

Representative-Based Cold Start for Adaptive SSVEP-BCI

Nanlin Shi¹, Student Member, IEEE, Xiang Li¹, Student Member, IEEE, Bingchuan Liu¹, Student Member, IEEE, Chen Yang¹, Member, IEEE, Yijun Wang¹, Member, IEEE, and Xiaorong Gao¹, Member, IEEE

Abstract—Objective: The tradeoff between calibration effort and model performance still hinders the user experience for steady-state visual evoked brain-computer interfaces (SSVEP-BCI). To address this issue and improve model generalizability, this work investigated the adaptation from the cross-dataset model to avoid the training process, while maintaining high prediction ability. **Methods:** When a new subject enrolls, a group of user-independent (UI) models is recommended as the representative from a multi-source data pool. The representative model is then augmented with online adaptation and transfer learning techniques based on user-dependent (UD) data. The proposed method is validated on both offline (N=55) and online (N=12) experiments. **Results:** Compared with the UD adaptation, the recommended representative model relieved approximately 160 trials of calibration efforts for a new user. In the online experiment, the time window decreased from 2 s to 0.56 ± 0.2 s, while maintaining high prediction accuracy of 0.89-0.96. Finally, the proposed method achieved the average information transfer rate (ITR) of 243.49 bits/min, which is the highest ITR ever reported in a complete calibration-free setting. The results of the offline result were consistent with the online experiment. **Conclusion:** Representatives can be recommended even in a cross-subject/device/session situation. With the help of represented UI data, the proposed method can achieve sustained high performance without a training process. **Significance:** This work provides an adaptive approach to the transferable model for SSVEP-BCIs, enabling a more

generalized, plug-and-play and high-performance BCI free of calibrations.

Index Terms—User dependent BCI, user independent BCI, cold start problem, adaptive BCI.

I. INTRODUCTION

A BRAIN-COMPUTER interface (BCI) is a human-machine interface (HMI) type that enables human-computer interaction by decoding intentions behind neural signals. BCI based on EEG signals has dominated the field owing to its non-invasive nature and ease of use. The paradigm of EEG-based BCIs utilizes steady-state evoked potentials (SSVEP) and has gained increasing attention mainly for its high-speed performance. The SSVEP paradigm achieves target discrimination by eliciting a response evoked by luminous modulated flickers [1]. Currently, the SSVEP-based BCI can achieve an information transfer rate of 325 bits/min, which is mainly attributed to the efficient system design and decoding algorithms like CCA and TRCA [2], [3].

However, the rising performance of decoding algorithms requires balancing other parameters. Training algorithms usually require high amounts of training data to render an accurate, personalized model. During the training process, the subjects are expected to follow the system instructions to collect personal neural responses under the simulation of different class flickers, hindering the free use of the BCI system. Hence, these algorithms can be referred as user-dependent (UD) [4]. The other type of method that does not require training is called user-independent (UI). The UI algorithms thus, rely on strong priors, such as the standard sine/cosine functions or cross-subject data. These types of methods show a significant reduction in the BCI performance because the variabilities across subjects and other experimental conditions are significant. Therefore, they cannot be represented by a set of standard templates.

Consequently, the conflict between calibration effort and system performance remains one of the most critical problems in BCI research. The EEG variability has determined that the classification model is and expected to when any experimental condition (i.e., subjects, devices and sessions) changes. Continuous efforts have been devoted to alleviating the calibration effort while maintaining a decent decoding accuracy. Two lines of approaches have proven to be effective.

Manuscript received 19 September 2022; revised 24 January 2023; accepted 13 February 2023. Date of publication 15 February 2023; date of current version 6 March 2023. This work was supported in part by the National Natural Science Foundation of China under Grant U2241208 and Grant 62171473 and in part by the Key Research and Development Program of Ningxia under Grant 2022CMG02026. (Corresponding author: Xiaorong Gao.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the IRB of Tsinghua University.

Nanlin Shi, Xiang Li, Bingchuan Liu, and Xiaorong Gao are with the Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing 100084, China (e-mail: snl18@mails.tsinghua.edu.cn; l-xiang16@mails.tsinghua.edu.cn; louislbc@hotmail.com; gxr-dea@mail.tsinghua.edu.cn).

Chen Yang is with the School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: yuanchouyc@gmail.com).

Yijun Wang is with the State Key Laboratory on Integrated Optoelectronics, Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083, China (e-mail: wangyj@semi.ac.cn).

Digital Object Identifier 10.1109/TNSRE.2023.3245654

The first is called transfer learning techniques and aims to minimize the distribution gap between well-labeled cross-domain data. Several studies have confirmed that the cross-condition SSVEP data can be informative [5], [6], [7] and that the calibration effort is alleviated with domain adaptation techniques [8], [9]. The second approach is called adaptive learning, and can rectify the classification model with incremental online data [5], [10]. Nevertheless, both approaches have shortcomings. Transfer learning techniques can only function when labeled data are collected, which means they can alleviate the calibration burden but are not free of it. Adaptive strategies can be calibration-free but still require an initiation model. The consequential adaptive process can be heavily affected by inaccurate and incomplete data labeling, when adapting from random parameters. The so-called weak labeling may yield unsatisfying results [11]. Thus, both approaches may not provide a fully calibration-free BCI experience with sustained high performance.

Similarly, the recommendation system, seeking to predict the ‘preference’ a user would give to an item, can be regarded as another HMI system [12]. The recommendation system faces the mutual calibration problem. The same problem has been termed as the Cold Start problem, which challenges a system to provide personalized recommendations without sufficient user-dependent information [13], [14]. Multiple strategies to mitigate this issue have been proposed over the years. One of the most important solutions to the Cold Start problem is to exploit inter-user knowledge. Nevertheless, unlike inter-user knowledge problems in other SSVEP research, this Cold Start problem solution strategy tends to mitigate calibration efforts based only on similar users. The inter-user similarity relationship helps maintain a valid recommendation performance [15], [16].

Inspired by the solutions to the Cold Start problem, this study proposed the Representative-based Cold Start (RCS) method. The core of the RCS involves the employment of transfer learning techniques and adaptation based on representatives. The representatives are an ensemble of UI users, whose data distribution would efficiently approximate the new users. Additionally, the representatives are selected from the pre-existing large-scale dataset with the distinct subject, device and session variability. The representative is suggested to own stronger transferability and thus higher potential performance. The representative model is adapted when online UD data is available to build a fully calibration-free BCI.

II. METHOD

A. Algorithms Used to Fit UD/UI Models

The SSVEP discrimination algorithms can be generalized into the template matching methods after spatially filtered data [17]. Their purpose is to compute a set of optimal spatial filters \mathbf{W} to maximize the respective optimization goal. Then the correlation coefficients between test data and the pre-defined template are calculated as:

$$\rho^{(f)} = \text{corr} \left(\mathbf{W}_Z^T \mathcal{Z}^{(f)}, \mathbf{W}_X^T \mathbf{X}^{(f)} \right) \quad (1)$$

TABLE I
LIST OF NOTATIONS

Notation	Description
\mathbf{Y}_i	SSVEP reference signal for f_i and ϕ_i
$\mathbf{X}_i^{[t]}$	t_{th} trial User dependent SSVEP data for subject i
$\mathcal{X}_i^{\{m\}}$	m_{th} User independent SSVEP data for subject i , where $i = 1, \dots, N_{sub}$
$\bar{\mathbf{X}}_i$	subject's i_{th} SSVEP template corresponding to the i_{th} stimulation class, where $i = 1, \dots, N_f$
$\bar{\mathbf{x}}_i$	representative's SSVEP template corresponding to the i_{th} stimulation class, where $i = 1, \dots, N_f$
\mathcal{X}_i	Multi-UI data for subject i , $\mathcal{X} = \hat{\mathbf{X}}_i^{[1]} \oplus \dots \oplus \hat{\mathbf{X}}_i^{[m]}$
ρ_i	i_{th} correlation coefficient
α	Type I Error, p value

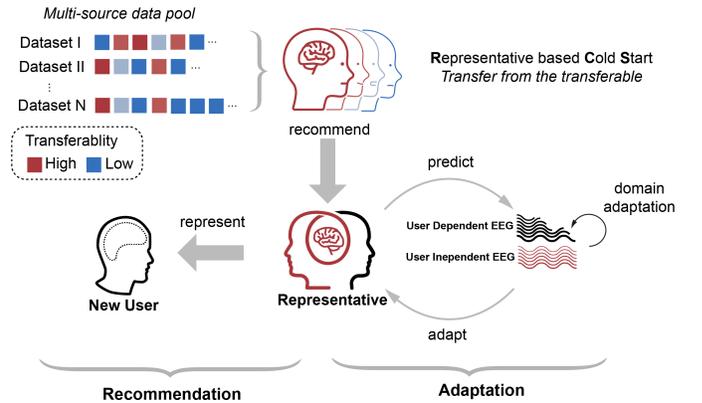


Fig. 1. Schematic flowchart of the RCS method: The proposed method is devised into the recommendation and adaptation stages. In recommendation stage, the ideal transferable UI representative is recommended to substitute as adaptation initiation. The representative model is further rectified by online UD adaptation in the adaptation stage.

