Semantic Latent Geometry Reveals Imagination–Perception Structure in EEG

applied sciences

MDPI

Article

Semantic Latent Geometry Reveals Imagination–Perception Structure in EEG

Hossein Ahmadi, Martina Impagnatiello and Luca Mesin *

Mathematical Biology and Physiology, Department of Electronics and Telecommunications, Politecnico di Torino, 10129 Turin, Italy; hossein.ahmadi@polito.it (H.A.); martinaimpa97@gmail.com (M.I.)
* Correspondence: luca.mesin@polito.it

Abstract

We investigate whether representation-level, semantic diagnostics expose structure in electroencephalography (EEG) beyond conventional accuracy when contrasting perception and imagination and relating outcomes to self-reported imagery ability. Using a task-independent encoder that preserves scalp topology and temporal dependencies, we learn semantic features from multi-subject, multi-modal EEG (pictorial, orthographic, auditory) and evaluate subject-independent decoding with lightweight heads, achieving state-of-the-art or better accuracy with low variance across subjects. To probe the latent space directly, we introduce threshold-resolved correlation pruning and derive the Semantic Sensitivity Index (SSI) and cross-modal overlap (CMO). While correlations between Vividness of Visual Imagery Questionnaire (VVIQ)/Bucknell Auditory Imagery Scale (BAIS) and leave-one-subject-out (LOSO) accuracy are small and imprecise at $n = 12$, the semantic diagnostics reveal interpretable geometry: for several subjects, imagination retains a more compact, non-redundant latent subset than perception (positive SSI), and a substantial cross-modal core emerges (CMO $\approx$ 0.5–0.8). These effects suggest that accuracy alone under-reports cognitive organization in the learned space and that semantic compactness and redundancy patterns capture person-specific phase preferences. Given the small cohort and the subjectivity of questionnaires, the findings argue for semantic, representation-aware evaluation as a necessary complement to accuracy in EEG-based decoding and trait linkage.

Keywords: brain–computer interface; EEG signal processing; semantic feature extraction; Vividness of Visual Imagery Questionnaire; Bucknell Auditory Imagery Scale

Academic Editor: Alexander N. Pisarchik

Received: 22 December 2025
Revised: 3 January 2026
Accepted: 5 January 2026
Published: 8 January 2026

1. Introduction

Brain–computer interface (BCI) systems seek reliable mappings between neural activity and user intent, with non-invasive electroencephalography (EEG) remaining the most practical sensing modality thanks to its millisecond temporal resolution, portability, and safety [1,2]. Among the most consequential BCI targets are imagination and perception, two cognitive regimes central to communication, assistive control, and neurorehabilitation, yet they are notoriously difficult to disentangle in EEG because their neural signatures often overlap and unfold across rapid, multiscale spatiotemporal patterns [3–5]. Establishing robust, generalizable decoding of imagination versus perception is therefore both scientifically important and technically challenging [6,7].

Progress is hampered by the intrinsic complexity of EEG: non-stationarity, low signal-to-noise ratios (SNR), cross-subject variability, and rich interactions across space, time,

and frequency that encode high-level cognitive context. Traditional pipelines (e.g., band-power summaries, spectral power analysis, common spatial patterns (CSP)/functional-connectivity matrices, event-related potential (ERP) components) capture low-level or task-tied statistics while discarding topology, nonlinearity, and long-range dependencies. Classic machine learning (ML) stacks (e.g., support vector machines (SVMs) or random forests (RFs)) built on handcrafted features often overfit to a dataset or paradigm and provide limited leverage on cross-modal and cross-task invariances [8,9]. Meanwhile, applying conventional ML/deep learning (DL) directly to raw or minimally processed EEG imposes rigid inductive biases, demands large labeled cohorts, and often fails to preserve joint spatiotemporal structure, precisely where imagination/perception distinctions and cross-modal regularities reside. Even modern DL models (convolutional neural networks (CNNs), recurrent neural networks (RNNs), and Transformers) can learn strong representations, but they are typically tuned to single tasks (e.g., motor imagery (MI), ERP, or specific sensory modalities) and still struggle to generalize across datasets and experimental paradigms [10]. The net effect is brittle decoders and limited interpretability when faced with nuanced cognitive contrasts.

At the same time, EEG signal processing has advanced along several application fronts, underscoring the value of principled feature design and robust learning. Ensemble learning has strengthened MI classification [11]; CNN-based pipelines have delivered clinically meaningful outcome prediction for disorders of consciousness [12]; and adversarially trained frameworks have improved robustness and security in BCI architectures [13,14]. These gains illustrate how architectural choices and training paradigms can translate into tangible improvements, yet they also highlight a persistent representational limitation: progress within task-specific ecosystems does not by itself yield feature spaces that are portable across tasks, modalities, and subjects, nor do such features naturally support analyses that tie neural separability to human traits.

A parallel line of inquiry concerns the link between subjective imagery ability and neural representations [15]. The Vividness of Visual Imagery Questionnaire (VVIQ) and the Bucknell Auditory Imagery Scale (BAIS) quantify individual differences in the phenomenological clarity of visual and auditory imagery, respectively [16]. These psychometrics are increasingly relevant for BCIs because vividness may modulate both the separability and stability of neural patterns engaged during imagination and perception, with downstream implications for personalization, training efficiency, and reliability. Yet principled relationships between self-reported vividness and EEG characteristics remain elusive [17]. One reason is a representational mismatch: VVIQ and BAIS quantify subjective imagery vividness, whereas commonly used EEG descriptors, such as band-limited power and ERP component amplitudes, capture low-level signal statistics that are only indirectly related to phenomenological content. This mismatch helps explain why correlations with VVIQ/BAIS are often weak, inconsistent, or confounded by noise, task specifics, and subject heterogeneity.

These converging limitations point to the need for representations that emphasize semantic regularities, those components of the signal that covary with cognitive content rather than session or stimulus-specific idiosyncrasies [18,19]. We refer to this as *semantic feature extraction*: task-independent embeddings that retain the spatial organization of EEG while summarizing high-level, context-bearing structure assumed to be closer to meaning than to raw morphology [20,21]. In practical terms, such embeddings should support (a) subject-independent imagination/perception decoding and (b) trait-aware analyses that quantify how individual differences in imagery vividness relate to both decoding performance and the geometry of the learned latent space [8].

In this paper, we use the term “semantic” in an operational BCI sense: the latent representation is considered “semantic” if it preferentially captures structure related to cognitive

content (here, the imagination vs. perception regime and its modality-invariant regularities) while suppressing nuisance variability such as exemplar-level stimulus differences and subject/session idiosyncrasies. Accordingly, we do not claim access to linguistic semantics from EEG; rather, we target content-linked representational organization that supports generalization and enables representation-level diagnostics beyond decoding accuracy.

Motivated by this, recent studies have begun to explore topography-preserving representations and task-independent EEG embeddings that retain spatial organization while capturing semantic structure [19–24]. These methods aim to encode the spatiotemporal morphology of EEG as it relates to cognitive content, thereby providing a more suitable substrate for robust imagination/perception decoding and for detecting trait-level relationships with VVIQ/BAIS [21,25]. However, two critical gaps remain:

- **(G1)** evidence that semantically aligned representations support subject-independent decoding of imagination vs. perception across multiple sensory modalities;
- **(G2)** evidence that such representations expose stable brain–behavior relationships with VVIQ/BAIS beyond what is observed with traditional features and standard classifiers.

We address these gaps by applying a task-independent semantic feature extraction framework to a multi-subject, multi-modality dataset of semantic concepts encompassing perception and imagination across visual–pictorial, visual–orthographic, and auditory modalities. The approach preserves topography and temporal structure while learning semantically oriented embeddings in an unsupervised manner; lightweight supervised heads are then trained for downstream decoding. Crucially, to the best of our knowledge, this is the first study to test whether learned semantic EEG representations align with individual differences in imagery vividness (VVIQ/BAIS), evaluating both outcome-level subject-independent decoding metrics and representation-level geometry (e.g., separability/compactness). Concretely, we pose two questions:

- **(Q1)** Do semantically aligned features enable robust, subject-independent decoding of imagination vs. perception within each modality?
- **(Q2)** Do VVIQ (visual) and BAIS (auditory) scores relate to (i) per-subject subject-independent decoding metrics and (ii) latent-space separability/compactness indices that quantify semantic geometry?

