HARMTSFORMER: HARMONIC ENCODING AND EVIDENCE-PRESERVING FUSION FOR UNSEEN-FREQUENCY SSVEP DECODING

Anonymous authors
Paper under double-blind review

ABSTRACT

Brain-computer interfaces based on steady-state visual evoked potentials (SSVEPs) identify attended flashing targets from electroencephalography (EEG) recordings. Decoders typically restrict their output to specific stimulus frequency categories, making it difficult to directly extend their performance to stimulus frequencies not encountered during training or not defined in the model. Although previous studies support unseen-frequency decoding, capturing within-candidate response structure and using information across candidates remain challenges for reliable cross-subject recognition. This paper proposes HarmTSFormer, a two-stage SSVEP decoding architecture. In the first stage, the Dual-Branch Harmonic Encoder (DBHE) constructs harmonic evidence from candidate frequencies and phases. Encoding parameters shared across candidates within each branch enable unified representations of seen and unseen stimuli, while low- and high-order Transformer branches model dependencies among channels, frequency bands, and harmonic responses. Aggregated candidate-set context further informs individual representations. In the second stage, the Evidence-Preserving Aggregation Module (EPAA) combines score normalization with learnable weighting to fuse learned scores and explicit reference-matching scores, integrating EEG response features with reference-matching information. HarmTSFormer uses only source participants' seen-frequency EEG for staged training, without target-user calibration. Evaluations on three public datasets demonstrate its effectiveness under cross-subject, unseen-frequency conditions. Ablation experiments, structural comparisons, and score decomposition analyses further examine the contributions of different information pathways to decoding performance.

1 INTRODUCTION

Electroencephalography (EEG) provides neural signals for non-invasive brain-computer interfaces with high temporal resolution (Lopes da Silva, 2013). Steady-state visual evoked potentials (SSVEPs) contain responses at stimulus frequencies and their harmonics, enabling recognition of attended targets (Norcia et al., 2015; Chen et al., 2015b). Deployment across users and stimulus configurations requires transferring learned discrimination while reducing calibration (Chen et al., 2023; Wang et al., 2023b).

Neural decoders with fixed classification heads assign outputs to predefined stimulus classes. New frequencies outside this set lack corresponding outputs (Figure 1(a)), while collecting additional EEG and retraining increases deployment costs (Waytowich et al., 2018; Chen et al., 2023). Differences between participants further complicate transfer. These challenges have motivated methods that reuse information from source participants and seen stimuli (Mai et al., 2025).

Known stimulus information allows analytic references to be constructed for new candidates, supporting direct EEG matching or template transfer (Lin et al., 2006; Wong et al., 2021). Shared-space correlation and reference-guided attention also enable learned matching for unseen frequencies (Wang et al., 2023b; 2025). However, the ability to score a new candidate does not ensure reliable discrimination across participants. Noise and response variability motivate improvements in how matching evidence is represented and combined. Although some progress has been made,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

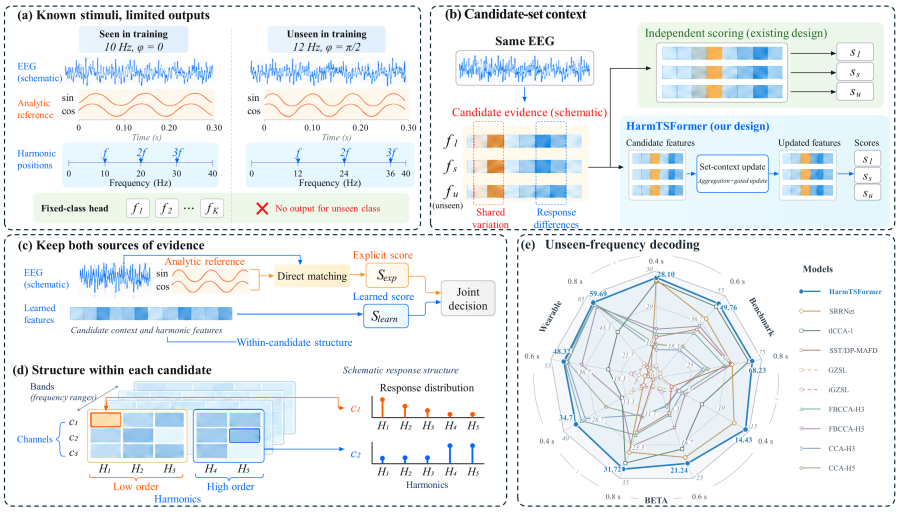

Figure 1: Motivation and overview. (a) A fixed-class head cannot directly represent unseen stimuli. (b) Pooled candidate context updates individual representations. (c) Explicit reference evidence and learned features support the joint decision. (d) Responses have channel, band and harmonic structure. (e) Unseen-frequency decoding across datasets and windows.

achieving reliable decoding of unseen frequencies without the need for user calibration remains a challenge.

Limited Modeling of Within-Candidate Harmonic Structure. Discriminative responses span channels, subbands, and harmonic orders with differing reliability. Filter banks and reference-guided attention exploit aspects of this structure (Chen et al., 2015a; Wang et al., 2025), but scalar matching strengths can obscure internal response patterns. Retaining these features supports joint modeling of their dependencies. Separate low- and high-order branches can accommodate their different characteristics before combining the resulting evidence.

Underutilized Candidate-Set Context. Candidate-wise matching does not directly incorporate other candidates' response features into the current representation. Yet all candidates are evaluated against the same EEG and may share background activity (Zehra et al., 2023; Tan et al., 2025). A summary of the candidate set provides a reference for its overall response pattern, which can inform individual representations when competing candidates have similar matching strengths.

Incomplete Integration of Learned and Explicit Reference Scores. Using references during feature learning does not necessarily preserve direct matching results as a separate decision pathway. Analytic references remain available for unseen frequencies and provide explicit measures of periodic consistency (Lin et al., 2006). Keeping these scores alongside learned scores permits both evidence sources to inform decisions. Their different scales and relative usefulness motivate normalization and learnable weighting.

To address these limitations, we propose HarmTSFormer, a two-stage SSVEP decoder that combines harmonic structure and candidate-set context with explicit reference evidence. In the first stage, the Dual-Branch Harmonic Encoder (DBHE) constructs harmonic evidence from candidate frequencies and phases and encodes it through low- and high-order Transformer branches. Shared parameters across candidates support scoring both seen and unseen frequencies, while gated updates incorporate pooled candidate-set context into individual representations (Figure 1(b), (d)). In the second stage, the Evidence-Preserving Aggregation Module (EPAA) preserves explicit reference-matching scores and fuses them with learned scores through normalization and shared learnable coefficients (Figure 1(c)). HarmTSFormer is trained only on source participants' seen-frequency EEG, without target-user calibration. Our main contributions are summarized as follows:

Dual-Branch Harmonic Encoding with Candidate Relationships. DBHE models harmonic response structure within candidates and incorporates information from the candidate set. A mean-

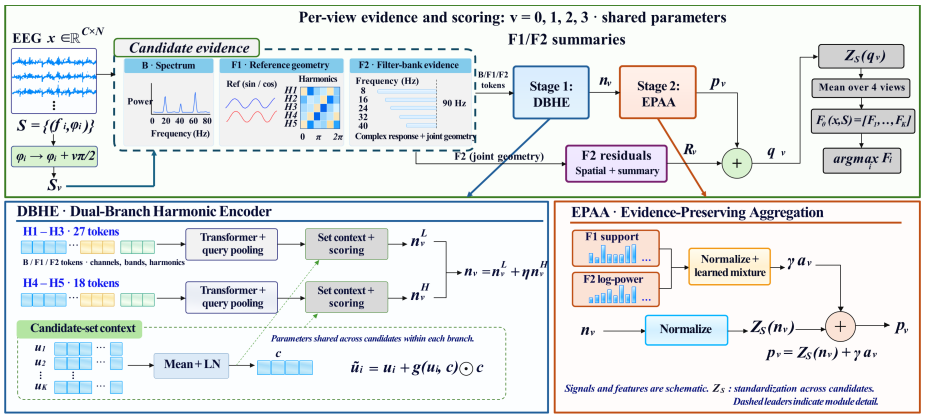

Figure 2: HarmTSFormer. DBHE encodes candidate-conditioned spectral, reference-matching and filter-bank evidence; EPAA fuses learned and explicit scores. F2 spatial/summary residuals precede candidate-axis normalization and four-view averaging. Lower panels detail DBHE, candidate context and EPAA. Here $n_v = e^{(v)}_{\text{learn}}$, n^L_v, n^H_v denote branch scores, and $R_v = R$. Two decoding stages use three training phases.

pooled summary and learnable gates allow the overall response pattern to inform each candidate's representation and learned score.

Score Aggregation Module Preserving Explicit Reference Evidence. EPAA preserves direct matching as a distinct scoring pathway alongside learned evidence. Coordinating their scales and learning their relative contributions enables both sources to participate in decoding, including for frequencies absent from training.

Cross-Subject Evaluation of Unseen-Frequency Recognition. HarmTSFormer achieves the highest mean accuracy in the main baseline comparison across Benchmark, BETA, and Wearable at 0.4, 0.6, and 0.8 s. Ablations, structural comparisons, and score-pathway analyses examine component contributions under this cross-subject, unseen-frequency protocol.

2 Methods

2.1 Model Architecture

HarmTSFormer has two decoding stages (Figure 2): DBHE learns scores from harmonic structure and candidate-set context; EPAA fuses learned and explicit reference scores. Spatial and summary residuals are added before per-view normalization and averaging across four shared-parameter phase views. These stages differ from the three optimization phases (Section 2.4).

2.2 Dual-Branch Harmonic Encoder and Learned Scoring

For EEG $x \in \mathbb{R}^{C\times N}$ (C channels, N samples), $\mathcal{S} = \{(f_i, \phi_i)\}_{i=1}^K$ specifies distinct candidate frequencies and known phases. DBHE produces $e_{\text{learn}} = f_{\text{DBHE}}(x, \mathcal{S}) \in \mathbb{R}^K$.

Harmonic evidence construction. Each phase view combines spectral (B), per-harmonic matching (F1), and subband joint (F2) evidence. B extracts harmonic and neighboring-frequency responses:

$$z^B_{cih\ell} = \mathbf{1}_{\{0<\nu_{ih\ell}<f_s/2\}} e^{-jh\phi_i} \sum_{n=0}^{N-1} x_{cn} w_n e^{-j2\pi\nu_{ih\ell}t_n}, \qquad \nu_{ih\ell} = hf_i + \Delta_\ell. \tag{1}$$

Here f_s is the sampling rate, $t_n = n/f_s$, w_n a normalized Hann window, Δ_ℓ a frequency offset, and $j^2 = -1$. Real, imaginary, and log-power components are standardized separately and concatenated. F1 matches x to a sine–cosine reference $Y_{ih} \in \mathbb{R}^{2\times N}$ at hf_i:

$$W_{ih} = \mathcal{W}(C_x) C_{xY_{ih}} \mathcal{W}(C_{Y_{ih}}), \qquad G_{ih} = W_{ih} W_{ih}^{\mathsf{T}}. \tag{2}$$

The covariance and cross-covariance estimates use row-centered $\widetilde{x}, \widetilde{Y}_{ih}$ and $d_N = \max(N-1, 1)$:

$$C_x = \frac{\widetilde{x}\widetilde{x}^\mathsf{T}}{d_N}, \qquad C_{Y_{ih}} = \frac{\widetilde{Y}_{ih}\widetilde{Y}_{ih}^\mathsf{T}}{d_N}, \qquad C_{xY_{ih}} = \frac{\widetilde{x}\widetilde{Y}_{ih}^\mathsf{T}}{d_N}. \tag{3}$$

For a $d \times d$ covariance matrix A, the regularized inverse square root is

$$\delta(A) = 10^{-4}\max\left(\frac{\mathrm{tr}(A)}{d}, 0\right) + 10^{-6}, \quad A + \delta(A)I_d = U\,\mathrm{diag}(\lambda_j)U^\mathsf{T}, \tag{4}$$

$$\mathcal{W}(A) = U\,\mathrm{diag}\Big(\max(\lambda_j, 10^{-6})^{-1/2}\Big)\,U^\mathsf{T}. \tag{5}$$

Thus $W_{ih} \in \mathbb{R}^{C\times 2}$ and $G_{ih} \in \mathbb{R}^{C\times C}$. F1 has $C^2 + 4$ features: $\mathrm{vec}(G_{ih})$, the two singular values s_1, s_2, $\log(1+s_1^2+s_2^2)$, and $(s_1-s_2)/(s_1+s_2+10^{-6})$. For five subband signals $x^{(b)}$, F2 concatenates complex harmonic-center responses l_{ihb} from Eq. (1) with joint matching features $j_{ib}^{[H]}$ from Eq. (2), using stacked references $Y_i^{[H]} = [Y_{i1}; \ldots; Y_{iH}]$. Joint features comprise the spatial Gram matrix and matching-spectrum summaries (expansion/standardization: Appendix A.2).

Within-candidate encoding. The low-order branch uses H1–H3; the high-order branch selects H4/H5 tokens from H1–H5 evidence, retaining all five harmonics in its joint reference. Projection and positional/type embeddings form D-dimensional tokens. Reliability is fixed to one for H1–H3; high-order evidence is modulated by

$$r_{ih} = \sigma\left(\frac{c - hf_i}{\tau}\right)\left[0.05 + 0.95\sigma\left(b + \sum_{m=1}^{5}\mathrm{softplus}(a_m)\widetilde{v}_{ihm}\right)\right], \quad h \in \{4, 5\}. \tag{6}$$

$$c = 40 + 50\sigma(a_c), \quad \tau = 2 + 18\sigma(a_\tau), \quad \widetilde{\boldsymbol{v}}_{ih} = \mathrm{clip}((\boldsymbol{v}_{ih} - \mu) \oslash \boldsymbol{s}, -6, 6)\,. \tag{7}$$

Here σ is sigmoid, $\oslash$ is elementwise division, $b, a_1, \ldots, a_5, a_c, a_\tau$ are shared learned parameters, and c, τ are in Hz. $\boldsymbol{v}_{ih}$ contains the leading F1 singular value, local log-SNR, temporal phase stability, log canonical energy, and canonical anisotropy. Appendix A.3 gives fixed centers/scales, statistic definitions, initialization, and modulation locations. Reliability is trial-dependent; its parameters use only source participants' seen-frequency EEG. The branches have independent parameters; for $d \in \{L, H\}$ and token matrix T_i^d,

$$H_i^d = \mathrm{Enc}_d(T_i^d), \qquad u_i^d = \mathrm{LN}_{d,p}\big[\mathrm{MHA}_{d,p}(\boldsymbol{\xi}_d, H_i^d, H_i^d)\big]\,. \tag{8}$$

The Transformer models harmonic/subband dependencies; learned query $\boldsymbol{\xi}_d$, multi-head attention (MHA), and layer normalization (LN) pool tokens into $u_i^d \in \mathbb{R}^D$.

Candidate-set context and scoring. The shared context $c^d = \mathrm{LN}_{d,c}(K^{-1}\sum_{j=1}^{K} u_j^d)$ updates candidates through gating:

$$\begin{aligned} \widetilde{u}_i^d &= u_i^d + \sigma\big(\mathrm{MLP}_{d,g}([u_i^d; c^d])\big) \odot c^d, \quad e_{\mathrm{learn},i}^d = \mathrm{MLP}_{d,s}\big(\mathrm{LN}_{d,s}(\widetilde{u}_i^d)\big)\,, \\ e_{\mathrm{learn},i} &= e_{\mathrm{learn},i}^L + \eta e_{\mathrm{learn},i}^H, \qquad\qquad \eta = \sigma(a_\eta). \end{aligned} \tag{9}$$

Here $\odot$ is elementwise multiplication and a_η a learned scalar: r_{ih} modulates evidence, whereas η weights the high-order score. Within-branch parameter sharing gives seen and unseen candidates the same scoring rule.

2.3 Score Fusion and Target Identification

EPAA fuses e_{learn} with explicit scores $e^{F1}, e^{F2} \in \mathbb{R}^K$.

Explicit reference scores. Let $s_{ih,1}$ be the largest singular value of W_{ih}. In F2, $L_i^{(3)}$ averages standardized H1–H3 log power over harmonics, subbands, and channels; $L_i^{(45)}$ averages the corresponding standardized H4/H5 log power weighted by r_{ih}. Low/high front ends standardize per channel over the joint candidate, harmonic, and subband axes using H1–H3/H1–H5 statistics, respectively:

$$e_i^{F1} = \sum_{h=1}^{3} s_{ih,1}^2 + \eta\sum_{h=4}^{5}(r_{ih}s_{ih,1})^2, \qquad e_i^{F2} = (1-\eta)L_i^{(3)} + \eta\frac{3L_i^{(3)} + 2L_i^{(45)}}{5}. \tag{10}$$

These scores bypass the Transformer/readout; the five-harmonic mixture reuses the low-order summary with each front end's normalization.

Score fusion. For $u \in \mathbb{R}^K$, candidate-axis standardization is

$$[Z_s(u)]_i = \frac{u_i - \bar{u}}{\max\{\mathrm{std}_s(u), 10^{-6}\}}. \tag{11}$$

Trial-wise candidate mean $\bar{u}$ and standard deviation (divisor K) express relative evidence and align score scales:

$$a = w_1 Z_s(e^{F1}) + w_2 Z_s(e^{F2}), \qquad p = Z_s(e_{\mathrm{learn}}) + \gamma a. \tag{12}$$

Coefficients are learned per model on source data and shared across trials/candidates, with $w_1 + w_2 = 1$, $w_t \geq 0.1$ for $t \in \{1, 2\}$, and $0.5 \leq \gamma \leq 2$ (Appendix A.4). Component scores remain trial-dependent, including high-order reliability modulation.

