SSVEP-freeTUNE: Sample-wise Domain Gating with Dual-Template Priors for Calibration-Free SSVEP Decoding

Kaisong Hu[1], Leilei Zhao[1], Fengcheng Wu[1], Yuxin Liu[1], Zhiguo Zhang[2], Zhenxi Song[2*]
[1]School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China
[2]School of Intelligence Science and Engineering, Harbin Institute of Technology, Shenzhen, China

***Abstract*—Calibration-free decoding is essential for cross-subject use of steady-state visual evoked potential (SSVEP)-based brain–computer interfaces (BCIs). However, generalization is still constrained by inter-subject variability, likely reflecting insufficient exploitation of the EEG's multi-domain characteristics. To address these challenges, we propose an adaptive domain-gated dual-template neural entrainment framework for calibration-free SSVEP decoding, termed SSVEP-freeTUNE. The gating mechanism adaptively balances spectral, temporal, and spatial representations. In parallel, neural entrainment priors implemented with dual templates combine explicit harmonic constraints and robust cross-subject response morphology, anchoring the model with physiologically grounded cues. At the decision level, frequency-locked correlation scores and fused multi-domain features are jointly integrated, bridging model-driven priors and data-driven representations. Experiments on the Benchmark and BETA datasets under leave-one-subject-out protocols demonstrate that SSVEP-freeTUNE achieves state-of-the-art accuracy and information transfer rate, advancing calibration-free SSVEP BCIs toward practical deployment.**

***Index Terms*—BCI, SSVEP, cross-subject generalization, EEG, adaptive learning**

I. Introduction

Steady-state visual evoked potential (SSVEP)-based brain-computer interfaces (BCIs) are a practical and widely used paradigm for non-invasive neural communication, enabling reliable communication and control of external devices from EEG signals [1]–[4]. Compared with other BCI paradigms, SSVEP-BCIs exhibit three key advantages: a high information transfer rate (ITR) enabled by stable phase-locked responses to periodic stimulation [5], [6], strong robustness to background noise and common physiological artifacts, and low deployment complexity requiring only standard EEG equipment and simple visual stimulation.

Owing to these advantages, SSVEP-BCIs have been successfully applied in various scenarios, including high-speed spelling systems for assistive communication [7]–[11], interactive entertainment and serious games for cognitive training [12]–[15], and assistive device control systems for motor function restoration [16]–[18]. However, their usability in real-world deployment critically depends on minimizing per-user calibration effort.

Despite these advancements, the practical scalability of SSVEP-BCIs is severely limited by inter-subject EEG variability. This variability may arise from differences in neuroanatomy and cortical topology, neural conduction properties, and cognitive states, leading to marked variations in SSVEP response characteristics. Conventional systems thus rely on time-consuming subject-specific calibration (repeated data acquisition, individualized parameter tuning), hindering rapid deployment and large-scale application—making high-performance calibration-free decoding methods essential for advancing practical SSVEP-BCIs.

Existing calibration-free decoders remain challenged by the trade-off between low latency and cross-subject robustness. Traditional linear approaches, such as canonical correlation analysis (CCA) and filter bank CCA (FBCCA) [19]–[21], rely on strict frequency priors and linear correlation assumptions. Although inherently calibration-free, they typically require long observation windows, limiting their suitability for low-latency BCI systems. In contrast, deep learning methods can model complex nonlinear neural responses within short time windows [22]–[26], as demonstrated by representative models such as SSVEPformer, TFF-Former, SSVEP-BiMA, and improved models for cross-subject generalization [27]–[31]. However, these models often exhibit limited cross-subject generalization and underexploit explicit topological cues (beyond implicit channel mixing) reflecting cortical topology and inter-regional coupling. Hybrid approaches (e.g., CCA-Net [32], [33]) attempt to combine frequency priors with deep representations, yet they primarily focus on temporal integration and fail to fully exploit the multidimensional structure of EEG signals. These gaps motivate a unified design that can balance multi-domain evidence per trial and integrate physiologically grounded priors in an interpretable calibration-free manner.

Key insight: Inter-subject variability can shift the relative reliability of spectral, temporal, and spatial evidence, making fixed fusion brittle in calibration-free settings. We therefore design a framework that adaptively balances multi-domain evidence per test sample, is anchored by physiologically grounded entrainment priors, and integrates the two at the

*Corresponding author: Zhenxi Song (songzhenxi@hit.edu.cn). This work was supported by the National Natural Science Foundation of China (Grant No. 62306089) and the Shenzhen Science and Technology Program (Grant No. RCBS20231211090800003 and ZDSYS20230626091203008).

