BRAINSTATE: GRAPH-BASED STATE-AWARE GENERALIZATION FOR NON-INVASIVE BRAIN-TO-TEXT

Anonymous authors
Paper under double-blind review

ABSTRACT

EEG-based brain-computer interfaces (BCIs) offer a promising pathway for neural language decoding and human-machine interaction. Yet, their generalization is limited by heterogeneous distribution shifts across recordings. Existing transfer learning methods often seek task-relevant representations that are less sensitive to physiological and recording differences. Nevertheless, state-dependent variation can persist in non-stationary EEG and affect how information is transferred across subjects. To address this, we propose BrainState, a human state aware state-space-based hierarchical graph transfer learning framework for EEG-to-language. BrainState first combines spatiotemporal convolutions with Mamba blocks to capture local EEG patterns and temporal dependencies, mapping each trial to task-relevant content and recording-state representations. It then organizes training examples into a hierarchical graph memory according to class, subject, and session, enabling condition-specific modeling and state-aware retrieval. Finally, an attention-weighted memory readout aggregates relevant prototypes, while compatibility-based retrieval and gated fusion limit contributions from poorly matched recording conditions. BrainState thereby provides a way to incorporate recording state into cross-subject EEG speech decoding. Comprehensive experiments demonstrate the SOTA performance of the framework against strong baselines under cross-subject settings. Our code is available online.[1]

1 INTRODUCTION

Decoding language from brain activity is a central goal of brain-computer interfaces (BCIs). Invasive speech neuroprostheses already reach high accuracy (Willett et al., 2023; Metzger et al., 2023; Card et al., 2024), but require neurosurgery. Non-invasive EEG is safe and portable, and recent work has begun to decode words and sentences from it (Défossez et al., 2023; d'Ascoli et al., 2025; Lévy et al., 2026). Yet, EEG-to-text remains highly challenging as scalp signals have a low signal-to-noise ratio (SNR) and limited spatial resolution, and apparent progress can vanish under rigorous evaluation (Jo et al., 2024). A key bottleneck is cross-subject generalization. Decoders degrade on new users, and per-user calibration is costly (Edelman et al., 2024; Li et al., 2026).

This degradation stems from recording heterogeneity. The same command sentence elicits different scalp distributions across subjects, sessions and recording states, owing to fatigue, attention, electrode impedance and individual neural dynamics (Gu et al., 2021b; Jiang et al., 2025). Training recordings are therefore not equally informative about a new user that some recording conditions share transferable evidence with the target trial, whereas others differ enough to mislead the decoder.

Transfer learning remains to be explored in EEG-to-language decoding. Existing decoders mostly pool subjects with subject-specific layers or unified topologies (Défossez et al., 2023; Huang et al., 2025), treating the subject as a coarse domain and rarely modeling how recording conditions determine what can be transferred. General EEG transfer methods fall into two families. Invariance methods align, adversarially remove or disentangle domain information (He & Wu, 2019; Özdenizci et al., 2020; Jeon et al., 2021). Under non-stationary and limited data this separation is only approximate (Locatello et al., 2019; Zhao et al., 2019), so residual state remains. Removing state also discards the cue for judging which training experience applies. Prototype, memory and graph methods

[1] https://anonymous.4open.science/r/BrainState

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.

(Snell et al., 2017; Li et al., 2025; Niu et al., 2026) aggregate neighbors by content similarity alone, and can therefore draw on state-mismatched experience and cause negative transfer. We argue that recording state should not only be removed, but should also decide which experience is transferred. To this end, we propose BrainState, a state-aware graph transfer learning framework for non-invasive brain-to-text decoding (Figure 2). We denote state as a trial-level description of the recording condition, distinct from the recurrent state inside Mamba. (a) Firstly, a Mamba-based encoder (Gu & Dao, 2023) combines spatiotemporal convolutions with selective state-space modeling to capture local waveforms and long-range temporal dependencies, and maps each trial to content and state representations. (b) Then, training trials naturally form a hierarchical class–subject–session graph memory based on metadata. (c) The retrieval ranks memory nodes by content similarity penalized by state distance, so that evidence from poorly matched recording conditions is down-weighted. An attention-weighted readout aggregates the retrieved evidence, and a gate fuses it with the direct prediction. (d) Lastly, content–state de-correlation, state reconstruction and gradient-reversal adversarial objectives keep class content out of the state representation. In short, content decides what to retrieve, and state decides where to retrieve it from.

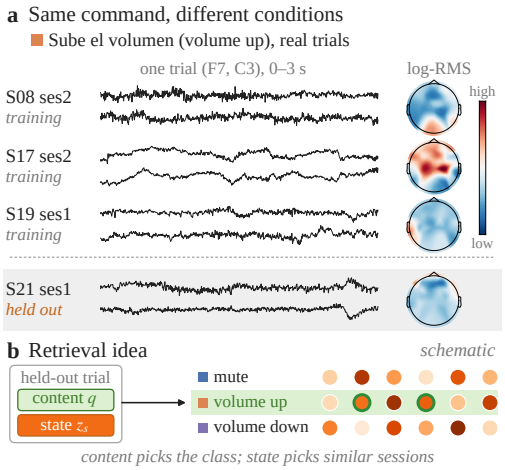


Figure 1: (a) Real trials of one covert command from three training sessions and a held-out subject, with condition-mean log-RMS maps on a shared scale. (b) Schematic of state-conditioned retrieval; dot shades are abstract state values.

Our main contributions are as follows:

1. To the best our knowledge, this is the first work to explicitly model recording-state heterogeneity for cross-subject transfer in non-invasive EEG-to-language decoding.
2. We propose a novel framework, BrainState. It treats the recording state as a retrieval condition rather than a nuisance to be suppressed. Whereas existing EEG decoders remove state variation as noise, BrainState disentangles each trial into content and state with a Mamba-based encoder and performs state-conditioned retrieval over a hierarchical graph memory, turning state variability into a source of information for decoding.
3. Under strict leave-one-subject-out (LOSO) evaluation, BrainState outperforms seven strong baselines on all six metrics, with ablation and attribution analyses validating the effectiveness.

2 Related Work

Decoding speech and language from non-invasive brain recordings is among the most challenging frontiers in brain–computer interfaces, with recent progress on perceived speech (Défossez et al., 2023), individual words (d'Ascoli et al., 2025), typed sentences (Lévy et al., 2026), multimodal semantics (Fan et al.), and covert or imagined speech (Nieto et al., 2022; Niu et al., 2026; Huang et al., 2025), often built on general EEG decoders (Lawhern et al., 2018; Song et al., 2023) or foundation models (Jiang et al., 2024; Wang et al., 2025). EEG remains the most demanding modality: within the same paradigm and cohort, EEG-based sentence decoding lags substantially behind MEG (Lévy et al., 2026). Recent covert speech decoders build graphs over electrodes arranged by scalp topology (Niu et al., 2026) or map electrodes to standardized brain regions (Huang et al., 2025), and some rely on eye-tracking data during pre-training (Huang et al., 2025).

A more fundamental obstacle is domain shift. Covert speech EEG is weak and exhibits heterogeneous non-stationarity across subjects, sessions, and recordings, limiting generalization to unseen subjects. Existing transfer methods, developed mostly for emotion recognition and motor imagery, either learn subject-invariant representations via mutual-information regularization (Jeon et al., 2021) or gradient-reversal adversarial training (Wang et al., 2024), or align brain dynamics

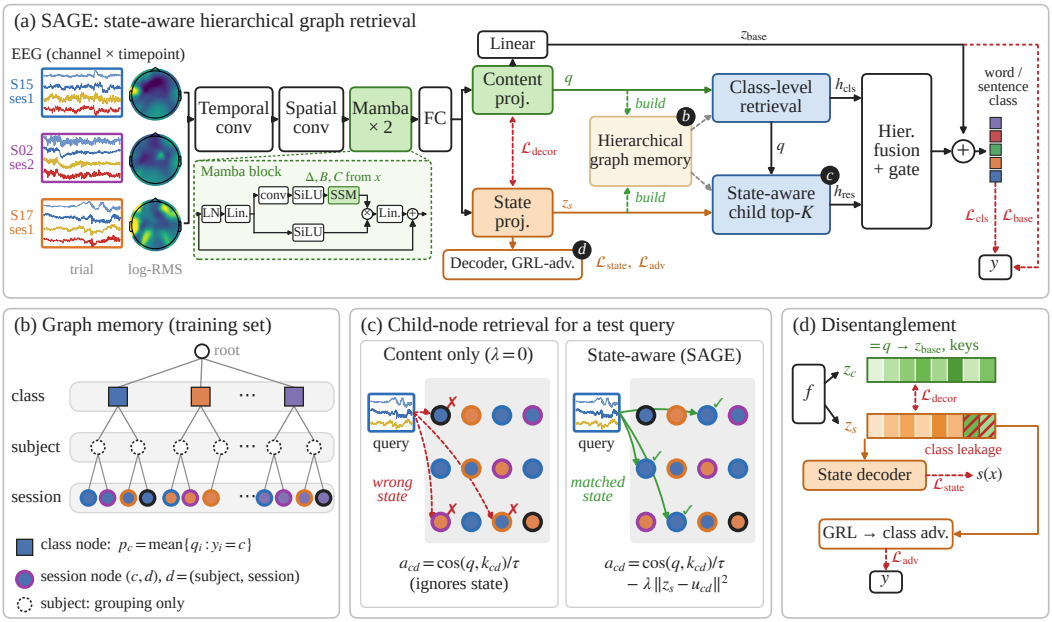


Figure 2: Overview of BrainState. (a) A Mamba-based encoder maps an EEG trial $x \in \mathbb{R}^{C \times T}$ to a content code $q = z_c$ and a state code z_s. The input examples are real TESSCCo trials of one command from three (subject, session) conditions, with their log-RMS maps. q retrieves class prototypes and, together with z_s, top-K session nodes from a training-set graph memory. The retrieved evidence is fused with the direct logits z_{base} to predict the word/sentence class. (b) Memory hierarchy, class $\to$ subject (grouping only) $\to$ session node (c, d), which stores a content key k_{cd} and a state u_{cd}. (c) BrainState ranks nodes by $a_{cd} = \cos(q, k_{cd})/\tau - \lambda\|z_s - u_{cd}\|^2$, which avoids state-mismatched experience that content-only retrieval ($\lambda = 0$) would select. (d) De-correlation, state reconstruction and gradient-reversal adversarial losses separate z_s from class content.

globally across subjects (Weng et al., 2026). The former discard subject- and session-specific state as nuisance, while the latter apply the same transfer to every test trial. Yet task-related EEG is itself modulated by the recording state, so decoding should be conditioned on the state rather than invariant to it. We therefore propose BrainState, a state-aware hierarchical graph transfer framework for EEG speech decoding. BrainState uses acquisition meta-information (subject and session) to organize training data into a hierarchical graph memory of class and class-by-condition nodes, and retrieves compatible experience for each test trial by combining content similarity with state matching, using signal-derived state without test-time identity labels. To our knowledge, BrainState is the first framework to condition cross-subject transfer on the recording state for EEG speech decoding. Appendix A provides a detailed review.

3 Methodology

Preliminary. State-space models (SSMs) have demonstrated strong performance in sequence modeling by evolving a compact latent state over time, including on long-range benchmarks (Gu et al., 2021a). Mamba has emerged as a prominent selective SSM, whose input-dependent dynamics retain relevant information while filtering less useful context, and its recurrent computation scales linearly with sequence length (Gu & Dao, 2023). Let $x \in \mathbb{R}^{C \times T}$ denote an EEG trial with C channels and T time samples, and let $y \in \{1, \ldots, K\}$ be its task label. For a token u_t after local convolution, a Mamba block updates its recurrent state and produces an output as

$$h_t = \exp(\Delta_t A) \odot h_{t-1} + \Delta_t B_t \odot u_t, \qquad o_t = C_t h_t + D \odot u_t, \tag{1}$$

where A and D are learned parameters, while Δ_t, B_t, and C_t depend on the input token. The input-dependent transition selects temporal information relevant to each EEG segment. We use h_t

for the recurrent state inside the encoder and z_s for the trial-level state representation used in graph retrieval.

The overall framework of BrainState is illustrated in Figure 2. In the following subsections, we first describe how the training trials form a graph memory, then the EEG encoder, and finally the state-conditioned readout.

3.1 HIERARCHICAL GRAPH MEMORY

Subject and session identifiers provide a physically grounded domain prior for organizing training EEG trials. A subject groups recordings from one individual, while a session identifies a particular acquisition period. We organize task evidence by class c and condition $d = (\text{subject}, \text{session})$.

To model the evidence shared across recording conditions at the class level while preserving condition-specific patterns in the child nodes, each observed class–condition pair forms a child node under its class, with subject identity, providing an intermediate grouping in the hierarchy. Given a content embedding q_i and a state embedding $z_{s,i}$ for each training trial, these nodes store

$$\begin{aligned} p_c &= \text{mean}_{i:y_i=c}\, q_i, \quad & k_{cd} &= \text{mean}_{i:y_i=c,\, d_i=d}\, q_i, \\ r_{cd} &= k_{cd} - p_c, & u_{cd} &= \text{mean}_{i:y_i=c,\, d_i=d}\, z_{s,i}. \end{aligned} \tag{2}$$

The class prototype p_c summarizes the content shared by trials of class c. The child-node key k_{cd} captures that class under condition d, and its residual r_{cd} describes the deviation from the class-wide prototype. The state prototype u_{cd} summarizes the states observed in the same class–condition group. Together, these quantities let a query retrieve shared class evidence and select condition-specific residuals according to its content and state. The memory is refreshed from training trials without gradients after each epoch. For a training query, its stored trial is excluded when computing the prototypes and residuals used in its retrieval. Validation and test queries read the frozen memory associated with the selected model checkpoint.

3.2 MAMBA ENCODING AND CONTENT–STATE REPRESENTATIONS

BrainState first applies a temporal convolution to capture local waveforms, a spatial convolution across EEG channels, and temporal average pooling to form a token sequence. Each of two Mamba blocks normalizes its input and projects it into a sequence branch and a gating branch. A causal depthwise convolution supplies local temporal context to the sequence branch; input-dependent Δ_t, B_t, and C_t then drive the selective state-space scan in Eq. 1. The scan output is modulated by the gate, projected back to the token dimension, and added to the block input through a residual connection. A fully connected layer maps the resulting sequence to a trial feature f. Separate projections of f produce the content query $q = z_c \in \mathbb{R}^{128}$ and the state representation $z_s \in \mathbb{R}^{128}$. A linear classifier on q produces direct logits ℓ_{base}; the same q supplies the content keys and queries in the graph memory.

The state branch is trained to reconstruct a standardized trial descriptor $s(x)$ comprising per-channel temporal means, log root-mean-square amplitudes, and log root-mean-square first differences. A cross-correlation penalty separates q and z_s across the minibatch. A class adversary connected through gradient reversal shapes z_s to carry less task-label information. Together, these constraints allocate task evidence to the content representation and recording-condition information to the state representation.