Residuals and target identification. Spatial and matching-spectrum features from $j_{ib}^{[5]}$, before reliability weighting/token projection, undergo independent encoding, subband aggregation, and readout to yield $r^{\mathrm{sp}}, r^{\mathrm{sum}}$:

$$R = \lambda_{\mathrm{sp}} Z_s(r^{\mathrm{sp}}) + \lambda_{\mathrm{sum}} Z_s(r^{\mathrm{sum}}), \qquad q = p + R, \tag{13}$$

where $\lambda_t = \mathrm{softplus}(\beta_t)$, $t \in \{\mathrm{sp}, \mathrm{sum}\}$, with learned β_t. Holding EEG fixed, four shared-parameter views apply $\phi_i^{(v)} = (\phi_i + v\pi/2) \bmod 2\pi$, $v \in \{0, 1, 2, 3\}$. Residuals do not use candidate phases, so $q_v = p_v + R$. Complete scores are standardized per view, then averaged:

$$F_\theta(x, \mathcal{S}) = \frac{1}{4} \sum_{v=0}^{3} Z_s(q_v), \qquad \widehat{i} = \underset{1 \leq i \leq K}{\mathrm{argmax}} [F_\theta(x, \mathcal{S})]_i, \tag{14}$$

where θ denotes model parameters and $F_\theta(x, \mathcal{S}) \in \mathbb{R}^K$.

2.4 Training Strategy

Three optimization phases exclude held-out-frequency EEG and use only source-subject EEG; held-out frequencies/phases enter solely as candidate metadata.

Base objective. For batch $\{(x_m, y_m)\}_{m=1}^M$, y_m is the true candidate index and $\ell = \{\ell_{mv}\}_{m,v}$ contains four-view scores $\ell_{mv} \in \mathbb{R}^K$:

$$\mathcal{L}_{\mathrm{base}}(\ell) = \mathcal{L}_{\mathrm{CE}}(\ell) + 0.1\mathcal{L}_{\mathrm{phase}}(\ell) + 0.1\mathcal{L}_{\mathrm{adj}}(\ell). \tag{15}$$

Cross-entropy is averaged over views. $\mathcal{L}_{\mathrm{phase}}$ constrains cross-view distributions, standardized scores, and target–competitor margins; $\mathcal{L}_{\mathrm{adj}}$ penalizes neighboring-frequency margin decreases against a stop-gradient original-view reference.

Low-order branch training. Phase 1 optimizes $\ell_{mv} = e^L_{\mathrm{learn},mv}$, then freezes all low-order parameters.

High-order branch and score-fusion training. Phase 2 uses p_{mv} to train the high-order branch, η, and EPAA. A student-initialized teacher is optimized independently:

$$\begin{aligned} \mathcal{L}_T &= \mathcal{L}_{\mathrm{base}}(p^T) + \mathcal{L}_\gamma(\gamma_T), \\ \mathcal{L}_S &= \mathcal{L}_{\mathrm{base}}(p^S) + \mathcal{L}_\gamma(\gamma_S) + \mathcal{L}_{\mathrm{comp}} + \mathcal{L}_{\mathrm{pres}}. \end{aligned} \tag{16}$$

Here p^T, p^S collect batch four-view scores; $\mathcal{L}_\gamma$ softly penalizes $\gamma < 1$. Constraints use target–competitor margins in $Z_s(e_{\mathrm{learn}})$ and γa. If student explicit evidence favors the target but learned evidence favors a competitor, $\mathcal{L}_{\mathrm{comp}}$ penalizes the negative learned margin. For the same teacher conflict, $\mathcal{L}_{\mathrm{pres}}$ penalizes student explicit-margin decreases relative to the stop-gradient teacher; otherwise, it penalizes learned-margin decreases. Later, both models apply source-subject-grouped adjustments to the high-order evidence front end and EPAA after AdamW, combining virtual-update classification gradients, four-view aggregated-score ranking gradients, and the actual update direction.

Auxiliary residual training. Phase 3 freezes the student and F2 extraction, training only residual pathways/scales with $q_{mv} = \mathrm{sg}(p_{mv}) + R_m$ in Eq. (15), where sg denotes stop-gradient. Inference retains the student and residuals (Eq. (14)). Appendix B provides complete losses, teacher/student and grouped-gradient updates, and stage-wise settings.

Table 1: Overview of the SSVEP datasets and input channels used in this study.

Dataset	Dataset subjects	Channels recorded / used	Targets	Frequency range (Hz)	Trials per subject
Benchmark (Wang et al., 2017)	35	64 / 9	40	8.0-15.8	240
BETA (Liu et al., 2020)	70	64 / 9	40	8.0-15.8	160
Wearable (wet) (Zhu et al., 2021)	102	8 / 8	12	9.25-14.75	120

3 EXPERIMENTS

3.1 DATASETS

SSVEP Datasets. We evaluate Benchmark (Wang et al., 2017), BETA (Liu et al., 2020), and wet-electrode Wearable (Zhu et al., 2021), spanning acquisition conditions and stimulus configurations (Table 1).

Cross-Subject and Unseen-Frequency Partitioning. Disjoint source/evaluation pools contain 16/12 Benchmark, 16/12 BETA, and 16/38 Wearable participants. Appendix D.1–D.3 details participant roles, frequency assignments (Tables 9–10), and selection rules.

3.2 EXPERIMENTAL SETUP

Experimental Configuration. 1. Training Setup. All experiments use an NVIDIA A100 GPU. Each HarmTSFormer training phase runs 1,200 updates with batch size 64 and AdamW weight decay 10^{-4}; learning rates are 2×10^{-3}, 2×10^{-3}, and 10^{-3}. High-order/EPAA and residual training freeze preceding paths; phases inherit model states, reset optimizers, and use final checkpoints. Appendix B.4 (Table 7) gives optimization settings, B.5 links runs to results, and D.3 distinguishes final-checkpoint evaluation from source-only configuration selection.

2. Evaluation Strategy. Training and model selection use only source participants' seen-frequency EEG; evaluation-participant EEG and held-out-frequency EEG are excluded. The five folds are frequency holdouts, not subject folds. Across folds, 38 of 40 Benchmark/BETA frequencies are each held out once while both endpoints always remain in training; all 12 Wearable frequencies are held out once (Appendix D.2, Table 10). Inputs use 250 Hz, 0.4/0.6/0.8-s windows, nine Benchmark/BETA channels, and eight wet Wearable channels. Testing retains all 40, 40, or 12 candidates, respectively, while accuracy is evaluated only on held-out-frequency trials. Contiguous-band, sparse-support, and boundary-frequency tests appear in Appendix F.1–F.3 (Figure 7); phase and candidate-set sensitivity in F.4 (Table 13, Figure 8) and F.5 (Figure 9).

Baselines. Reference-matching baselines are CCA (Lin et al., 2006) and FBCCA (Chen et al., 2015a), each using three or five harmonics. Transfer/learning comparators are tlCCA-1 (Wong et al., 2021), SST/DP-MAFD (Wang et al., 2023c), GZSL (Wang et al., 2023b), iGZSL (Wang et al., 2023a), and SRRNet (Mai et al., 2025). LINEAR is a source-trained linear scorer of ten explicit harmonic features. All methods use the same source-only cross-subject unseen-frequency protocol, participant/frequency splits, channels, windows, and full-candidate evaluation. Fitting, templates, and selection exclude target calibration and held-out-frequency EEG; calibration-dependent methods use source-only adaptations. Selection rationale, adaptations, and LINEAR are detailed in Appendix C.1–C.4; capacity and training-exposure controls in G.3. Appendix H.1 (Table 17) reports candidate-conditioned Transformer (CCT) and harmonic CNN matcher (HCM) configurations/results.

Evaluation Metrics. Recognition Performance. Main tables report accuracy as mean $\pm$ sample SD across three runs; Full and ablations use the same random seeds and final checkpoints. SD measures run variability; Appendix E separately reports participant-level uncertainty. Macro-F1 definitions and supplementary results appear in Appendix H.2 (Table 18). **Computational Efficiency.** Forward-inference and preprocessing-inclusive latency, operation counts, and training costs are reported in Appendix H.3 (Table 19, Figure 11).

Table 2: Unseen-frequency decoding without target-user calibration. Accuracy (0–1) is mean ± sample SD over three repeated runs for all methods.

Methods	0.4 s	0.6 s	0.8 s
Benchmark			
CCA-H3	0.0789 ± 0.0019	0.1860 ± 0.0082	0.3330 ± 0.0037
CCA-H5	0.0936 ± 0.0024	0.2105 ± 0.0049	0.3666 ± 0.0041
FBCCA-H3	0.1235 ± 0.0002	0.3015 ± 0.0087	0.4857 ± 0.0041
FBCCA-H5	0.1363 ± 0.0013	0.3461 ± 0.0099	0.5369 ± 0.0089
tlCCA-1	0.2703 ± 0.0002	0.4776 ± 0.0011	0.6587 ± 0.0009
SST/DP-MAFD	0.1144 ± 0.0037	0.3249 ± 0.0087	0.5296 ± 0.0009
GZSL	0.0339 ± 0.0048	0.0305 ± 0.0072	0.0328 ± 0.0072
iGZSL	0.0296 ± 0.0146	0.0318 ± 0.0225	0.0266 ± 0.0048
SRRNet	0.2747 ± 0.0031	0.3978 ± 0.0021	0.5411 ± 0.0029
LINEAR	0.1016 ± 0.0048	0.2186 ± 0.0034	0.3849 ± 0.0026
HarmTSFormer	$\mathbf{0.2810} \pm 0.0066$	$\mathbf{0.4976} \pm 0.0077$	$\mathbf{0.6823} \pm 0.0052$
BETA			
CCA-H3	0.0356 ± 0.0088	0.0713 ± 0.0078	0.1288 ± 0.0047
CCA-H5	0.0351 ± 0.0098	0.0685 ± 0.0045	0.1343 ± 0.0039
FBCCA-H3	0.0417 ± 0.0002	0.0921 ± 0.0091	0.1990 ± 0.0059
FBCCA-H5	0.0466 ± 0.0053	0.1245 ± 0.0070	0.2379 ± 0.0050
tlCCA-1	0.0738 ± 0.0008	0.1769 ± 0.0006	0.2966 ± 0.0005
SST/DP-MAFD	0.0247 ± 0.0085	0.0877 ± 0.0030	0.1897 ± 0.0021
GZSL	0.0296 ± 0.0113	0.0267 ± 0.0043	0.0289 ± 0.0105
iGZSL	0.0369 ± 0.0062	0.0347 ± 0.0136	0.0322 ± 0.0159
SRRNet	0.1261 ± 0.0016	0.1979 ± 0.0015	0.2588 ± 0.0048
LINEAR	0.0362 ± 0.0010	0.0822 ± 0.0045	0.1727 ± 0.0079
HarmTSFormer	$\mathbf{0.1443} \pm 0.0013$	$\mathbf{0.2124} \pm 0.0031$	$\mathbf{0.3172} \pm 0.0074$
Wearable			
CCA-H3	0.3070 ± 0.0079	0.4656 ± 0.0096	0.5822 ± 0.0050
CCA-H5	0.2164 ± 0.0053	0.4603 ± 0.0002	0.5853 ± 0.0005
FBCCA-H3	0.3224 ± 0.0013	0.4765 ± 0.0014	0.5811 ± 0.0040
FBCCA-H5	0.1544 ± 0.0061	0.3890 ± 0.0081	0.5607 ± 0.0005
tlCCA-1	0.1478 ± 0.0004	0.2510 ± 0.0003	0.3646 ± 0.0012
SST/DP-MAFD	0.0132 ± 0.0070	0.0425 ± 0.0030	0.0842 ± 0.0100
GZSL	0.1070 ± 0.0221	0.0909 ± 0.0584	0.1169 ± 0.0228
iGZSL	0.0632 ± 0.0323	0.1306 ± 0.0301	0.0911 ± 0.0807
SRRNet	0.0553 ± 0.0039	0.0596 ± 0.0033	0.0497 ± 0.0049
LINEAR	0.2993 ± 0.0030	0.4610 ± 0.0087	0.5897 ± 0.0023
HarmTSFormer	$\mathbf{0.3471} \pm 0.0043$	$\mathbf{0.4832} \pm 0.0081$	$\mathbf{0.5969} \pm 0.0039$

3.3 Comparison with Baseline Methods

Among the methods listed in Table 2, HarmTSFormer has the highest mean accuracy in all nine conditions, exceeding the strongest listed baseline by 0.63–2.47 percentage points (pp). At 0.8 s, Benchmark/BETA accuracies of 68.23/31.72% exceed tlCCA-1 (65.87/29.66%) by 2.36/2.06 pp; Wearable at 0.4 s reaches 34.71%, exceeding FBCCA-H3 (32.24%) by 2.47 pp. HarmTSFormer accuracy increases with window length, while relative gains vary by condition.

Gains over LINEAR range from 0.72 to 29.74 pp; on Wearable at 0.8 s, 59.69% versus 58.97% represents a modest advantage. Appendix E (Table 11) reports comparator-specific mean differences, confidence intervals, and counts of participants with improved accuracy. Supplementary tests retain an accuracy advantage over the prespecified comparator under contiguous-band holdouts, sparse fitting support, and targets outside the fitting-frequency range but within the recorded stimulus range, although sparser support reduces accuracy. Common phase offsets cause limited variation; larger candidate-specific phase errors and more competing candidates reduce accuracy, while equal-size composition changes have smaller effects (Appendix F.1–F.5, Table 13, Figures 7–9).

With 0.8-s windows, 40 candidates, and batch size one, median forward/preprocessing-inclusive latency is 65/83 ms per trial. Inference uses 0.80 million parameters regardless of candidate count, while computation increases with it. Three-stage training takes 0.64 GPU-hours per dataset/window/frequency-fold/seed fit (Appendix H.3, Table 19, Figure 11).

Table 3: Ablation and structural comparisons. Accuracy (0–1): mean ± SD.

Configuration	0.4 s	0.6 s	0.8 s
Benchmark			
w/o spatial residual	0.2421 ± 0.0046	0.4481 ± 0.0049	0.6459 ± 0.0097
w/o summary residual	0.1945 ± 0.0083	0.4258 ± 0.0056	0.6097 ± 0.0051
w/o H4/H5 path	0.1805 ± 0.0054	0.4202 ± 0.0076	0.5273 ± 0.0061
w/o F2 path	0.0977 ± 0.0015	0.3463 ± 0.0053	0.4519 ± 0.0020
A1: w/o candidate context	0.1854 ± 0.0027	0.4246 ± 0.0098	0.6185 ± 0.0097
A2: single-branch H1-H5	0.1943 ± 0.0056	0.4533 ± 0.0046	0.6084 ± 0.0078
A4: equal-weight fusion	0.1532 ± 0.0018	0.4574 ± 0.0073	0.6454 ± 0.0067
Full (HarmTSFormer)	$\mathbf{0.2810} \pm 0.0066$	$\mathbf{0.4976} \pm 0.0077$	$\mathbf{0.6823} \pm 0.0052$
BETA			
w/o spatial residual	0.1147 ± 0.0041	0.1829 ± 0.0057	0.2780 ± 0.0054
w/o summary residual	0.1153 ± 0.0032	0.1831 ± 0.0053	0.2750 ± 0.0099
w/o H4/H5 path	0.1093 ± 0.0011	0.1636 ± 0.0080	0.2818 ± 0.0096
w/o F2 path	0.0971 ± 0.0033	0.1324 ± 0.0077	0.1925 ± 0.0037
A1: w/o candidate context	0.1015 ± 0.0019	0.1835 ± 0.0042	0.2874 ± 0.0085
A2: single-branch H1-H5	0.1154 ± 0.0004	0.1743 ± 0.0091	0.2714 ± 0.0030
A4: equal-weight fusion	0.1079 ± 0.0074	0.1843 ± 0.0096	0.2243 ± 0.0013
Full (HarmTSFormer)	$\mathbf{0.1443} \pm 0.0013$	$\mathbf{0.2124} \pm 0.0031$	$\mathbf{0.3172} \pm 0.0074$
Wearable			
w/o spatial residual	0.3026 ± 0.0029	0.4698 ± 0.0064	0.5922 ± 0.0050
w/o summary residual	0.2890 ± 0.0042	0.4624 ± 0.0053	0.5863 ± 0.0051
w/o H4/H5 path	0.3063 ± 0.0022	0.4764 ± 0.0020	0.5832 ± 0.0033
w/o F2 path	0.2553 ± 0.0141	0.4338 ± 0.0073	0.5667 ± 0.0027
A1: w/o candidate context	0.2647 ± 0.0063	0.4479 ± 0.0083	0.5453 ± 0.0011
A2: single-branch H1-H5	0.3142 ± 0.0063	0.4631 ± 0.0044	0.5874 ± 0.0079
A4: equal-weight fusion	0.3044 ± 0.0092	0.4741 ± 0.0010	0.5747 ± 0.0067
Full (HarmTSFormer)	$\mathbf{0.3471} \pm 0.0043$	$\mathbf{0.4832} \pm 0.0081$	$\mathbf{0.5969} \pm 0.0039$

3.4 Ablation Experiments

Table 3 compares ablations and structural variants across all nine conditions. **Dual-branch harmonic encoding.** Removing H4/H5 paths or encoding H1–H5 in one branch with candidate-set context lowers mean accuracy throughout, supporting separate low/high-order modeling. Appendix G.3 (Table 16) gives capacity and training-budget controls. **Subband Joint Evidence (F2).** Removing F2 and dependent pathways lowers accuracy throughout, usually most strongly, supporting the combined subband-response and joint-matching paths. **Candidate-Set Context.** Disabling candidate aggregation/updates while retaining both branches and other scoring paths lowers accuracy throughout, supporting pooled context. **Score Fusion.** With normalization retained, fixing learned/explicit fusion to equal weights (A4) lowers mean accuracy throughout. Appendix G.1–G.2 (Table 14) details fixed/trainable parameters and score/normalization controls. **Auxiliary Residuals.** *Spatial* and *summary residuals* improve accuracy throughout, adding cross-channel structure and matching-spectrum statistics, respectively.

3.5 Analysis of Harmonic Representation and Score Fusion

We examine harmonic-branch readouts/sensitivity and score fusion. Figure 3(a) uses representation readouts, whereas Figure 4 uses decision-score pathways; their accuracies need not coincide. **Utilization of Different Harmonic Orders.** The dual-branch readout exceeds low-order-only accuracy in all nine conditions (Figure 3(a)), supporting additional high-order information. For three target frequencies, perturbing H1–H3 reduces the low-order branch's correct-target margin; the high-order branch also responds near H4/H5, especially H4 (Figure 3(b)). These differentiated sensitivities support joint low/high-order evidence.