The classification is made based on the maximum index of the correlation coefficient,

$$q = \text{argmax } \rho \quad (2)$$

where $\rho = \{\rho_1, \dots, \rho_{N_f}\}$. Various template matching algorithms differ in mainly two factors: templates and spatial filters. The template $\mathcal{Z}^{(f)}$ can be extracted from various sources and is the key to discriminating UD and UI models. The UD model is dependent on calibration data $\mathcal{Z}^{(f)} = \bar{\mathbf{X}}^{(f)}$. The model is considered to be a UD model, if personal specific knowledge $\bar{\mathbf{X}}^{(f)}$ is needed, regardless of the form of the applied algorithm. To substitute for calibration data, the models can introduce a wide range of sources to replace $\bar{\mathbf{X}}^{(f)}$. The most frequent category can be referential signal $\mathcal{Z}^{(f)} = \mathbf{Y}^{(f)}$:

$$\mathbf{Y}_i = \begin{bmatrix} \sin(2\pi f_i t^T) \\ \cos(2\pi f_i t^T) \\ \vdots \\ \sin(2\pi N_h f_i t^T) \\ \cos(2\pi N_h f_i t^T) \end{bmatrix}^T, \quad \mathbf{t} = [1/f_s, \dots, N_p/f_s]^T \quad (3)$$

where f_s is the sampling rate. Another frequently used source is the cross-condition (inter-subject/session/device) data: $\mathcal{Z}^{(f)} = \bar{\mathbf{X}}^{(f)}$. The cross-condition data can be actual EEG data, and thus may be a more efficient UD knowledge than $\mathbf{Y}^{(f)}$.

The spatial filters are the projection matrix calculated based on the temporal template \mathcal{Z} , which can potentially boost the signal-to-noise ratio in regards to the original signal space. This study deployed three spatial filtering algorithms, namely FBCCA, TRCA and TRCA w/R.

1) **FBCCA**: As a classical calibration-free algorithm [2], CCA takes standard templates \mathbf{Y} as matching templates and seeks a pair of spatial filters to maximize the weighted correlation of $\mathbf{W}_Y^T \mathbf{Y}^{(f)}$ and $\mathbf{W}_X^T \mathbf{X}$. The weight vectors are calculated by solving the optimization problem:

$$\mathbf{W}_Y, \mathbf{W}_X = \max_{\mathbf{W}_Y, \mathbf{W}_X} \frac{\mathbb{E}[\mathbf{W}_Y^T \mathbf{Y} \mathbf{X}^T \mathbf{W}_X]}{\sqrt{\mathbb{E}[\mathbf{W}_Y^T \mathbf{Y} \mathbf{Y}^T \mathbf{W}_Y] \mathbb{E}[\mathbf{W}_X^T \mathbf{X} \mathbf{X}^T \mathbf{W}_X]}} \quad (4)$$

Based on the basic CCA method, FBCCA filters X into N_{band} and perform CCA in each band. The correlation within each band is the combined correlation according to: $\tilde{\rho}_i = \sum_{n=1}^{N_{band}} (n^{-1.25} + 0.25) \cdot \rho_{k,i}$, where $i = 1, \dots, N_f$.

2) **Task-Related Component Analysis (TRCA)**: TRCA [3] sets a different optimization goal, compared to FBCCA. With the help of real multi-block SSVEP data, TRCA seeks a series of spatial filters to maximize task-related component consistency over trials. Thus, the optimization goal can be translated as the maximization of inter-trial covariance after applying the spatial filters as:

$$\mathbf{w}^T \mathbf{S} \mathbf{w} = \sum_{h_1, h_2=1}^{N_i} \sum_{h_1 \neq h_2}^{N_{ch}} w_{j_1} w_{j_2} \text{Cov}(x_{j_1}^{(h_1)}(t), x_{j_2}^{(h_2)}(t)) \quad (5)$$

where the h index representing the h_1 -th, h_2 -th trial. The optimization problem is solved under the constraint of:

$$\mathbf{w}^T \mathbf{Q} \mathbf{w} = \sum_{j_1, j_2=1}^{N_{ch}} w_{j_1} w_{j_2} \text{Cov}(x_{j_1}(t), x_{j_2}(t)) \quad (6)$$

Finally, the spatial filter can be solved from:

$$\hat{\mathbf{w}} = \arg \max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{S} \mathbf{w}}{\mathbf{w}^T \mathbf{Q} \mathbf{w}} \quad (7)$$

The linear coefficient $\hat{\mathbf{w}}$ is obtained by eigendecomposition of matrix $\mathbf{Q}^{-1} \mathbf{S}$, where $\hat{\mathbf{w}}$ is the eigenvector that corresponds to the biggest highest eigenvalue. Moreover, the simple TRCA can be upgraded to an ensemble version. The multi-class spatial filters are ensembled to utilize cross-class information and boost discriminability. After applying the same filterbank technique as in FBCCA, the ensembled spatial filter is:

$$\mathbf{W}^{(n)} = [\mathbf{w}_1^{(n)}, \mathbf{w}_2^{(n)}, \dots, \mathbf{w}_{N_c}^{(n)}] \quad (8)$$

where $i = 1, \dots, N_f$. The final predicted label is made by Eq1 where $\mathbf{Z} = \bar{\mathbf{X}}$, and $\mathbf{W}_X = \mathbf{W}_Z = \mathbf{W}$.

Based on the basic ensemble-TRCA, the sine/cosine reference knowledge of FBCCA is hereby used to boost performance. Previous studies have demonstrated that the

knowledge of standard sine/cosine templates could also contribute to discriminability. Hence, the final classification is made by summation of the correlation coefficients of both the FBCCA and TRCA methods as:

$$\rho = w_1 \rho_1 + w_2 \rho_2 \quad (9)$$

where ρ_1 and ρ_2 denotes the correlation coefficients of FBCCA and TRCA, respectively. The weight parameters $w_1 = w_2 = 1$ are not optimized, but optimization is also possible for performance enhancement. The combined algorithm is denoted as TRCA w/R.

B. Model Confidence Estimation Based on Hypothesis Testing

The model confidence evaluation techniques are crucial to the proposed method. Confidence reflects the prediction ability of the model and thus has already been used in various SSVEP-based BCI studies, primarily in dynamic window strategy [18], [19].

Similarly to previous studies, the confidence evaluation method was constructed around correlation coefficients ρ .

Since all template matching methods are classified by determining the biggest correlation coefficient ρ_q as the predicted target, the predictive confidence of ρ_q was investigated in specific in the hypothesis testing framework.

