FUNCTIONALMAP: LEARNING DATA-DRIVEN COORDINATES FOR CROSS-SUBJECT TRANSFORMER MODELING OF HETEROGENEOUS INTRACRANIAL RECORDINGS

Anonymous authors
Paper under double-blind review

ABSTRACT

Aggregating intracranial recordings across subjects is challenging because electrode count, placement, and regional coverage vary widely. Standard anatomical coordinates such as MNI provide a shared reference frame, but often fail to capture true functional similarity: even at matched anatomical coordinates, the targeted brain region and underlying neural dynamics can differ substantially between individuals. We propose **FunctionalMap**, a scalable representation-learning framework that learns a data-driven functional coordinate system for intracranial neural recordings. An encoder trained with contrastive objectives maps short local field potential segments into a subject-agnostic embedding space, inducing electrode identities that are locality-sensitive to region-specific neural signatures rather than anatomical location. These functional coordinates are then used as tokens for a transformer that models inter-regional relationships across variable channel sets. We evaluate this framework on a 20-subject basal ganglia–thalamic dataset collected during flexible rest/movement recording sessions with heterogeneous electrode layouts. The learned functional space forms clear, region-consistent clusters, supports within-subject discrimination, and transfers zero-shot to unseen channels. In cross-region LFP prediction, a single transformer using functional coordinates outperforms models using MNI coordinates or region labels, without subject-specific heads or fine-tuning. Performance improves with more training subjects; functional coordinates also improve spectral band-power prediction, and the masked-region result replicates with separately trained models on a public human cohort. On public mouse visual-behavior recordings, a trained functional encoder supports transformer modeling of pooled hit-versus-miss behavior prediction in unseen animals (zero-shot) and improves upon anatomical coordinates.

1 INTRODUCTION

Building models that generalize across subjects in neuroscience requires representations that remain stable despite variability in data acquisition. Intracranial neural recordings lack this stability: electrode locations, counts, sampling, and coverage differ across individuals, reflecting clinical needs rather than standardized layouts. Without a shared representational system, cross-subject aggregation is unreliable, limiting scalable modeling and clinical translation.

These difficulties are magnified in emerging invasive multi-region datasets, where electrodes simultaneously capture activity from distinct deep-brain structures(Cimbálník et al., 2025; Parmigiani et al., 2022; Shahabi et al., 2023; Wu et al., 2023; 2024; Sanger et al., 2018; Hernandez-Martin et al., 2023b;a). Such recordings are uniquely valuable for studying inter-regional communication, yet their heterogeneity makes them especially challenging to align. In practice, two obstacles dominate:

Anatomical variability and inconsistent electrode coverage. Electrode placement can vary significantly across subjects (Starr et al., 2006; Schönecker et al., 2015). Standard atlas-based alignment (e.g., MNI (Talairach, 1988)) assumes spatial correspondence implies functional similarity, yet recordings at matched coordinates often capture divergent functional roles(Nestor et al., 2014), and in some cases completely different brain regions (Ashkan et al., 2007; Daniluk et al., 2010).

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.

Multi-area recordings. Modern DBS procedures now routinely sample from several deep brain regions simultaneously (Wang et al., 2024; Liker et al., 2024). These datasets offer unprecedented opportunities to study inter-regional communication, but their heterogeneity magnifies the alignment problem: each region expresses distinct yet interdependent functional signatures.

Attempts to scale datasets, inspired by trends in language modeling (Wiggins & Tejani, 2022; Kaplan et al., 2020), are in many cases limited to multi-session data within a single subject and region (Pandarinath et al., 2018; Farshchian et al., 2018; Karpowicz et al., 2025; Degenhart et al., 2020), or combined, but through post-hoc latent alignment (Safaie et al., 2023; Ma et al., 2023), which precludes unified end-to-end training across multi-session, multi-subject datasets.