Fusion of learned scores and explicit reference scores. Learned scores lead on Benchmark; explicit scores lead at longer BETA/Wearable windows. EPAA exceeds the better individual pathway by 0.77–5.97 pp in all nine conditions (Figure 4). Full further improves accuracy throughout, supporting residual contributions beyond fusion. The comparison of EPAA with trial-adaptive gating



Figure 3: DBHE analysis. (a) H1–H3 and dual-branch readout accuracy. (b) Correct-target margin drop under frequency perturbations at three target frequencies; dashed lines indicate H1–H5. Positive drops denote reduced margins.

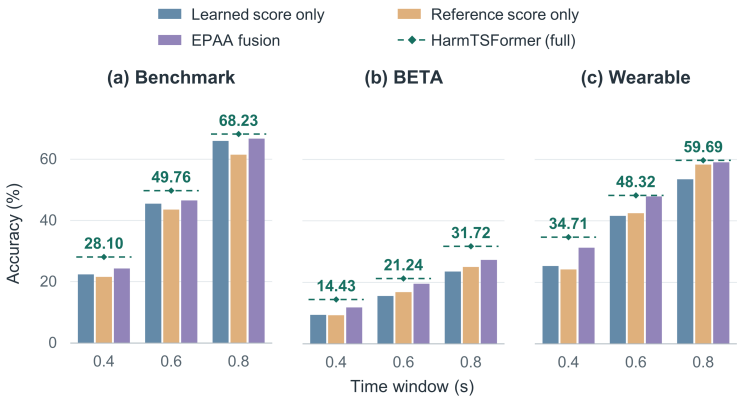

Figure 4: Score pathways on (a) Benchmark, (b) BETA and (c) Wearable. Dashed lines/diamonds indicate full HarmTSFormer with auxiliary residual paths.

is reported in Appendix G.1 (Figure 10); G.2 (Tables 14–15) reports additional fusion controls and weights.

4 Conclusions

HarmTSFormer combines candidate-conditioned DBHE and pooled context with EPAA for unseen-frequency recognition without target-user calibration. Fused scores receive residuals, per-view normalization, and four-view averaging. Three-dataset comparisons, ablations, and representation/score analyses support the architecture under the evaluated offline cross-subject protocols and stimulus configurations. Sharing scoring parameters across candidates allows unseen frequencies to be scored without adding class-specific outputs. Explicit reference matching remains a separate evidence pathway alongside representations learned from low- and high-order harmonic responses. Limitations and future work are discussed in Appendix J.

AI USE STATEMENT

Generative AI tools were used to assist with language editing, document preparation, literature search and discovery, research ideation or execution, and drafting parts of the manuscript and appendix. The authors take full responsibility for the scientific claims, experimental decisions, source verification, and the final manuscript.

ETHICS STATEMENT

This study uses previously released EEG datasets for offline SSVEP decoding research. No new participants were recruited, and no new EEG data were collected. The datasets were used in accordance with their documented consent and data-use conditions, permitted scope of secondary research, and applicable institutional ethics requirements. The evaluation is limited to controlled benchmark conditions and does not establish clinical effectiveness or the safety of an online BCI system. EEG recordings may contain sensitive information; therefore, access, processing, and any redistribution are subject to the applicable dataset terms and privacy protections.

REPRODUCIBILITY STATEMENT

Appendices A–B describe the model operators, objectives, training schedule, and run documentation. Appendix C describes the source-only baseline adaptations. Appendix D specifies participant allocation, frequency folds, candidate definitions, and configuration selection. Appendix E reports accuracy gains and participant-level uncertainty. Appendices F–H present generalization analyses, fusion and structural controls, additional metrics, and computational cost. Appendix I maps these materials to the corresponding main-text sections, and Appendix J discusses limitations and future work. The correspondence between the reproducibility artifacts and the reported experimental runs has been verified. Data sources and the planned artifact release are described in the Data and Code Availability Statement.

Data and Code Availability Statement. The EEG data are from the publicly released Benchmark (Wang et al., 2017), BETA (Liu et al., 2020), and Wearable (Zhu et al., 2021) datasets. Access and redistribution remain subject to the original dataset terms. The participant allocation, frequency folds, input channels, and preprocessing used here are documented in Appendices A and D. The code and associated materials are not included in the initial submission because their preparation for release is still ongoing. We plan to provide these materials in anonymized form during the discussion period, in accordance with conference policies.

REFERENCES

Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. Layer Normalization. arXiv preprint arXiv:1607.06450, 2016. doi: https://doi.org/10.48550/arXiv.1607.06450.

Guangyu Bin, Xiaorong Gao, Zheng Yan, Bo Hong, and Shangkai Gao. An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method. Journal of Neural Engineering, 6(4):046002, 2009. doi: https://doi.org/10.1088/1741-2560/6/4/046002.

Gavin C. Cawley and Nicola L. C. Talbot. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. Journal of Machine Learning Research, 11:2079–2107, 2010. URL https://www.jmlr.org/papers/v11/cawley10a.html.

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, 2015a. doi: https://doi.org/10.1088/1741-2560/12/4/046008.

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 Na-

tional Academy of Sciences, 112(44):E6058–E6067, 2015b. doi: https://doi.org/10.1073/pnas.1508080112.

Jianbo Chen, Yangsong Zhang, Yudong Pan, Peng Xu, and Cuntai Guan. A transformer-based deep neural network model for SSVEP classification. Neural Networks, 164:521–534, 2023. doi: https://doi.org/10.1016/j.neunet.2023.04.045.

Kuan-Jung Chiang, Chun-Shu Wei, Masaki Nakanishi, and Tzyy-Ping Jung. Boosting template-based SSVEP decoding by cross-domain transfer learning. Journal of Neural Engineering, 18(1):016002, 2021. doi: https://doi.org/10.1088/1741-2552/abcb6e.

Janez Demšar. Statistical Comparisons of Classifiers over Multiple Data Sets. Journal of Machine Learning Research, 7:1–30, 2006. URL https://www.jmlr.org/papers/v7/demsar06a.html.

Yang Deng, Zhiwei Ji, Yijun Wang, and S. Kevin Zhou. Cross-subject dual-domain fusion network with task-related and task-discriminant component analysis enhancing one-shot SSVEP classification. arXiv preprint arXiv:2311.07932, 2023. doi: https://doi.org/10.48550/arXiv.2311.07932.

Jean-Antoine Désidéri. Multiple-gradient descent algorithm (MGDA) for multiobjective optimization. Comptes Rendus. Mathématique, 350(5-6):313–318, 2012. doi: https://doi.org/10.1016/j.crma.2012.03.014.

B. Efron. Bootstrap Methods: Another Look at the Jackknife. The Annals of Statistics, 7(1):1–26, 1979. doi: https://doi.org/10.1214/aos/1176344552.

Bradley Efron. Better Bootstrap Confidence Intervals. Journal of the American Statistical Association, 82(397):171–185, 1987. doi: https://doi.org/10.1080/01621459.1987.10478410.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. Proceedings of the 34th International Conference on Machine Learning, Proceedings of Machine Learning Research, 70:1126–1135, 2017. URL https://proceedings.mlr.press/v70/finn17a.html.

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. doi: https://doi.org/10.1109/tbme.2006.889160.

Andrea Frome, Greg S. Corrado, Jonathon Shlens, Samy Bengio, Jeffrey Dean, Marc'Aurelio Ranzato, and Tomas Mikolov. DeViSE: A Deep Visual-Semantic Embedding Model. Advances in Neural Information Processing Systems, 26, 2013. URL https://papers.nips.cc/paper_files/paper/2013/hash/7cce53cf90577442771720a370c3c723-Abstract.html.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On Calibration of Modern Neural Networks. Proceedings of the 34th International Conference on Machine Learning, Proceedings of Machine Learning Research, 70:1321–1330, 2017. URL https://proceedings.mlr.press/v70/guo17a.html.

F.J. Harris. On the use of windows for harmonic analysis with the discrete Fourier transform. Proceedings of the IEEE, 66(1):51–83, 1978. doi: https://doi.org/10.1109/proc.1978.10837.

Dan Hendrycks and Kevin Gimpel. Gaussian Error Linear Units (GELUs). arXiv preprint arXiv:1606.08415, 2016. doi: https://doi.org/10.48550/arXiv.1606.08415.

H. Hotelling. Relations Between Two Sets of Variates. Biometrika, 28(3-4):321–377, 1936. doi: https://doi.org/10.1093/biomet/28.3-4.321.

Vinay Jayaram and Alexandre Barachant. MOABB: trustworthy algorithm benchmarking for BCIs. Journal of Neural Engineering, 15(6):066011, 2018. doi: https://doi.org/10.1088/1741-2552/aadea0.

Vinay Jayaram, Morteza Alamgir, Yasemin Altun, Bernhard Schölkopf, and Moritz Grosse-Wentrup. Transfer Learning in Brain-Computer Interfaces. IEEE Computational Intelligence Magazine, 11(1):20–31, 2016. doi: https://doi.org/10.1109/mci.2015.2501545.

Wei-Bang Jiang, Li-Ming Zhao, and Bao-Liang Lu. Large Brain Model for Learning Generic Representations with Tremendous EEG Data in BCI. International Conference on Learning Representations, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/47393e8594c82ce8fd83adc672cf9872-Abstract-Conference.html.

J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas. On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(3):226–239, 1998. doi: https://doi.org/10.1109/34.667881.

Demetres Kostas, Stéphane Aroca-Ouellette, and Frank Rudzicz. BENDR: Using Transformers and a Contrastive Self-Supervised Learning Task to Learn From Massive Amounts of EEG Data. Frontiers in Human Neuroscience, 15:653659, 2021. doi: https://doi.org/10.3389/fnhum.2021.653659.

Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. Journal of Neural Engineering, 15(5):056013, 2018. doi: https://doi.org/10.1088/1741-2552/aace8c.

Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis, 88(2):365–411, 2004. doi: https://doi.org/10.1016/s0047-259x(03)00096-4.

Juho Lee, Yoonho Lee, Jungtaek Kim, Adam Kosiorek, Seungjin Choi, and Yee Whye Teh. Set Transformer: A Framework for Attention-based Permutation-Invariant Neural Networks. Proceedings of the 36th International Conference on Machine Learning, Proceedings of Machine Learning Research, 97:3744–3753, 2019. URL https://proceedings.mlr.press/v97/lee19d.html.

Zhonglin Lin, Changshui Zhang, Wei Wu, and Xiaorong Gao. Frequency Recognition Based on Canonical Correlation Analysis for SSVEP-Based BCIs. IEEE Transactions on Biomedical Engineering, 53(12):2610–2614, 2006. doi: https://doi.org/10.1109/tbme.2006.886577.

Jiajun Lin, Liyan Liang, Xu Han, Chen Yang, Xiaogang Chen, and Xiaorong Gao. Cross-Target Transfer Algorithm Based on the Volterra Model of SSVEP-BCI. Tsinghua Science and Technology, 26(4):505–522, 2021. doi: https://doi.org/10.26599/tst.2020.9010015.

Bingchuan Liu, Xiaoshan Huang, Yijun Wang, Xiaogang Chen, and Xiaorong Gao. BETA: A Large Benchmark Database Toward SSVEP-BCI Application. Frontiers in Neuroscience, 14:627, 2020. doi: https://doi.org/10.3389/fnins.2020.00627.

Bingchuan Liu, Xiaogang Chen, Nanlin Shi, Yijun Wang, Shangkai Gao, and Xiaorong Gao. Improving the Performance of Individually Calibrated SSVEP-BCI by Task-Discriminant Component Analysis. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 29:1998–2007, 2021. doi: https://doi.org/10.1109/tnsre.2021.3114340.

Fernando Lopes da Silva. EEG and MEG: Relevance to Neuroscience. Neuron, 80(5):1112–1128, 2013. doi: https://doi.org/10.1016/j.neuron.2013.10.017.

Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization. International Conference on Learning Representations, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

F Lotte, L Bougrain, A Cichocki, M Clerc, M Congedo, A Rakotomamonjy, and F Yger. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. Journal of Neural Engineering, 15(3):031005, 2018. doi: https://doi.org/10.1088/1741-2552/aab2f2.

Ximing Mai, Jianjun Meng, Yi Ding, Xiangyang Zhu, and Cuntai Guan. SRRNet: Unseen SSVEP Response Regression From Stimulus for Cross-Stimulus Transfer in SSVEP-BCIs. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 33:1460–1472, 2025. doi: https://doi.org/10.1109/tnsre.2025.3560434.

Gernot R Müller-Putz, Reinhold Scherer, Christian Brauneis, and Gert Pfurtscheller. Steady-state visual evoked potential (SSVEP)-based communication: impact of harmonic frequency components. Journal of Neural Engineering, 2(4):123–130, 2005. doi: https://doi.org/10.1088/1741-2560/2/4/008.

Masaki Nakanishi, Yijun Wang, Yu-Te Wang, and Tzyy-Ping Jung. A Comparison Study of Canonical Correlation Analysis Based Methods for Detecting Steady-State Visual Evoked Potentials. PLOS ONE, 10(10):e0140703, 2015. doi: https://doi.org/10.1371/journal.pone.0140703.

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. doi: https://doi.org/10.1109/tbme.2017.2694818.

Anthony M. Norcia, L. Gregory Appelbaum, Justin M. Ales, Benoit R. Cottereau, and Bruno Rossion. The steady-state visual evoked potential in vision research: A review. Journal of Vision, 15(6):4, 2015. doi: https://doi.org/10.1167/15.6.4.

Bernardino Romera-Paredes and Philip Torr. An embarrassingly simple approach to zero-shot learning. Proceedings of the 32nd International Conference on Machine Learning, Proceedings of Machine Learning Research, 37:2152–2161, 2015. URL https://proceedings.mlr.press/v37/romera-paredes15.html.

Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for EEG decoding and visualization. Human Brain Mapping, 38(11):5391–5420, 2017. doi: https://doi.org/10.1002/hbm.23730.

Ozan Sener and Vladlen Koltun. Multi-Task Learning as Multi-Objective Optimization. Advances in Neural Information Processing Systems, 31, 2018. URL https://papers.nips.cc/paper/2018/hash/432aca3a1e345e339f35a30c8f65edce-Abstract.html.

Yonghao Song, Qingqing Zheng, Bingchuan Liu, and Xiaorong Gao. EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31:710–719, 2023. doi: https://doi.org/10.1109/tnsre.2022.3230250.

Ping Tan, Fengsheng Wang, Kaijun Zhou, and Yi Shen. VMD-FBCCA classification method for SSVEP brain–computer interfaces. Brain-X, 3(1):e70014, 2025. doi: https://doi.org/10.1002/brx2.70014.

Gaël Varoquaux. Cross-validation failure: Small sample sizes lead to large error bars. NeuroImage, 180:68–77, 2018. doi: https://doi.org/10.1016/j.neuroimage.2017.06.061.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. Advances in Neural Information Processing Systems, 30, 2017. URL https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

François-Benoît Vialatte, Monique Maurice, Justin Dauwels, and Andrzej Cichocki. Steady-state visually evoked potentials: Focus on essential paradigms and future perspectives. Progress in Neurobiology, 90(4):418–438, 2010. doi: https://doi.org/10.1016/j.pneurobio.2009.11.005.

Yijun Wang, Xiaogang Chen, Xiaorong Gao, and Shangkai Gao. A Benchmark Dataset for SSVEP-Based Brain–Computer Interfaces. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 25(10):1746–1752, 2017. doi: https://doi.org/10.1109/tnsre.2016.2627556.

Ze Wang, Chi Man Wong, Agostinho Rosa, Tao Qian, and Feng Wan. Adaptive Fourier Decomposition for Multi-Channel Signal Analysis. IEEE Transactions on Signal Processing, 70:903–918, 2022. doi: https://doi.org/10.1109/tsp.2022.3143723.

Xietian Wang, Aiping Liu, Le Wu, Ling Guan, and Xun Chen. Improving Generalized Zero-Shot Learning SSVEP Classification Performance From Data-Efficient Perspective. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31:4135–4145, 2023a. doi: https://doi.org/10.1109/tnsre.2023.3324148.

Xietian Wang, Aiping Liu, Le Wu, Chang Li, Yu Liu, and Xun Chen. A Generalized Zero-Shot Learning Scheme for SSVEP-Based BCI System. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31:863–874, 2023b. doi: https://doi.org/10.1109/tnsre.2023.3235804.

Ze Wang, Chi Man Wong, Agostinho Rosa, Tao Qian, Tzyy-Ping Jung, and Feng Wan. Stimulus-Stimulus Transfer Based on Time-Frequency-Joint Representation in SSVEP-Based BCIs. IEEE Transactions on Biomedical Engineering, 70(2):603–615, 2023c. doi: https://doi.org/10.1109/tbme.2022.3198639.

Xietian Wang, Aiping Liu, Heng Cui, Xingui Chen, Kai Wang, and Xun Chen. GZSL-Lite: A Lightweight Generalized Zero-Shot Learning Network for SSVEP-Based BCIs. IEEE Transactions on Biomedical Engineering, 72(9):2840–2850, 2025. doi: https://doi.org/10.1109/tbme.2025.3553204.

Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Haiteng Jiang, Shijian Li, Tao Li, and Gang Pan. CBraMod: A Criss-Cross Brain Foundation Model for EEG Decoding. International Conference on Learning Representations, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/bbbd6d915cb90be21c1254a82d45cedd-Abstract-Conference.html.

Nicholas Waytowich, Vernon J Lawhern, Javier O Garcia, Jennifer Cummings, Josef Faller, Paul Sajda, and Jean M Vettel. Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials. Journal of Neural Engineering, 15(6):066031, 2018. doi: https://doi.org/10.1088/1741-2552/aae5d8.

