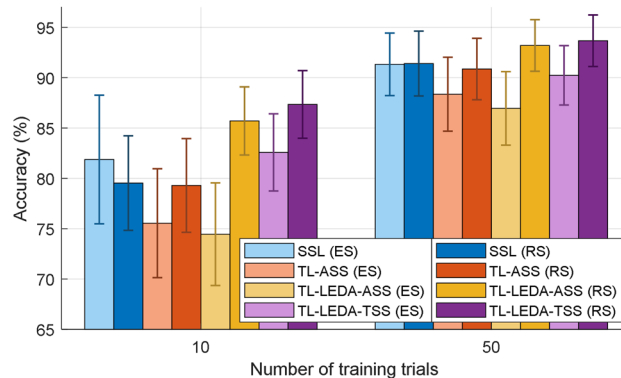
At all numbers of training trials, the accuracies of TL-LEDA-ASS are higher than those of SSL and the accuracies of TL-LEDA-TSS are higher than those of TL-LEDA-ASS, meaning that both data alignment and subject selection promoted the performance of transfer learning. The fewer the training trials, the greater is the performance improvement. Especially when the number of training trials is 5, the accuracies of TL-LEDA-TSS and TL-LEDA-ASS are 13.6% and 11.33% higher than that of SSL respectively. This makes sense for transfer learning because its goal is to decrease calibration time. Based on paired t-tests, statistically significant differences in accuracy ( $p$  values) among the four algorithms at 95% confidence level are shown in Table 2.



**Figure 8.** Averaged classification accuracies across all subjects of the four algorithms varying with the number of training trials, which ranges from 5 to 50 with the interval of 5 trials. The error bars represent the standard error.

NT	M1-M2	M1-M3	M1-M4	M2-M3	M2-M4	M3-M4
5	0.090	<b>0.001</b>	<b>0.0002</b>	<b>0.011</b>	<b>0.005</b>	0.294
10	0.938	<b>0.017</b>	<b>0.003</b>	<b>0.011</b>	<b>0.010</b>	0.160
15	0.212	0.126	<b>0.002</b>	<b>0.016</b>	<b>0.005</b>	0.056
20	0.186	0.119	<b>0.006</b>	<b>0.026</b>	<b>0.013</b>	0.109
25	0.250	0.229	<b>0.019</b>	<b>0.032</b>	<b>0.009</b>	0.123
30	0.252	0.508	<b>0.046</b>	0.052	<b>0.010</b>	0.135
35	0.634	0.291	<b>0.032</b>	0.078	<b>0.014</b>	0.357
40	0.832	0.070	<b>0.013</b>	<b>0.022</b>	<b>0.007</b>	0.094
45	0.800	0.052	<b>0.012</b>	<b>0.030</b>	<b>0.009</b>	0.237
50	0.531	0.070	<b>0.021</b>	<b>0.015</b>	<b>0.009</b>	0.211

**Table 2.** Statistically significant differences in accuracy ( $p$  values) among the four algorithms based on paired t-tests at the confidence level of 95%. NT: Number of training trials; M1: SSL; M2: TL-ASS; M3: TL-LEDA-ASS; M4: TL-LEDA-TSS. Significant values are in bold.



**Figure 9.** Averaged classification accuracies across subjects yielded by the two groups of algorithms derived from Euclidean and Riemannian space respectively in the cases of 10 and 50 training trials. ES and RS denote Euclidean space and Riemannian space respectively. The error bars represent the standard error.

- (5) Riemannian space vs Euclidean space: Fig. 9 shows the comparison of averaged classification accuracies across subjects yielded by the two groups of algorithms developed in Euclidean and Riemannian space respectively in the cases of 10 and 50 training trials. It is clearly observed from the figure that for the two numbers of training trials, the subject-specific learning algorithm from Euclidean space performs better than or equal to that from Riemannian space, whereas the three transfer learning algorithms from Riemannian space are superior to those from Euclidean space. Using the subject-specific algorithm from