Our contributions are threefold and map directly to the gaps above:

- **(C1)** A representation-centric framework for imagination/perception decoding that prioritizes semantic alignment, with multiscale embeddings learned without task labels, demonstrating subject-independent performance across three modalities (addresses G1).
- **(C2)** To our knowledge, the first trait-aware linkage between psychometric vividness (VVIQ/BAIS) and semantic EEG representations, assessed at two levels: (i) subject-independent decoding metrics and (ii) representation geometry (latent separability/compactness and cross-modal overlap), yielding interpretable brain–behavior associations that standard pipelines often miss (addresses G2).
- **(C3)** An empirical comparison against conventional models, showing that while direct accuracy–VVIQ/BAIS correlations are weak and imprecise in this cohort, representation-level semantic geometry diagnostics remain informative when outcome metrics saturate, revealing robust imagination/perception structure that conventional feature sets fail to expose.

The rest of this paper is organized as follows. Section 2 details the proposed methodology, including dataset description, task-independent preprocessing, semantic feature extraction, decoding heads, and correlation analysis with VVIQ/BAIS. Section 3 presents

decoding and correlation outcomes alongside latent-space diagnostics; Section 4 discusses implications for semantically driven BCI design and personalization; and Section 5 concludes with key insights and future directions.

2. Methodology

In this section, we first detail the dataset and trial structure, and then the preprocessing applied to the curator-provided EEG. Next, we present unsupervised semantic feature learning with a hierarchical dual AutoEncoder, followed by the supervised imagination vs. perception decoding heads. We then specify the evaluation protocol, define the latent-geometry diagnostics, describe the trait-aware correlation analysis (VVIQ/BAIS), lay out the statistical procedures, and finally report a mixed-effects analysis of Keep_Ratio.

2.1. Dataset

We use the open-access *EEG-based BCI Dataset of Semantic Concepts for Imagination and Perception Tasks* (12 participants; 124 scalp EEG electrodes; 1024 Hz; 24-bit), recorded with active-shielded gel caps laid out according to the 5% system (reference CPz; ground at the left mastoid) [26]. Recordings were conducted in a sound- and light-attenuated room; triggers were time-stamped using lab streaming layer (LSL). The release follows the brain imaging data structure (BIDS) and provides both raw and curated preprocessed data: raw `.cnt/.evt` converted to `.set/.fdt` with aligned event `.tsv`, plus per-session `.fif` files containing MNE-readable EEG and events. Event labels encode the exact stimulus instance (e.g., modality, concept, complexity/voice).

Each participant completed two imagery vividness questionnaires prior to the task: VVIQ (1–5) and BAIS-V (1–7). Two participants did not provide questionnaire data; across the cohort, the mean (std) vividness scores are VVIQ 3.75 (0.55) and BAIS 4.76 (0.85).

The paradigm orthogonally varies the task (perception, imagination) and the modality (visual–pictorial, visual–orthographic, auditory), while holding semantic identity constant across exemplars (flower, penguin, guitar). Condition epoch durations are as follows: visual perception 3 s, visual imagination 4 s; auditory perception 2 s, and auditory imagination 4 s. For auditory trials, a 1 s white-noise mask separates perception and imagination to prevent carryover. Participants completed 1 or 2 sessions (9 subjects, single sessions; 3 subjects, two sessions).

Figure 1 summarizes the stimulus types and the overall experimental structure. The dataset includes a diverse set of exemplars for each modality:

- **Pictorial:** Images with three levels of complexity, i.e., simple, intermediate, and naturalistic, with eight exemplars at the simple level and nine at both intermediate and naturalistic for flower and guitar, and nine at each level for penguin.
- **Orthographic:** Thirty text stimuli per category, presented in five colors and six font styles.
- **Audio:** Spoken words recorded in three voice tones (normal, low, high).

2.1.1. Trial Structure and Epoching

Each trial comprises two phases—perception (stimulus presentation) and imagination (mental recreation)—separated by a brief mask (visual: 500 ms; auditory: 1000 ms). We follow the dataset's event timing and define two non-overlapping EEG epochs per trial: one spanning the perception phase and one spanning the imagination phase. Epoch boundaries are aligned to the LSL time-stamps; the inter-phase mask and inter-trial intervals are excluded. For visual modalities, perception epochs are 3 s and imagination epochs are 4 s; for the auditory modality, perception epochs are 2 s and imagination epochs are 4 s.

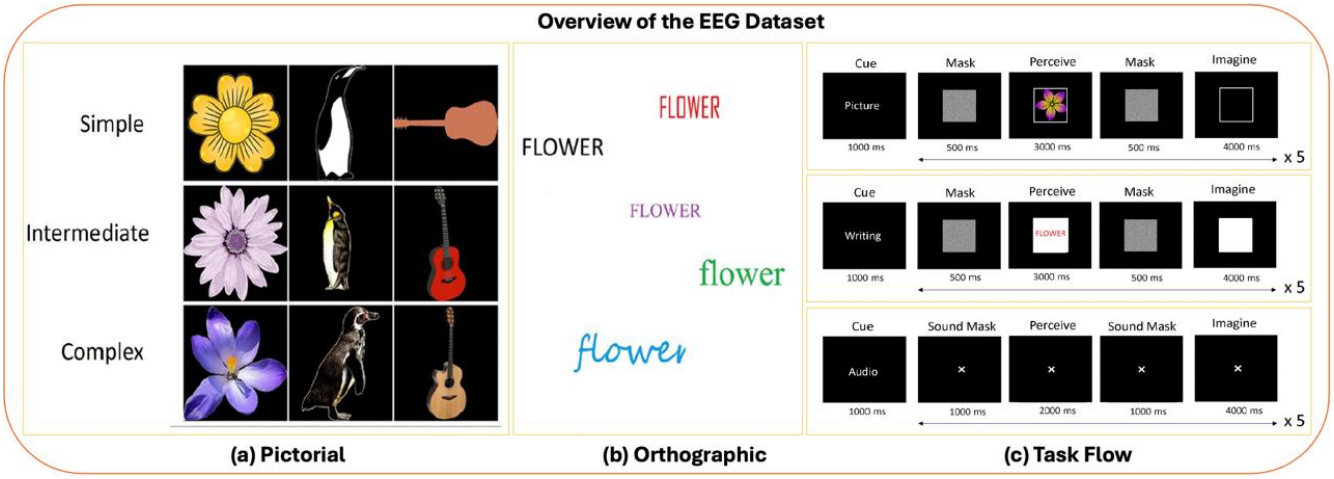

Figure 1. Overview of the EEG dataset and experimental protocol [26]. (**a**) Pictorial modality: Exemplar images of three concepts rendered at three visual-complexity levels. (**b**) Orthographic modality: The same concepts presented as words with variation in font, size, and color. (**c**) Trial flow for each modality: *Cue* (1000 ms) → *Mask* (visual: 500 ms; audio: 1000 ms) → *Perceive* (visual: 3000 ms; audio: 2000 ms) → *Mask* (visual: 500 ms; audio: 1000 ms) → *Imagine* (4000 ms); each block is repeated ×5 per concept. The design targets semantic identity across modalities and contrasts perception vs. imagination, enabling decoding of conceptual meaning from EEG.

2.1.2. Labeling

For each epoch, labels are derived from the orthogonal factors {*Task* ∈ [Perception, Imagination], *Modality* ∈ [Pictorial, Orthographic, Auditory], *Concept* ∈ [*flower*, *penguin*, *guitar*]}. Within-class exemplar variability (image complexity, text font/color, voice tone) is treated as nuisance variation and not used as a target label; these factors remain within-class to encourage semantic invariance.

2.1.3. Data Volume and Sessions

Across $N = 12$ participants (some subjects have 2 sessions), each condition (Task × Modality) contains up to 150 epochs per participant per session; actual completion typically ranges from 64 to 150 due to fatigue and breaks (see Table 2 for epoch durations and Table 4 for per-condition counts in [26]). Nine participants completed one session, and three completed two sessions. In all analyses, session identity is tracked for split integrity and session-transfer diagnostics.