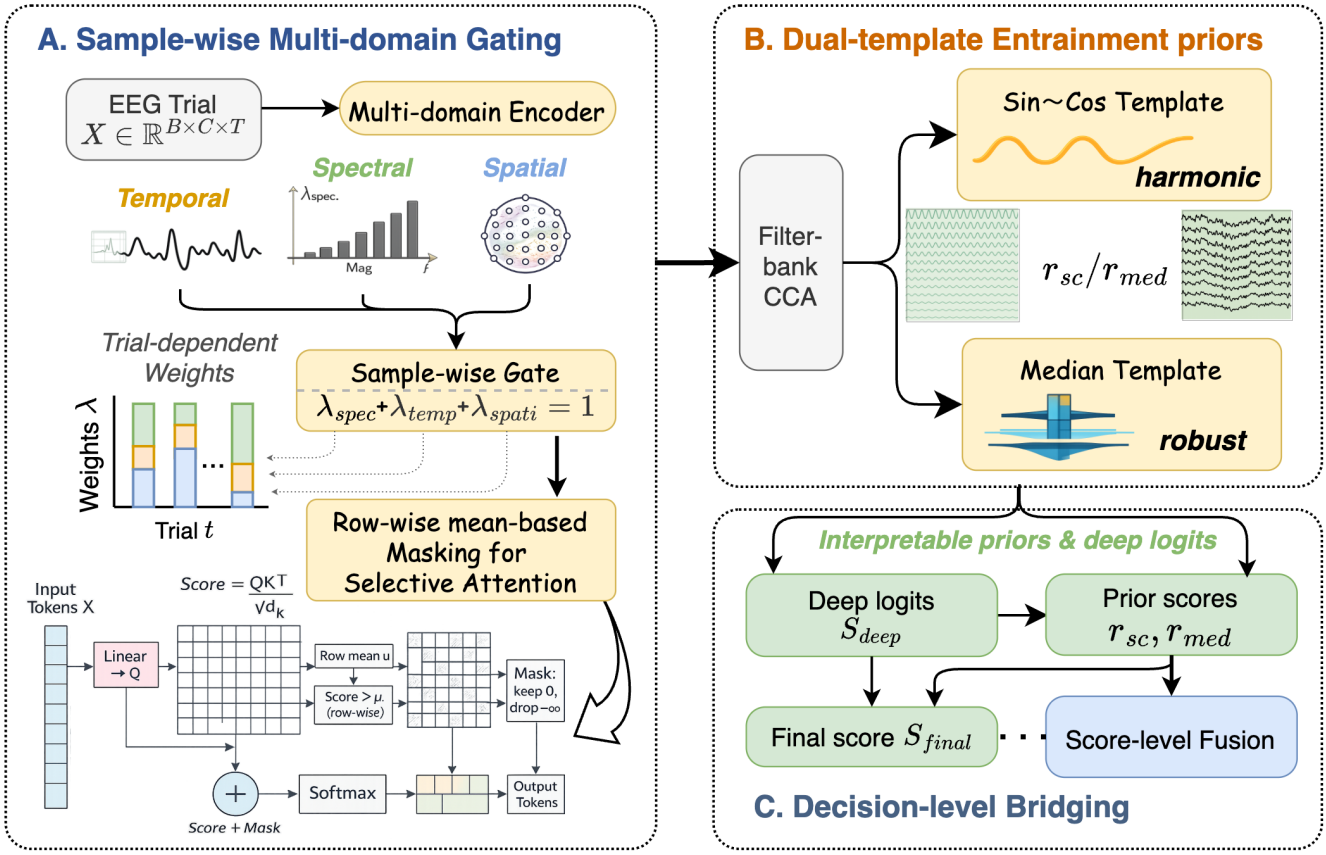

Fig. 1. **Overview of SSVEP-freeTUNE** with three contribution-aligned components. (A) A sample-wise domain gate adaptively fuses spectral/temporal/spatial representations to handle trial-level reliability shifts under cross-subject variability. (B) Dual-template entrainment priors (sin–cos harmonic constraints and median morphology templates) provide frequency-locked, physiologically grounded scores without target calibration. (C) A decision-level bridge combines normalized prior scores with deep logits, preserving interpretability while improving calibration-free decoding.

decision level for interpretability and robustness.

In this work, we propose **SSVEP-freeTUNE**, an adaptive domain-gated **du**al-template **n**eural **e**ntrainment framework for calibration-**free** SSVEP decoding. It unifies sample-wise multi-domain gating, dual-template entrainment priors, and modular decision-level bridging between correlation scores and learned features. Experiments on the Benchmark and BETA datasets under leave-one-subject-out (LOSO) protocols show state-of-the-art accuracy and ITR. The contributions of this work are threefold: (1) ***Sample-wise multi-domain gating*** over spectral, temporal, and spatial representations to improve cross-subject robustness. (2) ***Dual-template entrainment priors*** combining explicit harmonic constraints with a robust cross-subject morphology template, without target-subject calibration. (3) ***Modular decision-level bridging*** that integrates correlation-based priors and deep representations for calibration-free LOSO decoding.

II. Related Work

Calibration-free SSVEP decoding has been advanced along three main directions.

(1) Correlation-based methods exploit explicit frequency priors without user calibration, where CCA-based recognition [19], [20] and FBCCA [21] improve robustness via multi-subband decomposition; however, their linear correlation assumption and reliance on sufficiently long windows can limit performance under short-window or low signal-to-noise ratio (SNR) conditions.

(2) Deep models learn nonlinear representations to boost short-window decoding, ranging from compact CNN backbones [22], [24], [26] and CNN-RNN variants [25] to Transformer designs [27], [28]. While effective, generalization to unseen subjects remains challenging in practice, motivating cross-subject transfer learning [30], [31] and feature-fusion strategies explicitly targeting cross-subject robustness [28], [29].

(3) Hybrid approaches [32], [33] begin to combine frequency-locked scores with learned representations, yet it remains underexplored how to (i) balance spectral, temporal, and spatial evidence on a per-trial basis and (ii) integrate complementary entrainment priors while preserving interpretability. These gaps motivate our SSVEP-freeTUNE, which jointly emphasizes sample-wise multi-domain balancing, ex-

plicit entrainment priors, and modular decision-level bridging to improve zero-shot, calibration-free robustness, particularly under short windows.