3.3 STATE-CONDITIONED RETRIEVAL AND PREDICTION

Imagined-speech EEG exhibits subject-dependent spatial and spectral patterns Ma et al. (2025). For query (q, z_s), BrainState ranks each available class–condition node by

$$a_{cd} = \frac{\cos(q, k_{cd})}{\tau} - \lambda \frac{\|z_s - u_{cd}\|_2^2}{\dim(z_s)}, \tag{3}$$

where $\tau > 0$ controls content similarity and $\lambda \geq 0$ controls state agreement. Figure 1b sketches this ranking: content selects the class, and state favours training sessions recorded in a similar state. Softmax weights over the top-M scores aggregate projected residuals into h_{res}. In parallel, content

similarity to p_c aggregates class prototypes into h_{cls}. With retrieval-confidence features ξ, the model forms

$$\tilde{q} = \text{LN}(q + h_{\text{cls}} + \beta(\xi)h_{\text{res}}), \qquad \ell = \ell_{\text{base}} + g(q, \tilde{q}, \xi)\,\delta(\tilde{q}), \tag{4}$$

where β and g are sigmoid gates and δ maps fused evidence to a logit correction. Thus, the trial state changes which condition-specific residuals are retrieved.

Training minimizes the cross-entropy of both fused and direct logits, together with state reconstruction, content–state decorrelation, and the gradient-reversal adversarial objective:

$$\mathcal{L} = \mathcal{L}_{\text{cls}} + \mathcal{L}_{\text{base}} + w_s\mathcal{L}_{\text{state}} + w_d\mathcal{L}_{\text{decor}} + w_a\mathcal{L}_{\text{adv}}. \tag{5}$$

The adversary minimizes $\mathcal{L}_{\text{adv}}$ to predict the class, while gradient reversal sends its opposite gradient to the state encoder.

4 EXPERIMENTS

Datasets. We evaluate our method on the TV-control EEG-based Silent Speech Command Corpus (TESSCCo) Lobo-Alonso et al. (2026). TESSCCo is a bilingual command-level silent speech dataset. It contains EEG recorded while 24 healthy subjects (21 native and 3 non-native Spanish speakers) produced five TV-control commands (e.g., "Turn up volume", "Mute FireTV") in both English and Spanish, under both covert and overt speech conditions. Recordings were acquired with a 32-channel water-based EEG headset (Bitbrain Versatile E32) at 256 Hz over up to two sessions per subject, yielding 7,936 epochs (11.02 h) in total. In this work, we use the 21 native Spanish speakers and exclude the three non-native speakers.

Data Preprocessing and Splitting We follow the procedures specified in the original publication for each dataset and use the officially released preprocessed data when available. Further details are provided in Appendix B.2.

We primarily target at the transfer learning challenge in EEG-to-language, especially focus on the cross-subject performance. Therefore we use strict cross-subject leave-one-subject-out (LOSO) protocol to split the dataset. For a dataset with N subjects, each fold holds out one subject s for testing, assigns the preceding subject in cyclic order, $(s-1) \bmod N$, to validation, and uses the remaining $N-2$ subjects for training. All data-dependent preprocessing is fitted on the training subjects only.

Baselines and Evaluation. We compare BrainState with seven baselines in four categories: general EEG decoding models (EEGNet Lawhern et al. (2018) and EEGConformer Song et al. (2023)), EEG foundation models (LaBraM Jiang et al. (2024) and EEGMamba Wang et al. (2025)), methods for cross-subject representation learning (StateMamba Weng et al. (2026) and SICR Jeon et al. (2021)), and a carefully-designed imagined-speech decoding model BDR-GCL Niu et al. (2026). Further details are provided in Appendix J.

We evaluate the proposed method with comprehensive metrics, including accuracy, precision, recall, F1 score, macro AUROC, and Cohen's kappa. For each metric, we first average results across random seeds within each LOSO fold, then report the mean ± standard deviation across held-out subjects. For efficiency consideration, the number of trainable parameters in millions (to three decimal places) and FLOPs per sample are reported. Paired significance tests are used to compare models on the same held-out subjects.

Implementation Details BrainState uses a two-block Mamba EEG encoder with a 256-dimensional backbone representation. The content and state branches each produce 128-dimensional embeddings. Graph-memory retrieval uses the top 10 nodes, a temperature of 0.2, and a state-distance weight of 1.0. The memory is built from training trials and refreshed after each epoch. We train with AdamW optimizer Kingma & Ba (2014) using a learning rate of 10^{-3}, weight decay of 10^{-4}, and a batch size of 64. Training runs for at most 200 epochs, with early stopping after 15 epochs without improvement. Experiments are repeated with three random seeds: 0, 42, and 3407. Computation is performed on four nodes equipped with four NVIDIA L4, two NVIDIA RTX A5500, and two NVIDIA L40S GPUs in total. The software environment uses Python 3.10.18 and PyTorch 2.4.1 with CUDA 12.1.

Table 1: Performance comparison on the TESSCCo dataset under the cross-subject leave-one-subject-out (LOSO) setting. Results are reported as mean ± standard deviation across subjects. The best results are in **bold** and the second best are underlined.

Method	Accuracy (%) ↑	Precision (%) ↑	Recall (%) ↑	F1 (%) ↑	Macro AUROC (%) ↑	Cohen's κ ↑
EEGNet	26.16 ± 3.66	27.16 ± 6.43	26.09 ± 3.73	24.68 ± 3.77	54.10 ± 3.59	0.076 ± 0.046
EEGConformer	23.96 ± 4.20	20.83 ± 7.12	23.82 ± 4.27	20.18 ± 6.08	52.11 ± 2.97	0.048 ± 0.053
EEGMamba	21.64 ± 2.09	8.35 ± 5.45	21.34 ± 1.88	10.03 ± 4.35	52.23 ± 3.27	0.017 ± 0.024
LaBraM	24.57 ± 3.42	23.90 ± 8.81	24.51 ± 3.47	18.78 ± 4.49	52.64 ± 4.45	0.056 ± 0.044
SICR	22.76 ± 3.20	21.93 ± 4.47	22.78 ± 3.12	20.34 ± 2.92	51.30 ± 3.93	0.035 ± 0.039
BDRGCL	21.58 ± 1.89	12.79 ± 8.90	21.34 ± 2.01	10.67 ± 4.24	50.70 ± 4.21	0.016 ± 0.025
StateMamba	26.75 ± 2.15	27.27 ± 8.13	26.78 ± 2.48	21.25 ± 4.55	52.64 ± 2.87	0.084 ± 0.030
Ours	**29.94 ± 3.37**	**30.13 ± 3.86**	**29.93 ± 3.26**	**28.21 ± 3.79**	**54.52 ± 2.86**	**0.124 ± 0.041**

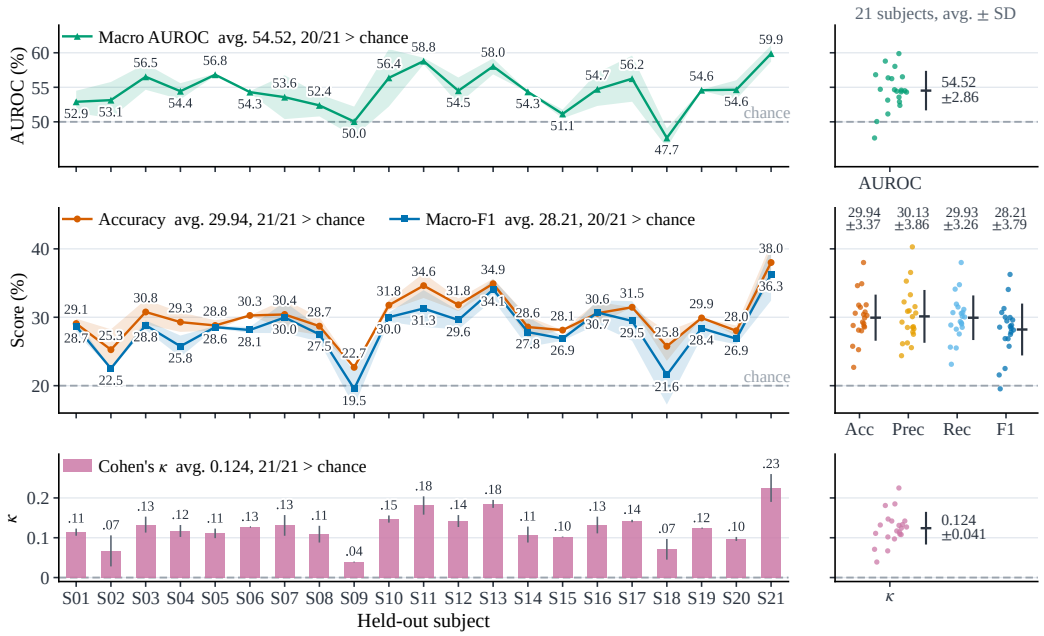


Figure 3: Per-subject results of BrainState on the TESSCCo dataset under cross-subject LOSO. Lines show the mean value, shaded bands show ±SD, and dashed lines mark chance level.

4.1 Results

Table 1 reports the cross-subject LOSO results on the native Spanish subset of the TESSCCo dataset, a five-class task with a chance level of 20%. We observe that this setting is especially challenging for all methods and we attribute this mainly to the limited scale of the recordings, with limited spatiotemporal resolution (32 channels, 256 Hz), and tiny data scale (fewer than 1,800 samples in total). These factors restrict the spatial and spectral detail available to the models, as well as the amount of data they can learn subject-invariant representations from. Even under these conditions, our method achieves the best performance on all six metrics, reaching $29.94 \pm 3.37\%$ accuracy, $30.13 \pm 3.86\%$ precision, $29.93 \pm 3.26\%$ recall, $28.21 \pm 3.79\%$ F1, $54.52 \pm 2.86\%$ macro AUROC, and a Cohen's κ of 0.124 ± 0.041. No single baseline is consistently second best. StateMamba is the strongest competitor on accuracy, precision, recall, and κ, while EEGNet is second on F1 and macro AUROC. Compared with the strongest baseline on each metric, our method improves accuracy by 3.19, precision by 2.86, recall by 3.15, and F1 by 3.53 percentage points. Relative to chance, our method is 9.94 points above the 20% chance level, while StateMamba is 6.75 points above it, a margin that is about 47% larger. The same gain appears in Cohen's κ, which rises from 0.084 to 0.124.

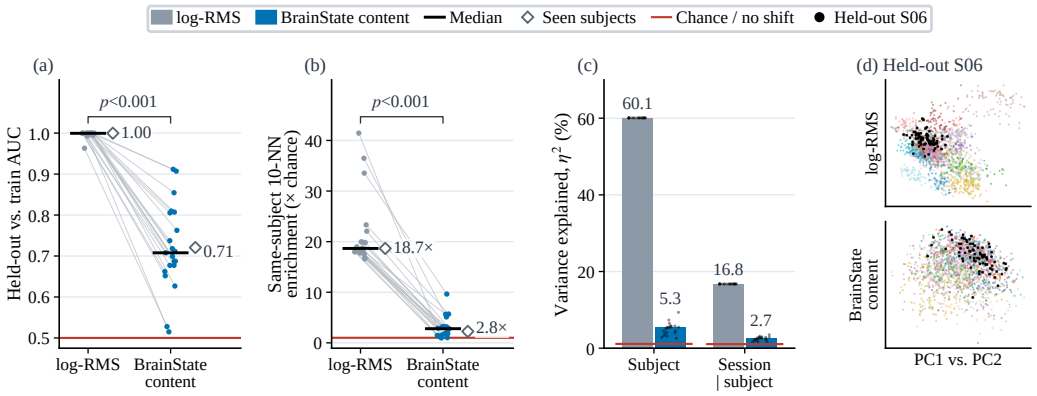


Figure 4: Subject shift on TESSCCo before (log-RMS) and after BrainState encoding (content representation z_c). (a) Cross-validated AUC for separating the held-out subject's trials from the training trials. (b) Enrichment of same-subject trials among the held-out subject's 10 nearest neighbours, relative to chance. (c) Variance explained (η^2) by subject and by session within subject over all 1792 trials. (d) PCA of one representative fold (held-out S06, median test accuracy); held-out trials in black, other subjects in colour.

The challenge of this dataset is also visible in the baseline results. Most baselines reach only 21.6–26.8% accuracy, which is just above chance, and several show a large gap between accuracy and the class-balanced metrics. EEGMamba and BDRGCL both reach about 21% recall, but their precision (8.35% and 12.79%) and F1 (10.03% and 10.67%) are much lower, and their κ values are close to zero (0.017 and 0.016). This pattern suggests that their predictions largely collapse onto one or two classes when tested on held-out subjects. Even the stronger baselines are unstable across subjects: the standard deviation of their precision ranges from 6.43 to 8.90 points. In contrast, our method keeps accuracy, precision, recall, and F1 closely aligned, between 28.2% and 30.1%, and has the lowest precision variance of all methods ($\pm$3.86). These results suggest that our model extracts more discriminative and subject-consistent representations even when channels, sampling rate, and training data are limited.

4.2 Cross-subject shift before and after encoding.

Figure 4 examines whether BrainState reduces the subject-specific shift that dominates the raw TESSCCo recordings. We describe each trial by its per-channel log-RMS amplitude and by its BrainState content representation z_c, and evaluate the held-out subject of each of the 21 LOSO folds. In the log-RMS space, a linear classifier separates the held-out subject from the training trials almost perfectly (median AUC 1.00), and its trials are 18.7 times more likely than chance to be each other's nearest neighbours. In the content space, the AUC drops to 0.71 and the enrichment to 2.8×, and both decrease in all 21 folds ($p < 0.001$). These values are close to those of the subjects seen during training (0.72 and 2.3×), which indicates that the reduction extends to unseen subjects rather than reflecting a fit to the training subjects. Across all trials, subject identity explains 60.1% of the log-RMS variance but only 5.3% of the content variance, and session within subject explains 16.8% and 2.7%, respectively. The state representation z_s retains more subject information than z_c, consistent with its role of carrying recording conditions. The shift is not removed completely: the AUC stays above 0.5 and reaches 0.91 for S04 and S12. Moreover, aligning the held-out subject to the training distribution does not by itself make its classes separable, since class explains 4.5% of its content variance, at the random-label level of 4.4%.

4.3 Feature-space separability

To examine what the encoder learns, we project the same trials into two dimensions with t-SNE, before encoding (raw EEG) and after encoding (BrainState content features) (Figure 5). In the raw EEG space, the five commands collapse into a single overlapping cloud. After encoding, trials of the same command gather in distinct regions (e.g., *power on*, *power off* and *volume down*), with each

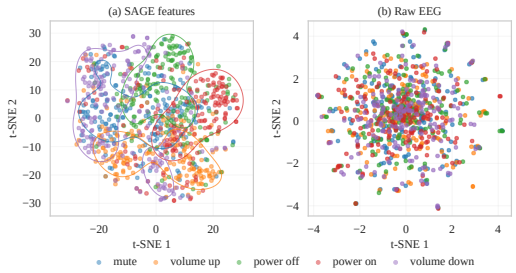


Figure 5: t-SNE of held-out subject S21 before encoding (raw EEG) and after encoding (BrainState content features), noted State-Aware GEneralization (SAGE).

