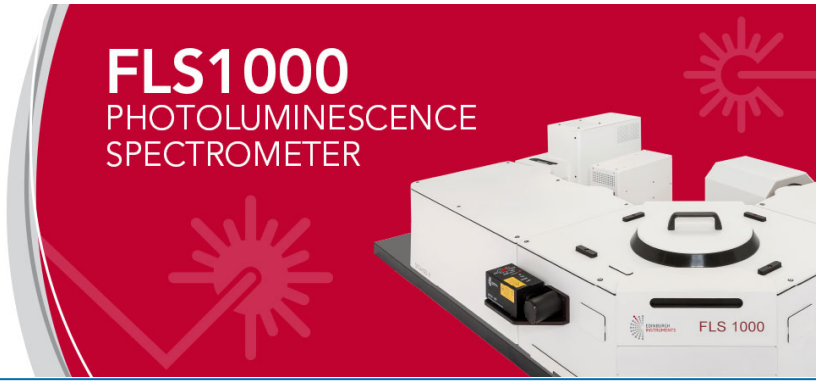
PAPER

Boosting template-based SSVEP decoding by cross-domain transfer learning



Journal of Neural Engineering

PAPER

Boosting template-based SSVEP decoding by cross-domain transfer learning

Kuan-Jung Chiang[1,2], **Chun-Shu Wei**[3], **Masaki Nakanishi**[2] **and Tzyy-Ping Jung**[2]

[1] Department of Computer Science and Engineering, University of California - San Diego, La Jolla, California 92122, United States of America
[2] Swartz Center for Computational Neuroscience, Institude of Neural Computation, University of California - San Diego, La Jolla, California 92122, United States of America
[3] Department of Computer Science, National Chiao Tung University, Hsinchu 300, Taiwan

E-mail: kuchiang@eng.ucsd.edu

RECEIVED
18 June 2020

REVISED
13 October 2020

ACCEPTED FOR PUBLICATION
16 November 2020

PUBLISHED
11 February 2021

Keywords: brain–computer interface, SSVEP, EEG, transfer learning

Abstract

Objective. This study aims to establish a generalized transfer-learning framework for boosting the performance of steady-state visual evoked potential (SSVEP)-based brain–computer interfaces (BCIs) by leveraging cross-domain data transferring. *Approach*. We enhanced the state-of-the-art template-based SSVEP decoding through incorporating a least-squares transformation (LST)-based transfer learning to leverage calibration data across multiple domains (sessions, subjects, and electroencephalogram montages). *Main results*. Study results verified the efficacy of LST in obviating the variability of SSVEPs when transferring existing data across domains. Furthermore, the LST-based method achieved significantly higher SSVEP-decoding accuracy than the standard task-related component analysis (TRCA)-based method and the non-LST naive transfer-learning method. *Significance*. This study demonstrated the capability of the LST-based transfer learning to leverage existing data across subjects and/or devices with an in-depth investigation of its rationale and behavior in various circumstances. The proposed framework significantly improved the SSVEP decoding accuracy over the standard TRCA approach when calibration data are limited. Its performance in calibration reduction could facilitate plug-and-play SSVEP-based BCIs and further practical applications.

1. Introduction

A brain-computer interface (BCI) is an immediate pathway that provides an intuitive interface for users, especially disabled ones, to translate their intention into commands to control external devices [1]. Among various neuroimaging modalities, electroencephalogram (EEG) has been widely used for developing BCI applications for the sake of unobtrusiveness, low cost, and high temporal resolution [1]. Researchers have attempted to develop a variety of real-world applications of an EEG-based BCI such as spellers that allow disabled users to communicate with others. Farewell and Donchin demonstrated a P300-based BCI as the first-ever BCI speller in the 1980s [2]. Since then, BCI spellers based on visual-evoked brain responses have been improved by different advancements in the decoding techniques. Recently, steady-state visual evoked potentials (SSVEPs), a type of neural responses to repetitive visual stimulation, have attracted increasing attention and been used in high-speed BCI spellers as the acquisition of SSVEP is relatively stable [3–5]. The performance of SSVEP-based BCIs has been significantly improved by advances in system design, signal processing, and decoding algorithms in the past decade [6].

As SSVEPs are oscillatory fluctuations that respond to flickering visual stimuli, an SSVEP-based BCI measures the user's SSVEP and uses an algorithm to identify the corresponding stimulus based on the SSVEP data. Analyzing SSVEPs in the frequency-domain has been an intuitive approach to detect the frequency peaks in the spectral response of SSVEP that correspond to the flickering frequency of stimuli. For example, power spectral density analysis [3] that uses features in frequency-domain was proposed to decode SSVEPs. Canonical correlation

analysis (CCA) [7, 8] that compares a test trial with computer-generated SSVEP models/templates consisting of sine-cosine signals for each target [9] was also proved to work on SSVEPs. However, these methods failed to maintain consistent performance across users because the individual differences in the SSVEP data were not taken into account [10]. Later studies exploited individualized training data, or SSVEP templates, and developed training-based SSVEP decoding schemes that better characterize individual SSVEP patterns using personalized calibration data. The training-based methods usually outperform the aforementioned calibration-free methods [11–16].

The success of training-based SSVEP decoding has significantly boosted the efficiency of SSVEP-based BCI spellers. Nonetheless, the calibration procedure is often laborious and time-consuming, hindering practical and wide-spread applications of BCIs in our daily life. Several studies have attempted to adopt transfer-learning techniques to reduce the individual-variability or session-variability so that a model for SSVEP can be tuned with existing data without repeat collections of calibration data before each use [17]. For instance, Yuan *et al* and Wong *et al* proposed subject-to-subject transfer-learning methods that transfer SSVEP data from existing subjects to new ones using a spatial filtering approach [18, 19]. Rodrigues *et al* also proposed a cross-subject transfer learning method using a Riemannian geometrical transformation [20]. Waytowich *et al* applied a convolutional neural network to train a subject-independent classifier for detecting SSVEPs [21]. In another study, Nakanishi *et al* proposed a session-to-session and device-to-device transfer-learning method using spatial filtering [22, 23]. Suefusa *et al* adopted a frequency shifting technique to synthesize calibration data at arbitrary frequencies from existing calibration data [24].

The transfer learning approaches have succeeded either in improving classification accuracy of SSVEPs compared with calibration-free methods without any calibration process or compared with fully-calibrated methods with reduced calibration cost. However, such methods focused only on transferring data across one domain such as cross-subject or cross-session transferring. For example, the aforementioned subject-to-subject transfer learning methods implicitly assumed that their subjects, stimulus parameters, and EEG devices are the identical between calibration and target data [18–21]. Similarly, the cross-session and the cross-device transfer learning methods proposed in our previous studies implicitly required calibration data to be from the same subjects and stimulus parameters with target data [22, 23]. Table 1 compares and summarizes current transfer-learning schemes for SSVEP-based BCIs. It is worth noting that a generalized transfer learning approach that can handle transferring existing data across multiple domains has yet to be proposed.