III. Methodology

A. Overview

The proposed **SSVEP-freeTUNE** framework (Fig. 1) comprises two parallel streams that are fused at the decision level. The ***adaptive multi-domain stream*** encodes an EEG segment $X \in \mathbb{R}^{B\times C\times T}$ into spectral, temporal, and spatial token sequences via domain-specific embedding and positional encoding, followed by masked self-attention. A sample-wise ***domain gate*** then outputs weights $(\lambda_{spec}, \lambda_{temp}, \lambda_{spati})$ to adaptively fuse the three domains into a unified representation and produce deep classification scores. In parallel, the ***dual-template stream*** instantiates ***neural entrainment priors*** by computing frequency-locked correlation scores between EEG signals and two complementary references (sine-cosine and median-averaged templates) under a filter-bank CCA scheme. Finally, the gated deep representation and the template-based correlation scores are jointly integrated through a lightweight fusion rule and an MLP head, yielding robust ***zero-calibration*** SSVEP decoding across subjects.

B. Multi-domain representations

The adaptive multi-domain module models EEG signals in spectral, temporal, and spatial latent spaces, each producing a token sequence for downstream fusion.

Spectral space. Each EEG segment is transformed from the time domain into the spectral domain via Fast Fourier Transform (FFT). To achieve high spectral resolution, the FFT length is set as $N_{\text{FFT}} = f_s/\Delta f$, where f_s is the sampling rate and $\Delta f = 0.2$ Hz. Zero-padding is applied if necessary, and the normalized complex spectral spectrum is computed. The amplitude spectrum within $[f_{\min}, f_{\max}] = [8, 64]$ Hz is extracted as:

$$\text{Mag} = \sqrt{\text{Re}(X_f)^2 + \text{Im}(X_f)^2} \in \mathbb{R}^{B\times C\times T'} \tag{1}$$

where $T' = (f_{\max} - f_{\min})/\Delta f$ denotes the number of frequency bins.

Temporal space. EEG segments are directly modeled as one-dimensional time series, preserving waveform structure without information loss from frequency transformation.

Spatial space. A Multi-Scale Spatial CNN with four parallel branches captures topological dependencies across EEG channels. For each branch j with kernel size $1 \times k_j$ ($k_j \in 15, 25, 45, 50$), we compute

$$H_j = \sigma(\text{Conv}_{1\times k_j}(X)), \quad j = 1, 2, 3, 4 \tag{2}$$

where $\sigma(\cdot)$ denotes the activation function. Branch outputs are concatenated along the temporal dimension as $H = \text{Concat}(H_1, H_2, H_3, H_4)$. Subsequently, a batch normalization layer and a $C \times 1$ convolution are applied to capture global spatial dependencies across channels, yielding $\text{Conv}_{C\times 1}(\text{BN}(H))$.

Embedding and self-attention. Spectral, temporal, and spatial features are each passed through domain-specific embedding layers, consisting of a 1D convolution followed by normalization, activation, and dropout. To retain sequential structure, sinusoidal positional encodings are added. The enriched representations are then processed by independent self-attention modules to learn domain-specific tokens. For a domain input $X \in \mathbb{R}^{N\times T}$, queries, keys, and values are computed as: $Q = X^\top W_Q$, $K = X^\top W_K$, and $V = X^\top W_V$, where $W_Q, W_K \in \mathbb{R}^{N\times d_k}$ and $W_V \in \mathbb{R}^{N\times d_v}$.

Row-mean masked self-attention. Given the attention logits $\text{Score} = QK^\top/\sqrt{d_k}$, we compute a row-wise threshold μ_i as the mean of the i-th row of Score. This yields a query-adaptive sparsification rule: for each query i, only keys with $\text{Score}_{i,j} \geq \mu_i$ are retained, while the remaining entries are suppressed by adding $-\infty$ before Softmax. Formally, we define an additive mask:

$$\text{Mask}_{i,j} = \begin{cases} 0, & \text{if Score}_{i,j} \geq \mu_i \\ -\infty, & \text{if Score}_{i,j} < \mu_i \end{cases} \tag{3}$$

and compute

$$\text{MaskAttn}(Q, K, V) = \text{Softmax}\left(\frac{QK^\top}{\sqrt{d_k}} + \text{Mask}\right)V \tag{4}$$

This row-wise mean-based masking encourages selective attention and reduces spurious low-confidence interactions in noisy EEG token sequences. After these operations, the three branches yield token sequences denoted as X_{spectral}, X_{temporal}, and X_{spatial}.

C. Sample-wise domain-gated fusion

To account for trial-level reliability shifts across domains under inter-subject variability, we introduce a **sample-wise** domain gate to adaptively fuse spectral, temporal, and spatial evidence. We first concatenate the three token sequences and apply mean pooling to form a global sample-level representation. This representation is fed into a gating network(MLP) that outputs three weights λ_{spec}, λ_{temp}, and λ_{spati}, subject to $\lambda_{spec} + \lambda_{temp} + \lambda_{spati} = 1$. A Softmax normalizes these weights to achieve adaptive balance across domains. The fused representation is computed as:

$$X_{\text{fused}} = \lambda_{spec} \cdot X_{\text{spectral}} + \lambda_{temp} \cdot X_{\text{temporal}} + \lambda_{spati} \cdot X_{\text{spatial}} \tag{5}$$

Finally, X_{fused} is passed through an MLP to produce deep classification scores S_{fused}.