Table 2: Class structure of representations for held-out subject S21 (chance = 20%). Pooled: all subjects; Transfer: fit on source subjects, tested on S21. $^{\dagger}p \leq 0.001$ (permutation).

	Raw EEG (24,576-d)	Raw PCA (50-d)	BrainState (128-d)
Pooled			
1-NN (%)	20.5	20.5	50.1†
10-NN LOO (%)	18.8	19.8	58.5†
Silhouette	-0.018	-0.028	$+0.010$
Transfer to S21			
10-NN vote (%)	18.0	20.0	30.0
Linear probe (%)	22.0	18.0	38.0

Table 3: Ablation of the core BrainState components on TESSCCo (Native, Spanish; 5-class) under cross-subject LOSO evaluation. † denotes significantly lower accuracy than the full BrainState (paired t-test across subjects, $p < 0.05$).

	Accuracy (%) per subject			Average			
Variant	S01	S04	S21	Accuracy (%) ↑	F1 (%) ↑	AUROC (%) ↑	Cohen's κ ↑
A: w/o graph	$22.79_{\pm 2.12}$	$26.60_{\pm 3.82}$	$32.67_{\pm 1.15}$	$27.35_{\pm 4.98}^{\dagger}$	$24.87_{\pm 3.14}$	$53.36_{\pm 2.67}$	$0.092_{\pm 0.062}$
B: + content graph	$25.51_{\pm 3.68}$	$25.93_{\pm 1.54}$	$32.67_{\pm 4.16}$	$28.03_{\pm 4.02}^{\dagger}$	$25.73_{\pm 3.38}$	$53.55_{\pm 2.22}$	$0.100_{\pm 0.050}$
C: + fixed state	$25.51_{\pm 3.68}$	$26.60_{\pm 2.54}$	$32.67_{\pm 4.16}$	$28.26_{\pm 3.86}^{\dagger}$	$26.16_{\pm 3.02}$	$53.80_{\pm 2.42}$	$0.103_{\pm 0.048}$
D: + learned state	$26.53_{\pm 2.89}$	$27.27_{\pm 1.43}$	$36.00_{\pm 0.00}$	$29.93_{\pm 5.27}^{\dagger}$	$27.78_{\pm 3.94}$	$54.87_{\pm 6.29}$	$0.124_{\pm 0.066}$
Ours (full BrainState)	$\mathbf{29.08}_{\pm \mathbf{0.72}}$	$\mathbf{29.29}_{\pm \mathbf{1.43}}$	$\mathbf{38.00}_{\pm \mathbf{2.83}}$	$\mathbf{32.12}_{\pm \mathbf{5.09}}$	$\mathbf{30.23}_{\pm \mathbf{5.42}}$	$\mathbf{55.74}_{\pm \mathbf{3.67}}$	$\mathbf{0.152}_{\pm \mathbf{0.063}}$

class's high-density region displaced from the others. Some overlap remains, reflecting the difficulty of non-invasive decoding, but BrainState turns nearly indistinguishable inter-class differences into locally separable structure.

Quantitatively (Table 2), the same-class nearest-neighbour rate and 10-NN accuracy are 20.5% and 18.8% in raw EEG space, both at chance ($p > 0.3$, permutation test), and rise to 50.1% and 58.5% in BrainState feature space ($p \leq 0.001$). On the target subject, 10-NN accuracy using source-subject trials rises from 18% to 30% and a linear probe from 22% to 38%, indicating that target-subject trials are better aligned with same-class source trials.

4.4 Ablation Study

We conduct the ablation study on three randomly selected held-out subjects (S01, S04 and S21) from the TESSCCo dataset. Table 3 evaluates progressively augmented variants that share the same Mamba-based encoder, content projection and hyperparameters (variant A-D). A classifies the content embedding directly without a graph. B adds the hierarchical graph memory built from training trials, whose retrieved evidence corrects the direct prediction through a gate. C further conditions retrieval on a fixed statistical state (channel-wise mean, log-RMS and first-difference log-RMS), while D replaces it with a learned state embedding trained with a state reconstruction loss. The full BrainState additionally adds direct classification, content–state decorrelation and class-adversarial objectives. Each component improves accuracy, from 27.35% (A) to 28.03% (B), 28.26% (C) and 29.93% (D), with the learned state yielding the largest single gain. The full BrainState performs best on all metrics(Table 7), reaching 32.12% accuracy (+4.77 pp over A).

4.5 Cross-Subject Error and Attribution Analysis

To characterize cross-subject generalization, we analyze test predictions for four held-out subjects and, for Subject-21, compare Input×Gradient (IxG) channel attributions in correct and incorrect decisions with scalp potential distributions (Figure 6). Across the four selected subjects (A), accuracy

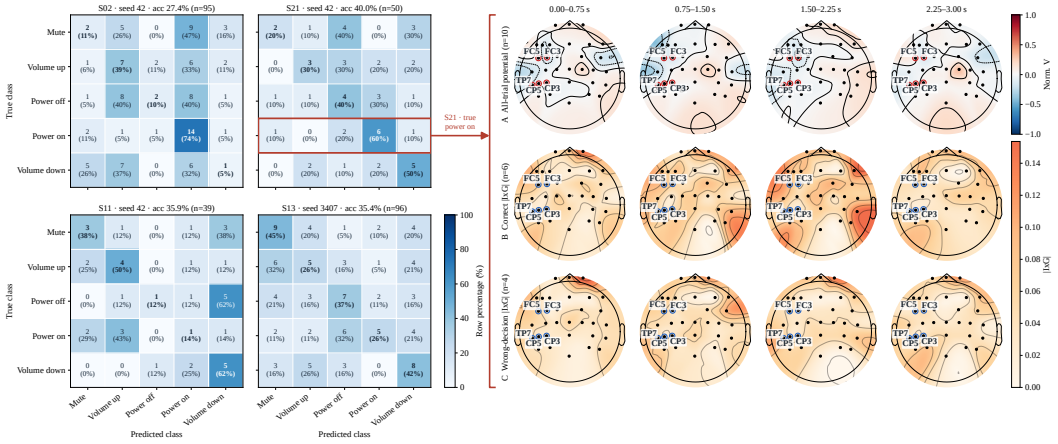


Figure 6: Correct and misclassified *power on* trials draw on different scalp regions. Left: confusion matrices of four held-out subjects; the outlined *power on* row is expanded on the right. Right: maps averaged over four 0.75-s windows (0–3 s after cue onset). (A) Baseline-corrected potential of all 10 *power on* trials (shared scaling, unitless). (B, C) Absolute Input×Gradient of the predicted-class logit, peak-normalized per trial, for correct (B) and misclassified (C) trials. Circled electrodes: significant channels from the TESSCCo ANOVA Lobo-Alonso et al. (2026).

ranges from 27.4% to 40.0% (chance 20%). Errors stem mainly from subject-specific output biases rather than a fixed set of hard commands. Each model favours a different class (e.g., *power on* for S02, 45% of predictions; *volume down* for S11, 38%), and confusions between paired commands such as *power on*/*power off* occur only for some subjects (S21, S13). Both correct (B) and incorrect (C) decisions show frontal IxG attribution, with frontopolar electrodes receiving 22–28% and 27–32% of attribution, respectively, versus 15.6% under a uniform distribution. Correct trials additionally show right temporo-parietal foci at 0.75–2.25 s. IxG correlates more strongly with absolute scalp potential for incorrect trials (Spearman $\rho = 0.28$–0.35) than for correct trials ($\rho = 0.14$–0.30), suggesting greater reliance on large-amplitude frontal channels during errors. To summarize, the model extracts class-relevant information from every held-out subject, but a large domain gap between subjects limits transfer. Each held-out subject shifts the decision boundaries differently, producing subject-specific output biases rather than confusions between similar commands. The model relies on a shared frontal pattern, and errors track the high-amplitude frontopolar channels more closely. This gap may partly stem from the covert paradigm itself: silent articulation produces orofacial, facial, and eye-movement activity at frontal and temporal electrodes that varies from person to person.

5 Conclusion

This work addresses cross-subject generalization in non-invasive EEG-to-language decoding, where recordings from different subjects and sessions differ in their recording state. Rather than only suppressing this variation, BrainState treats recording state as a retrieval condition. A Mamba-based encoder maps each trial to content and state representations, and a hierarchical class–condition graph memory is queried by content similarity penalized by state distance, so that evidence from poorly matched recording conditions contributes less to the prediction. Under strict leave-one-subject-out evaluation on the native Spanish subset of TESSCCo, BrainState outperforms seven baselines on all six metrics, and ablations indicate that state-conditioned retrieval with a learned state representation and the content–state separation objectives account for most of the improvement over a graph-free encoder.

Yet, there are limitations. Future work should evaluate state-conditioned retrieval on larger word- and sentence-level corpora and across sessions and recording devices, which is a prerequisite for EEG–language foundation models that transfer to new users without per-user calibration (Banville et al., 2026).

AI USE STATEMENT, ETHICS AND REPRODUCIBILITY

AI USE STATEMENT

In this work, Large Language Models (LLMs), including GPT-6 and Claude-Fable, were used as auxiliary tools to support manuscript preparation and implementation.

AI-assisted tools were used to support language refinement, improve the clarity and presentation of the manuscript, and assist with final checks of grammar, wording, consistency, and formatting. We apply a typical prompt: "Refine the following sentences to improve clarity and conciseness while preserving technical meaning."

We also consult AI tools to broaden the exploration of potentially relevant literature beyond the key references already known to the authors. All references included in the manuscript were independently verified by the authors against the original sources, including their existence, bibliographic information, relevance to this work, and the claims for which they are cited.

For experiments, LLMs were occasionally consulted to assist in coding tasks such as generating a specific function using Python.

All AI-assisted outputs used in preparing this work were carefully reviewed and verified by the authors. And the authors take full responsibility for the final content of this work, including all text, claims, figures, references, and other artifacts prepared with AI assistance.

ETHICS STATEMENT

This study conducts secondary analysis of previously released human EEG data and involves no new participant recruitment, intervention, or data collection. All data were obtained from publicly available research resources for which the original studies documented the relevant informed consent and ethical review procedures. We used only anonymized or pseudonymized data released by the original data providers and did not access personally identifiable information, attempt participant reidentification, or link EEG recordings with external personal records. All analyses rely solely on the identifiers and information provided in the official releases. Raw EEG recordings are not redistributed as part of this work, and subsequent use should follow the licenses, citation requirements, and participant protection conditions established by the original data providers.

The proposed method is intended solely for research on EEG-based language representation learning and neural decoding under controlled experimental conditions. Its predictions depend on the experimental paradigm, available training data, and predefined task space, and should not be interpreted as unrestricted access to private thoughts or mental content. Model-derived spatial patterns represent statistical contributions of scalp EEG measurements and, given volume conduction and the limited spatial resolution of EEG, should not be interpreted as precise cortical source localisation or used for clinical or neurological diagnosis. The present computational study introduces no additional physical or psychological risk to the original participants.

REPRODUCIBILITY STATEMENT

We work towards transparency, several steps are taken to facilitate reproducibility.

The proposed model architecture, learning objectives, and training procedures are described in detail in the main paper. Dataset preprocessing, experimental protocols, data splits, evaluation metrics, hyperparameter settings, and implementation details are documented in the sec 4 and supplementary materials. All experiments are conducted using explicitly reported random seeds and consistent evaluation protocols across all compared methods. Anonymized source code, configuration files, and scripts required to reproduce the reported results are also provided.

Following acceptance, we welcome interested researchers to contact us for further technical details, clarification, and discussion.

REFERENCES

Hubert Banville, Stéphane d'Ascoli, Simon Dahan, Jérémy Rapin, Marlène Careil, Yohann Benchetrit, Jarod Lévy, Saarang Panchavati, Antoine Ratouchniak, Elisa Cascardi, et al. Neuralbench: A unifying framework to benchmark neuroai models. *arXiv preprint arXiv:2605.08495*, 2026.

Tingrui Bao and Rumei Lu. Mutual information-guided subject disentanglement for cross-subject eeg speech envelope reconstruction. *Array*, pp. 101052, 2026.

Nicholas S. Card, Maitreyee Wairagkar, Carrina Iacobacci, Xianda Hou, Tyler Singer-Clark, Francis R. Willett, Erin M. Kunz, Chaofei Fan, Maryam Vahdati Nia, Darrel R. Deo, et al. An accurate and rapidly calibrated speech neuroprosthesis. *New England Journal of Medicine*, 391(7):609–618, 2024. doi: 10.1056/NEJMoa2314132.

Alexandre Défossez, Charlotte Caucheteux, Jérémy Rapin, Ori Kabeli, and Jean-Rémi King. Decoding speech perception from non-invasive brain recordings. *Nature Machine Intelligence*, 5 (10):1097–1107, 2023. doi: 10.1038/s42256-023-00714-5.

Stéphane d'Ascoli, Corentin Bel, Jérémy Rapin, Hubert Banville, Yohann Benchetrit, Christophe Pallier, and Jean-Rémi King. Towards decoding individual words from non-invasive brain recordings. *Nature Communications*, 16(1):10521, 2025.

Bradley J Edelman, Shuailei Zhang, Gerwin Schalk, Peter Brunner, Gernot Müller-Putz, Cuntai Guan, and Bin He. Non-invasive brain-computer interfaces: state of the art and trends. *IEEE reviews in biomedical engineering*, 18:26–49, 2024.

Jiahao Fan, Zongsheng Li, Benxiang Xiao, Qingzhu Zhang, Xinke Shen, and Quanying Liu. Semantic-guided contrastive learning for eeg multimodal decoding of listening and watching.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of machine learning research*, 17(59):1–35, 2016.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396*, 2021a.

Xiaotong Gu, Zehong Cao, Alireza Jolfaei, Peng Xu, Dongrui Wu, Tzyy-Ping Jung, and Chin-Teng Lin. Eeg-based brain-computer interfaces (bcis): A survey of recent studies on signal sensing technologies and computational intelligence approaches and their applications. *IEEE/ACM transactions on computational biology and bioinformatics*, 18(5):1645–1666, 2021b.

He He and Dongrui Wu. Transfer learning for brain–computer interfaces: A euclidean space data alignment approach. *IEEE Transactions on Biomedical Engineering*, 67(2):399–410, 2019.

Shuai Huang, Yongxiong Wang, and Huan Luo. Ccsumsp: A cross-subject chinese speech decoding framework with unified topology and multi-modal semantic pre-training. *Information Fusion*, 119:103022, 2025.

Eunjin Jeon, Wonjun Ko, Jee Seok Yoon, and Heung-Il Suk. Mutual information-driven subject-invariant and class-relevant deep representation learning in bci. *IEEE Transactions on Neural Networks and Learning Systems*, 34(2):739–749, 2021.

Wei-Bang Jiang, Li-Ming Zhao, and Bao-Liang Lu. Large brain model for learning generic representations with tremendous eeg data in bci, 2024. URL https://arxiv.org/abs/2405.18765.