The goal of this study is to propose a generalized framework of transfer learning that can leverage SSVEP data across multiple domains including subjects, sessions, and devices toward a practical BCI application.

Our preliminary study demonstrated that applying LST to the existing SSVEP datasets acquired from other users can augment the size of calibration data for new users and in turn enhance decoding performance without acquiring extra calibration process [25]. However, although the LST-based method could technically be employed to transfer SSVEPs across any domain such as cross-session and cross-device transferring, the effectiveness of such scenarios has not been investigated in our previous work [25].

Therefore, the generalizability of the LST-based method, especially in a real-world scenario where EEG data from different users are more likely recorded by different types of EEG recording systems, still remains unknown. The discrepancy among recording montages has posed a grand challenge in consolidating EEG data across domains and hindered the translation of BCI technologies toward practical applications. This study investigated the feasibility and the effectiveness of the LST-based method in transferring SSVEPs across multiple domains using the dataset acquired from multiple subjects and sessions with multiple EEG devices collected in our previous study [23]. This work further examined and compared the characteristics of data from different subjects with and without the LST-based transferring in feature spaces. In addition, this study also investigated the effects of parameters such as the number of data transferred via the LST on the classification accuracy.

2. Methods

2.1. EEG data

This study used the EEG data recorded and reported in a previous study [23]. The dataset consisted of the EEG recordings from ten healthy adults.

During the experiment, 40 visual stimuli were presented on a 27- inch liquid-crystal display. Each stimulus was modulated by a sampled-sinusoidal stimulation method with joint frequency-phase modulation [6, 12]. The stimulation frequencies ranged from 8 to 15.8 Hz with an interval of 0.2 Hz. The initial phase values started from 0 rad and the phase interval was 0.35 π rad.

In the experiment, the subjects performed two sessions of simulated online experiments [26]. The procedure of tasks in both sessions were identical except that the EEG signals were recorded with two different devices. The two devices were an ActiveTwo system (BioSemi, Inc.) as a high-density laboratory-oriented system and a Quick-30 (Q30)

Table 1. Comparison between different transfer learning methods for SSVEP-based BCIs.

Study	Year	Calibration data [1]	Transferring approach	Domain transferred			
				Subjects	Sessions	Devices	Stimuli
[18]	2015	Not required	Spatial filtering	**Yes**	No	No	No
[22]	2016	Not required	Spatial filtering	No	**Yes**	No	No
[24]	2017	Not required	Frequency shifting	No	No	No	**Yes**
[21]	2018	Not required	Convolutional neural network	**Yes**	No	No	No
[20]	2019	Not required	Geometrical transformation	**Yes**	No	No	No
[23]	2019	Not required	Spatial filtering	No	No	**Yes**	No
[25]	2019	Required	Channel-wise projection	**Yes**	No	No	No
[19]	2020	Required	Spatial filtering	**Yes**	No	No	No

[1] Additional calibration data from a target session

system (Cognionics, Inc.) as a mobile and wireless system for real-life applications. Figure 1 lists the characteristics of the two devices. The Q30 system was always tested in the first session, and the ActiveTwo system was tested in the second one to avoid the skin preparation required for the wet (gel) electrodes.

In each session, the subjects wore either one of the EEG devices and performed eight blocks of simulations. In each block, the subjects were asked to gaze at one visual stimulus indicated by the stimulus program for 1.5 s at a time until all 40 stimuli were gazed once. Therefore, the subjects performed 40 trials corresponding to 40 stimuli in a block, and the data consisted of eight trials per stimulus from each subject.

2.2. Preprocessing

Six channels (PO3, PO4, PO7, PO8, O1, O2) and eight channels (POz, PO3, PO4, PO7, PO8, Oz, O1, O2) of EEG signals were extracted from the recordings collected by the Q30 and the ActiveTwo systems, respectively. The signals extracted from both devices were resampled at 256 Hz (from 500 and 2048 Hz for the Q30 and the ActiveTwo system respectively) and then re-referenced to the Fz electrode. We employed a conventional sampling-rate conversion method, which includes upsampling, anti-aliasing filtering, and downsampling, to resample the data at a rational ratio between the original and the new sampling rates [27]. This process can be done by using the *resample* function in MATLAB. The re-referencing was done by simply subtracting the signals at target channels by that at the reference channel Fz.

The data were then extracted in $[L$ s, $L + 1.5$ s$]$, where time zero indicates the stimulus onset and L indicates latency delay in the experimental environment and the human's visual system. The latency L was set to 0.17 and 0.15 for the Q30 and the ActiveTwo systems, respectively, according to the previous study [23]. After epoching, the 60-Hz line noise was suppressed by applying an infinite impulse response notch filter to each epoch.

2.3. TRCA-based SSVEP detection

TRCA is a data-driven method aiming to find a spatial filter that maximizes the reproducibility in each trial during task periods [12, 28]. The task-related components extracted by the spatial filter obtained by TRCA have been shown to provide better SNR which can significantly improve the performance of training-based SSVEP detections [12]. In addition, the TRCA-based method could successfully be combined with the filter bank analysis, which decomposes EEG signals into multiple sub-band components so that independent information embedded in the harmonic components can be efficiently extracted [29].

In the procedure of the TRCA-based method with filter bank analysis, individual calibration data for the nth stimulus are denoted as $\mathbf{x}_n \in \mathbb{R}^{N_C \times N_S \times N_T}$, $n = 1, 2, ..., N_F$. Here N_C is the number of channels, N_S is the number of sampling points in each trial, N_T is the number of trials for each stimulus, and N_F is the number of visual stimuli (i.e. 40 in this study). In the training phase, the calibration data are divided into N_K sub-bands by a filter bank and become $\mathbf{x}_n^k \in \mathbb{R}^{N_C \times N_S \times N_T}$, $k = 1, 2, ..., N_K$. The N_K was set to five in this study. In the following parts of this paper, the ith trial of stimulus n in sub-band k will be denoted as $\mathbf{x}_{n,i}^k$. Spatial filters in each sub-band are obtained to maximize the sum of the inter-trial covariance after projecting the multi-channel signals into single-channel ones with the spatial filter. Therefore, the goal is finding the channel weights $\mathbf{w}_n^k \in \mathbb{R}^{N_C}$ to maximize the term

$$\begin{aligned} V_n^k &= \sum_{\substack{i,j \\ i \neq j}}^{N_T} \mathrm{Cov}\left((\mathbf{w}_n^k)^T \mathbf{x}_{n,i}^k, (\mathbf{w}_n^k)^T \mathbf{x}_{n,j}^k\right) \\ &= (\mathbf{w}_n^k)^T \left(\sum_{\substack{i,j \\ i \neq j}}^{N_T} \mathrm{Cov}\left(\mathbf{x}_{n,i}^k, \mathbf{x}_{n,j}^k\right) \right) \mathbf{w}_n^k \\ &= (\mathbf{w}_n^k)^T \mathbf{S}_n^k \mathbf{w}_n^k. \end{aligned} \quad (1)$$

Here, $\mathbf{S}_n^k$ is the sum of cross-covariance matrices between all pairs of trials of stimulus n in sub-band k.