D. Dual-template neural entrainment priors

Neural entrainment priors are modeled through a **dual-template** strategy combined with FBCCA. FBCCA decomposes EEG into multiple sub-bands and performs canonical correlation analysis with reference templates in each sub-band, yielding correlation coefficients that are integrated through weighted fusion. Across sub-bands, FBCCA correlation coefficients are aggregated by a weighted fusion rule to form frequency-locked evidence. The dual-template includes sine–cosine templates, which enforce harmonic frequency

TABLE I
CLASSIFICATION ACCURACY (%) ON **BENCHMARK** UNDER THE LOSO PROTOCOL WITH SIGNAL LENGTHS FROM 0.5 S TO 1.0 S. VALUES ARE MEAN±STD ACROSS TEST SUBJECTS. ASTERISKS INDICATE STATISTICAL SIGNIFICANCE OF OUR METHOD OVER THE BASELINES USING PAIRED t-TESTS WITH FDR CORRECTION ($^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$).

Method	0.5	0.6	0.7	0.8	0.9	1.0
FBCCA	19.36±1.60***	28.13±2.29***	37.93±2.95***	46.71±3.24***	55.01±3.41***	62.42±3.48***
TFF-Former	36.45±6.33***	39.43±5.92***	42.65±5.56***	57.76±6.31***	60.68±7.66***	69.94±7.72***
SSVEP-BiMA	39.39±3.22**	47.89±3.63***	67.43±3.70**	73.12±3.59***	76.30±3.63***	79.36±3.43**
SSVEPformer	50.29±8.64**	58.92±8.15**	67.85±8.77**	74.46±10.13**	79.37±10.27**	81.88±12.53**
CCA-Net	54.27±3.48*	62.52±3.59**	71.33±3.31*	76.99±3.22**	80.70±3.02*	84.02±2.88*
Ours	**57.86±3.29**	**67.86±3.42**	**75.67±3.30**	**81.11±3.22**	**83.75±2.94**	**86.42±2.81**

constraints, and median-averaged templates, which are more robust to noise and outliers than the conventional mean, providing a stable subject-specific representation. This strategy preserves universal frequency priors while enhancing cross-subject consistency, thereby enabling reliable calibration-free decoding. In SSVEP-freeTUNE, we instantiate dual-template entrainment priors: sine-cosine references enforce explicit harmonic constraints, while median-averaged templates provide a robust cross-subject response morphology that is less sensitive to noise and outliers than mean templates. Together, they preserve universal frequency priors while improving cross-subject consistency under calibration-free decoding.

Filter bank. We design 10 sub-band filters covering 6–90 Hz, with passband starts at [6, 14, 22, 30, 38, 46, 54, 62, 70, 78] Hz and a unified cutoff at 90 Hz; the corresponding stopband starts are [4, 10, 16, 24, 32, 40, 48, 56, 64, 72] Hz.

Sine–cosine templates. For each stimulus frequency f_k, templates constructed from the first N_h harmonics are:

$$\mathbf{Y}_{f_k} = \begin{bmatrix} \sin(2\pi f_k t + \phi_k) \\ \cos(2\pi f_k t + \phi_k) \\ \vdots \\ \sin(2\pi N_h f_k t + N_h \phi_k) \\ \cos(2\pi N_h f_k t + N_h \phi_k) \end{bmatrix}, \quad t = \left[\tfrac{1}{f_s}, \tfrac{2}{f_s}, \ldots, \tfrac{N_s}{f_s}\right] \tag{6}$$

where f_s is the sampling rate, N_s is the number of sampling points, and ϕ_k denotes the phase of the k-th stimulus.

TABLE II
DATASET SUMMARY OF **BENCHMARK** AND **BETA**.

Attribute	**Benchmark**	**BETA**
# Subjects	35	70
# Channels	64	64
Task	40-target virtual keyboard	40-target virtual keyboard
Stimulus frequencies	8–15.8 Hz (0.2 Hz step)	8–15.8 Hz (0.2 Hz step)
Phase shift	0.5π	0.5π
Blocks / subject	6 (40 trials / block)	4
Trial length	5 s	2 s (15 subj.) 3 s (55 subj.)
Mean latency	140 ms	130 ms
Recording setting	Controlled lab (higher SNR)	Naturalistic (lower SNR)

Median-averaged templates. To obtain a robust morphology template for each f_k, we compute the median using only the training subjects and their corresponding blocks in each leave-one-subject-out (LOSO) fold, thereby excluding any data from the held-out test subject:

$$\widetilde{Y}_{f_k}(c,t) = \text{median}\left\{X_{s,k,b}(c,t)\right\} \tag{7}$$

where $X_{s,k,b}(c,t)$ is the EEG signal at channel c and time t from subject s, frequency f_k, and block b. Each channel is standardized to mitigate inter-subject scale differences:

$$\widetilde{Y}_{f_k}(c,:) \leftarrow \frac{\widetilde{Y}_{f_k}(c,:) - \mu_c}{\sigma_c} \tag{8}$$

where μ_c and σ_c are the mean and standard deviation of channel c, respectively.

Applying FBCCA with sine–cosine and median-averaged templates yields two correlation scores, $r_{\text{sin–cos}}$ and $r_{\text{med-ave}}$.

