



OPEN

DATA DESCRIPTOR

# A 40-Class SSVEP Speller Dataset: Beta Range Stimulation for Low-Fatigue BCI Applications

Heegyu Kim<sup>1</sup>, Kyungho Won<sup>2</sup>, Minkyu Ahn<sup>3</sup> & Sung Chan Jun<sup>1,4</sup>  

The inherent non-stationarity of electroencephalography (EEG) signals necessitates large, consistent datasets for reliable brain–computer interface (BCI) research. In steady-state visual evoked potential (SSVEP) paradigms, prolonged exposure to visual stimuli can induce visual fatigue, leading to alterations in EEG patterns that degrade BCI performance. To mitigate fatigue-induced variability, this study employs visual stimulation in the beta frequency range (14–22 Hz), a range that appears less susceptible to the effects of fatigue. We present a comprehensive 40-class SSVEP speller dataset acquired from 40 participants, with EEG data recorded from 31 central-to-occipital channels. Each subject completed six sessions of the SSVEP speller task in addition to pre- and post-experiment resting-state recordings under both eyes-open and eyes-closed conditions. Subjective fatigue ratings combined with EEG band power analyses confirm that beta-range stimulation minimizes fatigue effects. Moreover, the high classification accuracy achieved by calibration-based algorithms indicates that the dataset is well-suited for training advanced SSVEP-based BCI systems.

## Background & Summary

The Brain–Computer Interface (BCI) is a technology that facilitates communication with external devices by decoding brain activity and transmitting commands<sup>1–3</sup>. BCI technology is being developed for clinical applications such as virtual keyboards, wheelchairs, and rehabilitation systems for patients with severe motor or communication impairments<sup>4</sup>. It is also being explored for everyday applications, including virtual reality and smart home control<sup>5</sup>.

The Steady-State Visual Evoked Potential (SSVEP) is a brain signal characterized by specific frequency components elicited by periodic visual stimuli<sup>6</sup>. Due to its high signal-to-noise ratio (SNR) and information transfer rate (ITR), SSVEP is widely used in BCI paradigms, allowing for the modulation of multiple commands via visual targets<sup>7</sup>. Typically, the high-speed performance of BCIs is achieved through multi-target visual spellers. The term “BCI illiteracy” describes the difficulty some individuals experience in effectively using BCI paradigms; however, the user-friendly design of SSVEP systems and their low incidence of BCI illiteracy make them promising candidates for real-world applications<sup>8</sup>.

Significant advancements have been made to enhance BCI performance in SSVEP frequency recognition<sup>9</sup>. Owing to its inherent properties, SSVEP can achieve high classification performance even with zero-training recognition methods such as canonical correlation analysis (CCA) and filter bank CCA (FB-CCA). Nevertheless, further improvements have been demonstrated using individual template (IT)-CCA, task-related component analysis (TRCA), and various machine learning techniques<sup>10</sup>. Recently, there has been an increase in attempts to use deep learning methods—including convolutional neural networks (CNNs), long short-term memory networks (LSTMs), Transformers, and generative adversarial networks (GANs)—for SSVEP classification<sup>8</sup>. To develop generalizable BCI systems, subject-independent classifiers are needed, which require data from multiple subjects.

In response to these requirements, numerous SSVEP BCI datasets have been made publicly available. Several notable datasets feature multi-class SSVEP speller data utilizing target frequencies within the alpha range (8–16 Hz)<sup>11,12</sup> as well as SSVEP paradigms that include four directional commands<sup>13,14</sup>. Moreover, there are

<sup>1</sup>School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju, South Korea. <sup>2</sup>Department of Embedded System Engineering, Incheon National University, Incheon, South Korea. <sup>3</sup>School of Computer Science and Electrical Engineering, Handong Global University, Pohang, South Korea. <sup>4</sup>AI Graduate School, Gwangju Institute of Science and Technology, Gwangju, South Korea. ✉e-mail: [scjun@gist.ac.kr](mailto:scjun@gist.ac.kr)

datasets that focus on SSVEP speller frequencies between 9 and 14 Hz, alongside those that span a wider frequency range (1–60 Hz) to explore the general characteristics of SSVEP<sup>15,16</sup>.

Due to the inherent exposure to periodic visual stimuli, SSVEP paradigms often induce visual fatigue, which can lead to significant alterations in EEG signals and compromise classification performance<sup>17,18</sup>. Studies have reported that visual fatigue particularly affects brain activity in the occipital region, causing increases in delta, theta, and alpha power, while beta and gamma bands tend to remain relatively stable<sup>19</sup>. Given the nonstationary nature of EEG signals, reducing such variability is crucial for improving model training and classification accuracy<sup>20,21</sup>. Fatigue can also reduce the signal-to-noise ratio (SNR) of SSVEP responses and impair individual calibration accuracy<sup>22,23</sup>. In contrast to alpha and theta bands, where amplitude fluctuations occur over time, the beta range exhibits more consistent characteristics. Prior research has shown that while alpha power increases and theta activity decreases under fatigue, beta band activity remains relatively unaffected<sup>24</sup>. Consistent with these findings, our results also showed notable increases in alpha power with minimal changes in the beta band.

A standard method for estimating visual fatigue in EEG analysis involves spectral power comparisons<sup>19,25</sup>, particularly focusing on the absolute and relative power increases in the alpha and theta bands, which are well-known indicators of fatigue. To quantify visual fatigue, we used the CCA coefficient, which is frequently employed for both qualitative and quantitative assessments of EEG-based fatigue<sup>26</sup>. Beyond its application in the SSVEP paradigm, CCA is widely used for the objective detection of various types of visual fatigue, enabling straightforward and objective quantitative analyses of test data. Furthermore, SSVEP responses typically show a reduction in amplitude due to visual fatigue<sup>27</sup>, with fatigue effects manifesting as changes in frequency peak amplitude or variations in the peak response's signal-to-noise ratio (SNR)<sup>28</sup>.

This study aims to collect EEG data from a large number of subjects using a massive-class SSVEP-based BCI speller. By employing stimuli within the beta wave range, which is less susceptible to visual fatigue, we aim to reduce inter-subject variability in EEG data. This dataset is designed to support deep learning applications by providing EEG data with reduced variability and improved suitability for training.

## Methods

**Participants.** A total of 40 subjects (25 males/15 females, with a mean age of  $22.8 \pm 3.34$  years, ranging from 20 to 35 years) participated in the experiment. Among these, three had previous experience with SSVEP-based BCI experiments. Importantly, none of the participants reported any history of neurological, psychiatric, or other conditions that could potentially affect the outcomes. All subjects were positioned at a distance of 70 cm from the monitor during the experiment (as Fig. 1B). The Institutional Review Board at Gwangju Institute of Science and Technology (GIST) approved this experiment (20211201-HR-64-02-04), and all subjects were informed of all experimental procedures and signed consent forms that included confidentiality of personal information (excluding questionnaire-related personal data) as well as consent for the use and open sharing of the data. Data sharing follows the appropriate ethical approval procedures to ensure that the privacy of participants is protected.