Cognionics, Inc.
Quick-30 system
- Active dry Ag/Agcl electrodes
- 30 channels
- Reference: Left earlobe

BioSemi, Inc.
ActiveTwo system
- Active wet Ag/Agcl electrodes
- 64 channels
- Reference: CMS/DRL

Figure 1. Specifications of two EEG devices used in this study.

To avoid arbitrary scaling with the weights, instead of finding the $\mathbf{w}_n^k$ that maximizes $\mathbf{V}_n^k$, a constraint term is needed:

$$\begin{aligned} C_n^k &= \sum_i^{N_T} \mathrm{Var}\left((\mathbf{w}_n^k)^T \mathbf{x}_{n,i}^k\right) \\ &= (\mathbf{w}_n^k)^T \left(\sum_i^{N_T} \mathrm{Cov}\left(\mathbf{x}_{n,i}^k\right)\right) \mathbf{w}_n^k \\ &= (\mathbf{w}_n^k)^T \mathbf{Q}_n^k \mathbf{w}_n^k \\ &= 1. \end{aligned} \tag{2}$$

Finally, the weights can be calculated as:

$$\begin{aligned} \mathbf{w}_n^k &= \mathrm{argmax}_{\mathbf{w}} \frac{V_n^k}{C_n^k} \\ &= \mathrm{argmax}_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{S}_n^k \mathbf{w}}{\mathbf{w}^T \mathbf{Q}_n^k \mathbf{w}}. \end{aligned} \tag{3}$$

The solution of equation (3) is equal to the eigenvector of the matrix $\mathbf{Q}^{-1}\mathbf{S}$ with the largest eigvenvalue.

In the ensemble TRCA as an extension version of TRCA, the final spatial filters for each sub-band $\mathbf{w}^k \in \mathbb{R}^{N_C \times N_F}$ are obtained by concatenating all the weights in each stimulus:

$$\mathbf{w}^k = \left[\mathbf{w}_1^k, \mathbf{w}_2^k, \ldots, \mathbf{w}_{N_F}^k\right]. \tag{4}$$

After obtaining the spatial filters, individual templates $\bar{\mathbf{x}}_n^k \in \mathbb{R}^{N_C \times N_S}$ are prepared. The training trials for nth stimulus in sub-band k are first averaged across trials as:

$$\bar{\mathbf{x}}_n^k = \frac{1}{N_T} \sum_i^{N_T} \mathbf{x}_{n,i}^k. \tag{5}$$

In the testing phase, single-trial testing data $\hat{\mathbf{x}} \in \mathbb{R}^{N_C \times N_S}$ are first pre-processed by the filter banks to be decomposed into N_K sub-bands as well. Then, the spatial filters $\mathbf{w}^k$ obtained in training phase are applied to the testing signals $\hat{\mathbf{x}}^k \in \mathbb{R}^{N_C \times N_S}$ in each sub-band. Feature values ρ_n^k are calculated as correlation coefficients between the testing signals and the individual templates as

$$\rho_n^k = r\left((\mathbf{w}^k)^T \hat{\mathbf{x}}^k, (\mathbf{w}^k)^T \bar{\mathbf{x}}_n^k\right), \tag{6}$$

where $r(\mathbf{a}, \mathbf{b})$ indicates the Pearson's correlation analysis between two variables $\mathbf{a}$ and $\mathbf{b}$. A weighted sum of the ensemble correlation coefficients corresponding to all the sub-bands was calculated as the final feature for target identification as:

$$\rho_n = \sum_{k=1}^{N_K} \alpha(k) \cdot \rho_n^k, \tag{7}$$

where $\alpha(k)$ was defined as $\alpha(k) = k^{-1.25} + 0.25$ according to [29]. Finally, the target stimulus τ can be identified as

$$\tau = \mathrm{argmax}_n \rho_n. \tag{8}$$

2.4. LST-based cross-domain transferring

This work proposes a transformation of SSVEP signals from one domain and another. Let $\mathbf{x} \in \mathbb{R}^{N_C \times N_S}$ and $\acute{\mathbf{x}} \in \mathbb{R}^{N_C' \times N_S}$ be the single-trial SSVEP data obtained in a domain and in another domain (i.e. subject, session, and/or device), respectively. Then, we aim to find a transformation matrix $\mathbf{P} \in \mathbb{R}^{N_C \times N_C'}$ such that $\mathbf{x}(t) = \mathbf{P}\acute{\mathbf{x}}(t) + \epsilon$, where $\mathbf{x}(t), \acute{\mathbf{x}}(t)$ represent the t-th column of $\mathbf{x}, \acute{\mathbf{x}}$, and $\epsilon \in \mathbb{R}^{N_C}$ is an error vector. Note that the numbers of channels N_C and N_C' are not necessary to be equal between two domains. The transformation matrix $\mathbf{P}$ can be obtained by minimizing the error term ϵ in the aforementioned equation with a multivariate least-squares regression given $\mathbf{x}$ and $\acute{\mathbf{x}}$ as follows:

$$\mathbf{P} = \mathrm{argmin}_{\mathbf{p}} \mathrm{Tr}\left[(\mathbf{x} - \mathbf{p}\acute{\mathbf{x}})(\mathbf{x} - \mathbf{p}\acute{\mathbf{x}})^T\right]. \tag{9}$$

This problem can be solved as follows:

$$\mathbf{P} = \mathbf{x}\acute{\mathbf{x}}^T(\acute{\mathbf{x}}\acute{\mathbf{x}}^T)^{-1}. \quad (10)$$

Because several studies have shown that trial-averaging can significantly improve the SNR of SSVEPs compared with single-trial SSVEPs by removing background EEG activities [11, 12], we used it to improve the transferability of SSVEPs across different domains. In the proposed method, therefore, instead of using the single-trial signals of the new domain $\mathbf{x}$, we use the averaged signals $\bar{\mathbf{x}}$ across several trials of the signals obtained from the new domain as the target of transformation from existing training pool. These calibration trials obtained from a new user are called transformation targets. Every trial of the existing domains in the training pool will be transformed to signals $\underline{\acute{\mathbf{x}}}$ which should be comparable to the transformation target $\bar{\mathbf{x}}$, i.e. $\bar{\mathbf{x}} \approx \underline{\acute{\mathbf{x}}}_i = \mathbf{P}\acute{\mathbf{x}}_i$ (i is a trial index). Finally, all trials from the new domain $\mathbf{x}$, which are used to construct the transformation targets, and in the existing data pool $\underline{\acute{\mathbf{x}}}$ are concatenated to form a larger training set than the one used in the conventional template-based algorithm (i.e. $\mathbf{x}$). With the new training set, the aforementioned TRCA-based method is performed to identify target stimuli. The procedure of the LST is also illustrated in figure 2.