E. Modular decision-level bridging

We bridge model-driven priors and data-driven representations at the decision level to preserve interpretability while improving robustness. Specifically, we min–max normalize the two template-based scores and the deep classifier output, yielding $\hat{r}_{\text{sin–cos}}$, $\hat{r}_{\text{med-ave}}$, and $\hat{S}_{\text{fused}}$. The two template-based scores are assigned equal weights because they play the same role as complementary prior terms, while this simple formulation avoids introducing additional weighting hyperparameters. The final fused score is computed as:

$$S_{\text{final}} = \tfrac{1}{2}\left(\hat{r}_{\text{sin–cos}} + \hat{r}_{\text{med-ave}}\right) + \hat{S}_{\text{fused}} \tag{9}$$

The predicted class is then obtained by $\arg\max S_{\text{final}}$.

Unlike feature-level mixing, decision-level fusion keeps correlation scores as explicit frequency-locked evidence, making the contribution of priors transparent under calibration-free evaluation.

IV. EXPERIMENTS AND RESULTS

A. Experimental setup

Datasets. We evaluate SSVEP-freeTUNE on two public SSVEP benchmarks, *Benchmark* and *BETA*, both collected under the Joint Frequency and Phase Modulation paradigm

TABLE III

CLASSIFICATION ACCURACY (%) ON **BETA** UNDER THE LOSO PROTOCOL WITH SIGNAL LENGTHS FROM 0.5 S TO 1.0 S. VALUES ARE MEAN±STD ACROSS TEST SUBJECTS. ASTERISKS INDICATE STATISTICAL SIGNIFICANCE OF OUR METHOD OVER THE BASELINES USING PAIRED t-TESTS WITH FDR CORRECTION ($^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$).

Method	0.5	0.6	0.7	0.8	0.9	1.0
FBCCA	21.29±1.28***	29.15±1.63***	36.96±2.01***	44.04±2.24***	49.83±2.34***	55.92±2.45***
TFF-Former	34.43±4.71***	37.67±5.52***	40.48±4.48***	49.13±6.13***	50.64±6.32***	56.74±6.51***
SSVEP-BiMA	38.57±2.19***	44.77±2.47***	56.35±2.58***	60.32±2.60***	63.61±2.59**	66.99±2.55***
SSVEPformer	51.56±6.52*	57.71±7.03*	63.84±7.10*	67.67±7.43*	71.83±8.96*	74.87±10.25*
CCA-Net	48.89±2.32**	55.76±2.41*	61.22±2.44**	65.33±2.37*	68.51±2.35*	72.01±2.20*
Ours	**52.47±2.41**	**58.36±2.55**	**64.36±2.40**	**68.30±2.35**	**72.21±2.25**	**75.69±2.12**

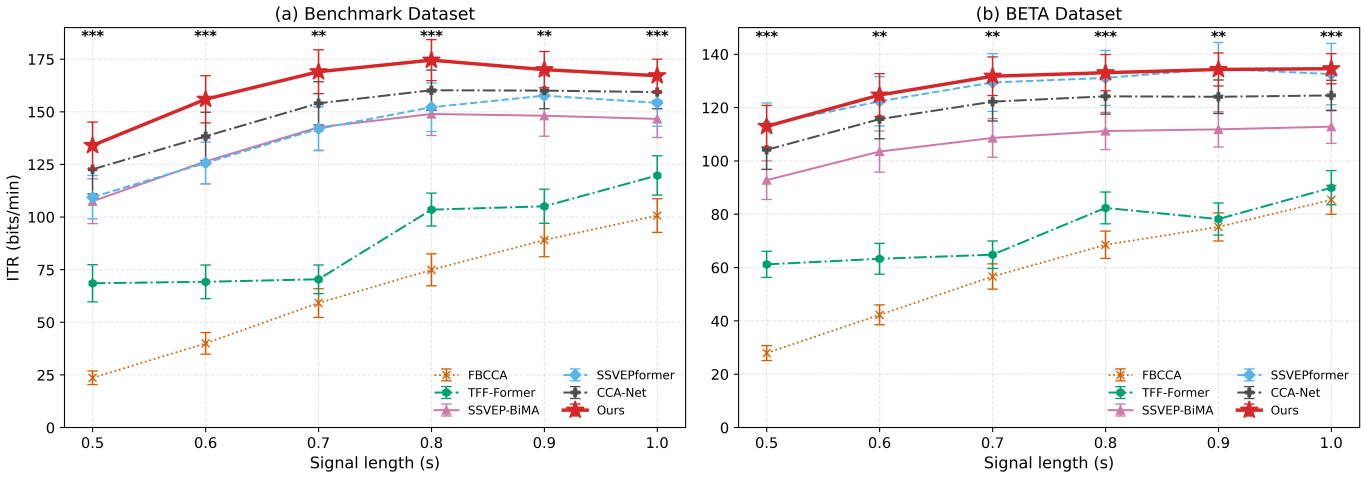

Fig. 2. ITR (bits/min) as a function of signal length (0.5–1.0 s) on (a) Benchmark and (b) BETA under the LOSO protocol. Error bars denote the standard deviation across test subjects. Asterisks indicate statistical significance of our method over the strongest baseline (SSVEPformer) using paired t-tests with FDR correction ($^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$).

to improve target separability and robustness. For a consistent and widely used posterior montage, we report results using nine occipital-parietal channels (Pz, PO3, PO5, PO4, PO6, POz, O1, Oz, and O2), which are known to capture the strongest SSVEP responses. Key dataset statistics are summarized in Table II.