The original hypothesis by Yang et al., [20] was simplified into two classes: $\mathcal{H}^{(q)}$, which means take the q_{th} target as result, while $\mathcal{H}^{(0)}$ is the hypothesis to reject the assumed q_{th} as targets [18].

$$\begin{aligned} \mathcal{H}^{(0)} : \mathbf{X} &= \mathbf{S}^{(i)} + \mathbf{W}, i \neq q, i = 1, \dots, N_f \\ \mathcal{H}^{(q)} : \mathbf{X} &= \mathbf{S}^{(q)} + \mathbf{W}, i = q \end{aligned} \quad (10)$$

where \mathbf{S}^i represents the evoked response of i_{th} target, and \mathbf{W} denotes background noise. Furthermore, the conditional distribution is assumed to be a normal distribution:

$$p(\rho | \mathcal{H}^{(j)}) \sim N(\mu_{(j)}, \sigma_{(j)}^2)$$

To determine which to accept, a likelihood ratio test can be performed over the two hypotheses: accept \mathcal{H}_0 if $\Lambda > c$ else reject, where Λ stands for the likelihood ratio over two hypotheses:

$$\Lambda(\rho) = \frac{\mathcal{L}(\rho | \mathcal{H}_0)}{\mathcal{L}(\rho | \mathcal{H}_q)} \quad (11)$$

which is equivalent to testing t -statistic over ρ

$$T(\rho) = \frac{\bar{\rho} - \mu_0}{s/\sqrt{n}} > c^* \quad (12)$$

where $c^* = \sqrt{(c-1)(n-1)}$. This becomes a classical Student's t -test scenario. If \mathcal{H}_0 is accepted under c , the Type I Error of \mathcal{H}_0 will be:

$$P(\Lambda < c | H_0) = \alpha \quad (13)$$

The Type I Error α is the conditional probability that the current data is insufficient to render ρ_q as a result but still makes a mistake and claims \mathcal{H}_q . Since α reflects the prediction

ability of the model over current data, $\gamma = 1 - \alpha$ can directly be treated as the model confidence.

From another perspective, the model confidence evaluation method is equivalent to performing a one-tail t -Test between ρ_q and $\tilde{\rho}_q$, where $\tilde{\rho}_q$ represents the remaining ρ excluding ρ_q . If the model is confident enough, ρ_q should be significantly higher than $\tilde{\rho}_q$. This technique is easy to implement by any statistical toolbox. In this study, the python statsmodels package was used to render $\alpha = ttest_ind(\rho_q, \tilde{\rho}_q, 'larger')$.

1) *Recommend Representatives*: The foundation of the proposed method is to recommend representatives without class labels. Based on the model evaluation method mentioned above, the theory, which was initially used in the dynamic stopping strategy to assess transferability/similarity, can be applied. If representatives exist, they can be recommended according to α ranking without the requirement of labeled data.

Specifically, during recommendation trials, α rankings are updated until the top k UI models are consistent for m trials. The initiation model will be trained by an ensemble of top k UI data. If the ideal representative data is not decided until a batch of 20 trials is finished, the top k candidates of the latest trial are treated as final representatives.

2) *Dynamic Window Over Batches*: A distinguished superiority of UD models over UI ones is that UD models can maintain high prediction ability even under a short time window of EEG data. Therefore, the Time Window (TW) can be reduced to an ideal length as the adaptation progresses, resulting in a greater Information Transfer Rate (ITR). To apply the confidence evaluation in dynamic stoppings, a unified threshold of $\alpha_0 = 0.025$ was set, under which the model is confident enough to yield predicted results once $\alpha < \alpha_0$. A condition for TRCA (or TRCA w/R) is that it should fit under templates with the same data length, so TW was shrunk for each block instead of adjusting each trial. This adjustment is a prominent distinction from other dynamic strategies. Finally, the evaluation was made from 2 s to 0.4s for every 0.2 s step, until $\alpha < \alpha_0$ for each batch.

C. Representative-Based Cold Start (RCS) Method

The full procedure of the RCS scheme can be divided into two stages, as seen in Fig1. The model is adapted for $\mathbf{X} \in \mathbb{R}^{N_f \times N_T \times N_C \times N_s}$ when N_T increases

- 1) *Recommendation*: The objective in this stage is to recommend transferable representatives for i_{th} subject as quickly as possible. The recommendation is made by sorting and ranking accumulated α based on refrecommmend. In this stage ($N_T \leq 20$), the predictions are made via the FBCCA model. When representative data $\chi_i \in \mathbb{R}^{k \times N_T \times N_C \times N_s}$ ($k = 5$) is determined, a TRCA w/R model is fit by χ_i .
- 2) *Adaptation*: The purpose of this stage is to modify the representative model with incremental UD data. The adaptation is conducted in a class-wise manner.
 - a) *Fusion*: Since all domain transfer techniques are required for accurate estimation on the class-wise template, the weak-labeled UD data $\mathbf{X}^{(f)}$ are

TABLE II
COMPARISON OF CALIBRATION-FREE METHODS

Scheme	Adaptive	Initialization	Knowledge	Main Algorithm
RCS	✓	χ	$\chi, \mathbf{X}, \mathbf{Y}$	TRCA w/R
SA	✓	None	\mathbf{X}, \mathbf{Y}	TRCA w/R
FBCCA	×	\	\mathbf{Y}	FBCCA

RCS, SA represent Representative-based Cold start and Self Adaptation, χ denotes representative UI data, \mathbf{X} denotes weak labeled UD data, \mathbf{Y} denotes standard sine/cosine template.

Initialization is the starting point of the adaptation and knowledge is the data incorporated to fit a discrimination model.

simply concatenated to $\chi_i^{(f)}$ as $\chi_i^{(f)} \oplus \mathbf{X}_i^{(f)}$. when $N_T < 2$. This stage is devised to boost performance when UD data is insufficient.

- b) *Domain Adaptation*: When $\mathbf{X}^{(f)}$ exceed two trials, a transfer learning technique termed Least Square Transformation (LST) [9] is performed to minimize the distribution gap between $\chi_i^{(f)}$ and $\mathbf{X}^{(f)}$.

$$\mathbf{P}^{(f)} = \operatorname{argmin}_p \operatorname{Tr} \left[(\bar{\mathbf{X}}^{(f)} - p\chi^{(f)})(\bar{\mathbf{X}}^{(f)} - p\chi^{(f)})^T \right] \quad (14)$$

The optimization problem is solved by:

$$\mathbf{P}^{(f)} = \bar{\mathbf{X}}^{(f)} \chi^{(f)T} \left(\chi^{(f)} \chi^{(f)T} \right)^{-1} \quad (15)$$

Finally, the representative data is transformed according to: $\chi_i'^{(f)} = \mathbf{P}^{(f)} \chi_i^{(f)}$. After adaptation, a new TRCA w/R model is the concatenation of two sources: $\chi_i' \oplus \mathbf{X}_i$.

The trials are segmented into batches ($N_T = 20$) for dynamic stopping when implementing the RCS method. The TW is shrunk monotonically for each batch based on Section II-B.2

Two other calibration-free methods were used for comparison. The first is the FBCCA model, to be used as a UI baseline. The other is Self-Adaptation (SA), which is another adaptive method, as a UD baseline. The SA method differs from RCS only in the representatives: only user-specific knowledge is included to fit a TRCA w/R model. Most importantly, the UD data is labeled by RCS for the SA method. This procedure is performed to avoid catastrophic performance decline caused by weak labeling. The comparison of the three approaches is illustrated in Table II

D. Offline and Online Experiment

1) *Data Pool*: The data pool contained 55 participants, 35 from dataset I (the public benchmark dataset in [21]) and the 20 from dataset II. Both datasets were acquired under the classical 40-target SSVEP BCI speller experiment (see [1], [21]). During the experiment, the subject was directed to shift their gaze to a random cued target before stimulation

started. The stimulation duration lasted 5 s, and once the cue was shown, the screen was black for 0.5 s. The visual stimulation was presented on a stimulator under a 60 Hz refresh rate. The flicker layout remained the same 5×8 matrix, containing 40 characters. These targets follow a JFPM coding rule, which comprises a series of flicker frequencies ranging from 8 Hz to 15.8 Hz with an interval of 0.2 Hz. Aside from that, the initial phase ranged from 0 to 2π with a step of 0.5π . Each share of data contained six blocks and each block consisted of 40 trials covering all the stimulation targets. Only the data in the occipital lobe (i.e., O1, Oz, O2, PO7, PO3, POz, PO4, PO8, Pz) were included in further analysis.

