ACCEPTED MANUSCRIPT • OPEN ACCESS

An SSVEP-based BCI with 112 targets using frequency spatial multiplexing

To cite this article before publication: Yaru Liu et al 2024 J. Neural Eng. in press https://doi.org/10.1088/1741-2552/ad4091

Manuscript version: Accepted Manuscript

Accepted Manuscript is "the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an 'Accepted Manuscript' watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors"

This Accepted Manuscript is © 2024 The Author(s). Published by IOP Publishing Ltd.

As the Version of Record of this article is going to be / has been published on a gold open access basis under a CC BY 4.0 licence, this Accepted Manuscript is available for reuse under a CC BY 4.0 licence immediately.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence <u>https://creativecommons.org/licences/by/4.0</u>

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected and is not published on a gold open access basis under a CC BY licence, unless that is specifically stated in the figure caption in the Version of Record.

View the article online for updates and enhancements.

An SSVEP-based BCI with 112 Targets Using Frequency Spatial Multiplexing

Yaru Liu¹, Wei Dai¹, Yadong Liu¹, Dewen Hu¹, Banghua Yang² and Zongtan Zhou^{1,*}

¹ College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410000, People's Republic of China

² Department of Automation, School of Mechatronics Engineering and Automation, V_{i} = L = $\int D_{i} = \int D_{i} + \int D_{i} = \int D_{i} = \int D_{i} + \int D_{i} = \int D_{i} = \int D_{i} + \int D_{i} = \int D_{i} = \int D_{i} + \int D_{i} = \int D_{i} =$

Key Laboratory of Power Station Automation Technology, Shanghai University,

Shanghai 200072, People's Republic of China

* Corresponding author

E-mail: narcz@nudt.edu.cn

November 2023

Abstract. Objective. Brain-computer interface (BCI) systems with large directly accessible instruction sets are one of the difficulties in BCI research. Research to achieve high target resolution (> 100) has not vet entered a rapid development stage, which contradicts the application requirements. Steady-state visual evoked potential (SSVEP) based BCIs have an advantage in terms of the number of targets, but the competitive mechanism between the target stimulus and its neighboring stimuli is a key challenge that prevents the target resolution from being improved significantly. Approach. In this paper, we reverse the competitive mechanism and propose a frequency spatial multiplexing method to produce more targets with limited frequencies. In the proposed paradigm, we replicated each flicker stimulus as a 2×2 matrix and arrange the matrices of all frequencies in a tiled fashion to form the interaction interface. With different arrangements, we designed and tested three example paradigms with different layouts. Further we designed a graph neural network that distinguishes between targets of the same frequency by recognizing the different electroencephalography (EEG) response distribution patterns evoked by each target and its neighboring targets. Main results. Extensive experiment studies employing eleven subjects have been performed to verify the validity of the proposed method. The average classification accuracies in the offline validation experiments for the three paradigms are 89.16%, 91.38%, and 87.90%, with information transfer rates (ITR) of 51.66, 53.96, and 50.55 bits/min, respectively. Significance. This study utilized the positional relationship between stimuli and did not circumvent the competing response problem. Therefore, other state-of-the-art methods focusing on enhancing the efficiency of SSVEP detection can be used as a basis for the present method to achieve very promising improvements.

Keywords: brain–computer interface (BCI), steady-state visual evoked potential (SSVEP), electroencephalogram (EEG), frequency spatial multiplexing, graph neural

networks (GNN)

Nomenclature

 $H_m, H_{\mathcal{N}(v_m^i)}, H_c$, Hidden layer features of nodes of $\mathcal{G}_m, \mathcal{N}(v_m^i)$, and \mathcal{G}_{c} .

 \mathbf{R}_k Reference template for the k-th stimulus frequency.

 $\boldsymbol{S}_{i_{sb}}$ i_{sb}-th subband filtered EEG signal.

 \boldsymbol{T}_m^{emb} Temporal embedding layer parameter in \mathcal{G}_m .

 \boldsymbol{W}_{α} Edge weights of attention aggregators.

 $X_m, X_{\mathcal{N}(v_m^i)}, X_c$, Features of $\mathcal{G}_m, \mathcal{N}(v_m^i)$ and \mathcal{G}_c .

 Y_c Predicted output of DSGAT.

 $\mathcal{G}_m, \mathcal{G}_c$ Intermediate and minimum graphs.

 $\mathcal{N}(v^i)$ The 1-step neighbors of node v^i

 $\tilde{\boldsymbol{W}}_{\alpha}$ Softmax normalized \boldsymbol{W}_{α} .

- $\varrho_k \qquad \text{FBCCA score at frequency } f_k.$
- c_{ij} *j*-th minimum cell in M_i .
- k Index of attention aggregator groups.
- l Index of DSGAT layers.

 L_{ij} L-shaped region of c_{ij} .

 M_i, M_{tg} *i*-th intermediate cell and target intermediate cell detected by FBCCA.

p Index of sub-networks.

1. Introduction

The researche on improving brain-computer interface (BCI) performance mainly takes two strategies: enhancing classification algorithms and designing more effective BCI paradigms. An effective paradigm can maintain more targets while eliciting sufficiently strong brain response activity. The design of a new paradigm seeks to build a larger instruction set to improve the efficiency of target selection, but the increase in the target number usually enhances classification difficulty. Therefore BCI systems need to make a compromise between the number of targets and classification performance [1, 2]. In terms of instruction set size, evoked potential-based BCI has an advantage over spontaneous BCI and thus is often applied to applications that require a large number of options, such as BCI-based spellers [3] or scenarios that require more refined intentions [4]. Steady-state visual evoked potential (SSVEP) based BCI usually has a higher information transfer rate (ITR) and larger instruction set than other BCI paradigms [5, 6]. When the human eye is stimulated by periodic flicker, the occipital region of the brain generates a modulated signal of the corresponding frequency, which produces energy enhancement at the frequency or the second or third harmonics [7].

/3

An SSVEP-based BCI with 112 Targets

The frequency range capable of evoking SSVEP components can be roughly divided into three intervals, the low-frequency band (6-12 Hz), the middle-frequency band (12-30 Hz), and the high-frequency band (30-60 Hz). The harmonic nature of SSVEP narrows the range of available frequencies for an SSVEP paradigm. On the other hand, there is a limit to the minimum difference between neighboring frequencies that can be discriminated, and the smaller the difference, the greater the difficulty posed to the classification algorithm. Therefore, the frequency band used for SSVEP-BCI is limited [8].

SSVEP paradigm innovations in recent years have focused on the design of mechanisms dedicated to generating more target options using limited stimulation frequencies. Hwang et al. proposed a dual-frequency stimulation method in which the black and white pattern in the checkerboard paradigm is set to flip at two different frequencies, thus enabling an SSVEP-BCI with 12 options using four frequencies [9]. Chen et al. used three luminance variation frequencies combined with two color variation modulation frequencies to evoke an intermodulation frequency response, achieving three frequencies to produce eight targets [10]. Yosuke et al. devised a new way of encoding stimuli using binary digits encoding of different frequency sequences to increase the number of visual stimuli with different characteristics [11]. Liang et al. proposed a new dual-frequency and phase modulation paradigm to optimize the combinations of two frequencies in the checkerboard-like dual-frequency paradigm [12]. Chen et al. implemented an SSVEP paradigm with 160 targets using the idea of multiple frequency sequential coding, in which a stimulus sequentially flickers at different frequencies [13].

To summarize these approaches, the main ideas for increasing the number of targets focus on 1) multiplexing stimulus frequencies, including temporal and spatial multiplexing, and 2) increasing the heterogeneous properties of same-frequency stimuli. Time-division multiplexing comes at the cost of increased time, In contrast, spatial multiplexing has the potential to further improve efficiency by being able to utilize different frequency information at one time and is, therefore, an approach worth exploring.

In addition, when a subject focuses on a target flicker, its neighboring stimuli also evoke SSVEP components. Therefore, previous studies typically increase the distance between adjacent stimuli and require subjects to reduce their attention to neighboring distractors [14, 15, 16, 17, 18]. In contrast, we reverse the use of the competition mechanism and propose a new method to produce more targets with limited frequencies.

This paper proposed a stimulus frequency spatial multiplexing method to design an SSVEP-BCI paradigm with 112 targets. The paradigm does not design for individual stimulus targets, but rather applies the location relationship between different stimuli to increase the attribute differences of targets with the same frequency. Specifically, different stimuli of the same frequency are in different locations, and the frequencies of the stimuli adjacent to them are different. This compound location relationship is used to encode targets, thus expanding 40 stimulus frequencies into 112 targets.