Xiaowei Jiang, Beining Cao, Liang Ou, Yu-Cheng Chang, Thomas Do, and Chin-Teng Lin. ifuzzytl: Interpretable fuzzy transfer learning for steady-state visual evoked potentials brain–computer interfaces system. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2025.

Hyejeong Jo, Yiqian Yang, Juhyeok Han, Yiqun Duan, Hui Xiong, and Won Hee Lee. Are EEG-to-text models working? *arXiv preprint arXiv:2405.06459*, 2024.

Urvashi Khandelwal, Omer Levy, Dan Jurafsky, Luke Zettlemoyer, and Mike Lewis. Generalization through memorization: Nearest neighbor language models. *arXiv preprint arXiv:1911.00172*, 2019.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Byung-Kwan Ko, Soowon Kim, and Seo-Hyun Lee. Subject-independent imagined speech detection via cross-subject generalization and calibration. In *2026 14th International Conference on Brain-Computer Interface (BCI)*, pp. 1–5. IEEE, 2026.

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: 10.1088/1741-2552/aace8c.

Jarod Lévy, Mingfang Zhang, Svetlana Pinet, Jérémy Rapin, Hubert Banville, Stéphane d'Ascoli, and Jean-Rémi King. Noninvasive decoding of typed sentences from human brain activity. *Nature Neuroscience*, pp. 1–7, 2026.

Guangli Li, Canbiao Wu, Zhehao Zhou, Na Tian, Li Zhang, and Zhen Liang. Learning domain- and class-disentangled prototypes for domain-generalized EEG emotion recognition. *arXiv preprint arXiv:2509.01135*, 2025.

Taida Li, Yujun Yan, Fei Dou, Wenzhan Song, and Xiang Zhang. Cross-subject generalization for eeg decoding: A survey of deep learning methods. *Progress in Biomedical Engineering*, 2026.

Guang Lin and Jianhai Zhang. Multi-subdomain adversarial network for cross-subject eeg-based emotion recognition. *arXiv preprint arXiv:2308.14059*, 2023.

Mario Lobo-Alonso, Iván Martín-Fernández, I Oropesa, Roberto Barra-Chicote, George Kontaxakis, and Rubén San-Segundo. A dataset with bilingual tv commands for silent speech interfaces using electroencephalographic signals. *Scientific Data*, pp. 1–1, 2026.

Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Rätsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. Challenging common assumptions in the unsupervised learning of disentangled representations. In *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 4114–4124, 2019.

Xinyu Ma, Yi Jiang, and Ning Jiang. 3m-cpseed, an eeg-based dataset for chinese pinyin production in overt, mouthed, and imagined speech. *Scientific Data*, 13(1):34, 2025.

Sean L. Metzger, Kaylo T. Littlejohn, Alexander B. Silva, David A. Moses, Margaret P. Seaton, Ran Wang, Maximilian E. Dougherty, Jessie R. Liu, Peter Wu, Michael A. Berger, Inga Zhuravleva, Adelyn Tu-Chan, Karunesh Ganguly, Gopala K. Anumanchipalli, and Edward F. Chang. A high-performance neuroprosthesis for speech decoding and avatar control. *Nature*, 620(7976):1037–1046, 2023. doi: 10.1038/s41586-023-06443-4.

Nicolás Nieto, Victoria Peterson, Hugo Leonardo Rufiner, Juan Esteban Kamienkowski, and Ruben Spies. Thinking out loud, an open-access eeg-based bci dataset for inner speech recognition. *Scientific data*, 9(1):52, 2022.

Yifan Niu, Ziyu Li, Li Yao, and Xia Wu. Bdr-gcl: Toward imagined speech decoding in naturalistic bci systems via brain dynamics representation enhanced graph contrastive learning. *Expert Systems with Applications*, 296:129058, 2026.

Ozan Özdenizci, Ye Wang, Toshiaki Koike-Akino, and Deniz Erdoğmuş. Learning invariant representations from EEG via adversarial inference. *IEEE Access*, 8:27074–27085, 2020.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017.

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: 10.1109/TNSRE.2022.3230250.

Weitong Sun, Yuping Su, Yumei Zhang, and Xiaojun Wu. Eeg emotion recognition through a domain-adversarial multi-feature fusion network. *Expert Systems with Applications*, pp. 129694, 2025.

Frederik Møllskov Trier, Xiaopeng Mao, and Sadasivan Puthusserypady. Decoding imagined speech: A strictly subject-independent approach using eeg, 2026. URL https://arxiv.org/abs/2609.29820.

Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Shijian Li, and Gang Pan. Eegmamba: An eeg foundation model with mamba. *Neural Networks*, pp. 107816, 2025.

Yiming Wang, Bin Zhang, and Yujiao Tang. Dmmr: Cross-subject domain generalization for eeg-based emotion recognition via denoising mixed mutual reconstruction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 628–636, 2024.

Weining Weng, Yang Gu, Yuan Ma, Yuchen Liu, Yingwei Zhang, and Yiqiang Chen. State mamba: Spatiotemporal eeg state-space model with dynamic brain alignment for cross-subject representation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 17850–17858, 2026.

Francis R. Willett, Erin M. Kunz, Chaofei Fan, Donald T. Avansino, Guy H. Wilson, Eun Young Choi, Foram Kamdar, Matthew F. Glasser, Leigh R. Hochberg, Shaul Druckmann, Krishna V. Shenoy, and Jaimie M. Henderson. A high-performance speech neuroprosthesis. *Nature*, 620 (7976):1031–1036, 2023. doi: 10.1038/s41586-023-06377-x.

Paolo Zanini, Marco Congedo, Christian Jutten, Salem Said, and Yannick Berthoumieu. Transfer learning: A riemannian geometry framework with applications to brain–computer interfaces. *IEEE Transactions on Biomedical Engineering*, 65(5):1107–1116, 2017.

Li Zhang, Jiyao Liu, and Lei Xie. Subject disentanglement neural network for speech envelope reconstruction from eeg. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 4002–4005. IEEE, 2024a.

Zihan Zhang, Xiao Ding, Yu Bao, Yi Zhao, Xia Liang, Bing Qin, and Ting Liu. Chisco: An EEG-based BCI dataset for decoding of imagined speech. *Scientific Data*, 11(1):1265, 2024b. doi: 10.1038/s41597-024-04114-1.

Han Zhao, Remi Tachet Des Combes, Kun Zhang, and Geoffrey J. Gordon. On learning invariant representations for domain adaptation. In *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 7523–7532, 2019.

Appendices

A Extended Related Work

A.1 Speech and Language Decoding from Brain Recordings

Implanted speech neuroprostheses have recently decoded attempted speech from intracortical microelectrode recordings (Willett et al., 2023) and restored communication through synthesized speech and a digital avatar (Metzger et al., 2023). While such systems require neurosurgery, non-invasive alternatives based on magneto- and electroencephalography (M/EEG) have attracted growing interest. Défossez et al. (2023) decoded perceived speech from M/EEG recordings of 175 participants, and d'Ascoli et al. (2025) decoded individual words from 723 participants who read or listened to five million words in three languages. Lévy et al. (2026) further decoded sentences that participants typed on a keyboard. Across these studies, EEG decoding is consistently less accurate than MEG decoding (Défossez et al., 2023; Lévy et al., 2026), which underlines the difficulty of the EEG setting targeted in this work.

Covert and imagined speech decoding has been supported by public EEG datasets, including Thinking Out Loud (Nieto et al., 2022), Chisco (Zhang et al., 2024b), and TESSCCo (Lobo-Alonso et al., 2026). TESSCCo contains five TV commands pronounced overtly and covertly in English and Spanish by 21 native and 3 non-native Spanish speakers, recorded with a 32-channel EEG system. Recent decoders include BDR-GCL (Niu et al., 2026), which applies graph contrastive learning to imagined speech, and CCSUMSP (Huang et al., 2025), which performs cross-subject Chinese speech decoding on EEG recorded during reading. Cross-subject evaluation has also received explicit attention. Ko et al. (2026) combine cyclic inter-subject training with few-shot calibration for imagined speech detection, and Trier et al. (2026) compare preprocessing pipelines under strictly subject-independent evaluation.

A.2 General EEG Decoders and Foundation Models

Compact convolutional networks such as EEGNet (Lawhern et al., 2018) and convolution–Transformer hybrids such as EEG Conformer (Song et al., 2023) remain widely used decoding backbones. EEG foundation models pre-train on large heterogeneous corpora and then transfer to downstream tasks. LaBraM (Jiang et al., 2024) segments signals into channel patches to handle datasets with different electrode configurations, and learns a vector-quantized neural tokenizer from about 2,500 hours of EEG drawn from around 20 datasets. EEGMamba (Wang et al., 2025) builds an EEG foundation model on the Mamba state-space architecture. State Mamba (Weng et al., 2026) models spatial and temporal brain-state transitions and aligns representations and transition patterns across subjects for cross-subject emotion and sleep recognition. We note that the term "state" in State Mamba refers to brain states and their state-space transitions, whereas in BrainState it denotes the recording condition of a trial.

A.3 Graph-Based EEG Modeling

Graph-based EEG decoders typically treat electrodes as nodes. In BDR-GCL, each node corresponds to one electrode, the adjacency matrix is initialized from electrode coordinates, and edge dynamics are derived from phase synchronization between electrodes (Niu et al., 2026). CCSUMSP instead maps individual electrodes to standardized regional representations to accommodate different montages, and uses eye-tracking data and text in its multimodal pre-training (Huang et al., 2025). In both cases, the graph is defined over the spatial layout of the sensors. BrainState uses a graph of a different kind: its nodes are classes and class-by-condition groups of training trials, so the graph organizes training experience rather than sensor space.

A.4 Cross-Subject and Cross-Session Transfer in EEG

A large body of work learns subject-invariant representations. Domain adversarial training inserts a gradient reversal layer so that features cannot discriminate between domains (Ganin et al., 2016). In EEG emotion recognition, DMMR combines a subject discriminator with gradient reversal and

Table 4: Fixed five-class label mapping for the two TESSCCo languages. Each experiment uses one language at a time.

Label	Class	English command	Spanish command
0	Mute	Mute FireTV	Silencia FireTV
1	Volume up	Turn up volume	Sube el volumen
2	Power off	Turn off FireTV	Apaga FireTV
3	Power on	Turn on television	Enciende la televisión
4	Volume down	Turn down volume	Baja el volumen

mixed mutual reconstruction (Wang et al., 2024), and related methods adopt multi-subdomain (Lin & Zhang, 2023) or multi-feature fusion (Sun et al., 2025) adversarial networks. Jeon et al. (2021) decompose motor imagery features into class-relevant and class-irrelevant parts by estimating mutual information, explicitly avoiding adversarial learning. For speech envelope reconstruction, SDN-Net uses a mutual-information estimator to remove subject identity from the reconstructed envelope (Zhang et al., 2024a), and Bao & Lu (2026) likewise propose mutual-information-guided subject disentanglement. A second line aligns data distributions across subjects, using Euclidean-space alignment (He & Wu, 2019), a Riemannian-geometry framework (Zanini et al., 2017), or learned alignment of brain-state dynamics (Weng et al., 2026). A third line adapts to new subjects with limited data, through model-agnostic meta-learning (Huang et al., 2025) or few-shot calibration (Ko et al., 2026). Most of these methods either remove subject- and session-specific information or apply the same transfer to every test trial. BrainState instead keeps condition information and uses it to decide which training experience is relevant to each test trial.

A.5 Retrieval, Memory, and Condition-Aware Decoding

Non-parametric memory has improved generalization in other domains: kNN-LM interpolates a language model with a k-nearest-neighbor model whose neighbors are retrieved by distance in the model's embedding space (Khandelwal et al., 2019), and prototypical networks classify by distances to class prototypes (Snell et al., 2017). In M/EEG decoding, Défossez et al. (2023) condition the model on the recording subject through a subject-specific layer, a matrix learned for each training subject and selected by the subject index. BrainState combines these ideas for cross-subject EEG decoding. Its class nodes act as prototypes, its class-by-condition nodes form a retrievable memory, and the condition used for retrieval is inferred from the signal itself, so that no subject index is needed for unseen test subjects.

B Dataset Details and Preprocessing

B.1 TESSCCo dataset

The TV-control EEG-based Silent Speech Command Corpus (TESSCCo) contains EEG from 24 healthy participants: 21 native Spanish speakers and three non-native Spanish speakers Lobo-Alonso et al. (2026). Participants performed five television-control commands in Spanish and English, each under overt speech (OS) and covert speech (CS) conditions. In this corpus, CS means silent articulation *with mouth movement*, rather than speech imagined without articulation. EEG was recorded with a 32-channel water-based Bitbrain Versatile E32 headset at 256 Hz. The released clean corpus contains 7,936 five-second epochs from 45 available sessions across both populations. Our main analysis uses all 21 native speakers, including Subject_21, and their 40 available sessions; Native Subjects 11 and 21 each have one available session, while the other 19 have two. We use CS only and treat the two languages as separate five-class tasks. The released native-speaker clean data contain 1,792 Spanish CS epochs and 1,787 English CS epochs; the main results table reports Spanish CS. Table 4 gives the fixed semantic label mapping used by our loader. We do not combine English and Spanish trials or include OS or non-native trials in this analysis. Figure 7 summarizes the recording protocol and electrode montage, and Table 5 lists the participant and epoch statistics.

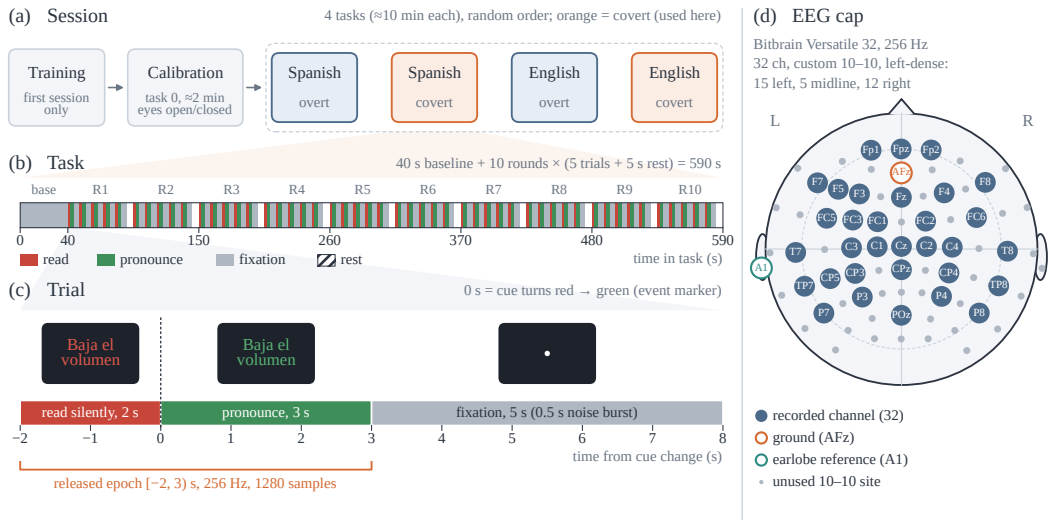


Figure 7: TESSCCo recording protocol and EEG montage, redrawn from Lobo-Alonso et al. (2026). (a) A session starts with one unrecorded training trial (first session only) and a 2-min calibration with eyes open and closed, followed by four tasks (Spanish or English, overt or covert) in random order; we use the two covert-speech (CS) tasks. (b) A task consists of a 40-s baseline and ten rounds; each round presents the five commands once in random order and ends with a 5-s rest. (c) In each trial, the command is shown in red for 2 s for silent reading and then in green for 3 s for articulation, followed by a 5-s fixation. The released epochs span $[-2, 3)$ s around the red-to-green change. (d) The 32 recorded channels of the custom 10–10 montage, with the ground at AFz and an earlobe reference; unused 10–10 positions are shown in grey.