The same visual paradigm was used in all experiments, while the other parameters were varied. First, participants in datasets I and II did not overlap. Second, the acquisition device for dataset-I was the Synamps2 EEG system (Neuroscan, Inc.), whereas a Brain AMP DC system (Brain Product) was used in dataset II. The paradigm in dataset I was presented on a 23.6-inch LCD monitor (Acer GD245 HQ), while for dataset-II, the paradigm was presented on a 19-inch (42 cm \times 24 cm) projection screen using a ProPixx projector.

Thus, the entire data pool is considered as variable as possible. Unlike the typical transfer learning studies where only one aspect of the experiment is varied, multiple conditions were varied simultaneously in this experiment.

2) *Offline Experiment*: The offline experiment was conducted on every subject in the data pool. Each subject was considered a new user for the system, meaning that the representatives were recommended among the remaining 54 subjects. For one of the six cross-validations, five blocks of SSVEP (200 trials in total) were treated like online incremental data with unknown labels to simulate the free-spelling process.

3) *Online Experiment*: Twelve healthy subjects (six males and six females) were recruited for the online experiment. All subjects were required to sign the consent form approved by the Research Ethics Committee of Tsinghua University.

The EEG data were acquired using a 9-channel montage (O1, Oz, O2, PO7, PO3, POz, PO4, PO8, Pz) by a Synamps2 EEG system (Neuroscan, Inc.) with a sampling rate of 1000 Hz. The electrode impedance was reduced under $20\text{ k}\Omega$. All data were notch filtered at 50 Hz to exclude power-line noises, then down-sampled to 250 Hz before decoding. No extra operation was performed. Also, unlike the prior studies, the stimulation programs were coded by PsychoPy [22] in Python instead of Psychtoolbox [23].

The stimulus configuration and experimental design remained consistent with the data pool. The 200-trial free-spelling process consisted of five blocks. Each block was segmented in two batches ($N_T = 20$). The subjects were asked to rest between batches until they felt ready to proceed. Prior to the free-spelling process, a block of test data was collected for evaluation.

4) *Method Evaluation*: The principle for method evaluation was to measure model performance of the RCS, SA and FBCCA models under the same experimental conditions, including data length, calibration and test data. The following procedures were followed during the experiment:

- 1) Scramble label: The labels were scrambled to fully simulate a free spelling process, meaning data from the same class may appear in successive order, and a “block” may not cover all targets. Scrambled order is the only way for a model to reflect real-life performance.
- 2) Complete test data: During each batch, the three alternatives in Table II were evaluated on ITR and accuracy on a separated test data, covering all the classes. The ITR (in bits per minute, bpm) was computed according to:

$$ITR = 60 \cdot \left(\log_2 M + P \log_2 P + (1 - P) \log_2 \frac{1 - P}{M - 1} \right) / T \quad (16)$$

where M denotes the number of classes, P denotes the accuracy, and T (in s) denotes the selection time, including gaze time (TW) and gaze shift time (0.5 s).

This procedure was performed to avoid popularity bias, meaning that the model can be biased over targets that appear frequently, but has no discrimination ability against the infrequent ones.

- 3) Same TW: In both online and offline experiments, the evaluation was conducted during batch intervals (20 trials). After each batch, the TW was shrunk according to the proposed RCS method based on Section II-B.2. The accuracies and ITRs were calculated for every batch on the separated test data under the same TW. In the offline experiment, the test data was shuffled by 6-fold cross-validation, while in the online experiment, the test data was collected before free-spelling starts.

III. RESULT

A. Offline Experiment

1) *Representative Model Performance*: The cross-model performance within the data pool was calculated to investigate the feasibility of selecting cross-dataset models as representatives. Each cell in Figure 2A illustrates the TRCA model accuracy (TW=2 s) for UD/UI test data. Row i of the matrix represents the model performances on the test data of subject i . The diagonal reflects the performance of the UD model for each subject, which has a substantial advantage over UI models. Nonetheless, the results confirmed the existence of high-performance representatives for the majority of subjects in the data pool, some of which matched the prediction ability of the UD model. Thus, the results in Figure 2A confirmed that cross-dataset SSVEP data shares similar distributions within groups. It is plausible to modify representative models for a swift transition towards UD models.

After confirming the existence of the representatives, it remains to be clarified they can be recommended without access to class labels. Figure 2B depicts the distributions of the coefficients (summation over filterbanks) for all subjects under two hypotheses. The shaded area shows the Type I Error for the likelihood ratio test, representing the concurrent prediction confidence level. The relationship between prediction accuracy and α was then explored. The results in Figure 2C show

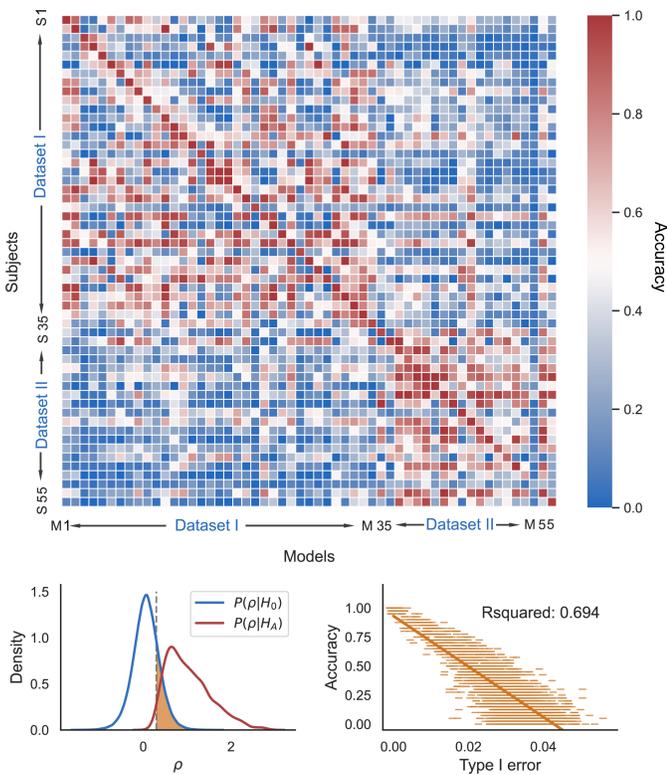


Fig. 2. Cross-dataset performance: A. Cross model prediction in the data pool. Cell (i, j) denotes the prediction accuracy of the TRCA model (TW=1s) trained on the i_{th} subject tested on j_{th} subject. B. Coefficient ρ under two hypotheses for all subjects. The shaded orange area represents the Type I Error α . C. Cross model accuracy is correlated with Type I Error.

a clear connection between Type I Error (average over test set) and prediction accuracy for all subjects, which indicates that recommending high-performance models according to α is possible. Since indicator α is calculated based on likelihood testing, no posterior class label is required during the recommendation.

Moreover, the superiority of recommended UI models to other adaptive initiation options remains to be explored. One can naturally start the adaptation process on a unified model trained on the entire data pool or with the standard template(FBCCA). In **Figure 3D**, these alternatives are compared. The recommended model was trained on the top k ($k=5$) transferable UD data, whereas the unified model was trained on the entire data set (excluding UD data, $k=54$). Both models were fit with the TRCA algorithm. The average accuracies are 0.79 ± 0.18 , 0.55 ± 0.26 , 0.59 ± 0.27 for the representative model, unified model and FBCCA, respectively (TW=1 s). The one-way ANOVA results showed a statistically significant distinction between these three models under TW<1 s. The post-hoc paired t -test also revealed that the representative model performance is significantly higher than the two other models ($p < 0.001$ with Bonferroni correction). The results show that small groups of representative data outperform the unified UI data with large data sizes when building a UI model under the state-of-the-art algorithm. Namely, it's wise to learn from 'the better few'.

Last, the number of trials (m) required to produce UI data with acceptable performance was considered. The upper panel in **Fig 3C** illustrates the recommended model accuracy as the number of recommendation trials increases (TW = 2 s). After about five trials of unlabeled UD data, stable recommendations are obtained, indicating that the proposed method can be rapidly initiated when free-spelling begins. For another aspect, we investigate the impact of the number of representatives (k) on the recommended model performance. As **Fig 3C** showed, around 5 UI subject need to be ensembled to maintain representativeness. Either too large or too small of the representatives can cause a decrease in model performance. Although the initiation TW=2 s was chosen in this study to achieve a stable transition from the UI to UD model, initiation TW can be adjusted according to the recommendation efficiency.