Training and evaluation protocol. To assess *calibration-free* cross-subject generalization, we adopt a leave-one-subject-out (LOSO) protocol: each fold holds out one subject for testing and uses the remaining subjects for training, and results are averaged across folds. We further evaluate multiple signal lengths (0.5 s–1.0 s) to characterize robustness under short-window decoding.

Optimization details. All deep models are trained with cross-entropy loss using AdamW, with a cosine-annealed learning rate from 1×10^{-3} to 1×10^{-5} and dropout set to 0.5. Unless otherwise specified, training runs for 8 epochs on a single NVIDIA RTX 3090 GPU.

B. Experimental results

Baselines and metrics. We compare against representative correlation-based, deep, and hybrid baselines, including FBCCA, SSVEPformer, TFF-Former, SSVEP-BiMA, and CCA-Net. We report both accuracy and ITR (bits/min) [34] to characterize recognition quality and practical communication efficiency, respectively. ITR jointly accounts for the number of targets, the decision accuracy, and the time required per selection, and is therefore more indicative of real-time usability than accuracy alone. Following [34], ITR is computed as:

$$\mathrm{ITR} = \left(\log_2 M + P \log_2 P + (1-P) \log_2 \frac{1-P}{M-1} \right) \cdot \frac{60}{T} \quad (10)$$

where M is the number of targets, P is the classification accuracy, and T (in seconds) is the time required to complete one selection, including gaze and shift time. All results below are reported under the LOSO protocol.

Comparative results under calibration-free LOSO. Tables I and III report the classification accuracy on the *Benchmark* and *BETA* datasets under signal lengths from 0.5 s to 1.0 s. Fig. 2 summarizes the corresponding ITR results under the same settings. As expected, longer windows generally improve all methods by providing richer SSVEP evidence. Across *all* window lengths and on *both* datasets, **SSVEP-freeTUNE** achieves the best accuracy and ITR, demonstrating consistent

calibration-free robustness. The advantage is most pronounced in short windows (0.5–0.7 s), where limited temporal evidence and low SNR make cross-subject decoding particularly challenging.

Statistical significance. We assess significance using paired t-tests with false discovery rate (FDR) correction. ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$). As shown by the asterisks in Tables I, III, and Fig. 2, our method significantly outperforms the strongest baseline (SSVEPformer) across most window lengths on both datasets ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$), confirming that the observed gains are reliable.

TABLE IV
COMPONENT-WISE ABLATION (ACCURACY, %) ON **BENCHMARK** UNDER LOSO WITH A 1.0 S WINDOW, EVALUATING MULTI-DOMAIN REPRESENTATIONS (SPEC/TEMP/SPA), SAMPLE-WISE GATING (AGM), AND DUAL-TEMPLATE PRIORS (S-CT, M-AT).

SPEC	TEMP	SPA	AGM	S-CT	M-AT	Acc
✗	✓	✓	✓	✓	✓	80.83±3.25
✓	✗	✓	✓	✓	✓	83.84±3.12
✓	✓	✗	✓	✓	✓	81.74±3.14
✓	✓	✓	✗	✓	✓	82.48±3.20
✓	✓	✓	✓	✗	✓	84.14±3.27
✓	✓	✓	✓	✓	✗	84.16±3.24
✓	✓	✓	✓	✗	✗	73.49±3.94
✓	✓	✓	✓	✓	✓	**86.42±2.81**

Mapping results to our contributions. (1) The gains under short windows support *sample-wise multi-domain gating* (Contribution 1), suggesting that re-weighting of spectral/temporal/spatial evidence improves robustness under inter-subject shift. (2) The consistent advantage over correlation-only baselines and purely data-driven models highlights the value of *dual-template entrainment priors* (Contribution 2), which provide physiologically grounded anchors under LOSO, especially on the lower-SNR BETA dataset. (3) The concurrent improvements in accuracy and ITR validate *modular decision-level bridging* (Contribution 3), indicating that combining frequency-locked correlation scores with learned multi-domain features is more effective than relying on either alone.

Ablation study. We further verify the role of each proposed component on the Benchmark dataset with a 1.0 s window (Table IV). Removing any of the three domain branches (SPEC/TEMP/SPA) degrades performance, with the largest drops observed when removing *spectral* or *spatial* representations, indicating that frequency-selective cues and explicit topological evidence are particularly discriminative for SSVEP decoding. Importantly, disabling the **adaptive gating mechanism** (AGM) leads to a clear accuracy decrease, directly validating *Contribution 1* that sample-wise domain re-weighting is critical beyond simply extracting multi-domain features. In addition, removing either sine–cosine templates (S-CT) or median-averaged templates (M-AT) consistently hurts performance, while removing both priors causes the most severe degradation (73.49±3.94%), which supports *Contribution 2* that the two templates offer complementary entrainment cues (explicit harmonic constraints vs. robust morphology). Finally, the best result is achieved only when the full model integrates the gated deep score with both template-based correlation scores, evidencing *Contribution 3* that decision-level bridging is necessary to realize the complementarity between priors and learned representations.