2.2. Preprocessing

We start from the curators' preprocessed `.fif` data (bad-channel handling via PyPREP, common average reference (CAR), 50/100/150 Hz notch, high-pass filter for independent component analysis (ICA), and ICA/Electrooculography (EOG) component removal) [26]. To that, we apply the minimal, leakage-safe steps required by our semantic framework [21]: (i) resample to 128 Hz; (ii) zero-phase finite impulse response (FIR) low-pass-filtered at 45 Hz; (iii) per-channel z-scoring with train-fold statistics only; (iv) epoching to the phase durations above; (v) 1 s windows with 50% overlap (128 × 124) excluding masks; (vi) light Gaussian noise ($\sigma \approx 0.01$) in train windows only. This preserves the curated denoising while standardizing bandwidth and windowing for our encoder.

2.3. Unsupervised Semantic Feature Learning

We reuse our hierarchical dual AutoEncoder from [21]: a CNN AutoEncoder captures local spatiotemporal structure and feeds a Transformer AutoEncoder for long-range dependencies (2 Transformer encoder/decoder layers, 8 heads, d_{model} = 128, feed-forward network (FFN) 256, dropout 0.1). Training minimizes mean squared error (MSE) with Adam (learning rate 10^{-4}, batch size 64), early stopping, and a `ReduceLROnPlateau` scheduler, which decreases the learning rate when the validation loss stops improving. In each leave-one-subject-out (LOSO) fold, models are re-initialized and trained only on windows from training subjects; the encoder is then frozen for that fold. The Transformer encoder outputs a 128-D latent vector per 1 s window; we ℓ_2-normalize and use the temporal mean as the window feature. A quality gate requires a relative reconstruction error of $<1\%$ before any supervised training. Component-level ablations of the encoder variants are reported in [21]; in the present work we treat the encoder as a fixed, previously validated semantic extractor and focus on representation-level diagnostics and trait-linkage analyses on this dataset.

2.4. Supervised Imagination (I) ↔ Perception (P) Decoding Heads

With the encoder frozen, we train a compact 2D-CNN head (64/128/256/512, 3×3 kernels, batch normalization, spatial attention, global average pooling (GAP), dropout 0.5) separately for each modality to discriminate between imagination and perception. Optimization used Adam (learning rate 10^{-3}, weight decay 10^{-4}) with early stopping on validation balanced accuracy. Decision thresholds were selected on the validation fold using Youden's J statistic. We also probe cross-phase transfer (train on perception, test on imagination P↠I, and vice versa I↠P) within LOSO to quantify invariance [21].

2.5. Evaluation Protocol

Unless stated otherwise, evaluation is performed using LOSO: all sessions of the held-out subject constitute the test set; samples for inner validation are drawn from training subjects for early stopping and thresholding. All normalization parameters (z-score mean/std) are estimated on the training fold only and then applied to validation/test folds to prevent data snooping. The metric for the supervised task is balanced accuracy. To express uncertainty, we report 95% confidence intervals (CIs) via stratified bootstrap (B = 2000) and assess above-chance via label permutations (1000 shuffles per fold).

2.6. Latent-Geometry Diagnostics

We analyze the frozen latents using the held-out subject to relate representational structure to traits. A correlation-pruning step removes redundancy before computing diagnostics. Throughout, the frozen encoder outputs a d = 128-dimensional latent vector per window; within each subject $\times$ phase $\times$ modality, we compute a 128×128 Pearson correlation matrix across windows and apply greedy pruning to obtain the retained index set $S(\tau)$. All diagnostics are computed post hoc on held-out-subject latents and are not used to tune the supervised decoder. Let $\mathcal{T}$ denote the set of correlation thresholds and $|\mathcal{T}|$ its cardinality:

- **Greedy correlation filtering:** Within each subject $\times$ phase $\times$ modality, we iteratively prune latent channels whose absolute Pearson correlation with an already-kept channel exceeds a threshold (τ). This yields a retained set $S(\tau)$ of latent-channel indices at each τ. We evaluate a grid ($\tau \in \{0.15, 0.25, \ldots, 0.85\}$) (8 thresholds).
- **Keep Ratio:** At each τ, the fraction of latent channels retained after pruning, $K(\tau)$. We define $K(\tau) = |S(\tau)|/d$ (with d=128). We also summarize the per-subject *mean keep ratio*, meanKR $= \frac{1}{|\mathcal{T}|} \sum_{\tau \in \mathcal{T}} K(\tau)$, which is later used as a sparsity covariate.

- **Semantic Sensitivity Index (SSI):** The signed area under the curve (AUC) of the phase-difference in Keep_Ratio, SSI $= \mathrm{AUC}\big(\Delta K(\tau)\big)$, where

$$\Delta K(\tau) = K_{\text{imagine}}(\tau) - K_{\text{perception}}(\tau). \quad (1)$$

  In practice, the SSI is computed by trapezoidal integration over the predefined τ grid, preserving the sign of $\Delta K(\tau)$.
- **Full $\Delta K(\tau)$ curves:** Subject-wise trajectories across (τ) to localize where phase differences emerge (e.g., only at aggressive pruning).
- **Cross-Modal Overlap (CMO):** Jaccard overlap of kept-feature sets between modality pairs at matched τ (Audio–Pictorial, Audio–Orthographic, Orthographic–Pictorial). We report per-subject means across thresholds and distributions across subjects. At each τ, for a modality pair (m_1, m_2), we compute $\mathrm{CMO}_{m_1,m_2}(\tau) = \frac{|S_{m_1}(\tau) \cap S_{m_2}(\tau)|}{|S_{m_1}(\tau) \cup S_{m_2}(\tau)|}$ and summarize it by averaging over the τ grid.
- Using a predefined grid avoids committing to a single arbitrary correlation cutoff and reduces the risk of cherry-picking. Accordingly, our primary indices (SSI and mean CMO) integrate information over τ, while threshold-resolved trajectories are reported to show where effects emerge and to verify robustness.

2.7. Trait-Aware Correlation Analysis (VVIQ/BAIS)

We test whether inter-individual imagery ability predicts (i) decodability and (ii) latent geometry:

- **Performance indices:** Modality-specific LOSO balanced accuracy. We also summarize P→I and I→P transfer within LOSO.
- **Geometry indices:** SSI (overall and per modality), full $\Delta K(\tau)$ behavior, and CMO. Questionnaires (VVIQ, BAIS) are used as released; subjects missing a questionnaire are excluded listwise for that trait. For partial correlations, we control for meanKR (global sparsity) and the other questionnaire (e.g., the partial correlation between SSI and VVIQ controlling for meanKR and BAIS).

2.8. Statistics

Because (n) is small and scales are ordinal, we report Spearman's ρ with bias-corrected bootstrap 95% CIs (B = 5000) and permutation (p); Kendall's τ_b is provided for robustness. We control the false discovery rate (FDR) across pre-registered primaries (visual: VVIQ vs. visual balanced accuracy and SSI; auditory: BAIS vs. auditory balanced accuracy and SSI) via Benjamini–Hochberg (BH) at q = 0.10. Sensitivity checks: (i) 5% Winsorization of extreme performance values; (ii) skipped correlations; (iii) rank-based regression with usable-trial count and second-session present as pragmatic covariates.

2.9. Mixed-Effects Analysis of Keep_Ratio

To leverage within-subject trajectories at a fixed pruning level, we fit a linear mixed-effects model to $K(\tau)$ at $\tau = 0.35$: fixed effects for modality, phase, VVIQ, and BAIS; random intercepts and random slopes for phase by subject. This yields the random-slopes restricted maximum likelihood (REML) fit reported (group variance, phase-slope variance, and best linear unbiased predictor (BLUP) slopes used for trait correlations).

To consolidate the steps introduced above, Figure 2 summarizes the entire pipeline from data acquisition and psychometrics through preprocessing, unsupervised semantic encoding, and quality gating to downstream decoding and latent-geometry/trait analyses. The blocks in the diagram align with the steps in this Methodology and provide a single visual reference for the procedures used in Section 3.