The findings suggest that the transferability relationship lies in the data pool. The representatives can be recommended rapidly based on α .

2) Cold Start Methods Comparison: The adaptation boost is initially reflected in the reduction of the optimal window. The link between TW and α was investigated by calculating the mean α on the test data for each adaptation batch. As **Fig 3B** shown, under TRCA w/R (blue line), α can drop drastically as TW increases. This tendency indicates that α can be an eligible indicator for the dynamic window strategy. The shading from green to blue signifies the adaption process over batches. When a fixed threshold α_0 is set, the optimal TW can be reduced with the increasing proportion of UD data. The choice of the threshold α_0 will consequently affect the optimal window. An inappropriate threshold either overestimates or underestimates the optimal time window. Moreover, **Fig 3B** also highlights the essential contribution of both standard templates and actual EEG data. The orange and green lines indicate the α trend for EEG data alone and standard templates over adaption. The Type I Error measured using the EEG template is substantially lower than using the standard template, especially at short TW, supporting the requirement of actual EEG data to build an effective classification model. However, the best trend was obtained by integrating EEG and standard templates (blue line). Under the combined knowledge of both templates, a shorter TW may be applied at a specified threshold for classification. In the offline experiment, the TW was reduced from 2 s to 0.82 ± 0.42 s on average ($\alpha=0.025$).

The accuracy and ITR of three strategies are compared in **Figure 3A**. The shaded background is mapped to the TW in each batch. Without the knowledge of UD data, FBCCA accuracy declines significantly from 0.86 ± 0.2 to 0.33 ± 0.17 during this process. Next, the main objective is to assess whether the representative model can contribute substantially to the adaptive process. The results in **Figure 3A** support the contribution of the representative model: the prediction accuracy on test data remained high, even when the TW was reduced. On the contrary, without the contribution of the representative model, the accuracy of the SA model decreased when calibration data was insufficient. With the accumulation of weak-labeled UD data,

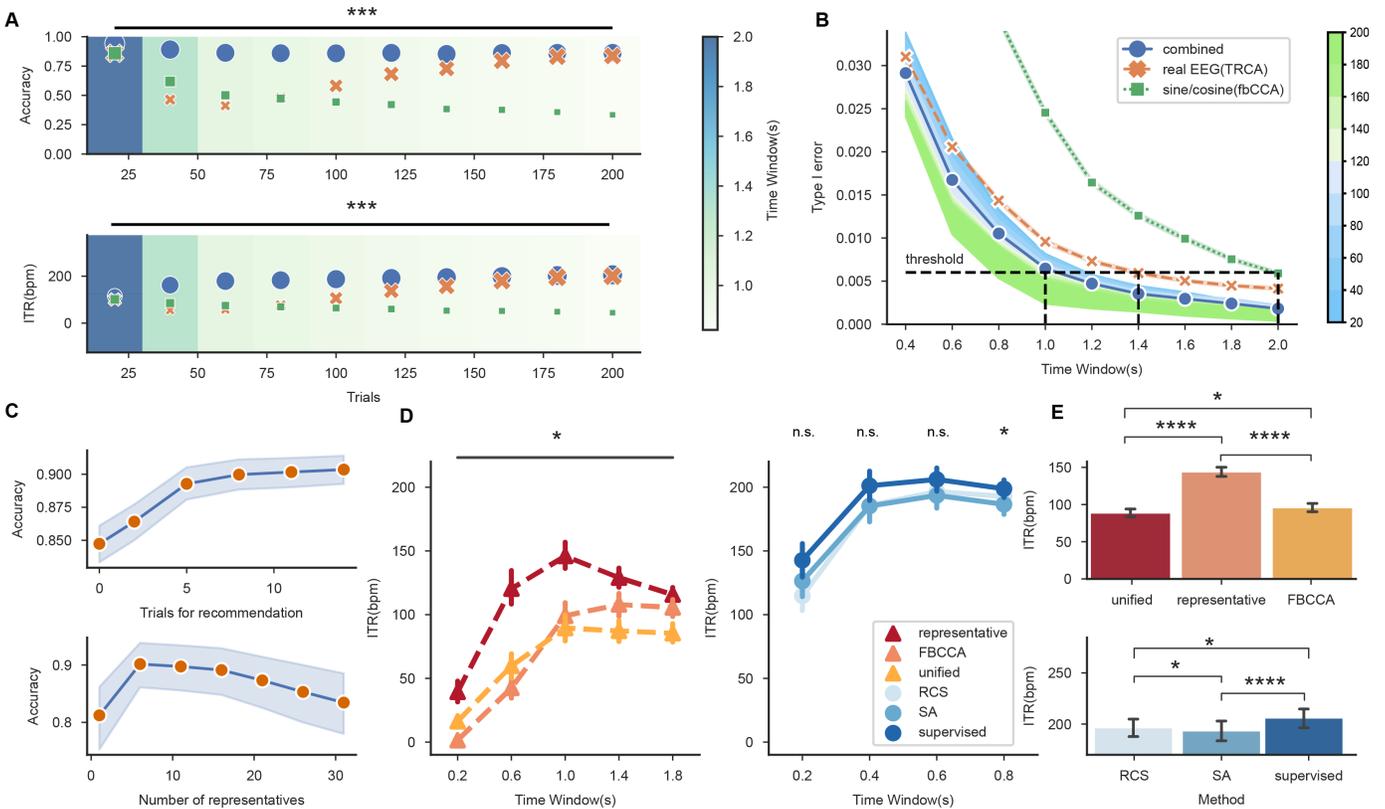


Fig. 3. Offline experiment results: A. Prediction accuracy and ITR on test data during the adaptation process. The background color represents the TW in each batch. Bars on top denote significance level ($p < 0.001$) by paired t -test between RCS and SA. B. α -TW relations. Three lines represent the average α tendency over adaptation batches, and background colors from blue to green signify the experiment process with the increase of online trials. C. Up: Num. of trials with the representative model accuracy (TW=2 s in this study); Below: Num. of representatives on the impact of UI model performance. D. Model performances of multiple methods before and after adaptation. The significance level is calculated by one-way ANOVA (TW). E. ITR comparison at TW=1s, tested by post-hoc t -Test.

the prediction accuracy steadily increased after the 60th trial. The accuracies at 60th trial (TW=1±0.45 s) were 0.86±0.13 and 0.41±0.18 for with/without representative model, respectively. The two adaptive methods can achieve approximate predictability ability under the 160th trial. This is consistent with former studies: at least four blocks of UD data need to be collected to build an ideal classification model. Consequently, although both models increased their performance, the adaptation based on the representative model outperformed SA throughout the experiment. The final ITRs of adaptation were 206.08 bpm and 200.20 bpm for RCS and SA. The t -test confirmed the significant difference between the two adaptive methods ($p < 0.001$).

As adaptation and transfer learning have proved beneficial to BCI performance boosting [5], [7], [10], that transfer is further elucidated in this study and the representative-based model can reduce the trial calibration effort by 80% (160/200). Therefore, it is possible to design calibration-free SSVEP BCI with sustained high performance.

3) Convergence of Weak Models: After 200 adaptation trials, the ability of the adapted model to converge to the conventional supervised model needs to be assessed. Similarly to the previous sections, the final weak labeled models were examined with the standard UD model on the separated test data (TW=0.2 to 1 s). The TRCA w/R algorithm was

used to build these models. The main distinction between these models is that the adapted ones are trained on weak labels, whereas the supervised ones are considered perfectly labeled. As Figure 3D and E shown, the supervised model outscored the weakly supervised models by a small margin on average. The ANOVA test showed a significant difference between the three alternatives under TW=0.8, 1 s ($p < 0.05$). The highest ITR was obtained under TW=0.6 s. The peak ITRs were 206±79.3 bpm, 196.4±78.8 bpm, and 193.4±83.5 bpm for standard supervised, RCS and SA models, respectively. A post-hoc paired t -test was additionally performed among the pairs of each group (TW=1 s). The results showed that the standard UD model remained significantly more advanced than the two weak labeled models ($p < 0.001$ with Bonferroni correction). The ITR performance of RCS is slightly lower in Section III-A.2, which can be attribute to the conduction of LST incrementally or all at once.