**SSVEP Paradigm.** The SSVEP speller was presented on a monitor (LG Electronics, Inc., LG24GL600F) with a  $1920 \times 1080$  resolution, 24 inches, and a 120 Hz refresh rate. Presentation software was developed in MATLAB (MathWorks, Inc., MA, USA; ver. 2020a) using Psychophysics Toolbox version 3 (PTB-3)<sup>29</sup>. The frequency and phase parameters for the 40 stimuli were designed using the joint frequency and phase modulation (JFPM) approach<sup>11</sup>. Each stimulus target measured  $140 \times 140$  pixels and was arranged in a  $5 \times 8$  matrix, with 50-pixel gaps between targets in both the vertical and horizontal directions. The flickering frequencies ranged from 14.0 Hz to 21.8 Hz, incremented by 0.2 Hz, with a phase difference of  $0.5\pi$  between adjacent stimuli (refer to Fig. 1C).

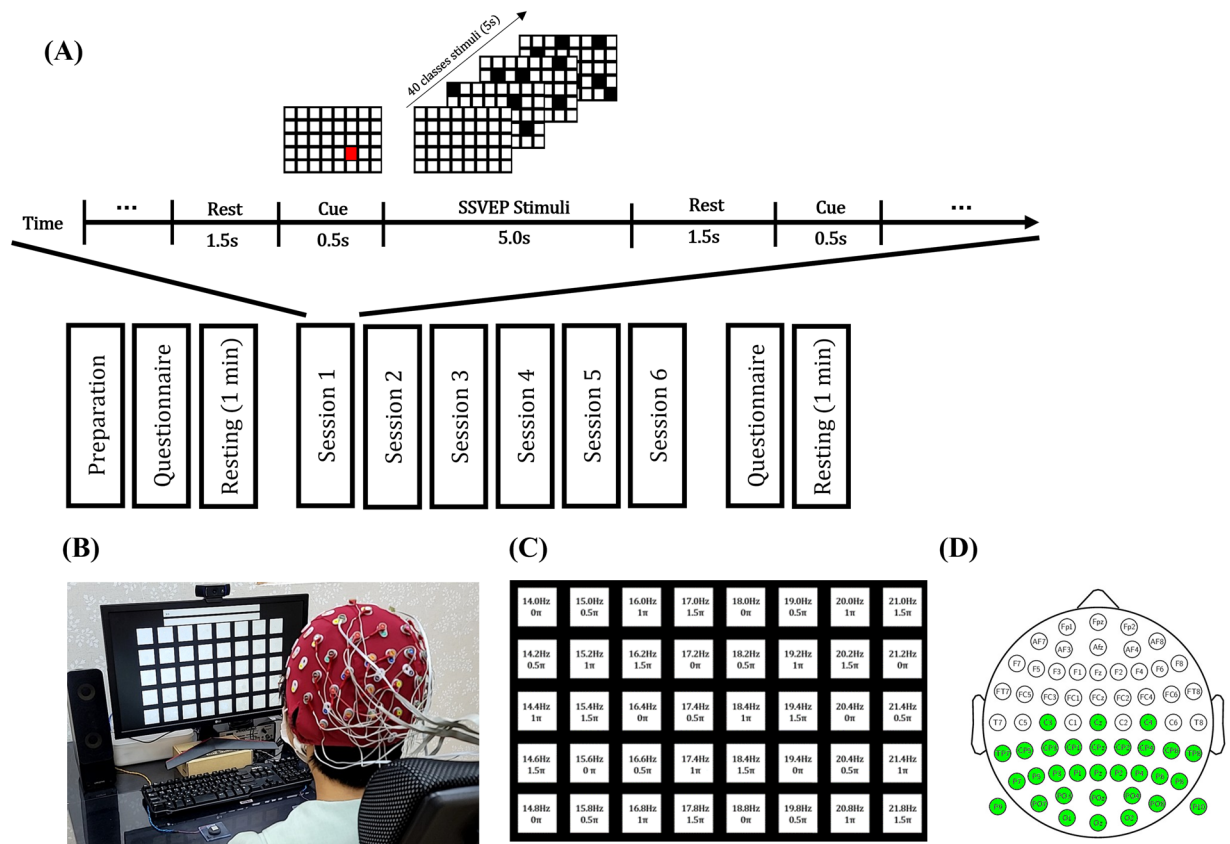
The experiment was structured as a cue-based target selection task. It comprised six blocks, each containing 40 trials that randomly presented all stimuli. Each trial began with a blank screen displayed for 1.5 seconds, followed by a 0.5-second target cue, and then a 5-second flickering period (see Fig. 1A). To maintain engagement, each session included a 1- to 2-minute break prior to the commencement of the next session.

In addition, questionnaires were administered both before and after the experiment. Pre-experimental surveys gathered personal and eye-related information, as well as measurements of the subjects' mental and physiological states (refer to Table 1). Following the surveys, a resting-state recording of over 1 minute was obtained while a black screen with a central fixation cross (+) was displayed.

**Experimental procedure.** The experiment was conducted in the order shown in Fig. 1A. Details are as follows: Each participant began the session with a thorough cleaning of the skin around the mastoid areas before the attachment of electrodes. A brief instructional video was shown to familiarize participants with the experimental procedure and the tasks that would follow. Following this, EEG electrodes were affixed, and checks for impedance and signal quality were conducted. The distance between the participant and the monitor was adjusted to approximately 70 cm to ensure uniform visual presentation. Participants then proceeded to complete the initial set of questionnaires (sections A to C). Subsequently, resting-state EEG data were recorded while a fixation cross was displayed on a black screen—first with participants' eyes open and then with eyes closed. The main experiment consisted of six sessions, each lasting no longer than five minutes. Short breaks of one to three minutes were provided between sessions to mitigate fatigue. After the final SSVEP session, participants completed the remaining questionnaire items (sections D to E). Following a brief one-minute rest period, a second resting-state EEG was recorded under the same conditions as the initial measurement.

## Data Records

**EEG data recording.** EEG signals were recorded at a sampling rate of 1024 Hz using 31 Ag-AgCl wet electrodes, with additional electrodes placed on both mastoids for reference. The BioSemi ActiveTwo system (BSM,



**Fig. 1** The SSVEP Experimental Paradigm. **(A)** Experiment Timeline: The study included pre- and post-surveys and a resting-state measurement lasting over one minute. It comprised six sessions, each presenting the 40-class stimuli in a randomized order, with each stimulus shown once per session in single day and single study. **(B)** Experimental setup: Participants sat approximately 70 cm away from a monitor while wearing EEG equipment during the experiment. **(C)** SSVEP Speller Design: The SSVEP speller was organized as an 5 × 8 matrix utilizing JFPM. **(D)** Electrode Configuration: The research utilized a Biosemi 64-channel system, collecting data from the 31 channels highlighted in green, along with both mastoid electrodes.