2.5. Performance evaluation

To validate the efficacy of the proposed LST-based method in transferring SSVEPs, we compared the performance of detecting SSVEPs using the following three schemes (figure 3):

(a) BASELINE: A self-decoding approach in which all the calibration data are collected from a new user (i.e. the conventional individual-template-based method).

(b) Transfer without LST (w/oLST): A cross-domain transferring approach in which the calibration data consist of calibration trials from the new domain and from other domains without any transformation.

(c) Transfer with LST (w/LST): A cross-domain transferring approach in which the calibration data consist of calibration trials from the new domain and from other domains transformed using LST. The transformation targets are obtained from the data obtained in the new domain.

This study ran a series of simulations to test the performance of the proposed LST-based method as a cross-domain transfer learning for an SSVEP-based BCI. A leave-one-subject-out cross-validation, in which a subject is treated as a new (i.e. target) user and all the other subjects are treated as existing (i.e. non-target) users, was employed to investigate the effectiveness of the proposed method under the cross-subject scenario. When one session of the new user is being tested, the eight trials for each stimulus was randomly divided into five and three as a calibration set and a testing set. We then trained three models including the TRCA-based spatial filtering and individualized template using different pools of training trials for each scheme. In the BASELINE, 2–5 calibration trials for each of 40 stimuli from a target subject are used to form training sets. In the w/oLST, all the eight trials for each stimulus from all nine non-target subjects (72 trials in total for each stimulus) are simply merged with the training sets used in the BASELINE. In the w/LST, the data from the non-target subjects are first transformed via the LST and then merged with the training sets used in the BASELINE. Then, three models were evaluated with SSVEP-decoding performance on the 120-trial (i.e. three trials $\times$ 40 stimuli) testing set. The w/LST scheme was further evaluated under the cross-device scenario. In the scenario, the leave-one-subject-out cross validation was also employed. In addition, different EEG devices were selected between target and non-target subjects. Note that the w/oLST cannot be applied to the cross-device scenario because the numbers of electrodes of the two EEG systems are different across devices. The random separation of the template/test set was repeated ten times. The decoding performance of each target subject was estimated by the average of ten repeats.

Lastly, the classification accuracy was statistically tested by factorial nonparametric permutation-based repeated measures analysis of variances (ANOVAs) [30]. The number of permutation was set to 5000. In the post-hoc analyses, different schemes were compared pair-wisely using Wilcoxon signed-rank tests.

3. Results

Figure 4 shows, for the three schemes, the averaged SSVEP-decoding accuracy across subjects with different numbers (from two to five) of calibration trials per stimulus under the cross-subject and cross-device scenarios. In general, the w/LST-based scheme outperformed the other two schemes regardless of the number of calibration trials.

In the cross-subject scenario, a three (schemes) $\times$ four (the number of calibration trials) two-way nonparametric permutation-based repeated measures ANOVA showed significant main effects of both schemes (Q30: $p < 0.001$; ActiveTwo: $p = 0.006$) and the number of calibration trials (Q30: $p < 0.001$; ActiveTwo: $p < 0.001$). The two-way ANOVA also showed a significant interaction between schemes and the number of calibration trials ($p < 0.001$). In the cross-device scenario, when transferring data from the ActvieTwo to the Q30 systems, a two (schemes) $\times$ four (the numbers of calibration trials per stimulus) two-way ANOVA showed signification main effects of both schemes ($p < 0.001$) and the number of calibration trials ($p < 0.001$), and a significant interaction between them ($p < 0.001$). On the other hand, when transferring from the Q30 to the ActiveTwo system,

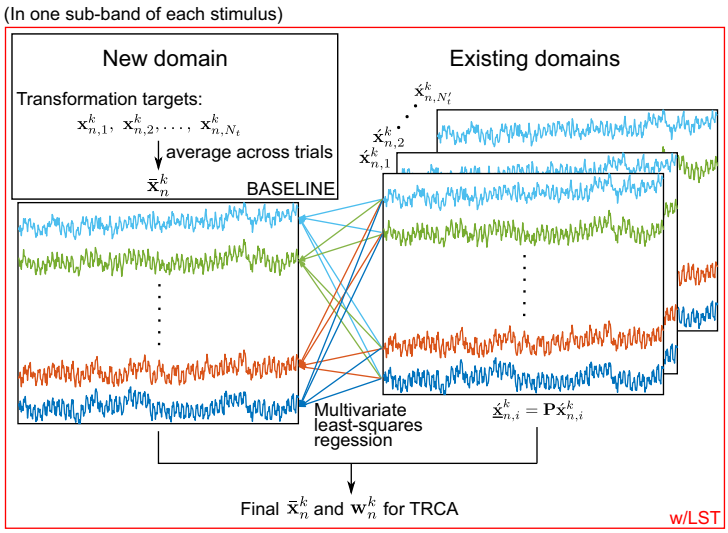


Figure 2. The procedure of transferring SSVEPs based on the least square error transformation. $\mathbf{w}^k_n$ refers to the TRCA-based spatial filer for nth stimulus in kth sub-band (see equation (3)).

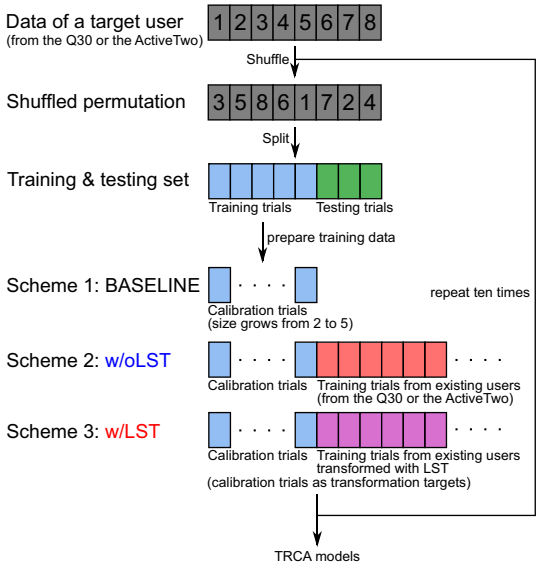


Figure 3. The flowchart of the preparation of the calibration data for three schemes.

the two-way ANOVA showed significant main effects of the number of calibration trials ($p < 0.001$), but no significant main effect of schemes ($p = 0.137$) and interaction between them ($p = 0.149$).