Although adaptation can significantly improve model performance, the inaccurate and incomplete weak labels still have a noticeable impact on the ultimate model prediction. The adaptation can only approximately converge to the ideal one. Moreover, the knowledge of representative data contributes marginally to UD data, indicating that SSVEP data is still highly personalized.

TABLE III
METHOD COMPARASION BEFORE AND
AFTER COMPARASION (TW=1s)

	Method	Accuracy	ITR(bpm)
Before Adaptation	FBCCA	0.59±0.27	96.96±60.24
	Representatives	0.78±0.28	144.12±57.39
	Unified	0.56±0.26	88.95±58.41
After Adaptation	RSA	0.89±0.18	177.42±48.05
	SA	0.87±0.21	172.00±53.20
	Standard	0.91±0.17	182.17±46.76

B. Online Performance

Figure 4 demonstrates the online experiment results at averaged and individual levels. Online experiments exhibited a similar tendency to offline results. The optimal TW was reduced from 2 s to 0.56 s on the averaged level. Thus, ITR was increased from 117.87 bpm to 243.49 bpm for the RCS method. Nevertheless, a significant difference over SA was only detected for the first two blocks ($p < 0.001$). This means that the representative model is statistically effective in the first 80 trials instead of throughout the entire experiment as in the offline experiment. Multiple reasons can lead to this, including the lack of extensive group validation ($N=12$) and the deficiency of cross-dataset recommendations.

On the individual level, 10 of the 12 subjects in total showed clear superiority of the RCS strategy (subjects No. 1, 2, 3, 4, 5, 6, 7, 8, 10, 12). For subjects No. 9 and No. 11, the ITR for SA surpassed the proposed scheme after UD data was sufficient, meaning that representative data possess a large distribution gap with the UD data. In this scenario, the representatives can hinder model prediction.

The contribution of representative recommendation for two separate datasets was also investigated. Figure 5B visualizes the cross-model accuracy similar to Figure 3A, which confirmed that the representative recommendation is still valid for the online dataset. In Figure 5A, each row represents the cross-model accuracy recommended by RCS method. On the subject averaged level, dataset I contributes approximately 60% (3 out of 5 ($k=5$)) of the recommendation. Furthermore, the averaged recommendation accuracy of dataset I was far superior than dataset II, 0.76 over 0.55. However, because representative recommendation can be extremely individualized, it is inaccurate to claim that dataset I is more attributive. The ultimate prediction ability of the representative UI model is denoted as the black triangular marker. Ensembled representative model accuracy is determined by the best representatives, meaning that the UI can tolerate few least ideal representatives.

IV. DISCUSSION

With year's devolution on leveraging calibration and performance, the expected ITR for a UI model([2], [10], [18], [5]) is around 150 bpm, whereas the UD model([7], [24], [3]) can achieve ITR 200~350 bpm. This study focused on smooth transition from the UI to UD model, providing a final ITR performance above 200 bpm (offline: 206.8 bpm, online: 243.49 bpm). To the best of the knowledge of the authors,

the ITR of 243.49 bpm is the highest ever reported in the online, calibration-free configuration. Moreover, rather than confining such scheme in deliberately controlled experimental conditions, this work revealed that such transition is still effective across multi-source public datasets.

A. Weak Labeling Effect in the Adaptation

However, being relieved from labeled calibrated data means weak labeling inevitably, and the inaccurate and incomplete labeling of UD data can lead to model deficiency [11]. Furthermore, extremely weak labeling can have a catastrophic impact on the following adaptive process, especially when UD data is insufficient. In this sense, the representatives not only provided a good starting point but also compensated for the weak labeling effect:

- 1) Inaccurate label: Inaccurately labeled UD data are combined with accurately labeled UI data to balance prediction ability between different classes.
- 2) Incomplete label: Complete set of UD data that covers each target class can rectify bias prediction when UD labeling is incomplete. Different target classes are not expected to appear in equal frequency in a real-world scenario, which means the model can be fed with an abundance of data on one target but has not seen an infrequent target for a long time.

The weak labeling impact in the offline experiment throughout the adaptation was then qualified in Figure 6A. Besides RCS and weak labeled SA, the self-adaptation process was simulated as if the labels were all correct (Note that this method also requires calibration because standard labels are needed). Within the two SA alternatives, the weak labeling impact was significant after the 100_{th} trial. With sufficient calibration, the standard labeled SA surpassed RCS for 20 bpm (same result in Figure 3E). An incomplete label effect is also evident at the beginning of the adaptation when SA developed a popularity bias with a few classes of calibration data.

Thus, other initiation options are theoretically more susceptible to weak labeling effect, for instance, if adaptation starts from random parameters, like in [10], the performance would likely decline in the first few trials, same with the SA methods in this study. Thus, rectifying a robust UI model covering all the classes (like the representative model in this study) is necessary to avoid catastrophic performance loss.

B. UI Model as Pre-Trained Model

The results so far show that representative data saves calibration/adaptation efforts. In this sense, the proposed work can be seen as a similar framework to the well-known pre-trained model strategy. Pre-trained models are typically Deep Neural Network (DNN) models trained on a large-scale dataset, used for the solution of similar problems. The biggest contribution of the pre-trained model is to avoid building a target-specific model from the beginning. Additionally, the pre-trained model can be further fine-tuned with target-specific data. Although the presented model was not a DNN-based model, the representative model in the proposed method acts