BioSemi, Amsterdam, Netherlands) served as the EEG amplifier. Prior to data collection, electrode impedances were confirmed to be below 5 kΩ for all channels. EEG data were collected and event-synchronized using OpenViBE<sup>30</sup>, with data acquisition conducted under conditions ensuring less than 2 ms of drift during the recording process. The electrodes were arranged according to the extended international 10–20 system<sup>31</sup>, covering the entire scalp, with the CMS/DRL positioned near the Pz region (see Fig. 1D). The data repository for this study is available at: [<https://doi.org/10.6084/m9.figshare.28806815.v2>]<sup>32</sup>.

**Data structure.** For each subject, the EEG data were organized into two distinct ‘mat’ files corresponding to the Resting State and SSVEP task recordings. The resting state data comprise four 60-second segments: pre-experiment; and eyes open and eyes closed conditions post-experiment. The SSVEP task data are epoched from –2000 ms to 5000 ms relative to the stimulus onset and are organized into 40 classes across 6 blocks, with stimulus parameters stored sequentially (see Fig. 2). Due to a personal issue, post-experiment resting state data for Subject 1 were not recorded; however, this subject successfully completed the pre-experiment resting state recording and fully participated in the SSVEP task. For information regarding data storage and access, please refer to the Code Availability section.

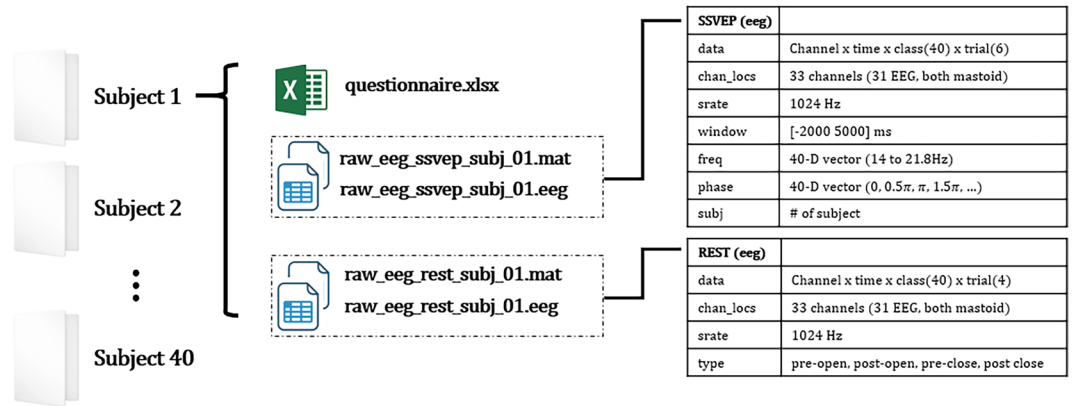
**Questionnaire.** In the experimental sequence illustrated in Fig. 1A, participants completed the questionnaires twice: once before and after the experiment. The questionnaire items, detailed in Table 1, were organized as follows: Sections A, B, and C were filled out before the experiment, while Sections D and E were completed afterward. All responses were saved in Excel (.xlsx) format, with each file labeled according to the participant’s number.

### Technical Validation

**Classification methods.** *Standard CCA (stdCCA).* Canonical Correlation Analysis (CCA) is a multivariate statistical method that reveals the underlying correlation between two matrices<sup>33</sup>. CCA calculates the relation between two multi-variable datasets after a linear combination of original data. In most studies, only the first maximum canonical coefficient is employed, and other canonical coefficients are worth considering.

| Index    | Question                                              | Answer                             |
|----------|-------------------------------------------------------|------------------------------------|
| <b>A</b> | <b>Personal information</b>                           |                                    |
| A01      | Wearing Glasses                                       | O/X                                |
| A02      | Physical/Mental Illness                               | O/X                                |
| A03      | Eye-related Disease                                   | O/X                                |
| A04      | Regular Use of Eye-related Medication                 | O/X                                |
| A05      | History of Eye-related Surgery                        | O/X                                |
| A06      | Experience with BCI Experiments                       | O/X                                |
| A07      | Experience with SSVEP Experiments                     | O/X                                |
| A08      | Preferred Handedness                                  | R/L/both                           |
| A09      | Gender                                                | M/F                                |
| <b>B</b> | <b>Physiological and Psychological condition</b>      |                                    |
| B01      | How many hours did you sleep last night?              | less 4, 4~6, 6~7, 7~8, over 8 hour |
| B02      | Did you drink coffee in the past 24 hours?            | 0, 0~6, 6~12, 12~18, 18~ 24 hour   |
| B03      | Did you consume alcohol in the past 24 hours?         | 0, 0~6, 6~12, 12~18, 18~ 24 hour   |
| B04      | Did you smoke in the past 24 hours?                   | 0, 0~6, 6~12, 12~18, 18~ 24 hour   |
| B05      | Display Exposure Environment                          | 0~1, 1~2, 2~3, 3~4, over 5 hour    |
| B06      | Daily Computer Usage                                  | 0~1, 1~2, 2~3, 3~4, over 5 hour    |
| B07      | Daily Smartphone Usage                                | 0~1, 1~2, 2~3, 3~4, over 5 hour    |
| <b>C</b> | <b>Questionnaire prior to experiments</b>             |                                    |
| C01      | Condition - Comfort                                   | Scale (1-5), 5: uncomfortable      |
| C02      | Condition -Drowsiness                                 | Scale (1-5), 5: drowsiness         |
| C03      | Condition - Concentration                             | Scale (1-5), 5: concentrate        |
| C04      | Condition - Physical condition                        | Scale (1-5), 5: high fatigue       |
| C05      | Condition - Mental condition                          | Scale (1-5), 5: high fatigue       |
| C06      | Condition - Eye fatigue - My eyes feel tired.         | Scale (1-5), 5: high fatigue       |
| C07      | Condition - Eye fatigue - I am blinking a lot.        | Scale (1-5), 5: high fatigue       |
| C08      | Condition - Eye fatigue - My eyes feel dry.           | Scale (1-5), 5: high fatigue       |
| C09      | Condition - Eye fatigue - My eyes feel bloodshot.     | Scale (1-5), 5: high fatigue       |
| C10      | Condition - Eye fatigue - I feel pain around my eyes. | Scale (1-5), 5: high fatigue       |
| <b>D</b> | <b>Questionnaire posterior to experiments</b>         |                                    |
| D01      | Condition - Comfort                                   | Scale (1-5), 5: uncomfortable      |
| D02      | Condition -Drowsiness                                 | Scale (1-5), 5: drowsiness         |
| D03      | Condition - Concentration                             | Scale (1-5), 5: concentrate        |
| D04      | Condition - Physical condition                        | Scale (1-5), 5: high fatigue       |
| D05      | Condition - Mental condition                          | Scale (1-5), 5: high fatigue       |
| D06      | Condition - Eye fatigue - My eyes feel tired.         | Scale (1-5), 5: high fatigue       |
| D07      | Condition - Eye fatigue - I am blinking a lot.        | Scale (1-5), 5: high fatigue       |
| D08      | Condition - Eye fatigue - My eyes feel dry.           | Scale (1-5), 5: high fatigue       |
| D09      | Condition - Eye fatigue - My eyes feel bloodshot.     | Scale (1-5), 5: high fatigue       |
| D10      | Condition - Eye fatigue - I feel pain around my eyes. | Scale (1-5), 5: high fatigue       |
| <b>E</b> | <b>Experimental factor</b>                            |                                    |
| E01      | Stiffness in the neck                                 | Scale (1-5), 5: high fatigue       |
| E02      | Stiffness in the shoulders                            | Scale (1-5), 5: high fatigue       |
| E03      | Stiffness in the lower back                           | Scale (1-5), 5: high fatigue       |
| E04      | Legs felt stiff or fatigued                           | Scale (1-5), 5: high fatigue       |
| E05      | Felt dazed                                            | Scale (1-5), 5: high fatigue       |
| E06      | Felt sleepy during the experiment                     | Scale (1-5), 5: high fatigue       |
| E07      | Felt dizziness during the experiment                  | Scale (1-5), 5: high fatigue       |
| E08      | Felt nervous or irritable                             | Scale (1-5), 5: high fatigue       |
| E09      | Were you able to focus well on the experiment?        | Well Manipulated (1) to Poorly (5) |
| E10      | Did you feel like you were actively controlling it?   | Well Manipulated (1) to Poorly (5) |
| E11      | Duration of one trial                                 | Short (1) to Long (5)              |
| E12      | Total experiment duration                             | Short (1) to Long (5)              |
| E13      | Software used for the experiment                      | Easy (1) to Difficult (5)          |
| E14      | Environment required for the experiment               | Good (1) to Poor (5)               |