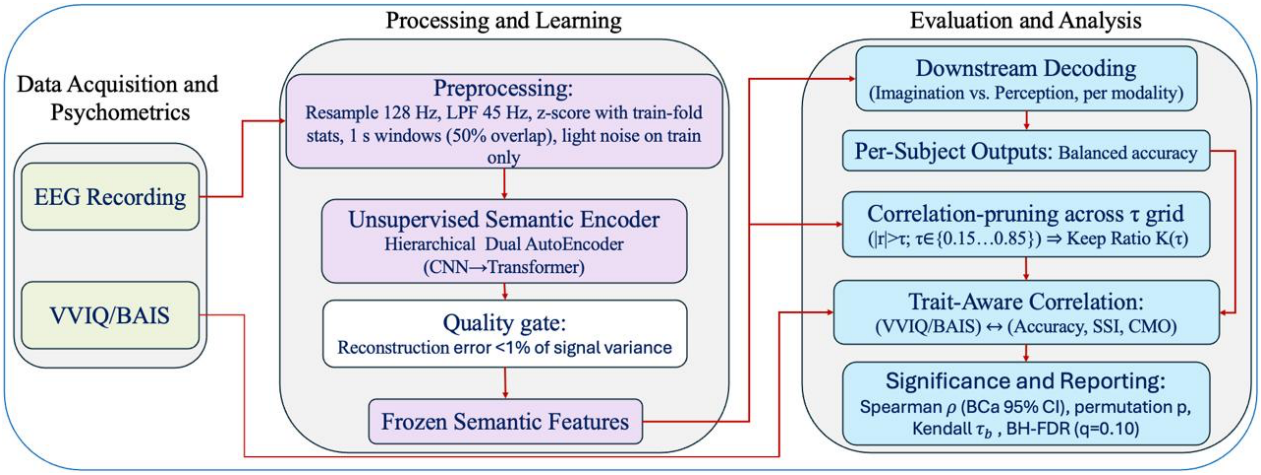

Figure 2. System model and analysis flow. EEG and psychometrics (VVIQ/BAIS) are collected (**left**). Signals are resampled to 128 Hz, low-pass-filtered at 45 Hz, z-scored with train-fold statistics, windowed into 1 s segments with 50% overlap, and lightly noised on train only. A hierarchical dual AutoEncoder (CNN→Transformer) learns unsupervised semantic features; windows passing a quality gate (reconstruction MSE $<$ 1% of signal variance) are frozen (**middle**). Evaluation (**right**) comprises (i) LOSO downstream decoding of imagination vs. perception per modality with per-subject balanced accuracy; (ii) correlation pruning across a threshold grid ($|r| > \tau$, $\tau \in [0.15, 0.85]$) to obtain Keep_Ratio ($K(\tau)$), its change $\Delta K(\tau)$, and the Semantic Sensitivity Index (SSI $=$ AUC(ΔK)); (iii) trait-aware correlations between VVIQ/BAIS and Accuracy/SSI/CMO. Statistical reporting uses Spearman's ρ with bias-corrected and accelerated (BCa) bootstrap 95% CIs (B = 5000), permutation p-values, Kendall's τ_b, and Benjamini–Hochberg (BH) FDR at q = 0.10.

3. Results

In this section, we report our results in three phases. First, we verify that the unsupervised encoder–decoder learns a stable latent space and reconstructs EEG within a pre-set error bound, reusing our previously introduced semantic feature extractor [21]. Next, we evaluate those latents in supervised decoding tasks. Finally, we analyze how both the learned semantic features and supervised outcomes relate to VVIQ/BAIS. Models were implemented in PyTorch 2.0+/Python 3.11 on a single NVIDIA RTX 4090 (CUDA 12.0).

3.1. Unsupervised Pretraining

The pretraining objective reconstructs short EEG windows while preserving scalp topology and temporal dependencies, producing task-agnostic latent features reused in all downstream experiments. Figure 3 compares original vs. reconstructed EEG for a random channel/trial across the three modalities. The waveform overlap is tight; per-panel MSEs (≈0.0026–0.0034) satisfy our gate of MSE $<$ 1% of the signal variance, indicating that the bottleneck retains structured signal content rather than fitting noise. For completeness, the training/validation loss and learning-rate schedule (Figure S1) and a subject-colored t-SNE visualization of the learned latents (Figure S2) are provided in the Supplementary Materials.

Here, topology preservation refers to respecting the spatial neighborhood structure of scalp EEG in the encoder's early (CNN) processing, while the Transformer models longer-range temporal dependencies. Moreover, the latent-channel correlation structure analyzed in SSI/CMO can be viewed as a representation-level analog of functional coupling (without claiming source-level connectivity), enabling interpretable redundancy/overlap diagnostics in a form consistent with network-style reasoning.

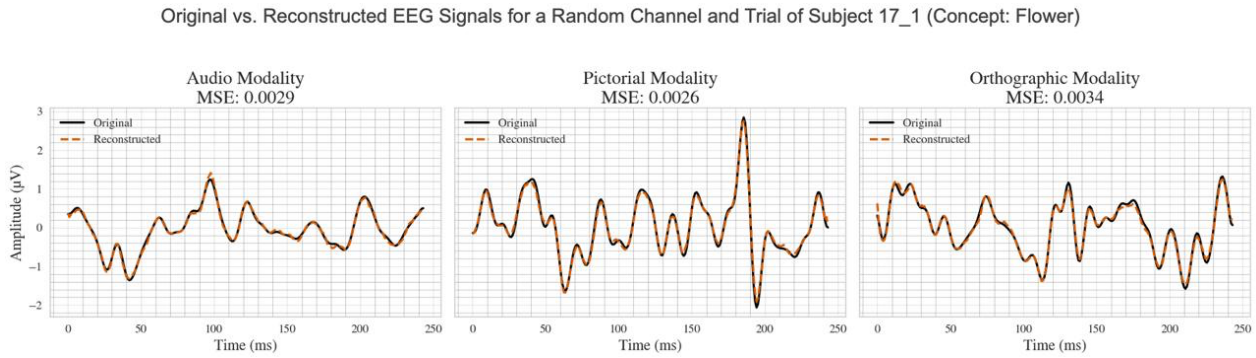

Figure 3. Original vs. reconstructed EEG for a random channel/trial across modalities. Black: original; orange dashed: reconstruction. Per-panel MSEs meet the MSE $< 1\%$ threshold, indicating that the encoder retains structured signal content while suppressing noise.

3.2. Supervised Downstream Decoding

We assess the utility of the learned latents for supervised decoding using a lightweight classifier (Section 2.4) trained on frozen representations under LOSO. Figure 4 summarizes balanced accuracy (mean $\pm$ SD across subjects) against representative baselines re-implemented following their original settings when applicable. Across all three modalities, our approach achieves the strongest mean performance with low across-subject variability (pictorial: $96.9 \pm 0.9\%$; orthographic: $96.3 \pm 1.1\%$; audio: $88.9 \pm 2.2\%$). The full per-subject breakdown (15 subjects) for each modality, including all baseline methods, is provided in the Supplementary Materials (Tables S1–S3).

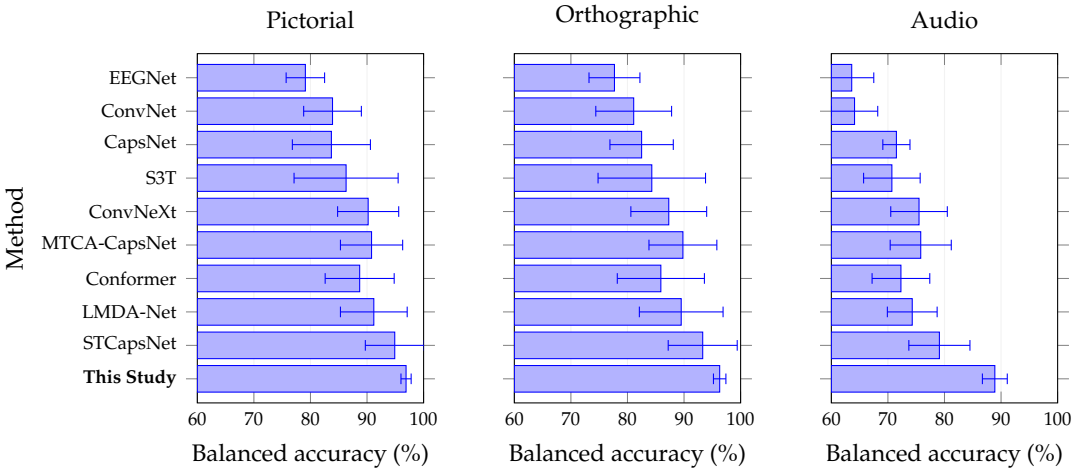

Figure 4. Balanced accuracy (mean $\pm$ SD) under LOSO for three stimulus modalities.