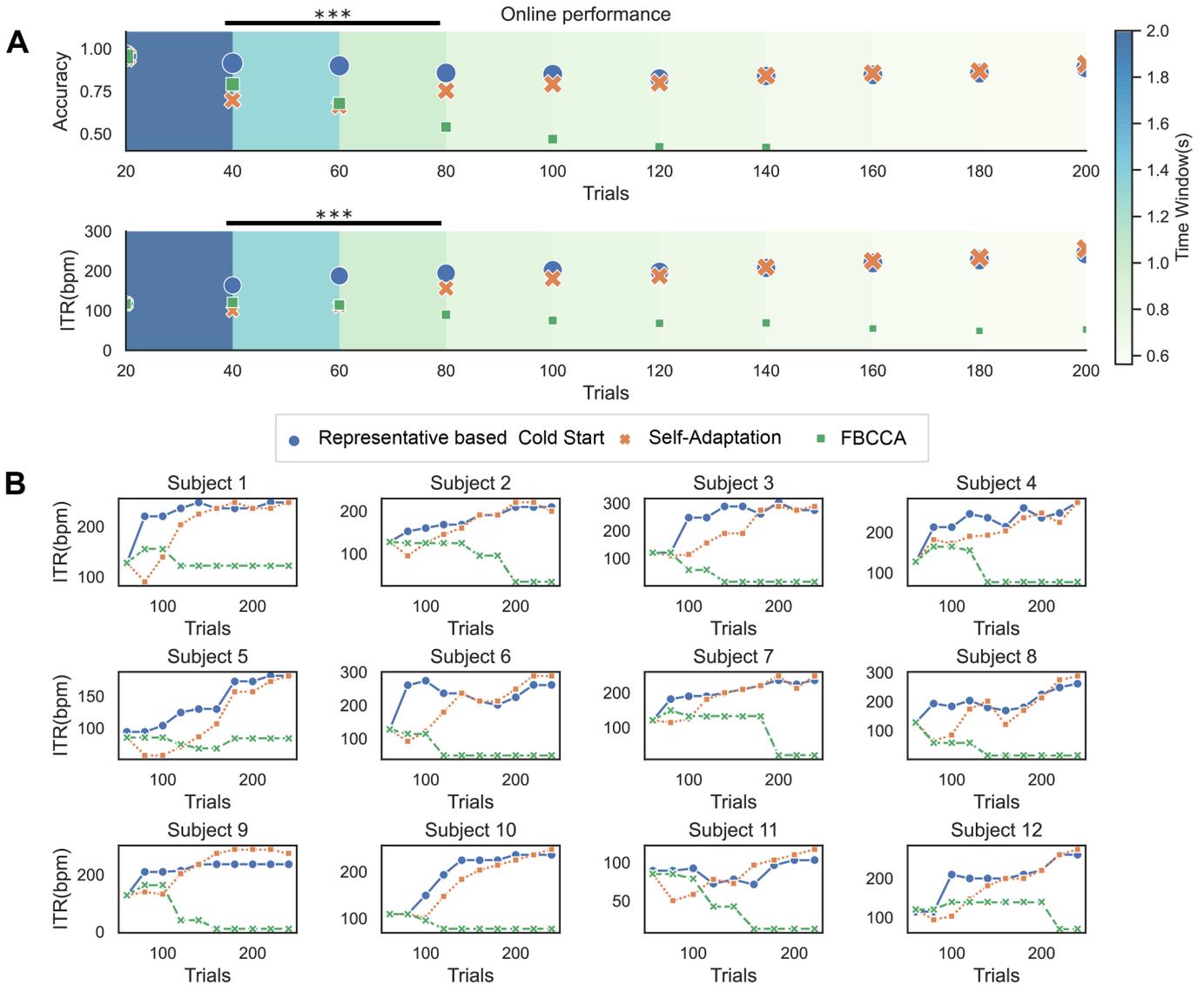


Fig. 4. Online experiment result: A. Prediction accuracy and ITR for three comparisons. B. ITR performance for all subjects.

TABLE IV
ONLINE EXPERIMENT PERFORMANCE

Trials		20	40	60	80	100	120	140	160	180	200
TW(s)		2.00±0.00	1.33±0.61	1.00±0.46	0.82±0.38	0.75±0.36	0.70±0.35	0.65±0.23	0.60±0.21	0.58±0.20	0.56±0.20
RCS	ITR (bpm)	117.87±13.69	163.17±57.20	187.75±53.99	194.23±52.10	201.96±57.00	198.49±55.51	207.45±45.08	219.77±46.82	227.56±46.42	243.49±28.21
	Accuracy	0.96±0.06	0.91±0.06	0.90±0.07	0.86±0.12	0.85±0.12	0.83±0.14	0.84±0.12	0.85±0.10	0.86±0.10	0.89±0.06
SA	ITR (bpm)	117.24±16.05	99.65±36.16	111.21±33.16	156.23±43.53	180.38±52.33	187.00±53.06	209.39±51.23	225.15±50.28	233.76±50.44	255.67±36.07
	Accuracy	0.95±0.08	0.70±0.18	0.66±0.11	0.75±0.10	0.79±0.11	0.80±0.14	0.84±0.11	0.86±0.09	0.87±0.09	0.91±0.04
FBCCA	ITR (bpm)	117.24±16.05	120.82±33.63	114.09±38.72	89.50±42.09	75.26±45.06	67.86±48.02	69.15±48.23	54.69±44.88	48.98±36.75	52.21±36.72
	Accuracy	0.95±0.08	0.79±0.17	0.68±0.16	0.54±0.19	0.47±0.22	0.42±0.23	0.42±0.23	0.35±0.21	0.32±0.17	0.34±0.17

in the same role as the pre-trained model. However, the pre-trained model framework is only suitable for two principles:

- 1) Large scale dataset: Years of effort from the entire BCI community have yielded large data sizes [25], especially for SSVEP-based BCIs [26], [27].
- 2) High similarity: In contrast to other paradigms like P300 and MI, the SSVEP paradigm has long been believed to be generalizable across populations, despite the fact that experimental settings are highly inconsistent.

The two principles are intertwined in this study. Because of the high variability in cross-condition datasets, every data share cannot be applied to fit a UD model (i.e., the unified

model). Instead, it is necessary to select a few data to boost similarity/transferability in the large-scale data pool. However, the proposed method does not guarantee that the representatives can be found. Figure 6 depicts the scenario where the ITR can decrease when pool size shrinks. This decline is likely attributed to decreased representative model performance caused by reduced pool size. On the other hand, it can be predicted that the transferability will increase with sufficiently large datasets.

Finally, when the pre-trained model is fitted on UI data with stronger transferability, fewer adjustments are needed to achieve rapid convergence to the ideal UD model. It is

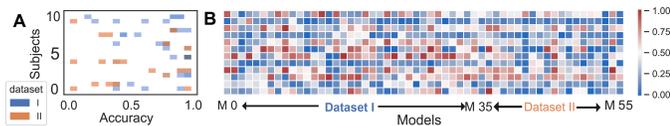


Fig. 5. Representative recommendation for the online experiment: A. Representative model accuracy for Dataset I, II. Each row denotes model accuracy of $k = 5$ recommendation for the two datasets (TRCA). The black triangle markers represent the final representative model accuracy. B. Cross model performance for the online experiment data ($N=12$, $TW=2s$).

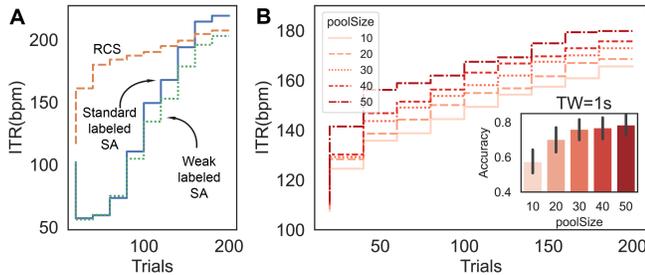


Fig. 6. The effect of weak labeling to model performance: A. Standard label SA over weak labeled SA. B. The proposed RCS method influenced by data pool size. Lines represent the RCS performance over the adaptation, bar plot denotes the respective representative model ($TW=1s$) performance for different pool sizes.

still possible that a large-scale unified DNN algorithm can be trained to render a more effective pre-trained model [28], or the linear combination of the representatives can be used to further boost performance [7]. However, the authors still want to underline that it is feasible to build an eligible pre-trained model with state-of-the-art algorithms, if inter-data transferability is respected.

C. Mechanism Behind Representatives

Figure 2A shows that the two datasets exhibited an apparent inter-dataset gap. The questions that need to be answered are, why are some data more transferable than others and why does cross-dataset still show a significant performance gap. This complicated question can be answered from a spatial filter perspective. In Figure 7, it is proposed that phase difference may also be a significant component in determining model transferability.

This inter-dataset phase delay was explored by evaluating the TRCA model trained on dataset I but tested on dataset II. The model was trained with a constant visual latency (140 ms) but with varying test latency from 132-164 ms. Each dot in Figure 7A represents the highest UI model accuracy ($TW = 2$) and the line in-between calculated the subject-averaged UI model accuracy. The peak group performance appeared at visual latency of 148 ms for the test data. These performance trends indicate a systematic performance lag (8 ms) between the two datasets.

The performance lag is presumably caused by temporal phase lags. Consequently, the cross-correlation between class-wise temporal templates of datasets I and II was computed. The averaged and class-wise cross-correlation is depicted in Figure 7B and C. The average cross-correlation also showed

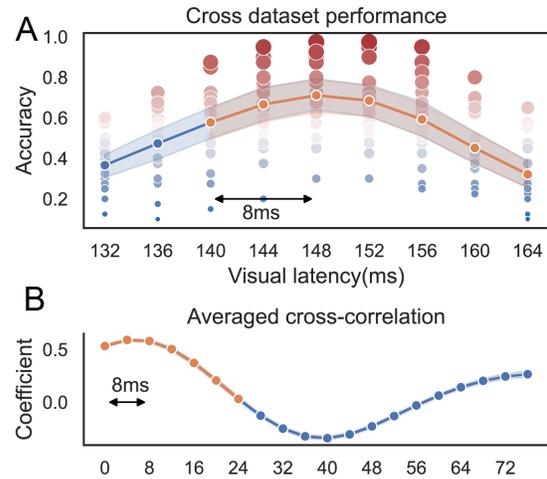


Fig. 7. Systematic phase lag causes performance delay: A. Inter-dataset prediction accuracy with the changes of visual latency. Each dot denotes the prediction accuracy of the TRCA model trained on dataset-I and tested on dataset-II. The line in-between is the average performance tendency of varying latency. B. Cross-correlation averaged over 40 targets between dataset-I and dataset-II.

an 8 ms phase lag, in line with the performance trend. Therefore, it is postulated that the performance delay results from the systematic temporal phase delay between the two datasets. The systematic phase delay is possibly due to various device configurations (graphic cards, screens, etc.). The existing domain adaptation method primarily focuses on the spatial filter domain, while only a part is about temporal template alignment. Temporal alignment may be the key to mapping the differences between untransferable data. Further studies need to be conducted to map the transferability from temporal characteristics.