Table 5: TESSCCo statistics. Each usable session has 4×50 trials; 8,908 remained after the authors' event and artifact checks and 7,936 after visual inspection. Per-participant counts are mean $\pm$ SD (native Spanish CS: 39–99). CS / OS: covert / overt speech.

	Native	Non-native	Total
Participants (male / female)	21 (13 / 8)	3 (3 / 0)	24 (16 / 8)
Age (years)	23.2 $\pm$ 2.1	23.3 $\pm$ 0.6	23.2 $\pm$ 2.0
Sessions (usable / recorded)	40 / 42	5 / 6	45 / 48
Trials in usable sessions $\rightarrow$ clean	8,000 $\rightarrow$ 7,007	1,000 $\rightarrow$ 929	9,000 $\rightarrow$ 7,936
Clean duration (h)	9.73	1.29	11.02
CS epochs (Spanish / English)	1,792 / 1,787	228 / 230	2,020 / 2,017
OS epochs (Spanish / English)	1,721 / 1,707	238 / 233	1,959 / 1,940
Spanish CS epochs per participant	85.3 $\pm$ 17.5	76.0 $\pm$ 28.2	84.2 $\pm$ 18.5
Spanish CS epochs per class (range)	355–367	44–47	399–414

B.2 Acquisition timing and data preprocessing.

Each task presented a command in red for 2 s for silent reading, then in green for 3 s to cue articulation. The released epochs span $[-2, 3)$ s relative to the red-to-green transition, so zero marks the instruction cue rather than a measured onset of articulation. We adopt the released *Clean* epochs. The preprocessing first synchronized EEG and event timestamps, trimmed recording boundaries, and retained the last of immediately repeated trials. The EEG values were converted from the device's mV output to EEGLAB's μV scale. Then apply EEGLAB finite-impulse-response high-pass and low-pass filters at 1 and 100 Hz, respectively, and a 49–51 Hz notch filter. Artifact Subspace Reconstruction repaired noisy segments with a burst threshold of 25 standard deviations, followed by common-average re-referencing and epoching around the cue. An adaptive narrow-band filter removed device-related peaks around 70–90 Hz. Extended `runica` was fitted to the epoched data; ICLabel components with an eye-artifact probability above 0.7 were removed. Finally, the authors

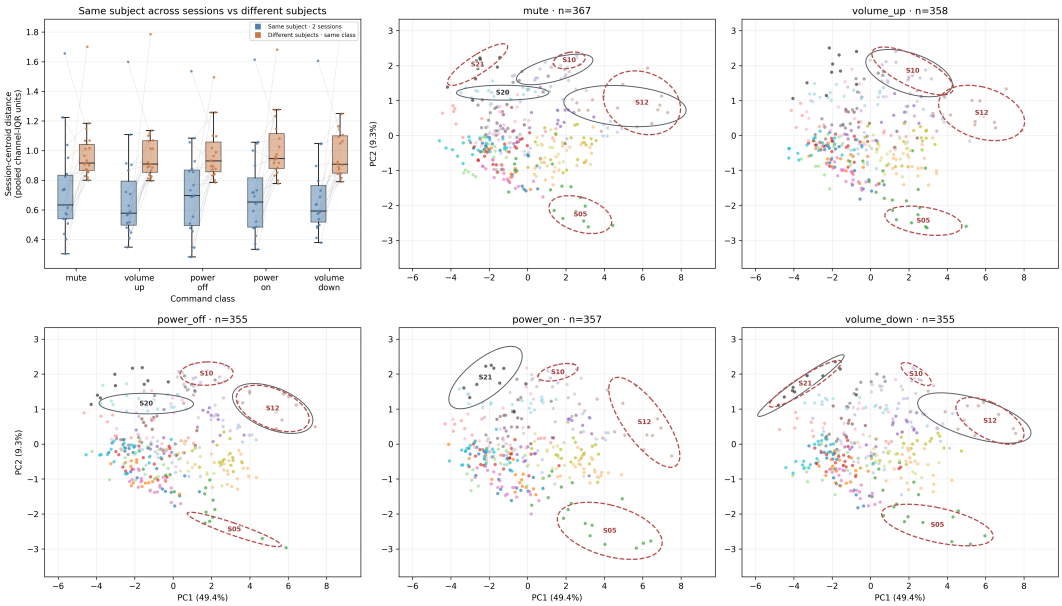


Figure 8: Cross-subject shift in the log-RMS feature of the TESSCCo recordings (21 Native subjects, covert speech). Upper left: for each class, distance between the two session centroids of the same subject (blue) and median distance to the other subjects (orange); dots are the 19 two-session subjects and grey lines connect the same subject. Other panels: all 1792 trials projected onto one shared PCA and split by class (n trials per class in the titles); colours denote subjects. Gray ellipses mark subject clusters (robust 80% coverage ellipse with at least 65% of the enclosed trials from that subject); red dashed ellipses enclose a subject's trials beyond the class-wise 90th percentile of robust Mahalanobis distance (subjects with at least five such trials).

rejected annotated bad epochs and visually inspected the remainder, retaining 7,936 clean epochs across the full corpus. These steps are described in the original paper and its released preprocessing script Lobo-Alonso et al. (2026).

Inputs and subject-level split. Our local cache preserves the released clean signal values and original command labels.The only change to the signal arrays is the axis order, now trial $\times$ channel $\times$ time. We select all 32 EEG channels and crop each epoch to the green-cue execution window $[0, 3)$ s, yielding 32×768 samples at the native 256 Hz for BrainState. In each leave-one-subject-out fold, one native speaker is held out for testing, the preceding native speaker in cyclic subject order is used for validation, and the remaining 19 speakers provide training data. The corpus's CS instructions permit mouth movement, so articulatory activity and residual muscle or movement artifacts may contribute to decoded signals.

B.3 Subject structure in the raw amplitude feature.

Figure 8 visualises the cross-subject shift in the recordings before any model is applied. Figure 4 quantifies the same shift and its reduction after encoding. The upper-left panel compares, for each command class, the distance between the two session centroids of a subject with the median distance from that subject to the other 20 subjects. Each centroid is the median log-RMS vector of one subject, session and class, with channels scaled by their interquartile range, pooled over all trials.

Across the 19 subjects recorded in two sessions, the median within-subject distance ranges from 0.58 to 0.70, and the median between-subject distance ranges from 0.91 to 0.95 over the five classes. The between-subject distance is larger in 86 of the 95 subject-class pairs. The remaining panels project all 1792 trials onto a single PCA of the log-RMS feature (PC1 49.4%, PC2 9.3%), shown separately for each class with a fixed colour per subject. Several subjects occupy similar regions in all five classes, for example, S12 at large PC1 and S05 at low PC2, whereas trials of different commands overlap. Grey ellipses mark concentrated subject clusters. For subjects with at least 10

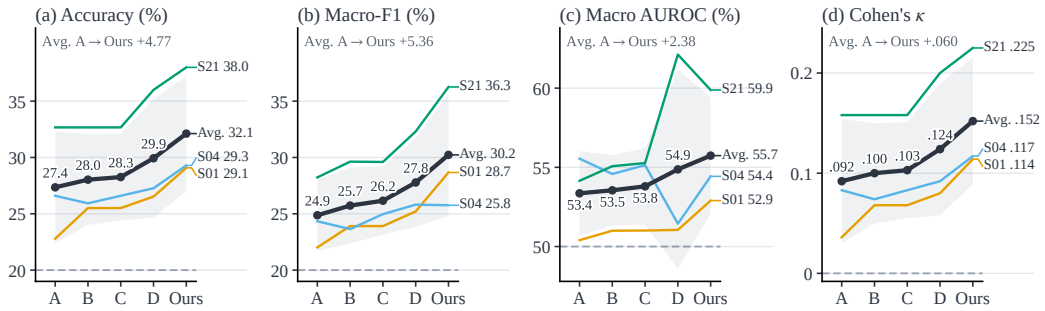


Figure 9: Ablation from variant A to the full model (Ours) on Accuracy, Macro-F1, Macro AUROC and Cohen's κ. Thin lines show each held-out subject and the thick line their average (Avg.), all as means over seeds 0, 42 and 3407; labels at the right end give the value of Ours. The shaded band shows $\pm$SD of Avg. across subjects, dashed lines mark chance level, and the corner text gives the average gain from A to Ours. Seed-level standard deviations are reported in Table 4.

trials in the class, we fit a robust (MinCovDet) ellipse that covers 80% of the subject's trials and keep it if at least 65% of the trials within it belong to that subject, showing at most three per class. Red dashed ellipses enclose outlying trials: within each class, trials whose robust Mahalanobis distance exceeds the 90th percentile of that class, drawn for subjects with at least five such trials (S05, S10 and S12 in every class, and S21 for mute and volume down).

Taken together, the distances and projections show that, within TESSCCo, each subject constitutes a distinct domain: the trials of a subject remain closer to each other across sessions than to those of other subjects and occupy consistent regions across commands, so that different subjects are separated by a clearly visible domain gap that exceeds the variation between sessions of the same subject.

C Detailed results

This section reports per-subject results that complement the averaged numbers in the main text. Table 6 lists the cross-subject LOSO performance of BrainState on all 21 subjects of the Native Spanish dataset. Table 7 breaks the ablation study down by subject. For every subject we report the mean $\pm$ standard deviation over seeds 0, 42 and 3407. The Avg. row gives the mean $\pm$ standard deviation across subjects.

Per-subject LOSO performance. BrainState reaches an average accuracy of $29.94 \pm 3.37\%$, a macro-F1 of $28.21 \pm 3.79\%$, a macro AUROC of $54.52 \pm 2.86\%$ and a Cohen's κ of 0.124 ± 0.041 (Table 6). The gains over chance hold for individual subjects and are not driven by a few of them. Accuracy is above the 20% chance level for all 21 held-out subjects. It ranges from 22.68% (S09) to 38.00% (S21), and ten subjects reach at least 30%. Cohen's κ is positive for every subject, from 0.039 (S09) to 0.225 (S21), and 17 of the 21 subjects reach $\kappa \geq 0.1$. Macro AUROC is above 50% for 20 subjects. The one exception is S18 (47.66%).

Across seeds, the standard deviation of accuracy is below 3% for every subject, and for most subjects it is below 2%. The spread across subjects (3.37%) is larger than the typical spread across seeds. Most of the variability therefore comes from differences between subjects, not from the random initialization. This is expected for cross-subject EEG decoding.

The weakest subjects, S09 and S18, also show the largest gap between accuracy and macro-F1 (22.68% vs. 19.54% and 25.76% vs. 21.57%). S09 also has a very unstable precision ($26.13 \pm 12.50\%$). These patterns suggest that, for these held-out subjects, the model's predictions concentrate on a subset of the classes. The best-decoded subjects, S21, S11 and S13, have consistently high values on all six metrics.

Table 6: Per-subject results of BrainState on the TESSCCo dataset under the cross-subject LOSO setting. Results are reported as mean ± standard deviation across seeds 0, 42 and 3407.

Subject	Accuracy (%) ↑	Precision (%) ↑	Recall (%) ↑	F1 (%) ↑	Macro AUROC (%) ↑	Cohen's κ ↑
S01	29.08 ± 0.72	28.52 ± 1.05	29.18 ± 0.78	28.67 ± 0.98	52.90 ± 1.60	0.114 ± 0.009
S02	25.26 ± 2.98	26.25 ± 0.85	25.51 ± 3.06	22.51 ± 0.06	53.14 ± 2.63	0.067 ± 0.039
S03	30.77 ± 1.81	28.30 ± 1.08	30.70 ± 1.51	28.84 ± 0.70	56.52 ± 1.88	0.133 ± 0.020
S04	29.29 ± 1.43	25.56 ± 2.11	29.37 ± 0.97	25.76 ± 1.57	54.43 ± 1.17	0.117 ± 0.015
S05	28.80 ± 0.77	29.29 ± 0.41	28.86 ± 1.04	28.55 ± 0.18	56.82 ± 0.15	0.111 ± 0.012
S06	30.26 ± 0.00	30.62 ± 3.79	30.04 ± 0.06	28.14 ± 0.44	54.30 ± 0.14	0.127 ± 0.000
S07	30.41 ± 2.19	30.09 ± 2.44	30.56 ± 1.85	29.97 ± 2.29	53.57 ± 3.18	0.131 ± 0.026
S08	28.70 ± 1.31	27.84 ± 0.58	29.02 ± 1.99	27.50 ± 1.06	52.38 ± 1.60	0.109 ± 0.021
S09	22.68 ± 0.00	26.13 ± 12.50	23.13 ± 0.05	19.54 ± 2.62	50.04 ± 2.17	0.039 ± 0.001
S10	31.77 ± 0.74	30.87 ± 0.86	31.92 ± 0.63	30.01 ± 0.69	56.37 ± 4.16	0.147 ± 0.009
S11	34.62 ± 1.81	35.28 ± 3.09	34.11 ± 1.77	31.28 ± 2.69	58.79 ± 0.48	0.181 ± 0.023
S12	31.82 ± 0.92	31.55 ± 1.75	31.15 ± 1.19	29.60 ± 1.38	54.50 ± 1.92	0.142 ± 0.015
S13	34.90 ± 0.74	36.55 ± 0.23	34.76 ± 0.78	34.06 ± 1.53	58.03 ± 1.16	0.185 ± 0.010
S14	28.57 ± 1.55	28.62 ± 3.46	28.53 ± 2.02	27.83 ± 2.19	54.34 ± 0.07	0.108 ± 0.020
S15	28.12 ± 0.00	28.51 ± 1.81	28.26 ± 0.15	26.88 ± 0.30	51.14 ± 0.64	0.102 ± 0.001
S16	30.59 ± 1.66	32.24 ± 2.42	30.68 ± 1.84	30.70 ± 1.77	54.71 ± 2.41	0.132 ± 0.021
S17	31.46 ± 0.00	30.95 ± 3.85	31.50 ± 0.03	29.46 ± 2.90	56.25 ± 3.34	0.142 ± 0.003
S18	25.76 ± 2.14	24.38 ± 4.08	25.63 ± 2.01	21.57 ± 4.34	47.66 ± 1.39	0.071 ± 0.026
S19	29.90 ± 0.00	27.51 ± 2.39	30.13 ± 0.04	28.39 ± 1.35	54.58 ± 0.09	0.124 ± 0.000
S20	28.02 ± 0.78	33.29 ± 2.85	27.56 ± 0.11	26.86 ± 0.81	54.65 ± 1.37	0.097 ± 0.005
S21	38.00 ± 2.83	40.29 ± 2.21	38.00 ± 2.83	36.26 ± 3.83	59.88 ± 1.17	0.225 ± 0.035
Avg.	29.94 ± 3.37	30.13 ± 3.86	29.93 ± 3.26	28.21 ± 3.79	54.52 ± 2.86	0.124 ± 0.041

Table 7: Per-subject Accuracy, Macro-F1, AUROC and Cohen's kappa of the ablation variants. Per-subject results are reported as mean ± std across seeds 0, 42 and 3407; Avg. reports mean ± std across subjects.