Number of trials	5	10	15	20	25	30	35	40	45	50
SSL	<b>0.002</b>	0.469	<b>0.011</b>	<b>0.013</b>	<b>0.045</b>	0.054	0.116	0.052	0.155	0.876
TL-ASS	0.287	0.277	0.327	0.354	0.243	0.202	0.101	0.159	0.098	0.075
TL-LEDA-ASS	<b>0.004</b>	<b>0.003</b>	<b>0.002</b>	<b>0.001</b>	<b>0.001</b>	<b>0.002</b>	<b>0.004</b>	<b>0.003</b>	<b>0.002</b>	<b>0.003</b>
TL-LEDA-TSS	<b>0.033</b>	<b>0.037</b>	<b>0.016</b>	<b>0.016</b>	<b>0.007</b>	<b>0.035</b>	<b>0.045</b>	<b>0.010</b>	<b>0.011</b>	<b>0.003</b>

**Table 3.** Statistically significant difference in accuracy ( $p$  values) between each pair of algorithms developed in Riemannian and Euclidean space respectively based on paired t-tests at 95% confidence level, at 10 numbers of training trials. Significant values are in bold.

- Euclidean space (SSL (ES)) as a baseline, the two transfer learning algorithms from Riemannian space (TL-LEDA-ASS (RS) and TL-LEDA-TSS (RS)) still improved the classification performance, especially when the number of training trials is small. For all 10 numbers of training trials, statistically significant difference in accuracy ( $p$  values) between each pair of algorithms derived from Riemannian and Euclidean space based on paired t-tests at 95% confidence level is listed in Table 3.
- (6) Computational complexity: As shown in Fig. 1, the classification process of this proposed algorithm TL-LEDA-TSS includes two phases of training and testing. In the training stage, the algorithm contains the five procedures of aligning data, creating super-trials, estimating SCMs, selecting source subjects and computing log-Euclidean mean per class; In the testing stage, the algorithm contains four procedures, the first three of which are the same as those in training stage, whereas the last one is to classify single-trial data based on the MDM classifier. The training stage is long because the shrinkage approach is used to estimate SCMs and numerous trials are required for training, whereas the testing stage is much shorter because data processing in online experiments is done trial by trial. The evaluation process was carried out on a computer with the configuration of Intel (R) Core (TM) i5-6500 CPU @ 3.20 GHz, 8.00 GB RAM and 64-bit OS. Assume that 20 training trials are available from a target subject and all source subjects are acted as transferred ones, the training stage takes about 9.8 min. Fortunately, the log-Euclidean mean per class used for testing stage can be calculated offline. Classifying a single-trial EEG data takes only 0.75 s, which is suitable for online processing.

## Discussion

A typical c-VEP BCI system requires a training stage to obtain sufficient calibration data. However, this cumbersome training procedure will limit its practical application. How to reduce the training time without sacrificing the accuracy is one of the main research directions. The results in this study indicated that compared to the subject-specific (baseline) algorithm, the transfer learning algorithm incorporating LEDA and TSS significantly increased the classification accuracy at each given number of training trials. In other words, the proposed algorithm achieved the similar accuracy to the baseline algorithm using much fewer training trials, and thus reduced the training time substantially.

The experimental results show that TL-LEDA-ASS is better than TL-ASS, and the TL-LEDA-TSS is better than TL-LEDA-ASS. These findings prove that both data alignment and source subject selection can improve transfer learning. At the same time, both TL-LEDA-TSS and TL-LEDA-ASS are better than SSL especially in the case of fewer training trials, where the improvement in accuracy is more obvious. These conclusions support our goal to reduce training time while maintaining high accuracy.

In the previous studies on c-VEP BCI, we did some meaningful work. We first made a study on stimulus specificity of c-VEP BCIs<sup>55</sup>. Five experiments were devised to investigate the effect of size, color and proximity of the stimuli, length of modulation sequence and the lag between two adjacent stimuli on target recognition. An optimal value of each parameter was attained in terms of classification performance, and thus provides a basis for designing a high-performance c-VEP BCI. To make BCI suitable for complex applications like word input, we then presented a novel c-VEP BCI paradigm for increasing the number of stimulus targets based on grouping modulation with different codes<sup>17</sup>. Using the paradigm, a BCI with 48 targets divided into three groups was implemented with a high ITR of 181 bits/min. Finally, we further expended the paradigm to four target groups and implemented a c-VEP BCI of 64 stimulus targets with an ITR of 184.6 bit/min<sup>56</sup>. All these methods were developed in Euclidian space. In this study, we developed a Riemannian geometry-based classification frame for c-VEP BCIs.