Transformer-based models have begun to address this by training foundation models for EEG decoding (Wan et al., 2023; Ding et al., 2025; Cui et al., 2024), but they typically assume fixed, high-density electrode grids and large labeled datasets. Extending these ideas to invasive recordings is harder: spatial coverage is sparse, contact locations vary across patients, and anatomical localization can be imprecise. Consequently, many approaches rely on atlas-based alignment (Mentzelopoulos et al., 2024; Chau et al., 2025), which inherits the limitations of standardized coordinate systems noted above. This motivates our central hypothesis: **Neural signals can be aligned across subjects more reliably by their functional characteristics than by their anatomical coordinates.** Recent work such as BaRISTA (Oganesian et al., 2025) further suggests that the choice of coordinate system for tokenizing intracranial recordings strongly affects transformer performance, with larger anatomical-scale tokenization outperforming channel-coordinate encoding in macro-scale iEEG. Here, we ask whether the coordinate system itself can instead be learned directly from neural dynamics.

We operationalize this hypothesis in *FunctionalMap*, a framework derives coordinate from neural data. A Siamese encoder (Hadsell et al., 2006; Bromley et al., 1993), trained with contrastive objectives learns subject-agnostic embeddings of short neural segments, mapping signals from the same brain region close together in latent space regardless of electrode placement. These embeddings serve as functional identifiers in tokens for a transformer (Vaswani et al., 2017; Dosovitskiy et al., 2020), which models inter-regional dependencies across variable channel sets without subject IDs or subject-specific heads. To our knowledge, we are the first to learn a subject-invariant, region-structured functional coordinate system from invasive, multi-region human LFPs and to use it for end-to-end training on a unified multi-session, multi-subject dataset. **Our contributions: Functional coordinates.** We introduce data-driven embeddings that capture region-specific dynamics, yielding a subject-agnostic functional identity that replaces unreliable atlas-based alignment. **Unified cross-subject modeling.** Embedding-informed tokenization lets a single transformer aggregate heterogeneous intracranial datasets' irregular electrode layouts, multiple regions, without per-subject models. **Empirical validation.** On a 20-subject multi-region intracranial LFP dataset functional embedding was able to cluster regions and transferred zero-shot to held-out channels. When ablating coordinate systems on the same 20-subject dataset, functional embeddings *significantly improve* cross-region prediction compared to MNI-coordinate alignment. Together, these results suggest that data-derived coordinate systems provide the missing backbone for large-scale modeling of invasive human recordings, opening the door to subject-agnostic pretraining, cross-subject decoding, and scalable representation learning in neuroscience. To make this work reproducible, all code used in this study are available at `https://github.com/Functional-Embedding/ICLR2027_Functional_Map`.

2 Related Work

Transformers for neural population modeling Transformer architectures have advanced modeling of neural population activity across spikes and field potentials, including NDT (Ye & Pandarinath, 2021; Ye et al., 2023), STNDT (Le & Shlizerman, 2022), and large-scale decoding frameworks (Azabou et al., 2023; Ding et al., 2025; Cui et al., 2024). These methods excel at spatiotemporal sequence modeling but generally assume stable channel identities/dense coverage or use neuronal level spiking activity, and do not provide mechanisms for aligning heterogeneous electrode sets across subjects. Our method complements these works: we retain transformer scalability while injecting *functional embeddings* that align signals across subjects and regions prior to aggregation.

Neural decoding with SEEG Recent SEEG studies demonstrate decoding for speech and other behaviors from intracranial recordings (Angrick et al., 2021; Meng et al., 2021; Petrosyan et al., 2022; Wu et al., 2024). These systems are often optimized per dataset or subject and rely on task-specific

supervision. In contrast, we aim to first learn subject-agnostic data-driven coordinates without behavioral labels, then plug these coordinates into transformers to scale across subjects and tasks.

Cross-subject SEEG decoding under electrode variability SEEG cohorts exhibit heterogeneous electrode numbers and placements across subjects, complicating multi-subject integration. Wang et al. (2023) proposes a subject and electrode agnostic neural representations using transformers trained on masked spectrogram reconstruction, our goal is not generic contextualization but functional alignment: we learn region-structured embeddings via contrastive objective, addressing the well-known gap between anatomical coordinates and functional identity. Mentzelopoulos et al. (2024) address this by tokenizing per-electrode activity, injecting 3D coordinates, and performing attention across time and electrodes with subject-specific heads. notably, Mentzelopoulos et al. (2024) found no significant performance loss when ablating positional encoding using MNI coordinates. Chau et al. (2025) pretrain a Population Transformer that aggregates arbitrary, spatially sparse channel ensembles via self-supervision, improving downstream decoding and few-shot transfer. In contrast, our approach learns *functional* embeddings directly from neural dynamics and uses them as universal coordinates for cross-subject aggregation, avoiding reliance on atlas coordinates. Most closely related to our tokenization question, BaRISTA (Oganesian et al., 2025) studies how spatial encoding scale affects intracranial transformer models and reports improved decoding using larger anatomical-scale tokenization relative to channel-coordinate encoding in macro-scale iEEG. Our work is complementary: rather than relying on predefined anatomical coordinate systems, we learn a data-driven functional coordinate system directly from LFP dynamics.