The post-hoc Wilcoxon signed-rank tests showed that the w/LST scheme consistently and significantly outperformed the others regardless of the number of calibration trials in the cross-subject scenario. In the cross-device scenario, when transforming the signals from the ActiveTwo system to the Q30 system, the signed-rank tests also showed that the w/LST scheme had significantly higher accuracy than the BASELINE regardless of the number of calibration trials. However, there was no statistically significant difference between w/LST and the BASELINE when transforming signals from the Q30 system to the ActiveTwo system.

Figures 5(a) and (b) show the decoding performance for each subject under the cross-subject and cross-device scenarios, respectively, with different numbers of calibration trials. The performance is represented as a logarithmic error rate in the 40-class classification. In figures 5(a) and (b), most of the data points fall within the lower right region (i.e. below the diagonal dashed line), which indicates the w/LST outperformed both the method w/oLST and the BASELINE schemes under most of the circumstances. In particular, when the testing data are from Q30, the w/LST consistently has lower error rates than the BASELINE among nearly all subjects. As for the circumstances when the testing data are from the ActiveTwo system, the w/LST can still outperform the BASELINE among most of the subjects when the size of transformation targets is small. When compared to the method w/oLST, the w/LST has higher accuracy under nearly all circumstances.

4. Discussions

This study demonstrated the efficacy of the LST-based transfer-learning method in mitigating the variability of SSVEP data across multiple domains. Figures 4 and 5 suggest that the LST-based method (i.e. w/LST) is capable of boosting the performance of the template-based SSVEP decoding with TRCA especially when the amount of calibration data from the target subject (new user) is insufficient. In addition, study results indicate the negative effects of using the naive transfer learning (w/oLST), compared to the standard TRCA (BASELINE) scheme. These results can be observed from the aspect of the signal characteristics. Figure 7 shows the scatter plots of sample EEG data recorded with the ActiveTwo system in two schemes (w/ and

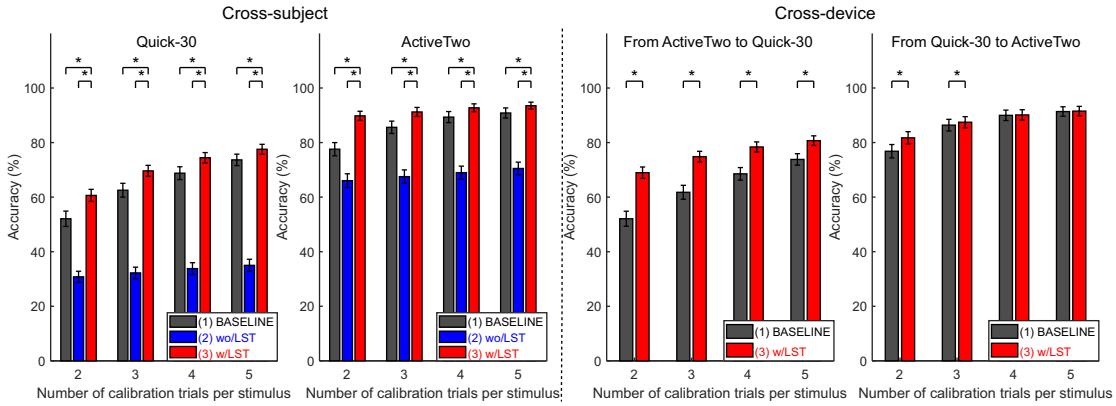


Figure 4. The averaged classification accuracy of different schemes across ten target subjects and ten cross-validation iterations at different numbers of calibration trials per stimulus. '*' indicates $p < 0.05$ of the Wilcoxon signed-rank test between two schemes.

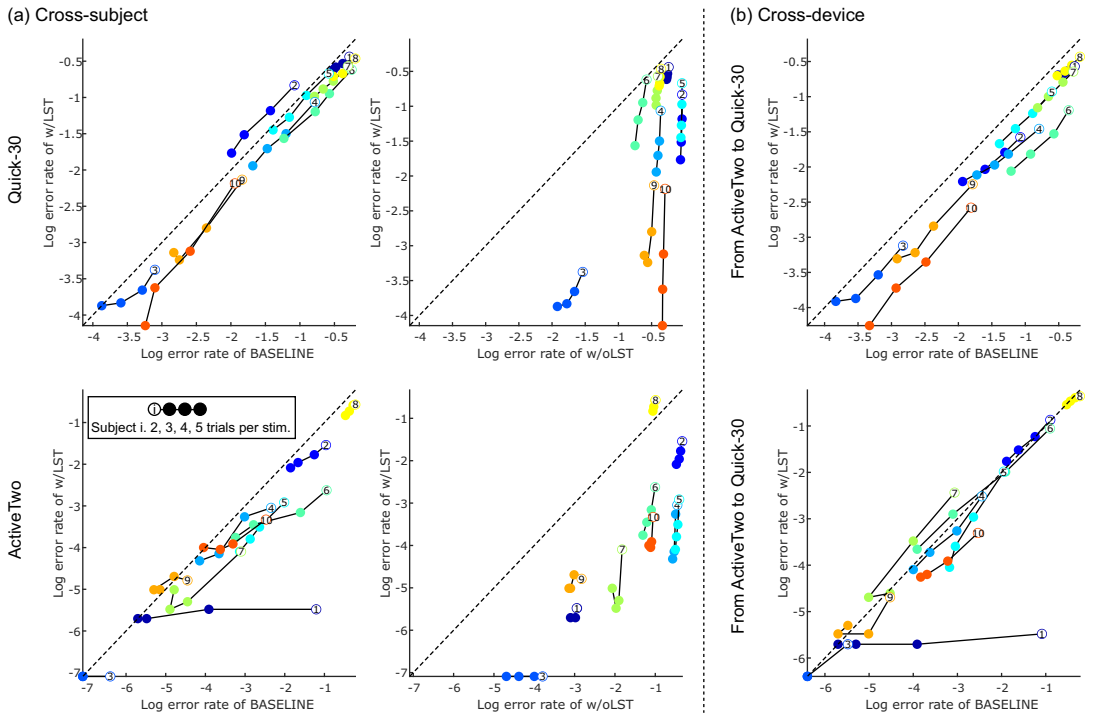


Figure 5. The comparison of the log error rate of different schemes when each of ten subjects plays as the target subject under: (a) cross-subject scenarios, and (b) cross-device scenarios. The number in the hollow dots indicates the subject id serving as the target subject. Four dots starting from the hollow dots are cases of number of calibration trials per stimulus starting from two to five.

w/oLST). The subject 1 served as the target subject in both cases, and the sizes of transformation targets were two and five trials per stimulus for the upper and lower plot respectively. The original EEG data and the EEG data transformed by LST onto the transformation targets were pooled together. All EEG trials were first averaged across channels and then processed with t-SNE [31] to reduce the dimension to 2D. The plots suggested that the LST can enhance the similarity between trials in the same stimulus across all subjects and reduce similarities across different stimuli. For better visualization, figure 7 only plots the trials of the first two of 40 stimuli (correspond to two colors, please see the figure caption).