As mentioned in section Introduction, the methodologies of Riemannian geometry have been successfully applied to BCIs based on MI, f-VEP and P300. For MI-based BCIs, Riemannian geometry is mainly used for matching the statistical distributions of two datasets so that transferring data from source subjects to a target subject becomes effective. Zanini et al.<sup>47</sup> proposed a RA approach that aligns the covariance matrices of a subject with the reference matrix estimated by the EEG data of resting states, i.e., the transitional periods between two trials. Rodrigues et al.<sup>24</sup> presented a Procrustes analysis-based method for aligning two data sets using geometrical transformations (translation, scaling and rotation) over the data points. Tang et al.<sup>26</sup> proposed a Riemannian mean-based rotation alignment (RMRA) domain adaptation method by rotating the SPD matrix in Riemannian space. Another study conducted by Kalaganis et al.<sup>25</sup> presented a method called discriminative

covariance reduction (DCR) to reduce the dimensionality of covariance matrices by turning the problem of sensor selection as a maximization of a function over the manifold of SPD matrices. In terms of f-VEP-based BCIs, Kalunga et al.<sup>27</sup> conducted the first study of online classification based on Riemannian geometry, in which they proposed a novel algorithm for asynchronous processing of brain signals, borrowing the principles of from semi-supervised approaches and following a dynamic stopping scheme to provide a prediction as soon as possible. In another study, Kalunga et al.<sup>28</sup> presented a transfer learning method for SSVEP-based BCIs by making use of available data from previous users, relying on Riemannian geometry to estimate the similarity with a multi-user dataset and borrowing the notion of composite mean to partition the space. As for P300-based BCI, Li et al.<sup>29</sup> proposed a transfer learning algorithm that combines XDAWN spatial filter and Riemannian geometry classifier. The XDAWN is used to enhance the P300 component in the raw signal as well as reduce its dimensions. The Riemannian mean is acted as the reference matrix to perform the affine transformation of SPD matrices that makes the data from different subjects comparable.

When Riemannian geometry methodologies are applied to c-VEP based BCIs, the dimensionality of covariance matrices is  $N_c(Z + 1) \times N_c(Z + 1)$  as shown in Eq. (3), which is high when both the number of channels  $N_c$  and the number of stimulus targets  $Z$  are large, resulting in high computational complexity of class means. To address the problem, a method for channel selection is usually necessary to select a small number of discriminative channels for given brain tasks. Recently, Jin et al. proposed two methods<sup>57,58</sup> for improving the classification performance of MI-based BCIs. They are based on the assumption that the channels related to MI should contain common information when participants are executing the same tasks. One method is called the correlation-based channel selection (CCS) and the other is named bispectrum-based channel selection (BCS). The former aims to select task-related channels and meanwhile exclude those channels containing redundant information and noise. A novel regularized CSP is used to extract effective features and an SVM classifier with the radial basis function is trained to accurately identify MI tasks; The latter aims at extracting non-linear and non-Gaussian information from EEG signals with bispectrum analysis. The method uses the sum of logarithmic amplitudes (SLA) and the first-order spectral moment (FOSM) to select EEG channels without redundant information. In the previous study on electrocorticogram (ECoG)-based MI BCIs, we proposed a wrapper method for selecting a suitable number of task-related channels to increase the discriminability of feature signals<sup>59</sup>. CSP algorithm was used for feature extraction, and channel selection was performed by genetic algorithm (GA) for optimizing the feature extraction. Fisher discriminant analysis (FDA) was used as the classifier, and the channel subset chosen at each generation was evaluated by classification accuracy. As for the current study, a channel selection algorithm can decrease the number of channels and resulting computational complexity in the estimation of SCMs, but this is beyond the scope of this study.

This study evaluated the proposed algorithm offline and applied it to a c-VEP BCI containing 16 stimulus targets. The effectiveness of the proposed algorithm is verified in terms of reducing training time. One limitation of the proposed algorithm is that the dimensionality of super-trials increases rapidly with the number of stimulus targets, leading to long time for estimating SCMs with the shrinkage approach and the consequent difficulty in estimating class means. Another limitation is that all subjects participated in the study are healthy ones without the involvement of the people with severe neural impairments. Future work will focus on generalizing the algorithm to c-VEP BCIs with more stimulus targets, experimenting on subjects with nerve injury and its online implementation.