3 METHOD

We propose a scalable framework (FunctionalMap) for learning functionally meaningful representations from multi-region human intracranial neural recordings, designed to generalize across subjects and sessions with variable electrode configurations (Figure 1). Our approach first constructs a functional embedding space using contrastive learning to map neural signals from the same brain region closer together in latent space. These learned embeddings are then used to tokenize input data and condition a transformer-based architecture for downstream tasks such as time-series reconstruction. This two-stage framework supports subject-agnostic modeling by capturing data-driven regional identity based on neural dynamics rather than anatomical location (i.e., MNI coordinates).

Data Collection and Preprocessing Our dataset comprises intracranial local field potential (LFP) recordings from 20 subjects with dystonia who underwent deep brain stimulation (DBS) procedures. Stereo-EEG (SEEG) electrodes were implanted in multiple basal ganglia and thalamic nuclei, includ-

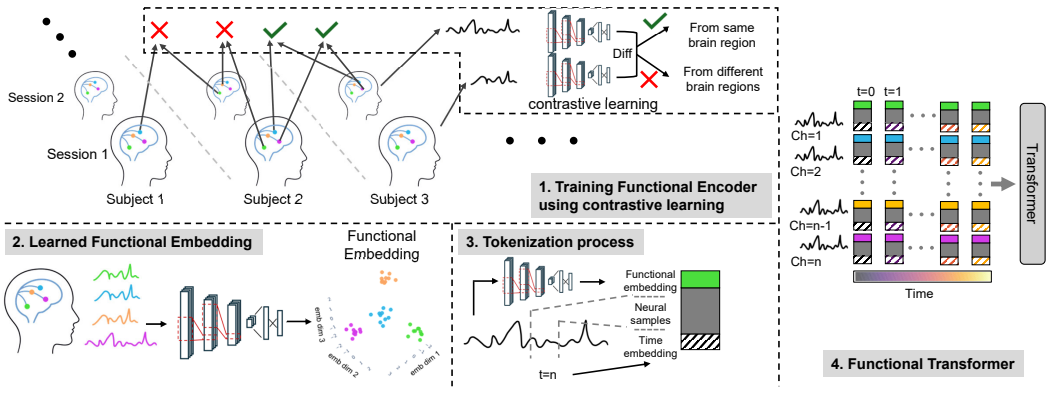


Figure 1: *Overview of Approach.* 1: Contrastive learning framework used to train the CNN-based functional encoder via a Siamese architecture, encouraging signals from the same brain region to map nearby in latent space. 2: Learned functional embeddings produced by the trained encoder, showing clustering by brain region. 3: Incorporating functional embeddings into the tokenization scheme. 4: Functional Transformer architecture that enables scalable modeling across subjects and sessions .

ing the internal segment of the globus pallidus (GPi), the subthalamic nucleus (STN), the ventralis oralis thalamic nucleus (VO), Ventral Anterior thalamic nucleus (VA), ventral intermediate thalamic nucleus (VIM), pedunculopontine nucleus (PPN), and substantia nigra pars reticulata (SNr), with electrode targets selected based on clinical criteria. All analyses use LFPs recorded from high-impedance 50-μm micro-contacts on the SEEG leads, rather than the larger clinical macro-contacts. In total, the dataset includes over 442.86 electrode-hours of neural recordings. Further details on datasets, electrode locations, and preprocessing steps are available in Appendices A.1 and A.2.