V. CONCLUSION

We presented **SSVEP-freeTUNE**, a calibration-free SSVEP decoding framework designed for robust cross-subject deployment under inter-subject variability. **SSVEP-freeTUNE** couples three complementary ingredients: (i) a *sample-wise multi-domain gate* that adaptively fuses spectral, temporal, and spatial representations to handle trial-level reliability shifts, (ii) *dual-template neural entrainment priors* that provide explicit frequency-locked evidence through harmonic-constrained sine–cosine references and robust morphology templates, and (iii) a *modular decision-level bridging* strategy that integrates template-based correlation scores with learned deep scores while preserving interpretability. Extensive LOSO evaluations on the *Benchmark* and *BETA* datasets show that **SSVEP-freeTUNE** consistently achieves state-of-the-art performance in both accuracy and ITR, with particularly strong gains under short time windows and in the lower-SNR BETA setting. Ablation results further verify that each component contributes meaningfully and that combining learned multi-domain features with physiologically grounded priors is critical for reliable calibration-free decoding. Future work will explore extending the framework to broader stimulus conditions and more diverse recording setups, as well as lightweight personalization that maintains the zero-calibration objective.

REFERENCES

[1] X.-Y. Liu, W.-L. Wang, M. Liu, M.-Y. Chen, T. Pereira, D. Y. Doda, *et al.*, "Recent applications of EEG-based brain–computer interfaces in the medical field," *Military Medical Research*, vol. 12, no. 1, p. 14, 2025. URL: https://doi.org/10.1186/s40779-025-00598-z.

[2] M. K. Islam and A. Rastegarnia, "Editorial: Recent advances in EEG (non-invasive) based BCI applications," *Frontiers in Computational Neuroscience*, vol. 17, p. 1151852, 2023. URL: https://doi.org/10.3389/fncom.2023.1151852.

[3] R. Janapati, V. Dalal, and R. Sengupta, "Advances in modern EEG-BCI signal processing: A review," *Materials Today: Proceedings*, vol. 80, pp. 2563–2566, 2023. URL: https://doi.org/10.1016/j.matpr.2021.06.409.

[4] F. Vialatte, M. Maurice, J. Dauwels, and A. Cichocki, "Steady-state visually evoked potentials: Focus on essential paradigms and future perspectives," *Progress in Neurobiology*, vol. 90, no. 4, pp. 418–438, 2010. URL: https://doi.org/10.1016/j.pneurobio.2009.11.005.

[5] R. Abiri, S. Borhani, E. W. Sellers, Y. Jiang, and X. Zhao, "A comprehensive review of EEG-based brain–computer interface paradigms," *Journal of Neural Engineering*, vol. 16, no. 1, p. 011001, 2019. URL: https://doi.org/10.1088/1741-2552/aaf12e.

[6] Y. Wang, R. Wang, X. Gao, B. Hong, and S. Gao, "A practical VEP-based brain–computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 234–240, 2006. URL: https://doi.org/10.1109/TNSRE.2006.875576.

[7] M. Li, D. He, C. Li, and S. Qi, "Brain–Computer Interface Speller Based on Steady-State Visual Evoked Potential: A Review Focusing on the Stimulus Paradigm and Performance," *Brain Sciences*, vol. 11, no. 4, p. 450, 2021. URL: https://doi.org/10.3390/brainsci11040450.

[8] X. Chen, Y. Wang, M. Nakanishi, X. Gao, T.-P. Jung, and S. Gao, "High-speed spelling with a noninvasive brain–computer interface," *Proc. Natl. Acad. Sci. USA*, vol. 112, no. 44, pp. E6058–E6067, 2015. URL: https://doi.org/10.1073/pnas.1508080112.

[9] Erwei Yin, Zongtan Zhou, Jun Jiang, Yang Yu, and Dewen Hu, "A Dynamically Optimized SSVEP Brain–Computer Interface (BCI) Speller," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 6, pp. 1447–1456, 2015. URL: https://doi.org/10.1109/TBME.2014.2320948.

[10] B.-K. Koo, H.-G. Lee, Y. Nam, and S. Choi, "Immersive BCI with SSVEP in VR head-mounted display," in *Proc. IEEE EMBS Conf.*, 2015, pp. 1103–1106. URL: https://doi.org/10.1109/EMBC.2015.7318558.

[11] X. Chen, Z. Chen, S. Gao, and X. Gao, "A high-ITR SSVEP-based BCI speller," *Brain–Computer Interfaces*, vol. 1, nos. 3–4, pp. 181–191, 2014. URL: https://doi.org/10.1080/2326263X.2014.944469.

[12] J. Patiño, I. Vega, M. A. Becerra, E. A. Duque-Grisales, and L. Jiménez, "Integration between serious games and EEG signals: A systematic review," *Applied Sciences*, vol. 15, no. 4, p. 1946, 2025. URL: https://doi.org/10.3390/app15041946.

[13] I. Martišius and R. Damaševičius, "A prototype SSVEP based real-time BCI gaming system," *Computational Intelligence and Neuroscience*, vol. 2016, pp. 1–15, 2016. URL: https://doi.org/10.1155/2016/3861425.

[14] A. Akhtar, J. J. S. Norton, M. Kasraie, and T. Bretl, "Playing checkers with your mind: An interactive multiplayer hardware game platform for brain-computer interfaces," in *Proc. IEEE EMBS*, 2014, pp. 1650–1653. URL: https://doi.org/10.1109/EMBC.2014.6943922.

[15] N. Chumerin, N. V. Manyakov, M. van Vliet, A. Robben, A. Combaz, and M. M. Van Hulle, "Steady-state visual evoked potential-based computer gaming on a consumer-grade EEG device," *IEEE Trans. Comput. Intell. AI Games*, vol. 5, no. 2, pp. 100–110, 2013. URL: https://doi.org/10.1109/TCIAIG.2012.2225623.