3.3. Correlation Analysis

We investigate whether inter-individual imagery ability (VVIQ) and auditory imagery control (BAIS) relate to (i) downstream decoding performance (LOSO accuracies from Section 3.2) and (ii) the geometry/robustness of the learned semantic features from Section 3.1. We report effect sizes with uncertainty and permutation tests; interpretation is deferred to the Discussion.

VVIQ and BAIS scores are available for $n = 13$ subjects; missing entries reduce the effective sample size in all trait analyses [26]. Full per-subject questionnaire values and dense ranks are provided in the Supplementary Materials (Table S4).

We test correlations between VVIQ/BAIS and LOSO accuracies (per modality and overall). Given small n and ordinal self-report scales, we emphasize Spearman correlations with bootstrap CIs and permutation p-values for the primary pairs (visual: pictorial vs.

VVIQ; auditory: audio vs. BAIS), alongside overall accuracy. The full correlation panel (including Pearson and partial correlations controlling for mean Keep_Ratio) is reported in the Supplementary Materials (Table S5).

To connect traits to representation structure, we use threshold-resolved correlation filtering in the latent space: latent channels with absolute Pearson correlation $|r| > \tau$ are pruned (greedy survivor selection), and we track the retained-feature fraction (Keep_Ratio) per phase and modality. This yields

$$\Delta K(\tau) = K_{\text{imagine}}(\tau) - K_{\text{perception}}(\tau) \quad (2)$$

as a function of threshold. Figure 5 visualizes the post-filter latent correlation structure for one subject at $\tau = 0.45$. Representative pre-filter latent correlation maps are provided in the Supplementary Materials (Figure S3).

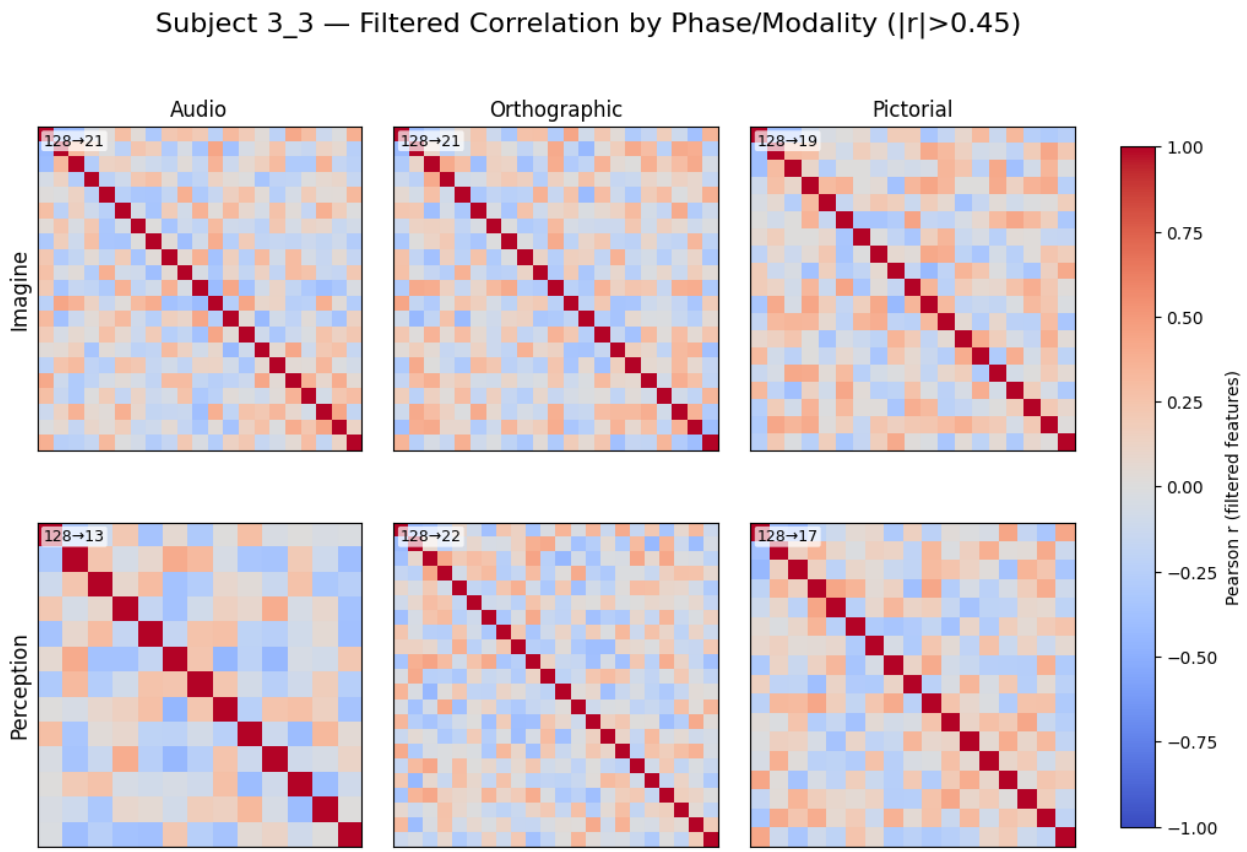

Figure 5. Post-filter latent structure ($|r| > \tau$ pruning). Feature–feature Pearson r matrices for one subject–session across phase (rows) and modality (columns) after greedy correlation filtering at $\tau = 0.45$. Many off-diagonal entries disappear, indicating a leaner set of non-redundant latent features used to compute Keep_Ratio and $\Delta K(\tau)$. Subject labels follow the dataset convention `subject_session` (e.g., 3_3 = subject 3, session 3) as released in [26].

3.3.1. Semantic Sensitivity Index (SSI)

To summarize phase differences across thresholds, we define the Semantic Sensitivity Index as the signed area under the curve of $\Delta K(\tau)$: SSI $= \text{AUC}(\Delta K(\tau))$. Figure 6 shows the distribution of the overall SSI distribution across subjects (thresholds: $\tau \in \{0.15, \ldots, 0.85\}$). To localize where phase differences emerge (e.g., only at aggressive pruning), subject-wise $\Delta K(\tau)$ trajectories are shown in the Supplementary Materials (Figure S4).

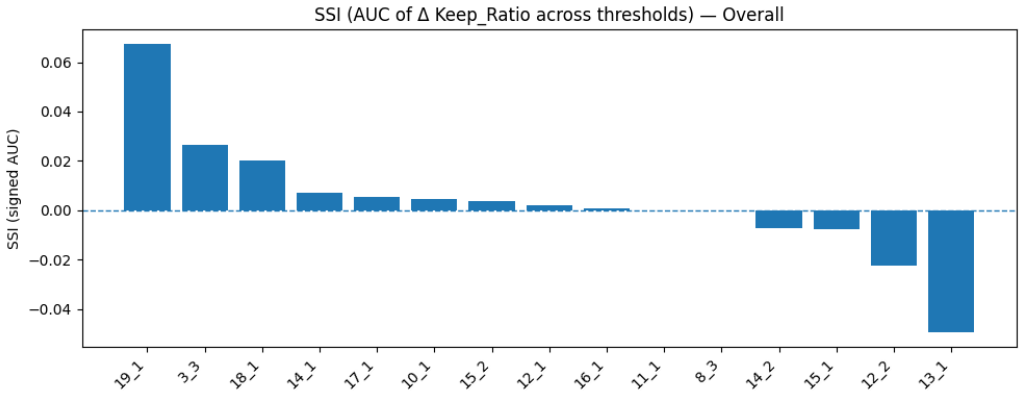

Figure 6. Overall SSI distribution by subject–session. Positive values indicate that imagination retains more latent structure than perception when features are pruned by correlation threshold; negative values indicate the opposite. The SSI summarizes threshold effects via AUC of $\Delta K(\tau)$ over $\tau \in \{0.15, 0.25, \ldots, 0.85\}$ (full $\Delta K(\tau)$ curves are shown in Supplementary Figure S4). Subject labels follow `subject_session` naming from [26].