2. Methods

2.1. Frequency spatial multiplexing paradigm

2.1.1. Neurological principles SSVEP responses are usually strongest for the stimulus located in the center of the visual field, with an approximately Gaussian distribution gradually decaying outward [7]. Fuchs et al. investigated the competitive neuronal dynamics in cortical networks of early visual processing in the human brain utilizing two frequency stimuli. The results showed that during attention to the target stimulus, competing stimulus proximity (visual angle less than $4.5^{\circ} - 5^{\circ}$) leads to an increase in the amplitude of the competing frequency response accompanied by a significant enhancement of the intermodulation frequency, but the intensity of the target frequency response is not diminished because the attentional mechanism releases the inhibitory effect on the target stimulus [14]. This mechanism of integrated neuronal processing of the target and competing stimuli provides the theoretical basis for this study, i.e., by designing the visual stimulus layout to match targets of the same frequency with different competing neighbors, forming a neighbor encoding where all visual stimuli serve as targets and competing stimuli for each other, thus making full use of the available frequencies.

2.1.2. Paradigm Design A total of 40 frequencies were used, ranging from 8Hz-15.8Hz, in 0.2Hz steps. Figure 1 illustrates the basic elements of the paradigm interface and figure 2 presents three interface examples. The minimum cell is a rectangular singlefrequency flickering stimulus, denoted as c. Each cell is replicated as a 2×2 matrix, as an intermediate cell (noted as M), c_{ij} denotes the j-th corner in M_i . The 40 intermediate cells are arranged in a tiled array of 10×16 to form the stimulation interface. The four minimum cells in an intermediate cell flicker at the same frequency, but the intermediate cells adjacent to different c_{ij} are differentiated. The frequency response evoked by other intermediate cells in the vicinity of c_{ij} is also enhanced when c_{ij} is the target, so gazing at different minimum cells leads to different SSVEP frequency response distributions; thus, a minimum cell can be located according to the distribution patterns. In this paper, three example interfaces are designed:

- 1) Aligned arrangement (AA) paradigm, 40 intermediate cells in sequential order with ranks aligned (figure 2a);
- 2) Nonadjacent arrangement (NA) paradigm, adjust the order of intermediate cells in the AA paradigm so that spatially adjacent intermediate cells are distanced in the frequency domain (figure 2b);
- 3) Row stagger arrangement (SA) paradigm, on the basis of the AA paradigm, the intermediate cells are arranged in rows misaligned by one minimum cell, and the minimum cells at the end of the row that is out of the range of Map are filled to the vacancy at the beginning of the row (figure 2c).



Figure 2: Schematic diagram of the three paradigms interfaces. Stimulus frequencies and phases marked are not shown in the practical experiments

The interface contains 160 minimum stimulus cells, but the cells outside the red box are not sufficiently distinguishable due to the lack of neighbors, so they serve as competing stimuli to assist in locating the inner cells. Therefore, the actual effective divisible cells are the inner 8×14 region, i.e., 112 divisible targets (the red boxed area

An SSVEP-based BCI with 112 Targets

in figure 1). The visual stimuli were presented using luminance sinusoidal variation, and the phase within the intermediate cells was consistent, with a phase difference of 0.5π between adjacent intermediate cells.

2.2. Target fixation method with a regional attention

In this paper, we used a target fixation with a regional attention approach, which gives selective attention to the neighborhood stimuli while gazing at the target. Taking the AA paradigm as an example (figure 1 and figure 2a), when gazing at a target c_{ij} , the eight nearest neighbors around are the competing stimuli closest to the visual field center, which can theoretically evoke stronger SSVEP responses. Among these neighbors, the stimuli that can distinguish c_{ij} from the other three corners $(c_{iq}, q \neq j)$ is the L-shaped region L_{ij} in the figure. Therefore, the simultaneous regional attention method is to give selective attention to L_{ij} while gazing at c_{ij} . Due to a large number of targets in the paradigm, it is difficult to quickly locate c_{ij} in M_i and the corresponding L_{ij} , and the target cell may even be lost during the selection process. To assist in locating the target, the colors of the intermediate cells were set to alternate between white and yellow in the experiments, as shown in figure 2.

2.3. Target classification

The target classification process consists of two steps: 1) 40 classification: Calculate the response distribution of the 40 intermediate cells by SSVEP detection algorithm; 2) 112 classification: Identify the minimum cell target according to the 40 classification results.

2.3.1. SSVEP detection of intermediate cells The filter bank canonical correlation analysis (FBCCA) algorithm [19] was used. FBCCA is a classical and efficient classification algorithm in SSVEP-BCIs, which decomposes the EEG signal into several sub-band components by frequency bands, calculates the correlation coefficient of each sub-band separately using CCA, and then classifies the weighted features. The 12channel EEG signals are divided into subbands by bandpass filters, each with the same upper cutoff frequency (88 Hz) and a different lower cutoff frequency. For the i_{sb} -th subband, the lower cutoff frequency is $i_{sb} \times 8$ Hz, the i_{sb} -th subband filtered EEG signal is \mathbf{S}_{i_b} . The sinusoidal reference template $\mathbf{R}_k \in \mathbb{R}^{2N_h \times N_s}$ for the k-th stimulus frequency is:

$$\boldsymbol{R}_{k} = \begin{bmatrix} \sin(2\pi f_{k}t) \\ \cos(2\pi f_{k}t) \\ \sin(2*2\pi f_{k}t) \\ \cos(2*2\pi f_{k}t) \\ \dots \\ \sin(N_{h}*2\pi f_{k}t) \\ \cos(N_{h}*2\pi f_{k}t) \end{bmatrix}, \quad t = \frac{1}{F_{s}}, \frac{2}{F_{s}}, \dots, \frac{N_{s}}{F_{s}}, \quad (1)$$

where f_k is the k-th frequency, N_h is the number of harmonics, N_s is the signal length, t is the current time point, and F_s is the sampling rate.

CCA finds a set of weight vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to maximize the correlation between $\boldsymbol{\alpha}^T \boldsymbol{S}_{i_{sb}}$ and $\boldsymbol{\beta}^T \boldsymbol{R}_k$:

$$\rho(\mathbf{S}_{i_{sb}}, \mathbf{R}_{k}) = \max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \frac{E[\boldsymbol{\alpha}^{T} \mathbf{S}_{i_{sb}} \mathbf{R}_{k}^{T} \boldsymbol{\beta}]}{\sqrt{E[\boldsymbol{\alpha}^{T} \mathbf{S}_{i_{sb}} \mathbf{S}_{i_{sb}}^{T} \boldsymbol{\alpha}] E[\boldsymbol{\beta}^{T} \mathbf{R}_{k} \mathbf{R}_{k}^{T} \boldsymbol{\beta}]}}$$

$$i_{sb} = 1, 2, \dots N_{sb} \quad k = 1, 2, \dots N_{f},$$
(2)

where N_{sb} is the number of filter banks and N_f is the number of stimulus frequencies.

The weighted sum of the squared correlation coefficients of all subbands is used as the correlation characteristic with the k-th frequency, and the weight of each subband is

$$w_{i_{sb}} = i_{sb}{}^{-a} + b, \tag{3}$$

where the optimal values of a and b have been optimized in the study [19] and set to 1.25 and 0.25, respectively. The response score of an EEG sample at frequency f_k is

$$\varrho_k = \sum_{i_{sb}=1}^{N_{sb}} w_{i_{sb}} \cdot \rho(\boldsymbol{S}_{i_{sb}}, \boldsymbol{R}_k)^2.$$
(4)

2.3.2. Minimum target classification algorithm We used a graph neural network (GNN) to classify the intermediate cell response results further. GNN abstracts irregular data as nodes, and each node establishes a relationship with its neighbor nodes through edges; thus GNNs can capture the complex structural relationships in graphs [20, 21, 22]. Graph Convolutional Network (GCN) is a representative GNN, especially the spatial-based GCN borrows the idea of the convolutional neural network to define graph convolution based on the spatial relationship of nodes, which is essentially a process of iteratively aggregating neighborhood information. The key to the minimum target classification of the proposed paradigm is the SSVEP response distributions of competing neighbors. Therefore, the properties of GCN neighbor aggregation fit well with the paradigm design. We designed a dual-scale graph attention network (DSGAT) for the global 112 classification problem.

Graph definition DSGAT involves two scales of graphs with two types of stimulus cells in the paradigm as graph nodes, respectively.

(i) Define intermediate graph $\mathcal{G}_m = (\mathcal{V}_m, \mathcal{E}_m)$, as shown in Figure 3a, and consider each intermediate cell as a node v_m (coarse-grained node, later referred to intermediate node). \mathcal{V}_m is the set of intermediate nodes, $|\mathcal{V}_m| = N_m$ is the number of intermediate nodes, and \mathcal{E}_m denotes the set of intermediate node edges. The SSVEP frequency responses of the intermediate cells are used as the initial node feature set of \mathcal{G}_m , which is denoted as $\mathbf{X}_m \in \mathbb{R}^{N_m \times N_{T_m}}$, and N_{T_m} is the number of feature channels.