## Conclusion

This paper proposes a transfer learning algorithm of c-VEP BCIs based on Riemannian geometry, which combines data alignment and subject selection to improve the performance of transfer learning. In the proposed algorithm, log-Euclidean data alignment (LEDA) is employed to reduce the difference in data distribution between subjects, whereas training accuracy-based subject selection (TSS) is utilized to select out the source subjects with higher similarity to the target subject. A comprehensive off-line analysis was conducted on the experimental data derived from 16 subjects. The results indicated that compared to the subject-specific algorithm, the proposed transfer learning algorithm effectively reduces the training time of the c-VEP BCI at the same performance level, and thereby facilitates its application in real world.

## Data availability

The data can be made available upon reasonable request by contacting the corresponding author.

Received: 10 February 2022; Accepted: 31 May 2022

Published online: 14 June 2022

## References

1. Wolpaw, J. R., Birbaumer, N., Mcfarland, D. J., Pfurtscheller, G. & Vaughan, T. M. Brain-computer interfaces for communication and control. *Clin. Neurophysiol.* **113**, 767–791 (2002).
2. Curran, E. A. & Stokes, M. J. Learning to control brain activity: a review of the production and control of EEG components for driving brain-computer interface (BCI) systems. *Brain Cognit.* **51**, 326–336 (2003).
3. Wang, H. et al. Cross-subject assistance: inter- and intra-subject maximal correlation for enhancing the performance of SSVEP-based BCIs. *IEEE Trans. Neural Syst. Rehabil. Eng.* **29**, 517–526 (2021).
4. Wei, Q. et al. A training data-driven canonical correlation analysis algorithm for designing spatial filters to enhance performance of SSVEP-based BCIs. *Int. J. Neural Syst.* **30**, 2050020 (2020).
5. Ge, S., Jiang, Y., Wang, P., Wang, H. & Zheng, W. Training-free steady-state visual evoked potential brain-computer interface based on filter bank canonical correlation analysis and spatiotemporal beamforming decoding. *IEEE Trans. Neural Syst. Rehabil. Eng.* **27**, 1714–1723 (2019).