3.3.2. Cross-Modal Overlap (CMO)

Finally, to quantify cross-modal stability of the learned latent core, we compute the Jaccard overlap of kept-feature sets between modality pairs at matched thresholds. Figure 7 summarizes CMO distributions across subjects and thresholds for each modality pair. Values cluster around ∼0.5–0.8, indicating a substantial cross-modal latent core.

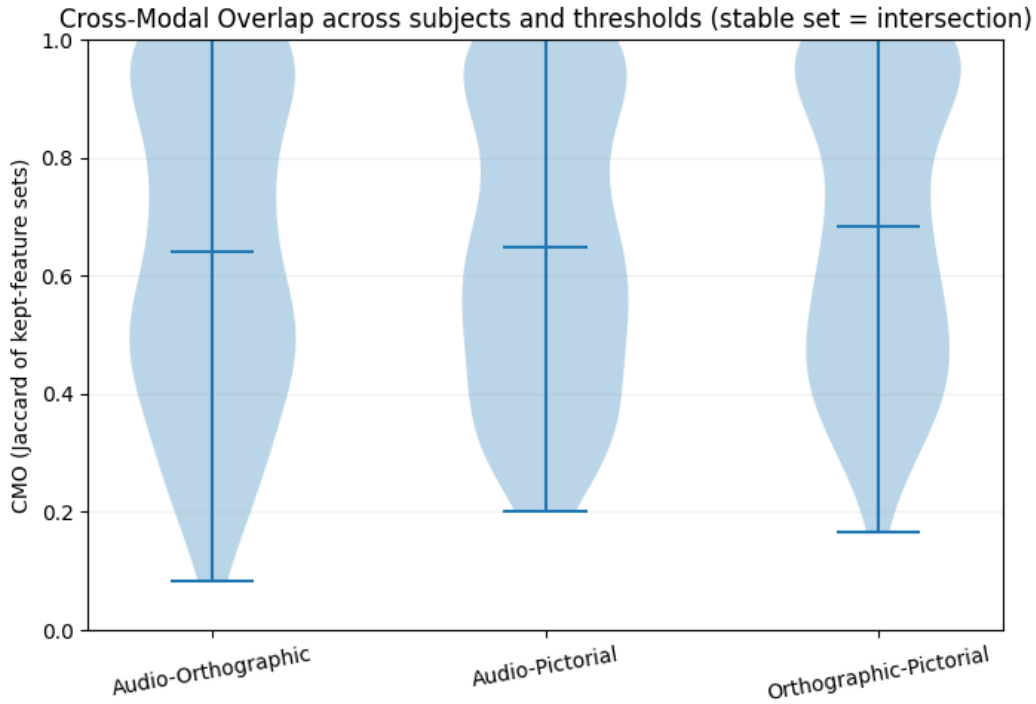

Figure 7. CMO distributions across subjects and thresholds for each modality pair. Horizontal bars denote medians; wider regions denote more frequent overlaps. Values around ∼0.5–0.8 indicate a substantial cross-modal latent core.

Across supervised outcomes and representation-level diagnostics, the associations with VVIQ/BAIS are small and imprecise at $n = 13$, yet the latent-space probes (e.g., SSI and CMO) expose structure that accuracy alone does not capture. In the Discussion, we revisit these patterns, emphasizing sample-size constraints, potential self-report biases, and the value of complementary semantic feature diagnostics for brain decoding evaluation.

4. Discussion

This study evaluated whether representation-level semantic diagnostics can reveal subject-specific structure that conventional outcome metrics (e.g., accuracy) and traditional feature pipelines often obscure. Three findings stand out:

- First, the proposed model achieves consistently high LOSO decoding performance across modalities (Figure 4), yet accuracy shows weak and imprecise associations with self-reported imagery traits (Supplementary Table S5);
- Second, latent-geometry probes based on correlation-aware sparsification expose systematic imagination–perception asymmetries and cross-modal structure (Figures 5–7) that are not visible at the outcome level;
- Third, these probes provide mechanistic descriptors of how cognition occupies the learned space (compactness/redundancy/overlap), complementing performance-centric evaluation.

In the following subsections, we elaborate on each of these findings, interpreting their implications for imagination–perception decoding and semantic EEG representations.

4.1. Accuracy–Vividness Mismatch

Across pictorial, orthographic, and audio tasks, the decoder attains strong LOSO performance with low across-subject variability (Figure 4). However, correlations between accuracy and VVIQ/BAIS are small, unstable in sign across pairs, and non-significant in this cohort (Supplementary Table S5). This pattern is consistent with a measurement mismatch: accuracy reflects discriminability under a specific training/testing protocol, whereas VVIQ and BAIS measure self-reported vividness/control on coarse ordinal scales and outside the decoding context. In addition, the effective sample for trait analyses is modest ($n = 13$; Supplementary Table S4), so even moderate associations would be difficult to estimate precisely.

A second, complementary perspective comes from subject-wise rank structure. In Supplementary Tables S1–S3, baseline pipelines induce a wide spread of ranks that separate "easy" and "hard" participants. In contrast, the rank ordering induced by this study is noticeably flatter, reflecting uniformly high performance across subjects. This compressed rank spread supports the interpretation that the learned semantic latents reduce subject-specific idiosyncrasies and yield a more subject-invariant representation than traditional pipelines, even though this invariance does not imply direct alignment with questionnaire scores.

4.2. Subject-Specific Heterogeneity and Threshold Sensitivity

Correlation-threshold filtering exposes substantial inter-individual variability in latent redundancy and in the proportion of retained features (Keep_Ratio) at a fixed threshold τ. While Figure 5 provides a representative post-filter visualization for one subject, the cohort-level behavior is not characterized by a universal pruning pattern. Instead, subjects exhibit distinct $\Delta K(\tau)$ trajectories across thresholds (Supplementary Figure S4), indicating that imagination–perception differences can emerge at different sparsification regimes and with different magnitudes. Thus, the reported threshold-resolved curves (Supplementary Figure S4) serve as an explicit sensitivity analysis with respect to τ, while SSI/mean CMO summarize effects without selecting a single cutoff.

This observation is conceptually important: traditional handcrafted pipelines typically return a fixed-size feature vector for every participant, which limits their ability to express person-specific differences in redundancy or effective dimensionality. In contrast, semantics-preserving latents admit variable cardinality after pruning, enabling a direct, interpretable readout of how much non-redundant structure each subject retains under a common criterion. Consequently, subject-conditioned analyses are more appropriate than assuming a single, cohort-wide semantic subset.

4.3. Latent Geometry vs. Vividness

The latent space learned by unsupervised pretraining is stable and structured: reconstructions are tight (Figure 3), and the latent manifold exhibits organized subject-wise clustering (Supplementary Figure S2). Raw feature–feature correlations show structured redundancy (Supplementary Figure S3), motivating correlation-based sparsification as a diagnostic lens rather than a classification trick. Because the downstream decoder is evaluated under LOSO, subject identity cannot be used as a shortcut for imagination–perception classification; thus, subject-wise clustering is more plausibly interpreted as evidence of a low-noise and geometrically organized embedding.

Building on this foundation, the threshold-resolved diagnostics provide a representation-level view of imagination versus perception:

- **SSI** indicates whether imagination tends to retain more (or less) non-redundant latent support than perception as sparsification tightens. The SSI distribution shows a meaningful between-subject spread around a near-zero center (Figure 6), consistent with heterogeneous phase preferences that are not captured by accuracy.
- **Full** $\Delta K(\tau)$ **curves** localize where differences occur across the threshold grid (Supplementary Figure S4). Several subjects exhibit positive imagination margins primarily at aggressive pruning (large τ), suggesting that the most selective latent channels may preferentially support imagery for those individuals.
- **CMO** indicates substantial cross-modal overlap of retained-feature sets (Figure 7), consistent with a modality-invariant latent core. This cross-modal stability provides a representation-level explanation for why frozen-latent decoders can perform strongly across modality-specific tasks.

Taken together, these diagnostics shift the interpretation from did the model classify? to what latent structure supports classification, and how does that structure differ across phases and individuals? This distinction matters when accuracy is already high and relatively uniform.