**Table 1.** Questionnaire prior to experiments.



**Fig. 2** The structure of dataset. Each ‘mat’ file contains the data stored under the variable named ‘eeg’. This ‘eeg’ variable is in MATLAB structure format, with individual fields such as ‘eeg.data’, ‘eeg.chan\_locs’, etc., storing the corresponding components of the dataset.

In the context of two multidimensional variables,  $\mathbf{X}$  and  $\mathbf{Y}$ , we are interested in finding linear combinations defined as  $\mathbf{x} = \mathbf{X}^T \boldsymbol{\omega}_x$  and  $\mathbf{y} = \mathbf{Y}^T \boldsymbol{\omega}_y$ , to maximize the correlation between  $\mathbf{X}$  and  $\mathbf{Y}$ . This is achieved by solving the following optimization problem:

$$\max_{\boldsymbol{\omega}_x, \boldsymbol{\omega}_y} \rho(\mathbf{x}, \mathbf{y}) = \frac{E[\mathbf{x}^T \mathbf{y}]}{\sqrt{E[\mathbf{x}^T \mathbf{x}] E[\mathbf{y}^T \mathbf{y}]}}$$

The maximum value of  $\rho$  with respect to  $\boldsymbol{\omega}_x$  and  $\boldsymbol{\omega}_y$  represents the maximum canonical correlation. The projections onto  $\boldsymbol{\omega}_x$  and  $\boldsymbol{\omega}_y$  are known as canonical variates. Here,  $\mathbf{X}$  denotes the set of multi-channel EEG signals, while  $\mathbf{Y}$  represents the corresponding set of reference signals that share the same length as  $\mathbf{X}$ .

$$\mathbf{Y}_{f_N} = \begin{bmatrix} \sin(2\pi f_N t) \\ \cos(2\pi f_N t) \\ \vdots \\ \sin(2\pi N f_N t) \\ \cos(2\pi N f_N t) \end{bmatrix}, t = \frac{1}{f_s}, \frac{2}{f_s}, \dots, \frac{N}{f_s}.$$

**FB-CCA.** Filter bank-CCA (FB-CCA) is widely used in feature extraction for SSVEP-based BCI systems. This method makes it difficult to accurately describe the mutative, complex, and different physiological SSVEP signals. Therefore, it is a suitable approach for massive target improvement in classification performance.

Apply a band-pass filter to the EEG signal ( $\mathbf{X}$ ) to obtain  $N$  sub-bands ( $\mathbf{X}_1$  to  $\mathbf{X}_N$ ). Then, calculate the following values from the vector composed of the coefficients between each  $\mathbf{X}_n$  and the reference signal ( $\mathbf{Y}$ ). The frequency of the reference signals with the maximal  $\rho_k$  is considered to be the frequency of the SSVEPs.

$$\boldsymbol{\rho}_k = \begin{bmatrix} \rho_k^1 \\ \rho_k^2 \\ \rho_k^2 \\ \vdots \\ \rho_k^N \end{bmatrix}, \tilde{\rho}_k = \sum \mathbf{w}(\mathbf{n}) * (\rho_k^n)^2, \mathbf{w}(\mathbf{n}) = \mathbf{n}^{-a} + \mathbf{b},$$

in this study,  $a = 1$ ,  $b = 0$ .

**IT-CCA.** Individual template-CCA (IT-CCA) improves the performance at the expense of additional training with post-stimulus data per subject<sup>34</sup>. One drawback of IT-CCA is excessive reliance on post-stimulus data in the training phase. The combination of CCA and IT-CCA is used for detecting SSVEP signals.

Instead of using the ideal signal ( $\mathbf{Y}$ ) as the reference signal, the classification is performed by utilizing an individual template ( $\tilde{\mathbf{X}}$ ) as the reference. CCA-based spatial filters ( $\boldsymbol{\omega}_{\mathbf{X}\tilde{\mathbf{Y}}}$ ) identify a CCA weight between  $\mathbf{X}$  and  $\mathbf{Y}$ .

$$\tilde{X}_k = \frac{1}{N} \sum \chi_k, \chi \text{ is training trials, } \max_{w_x, w_{\tilde{x}}} \rho(x, \tilde{x}) = \frac{E[x^T \tilde{x}]}{\sqrt{E[x^T x] E[\tilde{x}^T \tilde{x}]}}, \tilde{x} = \tilde{X}^T \omega_{\tilde{x}}$$

$$\rho_k = \begin{bmatrix} \rho(X^T, Y_f) \\ \rho(X^T \omega_{\tilde{X}}, \tilde{X}^T \omega_{\tilde{X}}) \\ \rho(X^T \omega_{\tilde{X}Y}, \tilde{X}^T \omega_{\tilde{X}Y}) \\ \rho(X^T \omega_{\tilde{X}Y}, \tilde{X}^T \omega_{\tilde{X}Y}) \end{bmatrix}, \tilde{\rho} = \sum \text{sign}(\rho_i) * \rho_i^2$$

**TRCA.** Task related component analysis (TRCA) as a user-dependent training method can obtain spatial filters that extract task related source activities from multi-channel EEG signals<sup>35</sup>. It extracts task related components by maximizing their reproducibility during task periods. Therefore, it is assumed that the EEG signal  $\mathbf{x}(t)$  consists of a task-related component  $\mathbf{s}(t)$  and unrelated components  $\mathbf{n}(t)$ , following as:

$$\mathbf{x}_j(t) = \mathbf{a}_1 \mathbf{s}(t) + \mathbf{a}_2 \mathbf{n}(t)$$

$$\mathbf{y}(t) = \sum \omega_j \mathbf{x}_j(t) \text{ and } \sum \omega_j \mathbf{a}_{1j} = \mathbf{1}, \sum \omega_j \mathbf{a}_{2j} = \mathbf{0}$$