Figure 3: Local diagram of \mathcal{G}_m and \mathcal{G}_c . The circles denote nodes, squares denote the minimum cells in the paradigm, and 2×2 squares denote intermediate cells. 3a is the local structure of \mathcal{G}_m , including a central intermediate node and its first-order neighbors. 3b is the local structure of \mathcal{G}_c , including a central minimum node and its first-order neighbors

The distributions of SSVEP frequency responses change over time, so SSVEP detection is performed multiple times for the duration of the stimulation. Therefore, the feature vector of the *i*-th intermediate node v_m^i denoted as $\mathbf{X}_m^i \in \mathbb{R}^{N_{T_m}}$ is a sequence of FBCCA detection scores throughout the stimulation:

$$\boldsymbol{X}_{m}^{i} = [\varrho_{i}(1), \varrho_{i}(2), \dots \varrho_{i}(N_{T_{m}})].$$
(5)

The feature vectors of all intermediate nodes form X_m . Thus the number of feature channels N_{T_m} of X_m is the number of SSVEP detection times during the whole stimulus process.

(ii) Define minimum graph $\mathcal{G}_c = (\mathcal{V}_c, \mathcal{E}_c)$, as shown in Figure 3b, with each minimum cell as node v_c (fine-grained node, later referred to minimum node), \mathcal{V}_c is the set of minimum nodes, $|\mathcal{V}_c| = N_c$. \mathcal{E}_c denotes the set of minimum node edges. The feature set of \mathcal{G}_c is denoted by $\mathbf{X}_c \in \mathbb{R}^{N_c \times N_{T_c}}$.

The first-order neighbors of node v^i are denoted as $\mathcal{N}(v^i)$, and the feature matrix of $\mathcal{N}(v^i)$ is denoted as $\mathbf{X}_{\mathcal{N}(v^i)}$. There is a corresponding relationship between V_m and V_c , consistent with the inclusive relationship of the two types of cells in the paradigm.

Structure of DSGAT model Figure 4 illustrates the architecture of DSGAT, consisting of an aggregation module, a fully connected layer, and a softmax output layer, where the aggregation module consists of a temporal embedding layer and a multi-head graph attention layer. The core idea of DSGAT is summarized in three points.



Figure 4: Structure of DSGAT model. The operations with the same color in the same sub-network in the figure have the same parameters

- (i) Operations on the graph are performed sequentially at \mathcal{G}_m and \mathcal{G}_c , so that feature information flows from coarse-grained nodes to fine-grained nodes, corresponding to the refinement process of SSVEP responses to global target localization.
- (ii) The relationship between the distributions of SSVEP responses to different competing stimuli and target localization is captured using a multi-head graph attention mechanism to cope with the contradiction between repetition and heterogeneity in the paradigm.
- (iii) A temporal embedding layer is added to handle the imbalance of SSVEP responses in the time dimension.

The graph structures of the three example interfaces are not identical, and the network structures here are exemplified by the AA paradigm (figure 2a).

The network input is the SSVEP detection results at all frequencies, and the input features are first subjected to a neighbor aggregation operation on \mathcal{G}_m . For the AA paradigm, the 1-step neighbors of an intermediate node are its 8-neighborhood intermediate cells as well as itself. Unlike \mathcal{G}_m , the nodes in \mathcal{G}_c have an additional attribute: the location detection of the minimum cell in the intermediate cell, and thus there are four location attributes. Therefore, X_m is fed into four sub-networks with the same structure but different parameters to distinguish four minimum nodes. The output of the p-th (p = 1, 2, 3, 4) sub-network is passed to the p-th group of minimum nodes, thus allowing the information to flow from the intermediate nodes to the more numerous minimum nodes.

The SSVEP detection results by FBCCA algorithm have fluctuating amplitudes over time, and the stimulus duration affects the accuracy of SSVEP detection. Therefore, a temporal embedding layer is designed before aggregation, which creates





Figure 5: Schematic diagram of the aggregation operation of \mathcal{G}_m . The edges in four colors and line types indicate the aggregation operations of four sub-networks

a learnable temporal embedding matrix $T_m^{emb} \in \mathbb{R}^{|\mathcal{N}(v_m^i)| \times N_{T_m}}$ to compensate for this temporal disequilibrium. For the *p*-th sub-network, the embedded feature matrix of node v_m^i and its 1-step neighbor nodes is

$$\boldsymbol{H}_{\mathcal{N}(v_m^i)}(p)^{emb} = \sigma_{ReLU}(\boldsymbol{X}_{\mathcal{N}(v_m^i)}(p) \circ \boldsymbol{T}_m^{emb}(p)),$$
(6)

where $X_{\mathcal{N}(v_m^i)} \in \mathbb{R}^{|\mathcal{N}(v_m^i)| \times N_{T_m}}$ is the feature matrix of $\mathcal{N}(v_m^i)$, \circ is the Hadamard product operation, σ_{ReLU} is the ReLU activation function.

The embedded feature signals are then neighborhood aggregated through the graph attention layer. There are two types of relationships between the intermediate nodes v_m^i and $v_m^j \in \mathcal{N}(v_m^i)$: frequency relationship and location relationship. When gazing at a target in SSVEP-BCI, the responses of its competing stimuli are also enhanced in both the spatial and frequency domains. For v_m^i , $v_m^j \in \mathcal{N}(v_m^i)$ have a regular arrangement in space and frequency, with left-right, up-down, and diagonal neighbors symmetric to each other and have equal frequency differences to the target stimulus.

An intermediate node is related to four minimum nodes, and when a minimum node is gazed at, the SSVEP responses of neighboring intermediate nodes in the direction of other minimum nodes are also affected; thus, the target localization needs to consider all 1-step neighbors of the intermediate node. The influence of neighboring intermediate nodes varies from different minimum nodes so the attention mechanism is applied to learn this difference adaptively, with different weights on the edges.

The global localization from 40 frequency responses to 112 targets is essentially a 4classification problem. The four feature distributions of the intermediate nodes and their 1-step neighbors determine the classification results of the four minimum nodes. We used a multi-headed graph attention layer to learn the four distribution patterns (unlike the standard graph attention layer, DSGAT restricts the attention calculation to 1-step neighbors, and the aggregation is also varied). Each sub-network uses K_a independent attention mechanisms. The schematic diagram of the aggregation operation on \mathcal{G}_m is shown in figure 5.

Suppose the input feature vector of the central node v_m^i at *l*-th layer is $\boldsymbol{H}_m^i(l)$ and

the weight coefficient of the neighbor node v_m^j to v_m^i is α_{ij} , then the parameters of layer l are

$$\boldsymbol{W}_{\alpha}(p)(l) = [\alpha_{i1}(p)(l), \alpha_{i2}(p)(l), \dots, \alpha_{i|\mathcal{N}(v_m^i)|}(p)(l)]$$

For the k-th $(k = 1, ..., K_a)$ group of attention aggregators, the edge weights are softmax normalized, and the new coefficients are

$$\tilde{\boldsymbol{W}}_{\alpha}(p)(l)(k) = softmax(\boldsymbol{W}_{\alpha}(p)(l)(k)).$$
(8)

7)

The new feature vector of node v_m^i is obtained by weighted summation of neighbors:

$$\boldsymbol{H}_{m}^{i}(p)(l+1)(k) = \sigma_{LReLU} \left(\sum_{\boldsymbol{v}_{m}^{j} \in \mathcal{N}(\boldsymbol{v}_{m}^{i})} \tilde{\alpha}_{ij}(p)(l)(k) \boldsymbol{H}_{m}^{j}(p)(l) \right),$$
(9)

 σ_{LReLU} is the LeakyReLU activation function. The K_a group outputs are combined by the concatenation operation into an aggregated feature vector of the *p*-th group of minimum nodes:

$$\boldsymbol{H}_{c}^{agg}(p)(l+1) = \prod_{k=1}^{K_{a}} \boldsymbol{H}_{m}(p)(l+1)(k),$$
(10)

where || is the concatenation operation. To avoid over-smooth, the new feature vector of minimum nodes are obtained by skip concatenating the pre-aggregation and aggregation feature vectors:

$$\boldsymbol{H}_{c}(p)(l+1) = \boldsymbol{H}_{m}(p)(l) || \boldsymbol{H}_{c}^{agg}(p)(l+1)$$
(11)

After the information is propagated from the intermediate nodes to the minimum nodes, the subsequent computation is performed on \mathcal{G}_c . The features of the aggregation operation are fed to a fully connected layer and a final softmax output layer. The final predicted output is:

$$\hat{\boldsymbol{Y}}_{c} = softmax\left(\sigma_{ReLU}(\boldsymbol{H}_{c}\boldsymbol{W}_{f} + \boldsymbol{b}_{f})\right), \qquad (12)$$

where $\boldsymbol{H}_c \in \mathbb{R}^{|\mathcal{N}(V_c)| \times N'_{T_c}}$ is the feature matrix of \mathcal{G}_c , N'_{T_c} is the number of feature channels after aggregation, and \boldsymbol{W}_f and \boldsymbol{b}_f are the fully connected layer parameters.