6. Zhao, H., Wang, Y., Liu, Z., Pei, W. & Chen, H. Individual identification based on code-modulated visual-evoked potentials. *IEEE Trans. Inf. Forensics Secur.* **14**, 3206–3216 (2019).
7. Nakanishi, M. *et al.* Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis. *IEEE Trans. Biomed. Eng.* **65**, 104–112 (2018).
8. Chen, X., Zhao, B., Wang, Y., Xu, S. & Gao, X. Control of a 7-DOF robotic arm system with an SSVEP-based BCI. *Int. J. Neural Syst.* **28**, 1850018 (2018).
9. Li, M., Yang, G., Liu, Z., Gong, M. & Lin, F. The effect of SOA on an asynchronous ERP and VEP-based BCI. *IEEE Access* **99**, 1–9 (2021).
10. Gembler, F. W., Benda, M., Rezeika, A., Stawicki, P. & Volosyak, I. Asynchronous c-VEP communication tools—efficiency comparison of low-target, multi-target and dictionary-assisted BCI spellers. *Sci. Rep.* **10**, 17064 (2020).
11. Shirzhiyan, Z., Keihani, A., Farahi, M., Shamsi, E. & Jafari, A. H. Toward new modalities in VEP-based BCI applications using dynamical stimuli: introducing quasi-periodic and chaotic VEP-based BCI. *Front. Neurosci.* **14**, 1067 (2020).
12. Bin, G., Gao, X., Wang, Y., Hong, B. & Gao, S. VEP-based brain-computer interfaces: time, frequency, and code modulations. *IEEE Comput. Intell. Mag.* **4**, 22–26 (2009).
13. Sutter, E. E. The visual evoked-response as a communication channel. *IEEE Trans. Biomed. Eng.* **31**, 583–583 (1984).
14. Sutter, E. E. The brain response interface: communication through visually-induced electrical brain responses. *J. Microcomput. Appl.* **15**, 31–45 (1992).
15. Bin, G. *et al.* A high-speed BCI based on code modulation VEP. *J. Neural Eng.* **8**, 025015 (2011).
16. Thielen, J., Marsman, P., Farquhar, J. & Desain, P. From full calibration to zero training for a code-modulated visual evoked potentials for brain-computer interface. *J. Neural Eng.* **18**, 056007 (2021).
17. Wei, Q. *et al.* A novel c-VEP BCI paradigm for increasing the number of stimulus targets based on grouping modulation with different codes. *IEEE Trans. Neural Syst. Rehabil. Eng.* **26**, 1178–1187 (2018).
18. Wittevrongel, B., van Wolputte, E. & van Hulle, M. M. Code-modulated visual evoked potentials using fast stimulus presentation and spatiotemporal beamformer decoding. *Sci. Rep.* **7**, 15037 (2017).
19. Riechmann, H., Finke, A. & Ritter, H. Using a cVEP-based brain-computer interface to control a virtual agent. *IEEE Trans. Neural Syst. Rehabil. Eng.* **24**, 692–699 (2016).
20. Waytowich, N. R. & Krusienski, D. J. Spatial decoupling of targets and flashing stimuli for visual brain-computer interfaces. *J. Neural Eng.* **12**, 036006 (2015).
21. Spüler, M., Rosenstiel, W. & Bogdan, M. Online adaptation of a c-VEP brain-computer interface (BCI) based on error-related potentials and unsupervised learning. *PLoS ONE* **7**, e51077 (2012).
22. Yger, F., Berar, M. & Lotte, F. Riemannian approaches in brain-computer interfaces: a review. *IEEE Trans. Neural Syst. Rehabil. Eng.* **25**, 1753–1762 (2017).
23. Congedo, M., Barachant, A. & Andreev, A. A new generation of brain-computer interface based on Riemannian geometry. *arXiv preprint arXiv:1310.8115* (2013).
24. Rodrigues, P. L. C., Jutten, C. & Congedo, M. Riemannian Procrustes analysis: transfer learning for brain-computer interfaces. *IEEE Trans. Biomed. Eng.* **66**, 2390–2401 (2019).
25. Kalaganis, F. P., Laskaris, N. A., Chatzilari, E., Nikolopoulos, S. & Kompatsiaris, I. A Riemannian geometry approach to reduced and discriminative covariance estimation in brain computer interfaces. *IEEE Trans. Biomed. Eng.* **67**, 245–255 (2020).
26. Tang, X., Li, X., Li, W., Hao, B. & Dang, X. Transfer learning: rotation alignment with Riemannian mean for brain-computer interfaces and wheelchair control. *IEEE Trans. Cogn. Dev. Syst.* **99**, 3082648 (2021).
27. Kalunga, E. K., Chevallier, S., Djouani, K., Monacelli, E. & Hamam, Y. Online SSVEP-based BCI using Riemannian geometry. *Neurocomputing* **191**, 55–58 (2016).
28. Kalunga, E. K., Chevallier, S., & Barthélemy Q. Transfer learning for SSVEP-based BCI using Riemannian similarities between users. In *Proc. of 26<sup>th</sup> European signal Processing Conference (EUSIPCO 2018)*, 1685–1689 (2018).
29. Li, F., Xia, Y., Wang, F., Zhang, D. & He, F. Transfer learning algorithm of P300-EEG signal based on XDAWN spatial filter and Riemannian geometry classifier. *Appl. Sci.* **10**, 1804 (2020).
30. Pan, S. J. & Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22**, 1345–1359 (2010).
31. Jayaram, V., Alamgir, M., Altun, Y., Scholkopf, B. & Grosse-Wentrup, M. Transfer learning in brain-computer interfaces. *IEEE Comput. Intell. Mag.* **11**, 20–31 (2016).
32. Azab, A. M., Mihaylova, L., Ang, K. K. & Arvaneh, M. Weighted transfer learning for improving motor imagery-based brain-computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* **27**, 1352–1359 (2019).
33. Bougrain, L., Rimbart, S., Rodrigues, P. L. C., Canron, G. & Lotte, F. Guidelines to use transfer learning for motor imagery detection: an experimental study. In *International IEEE/EMBS Conference on Neural Engineering*, 5–8 (2021).
34. Liang, Y. & Ma, Y. Calibrating EEG features in motor imagery classification tasks with a small amount of current data using multi-source fusion transfer learning. *Biomed. Signal Process.* **62**, 10210 (2020).
35. Miladinović, A. *et al.* Transfer learning improves MI BCI models classification accuracy in Parkinson's disease patients. In *European Signal Processing Conference*. *arXiv:2010.15899*. 1353–1356 (2020).
36. Bassi, P. R. A. S., Rampazzo, W. & Attux, R. Transfer learning and SpecAugment applied to SSVEP based BCI classification. *Biomed. Signal Process.* **67**, 102542 (2020).
37. Huang, Z., Zheng, W., Wu, Y. & Wang, Y. Ensemble or pool: a comprehensive study on transfer learning for c-VEP BCI during interpersonal interaction. *J. Neurosci. Methods* **343**, 108855 (2020).
38. Ledoit, O. & Wolf, M. A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.* **88**, 365–411 (2004).
39. Blankertz, B., Lemm, S., Treder, M., Haufe, S. & Müller, K.-R. Single-trial analysis and classification of ERP components—a tutorial. *Neuroimage* **56**, 814–825 (2011).
40. Schäfer, J. & Strimmer, K. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat. Appl. Genet. Mol. Biol.* **4**, 1–30 (2005).
41. Lee, J. *Introduction to Smooth Manifolds* 218 (Springer, 2012).
42. Barachant, A., Bonnet, S., Congedo, M. & Jutten, C. Riemannian geometry applied to BCI classification. In *Lecture Notes in Computer Science*. 629–636 (2010).
43. Chevallier, S., Kalunga, E. K., Barthélemy, Q. & Monacelli, E. Review of Riemannian distances and divergences, applied to SSVEP-based BCI. *Neuroinformatics* **19**, 93–106 (2020).
44. Arsigny, V., Fillard, P., Pennec, X. & Ayache, N. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **29**, 328–347 (2011).
45. Moakher, M. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J. Matrix Anal. Appl.* **26**, 735–747 (2005).
46. Jeuris, B., Vandebril, R. & Vandereycken, B. A survey and comparison of contemporary algorithms for computing the matrix geometric mean. *Electron. Trans. Numer. Anal.* **39**, 379–402 (2012).
47. Zanini, P., Congedo, M., Jutten, C., Said, S. & Berthoumieu, Y. Transfer learning: a Riemannian geometry framework with applications to brain-computer interfaces. *IEEE Trans. Biomed. Eng.* **65**, 1107–1116 (2018).