Chi Man Wong, Ze Wang, Agostinho C. Rosa, C. L. Philip Chen, Tzyy-Ping Jung, Yong Hu, and Feng Wan. Transferring Subject-Specific Knowledge Across Stimulus Frequencies in SSVEP-Based BCIs. IEEE Transactions on Automation Science and Engineering, 18(2):552–563, 2021. doi: https://doi.org/10.1109/tase.2021.3054741.

Yongqin Xian, Christoph H. Lampert, Bernt Schiele, and Zeynep Akata. Zero-Shot Learning—A Comprehensive Evaluation of the Good, the Bad and the Ugly. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(9):2251–2265, 2019. doi: https://doi.org/10.1109/tpami.2018.2857768.

Chaoqi Yang, M Westover, and Jimeng Sun. BIOT: Biosignal Transformer for Cross-data Learning in the Wild. Advances in Neural Information Processing Systems, 36:78240–78260, 2023. doi: https://doi.org/10.52202/075280-3420.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabás Póczos, Ruslan Salakhutdinov, and Alexander J. Smola. Deep Sets. Advances in Neural Information Processing Systems, 30, 2017. URL https://papers.nips.cc/paper/2017/hash/f22e4747da1aa27e363d86d40ff442fe-Abstract.html.

Syeda R. Zehra, Jing Mu, Anthony N. Burkitt, and David B. Grayden. Effect of alpha range activity on SSVEP decoding in brain-computer interfaces. Proceedings of the 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 1–4, 2023. doi: https://doi.org/10.1109/embc40787.2023.10340956.

Yu Zhang, Guoxu Zhou, Jing Jin, Xingyu Wang, and Andrzej Cichocki. Frequency Recognition in SSVEP-Based BCI Using Multiset Canonical Correlation Analysis. International Journal of Neural Systems, 24(4):1450013, 2014. doi: https://doi.org/10.1142/s0129065714500130.

Yangsong Zhang, Daqing Guo, Fali Li, Erwei Yin, Yu Zhang, Peiyang Li, Qibin Zhao, Toshihisa Tanaka, Dezhong Yao, and Peng Xu. Correlated Component Analysis for Enhancing the Performance of SSVEP-Based Brain-Computer Interface. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 26(5):948–956, 2018a. doi: https://doi.org/10.1109/tnsre.2018.2826541.

Yangsong Zhang, Daqing Guo, Fali Li, Erwei Yin, Yu Zhang, Peiyang Li, Qibin Zhao, Toshihisa Tanaka, Dezhong Yao, Peng Xu, and Masaki Nakanishi. Correction to "Correlated Component Analysis for Enhancing the Performance of SSVEP-Based Brain-Computer Interface". IEEE Transactions on Neural Systems and Rehabilitation Engineering, 26(8):1645–1646, 2018b. doi: https://doi.org/10.1109/tnsre.2018.2851318.

Danhua Zhu, Jordi Bieger, Gary Garcia Molina, and Ronald M. Aarts. A Survey of Stimulation Methods Used in SSVEP-Based BCIs. Computational Intelligence and Neuroscience, 2010:702357, 2010. doi: https://doi.org/10.1155/2010/702357.

Fangkun Zhu, Lu Jiang, Guoya Dong, Xiaorong Gao, and Yijun Wang. An Open Dataset for Wearable SSVEP-Based Brain-Computer Interfaces. Sensors, 21(4):1256, 2021. doi: https://doi.org/10.3390/s21041256.

A Detailed Model Specification

A.1 Notation and Tensor Shapes

For a single EEG trial, x has C channels and N samples. The supplied candidate table S contains K distinct frequencies and their stimulus phases. Candidate indices are local indices into this table; they are not a learned vocabulary of target identities. At a sampling rate of 250 Hz, the 0.4, 0.6, and 0.8 s windows contain 100, 150, and 200 samples, respectively. Benchmark and BETA use C = 9, whereas the wet-electrode Wearable input uses C = 8. The complete evaluation candidate table has K = 40 for Benchmark and BETA and K = 12 for Wearable. Section 3.2 summarizes the frequency coverage and candidate counts. Appendix D.2 (Table 10) specifies the fold assignments and candidate ordering used in the reported experiments.

The low-order branch constructs evidence using H = 3 harmonics. The high-order branch first constructs H = 5 evidence, then emits only the fourth- and fifth-harmonic tokens. The latter branch therefore has 18 tokens per candidate, but its joint F2 reference still contains all five harmonics. The two branches have independent parameters and each uses a 128-dimensional representation. Parameters are shared across candidates within a branch. There is no active target-ID embedding or learned frequency–phase code in the registered implementation; physical frequencies and phases instead enter the evidence operators.

The 27 low-order tokens and 18 high-order tokens are processed by separate encoders. They are not concatenated into one 45-token Transformer. Within each encoder, attention operates over the evidence tokens of one candidate; cross-candidate information is subsequently introduced by symmetric candidate-set pooling. Table 4 summarizes the single-trial shapes, omitting the batch axis.

Table 4: Evidence and representation dimensions for one EEG trial.

Quantity	**Shape or dimension**	**Meaning**
EEG x	$C \times N$	Channels $\times$ samples
Base spectral feature	$K \times H \times 3 \times 3C$	Three frequency offsets; real, imaginary, and log-power channel features
F1 feature	$K \times H \times (C^2 + 4)$	Spatial Gram matrix plus four matching summaries
F2 joint feature	$K \times 5 \times (C^2 + 3H + 3)$	Spatial geometry and joint harmonic summaries in five subbands
Expanded F2 feature	$K \times H \times 5 \times (3C + C^2 + 3H + 3)$	Local complex features concatenated with joint features
Low-order token matrix	$K \times 27 \times 128$	9 base + 3 F1 + 15 F2 tokens
High-order token matrix	$K \times 18 \times 128$	6 base + 2 F1 + 10 F2 tokens, after H4/H5 selection
Pooled branch representation	$K \times 128$	One representation per candidate and branch
Residual spatial / summary input	$K \times 5 \times C(C+1)/2$; $K \times 5 \times 18$	Upper triangular spatial geometry and H = 5 summaries
Per-view / final score	K / K	Candidate-aligned score vectors

A.2 Harmonic Evidence Construction and Standardization

Base spectral evidence uses complex demodulation at three offsets, -0.12, 0, and +0.12 Hz, around each harmonic center. The nonperiodic Hann window is normalized by its sum, and frequencies outside the open interval (0, fs/2) are masked. A phase rotation $\exp(-\mathrm{j}\, h\, \phi i)$ aligns each response with

the supplied candidate phase. Real responses, imaginary responses, and log(1 + squared magnitude) are standardized separately. For each trial and channel, the normalization axis combines candidates, harmonics, and offsets. The resulting three channel vectors are concatenated, linearly projected to dimension 128, and followed by LayerNorm, GELU, and dropout. Harmonic and offset embeddings are added to distinguish the token types.

F1 uses a two-row sine–cosine basis for each candidate harmonic. EEG and references are centered along time, and covariance matrices use the divisor $\max(N - 1, 1)$. Whitening uses a regularized inverse square root: the ridge is 10^{-4} times the nonnegative-clamped mean diagonal of the covariance, plus 10^{-6}; eigenvalues of the regularized covariance are additionally floored at 10^{-6} before inverse square rooting. With W denoting the whitened cross-covariance, the spatial feature is $\mathrm{vec}(WW^{\mathrm{T}})$. Its four accompanying summaries are the first and second singular values, log(1 + the sum of squared singular values), and the anisotropy $(s1 - s2)/(s1 + s2 + 10^{-6})$. Thus each F1 feature has $C^2 + 4$ entries. A bias-free linear projection, LayerNorm, GELU, and dropout produce its token before harmonic and evidence-type embeddings are added.

F2 uses five fourth-order Butterworth subbands: 8–90, 16–90, 24–90, 32–90, and 40–90 Hz. In the implementation, the squared magnitude of each filter response is precomputed at the window's rFFT frequencies, divided by its sampled peak, and multiplied into the EEG spectrum, followed by an inverse rFFT. This is the implemented window-level filtering operator. Complex responses at the harmonic centers are extracted from each filtered signal with the same normalized Hann window and phase rotation. Their real, imaginary, and log-power components are standardized separately for each trial and channel over the combined candidate, harmonic, and subband axes.

For the joint part of F2, the sine–cosine reference pairs for harmonics 1 through H are stacked into a 2H-row reference matrix. Whitening is performed jointly against this matrix within each subband. Let the resulting $C \times 2H$ matrix be W. The feature contains $\mathrm{vec}(WW^{\mathrm{T}})$, the 2H square roots of the nonnegative eigenvalues of $W^{\mathrm{T}}W$ in descending order (including structural zeros when $C < 2H$), the log-transformed energies of the H two-column harmonic blocks, log-transformed total joint strength, spectral dominance, and normalized effective rank. The last two quantities are calculated from the normalized squared singular values. The joint feature has $C^2 + 3H + 3$ entries and is repeated along the harmonic axis before concatenation with the 3C local spectral features. It is then projected to dimension 128, followed by LayerNorm, GELU, dropout, and harmonic, subband, and evidence-type embeddings.

The feature normalization described above and the score normalization in the main text are distinct operations. Base spectral and F2 local spectral features use the sample standard deviation over their flattened token axes, with 10^{-6} added to the denominator. Final candidate-score normalization uses the population standard deviation over K candidates, clamped from below at 10^{-6}. Neither operation estimates normalization statistics from target-subject training data. The joint canonical summaries are not subjected to the spectral-feature normalization before concatenation; their token projection includes LayerNorm.

$$\delta(A) = 10^{-4}\max(\mathrm{tr}(A)/d, 0) + 10^{-6}, \quad A + \delta(A)I_d = U\,\mathrm{diag}(\lambda_j)U^T. \tag{17}$$

$$\mathcal{W}(A) = U\,\mathrm{diag}\Big(\max(\lambda_j, 10^{-6})^{-1/2}\Big)\,U^T. \tag{18}$$

$$e_h = \|W_{:,2h-1:2h}\|_F^2, \quad J = \sum_j s_j^2, \quad p_j = \frac{s_j^2}{\max(J, 10^{-8})}, \quad d_{\mathrm{spec}} = p_1 \tag{19}$$

$$r_{\mathrm{eff}} = \frac{\exp\Big(-\sum_j p_j \log\max(p_j, 10^{-8})\Big)}{2H}. \tag{20}$$

$$\begin{aligned}\mathrm{Norm}_{\mathrm{feat}}(v) &= \frac{v-\bar{v}}{s_{\mathrm{sample}}(v)+10^{-6}},\\ [\mathcal{Z}_S(u)]_i &= \frac{u_i-\bar{u}}{\max\Big(\sqrt{K^{-1}\sum_j (u_j-\bar{u})^2}, 10^{-6}\Big)}.\end{aligned} \tag{21}$$

A.3 Dual-Branch Encoding and Reliability Weighting

Each branch contains two pre-normalized Transformer blocks with four attention heads, embedding dimension 128, feed-forward width 256, GELU, and dropout 0.1. A learned 128-dimensional query then pools the candidate's encoded tokens through four-head attention, followed by LayerNorm. The query is initialized with a truncated normal distribution of standard deviation 0.02. Candidate-set context is the LayerNorm of the arithmetic mean of all candidate representations. A 256→128→128 MLP with GELU and sigmoid generates a vector gate from the concatenation of the local and contextual representations. The gated context is added residually. A LayerNorm and 128→64→1 MLP with GELU and dropout generate the scalar score. This mean-based contextual path is permutation symmetric but makes the score depend on the competing candidate set. The attention backbone follows the Transformer architecture (Vaswani et al., 2017).

High-order reliability combines a frequency prior with five trial-dependent evidence summaries. The summaries are the leading F1 singular value, local spectral SNR, temporal phase stability, log canonical energy, and canonical anisotropy. Their fixed centering and scaling constants are shown in Table 5. The standardized summaries are clipped to $[-6, 6]$, multiplied by nonnegative learned weights, summed with a learned intercept, and passed through a sigmoid with a 0.05 evidence floor. The frequency prior is a sigmoid of the distance between the harmonic frequency and a learned cutoff. Low-order reliabilities are fixed to one; the frequency prior and evidence gate are multiplied only for H4 and H5.

Local SNR is the log ratio between power at the harmonic center and the average power at its two neighboring frequencies, with an additive 10^{-8} stabilizer and clipping to $[-6, 6]$. The neighbor offset is 1/T Hz for a T-second window. Responses are sampled by linear interpolation in the Hann-windowed rFFT. Temporal phase stability uses three consecutive temporal segments; segment phases are corrected for their start times, converted to unit complex values with a magnitude floor of 10^{-6}, averaged across segments, and then averaged in magnitude over channels. Reliability is computed from the current trial and candidate specifications; its parameters are fitted only on source-subject training frequencies.

Reliability modulation occurs at specified locations. Base high-order tokens are multiplied by reliability after token construction. For F1 and F2, reliability first scales the evidence features before projection and then scales the embedded tokens again. The high-order tokenizer computes the five-harmonic evidence before selecting H4/H5 tokens. This retains the joint reference geometry and is different from constructing an isolated two-harmonic canonical reference. The final learned score adds the high-order branch to the low-order branch through $\eta = \mathrm{sigmoid}(a\eta)$, initialized at $\eta = 0.10$.

$$r_{ih} = 1, \quad h \le 3. \tag{22}$$

$$\begin{aligned} r_{ih} &= \sigma((c - hf_i)/\tau)\left[0.05 + 0.95\sigma\left(b + \sum_{m=1}^{5} \mathrm{softplus}(a_m)\tilde{v}_{ihm}\right)\right],\\ &h \in \{4, 5\}.\end{aligned} \tag{23}$$

$$c = 40 + 50\sigma(a_c), \quad \tau = 2 + 18\sigma(a_\tau), \quad e_{\mathrm{learn}} = e^L_{\mathrm{learn}} + \sigma(a_\eta) e^H_{\mathrm{learn}}. \tag{24}$$

Table 5: Fixed feature scales and initial parameters of the reliability module.

Quantity	**Transformation or initialization**
Leading canonical singular value ρ	$(\rho - 0.18) / 0.10$
Local log-SNR	SNR / 0.75
Temporal phase stability	(stability $-$ 0.50) / 0.20
Log canonical energy	(energy $-$ 0.05) / 0.08
Canonical anisotropy	(anisotropy $-$ 0.22) / 0.20
Evidence weights, in the above order	0.70, 0.50, 0.50, 0.25, 0.25; nonnegative through softplus
Evidence intercept	-0.50
Frequency cutoff	60 Hz initially; bounded to 40–90 Hz
Frequency temperature	8 Hz initially; bounded to 2–20 Hz

A.4 Score Aggregation and Multi-View Inference

EPAA forms an explicit anchor by separately standardizing the F1 and F2 scores and combining them with two globally shared weights. The two unconstrained weight logits are initialized to zero, giving w1 = w2 = 0.5. The anchor scale is parameterized by a bounded sigmoid; its unconstrained parameter is initialized to $-\log 2$, giving γ = 1. These parameters are learned from source data and then shared across trials and candidates. The reliability module and harmonic mixing scale also affect the explicit scores. Consequently, this path preserves reference-matching evidence but is not a parameter-free analytic baseline.

The spatial and summary residuals form a separate auxiliary path outside EPAA. They use the unweighted five-harmonic joint F2 feature before token projection. The spatial Gram matrix is symmetrized, divided by its trace with a 10^{-6} floor, and standardized across candidates independently for every subband and matrix entry. Only its upper triangular entries are retained. The remaining 18 matching-summary features are separately standardized across candidates. Each path applies LayerNorm and a feature$\to$32$\to$32 MLP with GELU, adds a learned subband embedding, and uses a 32$\to$16$\to$1 attention MLP to obtain softmax weights over the five bands. The weighted feature is scored by an independent LayerNorm and 32$\to$16$\to$1 MLP. The two scores are independently candidate-standardized and multiplied by nonnegative softplus scales, each initialized to 0.05.

The joint canonical residual features depend on EEG and frequency, not candidate phase. Therefore the same residual R can be used across the four global phase views in the main-text formulation qv = pv + R. Each phase view offsets every candidate phase by vπ/2 modulo 2π while holding EEG fixed. All views share model parameters. The complete per-view score qv, including both residuals, is standardized over candidates before averaging. Averaging unnormalized view scores and normalizing afterward is a different computation.

For implementation, the complete inference sequence is: construct the candidate-conditioned evidence for each phase view; encode the low- and high-order branches and compute their shared-context scores; obtain the explicit F1/F2 scores and EPAA output; compute the phase-invariant F2 residual; add this residual to each EPAA view; standardize each complete candidate-score vector; average the four vectors; and return the maximum-scoring candidate. Candidate sorting is reversed before returning scores, so predictions remain aligned with the supplied table. Permuting the table permutes the mathematical output, but inserting or removing candidates can change context and normalization statistics and is not an invariance claim.

$$w_\ell = 0.1 + 0.8\,\mathrm{softmax}(\vartheta)_\ell, \quad \ell \in \{1, 2\}, \quad \gamma = 0.5 + 1.5\sigma(\rho). \tag{25}$$

$$a = w_1 \mathcal{Z}_S(e_{F1}) + w_2 \mathcal{Z}_S(e_{F2}), \quad p = \mathcal{Z}_S(e_{\mathrm{learn}}) + \gamma a. \tag{26}$$

$$R = \mathrm{softplus}(\beta_{\mathrm{sp}})\mathcal{Z}_S(r_{\mathrm{sp}}) + \mathrm{softplus}(\beta_{\mathrm{sum}})\mathcal{Z}_S(r_{\mathrm{sum}}),$$
$$q_v = p_v + R \tag{27}$$

$$F_\theta(x, S) = \frac{1}{4}\sum_{v=0}^{3} \mathcal{Z}_S(q_v). \tag{28}$$

B Optimization and Training Details

B.1 Complete Training Objectives

This section expands the three-phase training objective in Section 2.4. Let ℓ_{vbk} denote the stage-specific score for view v, trial b, and candidate k, with $V = 4$ phase views and B trials in a batch. Cross-entropy is computed on each view before averaging, rather than on the final view-averaged inference score. Define π_{vb} as its softmax distribution, $\bar{\pi}_b$ as the mean distribution across views, and z_{vb} as the candidate-standardized score vector. For rank j, $m_{vb}^{(j)}$ is the true standardized score minus the j-th largest wrong-candidate score. All variances below use population denominators.