4.4. Why Questionnaire Alignment Is Weak in This Cohort

Two non-exclusive explanations appear to be the most plausible. First, VVIQ and BAIS are context-free self-reports and may not track the task-specific neural selectivity that governs imagination–perception separability under pruning. Second, statistical power is limited: with $n = 13$, confidence intervals around correlations are necessarily wide (Supplementary Table S5). Under these constraints, the main scientific value of the present results is less about establishing strong trait correlations and more about demonstrating that latent geometry provides interpretable, phase- and modality-sensitive descriptors that remain informative when accuracy saturates.

4.5. Implications

These results suggest four practical implications for semantic EEG decoding:

1. High accuracy is necessary but not sufficient for scientific interpretation: strong LOSO performance can coexist with weak trait alignment when heterogeneous strategies are collapsed into a single scalar metric.
2. Representation-aware diagnostics provide interpretable mechanistic descriptors—compactness under pruning (SSI), modality-invariant structure (CMO), and threshold-localized phase differences ($\Delta K(\tau)$)—that complement outcome metrics.
3. The observed cross-modal overlap supports the design of decoders that explicitly target modality-invariant structure, which may be essential for generalization beyond tightly controlled laboratory settings.

4. The magnitude of CMO (0.5–0.8) implies that, after redundancy pruning at matched thresholds, roughly 50–80% of the retained latent channels are shared across modality pairs, suggesting a modality-invariant semantic core. Practically, this supports parameter sharing or a unified decoder across modalities and motivates geometry-aware regularization (or multi-modal training objectives) that explicitly encourages cross-modal alignment to improve robustness when the stimulus modality varies or is partially missing.

4.6. Limitations and Future Directions

The cohort is modest, and questionnaire availability further reduces the effective sample size, limiting correlation precision and richer hierarchical modeling. In addition, public imagination–perception EEG datasets that also provide self-report imagery questionnaires are scarce, which constrains external validation of the trait-linkage analysis; accordingly, we frame the VVIQ/BAIS findings as exploratory and treat replication on an independent cohort/dataset as an important direction for future work.

This study is also positioned as an offline evaluation framework rather than embedded real-time deployment: the encoder is pretrained offline and then frozen, online inference requires only a single forward pass through the encoder plus a lightweight head, and the SSI/CMO diagnostics are post hoc computations on 128-D latents (correlations and greedy pruning over a small τ-grid) that do not add cost to online decoding.

We also fixed the windowing scheme (1 s windows with 50% overlap) to match the semantic extractor configuration; an explicit sensitivity sweep over window length and overlap, as well as testing whether the SSI and CMO remain stable across multiple temporal scales, is an important next step.

Questionnaires were self-reported and were not paired with task-anchored behavioral imagery probes; adding objective measures (e.g., imagery interference tasks, mental rotation, or trial-wise vividness ratings) would sharpen construct validity. Moreover, subject structure in the latent space (Supplementary Figure S2) suggests that future work could explicitly disentangle subject identity factors from semantic factors, potentially improving robustness under cross-session or cross-site transfer. Extending the framework beyond binary imagination vs. perception within each modality to multiclass semantic decoding across concepts and modalities would further test whether the cross-modal latent core supports compositional and clinically relevant semantic generalization.

As a complementary direction to strengthen multiscale semantic structure, future work could incorporate wavelet-based time–frequency attention mechanisms (e.g., as explored in [27]) to explicitly model cross-scale temporal dynamics in EEG while retaining interpretability.

5. Conclusions

This study shows that semantic, representation-level analyses provide information that accuracy cannot: despite strong LOSO performance, VVIQ and BAIS exhibit weak, imprecise associations with accuracy, yet the learned semantic latents expose clear phase-dependent geometry, positive SSI in a subset of participants indicates that imagination can rely on a sparser, higher-salience subset of features, and consistently high CMO points to a cross-modal semantic core. Together, these results reframe evaluation from "did it classify?" to "how does cognition inhabit the latent space?", highlighting compactness and redundancy as meaningful axes for EEG interpretation. We stress two caveats: the questionnaire cohort is small ($n = 13$), limiting power, and self-report vividness is noisy and task-external. Even so, the semantic diagnostics are stable, mechanistic, and practically useful for personalization and model design. Future work should scale n, include task-anchored behavioral

imagery measures, quantify test–retest reliability of the SSI, and integrate geometry-aware regularization so that desirable properties, latent compactness, and cross-modal consistency are not only diagnosed but also learned.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/app16020661/s1: Figure S1. Training/validation loss and learning-rate (LR) schedule for the unsupervised encoder–decoder. Curves show rapid early gains followed by LR-driven refinement; validation tracks training with a small gap, and early stopping fires at epochs 302/350, indicating stable generalization. Figure S2. t-SNE of learned semantic features colored by subject. Latents cluster into compact, separable groups, suggesting a well-organized manifold suitable for downstream decoding and correlation analyses. Figure S3. Raw latent correlation structure. Feature–feature Pearson r matrices for one subject across phases (rows) and modalities (columns). Off-diagonal patterns reveal clusters of correlated latent channels (redundancy) that a simple classifier may or may not exploit. Figure S4. $\Delta K(\tau)$ per subject with mean $\pm$ standard error of the mean (SEM) band, split by modality (audio, orthographic, pictorial). Curves show where imagination–perception differences emerge: several subjects exhibit persistent positive margins at aggressive pruning (large τ), suggesting more compact, less-redundant semantic structure during imagination. Table S1. Comparisons with state-of-the-art methods on pictorial modality; Table S2. Comparisons with state-of-the-art methods on orthographic modality; Table S3. Comparisons with state-of-the-art methods on audio modality; Table S4. VVIQ/BAIS per subject with ranks (1 = best; dense ranking) [26]; Table S5. Correlations between VVIQ/BAIS and accuracies. References [19,28–35] are cited in the supplementary materials.

Author Contributions: Conceptualization, H.A., M.I., and L.M.; methodology, H.A.; software, H.A. and M.I.; validation, H.A.; formal analysis, H.A. and M.I.; investigation, H.A.; resources, L.M.; data curation, H.A. and M.I.; writing—original draft preparation, H.A.; writing—review and editing, L.M.; visualization, H.A.; supervision, L.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets utilized in this study are publicly accessible and can be found in the following reference: [26].

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

Acronym	Definition	Acronym	Definition
AUC	Area Under the Curve	GAP	Global Average Pooling
BAIS	Bucknell Auditory Imagery Scale	ICA	Independent Component Analysis
BAIS-V	BAIS—Vividness subscale	I↔P	Imagination ↔ Perception
BCa	Bias-Corrected and Accelerated (bootstrap)	I→P	Train on Imagination, Test on Perception
BCI	Brain–Computer Interface	LR	Learning Rate
BIDS	Brain Imaging Data Structure	LSL	Lab Streaming Layer
BH	Benjamini–Hochberg (FDR control)	LOSO	Leave-One-Subject-Out
BLUP	Best Linear Unbiased Predictor	MNE	MNE-Python toolkit
CAR	Common Average Reference	MSE	Mean Squared Error
CI	Confidence Interval	P→I	Train on Perception, Test on Imagination
CMO	Cross-Modal Overlap	PyPREP	Python PREP (bad-channel handling)
CNN	Convolutional Neural Network	REML	Restricted Maximum Likelihood
CPz	Midline Central–parietal Electrode	RF	Random Forest

CSP	Common Spatial Pattern	RNN	Recurrent Neural Network
EEG	Electroencephalography	SD	Standard Deviation
EOG	Electrooculography	SEM	Standard Error of the Mean
ERP	Event-Related Potential	SNR	Signal-to-Noise Ratio
FBCSP	Filter Bank Common Spatial Pattern	SSI	Semantic Sensitivity Index
FDR	False Discovery Rate	SVM	Support Vector Machine
FFN	Feed-Forward Network	t-SNE	t-Distributed Stochastic Neighbor Embedding
FIR	Finite Impulse Response	VVIQ	Vividness of Visual Imagery Questionnaire