To recover the task-related component  $\mathbf{s}(t)$ , the inter-trial covariance maximization is calculated as follows:

$$\sum C_{h1,h2} = \sum \sum \omega_{j1} \omega_{j2} \text{Cov}(\mathbf{x}_{j1}, \mathbf{x}_{j2}) = \omega^T S \omega$$

$$\text{Var}(\mathbf{y}(t)) = \sum \omega_{j1} \omega_{j2} \text{Cov}(\mathbf{x}_{j1}, \mathbf{x}_{j2}) = \omega^T Q \omega$$

The optimal coefficient vector for extracting the task-related component  $\mathbf{s}(t)$  from EEG is as follows:

$$\hat{\omega} = \underset{\omega}{\text{argmax}} \frac{\omega^T S \omega}{\omega^T Q \omega}$$

Frequency detection is performed using the data extracted as  $\mathbf{s}(t)$  from the individual calibration dataset  $\chi_n$  as the reference.

$$\mathbf{r}_n = \rho(X^T \omega_n, \chi_n^T \omega_n)$$

where,  $\rho(a, b)$  indicates the Pearson's correlation analysis between two signals  $a$  and  $b$ .

**EEGNet.** EEGNET (CNN-based model) designed specifically to extract both temporal and spatial features, and yields superior performance in BCI applications<sup>36</sup>.

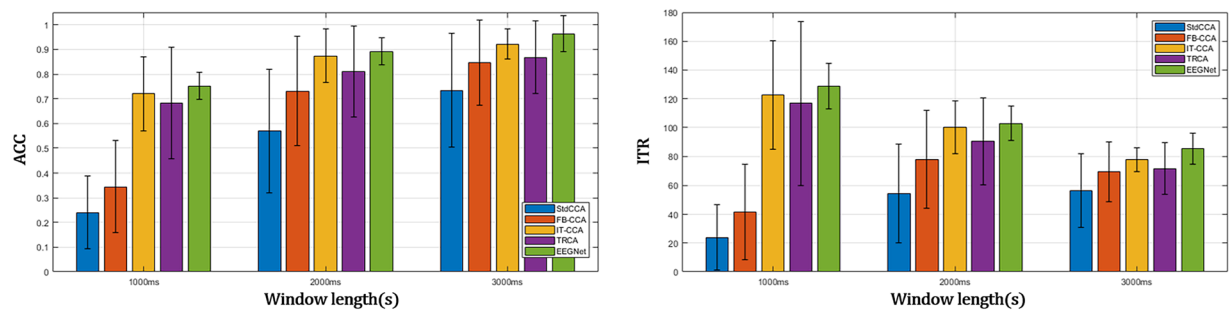
**Preprocessing.** To validate the performance of the SSVEP data, we selected 11 electrodes from the occipital to parietal regions (Oz, O1, O2, POz, PO3, PO4, PO7, PO8, Pz, P3, and P4) for optimal classification<sup>24</sup>. EEG signals were segmented into epochs ranging from 0–1000 ms up to 0–3000 ms relative to stimulus onset, and performance was compared for each subject across all 40 classes and 6 sessions.

Different classification methods employed distinct band-pass filter configurations, while a common notch filter at 59–61 Hz was applied to remove power line noise. The filter configurations were as follows:

- stdCCA: Designed to capture the fundamental and second harmonic components ([13–44] Hz).
- FB-CCA: Utilized a filter bank to extract up to fourth harmonic components, with sub-bands set as [13–89], [27–89], [41–89], and [55–89] Hz.
- IT-CCA, TRCA, EEGNet: Configured to include the fundamental up to the fourth harmonic components ([13–89] Hz).

For the training-based methods (IT-CCA, TRCA, and EEGNet), a 6-fold cross-validation was performed for each subject across all 6 sessions, employing a test-to-training ratio of 1:5.

**Information transfer rate (ITR).** The information transfer rate (ITR) in BCI systems quantifies the efficiency of communication between the user and the computer<sup>37</sup>. Both classification accuracy and ITR were calculated and compared as performance metrics, reflecting the speed and accuracy with which the BCI translates user intentions into commands. The ITR is computed as follows:



**Fig. 3** Classification performance and ITR results for each method were evaluated using epoch lengths of 0–1000 ms, 0–2000 ms, and 0–3000 ms. For ITR(bit/min) calculations, the command time was determined as the sum of the cue duration (500 ms) and the respective epoch length.

$$\text{ITR}(\text{bit}/\text{min}) = \frac{\log_2 N + P \log_2(P) + (1 - P) \log_2\left(\frac{1-P}{N-1}\right)}{T/60}$$

where:

- $N$  is the number of choices or targets (e.g., the number of classes in the SSVEP speller),
- $P$  is the accuracy (probability of correct classification), and
- $T$  is the time per trial (in seconds), defined in this study as the sum of the cue duration (0.5 s) and the stimulation time.

**Frequency properties from stimulation.** The square pulse-based visual stimulation used in this study elicited SSVEP responses. To confirm that the EEG accurately captured the stimulation frequency, we analyzed the correlation between the EEG signal at the fundamental harmonic component and the corresponding stimulus signal (the square pulse) using the Oz electrode. These analyses were conducted in both the time and frequency domains.

**Effect of visual fatigue.** To assess the impact of visual fatigue on EEG recordings during the experiment, we examined the absolute power across five frequency bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–50 Hz). For each session, the first 30 seconds before the start of the experimental trials were segmented into 5-second epochs using a 1-second sliding window to facilitate detailed calculations of spectral power changes.

**SSVEP Performance Evaluation.** SSVEP detection performance was evaluated for 40 participants using both classification accuracy and Information Transfer Rate (ITR). Across varying epoch lengths (1000 ms, 2000 ms, and 3000 ms), the classification accuracies(%) were as follows (refer to Fig. 3, left):