The phase penalty averages four quantities: divergence of each view's probability distribution from the view mean, variance of the standardized scores, variance of the top-one margin, and variance of the top-three-boundary margin. The rank used by the last term is min(3, K − 1). This is the margin against the third-largest wrong score when K ≥ 4; it is not a three-class cross-entropy. The probability floor is 10^{-6}, and no softmax temperature is introduced.

The adjacent-frequency penalty compares the true score against the largest score among all nearest-frequency neighbors. Frequency-distance ties are retained using tolerance $10^{-6} + 10^{-5}$ dmin. The zero-offset view's adjacent margin is detached and acts as a one-sided reference for the other three views. The term is retained even when that reference margin is negative. The base objective is LCE + 0.1 Lphase + 0.1 Ladj; the selected implementation does not add a subject-risk dispersion penalty.

During the second stage, write n = ZS(elearn), a for the explicit anchor, and p = n + γa for the parent score. Gamma contraction penalizes γ below one. The compatibility penalty uses the student's highest-scoring wrong candidate under its detached four-view parent inference score. Component margins are divided by the detached candidate-axis standard deviation of the parent score for each view, then averaged across views. It penalizes a negative learned-component margin only when the detached explicit-component margin is positive. The rival choice, support mask, explicit anchor, γ, and normalization scale are detached in this term.

The preservation penalty instead uses the teacher-selected wrong rival and teacher parent-score scale for both models. The teacher defines a conflict group when its explicit margin is positive and its learned margin is negative. For this group, the penalty preserves the student's explicit margin relative to the teacher; for the complement, it preserves the learned margin. Teacher quantities are detached. Student γ is also detached inside this penalty, while the student anchor itself receives gradients. The 0.1 coefficients are already included in Lcomp and Lpres and are not applied a second time.

$$L_{\mathrm{CE}} = -\frac{1}{VB}\sum_{v,b} \log \pi_{vby_b},$$
$$\pi_{vb:} = \mathrm{softmax}(l_{vb:}), \qquad \bar{\pi}_{b:} = \frac{1}{V}\sum_{v} \pi_{vb:}. \tag{29}$$

$$J = \frac{1}{VB}\sum_{v,b,k} \pi_{vbk}\left(\log \pi_{vbk} - \log\max(\bar{\pi}_{bk}, 10^{-6})\right),$$
$$V_z = \frac{1}{BK}\sum_{b,k} \mathrm{Var}_v(z_{vbk}) \tag{30}$$

$$V_j = \frac{1}{B}\sum_{b} \mathrm{Var}_v(m_{vb}^{(j)}). \tag{31}$$

$$L_{\text{phase}} = \frac{1}{4}(J + V_z + V_1 + V_r), \qquad r = \min(3, K-1),$$
$$L_{\text{adj}} = \frac{1}{(V-1)B} \sum_{v=1}^{V-1} \sum_{b} [\text{sg}(m_{0b}^{\text{adj}}) - m_{vb}^{\text{adj}}]_+^2. \tag{32}$$

$$m_{vb}^{\text{adj}} = z_{vby_b} - \max_{k \in \mathcal{N}(y_b)} z_{vbk}, \quad L_\gamma = \left(\frac{[1-\gamma]_+}{0.5} \right)^2. \tag{33}$$

$$m_{n,b} = \frac{1}{V} \sum_{v} \frac{n_{vby_b} - n_{vbj_b}}{s_{vb}},$$
$$m_{a,b} = \frac{1}{V} \sum_{v} \frac{\text{sg}(\gamma)\,\text{sg}(a_{vby_b} - a_{vbj_b})}{s_{vb}} \tag{34}$$

$$s_{vb} = \text{sg}\big(\max(\text{std}_k(p_{vbk}), 10^{-6})\big). \tag{35}$$

$$L_{\text{comp}} = \frac{0.1}{B} \sum_{b} \mathbf{1}[m_{a,b} > 0] \left(\frac{[-m_{n,b}]_+}{1 + [-m_{n,b}]_+} \right)^2. \tag{36}$$

$$m_{n,b}^M = \frac{1}{V} \sum_{v} \frac{n_{vby_b}^M - n_{vbj_b^T}^M}{s_{vb}^T},$$
$$m_{a,b}^M = \frac{1}{V} \sum_{v} \frac{\text{sg}(\gamma_M)(a_{vby_b}^M - a_{vbj_b^T}^M)}{s_{vb}^T}, \qquad M \in \{T, S\}. \tag{37}$$

$$A_b = \mathbf{1}[m_{a,b}^T > 0 \wedge m_{n,b}^T < 0], \quad h(u) = u/(1+u) \tag{38}$$

$$L_{\text{pres}} = \frac{0.1}{B} \sum_{b} \Big((1 - A_b) h([m_{n,b}^T - m_{n,b}^S]_+)^2 + A_b h([m_{a,b}^T - m_{a,b}^S]_+)^2 \Big). \tag{39}$$

$$L_T = L_{\text{base}}(T) + L_\gamma(T),$$
$$L_S = L_{\text{base}}(S) + L_\gamma(S) + L_{\text{comp}} + L_{\text{pres}}. \tag{40}$$

B.2 Parameter Updates and Gradient Flow

The teacher is a deep copy of the student at the start of stage two, after the low-order branch has been trained and frozen. Teacher and student have disjoint trainable parameters and separate AdamW states. Their losses are computed before either model is updated. Paired forward calls replay the same Torch random state for their stochastic masks. The teacher receives its own loss and optimizer update; detached teacher quantities guide the student penalties. There is no exponential moving average and no teacher–student ensemble at inference.

Starting at the 601st update of stage two, a source-consensus parameter adjustment is applied after AdamW to teacher and student separately. The adjusted subset comprises the high-order evidence tokenizer/front-end parameters and the three EPAA aggregation parameters. High-order attention, candidate-context fusion, the score head, and η continue to receive base gradients but are excluded from this additional adjustment. The adjustment is not another loss term accumulated into AdamW's moments.

For each of three source-role episodes, one virtual base-loss gradient step with step size 0.002 is applied to the selected parameter subset. A cross-entropy gradient on the episode's other source-subject group is then differentiated through this step, yielding a second-order gradient. Three further

first-order ranking gradients are computed at the unadapted model in evaluation mode, using the softplus of the difference between the second-highest wrong score and the true score under four-view parent inference. The rival is selected without gradient. The source groups on the two sides of an episode are disjoint, but both supply training information.

In this ranking-gradient rule, F^{parent} denotes the four-view parent score before adding the residual. The second-highest wrong-candidate term requires $K \geq 3$, as holds for all training candidate sets used here.

The actual AdamW displacement d on the same parameter subset is combined with these six auxiliary gradients. The seven unit directions are the three cross-fit gradients, the three ranking gradients, and $-\mathrm{d}$. A minimum-norm convex combination is obtained by solving over the probability simplex. Its negative direction proposes an auxiliary displacement whose magnitude before scaling is the smaller of the mean cross-fit gradient norm and the base-displacement norm. Multiplication by 0.1 caps the additional displacement at one tenth of the base displacement.

The implementation abstains from the auxiliary update when a required direction, the convex-hull solution, or the base displacement has norm at most 10^{-12}. It also checks that the proposed displacement has descent cosine at least -10^{-6} against all six auxiliary gradients and alignment cosine at least -10^{-6} with the base displacement. Failure retains the ordinary AdamW step. The simplex solver enumerates nonempty supports, solves the associated systems in float64 with least-squares rcond = 10^{-12}, allows weight feasibility tolerance -10^{-9}, clips and renormalizes weights, and uses objective-improvement tolerance 10^{-15}. These are local compatibility checks, not a guarantee that every nonlinear loss or held-out accuracy improves.

Stage three computes the frozen student's four parent scores and canonical F2 features without gradient, then trains only the spatial/summary residual modules and their scales. The same residual is added to all four parent views. Teacher training, virtual adaptation, and the source-consensus adjustment are absent from this stage and from inference.

$$\begin{aligned} \theta'_R &= \theta_R - 0.002 \nabla_{\theta_R} L_{\mathrm{base}}(D_r^{\mathrm{inner}}), \\ g_r &= \nabla_{\theta_R} L_{\mathrm{CE}}(D_r^{\mathrm{meta}}; \theta'_R), \qquad r \in \{1,2,3\}. \end{aligned} \tag{41}$$

$$F^{\mathrm{parent}} = \frac{1}{V} \sum_v \mathcal{Z}_S(p_v), \tag{42}$$

$$\begin{aligned} q_r &= \nabla_{\theta_R} \operatorname{mean}_{D_r^{\mathrm{meta}}} \operatorname{softplus}(F^{\mathrm{parent}}_{j^{(2)}} - F^{\mathrm{parent}}_{y}), \\ v &= (g_1, g_2, g_3, q_1, q_2, q_3, -d). \end{aligned} \tag{43}$$

$$\begin{aligned} w^* &= \operatorname*{arg\,min}_{w \geq 0,\, \sum_j w_j = 1} \left\| \sum_{j=1}^{7} w_j \hat{v}_j \right\|_2^2, \\ h^* &= \sum_{j=1}^{7} w_j^* \hat{v}_j. \end{aligned} \tag{44}$$

$$\begin{aligned} u &= -\frac{h^*}{\|h^*\|_2} \min\left(\frac{1}{3} \sum_{r=1}^{3} \|g_r\|_2, \|d\|_2 \right), \\ \theta_R &\leftarrow \theta_R^{\mathrm{afterAdamW}} + 0.1u. \end{aligned} \tag{45}$$

B.3 Stage-Wise Training Schedule

The three optimization stages are distinct from the two functional decoding stages shown in Figure 2. Each optimization stage performs 1,200 updates. Stage one trains the complete low-order branch. Stage two inherits and freezes that endpoint, then trains the high-order branch, η, and EPAA together with the independently optimized teacher. Stage three inherits the final stage-two student, freezes

the entire parent, and fits the two residual paths. A fresh optimizer is initialized at every stage; within a run, upstream states come from that same run.

The primary sampler repeatedly draws seeded permutations of the complete eligible source training-example list and takes consecutive batches of up to 64; the final short batch is retained. Thus, an update count is not an epoch count. We use three fixed random seeds for repeated runs. Stage-specific and sampling seeds are deterministically derived from each run seed. The exact seed values and offsets will be provided in the reproducibility configurations. The final checkpoint after 1,200 updates in each stage is retained. Periodic checkpoints are used for recovery rather than model selection based on outer evaluation scores.

For consensus optimization, sorted fitting-source IDs are alternated into two groups of eight for final fitting (16 sources), or six for inner configuration selection (12 sources). Each of the three episodes uses three cyclically selected subjects on each side, with episode offsets 0, 3, and 6. Odd episode steps use the first group for the inner update and the second for the cross-fit loss; even steps swap these roles. The cross-fit subject selection is offset by three positions. Each episode selects q = min(12, M) rotating training targets from M legal fit targets and one example per selected subject–target pair. Training-target subsets rotate across all M targets. The source subjects used for the cross-fit loss participate in optimization and are not an independent validation set.

At active stage-two update t, the target positions for episode r = 0, 1, 2 are $[(t - 601)3q + rq + j]$ modulo M, for $j = 0, \ldots, q - 1$. For episode $r \in \{0, 1, 2\}$, let $tr = t + 3r$. The inner subject selection starts at $(tr - 1)$ modulo g, where g is the group size (8 for final fitting; 6 for inner selection); all cyclic subject indices use modulo g. For a selected subject at position i and selected target at position j, the example slot is $(tr + o + i + j)$ modulo $n_{s,k}$, with $o = 0$ for the inner side and $o = 3$ for the cross-fit side. Here $n_{s,k}$ is the number of available examples in the fixed input order. The episode has no subject overlap between sides and no duplicate subject–target example within one side.

Table 6: Training stages and parameter scopes.

Stage	Updated parameters	Frozen parameters	Learning rate / objective
1: low-order branch	Low-order token projections, encoder, pooling, context, and scorer	No inherited trained branch	0.002; base objective
2: high-order branch and EPAA	High-order evidence and scorer, η, EPAA; independent teacher copy	Low-order branch	0.002; student/teacher objectives; consensus at updates 601–1200
3: auxiliary residuals	Spatial and summary encoders, attention readouts, score heads, and scales	Entire stage-two student and feature extraction	0.001; base objective on complete scores

B.4 Optimization Settings

Table 7 lists the optimization settings used in the experiments reported in the main text. Training uses FP32, without automatic mixed precision or gradient accumulation. AdamW uses constant learning rates within each stage, a weight decay of 10^{-4}, and global gradient-norm clipping at 5. No outer evaluation metric is used for checkpoint selection. EEG from evaluation subjects and held-out frequencies is excluded from parameter updates; candidate stimulus specifications are available through the defined inference interface. The AdamW optimizer follows Loshchilov and Hutter (2019).

Table 7: Model and optimization settings used in the reported experiments.

Setting	Value
Embedding dimension / encoder blocks per branch	128 / 2

Continued on next page

Table 7 (continued)

Setting	Value
Attention heads / feed-forward hidden dimension	4 / 256
Dropout	0.1
Residual embedding dimension	32
Optimizer / weight decay / clipping	AdamW / 10^{-4} / 5
Batch size / updates per stage	64 / 1,200
Stage learning rates	0.002 / 0.002 / 0.001
Phase offsets	0, $\pi/2$, π, $3\pi/2$
Number of random seeds	3
Phase-consistency / adjacent-margin weights	0.1 / 0.1
Compatibility / preservation weights	0.1 / 0.1, already included in their definitions
Virtual inner step / auxiliary-update multiplier	0.002 / 0.1
Source-consensus activation	Updates 601–1200 of stage two
Initial η / initial EPAA weights / initial γ	0.10 / (0.5, 0.5) / 1.0
Initial spatial / summary residual scales	0.05 / 0.05
Checkpoint rule	Final update of each stage; fresh optimizer between stages

B.5 Run Correspondence and Reproducibility Records

The correspondence between the reported results and the underlying experimental runs has been verified. The planned anonymized release will include a run manifest linking the reported results to their datasets, observation windows, frequency folds, random seeds, model variants, source and evaluation participant lists, preprocessing procedures, candidate ordering, trainable-parameter masks, update counts, and checkpoint-selection rules. The release package will also document the code revision, software environment, checkpoint and prediction-file fingerprints, and aggregation scripts. These materials are being prepared for release during the discussion period, in accordance with conference policies.

The training configuration is summarized in Appendix B.4. The configuration-selection procedure is described in Appendix D.3. Participant-level uncertainty and its distinction from variability across random seeds are discussed in Appendix E.

C Baseline Selection and Implementation

C.1 Common data access and implementation scope

The baselines use the same registered source subjects, frequency folds, EEG channels, window lengths, and full candidate tables. Benchmark and BETA use nine channels and 40 candidates; Wearable uses eight channels and 12 candidates. At 250 Hz, the 0.4, 0.6, and 0.8 s windows contain 100, 150, and 200 samples. Fitting uses only the 16 registered source subjects and seen-frequency (FIT) trials. Target-subject EEG and real EEG from held-out frequencies are excluded from fitting, learned normalization, template estimation, and model selection. Known candidate frequencies and stimulus phases remain available for analytic references. Each method receives only EEG within the evaluated window; a longer analytic reference does not provide additional EEG samples.

Method-specific filtering and scoring are retained within the common data-access protocol. The descriptions below specify the source-only adaptations used for the results reported in Table 2 and identify their implementation origins. These adaptations follow the evaluation protocol defined in

this paper, which differs from some original experimental protocols. SSVEPformer (Chen et al., 2023) uses a fixed class-output head, whereas the original CSDuoDFN protocol (Deng et al., 2023) includes one calibration trial per stimulus target from the target user; these settings differ from zero-target-calibration decoding with held-out stimulus frequencies.

C.2 Analytic and Transfer Baselines

CCA-H5 uses time-centered EEG and five-harmonic sine/cosine references, a covariance regularizer of 10^{-4}, and the leading canonical correlation for each candidate. FBCCA-H5 additionally combines squared correlations across filter-bank outputs. These analytic scores require no learned subject template.

tlCCA-1 is adapted from the SSVEP-tlCCA implementation. Each source subject supplies mean FIT-frequency responses; five nearby FIT frequencies support multistimulus CCA and alternating estimation of spatial filters and impulse responses. An unseen candidate is synthesized from its nearest FIT anchor using its own frequency and phase, whereas a seen candidate retains the measured source mean and its projection. Complete candidate-score vectors are then averaged equally across the 16 sources. This source-score averaging and direct seen-frequency branch are explicit adaptations of the original within-subject transfer setting. The implementation uses five harmonics, five Chebyshev-I filter bands, and filter weights $b^{-1.25} + 0.25$; the alternating solver uses a tolerance of 10^{-4} and at most 199 iterations.

SST follows the time-frequency-joint-representation repository. DP-MAFD is applied separately to each source subject and physical channel using FIT-frequency means and the current-window Hilbert transform. The implementation uses 50 decomposition layers and a cumulative-coefficient threshold of 0.9. Nearest-FIT coefficients and each candidate's own phase grid reconstruct unseen-frequency templates; seen candidates retain empirical templates. Grouped CCA uses templates sharing a FIT anchor together with five-harmonic references. Scoring follows the author-code combination $R_4 = R_1 + R_3$, followed by equal averaging of source-subject scores. Adaptive Fourier decomposition is described by Wang et al. (2022).