The aggregation operation in DSGAT is consistent for all nodes, but the nodes located at the outermost circle of the paradigm lack sufficient neighbors. For this case, borrowing from the padding method in convolutional neural networks, virtual nodes are constructed to fill the missing neighbors. Suppose a virtual node used for padding is v^{pad} , and the initial feature vector is $\boldsymbol{X}_m^{pad} \in \mathbb{R}^{N_{T_m}}$. The *t*-th value of \boldsymbol{X}_m^{pad} is the minimum value of the *t*-th channel of \boldsymbol{X}_m .

3. Experiments

3.1. Subjects

Eleven healthy subjects participated in the experiment, including six males, aged 23-32 years old, mean aged of 26.6 years, and all participants had normal or corrected normal



vision. Two subjects had experience with SSVEP-BCI experiments, and the others did not. This study was performed in accordance with the Declaration of Helsinki. This human study was approved by The Ethics Committee of the Xiangya Hospital of Central South University - approval: 2021111249. All adult participants provided written informed consent to participate in this study.

3.2. Data acquisition

EEG signals acquisition utilized a BrainAmp Amplifier (Brain Products GmbH, Germany) through 12 electrodes placed at Oz, O1, O2, POz, PO3, PO4, PO7, PO8, P1, P2, P5, and P6, referenced to TP10 and grounded to Fpz, according to the International 10/20 system. The signal sampling rate is 500 Hz. Before data recording, the electrode impedances were kept below $20k\Omega$.

3.3. Experiment process

The visual paradigm was presented on a 27 inch LCD monitor with a 240 Hz refresh rate and a $1920px \times 1080px$ resolution. Frequency stimulus presentation was implemented using the PsychoPy toolkit [23], with 40 frequencies set to 8-15.8 Hz at 0.2 Hz intervals. The experimental procedure was controlled using the BCI2000 platform [24]. Subjects seated at a distance of 60cm from the monitor, with the horizontal viewpoint at the monitor center. The single minimum cell spanned 2.81° horizontally and 2.53° vertically and was separated from adjacent cells by approximately 0.65° and 0.59° in horizontal

An SSVEP-based BCI with 112 Targets

and vertical viewing angles, satisfying the competing stimulus condition.

Each subject performed at least one training experiment before the formal experiments to understand the experimental process and adapt to the new paradigm. In the formal experiments, each subject performed three sessions of experiments corresponding to the three paradigms. Each session consists of two groups, with 112 trials in each group, i.e., one data collection for each target in the effective area. 112 trials were divided into 12 blocks, each block containing ten trials (the last group contained two trials). The rest period between groups was determined by the subjects themselves according to their status. Each trial consisted of a visual cue and a stimuli flicker phase. In the target cue phase, all stimuli were presented at maximum luminance, and the target to be selected was covered by a red square for 2 s. During the flicker phase, all stimuli started to flicker for 6.5 s. The target cue turned from a solid square to a red wireframe during the stimulus phase. To improve the subjects' attention level and their sense of control over the experiment, we used a real-time feedback mechanism. Since the global classification parameters had not been trained yet, and the FBCCA algorithm for SSVEP detection is an unsupervised method, the current detected 40 classification results were fed back in real-time during the flicker phase. SSVEP detection is performed every 0.2 s. The intermediate cell corresponding to the maximum value of each FBCCA score is marked with a red dot. To avoid the effect of experimental sequence, each subject performed a session within one day, and the order of trials was arranged in a randomized form. Before the formal experiment, subjects were asked to perform several practice trials to familiarize themselves with the paradigm, the gaze method, and the experimental process. During the cueing phase in the practice experiment, the target minimum cell was marked by a red square and its adjacent L-shaped area was framed by a red wireframe, indicating the area to which the subject needed to pay attention simultaneously while gazing at the target. The target cue turned to a red box during the flicker phase, and the attention cue disappeared.

For DSGAT training, the data collected for one trial is one sample, and the data of a subject is divided into training and test sets in the ratio of 8:2. The network model uses two attention operations ($K_a = 2$) for each sub-network and contains a 1-layer aggregation module. The model is trained with an Adam optimization algorithm to minimize the cross-entropy on the training data for 100 epochs. The initial learning rate is 0.1 with a decay rate of 0.7 after every 20 epochs. The batch size is 128.

In addition, we used two baselines for comparison with DSGAT, using global classification accuracy as the evaluation metric.

baseline 1 An unsupervised method to locate the minimum cell in the intermediate cell based on the SSVEP detection results. For an EEG sample, the intermediate cell detected by FBCCA at t is

$$M_{tg}(t) = \underset{k \in [1, \dots, N_m]}{\arg \max} \varrho_k(t).$$
(13)

An SSVEP-based BCI with 112 Targets

For c_{ij} in M_{tg} , the sum of the SSVEP response scores of c_{ij} and its L-shape region neighbors at moment t is taken as the new feature value of c_{ij} . The minimum cell with the highest number of maximum feature values over a period is taken as the final global target.

baseline 2 Logistic Regression (LR) algorithm is used to classify the four minimum cells in M_{tg} . The SSVEP score vectors of M_{tg} and its 1-step neighbors in \mathcal{G}_m are concatenated as the feature vector $\mathbf{X}_{\mathcal{N}(M_{tg})}{}^{LR} \in \mathbb{R}^{|\mathcal{N}(M_{tg})| \times N_{T_m}}$, and the sequence number of the minimum cell in the intermediate cell is used as the 4-classification label. The training data were screened before training because wrong M_{tg} localization could interfere with the LR algorithm training. A data sample is considered valid if the intermediate cell containing the true target gets the highest FBCCA score for the most times in a period of time, and conversely, the sample is rejected in training.

4. Results

4.1. Qualitative relsults of SSVEP responses

We first verified whether the frequency response components of the target and competing stimuli could be evoked simultaneously in the EEG signal by time-frequency analysis.

Figure 6b shows example spectra of four data samples using Fast Fourier Transform (FFT) on a typical subject (S1) under the AA paradigm (figure 2a). The spectrum in each subplot is the average of all channels of the EEG signal. The targets of the four samples belong to the same intermediate cell with a stimulus frequency of 11.6 Hz, noted as $c_{19,1}$, $c_{19,2}$, $c_{19,3}$ and $c_{19,4}$ (11.6 Hz is the 19th frequency), and the mean spectra of all signal channels for each data sample is shown in the figure.

For ease of observation, the paradigm is simplified to a schematic diagram with only the intermediate cells, as shown in figure 6a. It is clearly observed from every subplot that the 11.6 Hz peak is evoked, while amplitude peaks at competing and harmonic frequencies are also observed. For $c_{19,1}$, the spectra show amplitude peaks at the target stimulus frequency 11.6 Hz, the 2nd harmonic frequency of 23.2 Hz, and the competing stimulus frequency of 9.8 Hz. The 9.8 Hz stimuli located in the neighboring L-shaped region $L_{19,1}$, whereas the amplitudes of competing frequencies farther away do not show a significant increase. Similarly, for the target $c_{19,2}$, an amplitude peak occurs at 10.4 Hz in the $L_{19,2}$ region, while the neighboring competing stimulus responses for $c_{19,3}$ and $c_{19,4}$ show peaks at 13.2 Hz and 13.4 Hz, respectively.

In addition, the spectra of $c_{19,1}$ and $c_{19,2}$ also show amplitude peaks at frequencies such as 20 Hz ($c_{19,1}$) and 19 Hz ($c_{19,2}$). These peaks may be the intermodulation and harmonic components of competing stimulus frequencies. The 20 Hz peak in $c_{19,1}$ may be the 2nd harmonic of the 10 Hz stimulus, or intermodulations of different frequencies, such as the sum of 10.2 Hz and 9.8 Hz, or the sum of 8.6 Hz and 11.4 Hz. The 19 Hz peak in $c_{19,2}$ may be caused by the complex intermodulations, such as the sum of 8.8



Hz and 10.2 Hz.

These results suggest that the competing response enhancement has different distributions when different minimum cells in the same intermediate cell are targeted, which initially verifies the validity of the proposed paradigm. The enhancement is not limited to 1-step neighbors; the 2-step neighbors in this example also show different degrees of amplitude increase, such as 9.6 Hz and 10.4 Hz. This phenomenon may be caused by a combination of frequency, spatial distance, and subjects' selective attention. Further, we use the FBCCA scores to analyze the time-frequency pattern, still using the four examples above, and the results are shown in figure 7. The upper panels

An SSVEP-based BCI with 112 Targets

show the FBCCA scores of each frequency at different time points. The scores for all frequencies at the beginning phase are higher than those afterward (overfitting due to the short signal length [25]), resulting in unintuitive visualization of the stabilization phase later. Thus the figure shows the normalization results of all frequency scores at each moment. we randomly selected the FBCCA detection results at 4.6 s as an example for visualization, as shown in the lower panels, which display the scores according to the arrangement of intermediate cells.