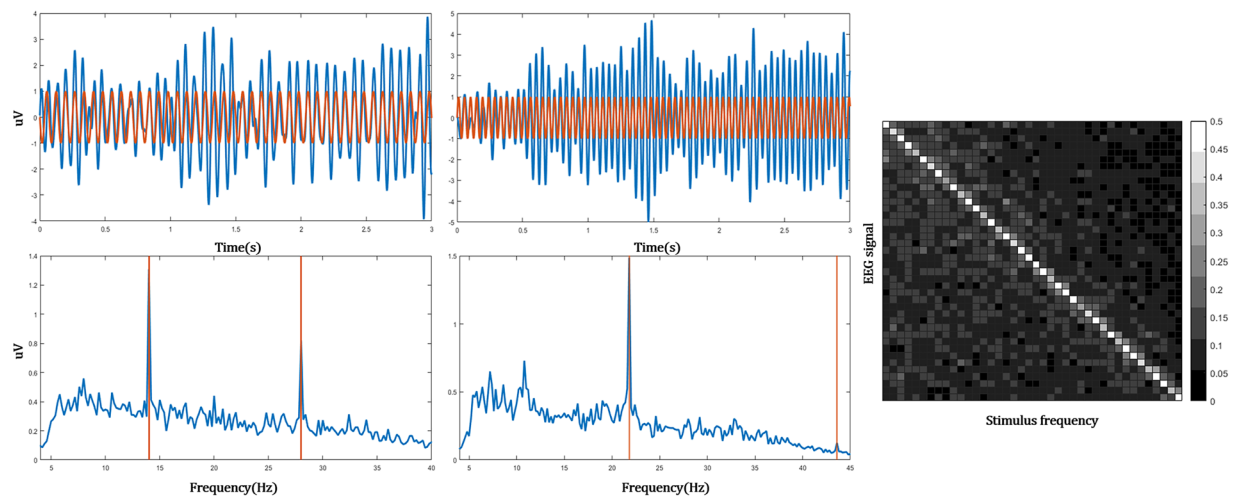
- stdCCA:  $24.06 \pm 14.70\%$ ,  $56.99 \pm 24.90\%$ ,  $73.44 \pm 22.97\%$
- FB-CCA:  $34.44 \pm 18.56\%$ ,  $73.08 \pm 22.17\%$ ,  $84.72 \pm 17.26\%$
- IT-CCA:  $72.05 \pm 14.95\%$ ,  $87.42 \pm 10.79\%$ ,  $92.17 \pm 6.11\%$
- TRCA:  $68.29 \pm 22.73\%$ ,  $81.08 \pm 18.42\%$ ,  $86.84 \pm 14.75\%$
- EEGNet:  $75.23 \pm 5.59\%$ ,  $89.23 \pm 6.78\%$ ,  $96.37 \pm 7.28\%$

The corresponding ITR(bit/min) values are shown in Fig. 3, right:

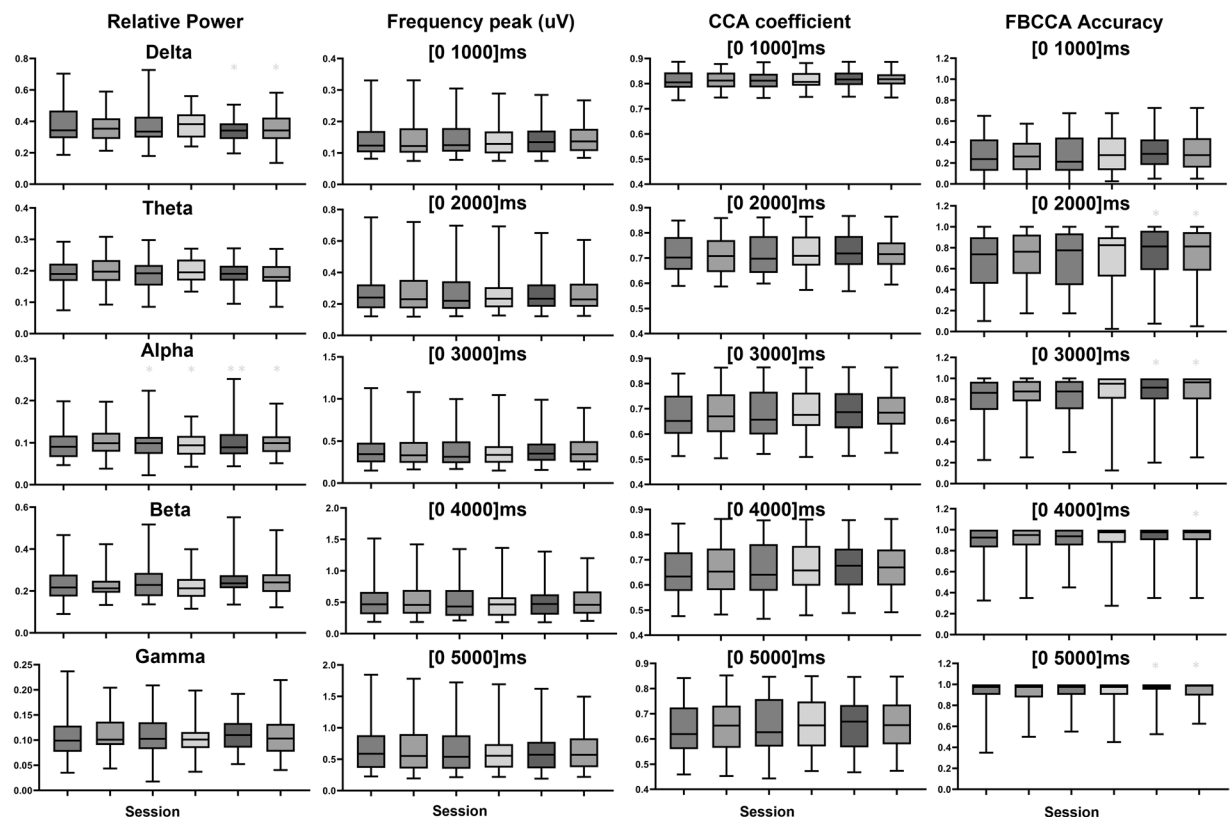
- stdCCA:  $19.67 \pm 17.67$ ,  $57.12 \pm 34.11$ ,  $59.26 \pm 24.50$
- FB-CCA:  $41.60 \pm 33.14$ ,  $77.86 \pm 33.88$ ,  $69.38 \pm 20.62$
- IT-CCA:  $122.67 \pm 37.51$ ,  $100.10 \pm 18.51$ ,  $77.78 \pm 8.01$
- TRCA:  $116.86 \pm 56.96$ ,  $90.51 \pm 30.23$ ,  $71.59 \pm 17.71$
- EEGNet:  $128.69 \pm 15.77$ ,  $102.93 \pm 11.90$ ,  $85.26 \pm 10.60$

Notably, stdCCA achieved the highest ITR with a 3-second window, FB-CCA peaked at a 2-second window, whereas IT-CCA, TRCA, and EEGNet showed optimal performance at a 1-second window. The calibration-based approaches (IT-CCA, TRCA, and EEGNet) demonstrated high-speed detection capabilities, highlighting their potential for efficient, real-time SSVEP-based BCI applications.