C.3 Neural Baseline Adaptations and Selection

The GZSL adaptation uses the GZSL architecture in the MandlerJ/SRRNet repository with source/FIT training data. It retains full-table inference and uses a fixed terminal checkpoint. The separate SRRNet baseline follows MaiXiming/SRRNet; its templates and spatial filters are estimated from source/FIT EEG, and its training schedule uses a fixed terminal epoch. These implementation identities distinguish GZSL, iGZSL, and the SRRNet baseline.

iGZSL uses the SRRNet-repository architecture with the paper's waveform and classification objectives. For selection, the first 12 registered sources provide normalization statistics, FIT templates, and training data; FIT trials from the remaining four sources are used to select once among the aggregate, Transformer, extractor, and classifier strategies based on subject-macro accuracy. The normalization and selected strategy are then frozen, and the final model is trained on all 16 sources for 25 epochs using five cycles of three waveform-training and two classification-training epochs. Training uses AdamW with a learning rate of 0.01 and weight decay of 0.001; the terminal checkpoint is retained. Inference preserves iGZSL's routing rule: predictions outside FIT retain the fused prediction, while predictions within FIT use the selected strategy. This routing uses candidate membership, not target labels. No target-calibration EEG or held-out-frequency EEG is used for selection.

C.4 LINEAR physical-score control

LINEAR is a ten-dimensional linear candidate scorer implemented independently of the learned HarmTSFormer representation. For each candidate, it concatenates five ungated squared reference-correlation features and five ungated F2 log-power features, the latter averaged across bands and channels for each harmonic. Each feature is standardized across the candidate axis using population standard deviation. A single ten-weight vector, shared by every candidate and with no intercept, is fitted to source/FIT trials by mean cross-entropy over four common phase views (0, $\frac{\pi}{2}$, π, and $3\pi/2$), with an L2 penalty of ($\frac{10^{-3}}{2}$) times the squared weight norm. The deterministic float64 L-BFGS-B solver starts from zero.

Training features are computed using only the FIT candidate set, rather than slicing features normalized over the full table. Evaluation uses the full candidate table. Within each view, the resulting candidate scores are standardized across candidates; the four standardized score vectors are then averaged. No HarmTSFormer parameters or checkpoints are used. LINEAR therefore tests a simple learned combination of explicit physical evidence, while differing from a nonlinear candidate-conditioned network or joint end-to-end training.

C.5 Configuration summary

Table 8 summarizes the baseline configurations. GZSL uses a 3×3 learning-rate/weight-decay grid (0.0001/0.0003/0.001 × 0/0.0001/0.001); SRRNet uses the same learning rates with weight decay 0/0.001. Source-only selection follows D.3; lower cost breaks validation ties. The iGZSL schedule is specified in C.3.

Table 8: Baseline configuration summary.

Method	Source-only scope	Fit / selection / checkpoint	Configuration ID
CCA / FBCCA	Analytic candidate references; no learned target template.	No gradient fit; retain each named H3/H5 setting.	C1-01
tlCCA-1	Source FIT templates; nearest-anchor synthesis; source-score averaging.	$\leq$199 solver iterations; tolerance 10^{-4}; terminal state.	C1-02
SST	FIT-source reconstruction and shared-anchor CCA.	50 decomposition layers; threshold 0.9; terminal decomposition.	C1-03
GZSL	Source/FIT training; full-table inference.	3,600 updates per fit; 9 LR/WD configurations; terminal update.	C1-04
iGZSL	12-source fit / four-source strategy selection, then 16-source refit.	25 epochs; four routing strategies; AdamW 0.01/0.001; terminal checkpoint (C.3).	C1-05
SRRNet	Source/FIT templates and spatial filters; fixed terminal checkpoint.	50 epochs per fit; 6 LR/WD configurations; final epoch.	C1-06
LINEAR	Ten physical features; shared deterministic scorer.	L-BFGS-B cap 500 iterations; fixed L2 and no random restart.	C1-07
Shared MLP	H1–H5, candidate mean, four views and explicit-score fusion (G.3).	3 × 1,200 updates. Width search: 64/96/128; terminal stages.	C1-08
CCT / HCM	Protocol adaptations with 12/4 inner source selection (H.1).	LR 0.0001/0.0003/0.001 × WD 0/0.001; terminal checkpoint.	C1-09

Baseline diagnostics separately examine candidate-ID round trips, finite nonconstant scores, source-FIT learning curves, and raw versus routed predictions. The diagnostic records include source-validation selection and the final checkpoint for each configuration.

C.6 Contextual Literature and Technical Foundations

Harmonics and multichannel reference matching. Physiological reviews and stimulation surveys emphasize the role of the stimulation paradigm in SSVEP responses (Vialatte et al., 2010; Zhu et al., 2010). Harmonic components can contribute useful stimulus information, while multichannel detection combines spatial observations to improve frequency recognition (Müller-Putz et al., 2005; Friman et al., 2007). Canonical correlation analysis provides a general framework for relating two sets of variables (Hotelling, 1936); its SSVEP application matches EEG to sinusoidal references (Lin et al., 2006; Bin et al., 2009). These foundations motivate candidate-specific harmonic evidence in

HarmTSFormer. Analytically constructing a reference for a new frequency makes that candidate representable, but representability alone does not establish accurate decoding of its EEG response.

Spatial filtering and calibration requirements. CCA variants differ in how they incorporate recorded templates and individual training data (Nakanishi et al., 2015). Multiset CCA learns common structure across repeated observations (Zhang et al., 2014), TRCA emphasizes reproducible task-related components (Nakanishi et al., 2018), and TDCA introduces discriminant information for individually calibrated SSVEP decoding (Liu et al., 2021). CORRCA offers another correlated-component formulation; its original publication and subsequent correction should be read together (Zhang et al., 2018a;b). Their access to calibration recordings is an essential part of the experimental setting, so transferring an operator into a source-only protocol requires specifying which data construct its filters and templates.

Transfer across domains and stimulus targets. EEG transfer learning addresses subject/session variability and calibration burden (Lotte et al., 2018; Jayaram et al., 2016). Cross-domain template transfer (Chiang et al., 2021) and Volterra-model cross-target transfer (Lin et al., 2021) address related routes for reducing calibration burden. Frequency-transfer methods such as tlCCA and SST (Wong et al., 2021; Wang et al., 2023c) further motivate separating transferable structure from target-specific observations; the adaptive Fourier decomposition underlying the latter has its own multichannel signal-analysis formulation (Wang et al., 2022). CSDuDoFN instead studies a one-shot setting with a calibration example for each target (Deng et al., 2023). Consequently, subject transfer, stimulus transfer, and target-calibration reduction are distinct protocol dimensions. The adaptations described in this appendix specify the information available to each implemented comparator and do not imply a reproduction of every original evaluation protocol.

Neural EEG representations. Convolutional EEG decoding (Schirrmeister et al., 2017) and the compact EEGNet architecture (Lawhern et al., 2018) provide established approaches to learning temporal and spatial features. EEG Conformer combines convolutional features with attention for EEG decoding (Song et al., 2023), while Transformer attention provides a general sequence-processing mechanism (Vaswani et al., 2017). These precedents motivate architectural ingredients, but an encoder alone does not determine how unseen stimulus candidates enter a classifier. HarmTSFormer scores candidate-conditioned representations with shared parameters. The CCT and HCM comparators in Appendix H are the specific adaptations defined there; the broader neural literature cited here supplies technical context and should not be read as a list of additional measured baselines.

Pretraining and transferable EEG representations. Large-scale EEG pretraining provides a complementary direction: BENDR uses contrastive self-supervision (Kostas et al., 2021), BIOT supports heterogeneous biosignal formats (Yang et al., 2023), LaBraM learns generic EEG representations (Jiang et al., 2024), and CBraMod models spatial and temporal dependencies separately (J. Wang et al., 2025). Such transfer concerns learned representations; an unseen-frequency decision still requires an explicit candidate interface and a specified downstream adaptation protocol.

Candidate-set context and evidence fusion. Semantic zero-shot models use descriptions of unobserved classes (Frome et al., 2013; Romera-Paredes and Torr, 2015), with evaluation sensitive to training and candidate-label protocols (Xian et al., 2019); our descriptors are physical stimulus references. Deep Sets formalizes permutation-invariant set processing through symmetric aggregation (Zaheer et al., 2017); Set Transformer instead uses attention to model interactions among set elements (Lee et al., 2019). HarmTSFormer's pooled candidate context and gate belong to the former broad design principle, without introducing pairwise candidate attention. Reordering candidates and changing their membership are different operations: a symmetric pool can preserve order equivariance while its value changes when candidates are added. Classical classifier-combination work supplies context for combining multiple evidence sources (Kittler et al., 1998). Probability calibration concerns agreement between confidence and correctness (Guo et al., 2017); candidate-axis score standardization in EPAA does not by itself establish that property or turn fusion weights into causal contributions.

Numerical and signal-processing foundations. Windowing affects leakage and the separation of nearby harmonic components in finite observations (Harris, 1978). Covariance conditioning is likewise relevant when estimating correlation-based evidence from short multichannel segments (Ledoit and Wolf, 2004). The fixed regularization specified for our operators should be distinguished from an estimated Ledoit–Wolf shrinkage coefficient. Layer normalization (Ba et al., 2016) and

GELU activation (Hendrycks and Gimpel, 2016) provide standard neural-network operations used in the representation pipeline. These operations have different normalization domains and purposes from candidate-axis score standardization.

Optimization context. AdamW separates weight decay from the adaptive gradient update (Loshchilov and Hutter, 2019). Differentiation through a virtual gradient step is also a familiar optimization construction, including in meta-learning (Finn et al., 2017); here, the documented source-data updates do not introduce target-participant adaptation at inference. Minimum-norm combinations of task gradients connect to multiobjective descent (Désidéri, 2012; Sener and Koltun, 2018). That connection explains the local gradient-combination operation used during training. It does not establish global Pareto optimality, particularly when the actual procedure restricts the updated parameter subset and caps update displacement.

Selection and statistical interpretation. MOABB illustrates the value of consistent preprocessing and evaluation across datasets (Jayaram and Barachant, 2018). Model-selection bias motivates separating configuration selection from final evaluation (Cawley and Talbot, 2010), while small-sample neuroimaging evaluations can have substantial uncertainty (Varoquaux, 2018). Bootstrap resampling provides a general basis for uncertainty estimation (Efron, 1979), with different interval constructions having different properties (Efron, 1987). Appendix E specifies paired participant resampling and percentile intervals; these should not be interpreted as bias-corrected and accelerated intervals or as independent trial-level replications. Work on comparisons across multiple datasets further emphasizes the unit of analysis (Demšar, 2006). Our nine dataset–window conditions are related evaluation settings, and their count should not be read as nine independent datasets or as a multiplicity-adjusted significance result.

D Dataset Partitions and Source-Only Selection

The protocol uses 16 source participants per dataset and 12, 12, and 38 evaluation participants for Benchmark, BETA, and Wearable, respectively. Only source-participant EEG at training frequencies is used for fitting, learned normalization, template construction, or model selection.

D.1 Complete participant allocation

Table 9 specifies the participant allocation used in the reported experiments. Participant IDs are one-based, and all ranges are inclusive. The source and evaluation sets are disjoint. The outside-pool column lists the remaining released participants outside these two sets.

Table 9: Complete participant allocation used in the reported experiments.

Dataset	Released N	Source IDs	Evaluation IDs	Outside-pool IDs
Benchmark	35	1–16	17–28	29–35
BETA	70	1–16	17–28	29–70
Wearable wet	102	1–16	17–54	55–102

The ID ranges cover all released participants, and the source and evaluation counts match the cohort sizes reported in the main text.

Figure 5 summarizes the participant and frequency holdouts and the full-candidate decision rule.

D.2 Frequency folds and candidate definitions

The five folds are frequency holdouts, not subject folds. Training frequencies are the complement of each holdout. All 40 Benchmark/BETA or 12 Wearable candidates remain active at inference, while only held-out true labels enter unseen-frequency evaluation. Errors toward training-frequency candidates remain errors. The main Benchmark/BETA protocol holds out 38 interior frequencies once and retains the 8.0/15.8 Hz endpoints in training; it principally tests interpolation. Wearable holds out all 12 frequencies once.

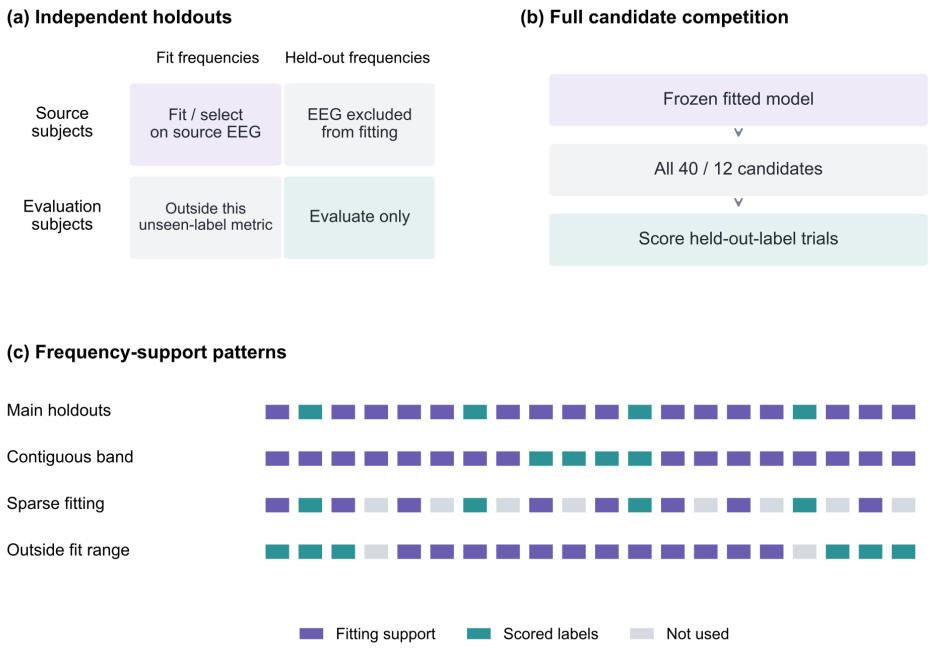

Figure 5: Data partitions and evaluation protocol. (a) Participant and frequency holdouts. (b) Full-candidate inference with 40 or 12 candidates, scoring trials with held-out labels. (c) Fitting and scored frequencies across four settings. Candidate stimulus frequencies and phases are available as input descriptors; held-out EEG is excluded from fitting.

Table 10: Frequency holdouts and training candidate counts used in the main protocol.

Fold	Benchmark / BETA holdout Hz	Train K	Wearable holdout Hz	Train K
1	8.2, 9.2, . . . , 15.2	32	9.25, 11.75, 14.25	9
2	8.4, 9.4, . . . , 15.4	32	9.75, 12.25, 14.75	9
3	8.6, 9.6, . . . , 15.6	32	10.25, 12.75	10
4	8.8, 9.8, . . . , 14.8	33	10.75, 13.25	10
5	9.0, 10.0, . . . , 15.0	33	11.25, 13.75	10

Each Benchmark/BETA holdout sequence advances by exactly 1 Hz. The training frequencies are the complement of the listed holdout within the full candidate table.

Candidate indices i start at zero. For Benchmark, a = i mod 8, b = floor(i/8), frequency = 8 + a + 0.2b Hz, and phase/π = 0.5[(a + b) mod 4]. For BETA, frequency = 8 + 0.2[(i + 3) mod 40] Hz and phase/π = 0.5[(i + 3) mod 4]. For Wearable, a = i mod 3, b = floor(i/3), frequency = 9.25 + 2a + 0.5b Hz, and phase = 0. All candidate-conditioned quantities share the corresponding candidate order.

D.3 Configuration selection

The final-checkpoint rule retains the state after each 1,200-update stage. Hyperparameters are selected using the source-only procedure below. Source groups used in consensus optimization are training groups, not independent validation participants.

Configuration selection uses 12 source participants for inner fitting and four disjoint source participants for validation, with an inner frequency holdout restricted to the outer training frequencies. Mean participant accuracy selects the configuration, and the choice is frozen before final fitting on all 16 sources. The width/layer configurations 64/1, 128/2, and 256/3 yield inner-validation accuracies of 41.8%, 43.8%, and 43.1%, respectively, selecting 128/2. Outer evaluation EEG and outer held-out-frequency EEG are excluded from selection. All fits retain the terminal checkpoint.

E STATISTICAL UNITS AND UNCERTAINTY

Table 11 aligns each reported Full mean with the strongest observed Table 2 comparator and its arithmetic difference, together with the confidence interval and number of participants showing a positive difference. Across three fitted runs, sample SD uses denominator $3 - 1$ and measures seed variation, not participant uncertainty. With fixed inputs and preprocessing, CCA/FBCCA do not generate training-seed variation; dispersion arising from sampling, preprocessing, or outer partitions is distinct from training-seed variation.

For participant-level comparisons, mutually exclusive holdout trials are pooled within each participant and seed, seeds are averaged within participant, and participants receive equal weights. Seeds, folds, and trials are not treated as independent participants. Participant-level uncertainty is conditional on the seed set and candidate grid. The strongest observed comparator is a descriptive reference; selecting a comparator from outer outcomes requires multiplicity control across all eligible comparisons for confirmatory inference.

The confidence intervals are computed by jointly resampling paired participants 10,000 times after the stated within-participant seed averaging, using random seed 42 and the 2.5th/97.5th percentiles of mean differences with linear quantile interpolation. The same resampled participant IDs are used for both methods.

For fitted models, SD is computed across three training seeds. For deterministic baselines, nonzero dispersion is computed across three prespecified source/trial-resampling repeats, paired across methods. With identical inputs and configuration, a deterministic baseline's training-seed SD is zero.

Table 11: Accuracy means, derived gains, confidence intervals, and counts of participants showing improvement. Accuracy is percent; differences and intervals are percentage points.

Dataset / s	**Best baseline**	**Full / baseline (%)**	**Δ (pp)**	**95% CI**	**improved / N**
Benchmark / 0.4	SRRNet	28.10 / 27.47	+0.63	[−0.90, +2.16]	7/12
Benchmark / 0.6	tlCCA-1	49.76 / 47.76	+2.00	[+0.30, +3.70]	9/12
Benchmark / 0.8	tlCCA-1	68.23 / 65.87	+2.36	[+0.45, +4.27]	9/12
BETA / 0.4	SRRNet	14.43 / 12.61	+1.82	[+0.15, +3.49]	8/12
BETA / 0.6	SRRNet	21.24 / 19.79	+1.45	[+0.05, +2.85]	8/12
BETA / 0.8	tlCCA-1	31.72 / 29.66	+2.06	[+0.20, +3.92]	9/12
Wearable / 0.4	FBCCA-H3	34.71 / 32.24	+2.47	[+0.35, +4.59]	25/38
Wearable / 0.6	FBCCA-H3	48.32 / 47.65	+0.67	[−0.80, +2.14]	22/38
Wearable / 0.8	LINEAR	59.69 / 58.97	+0.72	[−0.70, +2.14]	23/38

Confidence intervals and improved-participant counts summarize different aspects of the comparison. A positive aggregate difference does not imply improvement for every participant. The strongest observed comparator is used descriptively; the table does not by itself establish multiplicity-adjusted significance.