The improvement in the similarity was also reflected in the EEG spectra. Figure 8 shows the mean spectra of the means across all EEG signals in response to the 12 Hz stimulus in three schemes when subject 1 was the target subject. First, the peak of the spectrum of the BASELINE scheme when the number of calibration trials per stimulus was two (the top panel) did not appear in the target frequency due to lack of training trials, while the peak became centered at 12 Hz when the number was increased to five (the bottom panel). Note that this phenomenon was reflected in the classification accuracy (figure 5(b)). Furthermore, the fact that the increment in training trials resulted in a more steady spectrum demonstrated the benefit

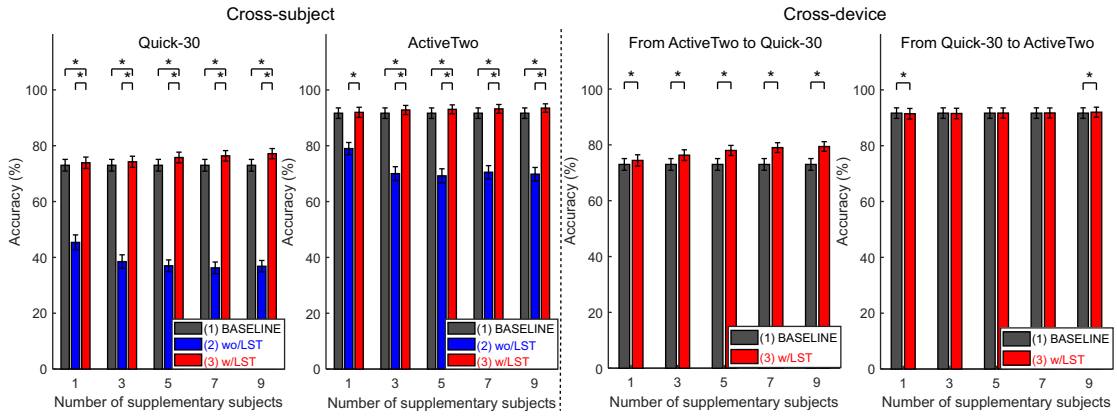


Figure 6. The averaged classification accuracy of different schemes across ten target subjects and ten iterations using different numbers of supplementary subjects. The number of trials per stimulus was fixed to five. '*' indicates $p < 0.05$ of the Wilcoxon signed-rank test between two schemes.

provided by the w/LST scheme. Because the w/LST scheme makes pooling a large number of training trials possible, the SNR can be significantly increased. On the other hand, it can be seen that because the w/oLST scheme simply pooled many trials with high variability, the peak at the target frequency was less prominent. This implies that only with proper transformation on the trials of an existing domain, pooling these trials could lead to positive transfer and improve the SNR. Finally, figure 9 shows the averaged Pearson's correlation coefficients of time-domain signals across channels between training and testing trials in all cases. The similarity in the frequency domain matched the magnitude of the correlation coefficients in the time domain. The trend in figure 9 also matches the one in the classification accuracy (figure 4).

The classification results shown in figures 4 and 5 suggest the efficacy of the proposed LST-based method, which significantly enhanced SSVEP-decoding performance, particularly when the performance of the original model (BASELINE) was limited. While the leading-edge SSVEP-decoding method, template-based method with TRCA-based spatial filtering [12], struggles with time-consuming calibration sessions, the LST-based method can leverage existing data from other domains (subjects and recording devices) and improve decoding performance. As shown in the figures, when the number of template trials was limited in all four scenarios, the w/LST scheme offered high accuracy without requiring many templates.

Comparable performance was found using the conventional TRCA approach (BASELINE) and the w/LST scheme in some cases when testing trials were from the ActiveTwo system (figure 5(a), the lower-left panel). In the cross-subject scenarios with the ActiveTwo system, as the number of calibration trials per stimulus was large enough (four and five), the performance of the BASELINE model nearly reached the ceiling, and therefore, the LST-based method could not improve the performance, suggesting that leveraging a large amount of data from other subjects has no observable benefit when newly collected individual calibration trials are sufficient. This is in line with the rationale of training-based SSVEP methods, which emphasizes the importance of individualized calibration for SSVEP decoding.

In comparison with the naïve transferring (w/oLST), for most of the subjects, the performance of the LST-based method improved along with acquiring additional calibration trials from the target user while the w/oLST scheme did not. In addition, another big advantage of the LST is that the numbers and locations of the EEG channels of new calibration data and supplementary data can be different. In the cross-device scenario, while the naïve data pooling is not even allowed, the LST could help expand the number of training trials.

In the cross-device scenario, when the EEG signals of the target subjects were from the ActiveTwo system and the ones of the existing subjects were from the Q30 system, the increment in accuracy that the LST can bring was less than transferring within the ActiveTwo system (figure 4, the second and the fourth panel from the left). This implies the limitation of the LST-based method that it still relies on the supplementary data from existing domains with good quality. Nonetheless, the LST did not bring any negative impact either. In a more practical scenario, in which the dry-EEG-based Q30 was used as the recording device for a new target subject, the LST can leverage the existing data collected from other subjects using a standard wet-electrode-based EEG system in a well-controlled laboratory to improve the SSVEP-decoding accuracy (figure 4, the third panel from the left). In other words, if there is sufficient data collected by gel-based EEG systems in well-controlled laboratories or even from the publicly available datasets on the Internet, the LST can leverage these data and a small number of calibration data collected from