It can be seen that all plots have the strongest response at 11.6 Hz, and the SSVEP responses are gradually ordered from about 1 second onwards. Two frequencies, 11.4 Hz and 11.8 Hz, also show higher response enhancement than the other frequencies because they are closest to the target 11.6 Hz in both frequency and spatial distances. This dual effect of space and frequency is not limited to 11.4 Hz and 11.8 Hz, as the 2-step neighbors with a 0.4 Hz frequency difference also have different degrees of response enhancement. In addition, the neighbors located on the left and right sides are slightly different from each other, and intuitively the response on the side of the minimum cell is slightly higher than the other side. Other competing stimuli also have response enhancement, and the degree of enhancement and the spacial distance to the target minimum cell showed correlations, such as 9.8 Hz, 10 Hz, 10.2 Hz for $c_{19,1}$ and $c_{19,2}$, and 13.2 Hz, 13.4 Hz for $c_{19,3}$ and $c_{19,4}$. These competing responses lag slightly compared to the target frequency and fluctuate over time, which is related to the fixation method with regional attention, with selective attentional wandering leading to instability of competing neighbor responses.

To summarize the time-frequency analysis, the target frequency evokes the strongest SSVEP response, and the 1-step neighboring intermediate cells could elicit higher responses than other stimuli, especially in the direction of the minimum target cell. Enhanced responses are also observed for 2-step neighbor stimuli, but are generally lower than for 1-step neighbors. This subsection is just a qualitative analysis of the SSVEP response of the intermediate cells through an example to visualize the feasibility of the method. The performance of the proposed method needs to be statistically analyzed further.

4.2. Statistical results of SSVEP detection

Figure 8 shows the intermediate cell classification accuracy-data length curves for each subject, with accuracy calculated every 0.2 s for 224 trials in each paradigm. This metric reflects the performance of the first step of signal processing. A higher accuracy result only represents a more accurate identification of the intermediate cell containing the target, and does not reflect the responses to the competing frequencies. Overall, for three paradigms, the accuracy tends to increase with data length. In the beginning, the accuracy increases rapidly and basically stabilizes.

In terms of accuracy, most of the subjects were able to achieve high accuracy, with some subjects achieving a maximum accuracy close to 100%. There are also individual



Figure 7: Time-frequency analysis of a typical subject (S1) using FBCCA results





Figure 8: Accuracy-data length curves for the classification of 40 intermediate cells per subject. Red, blue and green curves represent AA, NA and SA paradigms, respectively.

subjects, such as S8, with an average classification accuracy of only 67.97% for the 40 intermediate cells. In terms of response speed, there were individual differences between subjects, but basically they all stabilized from about 2 s. Thus reliable SSVEP detection results were obtained most of the time throughout the stimulation, providing a data base for subsequent global classification. It can also be observed in the resuls of some subjects that the accuracy shows a slight decrease in the later stage. This phenomenon was particularly obvious for S6, which shows a decreasing trend after the accuracy reached the highest point. The reason may be that, on the one hand, the proximity of the stimuli and their mutual influence make detection difficult, but the main reason should be the subjects' selective attention operation, which indicates that the method of accompanying attentional fixation puts requirements on the subjects. It is potentially possible for the subjects to further enhance the efficiency of the method in this study by training and mastering the skill.

In terms of the of variability in the performance across the three paradigms, we record the statistical accuracy for each subject in Table 1 (SSVEP columns). The



Figure 9: Average accuracy histogram of the three paradigms using different algorithms

statistical accuracy was calculated as follows: for each 6.5-second sample, the sequence of SSVEP detection results was counted (every 0.2 s), and if the frequency with the highest number of occurrences was consistent with the ground-truth frequency, the sample result was considered correct. The 11 subjects' average accuracies for AA, NA, and SA are 95.01±7.77%, 94.30±9.34%, and 90.80±10.67%, respectively. One-way repeated measures analysis of variance (RANOVA) was used to test the difference in the classification performances across paradigms, with a Greenhouse–Geisser correction and statistical significance defined as p < 0.05. The one-way RANOVA shows that there is a statistically significant difference in accuracies among the three paradigms [F(2,20)=4.397, p=0.047]. In the AA and NA paradigm, there are eight competing stimuli around an intermediate cell, and in the SA paradigm, the number is six. We expected that the reducing competing stimuli would promote the efficiency of SSVEP detection, but the experimental results were contrary to our expectations. We also note that the accuracy of S9 in the SA paradigm has a gap with AA and NA paradigms. The reason may be that the row stagger design of the SA interfered with the subjects' observations, thus affecting the SSVEP detection.

4.3. Minimum target classification results

Table 1 records the minimum target classification accuracy of each subject applying three paradigms and using different algorithms. The minimum target classification accuracy for each subject is a 5-fold cross-validation average. The table shows the SSVEP detection accuracy, the accuracy results of the DSGAT method, two baselines, and the DSGAT without temporal embedding (w/o T_m^{emb}). Figure 9 shows the average accuracy histogram of the three paradigms using different algorithms.

It can be seen that the DSGAT approach achieves higher accuracy in three paradigms (AA: DSGAT vs BS1, p < 0.001, DSGAT vs BS2, p < 0.01; NA: DSGAT vs BS1, p < 0.001, DSGAT vs BS2, p < 0.05; SA: DSGAT vs BS1, p < 0.001, DSGAT vs BS2, p < 0.05; SA: DSGAT vs BS1, p < 0.001, DSGAT vs BS2, p < 0.05; SA: DSGAT vs BS1, p < 0.001, DSGAT vs BS2, p < 0.05; SA: DSGAT vs BS1, p < 0.001, DSGAT vs BS2, p < 0.05; SA: DSGAT vs BS1, p < 0.001, DSGAT vs BS2, p < 0.05; SA: DSGAT vs BS1, p < 0.001, DSGAT vs BS2, p < 0.05; SA: DSGAT vs BS1, p < 0.001, DSGAT vs BS2, p < 0.05; SA: DSGAT vs BS1, p < 0.001, DSGAT vs BS1, p < 0.001, DSGAT vs BS2, p < 0.05; SA: DSGAT vs BS1, p < 0.001, DSGAT vs B