Table 12 summarizes the 0.6 s comparison across two alternative participant partitions and reports counts of positive participant and frequency differences. Alternative pools preserve the reported cohort sizes, remain disjoint within each split, and use identical method-specific inputs. All participant and frequency results are retained, including nonpositive differences; frequency summaries use recall under the full candidate competition. Repeated partitions of the same public dataset are sensitivity analyses, not new external test sets.

Table 12: Sensitivity to alternative participant partitions at 0.6 s; differences are Full minus the named comparator.

Dataset / comparator	Split A gap pp	Split B gap pp	Positive subjects(Split A)	Positive frequencies(Split A)
Benchmark / tlCCA-1	+1.60	+1.80	8 / 12	28 / 38
BETA / SRRNet	+1.20	+0.20	8 / 12	27 / 38
Wearable / FBCCA-H3	+0.60	+0.30	25 / 38	8 / 12

Subject and frequency denominators follow the stated protocol. Positive counts denote participants or frequencies for which Full exceeds the named comparator.

The alternative participant partitions rotate the ordered union of Table 9 source and evaluation IDs left by four positions for Split A and eight for Split B. The first 16 are source participants and the remaining 12/12/38 are evaluation participants. Inner selection and frequency exclusions are reapplied within each split. The two positive-count columns in Table 12 refer to Split A and are not pooled across A and B.

These results show small positive mean gains across the examined partitions with heterogeneous individual effects. They do not establish universal or uniformly significant superiority.

Figure 6 reports separate frequency-wise and participant-wise comparisons; Tables 11–12 report their respective aggregate and partition analyses.

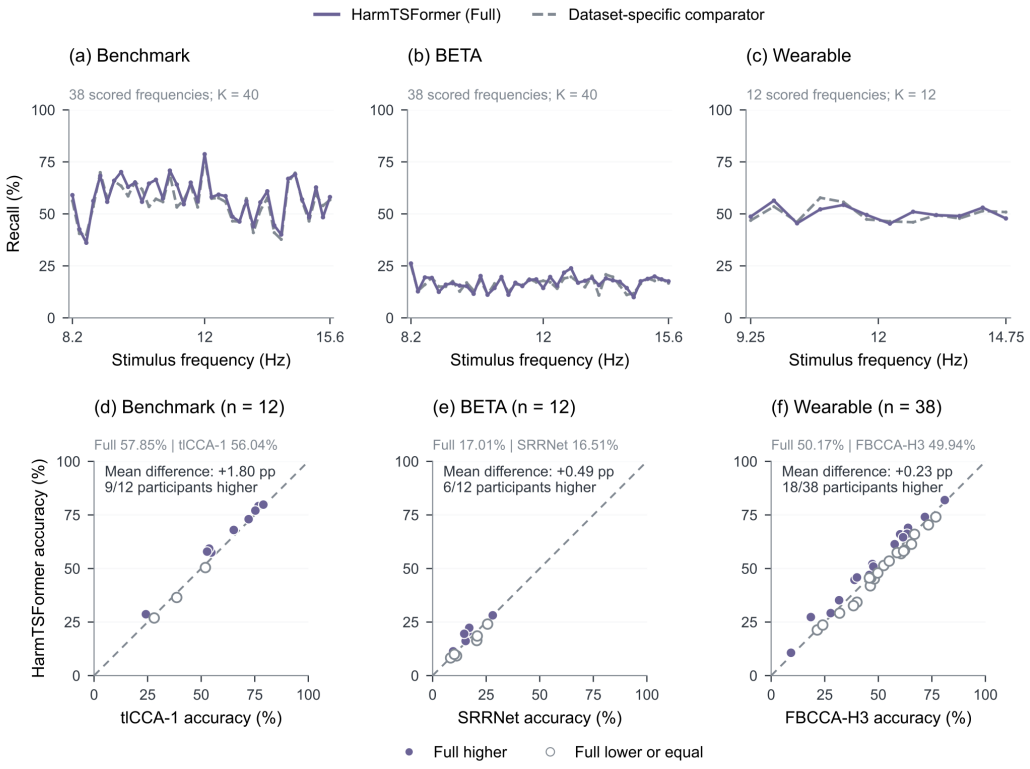

Figure 6: Frequency-wise and participant-wise comparisons at 0.6 s, with 12/12/38 evaluation participants, 38/38/12 scored frequencies, and 40/40/12 candidates for Benchmark/BETA/Wearable, respectively. Each lower-panel point represents one participant. The comparators are tlCCA-1, SRRNet, and FBCCA-H3, respectively. Full mean accuracies are 57.85%, 17.01%, and 50.17%. The diagonal indicates equal accuracy. These comparisons are reported separately from Tables 11–12.

F GENERALIZATION AND SENSITIVITY

The following analyses use the same source-only exclusion rules as Appendix D. Each comparison fixes the participant pools, channel set, observation window, candidate metadata, and selection budget before evaluating held-out labels.

F.1 CONTIGUOUS FREQUENCY BANDS

For the 40-target datasets, one of five consecutive eight-frequency bands is reserved: 8.0–9.4, 9.6–11.0, 11.2–12.6, 12.8–14.2, or 14.4–15.8 Hz, in 0.2-Hz steps. Models are fitted on its complement and score all 40 candidates. The middle bands test missing regions inside the fitting range; the end bands test targets beyond a fitting boundary. For Wearable, the analogous construction reserves one of four consecutive three-frequency groups and fits on the remaining nine, while scoring all 12 candidates. The comparator is fixed using source validation before inspecting test outcomes.

F.2 SPARSE FITTING SUPPORT

The test-frequency set is fixed while nested, approximately evenly spaced subsets of its fitting complement are varied. For the 40-target case, the support sizes are 8, 12, 16, 24, and 32 frequencies; the analogous Wearable sizes are 3, 5, 7, and 9. The update budget and complete test candidate set remain fixed. Fewer fitting frequencies also reduce the number of available training trials. Relative gains are retained despite lower absolute accuracy at the sparsest support; changing test candidate count alone would not establish this result.

F.3 TARGETS OUTSIDE THE FITTING RANGE

Lower and upper target bands with recorded EEG are reserved, and models are fitted only on the interior stimulus frequencies. These targets remain inside the dataset stimulus range but lie outside the training-frequency interval. Each side is evaluated separately using distance to the closest fitting frequency and target-wise recall against the full candidate set. This differs from adding an off-grid distractor without corresponding test EEG. The three panels summarize frequency-support analyses; panel-specific results are distinct from the main-protocol five-fold aggregate.

Figure 7 reports Benchmark results at 0.6 s with 12 evaluation participants and all 40 candidates. Panel (a) uses the five eight-frequency holdouts in F.1. Panel (b) fixes targets 8.4, 9.4, …, 15.4 Hz and nested fitting supports of 8/12/16/24/32 frequencies from their complement. Panel (c) fits 9.6–14.2 Hz and tests eight frequencies on each outside side. Full is evaluated over three runs; deterministic tlCCA-1 has no training-seed error bar.

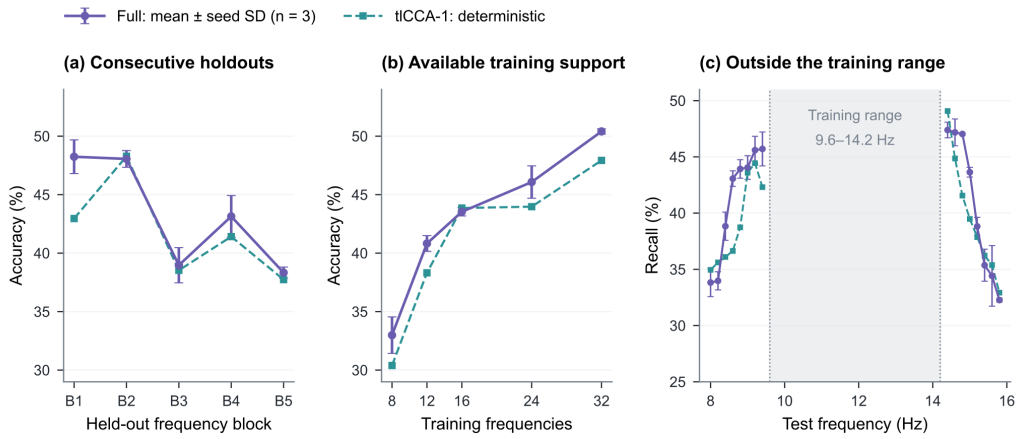

Figure 7: Benchmark, 0.6 s, 12 evaluation participants, 40 candidates. (a) Contiguous-band holdouts. (b) Fixed test labels with varying fitting support. (c) Recall outside the 9.6–14.2 Hz fitting range; gray shading denotes fitting support. Full error bars are sample SD over three runs, not participant confidence intervals.

F.4 CANDIDATE PHASE INFORMATION

For the common-offset analysis, the same phase offset is added to every candidate, the model is frozen, EEG and crop timing are unchanged, and four-view score averaging is retained. The interval from $-\pi/4$ to $\pi/4$ spans one period; adding $\pi/2$ only permutes the four views and is not an independent stress point. The reference control uses the complete sine–cosine span. The results show small off-grid variation; this common-offset construction does not test independent errors in the phase of each candidate.

A separate candidate-specific test draws each phase error independently from a uniform interval with half-width $\pi/12$, $\pi/6$, or $\pi/3$, reuses the same prespecified draws across methods, and averages over draws. Table 13 reports the 0.6-s results for this perturbation.

Table 13: Candidate-specific phase-error sensitivity, accuracy (%).

Dataset	**0**	**π/12**	**π/6**	**π/3**
Benchmark	49.76	49.16	47.56	43.96
BETA	21.24	20.84	20.04	18.54
Wearable	48.32	47.82	46.72	44.62

The common-offset figure uses a different perturbation and is not the numerical source of Table 13.

Table 13 reports candidate-specific phase-error sensitivity at 0.6 s. Zero-offset accuracies are 49.76%, 21.24%, and 48.32% for Benchmark, BETA, and Wearable, respectively. The protocol uses 20 fixed independent candidate-error draws shared across methods, then averages draws within participant and seed before aggregation. This analysis is separate from Figure 8's common-offset analysis.

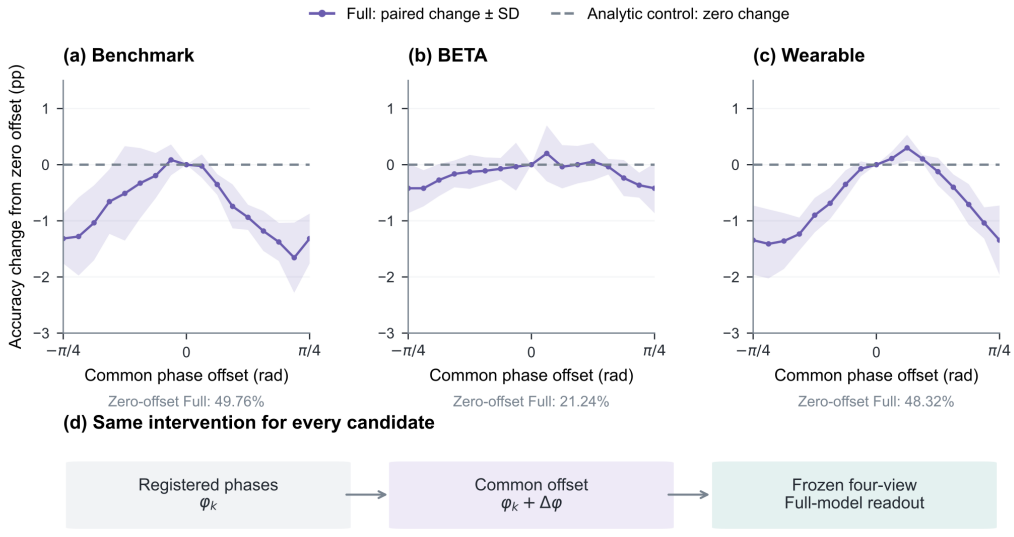

Figure 8: Change in accuracy relative to each seed's own zero-offset value, in percentage points, at 17 common offsets over $-\pi/4$ to $\pi/4$. Bands are sample SD of three paired changes; the zero-offset change and its SD equal zero by definition. Zero-offset accuracies are 49.76%, 21.24%, and 48.32%. The endpoints coincide by the four-view $\pi/2$ period. (d) The same offset is applied to every candidate; the analytic control is phase-invariant. This differs from Table 13's independent candidate errors.

F.5 CANDIDATE COUNT AND COMPOSITION

Define A as recomputing scores on a CORE candidate subset and choosing within CORE; B as computing scores on FULL but choosing within CORE; and C as computing and choosing on FULL. All comparisons reuse the same checkpoint and scored trials, with every true target retained in

CORE. B−A measures the combined change in candidate-dependent computation, including feature context and normalization; C−B measures added decision competition. B−A is not an isolated causal effect of the EPAA weights.

For count sensitivity, prespecified nested candidate sets are compared while retaining the same true-label trials; for composition sensitivity, several prespecified sets are compared at the same K. Repeated seeds and subset draws are paired, and their sources of uncertainty remain distinct. Candidate insertion produces a modest accuracy loss, and equal-sized compositions remain close; this conclusion is bounded to the displayed sets.

Figure 9 reports candidate-set analyses at 0.6 s. Benchmark candidate counts are 12/20/28/40; the three CORE/FULL comparisons are 20/40, 20/40, and 6/12 for Benchmark/BETA/Wearable. Every CORE retains eligible true labels. Panel (c) links the same three seeds across three K=20 Benchmark compositions; Set 1 is shared with (a,b). Full-candidate means and SDs match the Figure 8 zero-offset results.

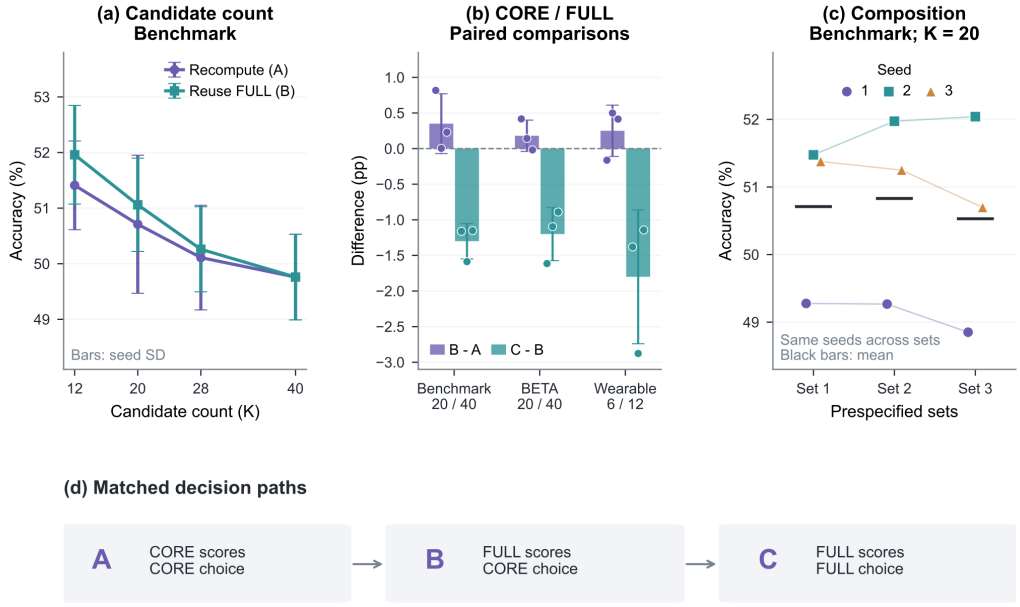

Figure 9: Candidate count, composition, and decision competition. (a) Means and sample SD over three seeds. (b) Paired B−A and C−B differences: B−A combines context and normalization; fixed-score extra competition gives C−B≤0. (c) The same seeds across K=20 sets; black bars are means. (d) CORE/FULL scoring and choice paths. At the full candidate set, A=B=C.

G Fusion and Structural Controls

The A4 means and Figure 4 readouts are reported in the main text. The following controls specify which components are fixed or learned and compare fusion alternatives at matched endpoints.

G.1 Equal weights and matched fusion endpoints

The equal-weight control retains candidate-score normalization, fixes the outer learned/explicit scale γ to one, and keeps the inner explicit mixture weights trainable under their original constraints. Thus the learned score and composite explicit score have a 1:1 outer coefficient; the two raw explicit scores are not necessarily equally weighted.

For the adaptive comparison, a small shared gate reads the normalized learned, F1, and F2 scores and their candidate-set summaries, then produces trial-dependent fusion coefficients. Static and adaptive alternatives use the same source-only validation and tuning budget. Both are compared at the EPAA-only endpoint, with all residuals absent and four-view inference retained. A full-

model comparison would additionally require matched residuals in both arms. The comparison characterizes the relative performance of shared and trial-dependent coefficients under this protocol.

A4 is the main Table 3 full-output intervention; Figure 4 reports separate learned/reference/EPAA/Full readouts. The matched fusion comparison uses identical source selection, update exposure, and the same low-order state, removes both residual paths, and retains four views. Figure 10 reports three paired runs; the gate mean exceeds EPAA in five of nine conditions. Run 1, Run 2, and Run 3 identify the paired repeats within this analysis.