48. He, H. & Wu, D. Transfer learning for brain-computer interfaces: a Euclidean space data alignment approach. *IEEE Trans. Biomed. Eng.* **67**, 399–410 (2020).
49. Xu, L., Xu, M., Ke, Y., An, X. & Ming, D. Cross-dataset variability problem in EEG decoding with deep learning. *Front. Hum. Neurosci.* **14**, 00103 (2020).
50. Lotte, F. & Guan, C. Learning from other subjects helps reducing brain-computer interface calibration time. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2010)*, Dallas, USA. 614–617 (2010).
51. Qi, H., Xue, Y., Xu, L., Cao, Y. & Jiao, X. A speedy calibration method using Riemannian geometry measurement and other-subject samples on A P300 speller. *IEEE Trans. Neural Syst. Rehabil. Eng.* **26**, 602–608 (2018).
52. Giles, J., Ang, K. K., Mihaylova, L. S. & Arvaneh, M. A subject-to-subject transfer learning framework based on Jensen-Shannon divergence for improving brain-computer interface. In *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2019)*, Brighton, UK. 3087–3091 (2019).
53. Yuan, P. *et al.* A study of the existing problems of estimating the information transfer rate in online brain-computer interfaces. *J. Neural Eng.* **10**, 026013 (2013).
54. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
55. Wei, Q., Feng, S. & Lu, Z. Stimulus specificity of brain-computer interfaces based on code modulation visual evoked potentials. *PLoS One* **11**, 0156416 (2016).
56. Liu, Y., Wei, Q. & Lu, Z. A multi-target brain-computer interface based on code modulated visual evoked potentials. *PLoS ONE* **13**, 0202478 (2018).
57. Jin, J. *et al.* Correlation-based channel selection and regularized feature optimization for MI-based BCI. *Neural Netw.* **118**, 262–270 (2019).
58. Jin, J. *et al.* Bispectrum-based channel selection for motor imagery based brain-computer interfacing. *IEEE Trans. Neural Syst. Rehabil. Eng.* **28**, 2153–2163 (2020).
59. Wei, Q., Lu, Z., Chen, K. & Ma, Y. Channel selection for optimizing feature extraction in an electrocorticogram-based brain-computer interface. *J. Clin. Neurophysiol.* **27**, 321–327 (2010).

## Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant 62066028 and Grant 61663025.

## Author contributions

J.Y. performed the experiment, analyzed the data and wrote the manuscript. Q.W. conceived the study, designed the experiment and revised the manuscript. X.Z. performed the experiment. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to Q.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022