Subject	A	B	C	D	Ours
Accuracy (%)					
S01	22.79 ± 2.12	25.51 ± 3.68	25.51 ± 3.68	26.53 ± 2.89	29.08 ± 0.72
S04	26.60 ± 3.82	25.93 ± 1.54	26.60 ± 2.54	27.27 ± 1.43	29.29 ± 1.43
S21	32.67 ± 1.15	32.67 ± 4.16	32.67 ± 4.16	36.00 ± 0.00	38.00 ± 2.83
Avg.	27.35 ± 4.98	28.03 ± 4.02	28.26 ± 3.86	29.93 ± 5.27	32.12 ± 5.09
Macro-F1 (%)					
S01	22.02 ± 1.75	23.91 ± 5.76	23.91 ± 5.76	25.21 ± 3.91	28.67 ± 0.98
S04	24.34 ± 4.03	23.64 ± 1.02	24.97 ± 1.55	25.81 ± 1.22	25.76 ± 1.57
S21	28.23 ± 2.51	29.63 ± 5.87	29.60 ± 5.85	32.31 ± 1.76	36.26 ± 3.83
Avg.	24.87 ± 3.14	25.73 ± 3.38	26.16 ± 3.02	27.78 ± 3.94	30.23 ± 5.42
Macro AUROC (%)					
S01	50.40 ± 1.16	51.00 ± 2.23	51.01 ± 2.23	51.05 ± 0.98	52.90 ± 1.60
S04	55.55 ± 1.26	54.58 ± 1.25	55.14 ± 0.68	51.44 ± 2.62	54.43 ± 1.17
S21	54.15 ± 1.99	55.07 ± 5.39	55.27 ± 5.37	62.12 ± 4.35	59.88 ± 1.17
Avg.	53.36 ± 2.67	53.55 ± 2.22	53.80 ± 2.42	54.87 ± 6.29	55.74 ± 3.67
Cohen's κ					
S01	0.036 ± 0.028	0.068 ± 0.047	0.068 ± 0.047	0.080 ± 0.038	0.114 ± 0.009
S04	0.083 ± 0.049	0.074 ± 0.020	0.083 ± 0.034	0.092 ± 0.019	0.117 ± 0.015
S21	0.158 ± 0.014	0.158 ± 0.052	0.158 ± 0.052	0.200 ± 0.000	0.225 ± 0.035
Avg.	0.092 ± 0.062	0.100 ± 0.050	0.103 ± 0.048	0.124 ± 0.066	0.152 ± 0.063

Per-subject ablation results. Table 7 compares the full model with variants A–D on subjects S01, S04 and S21. Note that the Avg. rows in this table are computed over these three subjects only, so they differ from the 21-subject averages in Table 6.

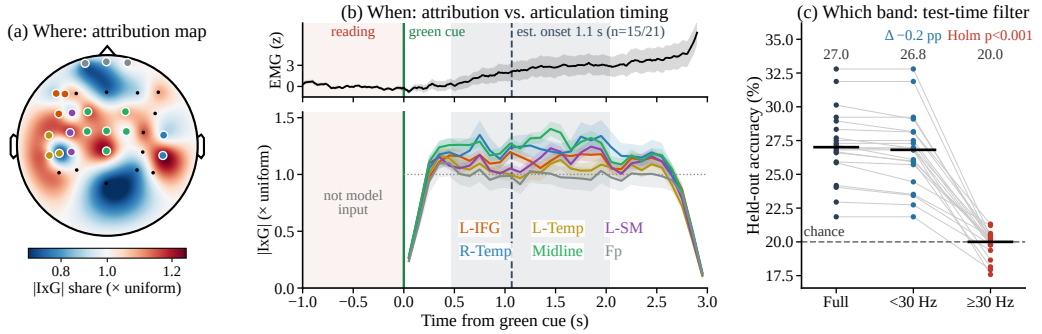


Figure 10: Evidence used by BrainState on held-out subjects (21 LOSO folds × 5 seeds). (a) Mean |IxG| share relative to uniform on a template montage; coloured electrodes mark the regions in (b). (b) EMG proxy (30–100 Hz envelope at temporal-edge electrodes; top) and regional attribution (bottom, mean ± s.e.m.). The green line is the articulation cue; the dashed line and shading show the median and IQR of the estimated onset. (c) Accuracy with the full input, with ≥30 Hz removed, and with only ≥30 Hz kept.

The full model has the best average on all four metrics: accuracy 32.12%, macro-F1 30.23%, macro AUROC 55.74% and $\kappa = 0.152$. The strongest variant, D, reaches 29.93%, 27.78%, 54.87% and 0.124, respectively.

On accuracy and Cohen's κ, the full model is the best on each of the three subjects individually. Relative to the best ablated variant, it improves accuracy by 2.55, 2.02 and 2.00 percentage points on S01, S04 and S21.

On the threshold-free metrics the picture is less uniform:

- Variant D obtains a higher AUROC on S21 (62.12% vs. 59.88%).
- Variants A and C are slightly higher on S04 (55.55% and 55.14% vs. 54.43%).
- Variant D is marginally higher in macro-F1 on S04 (25.81% vs. 25.76%).

These components therefore mainly improve the calibration of the decision boundary, and so the thresholded predictions, more than the ranking of class scores.

On average, the variants follow a consistent order, A $<$ B $\approx$ C $<$ D $<$ Ours, for accuracy, macro-F1 and κ. B and C give identical accuracy and macro-F1 on S01, and identical accuracy on S21. The full model also has the smallest or joint-smallest seed-level standard deviation of accuracy and κ on S01 and S04. This indicates that the complete design improves both performance and robustness to initialization.

D Does cross-subject transfer rely on articulation artifacts?

Covert speech in TESSCCo involves silent mouth movement, so accuracy on unseen speakers could come from articulatory muscle activity rather than from speech-related brain activity. To test this, we asked where, when and in which frequency band BrainState finds its evidence. We reused the 21 LOSO folds × 5 seeds. On every held-out trial (1,792 native Spanish CS trials), we computed the absolute Input×Gradient (|IxG|) of the predicted class. We estimated articulation onset from the 30–100 Hz envelope at temporal-edge electrodes, used as an EMG proxy. We also re-evaluated the frozen models after removing or keeping only the ≥30 Hz band at test time. Attribution was spread across the scalp, with channels at 0.66–1.24 times the uniform share and no left lateralization (Fig. 10a). It stayed flat over the articulation window and did not follow the estimated onset (median 1.07 s; Fig. 10b). Removing ≥30 Hz changed accuracy by only −0.2 percentage points, whereas the ≥30 Hz band alone gave chance-level accuracy (20.0%; Fig. 10c). The transferred information

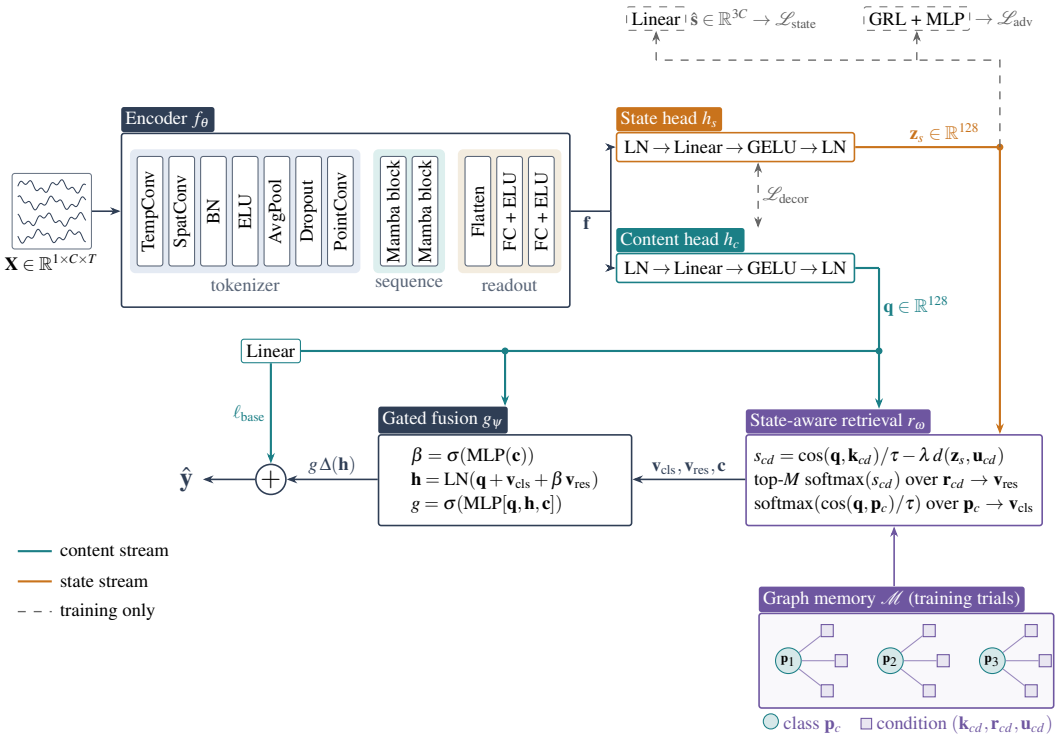


Figure 11: BrainState architecture. The encoder f_θ maps an EEG trial $\mathbf{X} \in \mathbb{R}^{1\times C\times T}$ through a convolutional tokenizer, two Mamba blocks, and a fully connected readout to a trial feature $\mathbf{f} \in \mathbb{R}^{256}$. The content head h_c and state head h_s produce the content query $\mathbf{q}$ (teal) and the learned state $\mathbf{z}_s$ (orange). The state-aware retrieval r_ω ranks condition nodes of the training-set graph memory $\mathcal{M}$ by content similarity and state distance and returns the aggregated class and residual values $\mathbf{v}_{\text{cls}}$, $\mathbf{v}_{\text{res}}$ with their confidence features $\mathbf{c}$. The gated fusion g_ψ adds the correction $g\,\Delta(\mathbf{h})$ to the direct logits ℓ_{base}. Dashed components (state reconstruction, gradient-reversal class adversary, and the content–state decorrelation $\mathcal{L}_{\text{decor}}$) are used only during training.

therefore lies below 30 Hz and is not tied to high-frequency EMG or to a single articulatory event, although low-frequency movement artifacts cannot be excluded.

E A PROBABILISTIC ANALYSIS OF STATE-CONDITIONED GRAPH RETRIEVAL

E.1 NOTATION

The symbols below follow the main methodology where applicable. A node index $j = (c, d)$ always denotes a class–condition child node.

Symbol	Meaning
$c, d, j, \mathcal{J}$	Task class; subject–session condition; their joint node $j = (c, d)$; set of available training nodes.
i, e, m	Trial index; dummy condition index; competing or dummy node index.
$q_i, z_{s,i}, \theta$	Content and trial-level state embeddings of training trial i; learned encoder parameters. Unsubscripted q, z_s denote a query.
$p_c, k_{cd}, r_{cd}, u_{cd}$	Class prototype; child-node content key; residual $k_{cd} - p_c$; child-node state prototype. Subscript j abbreviates (c, d).
$h_{\text{cls}}, h_{\text{res}}$	Aggregated class-prototype evidence and condition-residual evidence in the main model.
$\mu_c, \rho_{cd}, \epsilon_i$	Shared class component, condition displacement, and zero-mean trial noise in Eq. 6.
$N_{cd}, N_c, \pi_{d\|c}$	Number of training trials in node (c, d); class total; empirical condition proportion N_{cd}/N_c.
$\hat{q}, \hat{k}_j$	Unit-normalized content query and child-node key.
κ, σ^2, I	Directional concentration, isotropic state variance, and identity matrix in Eq. 8.
$\tau, \lambda, a_j, \dim(z_s)$	Content temperature, state-distance weight, combined retrieval score, and state-embedding dimension.
$P(j \mid \hat{q}, z_s)$	Posterior probability assigned to child node j under the stated likelihood model.
$b_j, D_j, \gamma, \varepsilon$	Content score, mean-squared state distance, state-distance margin, and upper bound on a competitor's content advantage.
M, N, S_M, G, ϵ_M	Top-node budget; total node count $N = \|\mathcal{J}\|$; selected set; score gap at rank M; omitted posterior mass.
V, R	Learned residual-value projection and an upper bound on $\|V(r_j)\|_2$.
$\mathbb{E}, \perp, \text{vMF}, \mathcal{N}$	Expectation, conditional independence, von Mises–Fisher distribution, and Gaussian distribution.

We analyze the class–condition memory and retrieval score in Eqs. 2 and 3. We condition on the learned encoder and treat its content and state embeddings as random variables across EEG trials.

E.2 WHAT THE HIERARCHICAL RESIDUAL REPRESENTS

Suppose that the content embedding of a trial from node (c, d) admits the decomposition

$$q_i = \mu_c + \rho_{cd} + \epsilon_i, \qquad \mathbb{E}[\epsilon_i \mid c, d] = 0, \tag{6}$$

where μ_c is shared class evidence and ρ_{cd} is its condition-dependent displacement. Write N_{cd} for the number of training trials in the node, $N_c = \sum_d N_{cd}$, and $\pi_{d|c} = N_{cd}/N_c$. The memory in Eq. 2 then has the following interpretation.

Proposition 1 (class-centered condition residual). Conditioned on the node counts and the encoder, the stored residual $r_{cd} = k_{cd} - p_c$ satisfies

$$\mathbb{E}[r_{cd} \mid \{N_{ce}\}_e] = \rho_{cd} - \sum_e \pi_{e|c}\rho_{ce}. \tag{7}$$

Thus, each child node stores its departure from the training-set average expression of its class.

Proof. By Eq. 6, $\mathbb{E}[k_{cd}] = \mu_c + \rho_{cd}$. Since $p_c = \sum_e \pi_{e|c} k_{ce}$, we have $\mathbb{E}[p_c] = \mu_c + \sum_e \pi_{e|c}\rho_{ce}$. Subtracting gives Eq. 7. □

The decomposition explains the two-level readout in Eq. 4: class prototypes provide shared task evidence, while retrieved residuals supply condition-specific corrections. During training, BrainState recomputes both levels with the query trial removed from their means.