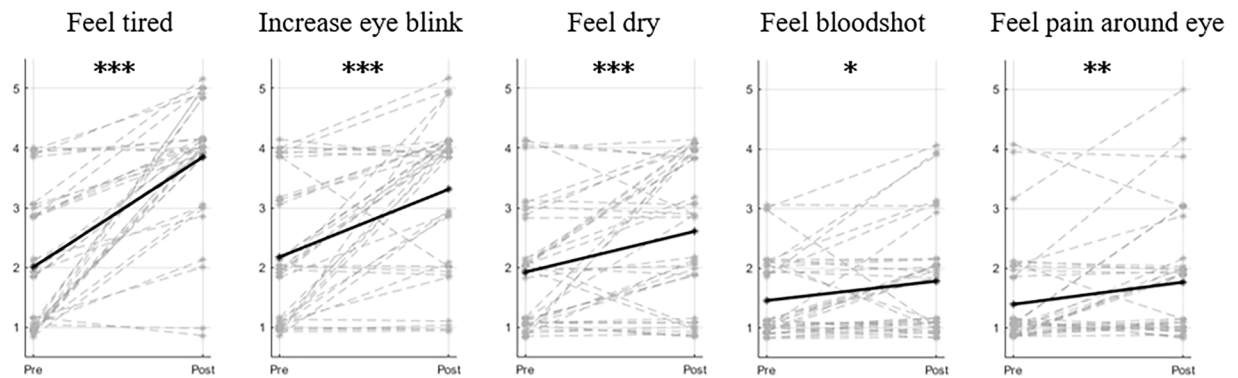
We also evaluated individual performance and assessed cross-subject generalization by validating the IT-CCA and TRCA methods using a leave-one-subject-out (LOSO) cross-validation approach. The performance results are presented below. For more detailed information, please refer to the Supplementary Material 1.



**Fig. 4** The left panel displays the time-domain and frequency-domain session-averaged EEG signal for Subject 1 at the Oz channel, with frequency analysis based on 14 Hz and 21.8 Hz stimuli over a 0–3000 ms window. The right panel presents the subject-averaged correlation between the stimulus and EEG signals.



**Fig. 5** This figure presents relative power, stimulus frequency power, CCA coefficients, and FB-CCA-based classification performance across Sessions 1 to 6. Relative Power; This was calculated across five EEG frequency bands to evaluate spectral changes throughout the sessions. Stimulus Frequency Power; This represents the average peak variations in the SSVEP first harmonic frequency across EEG signals corresponding to each stimulus frequency. CCA Coefficient; This metric measures the correlation between the first harmonic sinusoidal reference and EEG signals across 11 selected channels. Classification Performance; Performance was evaluated using a calibration-free classification method to compare variations in SSVEP decoding across sessions. Statistical comparisons were conducted using paired t-test between Session 1 and other sessions (\* $p < 0.05$ , \*\* $p < 0.01$ ).



**Fig. 6** The survey result shows visual fatigue before and after the experiment, related to items C06 to C10 and D06 to D10 in Table 1. The gray lines represent the changes for each subject, while the black line indicates the subject average. Statistical comparisons were conducted using paired t-test (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.005$ ).

- LOSO templet-CCA:  $31.80 \pm 18.16\%$ ,  $63.30 \pm 23.97\%$ ,  $76.47 \pm 21.81\%$
- LOSO with shifting templet CCA:  $62.21 \pm 26.15\%$ ,  $79.42 \pm 18.78\%$ ,  $88.22 \pm 16.35\%$
- LOSO TRCA:  $65.48 \pm 18.10\%$ ,  $77.01 \pm 19.36\%$ ,  $82.26 \pm 20.68\%$

**Frequency properties from stimulation.** To assess the level of SSVEP elicitation in each subject, a band-pass filter (5–45 Hz) was applied to the Oz electrode recordings, capturing up to the second harmonic components. We analyzed the correlation between the EEG signal and the stimulus signal (a square pulse containing the designated frequency) in both the time and frequency domains. The correlation at the stimulation frequency was  $0.5124 \pm 0.1012$ , in contrast to  $0.0890 \pm 0.0446$  at non-stimulation frequencies, indicating a significant difference (see Fig. 4).

**Effect of visual fatigue.** To investigate the impact of visual fatigue throughout the experiment, we examined session-wise changes in EEG features, focusing on the relative power across five frequency bands: Delta (1–4 Hz), Theta (4–8 Hz), Alpha (8–12 Hz), Beta (12–30 Hz), and Gamma (30–50 Hz). In addition, we analyzed changes in the SSVEP frequency peak, CCA coefficient, and FB-CCA performance across windows ranging from 1000 ms to 5000 ms. To minimize the influence of stimulus-driven responses, the relative power was computed over 30-second intervals during which participants viewed a black screen prior to each session. For detailed analysis, EEG data were segmented into non-overlapping 5-second epochs, and the frequency peak, CCA coefficient, and FB-CCA performance were compared as the window size increased from the onset of stimulation. To further reduce session-dependent calibration effects, classification performance was assessed solely using the calibration-free FB-CCA method Fig. 3.

A paired t-test, comparing Session 1 with subsequent sessions, revealed statistically significant increases in the relative power of the delta and alpha bands during Sessions 5 and 6 (see Fig. 5). Moreover, classification performance showed a statistically significant improvement in the final session. Changes in visual fatigue were further evaluated through survey items (C06–C10, D06–D10) administered before and after the experiment, with participants reporting significant increases in fatigue, particularly regarding tiredness, eye blinking, and eye dryness (see Fig. 6).

### Data availability

All research data has been consented to by participants, with explicit agreement for public sharing. Data sharing follows the appropriate ethical approval procedures to ensure the privacy of participants is fully protected. It is available at [https://figshare.com/articles/dataset/SSVEP\\_40class\\_dataset/28806815/2](https://figshare.com/articles/dataset/SSVEP_40class_dataset/28806815/2).

### Code availability

The GitHub repository [<https://github.com/heegyukim4043/SSVEP-dataset-processing>] contains the preprocessing scripts, classification models, and survey materials used in this study. While specific model parameters and training procedures are not detailed in this paper, they can be fully replicated using the resources provided in the repository. In our analysis, we utilized MATLAB version 2020a, EEGLAB toolbox version 2022.0, FieldTrip toolbox version 20190618, Python 3.6, and TensorFlow 2.2.0. For detailed information about the requirements for each toolbox and library, please refer to the documentation available in the repository.

Received: 28 April 2025; Accepted: 24 September 2025;

Published online: 05 November 2025

### References

1. McFarland, D. J. & Wolpaw, J. R. Brain-computer interfaces for communication and control. *Commun. ACM* **54**, 60–66 (2011).
2. Lebedev, M. A. & Nicolelis, M. A. Brain-machine interfaces: past, present and future. *TRENDS Neurosci.* **29**, 536–546 (2006).