3.1 Functional Embedding Network

To address anatomical variability and heterogeneous electrode coverage across subjects, we learn a 32-dimensional (32-D) functional embedding for each recording channel using a Siamese neural network trained with contrastive loss. The objective is to construct a latent space in which signals from the same brain region, regardless of subject or session, are mapped close together (locality-sensitive functional map), thereby enabling data aggregation across sessions and individuals.

The network receives pairs/groups of LFP segments as input and minimizes a contrastive loss function that brings embeddings of positive pairs (same region) closer while pushing negative pairs (different regions) apart. The encoder architecture is a lightweight convolutional neural network (CNN) with batch normalization, GELU activation function (Hendrycks & Gimpel, 2016), and dropout (Srivastava et al., 2014). We implemented two contrastive methods:

Method 1: Pairwise Siamese Contrastive (PSC)
Let $f_\theta : \mathbb{R}^T \to \mathbb{R}^d$ be the encoder that maps a 10 s segment x to an embedding $z = f_\theta(x)$. Given a balanced set of pairs $\{(x_i, x_j, y_{ij})\}$ with $y_{ij} \in \{0, 1\}$ (0 similar: same region; 1 dissimilar: different regions), let z_i and z_j be their embeddings and $d_{ij} = \|z_i - z_j\|_2$. Here, $\mathcal{B}$ denotes the set of index pairs in the current mini-batch. We minimize the standard contrastive loss with margin $m > 0$:

$$\mathcal{L}_{\text{pair}} \;=\; \frac{1}{|\mathcal{B}|} \sum_{(i,j)\in\mathcal{B}} \Big[(1 - y_{ij})\, d_{ij}^2 \;+\; y_{ij} \left(\max(0,\, m - d_{ij})\right)^2\Big].$$

In the case of similar pairs ($y_{ij} = 0$) only the first term in the summation is active making the loss proportional to the distance of embeddings (d_{ij}^2), penalizing long distances in the embedding space from same-source sample pairs. In the case of dissimilar pairs ($y_{ij} = 1$), the second term in the summation is active making the loss function proportional to the closeness of embeddings $\big((m - d_{ij})^2\big)$, if distance is bigger than margin ($m = 0.5$), the loss becomes zero. This encourages dissimilar pair embeddings to be at least m apart(Hadsell et al., 2006).

Method 2: Modified SupCon (MSC)
We utilized a modified Supervised Contrastive approach (SupCon; Khosla et al. (2020)) in a single-view, projection-free setting. This method uses multi-positive InfoNCE (van den Oord et al., 2018) with label-defined positives. Let $f_\theta : \mathbb{R}^T \to \mathbb{R}^d$ be the encoder (same as PSC) producing an embedding $z_i = f_\theta(x_i)$ for segment x_i with region label y_i. For a mini-batch $\mathcal{B} = \{(x_i, y_i)\}_{i=1}^N$, define the positive set for anchor i as $\mathcal{P}(i) = \{\, p \in \{1, \ldots, N\} \setminus \{i\} \;:\; y_p = y_i \,\}$. Using inner-product similarity $s_{ij} = z_i^\top z_j$ and temperature $\tau > 0$ in our case $\tau = 0.2$:

$$\mathcal{L}_{\text{sup}} = \frac{1}{N} \sum_{i=1}^{N} \left(-\frac{1}{|\mathcal{P}(i)|} \sum_{p\in\mathcal{P}(i)} \log \frac{\exp\big(s_{ip}/\tau\big)}{\sum\limits_{a\neq i} \exp\big(s_{ia}/\tau\big)} \right).$$

To further tighten same-region clusters, we add a batchwise intra-class variance term. Let $\mathcal{R}_\mathcal{B}$ be the set of region labels present in the batch and $\mathcal{B}_r = \{i \in \{1, \ldots, N\} : y_i = r\}$, with class mean $\mu_r = \frac{1}{|\mathcal{B}_r|} \sum_{i\in\mathcal{B}_r} z_i$. The variance penalty is defined as:

$$\mathcal{L}_{\text{var}} = \frac{1}{|\mathcal{R}_\mathcal{B}|} \sum_{r\in\mathcal{R}_\mathcal{B}} \frac{1}{|\mathcal{B}_r|} \sum_{i\in\mathcal{B}_r} \|z_i - \mu_r\|_2^2 .$$

Our final objective is

$$\mathcal{L} = \mathcal{L}_{\text{sup}} + \lambda_{\text{var}}\, \mathcal{L}_{\text{var}},$$

with $\lambda_{\text{var}} > 0$ ($\lambda_{\text{var}} = 0.05$ in our case) a tuning coefficient. No projection head and no view augmentations are used.