Page 20 of 28

An SSVEP-based BCI with 112 Targets

						Ť			· · ·	· · ·					
Subject	AA paradigm				NA paradigm				SA paradigm						
5	SSVEP	BS1	BS2	DSGAT	DSGAT*	SSVEP	BS1	BS2	DSGAT	DSGAT^*	SSVEP	BS1	BS2	DSGAT	DSGAT
S1	99.11	$\substack{69.13\\\pm8.61}$	$\substack{81.74\\\pm5.87}$	$\substack{\textbf{84.35}\\\pm 3.23}$	$\begin{array}{c c} 80.43 \\ \pm 2.66 \end{array}$	97.32	$\begin{array}{c} 80.00 \\ \pm 6.03 \end{array}$	$\begin{array}{c} 86.52 \\ \pm 7.28 \end{array}$	93.91 ±2.38	89.57 ± 5.19	93.75	$\begin{array}{c} 74.78 \\ \pm 7.93 \end{array}$	77.39 ± 3.95	85.65 ±5.67	$\begin{array}{c} 80.00 \\ \pm 5.63 \end{array}$
S2	89.29	54.35 ± 10.54	$\begin{array}{c} 77.83 \\ \pm 9.14 \end{array}$	90.00 ±3.30	87.39 ±5.63	88.39	65.65 ± 10.24	$\begin{array}{c} 75.65 \\ \pm 4.46 \end{array}$	$\begin{array}{c} \textbf{93.48} \\ \pm 2.67 \end{array}$	91.74 ± 3.57	89.29	$\substack{63.91\\\pm4.51}$	$\begin{array}{c} 80.87 \\ \pm 4.71 \end{array}$	87.39 ±3.89	$84.08 \\ \pm 3.94$
S3	93.75	$\begin{array}{c} 70.43 \\ \pm 9.17 \end{array}$	$\begin{array}{c} 79.13 \\ \pm 6.26 \end{array}$	$\substack{88.26\\\pm4.51}$	88.69 ±5.63	99.11	76.96 ± 9.42	$\substack{82.17\\\pm5.63}$	$\frac{86.52}{\pm 3.89}$	86.96 ±1.54	94.64	66.52 ± 9.55	78.26 ± 4.07	84.35 ±4.96	83.05 ± 4.46
S4	98.20	$^{86.96}_{\pm 5.54}$	$^{89.56}_{\pm 6.41}$	96.52 ±4.77	$94.35 \\ \pm 3.95$	99.11	$^{87.39}_{\pm 5.83}$	92.17 ± 3.30	96.52 ±1.19	$95.22 \\ \pm 0.97$	100.00	74.78 ± 7.93	$\begin{array}{c} 79.56 \\ \pm 4.24 \end{array}$	83.48 ±3.95	$ 80.43 \\ \pm 3.44 $
S5	99.11	$\begin{array}{c} 66.96 \\ \pm 7.90 \end{array}$	$\substack{69.13\\\pm6.41}$	$\begin{array}{c} \textbf{75.12} \\ \pm 4.82 \end{array}$	$\begin{array}{c c} 70.43 \\ \pm 5.87 \end{array}$	98.21	$57.83 \\ \pm 5.67$	$\substack{63.04\\\pm5.10}$	68.26 ±3.64	68.26 ± 3.95	99.11	$\begin{array}{c} 55.56 \\ \pm 5.88 \end{array}$	$68.70 \\ \pm 8.36$	76.52 ±7.90	70.00 ± 8.47
S6	98.21	$\begin{array}{c} 70.00 \\ \pm 7.11 \end{array}$	$\begin{array}{c} 86.52 \\ \pm 4.45 \end{array}$	92.17 ±5.87	$ \begin{array}{c c} 86.52 \\ \pm 5.83 \end{array} $	99.11		$\begin{array}{c} 87.83 \\ \pm 8.64 \end{array}$	93.91 ±4.18	$87.39 \\ \pm 8.19$	91.96	$76.89 \\ \pm 5.76$	84.78 ± 2.67	94.35 ±3.30	92.17 ± 3.95
S7	99.13	$^{74.35}_{\pm 6.77}$	$\substack{84.35\\\pm7.11}$	$90.43 \\ \pm 6.80$	90.87 ±5.83	98.70	73.48 ± 7.11	$\begin{array}{c} 88.26 \\ \pm 5.00 \end{array}$	96.09 ±2.38	$\begin{array}{c c}93.04\\\pm4.46\end{array}$	93.04	64.78 ± 9.40	$^{79.57}_{\pm 5.67}$	90.87 ±6.03	89.57 ± 4.18
S8	73.48	$^{44.35}_{\pm 6.26}$	$\begin{array}{c} 63.91 \\ \pm 7.47 \end{array}$	90.00 ±1.94	$ \begin{array}{c c} 89.13 \\ \pm 2.66 \end{array} $	67.83	$40.00 \\ \pm 9.30$	55.65 ± 5.67	92.17 ±2.48	$ \begin{array}{c} 84.35 \\ \pm 1.82 \end{array} $	62.61	$29.13 \\ \pm 9.04$	56.96 ± 2.83	$\begin{array}{c} \textbf{93.91} \\ \pm 5.19 \end{array}$	$90.43 \\ \pm 4.24$
S9	99.57	$\begin{array}{c} 74.35 \\ \pm 9.02 \end{array}$	$^{86.96}_{\pm 2.66}$	92.17 ±5.46	87.83 ±3.30	98.70	70.87 ± 6.98	$\begin{array}{c} 78.70 \\ \pm 4.18 \end{array}$	88.70 ±3.89	$ \begin{array}{c c} 84.35 \\ \pm 4.18 \end{array} $	81.74	$55.22 \\ \pm 5.24$	72.17 ± 3.89	90.87 ±4.18	83.48 ± 2.92
S10	97.83	$\begin{array}{c} 70.00 \\ \pm 7.90 \end{array}$	$\begin{array}{c} 78.26 \\ \pm 4.07 \end{array}$	86.96 ±4.35	83.48 ±7.14	94.78	79.13 ± 6.63	$\begin{array}{c} 88.70 \\ \pm 4.18 \end{array}$	96.52 ±3.30	$91.30 \\ \pm 2.17$	99.13	66.96 ± 12.05	73.48 ± 11.02	83.04 ±3.57	
S11	97.39		$^{87.39}_{\pm 5.19}$	94.78 ±2.92	$93.48 \\ \pm 2.66$	96.09	$^{81.30}_{\pm 3.30}$	$^{86.96}_{\pm 4.61}$	99.13 ±1.94	92.17 ± 2.48	93.48	76.09 ± 6.15	$\begin{array}{c} 87.83 \\ \pm 3.30 \end{array}$	96.52 ±3.63	95.65 ± 3.44
AVG	95.01	69.25	80.43	89.16	86.60	94.30	72.10	80.51	91.38	87.67	90.80	64.06	76.32	87.90	84.44
STD	7.77	11.62	8.01	5.77	6.69	9.34	13.43	11.61	8.48	7.34	10.67	13.87	8.42	6.00	7.17

Table 1: Global localization accuracy of each subject (%)(DSGAT*: DSGAT w/o T_m^{emb})

vs BS2, p < 0.005). For the three paradigms, the average 112 classification accuracies across all subjects using DSGAT are 89.16%, 91.38%, and 87.90%, indicating that all three paradigms are feasible and effective. For each subject, at least one paradigm was able to achieve a classification accuracy of over 70%, and the highest accuracy of each subject (using any paradigm and approach) reaches an average of 92.84%, with subject S11 being able to achieve 99.13% using the AA and NA paradigms.

Comparing the three paradigms, it can be concluded from the overall average results that the NA paradigm obtains the best performances using different algorithms, the AA paradigm comes second, and the SA paradigm has the lowest average classification accuracy. But the one-way RANOVA shows that there is no statistically significant difference in the accuracies among the three paradigms [F(2,20)=2.228, p > 0.05].

Compared to the SSVEP detection accuracies of intermediate cells, the classification results of the minimum cell by DSGAT do not show obvious correlation with them, with a Pearson correlation coefficient between the two types of accuracies of -0.206, i.e., the SSVEP detection accuracy does not directly affect the results of DSGAT. Especially for S8, the SSVEP detection accuracies under the three paradigms are 73.48%, 67.83%, and 62.61%, while the accuracies obtained by the DSGAT algorithm are 90.00%, 92.17%, and 93.91%, respectively. This interesting result indicates that DSGAT is able to compensate for the detection error of FBCCA. In the case of unsatisfactory SSVEP detections, it is still possible for the DSGAT algorithm to learn the correct mapping relation from the intermediate cell response distribution.

Compare the performance of different global localization methods. BS1 is

An SSVEP-based BCI with 112 Targets

an unsupervised method that intuitively reflects the approximate distribution of competitive stimulus responses. This method obtained the lowest accuracy, indicating that it is difficult to comprehensively capture the underlying relationships and distribution patterns among competing stimuli by analyzing them in a piecemeal fashion. Some subjects can achieve an accuracy of 80% or 90% or even higher through the unsupervised method, while some others are much lower; for example, the data of S4 has excellent separability, which directly demonstrates the feasibility of the paradigm. BS2 has higher accuracy compared to BS1, and a few subjects are able to achieve results close to those of DSGAT. Compared to the two baselines, DSGAT is more stable and has the smallest standard deviation over 5-fold cross-validation, while the baselines are more affected by the data. For example, for S2 and S8, both subjects have very limited baseline results, but achieve good results with DSGAT.

Comparing the three algorithms, two baselines perform 4-classification on the basis of a single intermediate cell screened by FBCCA, and thus global localization is directly affected by the SSVEP detection. In contrast, DSGAT applies global SSVEP response information, which makes it possible to analyze subjects' SSVEP distribution preferences and thus locate the correct target minimum cell, even if the intermediate cell is mislocalized. We further analyzed the relationship between the two-step localization results. For DSGAT and the two baselines, we calculated the proportion of samples with correct global localization (TP_C) among those with correct SSVEP detection results (TP_M) , denoted as $P(TP_C|TP_M)$, and among those with incorrect SSVEP detection results (TN_M) , denoted as $P(TP_C|TN_M)$, respectively. The results are recorded in table 2. It can be seen that DSGAT can accomplish correct global localization even in the case of incorrect SSVEP detection, while the baselines must base on the right intermediate cell.

Temporal embedding operation was added to the DSGAT network. To verify the impact of this operation, we removed the temporal embedding layer to perform the ablation study. The results are presented in table 1. The temporal embedding operation improved the average accuracy of all subjects by 2.56%, 3.71%, and 3.46% for AA, NA, and SA paradigms, respectively, demonstrating that the temporal embedding further exploits the process information of SSVEP detection to optimize the classification performance. DSGAT w/o T_m^{emb} outperforms two baselines and can get even higher accuracy than DSGAT for some subjects, further demonstrating the superiority of treating the paradigm as graph data and showing that for the proposed paradigm, the key to minimum target classification lies in dealing with the relationship between competing and target stimuli.

Table 3 lists the average ITRs by different algorithms using the three paradigms. Since the data length used was the same for all subjects, only the average ITRs are reported in the table. The highest average ITRs for the three paradigms are 51.66 ± 5.07 bits/min, 53.96 ± 7.33 bits/min, and 50.55 ± 5.36 bits/min, respectively.