[16] X. Chen, B. Zhao, Y. Wang, and X. Gao, "Combination of high-frequency SSVEP-based BCI and computer vision for controlling a robotic arm," *Journal of Neural Engineering*, vol. 16, no. 2, p. 026012, 2019. URL: https://doi.org/10.1088/1741-2552/aaf594.

[17] A. Guneysu and H. L. Akin, "An SSVEP based BCI to control a humanoid robot by using portable EEG device," in *Proc. IEEE EMBS*, 2013, pp. 6905–6908. URL: https://doi.org/10.1109/EMBC.2013.6611145.

[18] A. Nourmohammadi, M. Jafari, and T. O. Zander, "A survey on unmanned aerial vehicle remote control using brain–computer interface," *IEEE Trans. Human-Mach. Syst.*, vol. 48, no. 4, pp. 337–348, 2018. URL: https://doi.org/10.1109/THMS.2018.2830647.

[19] G. Bin, X. Gao, Z. Yan, B. Hong, and S. Gao, "An online multi-channel SSVEP-based brain–computer interface using a canonical correlation analysis method," *Journal of Neural Engineering*, vol. 6, no. 4, p. 046002, 2009. URL: https://doi.org/10.1088/1741-2560/6/4/046002.

[20] Z. Lin, C. Zhang, W. Wu, and X. Gao, "Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2610–2614, Dec. 2006. URL: https://doi.org/10.1109/TBME.2006.886577.

[21] X. Chen, Y. Wang, S. Gao, T.-P. Jung, and X. Gao, "Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface," *J. Neural Eng.*, vol. 12, no. 4, p. 046008, Aug. 2015. URL: https://doi.org/10.1088/1741-2560/12/4/046008.

[22] Y. Pan, N. Li, Y. Zhang, P. Xu, and D. Yao, "Short-length SSVEP data extension by a novel generative adversarial networks based framework," arXiv:2301.05599 [eess.IV], 2023. URL: https://arxiv.org/abs/2301.05599.

[23] J. Wu and J. Wang, "An Analysis of Traditional Methods and Deep Learning Methods in SSVEP-Based BCI: A Survey," *Electronics*, vol. 13, no. 14, p. 2767, 2024. URL: https://doi.org/10.3390/electronics13142767.

[24] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, p. 056013, 2018. URL: https://doi.org/10.1088/1741-2552/aace8c.

[25] Y. Pan, J. Chen, Y. Zhang, Y. Zhang, J. Li, and D. Yao, "An efficient CNN-LSTM network with spectral normalization and label smoothing technologies for SSVEP frequency recognition," *J. Neural Eng.*, vol. 19, no. 5, p. 056014, 2022. URL: https://doi.org/10.1088/1741-2552/ac8dc5.

[26] A. Ravi, N. H. Beni, J. Manuel, and N. Jiang, "Comparing user-dependent and user-independent training of CNN for SSVEP BCI," *J. Neural Eng.*, vol. 17, no. 2, p. 026028, 2020. URL: https://doi.org/10.1088/1741-2552/ab6a67.

[27] J. Chen, Y. Zhang, Y. Pan, P. Xu, and C. Guan, "A transformer-based deep neural network model for SSVEP classification," *Neural Networks*, vol. 164, pp. 521–534, 2023. URL: https://doi.org/10.1016/j.neunet.2023.04.045.

[28] X. Li, W. Wei, S. Qiu, and H. He, "TFF-Former: Temporal-frequency fusion transformer for zero-training decoding of two BCI tasks," in *Proc. ACM Int. Conf. Multimedia*, 2022, pp. 51–59. URL: https://doi.org/10.1145/3503161.3548269.

[29] Y. Liu, Z. Song, G. Xu, Z. Wang, F. Wan, Y. Hu, M. Zhang, and Z. Zhang, "SSVEP-BiMA: Bifocal Masking Attention Leveraging Native and Symmetric-Antisymmetric Components for Robust SSVEP Decoding," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2025, pp. 1–5. URL: https://doi.org/10.1109/ICASSP49660.2025.10890554.

[30] Y. Zhang, S. Q. Xie, C. Shi, J. Li, and Z. Zhang, "Cross-subject transfer learning for boosting recognition performance in SSVEP-based BCIs," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1574–1583, 2023. URL: https://doi.org/10.1109/TNSRE.2023.3250953.

[31] Y. Dai, Z. Chen, T.-A. Cao, H. Zhou, M. Fang, and Y. Dai, *et al.*, "A time-frequency feature fusion-based deep learning network for enhancing cross-subject generalization in SSVEP decoding," *Frontiers in Neuroscience*, vol. 19, p. 1679451, 2025. URL: https://doi.org/10.3389/fnins.2025.1679451.

[32] Y. Deng, Z. Ji, Y. Wang, and S. K. Zhou, "CCA-Net: Zero-shot SSVEP classification via an integration of canonical correlation analysis and deep neural network," *IEEE Trans. Instrum. Meas.*, (Early Access), 2025. URL: https://doi.org/10.1109/TIM.2025.3571139.

[33] Y. Li and T. Kesavadas, "SSVEP-Based Brain–Computer Interface for Part-Picking Robotic Co-Worker," *Journal of Computing and Information Science in Engineering*, vol. 22, no. 2, p. 021001, 2022. URL: https://doi.org/10.1115/1.4051596.

[34] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, T. M. Vaughan, and J. R. Wolpaw, *et al.*, "Brain–computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002. URL: https://doi.org/10.1016/S1388-2457(02)00057-3.