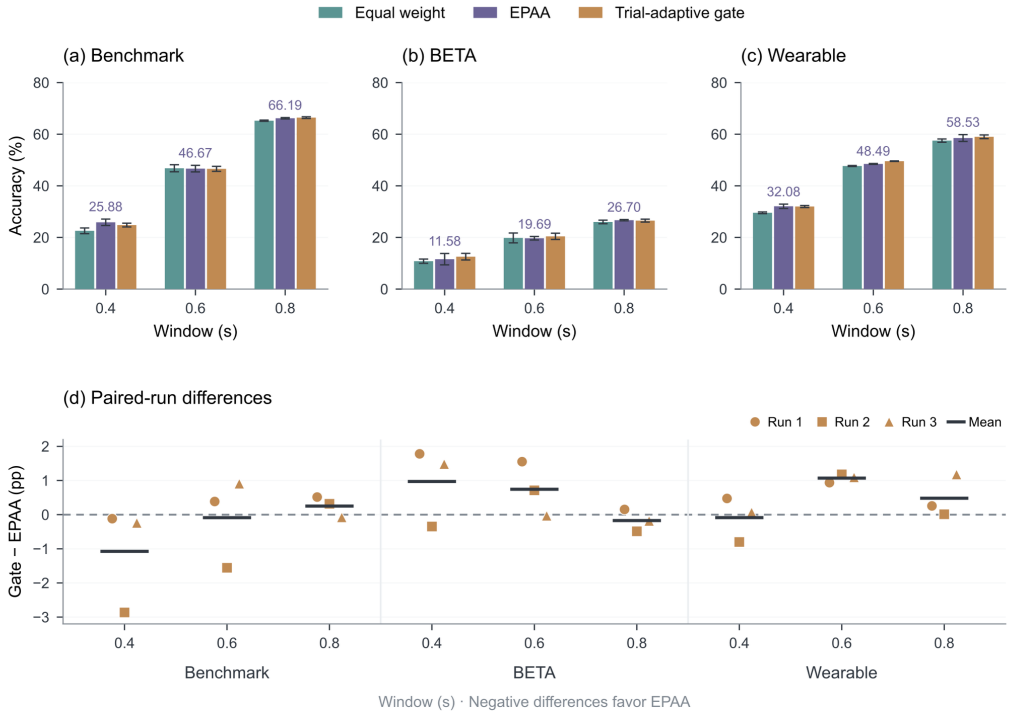

Figure 10: Fusion comparison across datasets and observation windows. (a–c) Bars show mean ± sample SD (ddof=1) over three paired runs. (d) Gate-minus-EPAA differences from the same runs; negative values favor EPAA. Method ordering varies by condition. These results are reported separately from A4, Figure 4, and Table 14.

G.2 Isolated evidence and scale controls

The controls below use matched low-order states and the same stage-2 budget, with residual pathways excluded from every endpoint. This separates scalar F1/F2 evidence from token-level F2 removal and separates normalization from the outer anchor scale. Fixed-score fusion and complete retraining are treated as separate experiments.

Table 14: One-factor fusion controls and accuracy at 0.6 s.

Control	Intervention	Benchmark	BETA	Wearable
F1 only	Retain F1 explicit anchor only	44.30	17.00	43.70
F2 only	Retain F2 explicit anchor only	43.80	18.00	46.00
No score normalization	Remove candidate-axis Z only	43.50	16.90	44.20
Fixed outer scale	γ=1; inner mixture learned	45.00	18.40	46.20
Small MLP	Shared nonlinear score mixer	45.20	18.70	46.80
Shared EPAA	Both anchors; normalized; γ learned	46.20	19.20	47.50

These means correspond to the controls defined in this table and are distinct from the A4 and Figure 4 readouts. The matched-endpoint comparison in Figure 10 is a separate analysis.

Weights reflect a parameterization and score scaling; they are not percentages of each pathway's causal contribution. Table 15 reports the shared weights.

Table 15: Shared weights at 0.6 s.

Dataset	**w1**	**w2**	γ
Benchmark	0.60	0.40	0.90
BETA	0.45	0.55	1.15
Wearable	0.40	0.60	1.20

Each row satisfies w1+w2=1 and the registered parameter bounds.

G.3 Capacity and Training Exposure Controls

A shared two-layer MLP encodes H1–H5 evidence tokens, pools within each candidate, concatenates the candidate-set mean, and produces a shared scalar score fused with the same explicit evidence. Full, context-disabled A1, single-branch A2, and this shared encoder use identical legal source trials, candidate tables, four views, and prespecified source-only selection budgets. Table 16 lists the capacity-matching targets and training exposure for these controls.

Each arm receives three stages of 1,200 updates. A2 and the shared MLP train the encoder/head first, continue with fusion second, then freeze these paths and fit residuals third. Widths are adjusted through source-only configuration selection to approach the capacity target. Active inference parameters, trainable parameters by stage, sampled examples, and training cost are distinct quantities; equal update counts do not establish equal computation. Table 19 and Figure 11 report computational cost separately.

Table 16: Capacity targets and training exposure. Ratios are targets relative to Full, rather than parameter counts.

arm	**Structure / context**	**Target parameter ratio**	**updates**
Full	Dual / on	1.00	$3 \times 1{,}200$
A1	Dual / off	1.00	$3 \times 1{,}200$
A2	H1-H5 single / on	1.00	$3 \times 1{,}200$
shared MLP	Pooled H1-H5 / on	1.00	$3 \times 1{,}200$

Table 16 specifies the matching targets and exposure for the comparison. It does not introduce additional shared-encoder accuracy results or redefine the Table 3 ablations.

H Direct Baselines, Additional Metrics, and Computational Cost

This appendix reports direct candidate-conditioned comparisons, Macro-F1, and computational cost. The main-table results and LINEAR values are reported in the main text.

H.1 Direct candidate-conditioned comparators

CCT denotes a Candidate-conditioned Transformer with shared EEG/reference encoders, candidate-set attention, and a shared scalar head. HCM denotes a harmonic CNN matcher with shared EEG–reference matching and no candidate-set attention. Both are adaptations to the present protocol, not official reproductions. Source-only selection follows Appendix D: 12 inner-fitting and four inner-validation sources, inner frequency holdouts drawn only from outer FIT frequencies, equal-subject

validation accuracy, and final fitting on all 16 sources at a frozen terminal checkpoint. Each method's grid combines learning rates 0.0001/0.0003/0.001 with weight decays 0/0.001; outer evaluation EEG is excluded.

Table 17: Candidate-conditioned baseline accuracy (%).

Dataset / window	CCT	HCM	Full − stronger (pp)
Benchmark / 0.4 s	27.60	27.40	+0.50
Benchmark / 0.6 s	48.90	47.80	+0.86
Benchmark / 0.8 s	67.70	66.80	+0.53
BETA / 0.4 s	13.90	13.50	+0.53
BETA / 0.6 s	20.80	20.30	+0.44
BETA / 0.8 s	30.90	30.40	+0.82
Wearable / 0.4 s	33.90	34.10	+0.61
Wearable / 0.6 s	47.80	47.70	+0.52
Wearable / 0.8 s	59.10	59.30	+0.39

Across the nine conditions, HarmTSFormer achieved mean accuracy gains of 0.39–0.86 percentage points over the stronger of CCT and HCM in each condition (Table 17). These gains were modest and should be considered alongside the computational costs reported in Appendix H.3.

Differences are computed from the reported Full mean and the stronger CCT/HCM comparator in each condition. This table reports point estimates; no SD, confidence interval, or p-value is reported.

H.2 Macro-F1 and scoring labels

True labels belong to the fold's held-out set U_f, but predictions retain the complete candidate set K without an oracle mask. For each $c \in U_f$, $F_{1,c} = \frac{2\mathrm{TP}_c}{2\mathrm{TP}_c+\mathrm{FP}_c+\mathrm{FN}_c}$. Predicting outside U_f still adds a false negative for the true class; classes outside U_f are excluded from that fold's macro average. A zero denominator gives zero. Compute each subject's within-fold class macro, weight folds by held-out-class count, then average subjects and registered seeds.

Table 18: Macro-F1 across nine conditions (%).

Dataset / window	Traditional comparator	Trad.	Full	CCT	HCM
Benchmark / 0.4 s	tlCCA-1	31.50	32.80	32.40	31.90
Benchmark / 0.6 s	tlCCA-1	52.20	54.10	53.30	52.10
Benchmark / 0.8 s	tlCCA-1	69.70	71.70	71.00	70.20
BETA / 0.4 s	tlCCA-1	8.60	17.00	16.20	15.80
BETA / 0.6 s	tlCCA-1	20.70	25.10	25.00	24.20
BETA / 0.8 s	tlCCA-1	34.80	37.00	35.90	35.40
Wearable / 0.4 s	FBCCA-H3	37.20	39.30	38.30	38.80
Wearable / 0.6 s	FBCCA-H3	53.10	53.70	53.50	53.00
Wearable / 0.8 s	CCA-H5	63.40	64.20	63.80	63.70

The traditional comparator is selected from the reported accuracy comparison, rather than by maximizing Macro-F1. Macro-F1 ranges from 0 to 100%. The table reports point estimates without uncertainty estimates.

H.3 COMPUTATIONAL COST AND TIMING SCOPE

Figure 11 reports parameter counts, computational cost, and latency across candidate counts and observation windows. The timing protocol uses one CUDA-capable GPU, batch one, and FP32. The hardware and timing settings are specified below. Preprocessing-inclusive latency comprises EEG filtering/normalization, transfer and the complete scoring path; forward latency isolates candidate construction and scoring. Neither includes EEG acquisition. Parameters count active inference weights; FLOPs count operations per trial, with two operations per multiply–add. Parameter millions and GFLOPs use 10^6 and 10^9, respectively.

The hardware environment comprises an NVIDIA A100 GPU, an eight-core x86 CPU, 64 GB RAM, Windows 11, Python 3.10, PyTorch 2.x, and CUDA 12.x. Timing uses 20 warm-up calls and 1,000 timed calls per condition, batch one, FP32, and GPU synchronization before and after timing. CUDA events cover GPU-only work; synchronized wall time covers filtering/normalization, materialization, transfer, and four-view scoring. Acquisition, display, disk I/O, and model loading are excluded.

Figure 11 uses Benchmark geometry (nine channels, 250 Hz), four views, and residuals. Active/allocated parameters are 0.800/0.806 M; the operation-count formula is 0.014+(T/0.8)(0.020+0.038K) GFLOPs. Timing summaries contain 1,000 calls per condition. Their medians and IQRs describe variability across timed calls. Table 19 separately reports stage-training costs.

Table 19: Stage-training cost per final fit (GPU-hours).

Method	**Stage**	**GPU-hours**
Full	H3	0.040
Full	H45 / aggregation	0.550
Full	Residual	0.050
Full	Total	0.640
CCT	End-to-end total	0.320
HCM	End-to-end total	0.180

Costs are reported for one dataset, window, frequency fold, and seed; Full's total is the sum of its three stages. Inner searches, failed attempts, and final evaluation are excluded. All costs use the single-GPU setup described above.

I SUPPLEMENTARY MATERIAL REFERENCE MAP

Table 20 maps the methods, experimental settings, analyses, and discussion in the main text to the corresponding appendix sections, tables, and figures. The code and supporting reproducibility materials are being prepared for anonymized release as described in the Data and Code Availability Statement and Appendix B.5.

Table 20: Main-text locations and the corresponding supplementary material.

Topic	**Main text**	**Appendix location**
Model definitions and score normalization	2.2–2.3	A.1–A.4; Tables 4–5
Training and run correspondence	2.4; 3.2	B.1–B.5; Tables 6–7
Source-only baseline adaptation and LINEAR	3.2–3.3	C.1–C.4; H.1

Continued on next page

Table 20 (continued)

Topic	Main text	Appendix location
Participant roles, frequency folds and selection	3.1–3.2	D.1–D.3; Tables 9–10; Figure 5
Mean gains and uncertainty	3.2–3.3	E; Tables 11–12; Figure 6
Frequency, phase and candidate-set analyses	3.2	F.1–F.5; Table 13; Figures 7–9
A4, dynamic fusion and capacity controls	3.4–3.5	G.1–G.3; Tables 14–16; Figure 10
Additional metrics and cost	3.2	H.2–H.3; Tables 18–19; Figure 11
Limitations and future work	4	J.1–J.2

J Limitations and Future Work

J.1 Limitations

Evaluation scope. The experiments use offline EEG from selected source and evaluation participants in three public datasets, with the specified channel subsets and wet-electrode Wearable recordings. The five main-protocol frequency folds retain the same participant allocation and are not independent subject partitions. Appendix E reports sensitivity analyses using alternative participant partitions within the same public datasets. Cross-subject evaluation within each dataset does not establish transfer to a new acquisition device, electrode montage, dry-electrode system, or clinical population. Closed-loop operation, asynchronous control, and reliable detection of periods without an intended target are outside the present evaluation.

Stimulus and candidate assumptions. Inference receives a candidate list containing distinct frequencies and their phases and selects one candidate from that list. The formulation therefore assumes that the intended stimulus is represented in the supplied set. The primary Benchmark and BETA folds retain the frequency endpoints in training and mainly assess interpolation. The outside-range tests in Appendix F use recorded targets beyond the fitting interval but still within the dataset stimulus range. They do not establish arbitrary-frequency or open-set recognition. Candidate-dependent feature normalization, pooled context, and score normalization can change scores when candidates are added or removed. Invariance to candidate ordering would not imply invariance to adding or removing candidates. The phase and candidate-set analyses characterize the tested interventions rather than every possible stimulus mismatch.

Interpretation and uncertainty. The harmonic branch split is an architectural prior, and the high-order branch's joint F2 reference still includes the first five harmonics. Perturbation sensitivity and readout accuracy describe model behavior but do not establish a unique physiological mechanism. Removing F2 changes several dependent paths, and the full model includes residual components beyond EPAA. These distinctions limit attribution of a combined gain to one isolated operator. Repeated-run standard deviations describe run variability, while participant-level comparisons address uncertainty conditional on the chosen cohorts and protocol. Baseline rankings apply to the stated source-only adaptations and do not rank methods under their original calibration protocols. Capacity targets in Table 16 and equal update counts also do not establish exact equality of parameter counts or computational cost.

Deployment and cost. The cost measurements describe the reported single-GPU configuration and workload. Forward and preprocessing-inclusive latency exclude EEG acquisition. The reported training cost covers one final fit for one dataset, observation window, frequency fold, and seed, and excludes inner searches, failed attempts, and final evaluation. A fixed parameter count across candidate sets therefore does not imply fixed computation, total development cost, or end-to-end interaction latency. The results do not establish real-time performance on a CPU or wearable processor, user comfort, or safety in an operational BCI.

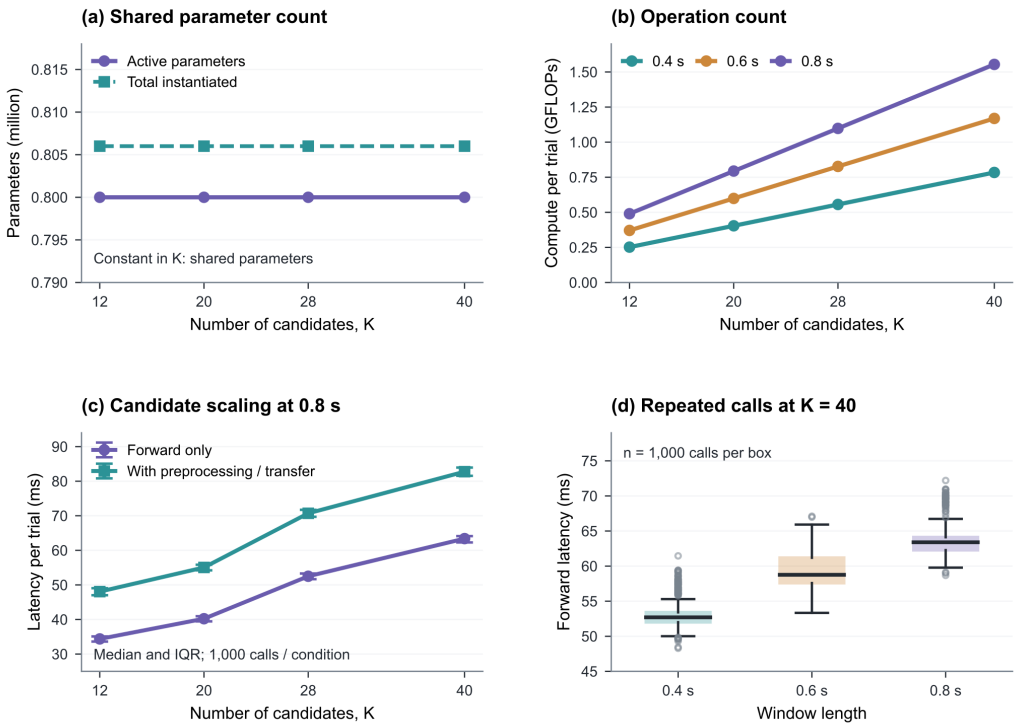

Figure 11: Parameter counts, GFLOPs, and latency. (c) Median and IQR at 0.8 s. (d) K=40 median, IQR, 1.5-IQR whiskers, and outliers. The 0.8 s/K=40 calls are identical across panels. Table 19 separately reports stage-training costs.

J.2 Future Work

Future evaluation should use prospectively specified cohorts and collect data across sessions, devices, electrode configurations, and stimulus schedules. Online studies should report selection accuracy together with acquisition-inclusive latency, unintended selections, and participant usability. Any target-data adaptation should be reported separately from the calibration-free setting so that its additional information and effort remain explicit.

Methodological priorities include sensitivity to uncertain stimulus timing and candidate metadata, broader candidate-set changes, and confidence-based rejection when the intended target is absent or evidence is insufficient. These evaluations should distinguish changes in candidate-dependent computation, including feature and score normalization and pooled context, from changes in decision competition. Model selection should remain source-only when assessing calibration-free decoding.

Further analysis should compare harmonic encoders and fusion rules under jointly controlled parameter capacity, training exposure, and measured computation, and use additional participant-allocation sensitivity analyses, repeated runs, and prospectively independent cohorts to assess the stability of gains. Prespecified participant-level analyses and the planned release of configurations, split files, checkpoints, and predictions would enable closer examination of failure cases and reproducibility.