Table 2: (Comparison	analysis	of SSVEP	detection	and	global	localization	results
------------	------------	----------	----------	-----------	-----	--------	--------------	---------

Table	2: Compariso	on anal	lysis of	SSVEP	detect	ion and	d global I	localız	ation re	esults
Subject	Sample type	AA paradigm			NA paradigm			SA paradigm		
Subject	Sumple type	BS1	BS2	DSGAT	BS1	BS2	DSGAT	BS1	BS2	DSGAT
S1	$P(TP_C TP_M)$	0.70	0.82	0.86	0.82	0.89	0.94	0.80	0.83	0.86
	$P(TP_C TN_M)$	0	0	0.50	0	0	1.00	0	0	0.56
S2	$P(TP_C TP_M)$	0.61	0.87	0.91	0.74	0.86	0.93	0.72	0.91	0.89
~	$P(TP_C TN_M)$	0	0	0.89	0	0	0.97	0	0	0.75
S3	$P(TP_C TP_M)$	0.75	0.84	0.88	0.78	0.83	0.87	0.70	0.83	0.84
~~ _	$P(TP_C TN_M)$	0	0	0.88	0	0	0.75	0	0	0.88
S4	$P(TP_C TP_M)$	0.89	0.91	0.96	0.88	0.93	0.96	0.75	0.80	0.84
	$P(TP_C TN_M)$	0	0	1.00	0	0	1.00	0	0	1.00
S5	$P(TP_C TP_M)$	0.68	0.70	0.76	0.59	0.64	0.68	0.56	0.69	0.79
	$P(TP_C TN_M)$	0	0	0	0	0	0.75	0	0	0.50
S6	$P(TP_C TP_M)$	0.71	0.88	0.93	0.81	0.89	0.94	0.84	0.92	0.96
	$P(TP_C TN_M)$	0	0	0.78	0	0	0.67	0	0	0.88
S7	$P(TP_C TP_M)$	0.75	0.85	0.91	0.74	0.89	0.96	0.70	0.86	0.93
	$P(TP_C TN_M)$	0	0	0.50	0	0	1.00	0	0	0.63
S 8	$P(TP_C TP_M)$	0.60	0.87	0.91	0.59	0.82	0.96	0.47	0.91	0.94
	$P(TP_C TN_M)$	0	0	0.89	0	0	0.85	0	0	0.93
S9	$P(TP_C TP_M)$	0.72	0.87	0.93	0.72	0.80	0.90	0.68	0.88	0.94
	$P(TP_C TN_M)$	0	0	0.00	0	0	0.67	0	0	0.79
S10	$P(TP_C TP_M)$	0.72	0.80	0.87	0.83	0.94	0.97	0.68	0.74	0.84
	$P(TP_C TN_M)$	0	0	0.80	0	0	0.92	0	0	0.50
S11	$P(TP_C TP_M)$	0.83	0.90	0.95	0.85	0.90	0.989	0.81	0.94	0.98
NTT -	$P(TP_C TN_M)$	0	0	1.00	0	0	1.00	0	0	0.87

Table 3: The average ITR (bits/min) for the three paradigms using different algorithms

Paradigm BS1	BS2	DSGAT	DSGAT*
AA 35.72 ± 8.33	44.22 ± 6.37	51.66 ± 5.07	49.41 ± 5.65
NA 37.97 ± 9.52	44.53 ± 9.01	53.96 ± 7.33	50.39 ± 6.15
SA 32.09 ± 9.06	40.93 ± 6.42	50.55 ± 5.36	47.57 ± 6.13

5. Discussion

Frequency spatial multiplexing can also be seen as a segmentation of the intermediate cell. Then the 112 target classification is achieved merely by relying on the relative position of the stimuli and the position of the fixation point. Based on the AA paradigm, the NA paradigm was proposed to decouple the frequency and spatial relationship to

An SSVEP-based BCI with 112 Targets

explore whether the same efficient global classification can be achieved through the spatial relationship only. NA paradigm keeps similar frequencies as far away from each other as possible, which helps to improve the SSVEP detection performance, but in turn, introduces the risk of locating completely off-target areas. SA paradigm was proposed to increase the distinguishability between adjacent targets. The experimental results show that the NA paradigm has better performance, but the three paradigms do not show significant differences (p > 0.1), demonstrating that all three paradigms can achieve high-resolution target selection and that global classification relies more on the spatial relationships of competing stimuli.

In this study, the minimum target classification used the strategy of first detecting the SSVEP response distribution of the intermediate cells and then locating the minimum target. Therefore, the supervision part of the location is essentially a 4classification problem. If minimum target classification were performed directly, it transformed into a 112 classification problem, then a massive amount of data needs to be collected to ensure each class has enough data.

The classification accuracy of the GCN method can reach a satisfactory level. Although the paradigm layout is regular, we still treated it as graph data because the properties of each minimum cell are different, determined by its relative position in the intermediate cell, which is the irregularity of the paradigm. The advantages of DSGAT are:

- 1) Spatially-based graph convolutional network models perform the same local convolution for each node, thus require fewer parameters by sharing weights, and are highly scalable, allowing the paradigm to scale up or down in size freely.
- 2) Although the aggregation operation is local inductive learning, but the local aggregation covers the whole paradigm. Therefore, DSGAT is able to handle 1-step and 2-step neighbor informations as well as more distant response informations, which takes into account the global while focusing on the local.
- 3) DSGAT can adaptively demarcate the competing influence scale. Multi-layer GCNs are able to obtain a larger perceptual domain. In this study, the aggregation operation was performed only on \mathcal{G}_m , and we have attempted to make a second aggregation layer on \mathcal{G}_c , which resulted in training overfitting due to the contradiction between the parameter increase and the current amount of data. And the single-layer aggregation operation was sufficient to verify the paradigm feasibility; thus, finally, we did not use the multilayer aggregation structure.

Technically, DSGAT differs from the standard attention mechanism because classical attention is a shared mechanism, i.e., all neighbors share the same parameters to calculate the correlation coefficient through the features themselves. In this work, the neighbor features are determined by the neighborhood relationship, so the shared attention aggregation can only invert the established node relationships. Therefore, for the specificity of the paradigm, DSGAT directly takes the edge weights W_{α} as the learnable parameters.

Table 4: Characteristics of BCI study focusing on multiple targets

			-	-	
Authors (Year)	Number of targets	Number of frequencies	Accuracy (%)	ITR (bits/min)	Method
Nakanish (2018) [5]	40	40	89.83	325.33	single-frequency
Chen et al (2014) [26]	45	45	84.1	105	single-frequency
Hwang et al (2013) [9]	12	4	87.23	33.26	frequency spatial multiplexing
Liang (2020) [12]	40	11	96.06	196.09	frequency spatial multiplexing
Yosuke et al (2013) [11]	10	2	81.75	28.90	frequency temporal multiplexing
Chen et al (2021) [13]	160	40	87.16	78.84	frequency temporal multiplexing
Benítez et al (2018) [27]	5	4	97.78	not mentioned	frequency control signals superposition
Chen et al (2013) [10]	8	3+2	93.83	33.80	other attributes frequency modulation
Shi et al (2023) [6]	160	-	95	366.05	broadband code modulated VEP
This study	112	40	91.38	53.96	frequency spatial multiplexing

As can be seen from the results, the classification accuracies of the two stages do not show correlation and have individual differences. The average intermediate cell classification accuracy of most subjects is higher than the final minimum target classification accuracy, and there are also some experimental results that the final accuracies are higher, indicating the effectiveness of the algorithm. The reasons for the lower final classification accuracy could be, on the one hand, insufficient data samples resulting in the algorithm not fully capturing the pattern features. On the other hand, it is also necessary to consider the individual data differences, which may be the different adaptability of subjects to the new paradigm, or the differences in the subjects' response patterns to the target and competing stimuli.

As shown in table 4, we compared the proposed method with representative studies that increase the target resolution of SSVEP-BCI. The table lists the methods of each study, the number of targets, the number of frequencies used, and the performance of the experiments. In addition, a cVEP method is also listed in the table, which illustrates the ability of the cVEP method to enhance target resolution, but we believe that SSVEP BCIs also hold promise for further enhancement and is worth continuing to explore. Compared to other methods, our method does not currently stand out in terms of number of targets and ITR. But the proposed method has its unique advantages. First of all, we use the method of frequency spatial multiplexing, a line of research that is only attempted with a small number of targets, whereas our method has large number of targets. On the other hand, the method we proposed does not conflict with other excellent methods. While other paradigms focus on the design of individual target

25

An SSVEP-based BCI with 112 Targets

stimuli, this study relies on the positional relationship between the stimuli and does not circumvent the competing response problem. Therefore, not only these methods, but also other state-of-the-art methods focusing on enhancing the efficiency of SSVEP detection can be used as a basis for the present method to achieve very promising enhancements to make further contributions.

Some limitations of this study should to be discussed further. 1) Firstly, the proposed method has only achieved good results in offline experimental evaluations, in the future, we need to further validate the feasibility under online conditions. 2) In addition, there is still a gap in the ITR with some cutting-edge SSVEP-BCI systems, due to the long stimulus duration. There is a significant difference between the laboratory environment and the real-world. For applications in the real-word, methods with excellent response speed could ensure more robust performance. And for the SSVEP BCI study, longer duration of visual stimulation inevitably brought about visual fatigue. These factors prompted us to further explore more strategies to improve the performance under short time windows. 3) Considering that in some practical applications, visual stimuli may be superimposed on top of the image, but the current large size of the stimuli will affect the observation of other tasks. If subjects are required to balance observation and control, the situation of long-time operation risks aggravating the brain load. Therefore lower burden interactivity is also an area that should be further optimized. 4) Currently our algorithm requires within-subjects training, which puts pressure on data collection and a relatively heavier burden on the subjects. The next step is to explore cross-subjects classification methods, employing transfer learning techniques to make full use of the old data to transfer to the new subject data. 5) This work focuses on validating the effectiveness of the proposed method through three paradigms, and the next step will be to further optimize the performance for a specific paradigm.