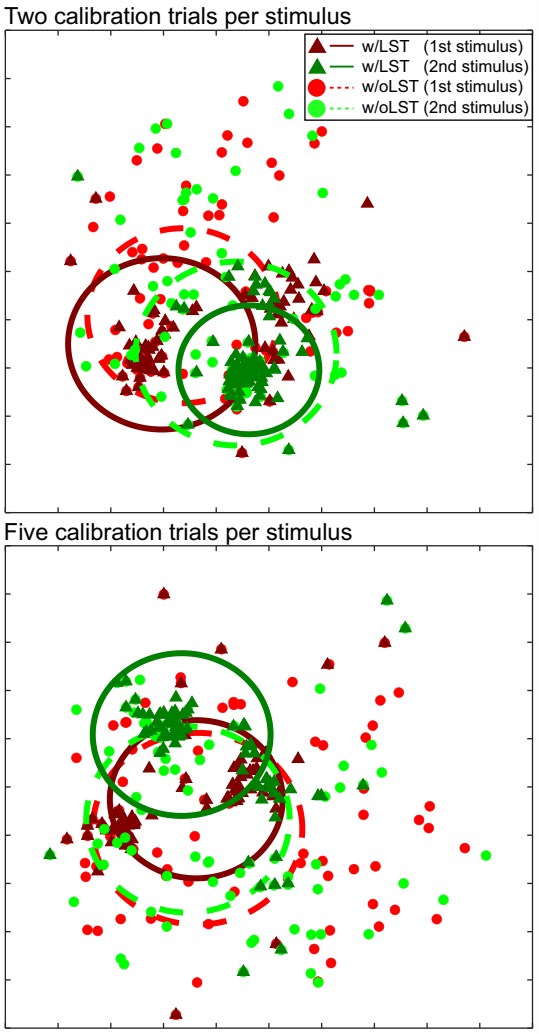


Figure 7. The scatter plot of EEG trials after dimension reduction with t-SNE. For the easiness of visualization, only trials of the first two stimuli are plotted. The triangular dots with darker colors are trials after the LST (subject 1 as the target subject), and the circular dots with lighter colors are original trials. The circles with solid or dashed lines indicate the medians and the standard deviation of trials w/ or w/oLST with corresponding colors. The Silhouette scores of w/oLST and w/LST are 0.0287 and 0.2262 when the number of calibration trials per stimulus is two, and are −0.0071 and 0.1386 when the number of calibration trials per stimulus is five.

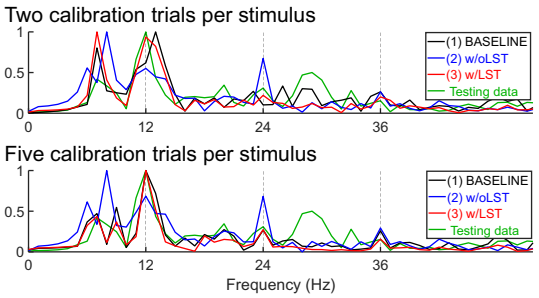


Figure 8. Normalized spectra of averaged EEG signals across all training trials in each scheme (subject 1 as the target subject) and testing trials.

the test subject to build a practical SSVEP BCI, significantly improving the practicality of real-world SSVEP BCIs.

In the two scenarios, the cross-subject transferring within the ActiveTwo system and the cross-device transferring from the ActiveTwo to the Q30 system, the accuracy of the w/LST with two calibration trials per stimulus was equal to or higher than that of the BASELINE with five calibration trials per stimulus. Therefore, in such cases, 60% of calibration time could be saved. Assuming a 40-command BCI speller, the proposed method could save 5 min to collect training data with a trial length of 2.5 s (Stimulation time: 1.5 s; Inter-stimulus interval (ISI): 1.0 s; Three trials per stimulus).

In addition to the simulation validating the LST-based method with different numbers of calibration trials, another simulation study that varied the number of supplementary subjects was also conducted. In this simulation, a leave-one-subject-out cross-validation was also employed, but when preparing the calibration data of the w/oLST and w/LST schemes, different numbers (1, 3, 5, 7, and 9) of other subjects were randomly selected as the supplementary subjects. For each number of supplementary subject, the random selections were repeated ten times. Unlike the first simulation, the number of trials per stimulus was fixed to five, and the first five trials were always used to form the calibration data and the last three rials were used as testing trials. As figure 6 shows, the performance of the w/LST scheme method improved slightly as the number of supplementary subjects increased, while the w/oLST scheme did not. These results suggested that the number of supplementary subjects does not heavily affect the performance of the LST-based method but the more the better. These results are important for the practical scenarios, because in real-world, the number of supplementary subjects (i.e. existing users) is not limited, and it is important to show that the LST-based method does not rely on very specific parameters.

In the comparison of the proposed method with existing approaches listed in table 1, our work stands out as the only generalized framework for multiple cross-domain transfer learning for SSVEP decoding. Although some of the methods do not require any additional calibration data from a target session, the others require a small amount of calibration data from a target session. In general, the methods that do not require additional data can achieve higher accuracy than the calibration-free method, but they are far inferior to fully-calibrate methods such as the TRCA-based method [23]. The ones that require calibration data including the LST-based method employ transfer learning to achieve better performance than fully-calibration methods. Most importantly, the most of existing transfer learning methods can only transfer SSVEPs across one domain

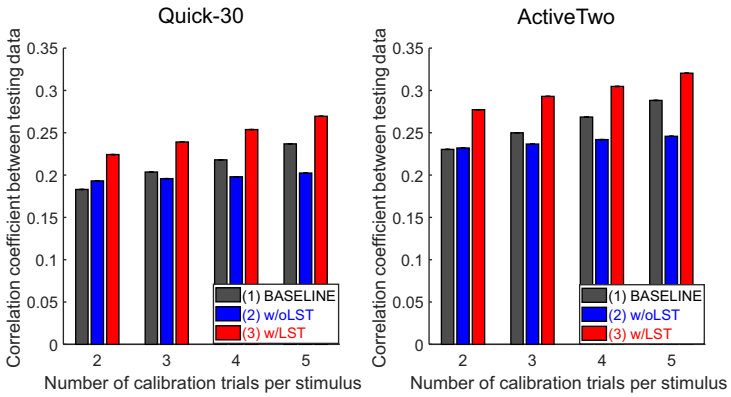


Figure 9. Pearson's correlation coefficients of time domain signals averaged across channels between training trials and testing trials. The correlation is computed by averaging across all target subjects.

such as either cross-subject, cross-session or cross-device scenarios, whereas this study validated that the LST-based method can transfer SSVEPs across multiple domains except cross-stimulus transferring. It indicates that any user could reach a higher accuracy than fully-calibrated methods with a small amount of calibration data collected from his/her own EEG device even if it was the first time for him/her to use the system.

It is also worth noting that the classification accuracy obtained in the simulated online analysis could be generalized to actual online performance. The preprocessing pipeline including notch and band-pass filters was applied to each data epoch separately after the data were segmented. In addition, the ISI was set to 1.0 s in the experiment, which has been commonly and reasonably used in previous studies with online analyses [6, 12].

The main limitation of the proposed method is that it still requires a small number of calibration trials from the new user, and therefore, it is still yet a calibration-free method. In addition, when the quality of the supplemental data is worse than the targeting data, the improvement from the LST-based method is limited. However, in practice, it is more likely that the supplemental data have better signal quality since they can be prepared in a well-controlled environment while the data of the targeting user can be acquired in any general environment.

In a nutshell, the LST enables an effective consolidation of EEG data across users and devices and consistently outperforms the standard TRCA approach and the naive integration of data without LST. Our results suggest using the LST-based method should be taken into account for augmenting calibration data when using TRCA-based SSVEP spellers.

5. Conclusions

This study proposed a cross-domain transfer method based on the LST for transforming SSVEP data across users and devices. The experimental results demonstrated the efficacy of the LST-based method in alleviating the inter-subject and inter-device variability in the SSVEPs. The LST-based method also improved the SSVEP-decoding accuracy by leveraging data from other subjects and/or devices and a small number of calibration data from a new subject. These findings considerably improve the practicality of a real-world SSVEP BCI.