References

1. Pulvermüller, F.; Shtyrov, Y. Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. *Prog. Neurobiol.* **2006**, *79*, 49–71. [CrossRef]
2. Michel, C.M.; Brunet, D. EEG Source Imaging: A Practical Review of the Analysis Steps. *Front. Neurol.* **2019**, *10*, 446653. [CrossRef]
3. Farah, M.J. Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychol. Rev.* **1988**, *95*, 307–317. [CrossRef]
4. Kosslyn, S.M.; Ganis, G.; Thompson, W.L. Neural foundations of imagery. *Nat. Rev. Neurosci.* **2001**, *2*, 635–642. [CrossRef] [PubMed]
5. Dijkstra, N.; Bosch, S.E.; van Gerven, M.A. Vividness of Visual Imagery Depends on the Neural Overlap with Perception in Visual Areas. *J. Neurosci.* **2017**, *37*, 1367–1373. [CrossRef] [PubMed]
6. Binder, J.R.; Desai, R.H.; Graves, W.W.; Conant, L.L. Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. *Cereb. Cortex* **2009**, *19*, 2767–2796. [CrossRef] [PubMed]
7. Schacter, D.L.; Addis, D.R.; Hassabis, D.; Martin, V.C.; Spreng, R.N.; Szpunar, K.K. The Future of Memory: Remembering, Imagining, and the Brain. *Neuron* **2012**, *76*, 677–694. [CrossRef]
8. Rybář, M.; Daly, I. Neural Decoding of Semantic Concepts: A Systematic Literature Review. *J. Neural Eng.* **2022**, *19*, 021002. [CrossRef]
9. Rekrut, M.; Sharma, M.; Schmitt, M.; Alexandersson, J.; Krüger, A. Decoding Semantic Categories from EEG Activity in Object-Based Decision Tasks. In Proceedings of the 2020 8th International Winter Conference on Brain-Computer Interface (BCI), Gangwon, Republic of Korea, 26–28 February 2020; pp. 1–7. [CrossRef]
10. Lee, K.-W.; Lee, D.-H.; Kim, S.-J.; Lee, S.-W. Decoding Neural Correlation of Language-Specific Imagined Speech using EEG Signals. In Proceedings of the 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, UK, 11–15 July 2022; pp. 1977–1980. [CrossRef]
11. Ahmadi, H.; Mesin, L. Enhancing MI EEG Signal Classification With a Novel Weighted and Stacked Adaptive Integrated Ensemble Model: A Multi-Dataset Approach. *IEEE Access* **2024**, *12*, 103626–103646. [CrossRef]
12. Ahmadi, H.; Costa, P.; Mesin, L. A Novel Hierarchical Binary Classification for Coma Outcome Prediction Using EEG, CNN, and Traditional ML Approaches. *TechRxiv* **2024**. [CrossRef]
13. Ahmadi, H.; Kuhestani, A.; Mesin, L. Adversarial Neural Network Training for Secure and Robust Brain-to-Brain Communication. *IEEE Access* **2024**, *12*, 39450–39469. [CrossRef]
14. Ahmadi, H.; Kuhestani, A.; Keshavarzi, M.; Mesin, L. Securing Brain-to-Brain Communication Channels Using Adversarial Training on SSVEP EEG. *IEEE Access* **2025**, *13*, 14358–14378. [CrossRef]
15. Marks, D.F. Visual imagery differences in the recall of pictures. *Br. J. Psychol.* **1973**, *64*, 17–24. [CrossRef] [PubMed]
16. Shinkareva, S.V.; Malave, V.L.; Mason, R.A.; Mitchell, T.M.; Just, M.A. Commonality of neural representations of words and pictures. *NeuroImage* **2011**, *54*, 2418–2425. [CrossRef] [PubMed]
17. Halpern, A.R. Differences in auditory imagery self-report predict neural and behavioral outcomes. *Psychomusicology Music Mind Brain* **2015**, *25*, 37. Available online: https://api.semanticscholar.org/CorpusID:143273929 (accessed on 12 November 2025). [CrossRef]
18. Rueschemeyer, A. Cross-Modal Integration of Lexical-Semantic Features during Word Processing: Evidence from Oscillatory Dynamics during EEG. *PLoS ONE* **2014**, *9*, e101042. [CrossRef]
19. Huang, J.; Chang, Y.; Li, W.; Tong, J.; Du, S. A Spatio-Temporal Capsule Neural Network with Self-Correlation Routing for EEG Decoding of Semantic Concepts of Imagination and Perception Tasks . *Sensors* **2024**, *24*, 5988. [CrossRef]
20. Chen, H.; He, L.; Liu, Y.; Yang, L. Visual Neural Decoding via Improved Visual-EEG Semantic Consistency. *arXiv* **2024**. Available online: https://arxiv.org/abs/2408.06788 (accessed on 24 October 2025).
21. Ahmadi, H.; Mesin, L. Universal semantic feature extraction from EEG signals: A task-independent framework. *J. Neural. Eng.* **2025**, *22*, 036003. [CrossRef]

22. Fahimi Hnazaee, M.; Khachatryan, E.; Van Hulle, M.M. Semantic Features Reveal Different Networks During Word Processing: An EEG Source Localization Study. *Front. Hum. Neurosci.* **2018**, *12*, 503. [CrossRef]
23. Zeng, H.; Xia, N.; Qian, D.; Hattori, M.; Wang, C.; Kong, W. DM-RE2I: A Framework Based on Diffusion Model for the Reconstruction from EEG to Image. *Biomed. Signal Process. Control* **2023**, *86*, 105125. [CrossRef]
24. Ahmadi, H.; Mesin, L. Decoding Visual Imagination and Perception from EEG via Topomap Sequences. In Proceedings of the 2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Copenhagen, Denmark, 14–17 July 2025; pp. 1–7. [CrossRef]
25. Zeng, H.; Xia, N.; Tao, M.; Pan, D.; Zheng, H.; Wang, C.; Xu, F.; Zakaria, W.; Dai, G. DCAE: A Dual Conditional Autoencoder Framework for the Reconstruction from EEG into Image. *Biomed. Signal Process. Control* **2023**, *81*, 104440. [CrossRef]
26. Wilson, H.; Golbabaee, M.; Proulx, M.J.; Charles, S. EEG-based BCI Dataset of Semantic Concepts for Imagination and Perception Tasks. *Sci. Data* **2023**, *10*, 386. [CrossRef]
27. Wang, F.; Ke, H.; Cai, C. Deep Wavelet Self-Attention Non-negative Tensor Factorization for non-linear analysis and classification of fMRI data. *Appl. Soft Comput.* **2025**, *182*, 113522. [CrossRef]
28. Lawhern, V.J.; Solon, A.J.; Waytowich, N.R.; Gordon, S.M.; Hung, C.P.; Lance, B.J. EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces. *J. Neural Eng.* **2018**, *15*, 056013. [CrossRef]
29. Schirrmeister, R.T.; Springenberg, J.T.; Fiederer, L.D.J.; Glasstetter, M.; Eggensperger, K.; Tangermann, M.; Hutter, F.; Burgard, W.; Ball, T. Deep learning with convolutional neural networks for EEG decoding and visualization. *Hum. Brain Mapp.* **2017**, *38*, 5391–5420. [CrossRef]
30. Sabour, S.; Frosst, N.; Hinton, G.E. Dynamic Routing Between Capsules. *arXiv* **2017**. Available online: https://arxiv.org/abs/1710.09829 (accessed on 26 October 2025).
31. Song, Y.; Jia, X.; Yang, L.; Xie, L. Transformer-based Spatial-Temporal Feature Learning for EEG Decoding. *arXiv* **2021**. Available online: https://arxiv.org/abs/2106.11170 (accessed on 11 November 2025).
32. Liu, Z.; Mao, H.; Wu, C.; Feichtenhofer, C.; Darrell, T.; Xie, S. A ConvNet for the 2020s. *arXiv* **2022**. Available online: https://arxiv.org/abs/2201.03545 (accessed on 26 October 2025).
33. Li, C.; Wang, B.; Zhang, S.; Liu, Y.; Song, R.; Cheng, J.; Chen, X. Emotion recognition from EEG based on multi-task learning with capsule network and attention mechanism. *Comput. Biol. Med.* **2022**, *143*, 105303. [CrossRef] [PubMed]
34. Song, Y.; Zheng, Q.; Liu, B.; Gao, X. EEG Conformer: Convolutional Transformer for EEG Decoding and Visualization. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2023**, *31*, 710–719. [CrossRef]
35. Miao, Z.; Zhao, M.; Zhang, X.; Ming, D. LMDA-Net: A lightweight multi-dimensional attention network for general EEG-based brain-computer interfaces and interpretability. *NeuroImage* **2023**, *276*, 120209. [CrossRef] [PubMed]