Next, we aim to address the current shortcomings and systematically improve the performance. In terms of time reduction, according to the time-frequency analysis, the SSVEP response of competing neighbors has increased in the first half of the stimulation phase. The next step is to use graph spatial-temporal networks (GSTN) to study the real-time global classification method. GSTN better captures the temporal variation of the relationship and is suitable for the temporal continuity of SSVEP detection results. Thus, GSTN is expected to learn the changing pattern of competing stimuli better to shorten the stimulus duration. Regarding raising the number of targets, 1) In the proposed paradigm, a frequency is assigned only to one intermediate cell. Further, the frequency spatial multiplexing mechanism can be extended to intermediate cells, i.e., a frequency is assigned to multiple intermediate cells at different locations, further increasing the number of targets through a richer frequency scheduling relationship. 2) Combine with other methods to advance research based on better performing SSVEP BCIs. Finally, the present method has application prospects in the fields of medical health, computer vision assistance, spatial navigation, complex cognitive decoding, and human-machine shared control, etc. However, the current research is synchronous, and in practical asynchronous applications, the spatio-temporal pattern of the SSVEP

An SSVEP-based BCI with 112 Targets

response changes when the subject needs to switch the target, and how to adapt to and discriminate such a situation is an important aspect that needs to be further optimized.

6. Conclusion

This study designed a novel unimodal SSVEP-BCI paradigm with 112 targets based on frequency spatial multiplexing and neuronal competing mechanism. Instead of designing individual visual stimuli, the present work distinguishes the minimum cells relying on the location relationship of the stimulus arrangement. Three specific interfaces were designed, namely AA, NA, and SA paradigms. A dual-scale graph attention network was constructed as a global localization algorithm based on the SSVEP detection using FBCCA. Eleven subjects participated in the offline validation experiments and obtained an average global localization accuracy of 91.38% and ITR of 53.96 bits/min using the DSGAT algorithm in the NA paradigm. This study is applicable to BCI application scenarios with a large number of targets and has the potential to expand the number of targets for SSVEP-BCI further.

Acknowledgments

This work is supported by the National Natural Science Foundation of China [grant numbers: 62203460, 62036013, 62376149]. The authors gratefully acknowledge this support.

References

- Ola Friman, Ivan Volosyak, and Axel Graser. Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*, 54(4):742–750, 2007.
- [2] Yu Zhang, Guoxu Zhou, Jing Jin, Minjue Wang, Xingyu Wang, and Andrzej Cichocki. L1regularized multiway canonical correlation analysis for ssvep-based bci. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 21(6):887–896, 2013.
- [3] Xiaogang Chen, Yijun Wang, Masaki Nakanishi, Xiaorong Gao, Tzyy-Ping Jung, and Shangkai Gao. High-speed spelling with a noninvasive brain-computer interface. Proceedings of the National Academy of Sciences, 112(44):6058–6067, 2015.
- [4] Yaru Liu, Wei Dai, Huimin Lu, Yadong Liu, and Zongtan Zhou. Brain-computer interface for human-multirobot strategic consensus with a differential world model. Applied Intelligence, 51(6):3645–3663, 2021.
- [5] Masaki Nakanishi, Yijun Wang, Xiaogang Chen, Yu-Te Wang, Xiaorong Gao, and Tzyy-Ping Jung. Enhancing detection of ssveps for a high-speed brain speller using task-related component analysis. *IEEE Transactions on Biomedical Engineering*, 65(1):104–112, 2018.
- [6] Nanlin Shi, Yining Miao, Changxing Huang, Xiang Li, Yonghao Song, Chen Xiaogang, Wang Yijun, and Gao Xiaorong. Estimating and approaching maximum information rate of noninvasive visual brain-computer interface. arXiv preprint arXiv:2308.13232, 2023.
- [7] Erich E. Sutter. The brain response interface: communication through visually-induced electrical brain responses. *Journal of Microcomputer Applications*, 15(1):31–45, 1992. Special Issue on Computers for Handicapped People.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
50
57
52 52
57
54
55
50
57 50
50
59 60
oυ

An SSVEP-based BCI with 112 Targets

- [8] Wei Dai, Yaru Liu, Huimin Lu, Zhiqiang Zheng, and Zongtan Zhou. A shared control framework for human-multirobot foraging with brain-computer interface. *IEEE Robotics and Automation Letters*, 6(4):6305–6312, 2021.
- [9] Han-Jeong Hwang, D. Kim, Changhee Han, and C. Im. A new dual-frequency stimulation method to increase the number of visual stimuli for multi-class ssvep-based brain-computer interface (bci). Brain Research, 1515:66–77, 2013.
- [10] Xiaogang Chen, Zhikai Chen, Shangkai Gao, and Xiaorong Gao. Brain-computer interface based on intermodulation frequency. *Journal of Neural Engineering*, 10(6):066009, oct 2013.
- [11] Yosuke Kimura, Toshihisa Tanaka, Hiroshi Higashi, and Naoki Morikawa. Ssvep-based brain-computer interfaces using fsk-modulated visual stimuli. *IEEE Transactions on Biomedical Engineering*, 60(10):2831–2838, 2013.
- [12] Liyan Liang, Jiajun Lin, Chen Yang, Yijun Wang, Xiaogang Chen, Shangkai Gao, and Xiaorong Gao. Optimizing a dual-frequency and phase modulation method for SSVEP-based BCIs. 17(4):046026, aug 2020.
- [13] Yonghao Chen, Chen Yang, Xiaochen Ye, Xiaogang Chen, Yijun Wang, and Xiaorong Gao. Implementing a calibration-free SSVEP-based BCI system with 160 targets. *Journal of Neural Engineering*, 18(4):046094, 2021.
- [14] Sandra Fuchs, Søren K. Andersen, Thomas Gruber, and Matthias M. Müller. Attentional bias of competitive interactions in neuronal networks of early visual processing in the human brain. *NeuroImage*, 41(3):1086–1101, 2008.
- [15] Yonghao Chen, Chen Yang, Xiaogang Chen, Yijun Wang, and Xiaorong Gao. A novel training-free recognition method for SSVEP-based BCIs using dynamic window strategy. *Journal of Neural Engineering*, 18(3):036007, mar 2021.
- [16] Kian B Ng, Andrew P Bradley, and Ross Cunnington. Stimulus specificity of a steady-state visualevoked potential-based brain-computer interface. Journal of Neural Engineering, 9(3):036008, may 2012.
- [17] Peng Yuan, Xiaogang Chen, Yijun Wang, Xiaorong Gao, and Shangkai Gao. Enhancing performances of SSVEP-based brain-computer interfaces via exploiting inter-subject information. *Journal of Neural Engineering*, 12(4):046006, jun 2015.
- [18] Nicholas R Waytowich, Josef Faller, Javier O Garcia, Jean M Vettel, and Paul Sajda. Unsupervised adaptive transfer learning for steady-state visual evoked potential brain-computer interfaces. In 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pages 004135– 004140, 2016.
- [19] Xiaogang Chen, Yijun Wang, Shangkai Gao, Tzyy-Ping Jung, and Xiaorong Gao. Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain-computer interface. *Journal of Neural Engineering*, 12(4):046008, jun 2015.
- [20] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S. Yu. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1):4–24, 2021.
- [21] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications. AI Open, 1:57–81, 2020.
- [22] Feng Xia, Ke Sun, Shuo Yu, Abdul Aziz, Liangtian Wan, Shirui Pan, and Huan Liu. Graph learning: A survey. *IEEE Transactions on Artificial Intelligence*, 2(2):109–127, 2021.
- [23] Jonathan Peirce, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1):195–203, 2019.
- [24] Graimann Bernhard, Pfurtscheller Gert, and Allison Brendan, editors. Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction. Springer, Berlin, Heidelberg, Berlin, Heidelberg, 2010.
- [25] Jie Pan, Xiaorong Gao, Fang Duan, Zheng Yan, and Shangkai Gao. Enhancing the classification

An SSVEP-based BCI with 112 Targets

accuracy of steady-state visual evoked potential-based brain–computer interfaces using phase constrained canonical correlation analysis. *Journal of Neural Engineering*, 8(3):036027, may 2011.

- [26] Xiaogang Chen, Zhikai Chen, Shangkai Gao, and Xiaorong Gao. A high-itr ssvep-based bci speller. Brain-Computer Interfaces, 1(3-4):181–191, 2014.
- [27] J.L. Pérez-Benítez, J.A. Pérez-Benítez, and J. H. Espina-Hernández. Development of a brain computer interface using multi-frequency visual stimulation and deep neural networks. In 2018 International Conference on Electronics, Communications and Computers (CONIELECOMP), pages 18–24, 2018.