D. Inspiration to the BCI Field

Improving model generalizability has been a universal challenge for all BCI studies. A growing amount of non-invasive BCI data is made public. The abundance of data can contribute to the development of a more generalized BCI system, but the enormous underlying variability remains a formidable obstacle. Many works have been devoted to methodologies like transfer learning techniques. While previous works aim to solve problems like ‘How to transfer?’, the present study adopts a data perspective to solve the problem as ‘Where to transfer?’. The solution suggests that only the transferable should be transferred. Those with more substantial underlying similarities were more capable of achieving satisfactory results.

When assessing transferability, previous research commonly employed a class-based distance measuring perspective [29]. Nevertheless, this method still requires labeled data and does not directly evaluate model performance. The assumption hardly stands in multi-class BCI cases like 40-target SSVEP in this study, because SSVEP typically shares distinct frequency preference characteristics, which means similar data in low-frequency stimulation does not necessarily lead to high similarity in other bands. Therefore, the likelihood testing

techniques presented here are a useful similarity measurement and can be applied in further BCI research.

V. CONCLUSION

This study proposed a Representative-based Cold start method that can deploy a calibration-free brain-computer interface (BCI) with sustained high performance. The representative model is trained on recommended transferable user independent data, enabling rapid and smooth convergence to the optimal user-dependent model. The proposed method can be a critical in building a generalized and calibration-free SSVEP BCI in real-life applications.

REFERENCES

- [1] X. Chen, Y. Wang, M. Nakanishi, X. Gao, T.-P. Jung, and S. Gao, "High-speed spelling with a noninvasive brain-computer interface," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 44, pp. E6058–E6067, Nov. 2015.
- [2] X. Chen, Y. Wang, S. Gao, T.-P. Jung, and X. Gao, "Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain-computer interface," *J. Neural Eng.*, vol. 12, no. 4, Aug. 2015, Art. no. 046008.
- [3] M. Nakanishi, Y. Wang, X. Chen, Y. Wang, X. Gao, and T.-P. Jung, "Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 1, pp. 104–112, Jan. 2018.
- [4] A. Ravi, N. H. Beni, J. Manuel, and N. Jiang, "Comparing user-dependent and user-independent training of CNN for SSVEP BCI," *J. Neural Eng.*, vol. 17, no. 2, Apr. 2020, Art. no. 026028.
- [5] P. Yuan, X. Chen, Y. Wang, X. Gao, and S. Gao, "Enhancing performances of SSVEP-based brain-computer interfaces via exploiting inter-subject information," *J. Neural Eng.*, vol. 12, no. 4, Aug. 2015, Art. no. 046006.
- [6] C. M. Wong et al., "Transferring subject-specific knowledge across stimulus frequencies in SSVEP-based BCIs," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 2, pp. 552–563, Apr. 2021.
- [7] C. M. Wong et al., "Inter-and intra-subject transfer reduces calibration effort for high-speed SSVEP-based BCIs," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 10, pp. 2123–2135, Oct. 2020.
- [8] B. Liu, X. Chen, X. Li, Y. Wang, X. Gao, and S. Gao, "Align and pool for EEG headset domain adaptation (ALPHA) to facilitate dry electrode based SSVEP-BCI," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 2, pp. 795–806, Feb. 2022.
- [9] K.-J. Chiang, C.-S. Wei, M. Nakanishi, and T.-P. Jung, "Boosting template-based SSVEP decoding by cross-domain transfer learning," *J. Neural Eng.*, vol. 18, no. 1, Feb. 2021, Art. no. 016002.
- [10] C. M. Wong et al., "Online adaptation boosts SSVEP-based BCI performance," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 6, pp. 2018–2028, Jun. 2022.
- [11] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *Nat. Sci. Rev.*, vol. 5, no. 1, pp. 44–53, Jan. 2018.
- [12] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowl. Syst.*, vol. 46, pp. 109–132, Jul. 2013.
- [13] J. Gope and S. K. Jain, "A survey on solving cold start problem in recommender systems," in *Proc. Int. Conf. Comput., Commun. Autom. (ICCCA)*, May 2017, pp. 133–138.
- [14] B. Lika, K. Kolomvatsos, and S. Hadjiefthymiades, "Facing the cold start problem in recommender systems," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 2065–2073, Mar. 2014.
- [15] L. Chen, F. Yuan, J. Yang, X. He, C. Li, and M. Yang, "User-specific adaptive fine-tuning for cross-domain recommendations," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 1, pp. 3239–3252, Mar. 2023.
- [16] H. Liu, Z. Hu, A. Mian, H. Tian, and X. Zhu, "A new user similarity model to improve the accuracy of collaborative filtering," *Knowl.-Based Syst.*, vol. 56, pp. 156–166, Jan. 2014.
- [17] C. M. Wong, B. Wang, Z. Wang, K. F. Lao, A. Rosa, and F. Wan, "Spatial filtering in SSVEP-based BCIs: Unified framework and new improvements," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 11, pp. 3057–3072, Nov. 2020.
- [18] Y. Chen, C. Yang, X. Chen, Y. Wang, and X. Gao, "A novel training-free recognition method for SSVEP-based BCIs using dynamic window strategy," *J. Neural Eng.*, vol. 18, no. 3, Jun. 2021, Art. no. 036007.
- [19] S. Nagel and M. Spüler, "Asynchronous non-invasive high-speed BCI speller with robust non-control state detection," *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, Jun. 2019.
- [20] C. Yang, X. Yan, Y. Wang, Y. Chen, H. Zhang, and X. Gao, "Spatio-temporal equalization multi-window algorithm for asynchronous SSVEP-based BCI," *J. Neural Eng.*, vol. 18, no. 4, Aug. 2021, Art. no. 0460b7.
- [21] Y. Wang, X. Chen, X. Gao, and S. Gao, "A benchmark dataset for SSVEP-based brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1746–1752, Oct. 2016.
- [22] J. Peirce et al., "PsychoPy2: Experiments in behavior made easy," *Behav. Res. Methods*, vol. 51, no. 1, pp. 195–203, Feb. 2019.
- [23] D. H. Brainard, "The psychophysics toolbox," *Spatial Vis.*, vol. 10, no. 4, pp. 433–436, 1997.
- [24] B. Liu, X. Chen, N. Shi, Y. Wang, S. Gao, and X. Gao, "Improving the performance of individually calibrated SSVEP-BCI by task-discriminant component analysis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1998–2007, 2021.
- [25] M. Kaya, M. K. Binli, E. Ozbay, H. Yanar, and Y. Mishchenko, "A large electroencephalographic motor imagery dataset for electroencephalographic brain computer interfaces," *Sci. Data*, vol. 5, no. 1, pp. 1–16, Oct. 2018.
- [26] B. Liu, X. Huang, Y. Wang, X. Chen, and X. Gao, "BETA: A large benchmark database toward SSVEP-BCI application," *Frontiers Neurosci.*, vol. 14, p. 627, Jun. 2020.
- [27] F. Zhu, L. Jiang, G. Dong, X. Gao, and Y. Wang, "An open dataset for wearable SSVEP-based brain-computer interfaces," *Sensors*, vol. 21, no. 4, p. 1256, Feb. 2021.
- [28] Y. Song, Q. Zheng, B. Liu, and X. Gao, "EEG conformer: Convolutional transformer for EEG decoding and visualization," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 710–719, 2023.
- [29] C.-S. Wei, Y.-P. Lin, Y.-T. Wang, C.-T. Lin, and T.-P. Jung, "A subject-transfer framework for obviating inter-and intra-subject variability in EEG-based drowsiness detection," *NeuroImage*, vol. 174, pp. 407–419, Jan. 2018.