Acknowledgement

This work was partially supported by the US Army Research Laboratory under Cooperative Agreement W911NF-10-2-0022. T-P Jung is also supported, in part, by the US National Science Foundation (NSF) under Grant CBET-1935860. This work was also supported in part by the Higher Education Sprout Project of the National Chiao Tung University and Ministry of Education, Taiwan, the Ministry of Science and Technology, Taiwan, under MOST(NSC) 109-2222-E-009-006-MY3.

ORCID iD

Kuan-Jung Chiang https://orcid.org/0000-0002-2163-071X

References

[1] Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 Brain–computer interfaces for communication and control *Clin. Neurophysiol.* **113** 767–91
[2] Farwell L A and Donchin E 1988 Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials *Electroencephalogr. Clin. Neurophysiol.* **70** 510–23
[3] Wang Y, Gao X, Hong B, Jia C and Gao S 2008 Brain-computer interfaces based on visual evoked potentials *IEEE Eng. Med. Biol. Mag.* **27** 64–71
[4] Vialatte F, Maurice M, Dauwels J and Cichocki A 2010 Steady-state visually evoked potentials: focus on essential paradigms and future perspectives *Prog. Neurobiol.* **90** 418–38

[5] Gao S, Wang Y, Gao X and Hong B 2014 Visual and auditory brain–computer interfaces *IEEE Trans. Biomed. Eng.* **61** 1436–47
[6] Chen X, Wang Y, Nakanishi M, Gao X, Jung T-P and Gao S 2015 High-speed spelling with a noninvasive brain–computer interface *Proc. Natl Acad. Sci. USA.* **112** E6058–67
[7] Lin Z, Zhang C, Wei W and Gao X 2006 Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs *IEEE Trans. Biomed. Eng.* **53** 2610–14
[8] Bin G, Gao X, Yan Z, Hong B and Gao S 2009 An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method *J. Neural Eng.* **6** 046002
[9] Wang Y, Gao X and Gao S 2015 Computational modeling and application of steady-state visual evoked potentials in brain-computer interfaces *Sci. Suppl.* **350** 43–6
[10] Wei C-S, Lin Y-P, Wang Y-T, Lin C-T and Jung T-P 2018 A subject-transfer framework for obviating inter-and intra-subject variability in EEG-based drowsiness detection *NeuroImage* **174** 407–19
[11] Nakanishi M, Wang Y, Wang Y-T, Mitsukura Y and Jung T-P 2014 A high-speed brain speller using steady-state visual evoked potentials *Int. J. Neural Syst.* **24** 1450019
[12] Nakanishi M, Wang Y, Chen X, Wang Y-T, Gao X and Jung T-P 2017 Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis *IEEE Trans. Biomed. Eng.* **65** 104–12
[13] Nakanishi M, Wang Y, Wang Y-T and Jung T-P 2015 A comparison study of canonical correlation analysis based methods for detecting steady-state visual evoked potentials *PloS One* **10** e140703
[14] Zhang Y, Zhou G, Jin J, Wang M, Wang X and Cichocki A 2013 L1-Regularized multiway canonical correlation analysis for SSVEP-based BCI *IEEE Trans. Neural Syst. Rehabil. Eng.* **21** 887–96
[15] Zhang Y, Zhou G, Jin J, Wang X and Cichocki A 2014 Frequency recognition in SSVEP-based BCI using multiset canonical correlation analysis *Int. J. Neural Syst.* **24** 1450013
[16] Zerafa R, Camilleri T, Falzon O and Camilleri K P 2018 To train or not to train? A survey on training of feature extraction methods for SSVEP-based BCIs *J. Neural Eng.* **15** 051001
[17] Wu D, Xu Y and Lu B-L 2020 Transfer learning for EEG-based brain-computer interfaces: a review of progress made since 2016 *IEEE Transactions on Cognitive and Developmental Systems* 1
[18] Yuan P, Chen X, Wang Y, Gao X and Gao S 2015 Enhancing performances of SSVEP-based brain–computer interfaces via exploiting inter-subject information *J. Neural Eng.* **12** 046006
[19] Wong C M, Wang Z, Wang B, Lao K F, Rosa A, Xu P, Jung T-P, Chen C L P and Wan F 2020 Inter-and intra-subject transfer reduces calibration effort for high-speed SSVEP-based BCIs *IEEE Trans. Neural Syst. Rehabil. Eng.* **28** 2123–35
[20] Rodrigues P L C, Jutten C and Congedo M 2019 Riemannian procrustes analysis: transfer learning for brain–computer interfaces *IEEE Trans. Biomed. Eng.* **66** 2390–401
[21] Waytowich N, Lawhern V J, Garcia J O, Cummings J, Faller J, Sajda P and Vettel J M 2018 Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials *J. Neural Eng.* **15** 066031
[22] Nakanishi M, Wang Y and Jung T P 2016 Session-to-session transfer in detecting steady-state visual evoked potentials with individual training data *Int. Conf. on Augmented*
[23] Nakanishi M, Wang Y-T, Wei C-S, Chiang K-J and Jung T-P 2019 Facilitating calibration in high-speed BCI spellers via leveraging cross-device shared latent responses *IEEE Trans. Biomed. Eng.* **67** 1105–13
[24] Suefusa K and Tanaka T 2017 Reduced calibration by efficient transformation of templates for high speed hybrid coded SSVEP brain-computer interfaces *2017 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* pp 929–33
[25] Chiang K-J, Wei C-S, Nakanishi M and Jung T-P 2019 Cross-subject transfer learning improves the practicality of real-world applications of brain-computer interfaces *Proc. 9th Int. IEEE/EMBS Neural Eng. Conf.* pp 424–7
[26] Nakanishi M, Wang Y, Wang Y-T, Mitsukura Y and Jung T-P 2014 Generating visual flickers for eliciting robust steady-state visual evoked potentials at flexible frequencies using monitor refresh rate *PloS One* **9** e99235
[27] Crochiere R E 1979 A general program to perform sampling rate conversion of data by rational ratios. *Speech Digital Signal Processing Committee of the IEEE Acoustics and Signal Processing Society,* ed Programs for Digital Signal Processing (Piscataway, NJ: IEEE)
[28] Tanaka H, Katura T and Sato H 2013 Task-related component analysis for functional neuroimaging and application to near-infrared spectroscopy data *NeuroImage* **64** 308–27
[29] Chen X, Wang Y, Gao S, Jung T-P and Gao X 2015 Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface *J. Neural Eng.* **12** 046008
[30] Anderson M and Ter Braak C 2003 Permutation tests for multi-factorial analysis of variance *J. Stat. Comput. Simul.* **73** 85–113
[31] van der Maaten L and Hinton G 2008 Visualizing data using t-SNE *J. Mach. Learn. Res.* **9** 2579–605