3. Wolpaw, J. R. *et al.* Brain-computer interface technology: a review of the first international meeting. *IEEE Trans. Rehabil. Eng.* **8**, 164–173 (2000).
4. Ang, K. K. & Guan, C. Brain-Computer Interface in Stroke Rehabilitation. *J. Comput. Sci. Eng.* **7**, 139–146 (2013).
5. Nicolas-Alonso, L. F. & Gomez-Gil, J. Brain Computer Interfaces, a Review. *Sensors* **12**, 1211–1279 (2012).
6. Gao, S., Wang, Y., Gao, X. & Hong, B. Visual and auditory brain–computer interfaces. *IEEE Trans. Biomed. Eng.* **61**, 1436–1447 (2014).
7. Kritzman, L. *et al.* Steady-state visual evoked potentials differentiate between internally and externally directed attention. *NeuroImage* **254**, 119133 (2022).
8. Al-Qaysi, Z. T. *et al.* A comprehensive review of deep learning power in steady-state visual evoked potentials. *Neural Comput. Appl.* **36**, 16683–16706 (2024).
9. Zerafa, R., Camilleri, T., Falzon, O. & Camilleri, K. P. To train or not to train? A survey on training of feature extraction methods for SSVEP-based BCIs. *J. Neural Eng.* **15**, 051001 (2018).
10. Hong, J. & Qin, X. Signal processing algorithms for SSVEP-based brain computer interface: State-of-the-art and recent developments. *J. Intell. Fuzzy Syst.* **40**, 10559–10573 (2021).
11. Wang, Y., Chen, X., Gao, X. & Gao, S. A Benchmark Dataset for SSVEP-Based Brain–Computer Interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.* **25**, 1746–1752 (2017).
12. Liu, B., Huang, X., Wang, Y., Chen, X. & Gao, X. BETA: A Large Benchmark Database Toward SSVEP-BCI Application. *Front. Neurosci.* **14** (2020).
13. Lee, M.-H. *et al.* EEG dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy. *GigaScience* **8**, giz002 (2019).
14. Acampora, G., Trinchese, P. & Vitiello, A. A dataset of EEG signals from a single-channel SSVEP-based brain computer interface. *Data Brief* **35**, 106826 (2021).
15. Zhu, F., Jiang, L., Dong, G., Gao, X. & Wang, Y. An Open Dataset for Wearable SSVEP-Based Brain–Computer Interfaces. *Sensors* **21**, 1256 (2021).
16. Gu, M., Pei, W., Gao, X. & Wang, Y. Optimizing Visual Stimulation Paradigms for User-Friendly SSVEP-Based BCIs. *IEEE Trans. Neural Syst. Rehabil. Eng.* **32**, 1090–1099 (2024).
17. Volosyak, I., Gembler, F. & Stawicki, P. Age-related differences in SSVEP-based BCI performance. *Neurocomputing* **250**, 57–64 (2017).
18. Azadi Moghadam, M. & Maleki, A. Fatigue factors and fatigue indices in SSVEP-based brain-computer interfaces: a systematic review and meta-analysis. *Front. Hum. Neurosci.* **17** (2023).
19. Diez, P., Orosco, L., Garcés Correa, A. & Carmona, L. Assessment of visual fatigue in SSVEP-based brain-computer interface: a comprehensive study. *Med. Biol. Eng. Comput.* **62**, 1475–1490 (2024).
20. Zhao, H., Zheng, Q., Ma, K., Li, H. & Zheng, Y. Deep Representation-Based Domain Adaptation for Nonstationary EEG Classification. *IEEE Trans. Neural Netw. Learn. Syst.* **32**, 535–545 (2021).
21. Kamrud, A., Borghetti, B. & Schubert Kabban, C. The Effects of Individual Differences, Non-Stationarity, and the Importance of Data Partitioning Decisions for Training and Testing of EEG Cross-Participant Models. *Sensors* **21**, 3225 (2021).
22. Peng, Y. *et al.* Fatigue Detection in SSVEP-BCIs Based on Wavelet Entropy of EEG. *IEEE Access* **9**, 114905–114913 (2021).
23. Chang, M. H., Baek, H. J., Lee, S. M. & Park, K. S. An amplitude-modulated visual stimulation for reducing eye fatigue in SSVEP-based brain–computer interfaces. *Clin. Neurophysiol.* **125**, 1380–1391 (2014).
24. Choi, G.-Y., Han, C.-H., Jung, Y.-J. & Hwang, H.-J. A multi-day and multi-band dataset for a steady-state visual-evoked potential-based brain-computer interface. *GigaScience* **8**, giz133 (2019).
25. Zheng, X. *et al.* Anti-fatigue Performance in SSVEP-Based Visual Acuity Assessment: A Comparison of Six Stimulus Paradigms. *Front. Hum. Neurosci.* **14** (2020).
26. Tian, P. *et al.* A quantization algorithm of visual fatigue based on underdamped second order stochastic resonance for steady state visual evoked potentials. *Front. Neurosci.* **17** (2023).
27. Xie, J. *et al.* Effects of Mental Load and Fatigue on Steady-State Evoked Potential Based Brain Computer Interface Tasks: A Comparison of Periodic Flickering and Motion-Reversal Based Visual Attention. *PLOS ONE* **11**, e0163426 (2016).
28. Peng, Y. *et al.* Fatigue evaluation using multi-scale entropy of EEG in SSVEP-based BCI. *IEEE Access* **7**, 108200–108210 (2019).
29. Brainard, D. H. The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436 (1997).
30. Renard, Y. *et al.* OpenViBE: An Open-Source Software Platform to Design, Test, and Use Brain–Computer Interfaces in Real and Virtual Environments. *Presence* **19**, 35–53 (2010).
31. Homan, R. W., Herman, J. & Purdy, P. Cerebral location of international 10–20 system electrode placement. *Electroencephalogr. Clin. Neurophysiol.* **66**, 376–382 (1987).
32. Kim, H. *et al.* A 40-class SSVEP speller dataset: Beta range stimulation for low-fatigue BCI applications. *figshare* <https://doi.org/10.6084/m9.figshare.28806815.v2> (2025).
33. Lin, Z., Zhang, C., Wu, W. & Gao, X. Frequency Recognition Based on Canonical Correlation Analysis for SSVEP-Based BCIs. *IEEE Trans. Biomed. Eng.* **53**, 2610–2614 (2006).
34. Nakanishi, M., Wang, Y., Wang, Y.-T., Mitsukura, Y. & Jung, T.-P. A high-speed brain speller using steady-state visual evoked potentials. *Int. J. Neural Syst.* **24**, 1450019 (2014).
35. Nakanishi, M. *et al.* Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis. *IEEE Trans. Biomed. Eng.* **65**, 104–112 (2017).
36. Lawhern, V. J. *et al.* EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *J. Neural Eng.* **15**, 056013 (2018).
37. Wolpaw, J. R., Ramoser, H., McFarland, D. J. & Pfurtscheller, G. EEG-based communication: improved accuracy by response verification. *IEEE Trans. Rehabil. Eng.* **6**, 326–333 (1998).

## Acknowledgements

This work was supported by the National Research Foundation of Korea (No. RS-2024-00361688) and Institute of Information & communications Technology Planning & Evaluation (IITP) (No. 2019-0-01842).

## Author contributions

H. Kim: contributed to the design of the experiment, data collection, software programming, data validation, and preparation of the manuscript. K. Won: contributed to the design of the experiment, data collection, and software programming. M. Ahn: contributed to the design of the experiment, data validation, and preparation of the manuscript. S. C. Jun: contributed to supervision of the project and editing the manuscript.

## Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-025-06032-2>.

**Correspondence** and requests for materials should be addressed to S.C.J.

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

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



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

© The Author(s) 2025