E.3 Retrieval as posterior inference

Let $\hat{q} = q/\|q\|_2$ and $\hat{k}_j = k_j/\|k_j\|_2$. Assume nonzero content vectors and the following conditional model for a query generated under node j:

$$\hat{q} \mid j \sim \mathrm{vMF}(\hat{k}_j, \kappa), \qquad z_s \mid j \sim \mathcal{N}(u_j, \sigma^2 I), \qquad \hat{q} \perp z_s \mid j. \tag{8}$$

The von Mises–Fisher distribution assigns higher density to content directions aligned with $\hat{k}_j$; the Gaussian assigns higher density to states near u_j. Both concentrations are shared across nodes. We use a uniform prior on the available nodes, matching the unweighted node ranking in BrainState.

Proposition 2 (posterior form of the retrieval score). Set $\kappa = 1/\tau$ and $\sigma^2 = \dim(z_s)/(2\lambda)$ for $\lambda > 0$. Under Eq. 8, the posterior over nodes is

$$P(j \mid \hat{q}, z_s) = \frac{\exp(a_j)}{\sum_{m \in \mathcal{J}} \exp(a_m)}, \qquad a_j = \frac{\hat{q}^\top \hat{k}_j}{\tau} - \lambda \frac{\|z_s - u_j\|_2^2}{\dim(z_s)}. \tag{9}$$

The score a_j is exactly the state-conditioned node score in Eq. 3. At $\lambda = 0$, retrieval reduces to the content likelihood alone.

Proof. The von Mises–Fisher log density is $\kappa \hat{q}^\top \hat{k}_j$ plus a normalizing constant shared by all nodes. The Gaussian log density is $-\|z_s - u_j\|_2^2/(2\sigma^2)$ plus another shared constant. Conditional independence adds the two log densities. Substitution of κ and σ^2, followed by Bayes' rule with the uniform prior, yields Eq. 9. □

This interpretation connects the model's representation constraints to retrieval: the state-reconstruction objective encourages z_s to encode observable trial conditions, while the class adversary and content–state decorrelation encourage distinct roles for z_s and q.

E.4 When state agreement changes the selected evidence

Define the content score $b_j = \cos(q, k_j)/\tau$ and the mean-squared state distance $D_j = \|z_s - u_j\|_2^2/\dim(z_s)$, so $a_j = b_j - \lambda D_j$. Consider a state-matched node j and a competing node m.

Proposition 3 (state-matching margin). If $D_m - D_j \geq \gamma > 0$ and $b_m - b_j \leq \varepsilon$, then

$$\log \frac{P(j \mid \hat{q}, z_s)}{P(m \mid \hat{q}, z_s)} = a_j - a_m \geq \lambda\gamma - \varepsilon. \tag{10}$$

In particular, $\lambda\gamma > \varepsilon$ makes the state-matched node outrank m, even when m has a content-similarity advantage.

Proof. Substitute $a_j = b_j - \lambda D_j$ and $a_m = b_m - \lambda D_m$ into the posterior odds from Eq. 9. The assumed bounds give $(b_j - b_m) + \lambda(D_m - D_j) \geq -\varepsilon + \lambda\gamma$. □

BrainState aggregates only the top-M nodes. The following bound describes the approximation induced by this truncation. Order the $N = |\mathcal{J}|$ scores as $a_{(1)} \geq \cdots \geq a_{(N)}$, let S_M contain the first $M < N$ nodes, and let $G = a_{(M)} - a_{(M+1)}$. If $\epsilon_M = \sum_{j \notin S_M} P(j \mid \hat{q}, z_s)$ is the omitted posterior mass, then

$$\epsilon_M \leq \frac{(N - M)e^{-G}}{M + (N - M)e^{-G}}. \tag{11}$$

Indeed, the selected nodes have total unnormalized mass at least $Me^{a_{(M)}}$, while the remaining nodes have mass at most $(N - M)e^{a_{(M+1)}}$. Let V denote the learned residual projection. If every projected value $V(r_j)$ has norm at most R, the difference between its full-posterior average and the renormalized top-M average is at most $2R\epsilon_M$. Consequently, a larger retrieval margin concentrates the residual evidence on fewer condition nodes.

F Model architecture and Implementation

Figure 11 shows the complete BrainState architecture, and Table 8 lists the corresponding layer dimensions.

Algorithm 1 BrainState graph-memory refresh and state-conditioned prediction

procedure REFRESHMEMORY($\mathcal{D}_{\mathrm{tr}}, \theta$)
  for all $(x_i, y_i, d_i, \mathrm{id}_i) \in \mathcal{D}_{\mathrm{tr}}$ **do**
    $(q_i, z_{s,i}) \leftarrow$ ENCODEEVALNOGRAD($x_i; \theta$)
    Store $(q_i, z_{s,i}, y_i, d_i, \mathrm{id}_i)$ and class/node sums
  end for
  Compute p_c, k_{cd}, u_{cd} and $r_{cd} = k_{cd} - p_c$ from training sums
  return memory $\mathcal{M}$ containing stored embeddings, sums, counts, and IDs
end procedure

procedure PREDICT($x, \mathcal{M}, \mathrm{id}$)
  $f \leftarrow$ MAMBAEEGENCODER(x); $q \leftarrow$ CONTENTPROJECT(f); $z_s \leftarrow$ STATEPROJECT(f)
  $\ell_{\mathrm{base}} \leftarrow$ LINEARCLASSIFIER(q)
  if id is a training-trial ID **then**
    $(p_c, k_{cd}, r_{cd}, u_{cd}) \leftarrow$ READLEAVEONEOUT($\mathcal{M}, \mathrm{id}$)
      ▷ Subtract the stored trial, then recompute affected means and residuals.
  else
    $(p_c, k_{cd}, r_{cd}, u_{cd}) \leftarrow$ READFROZEN($\mathcal{M}$)
  end if
  $b_c \leftarrow \cos(q, p_c)/\tau$; $\alpha_c \leftarrow \mathrm{softmax}_c(b_c)$
  $h_{\mathrm{cls}} \leftarrow \sum_c \alpha_c V_{\mathrm{cls}}(p_c)$
  for all valid class–condition nodes $j = (c, d)$ **do**
    $a_j \leftarrow \cos(q, k_j)/\tau - \lambda \|z_s - u_j\|_2^2 / \dim(z_s)$
  end for
  $S \leftarrow \mathrm{TopM}(\{a_j\})$; $w_j \leftarrow \mathrm{softmax}_{j \in S}(a_j)$
  $h_{\mathrm{res}} \leftarrow \sum_{j \in S} w_j V_{\mathrm{res}}(r_j)$
  $\xi \leftarrow$ RETRIEVALCONFIDENCE($\{b_c\}, \{a_j\}_{j \in S}, \alpha, w$)
  $\beta \leftarrow \sigma(f_\beta([\mathbf{0}, \xi]))$; $\tilde{q} \leftarrow \mathrm{LN}(q + h_{\mathrm{cls}} + \beta h_{\mathrm{res}})$
  $g \leftarrow \sigma(f_g([q, \tilde{q}, \xi]))$; $\ell \leftarrow \ell_{\mathrm{base}} + g\, \delta(\tilde{q})$
  return $(\ell, \ell_{\mathrm{base}}, q, z_s)$
end procedure

For an EEG trial $x \in \mathbb{R}^{C \times T}$, BrainState first applies a temporal convolution with 40 filters of length 25, followed by a spatial convolution spanning all C electrodes. Batch normalization, ELU, average pooling (window 75, stride 15), and dropout produce $L = \lfloor (T - 99)/15 \rfloor + 1$ tokens of width 40; a 1×1 convolution completes the token projection. Two residual Mamba blocks then model the token sequence, each using a 40-dimensional input, an 80-dimensional expanded branch, a 16-dimensional selective state, and a causal depthwise convolution of width four. A two-layer fully connected readout maps the flattened sequence to a 256-dimensional trial feature f. Separate layer-normalized projections produce the 128-dimensional content query q and learned state z_s; a linear classifier on q yields the direct logits ℓ_{base}. The state branch also predicts a $3C$-dimensional descriptor of per-channel means, log root-mean-square amplitudes, and log root-mean-square first differences, and feeds a gradient-reversal class adversary during training.

The graph memory is constructed from training trials only. For each class c it stores a content prototype p_c, and for each available class–subject–session condition d it stores a content key k_{cd}, its residual $r_{cd} = k_{cd} - p_c$, and a mean state u_{cd}. Retrieval combines cosine content similarity with the mean squared distance between z_s and u_{cd}, then aggregates the top M condition residuals and the class prototypes. Confidence-dependent gates combine these memory values with q and add a learned correction to ℓ_{base}. During training, the query trial is excluded when its memory statistics are read; at validation and test time the frozen training memory is used. Table 8 gives the layer dimensions, and Figure 11 shows the information flow. Here K denotes the number of classes and N the number of available class–condition nodes.

Algorithm 2 BrainState training and inference

Require: Training trials $\mathcal{D}_{\mathrm{tr}}$ with class and subject–session labels; validation trials $\mathcal{D}_{\mathrm{val}}$; retrieval parameters M, τ, λ

Fit input normalization and state-descriptor normalization on $\mathcal{D}_{\mathrm{tr}}$
For each training trial, compute standardized $s(x)$ from pre-normalization EEG
Initialize model parameters θ and $\mathcal{M} \leftarrow \textsc{RefreshMemory}(\mathcal{D}_{\mathrm{tr}}, \theta)$
for epoch $= 1, \ldots, E$ **do**
  for all minibatches $\mathcal{B} \subset \mathcal{D}_{\mathrm{tr}}$ **do**
    for all $(x_i, y_i, \mathrm{id}_i) \in \mathcal{B}$ **do**
      $(\ell_i, \ell_{\mathrm{base},i}, q_i, z_{s,i}) \leftarrow \textsc{Predict}(x_i, \mathcal{M}, \mathrm{id}_i)$
    end for
    $\mathcal{L}_{\mathrm{cls}} \leftarrow \mathrm{CE}(\ell, y)$; $\mathcal{L}_{\mathrm{base}} \leftarrow \mathrm{CE}(\ell_{\mathrm{base}}, y)$
    $\mathcal{L}_{\mathrm{state}} \leftarrow \mathrm{MSE}(\textsc{StateDecoder}(z_s), s(x))$
    $\mathcal{L}_{\mathrm{decor}} \leftarrow \textsc{SquaredCrossCorrelation}(q, z_s)$
    $\mathcal{L}_{\mathrm{adv}} \leftarrow \mathrm{CE}(\textsc{ClassAdversary}(\mathrm{GRL}(z_s)), y)$
    $\mathcal{L} \leftarrow \mathcal{L}_{\mathrm{cls}} + \mathcal{L}_{\mathrm{base}} + w_s\mathcal{L}_{\mathrm{state}} + w_d\mathcal{L}_{\mathrm{decor}} + w_a\mathcal{L}_{\mathrm{adv}}$
    Update θ by backpropagation; keep $\mathcal{M}$ fixed within the epoch
  end for
  $\mathcal{M} \leftarrow \textsc{RefreshMemory}(\mathcal{D}_{\mathrm{tr}}, \theta)$
  $v \leftarrow$ validation score using $\textsc{Predict}(x, \mathcal{M}, \varnothing)$
  if v improves the best validation score **then**
    Save the paired model and memory $(\theta, \mathcal{M})$
  end if
end for
Restore the selected $(\theta, \mathcal{M})$
for all new EEG trials x **do**
  $(\ell, \ell_{\mathrm{base}}, q, z_s) \leftarrow \textsc{Predict}(x, \mathcal{M}, \varnothing)$
  Output predicted class $\arg\max_c \ell_c$
end for

Table 8: BrainState architecture for one trial. Dimensions omit the batch axis. $L = \lfloor (T-99)/15 \rfloor + 1$; $M = \min(10, N)$ with the default top-k setting.

Stage	Output	Parameter / setting	Operation
Input	$C \times T$	–	EEG trial
Temporal convolution	$40 \times C \times (T-24)$	$1 \to 40$, kernel 1×25	Local temporal features
Spatial convolution	$40 \times 1 \times (T-24)$	$40 \to 40$, kernel $C \times 1$	Cross-electrode features
Normalization and pooling	$40 \times L$	BN, ELU; pool $75/15$; dropout 0.5	Temporal tokenization
Token projection	$L \times 40$	$40 \to 40$, kernel 1×1	Channel projection, rearrange
Mamba encoder $\times 2$	$L \times 40$	$d_{\mathrm{state}} = 16$, $d_{\mathrm{conv}} = 4$, expansion 2	Selective scan, gate, residual
Trial readout	256	$40L \to 256 \to 256$	Flatten, ELU, dropout
Content and state	$q, z_s \in \mathbb{R}^{128}$	Separate $256 \to 128$	LN, linear, GELU, LN
Direct classifier	K	$128 \to K$	Base logits
Auxiliary state heads	$3C$ and K	$128 \to 3C$; $128 \to 64 \to K$	Reconstruction; adversary
Graph memory	K class, N condition nodes	128-dimensional stored values	Training-set means
Retrieval and fusion	K	top M; 128-dimensional fusion	State-aware rank, gated correction

G Default hyperparameters

Table 9 lists the default hyperparameters of BrainState. The layer dimensions of the network are given in Table 8 and are not repeated here. All data-dependent statistics, including the channel normalization, the state-descriptor normalization, and the graph memory, are computed from the training subjects of each fold only.

Table 9: Default hyperparameters of BrainState. Values are the defaults of the released code. C is the number of EEG channels and K the number of classes.

Group	Hyperparameter	Default
Input	Channels, window, sampling rate	32, $[0, 3)$ s after the cue, 256 Hz (32×768)
	Normalization	Per-channel z-score, fitted on training trials
	Data augmentation	None
Optimization	Optimizer	AdamW
	Learning rate	1×10^{-3}
	Weight decay	1×10^{-4}
	Batch size	64
	Maximum epochs	200
	Learning-rate schedule	Reduce on plateau, factor 0.5, patience 4
	Early stopping patience	15 epochs
	Gradient clipping	Global ℓ_2 norm 5.0
	Class weighting	None
	Numerical precision	float32, deterministic kernels
	Model selection	Validation accuracy
	Random seeds	0, 42, 3407
Encoder	Temporal filters, kernel length	40, 25
	Pooling window, stride	75, 15
	Mamba blocks	2
	State size, convolution width, expansion	16, 4, 2
	Dropout (tokenizer, Mamba residual)	0.5
	Dropout (readout layers 1, 2)	0.5, 0.3
	Trial feature, content, state dimensions	256, 128, 128
Memory and retrieval	Condition node	Class $\times$ subject $\times$ session
	Memory refresh	Before epoch 1 and after every epoch
	Retrieved condition nodes M	10
	Retrieval temperature τ	0.2
	State-distance weight λ	1.0
	Fusion width	128
	Correction-head dropout	0.1
	Gate bias initialization (β, g)	-2, -4
Objective	Weights of $\mathcal{L}_{\mathrm{cls}}$, $\mathcal{L}_{\mathrm{base}}$	1.0, 1.0
	State reconstruction w_s	0.1
	Content–state de-correlation w_d	0.1
	State class adversary w_a	0.1
	Gradient-reversal coefficient in $\mathcal{L}_{\mathrm{adv}}$	1.0
	State descriptor $s(x)$	Per-channel mean, log-RMS, log-RMS of first difference ($3C$), standardized on training trials

H Hyperparameter Sensitivity Analysis

We vary the learning rate, the batch size, and the amount of labelled data from the held-out subject on three TESSCCo folds (held-out subjects S21, S10, and S04). Each fold starts from its configuration used for the main results, one factor is changed at a time, and every setting is trained with seeds 0, 42, and 3407.

Learning rate and batch size. Over learning rates from 5×10^{-5} to 10^{-3} and batch sizes from 16 to 128, neither factor shows a direction that is consistent across folds (Figure 12, left). Within a fold, all settings lie within about 6 percentage points, and most differences are smaller than one seed standard deviation, so the results do not depend on a narrowly tuned optimum.

Calibration data from the held-out subject. Adding a small fraction of target trials to training changes accuracy little, and even at 80% it stays below 40%. This suggests that simply pooling a few calibration trials may not be enough to close the gap between subjects, although the small test sets at high fractions make this estimate uncertain.

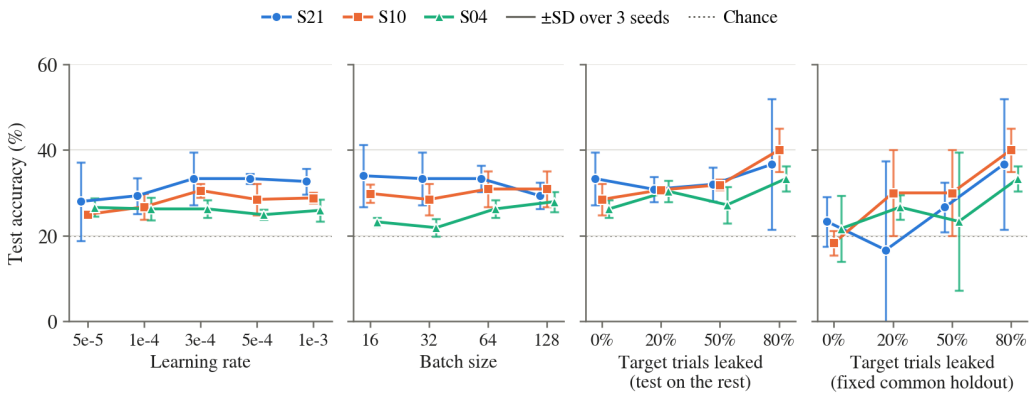


Figure 12: Sensitivity of BrainState on three TESSCCo folds (held-out subjects S21, S10, S04). From left to right: learning rate, batch size, fraction of the held-out subject's trials moved into training evaluated on the remaining trials, and the same fractions evaluated on a fixed common holdout of trials that are never moved. Markers are means over seeds 0, 42, and 3407, and error bars show ± 1 standard deviation over seeds. The tick 3e-4 marks 2.76×10^{-4}. The dotted line is chance (20%).

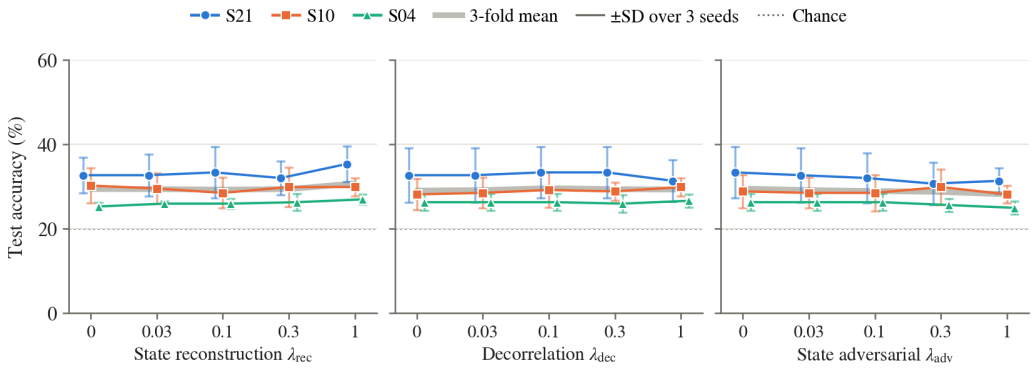


Figure 13: Test accuracy of BrainState as each auxiliary loss weight varies (mean $\pm$ SD over three seeds; grey line: mean over the three held-out subjects).

Sensitivity to the loss weights. We varied each auxiliary loss weight (λ_{rec}, λ_{dec}, λ_{adv}) over $\{0, 0.03, 0.1, 0.3, 1\}$ with all other settings fixed (Fig. 13). Subject-averaged accuracy stayed within 1.3 percentage points of the reported configuration (29.4%), below the seed-to-seed variation, with only a mild decline at large λ_{adv}. BrainState is therefore robust to the choice of these weights, although no single auxiliary term individually drives accuracy.

I Computational efficiency

We compare the decoding accuracy of BrainState and the seven baselines against their parameter counts and inference cost on TESSCCo. All results use the Spanish covert-speech (CS) task under the cross-subject leave-one-subject-out (LOSO) protocol. Figure 14 plots the mean LOSO accuracy of each model against its forward FLOPs per trial on a logarithmic axis. Bubble area encodes the parameter count, and each label gives the exact count in thousands.

Measurement protocol. We measure every model on a single trial with batch size one, in evaluation mode, without gradient computation, and on CPU. Each model receives the input used in our experiments. To clarify, both inputs cover the same 3-s window with 32 channels, but at different

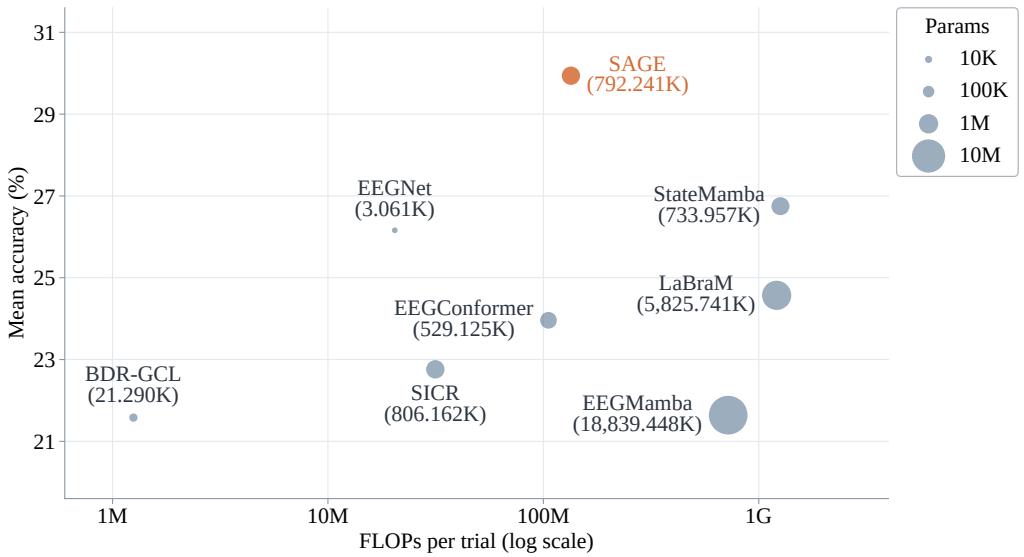


Figure 14: Accuracy and computational cost on TESSCCo (Native Spanish CS, cross-subject LOSO). The horizontal axis shows counted forward FLOPs for one trial on a logarithmic scale, and the vertical axis shows mean LOSO accuracy. Bubble area is proportional to the square root of the parameter count, and the labels give exact parameter counts in thousands. The counting protocol and its limitations are described in Appendix I.

sampling rates: the baselines receive 32×600 samples at 200 Hz, and BrainState receives 32×768 samples at 256 Hz. FLOPs are counted with the PyTorch profiler (`with_flops=True`), where one multiply–accumulate counts as two FLOPs. For the baselines, the profiler misses some convolutions, including `padding='same'` and several one-dimensional convolutions. We therefore remove the profiler's convolution counts and recount all convolution modules with forward hooks. For BrainState, the profiler covers the two-dimensional convolutional front end. The only convolutions it misses are the depthwise one-dimensional convolutions inside the two Mamba blocks, which add 61,440 FLOPs, less than 0.05% of the total. The BrainState count includes the state-conditioned retrieval over the 185 class–condition child nodes of a representative fold. It excludes the memory construction from training trials, which is not part of per-trial inference. Parameter counts cover the whole constructed model, including classification heads and auxiliary training branches.

Accuracy relative to cost. BrainState reaches the highest mean accuracy, 29.94%. This is 3.19 percentage points above the strongest baseline, StateMamba (26.75%), and 3.78 points above EEGNet (26.16%). BrainState also has a moderate cost. It requires 134.3M FLOPs per trial, 9.4 times fewer than StateMamba (1,260.8M) and 9.0 times fewer than LaBraM (1,210.9M). Its 792.241K parameters are comparable to those of StateMamba (733.957K) and SICR (806.162K). They are 7.4 times fewer than those of LaBraM (5,825.741K) and 23.8 times fewer than those of EEGMamba (18,839.448K). Among the models evaluated here, higher computational cost is not associated with higher accuracy. The two EEG foundation models are the most expensive, yet EEGMamba has the second-lowest accuracy (21.64%), and LaBraM (24.57%) remains below both EEGNet and StateMamba.

Cost of the state-aware components. Most of the BrainState cost comes from its convolutional front end. Two-dimensional convolutions account for 123.9M of the 134.3M counted FLOPs (92.3%). Relative to the memory-free variant in the ablation study (variant A), which uses the same backbone and input, the full model adds 6.45M FLOPs (5.0%) and 147,964 parameters (23.0%). These additions come from the class–condition memory, the state encoder, and the retrieval pathway. The accuracy gain therefore comes at a small increase in inference computation.

Table 10: Summary of baselines.

Method	Category	Backbone	Init.	Code
EEGNet Lawhern et al. (2018)	General EEG	CNN	Scratch	✓
EEGConformer Song et al. (2023)	General EEG	CNN+Transformer	Scratch	✓
LaBraM Jiang et al. (2024)	EFM	Transformer	Frozen	✓
EEGMamba Wang et al. (2025)	EFM	Mamba2	Frozen	✓
State Mamba Weng et al. (2026)	TL	SSM	Fold PT	–
SICR Jeon et al. (2021)	TL	MI decomposition	Fold PT	–
BDR-GCL Niu et al. (2026)	Brain-to-text	Graph contrastive	Fold PT	–

Lightweight baselines. EEGNet (3.061K parameters, 20.4M FLOPs) and BDR-GCL (21.290K parameters, 1.25M FLOPs) are cheaper than BrainState, but their accuracies are 3.78 and 8.36 points lower, respectively. The FLOP count for BDR-GCL substantially underestimates its true inference cost. The Fourier-transform, phase, and graph operations that dominate BDR-GCL are not fully covered by the profiler. BDR-GCL should therefore not be regarded as the least expensive baseline.

Interpretation of parameter counts and Flops. The whole-model count can differ from the number of parameters that are trained or used at inference. For LaBraM and EEGMamba, we fine-tune only the classification head (1,405 and 15,522,005 trainable parameters, respectively), but the full pretrained backbone still runs in every forward pass. SICR's 806.162K parameters include the mutual-information and discriminator modules used only during training; its classification path uses 3,669 parameters, and its FLOPs are measured along that path.

The reported FLOPs count only the covered operators and are not complete totals. The selective state-space scans in BrainState, StateMamba, and EEGMamba are not fully counted. Parts of the normalization, nonlinearities, and element-wise operations are also missing. The EEGMamba count uses the CPU reference implementation of Mamba2, so it does not reflect the runtime of the fused CUDA kernel. BrainState is measured at 256 Hz, whereas the baselines are measured at 200 Hz. Its front-end convolutions scale approximately linearly with the number of samples, so the BrainState figure is conservative relative to the baselines. Finally, FLOPs and parameter counts are hardware-independent proxies. They do not measure wall-clock latency or training cost, which also depend on the implementation and on memory access.

J Baseline Selection

We compare BrainState with seven baselines grouped into four categories: general-purpose EEG decoding models, EEG foundation models (EFMs), transfer learning (TL) methods, and a brain-to-text speech decoding model.

General EEG decoding models. EEGNet Lawhern et al. (2018) is a compact CNN combining temporal, depthwise spatial, and separable convolutions. We use the EEGNet-8,2 configuration as a conventional convolutional reference. EEGConformer Song et al. (2023) combines a convolutional feature extractor with a Transformer encoder to capture local EEG patterns and dependencies between temporal tokens. It provides a comparison with attention-based contextual modeling.

EEG foundation models. LaBraM Jiang et al. (2024) learns transferable EEG representations through channel-wise temporal patches and masked neural-code prediction, using a tokenizer trained by vector-quantized neural spectrum prediction. We use LaBraM Base to assess the utility of representations pretrained on heterogeneous EEG datasets. EEGMamba Wang et al. (2025) combines temporal and spectral patch embeddings with Mamba2 blocks. It provides a state-space foundation-model counterpart to the Transformer-based LaBraM.

Transfer learning methods. State Mamba Weng et al. (2026) models spatial and temporal EEG state transitions using a multichannel state-space architecture. It combines local–global attention with representation and pattern alignment objectives to capture consistent neural responses across subjects. Originally evaluated on emotion recognition and sleep staging, it provides a comparison

for cross-subject state modeling and alignment. Mutual Information, implemented as SICR Jeon et al. (2021), combines feature decomposition with mutual-information-based feature enrichment to learn subject-invariant and class-relevant representations. It provides an information-theoretic comparison for reducing subject-dependent variability while preserving task-relevant information.

Brain-to-text model. BDR-GCL Niu et al. (2026), abbreviated as BDR, incorporates brain dynamics representations into graph contrastive learning for imagined speech decoding. It combines EEG connectivity and dynamical information to model interactions between brain regions, providing a speech-specific graph learning baseline. Our comparison concerns speech-label decoding rather than unrestricted text generation.

Initialization and adaptation. EEGNet, EEGConformer, and SICR are initialized randomly and trained using the training subjects in each fold. For the two EEG foundation models, we load the authors' released pretrained checkpoints, which are the LaBraM Base student weights and the EEGMamba backbone weights. Their pretrained encoders remain frozen in evaluation mode, while newly initialized downstream classification heads are trained for our task. For LaBraM, the downstream `fc_norm` layer is also trained. This setting evaluates transfer through frozen pretrained representations rather than end-to-end encoder fine-tuning. State Mamba and BDR-GCL are initialized randomly for each fold and undergo pretraining using only that fold's training subjects before supervised classification; neither uses external pretrained weights.

Our EEGNet, EEGConformer, LaBraM, EEGMamba, and SICR implementations follow publicly released author code. For State Mamba and BDR-GCL, we use local implementations based on the papers, since an official public implementation could not be verified. These two results therefore represent paper-based adaptations rather than verified reproductions of the authors' algorithm.