During training, inputs are sampled from either a single session or across sessions and subjects. In the single session setting, input pairs are not time-synchronized. In the multi-subject and session setting, pairs may originate from different individuals, sessions, and time points. This randomized, region-consistent sampling strategy ensures that the encoder learns region-specific neural signatures that are robust to subject and session variability. **Details.** All model details and training parameters are provided in Appendix A.4, while training resources and times can be found in Appendix A.17.

3.2 Functional Transformer: cross-region LFP prediction

To directly test our hypothesis that functional data-driven embeddings capture circuit-level relationships usable across subjects, we adopt cross-region LFP prediction: withhold all channels from one region and require the model to predict them from the remaining regions. This task is label-free, runs on any multi-region dataset, isolates purely neural-circuit information. For each window, multichannel LFP $X \in \mathbb{R}^{C\times T}$, where C is the number of channels and T is the number of time samples, is split into source channels S and target channels $\mathcal{T}$, and each channel has a 32-D functional embedding extracted once from a 10 s snippet. Because the same waveform feature can have different functional meaning depending on its recording site, each token must carry channel identity in addition to local temporal content.

Tokenization. A 1D conv tokenizer converts each source channel into P time-patch features (width d). The channel's functional embedding is mapped to width d, broadcast across its patches, concatenated with the conv features, reduced back to d, and augmented with a time-only positional code to form *source tokens*. Targets have no signal; we build *query tokens* by fusing a learned per-patch query base with the target's functional embedding and the same positional code. Variable numbers of sources/targets are batched with padding masks. In the pooled comparisons, functional (32-D), MNI (3-D), and one-hot region (4-D) descriptors are projected and fused to the same 128-D Transformer token width (Appendix A.11).

Architecture. A standard pre-LN encoder–decoder Transformer operates on these sequences: the encoder applies self-attention over source tokens to produce a memory; the decoder applies self-attention over queries followed by cross-attention to the memory, then a linear head maps each decoder token to a waveform patch that we concatenate to form the reconstruction. No subject IDs, subject-specific heads, or per-subject fine-tuning are used; a single shared model is trained across all subjects.

Objective. We optimize a mixed loss that combines sample-wise MSE with a correlation term to emphasize waveform shape:

$$\mathcal{L} = \mathrm{MSE}(\widehat{\mathbf{Y}}, \mathbf{Y}) \;+\; \lambda\big(1 - \rho(\widehat{\mathbf{Y}}, \mathbf{Y})\big), \quad \lambda = 0.05,$$

where ρ is mean Pearson correlation across targets. The correlation term counteracts the tendency of pure MSE to favor amplitude shrinkage/flat predictions under scale or SNR mismatch. All token shapes, attention formulas (MHSA/MHA with Q/K/V sources), and masking/batching mechanics, model details and training parameters are provided in Appendix A.5. Training resources and times can be found in Appendix A.17

4 Results

4.1 Functional embeddings recover region-specific neural signatures

To assess whether the functional embedding network captures physiologically meaningful neural signatures, we first performed controlled simulation analyses with known ground-truth region signatures, then evaluated whether the same idea holds in single-subject human recordings.

Simulation-based validation with ground-truth neural signatures. We generated a simulated dataset in which each brain region was assigned a parameterized neural signature, primarily defined in the spectral domain, and recordings were synthesized across 10 subjects and 2 sessions (additional details in Appendix A.3). Subject/session variability was introduced by scaling, frequency shifts, and noise (Fig. A.2). Training on a subset of subjects and evaluating on held-out subjects, the functional embedding network trained with contrastive learning (PSC) correctly clustered recordings by region with high accuracy (Fig. 2A). This demonstrates that the embedding framework can extract stable, region-specific neural signatures even in the presence of cross-subject variability.

Perturbation-based attribution. We next probed which input features contributed most strongly to embedding location. Short temporal segments of the input were selectively frozen while other parts of

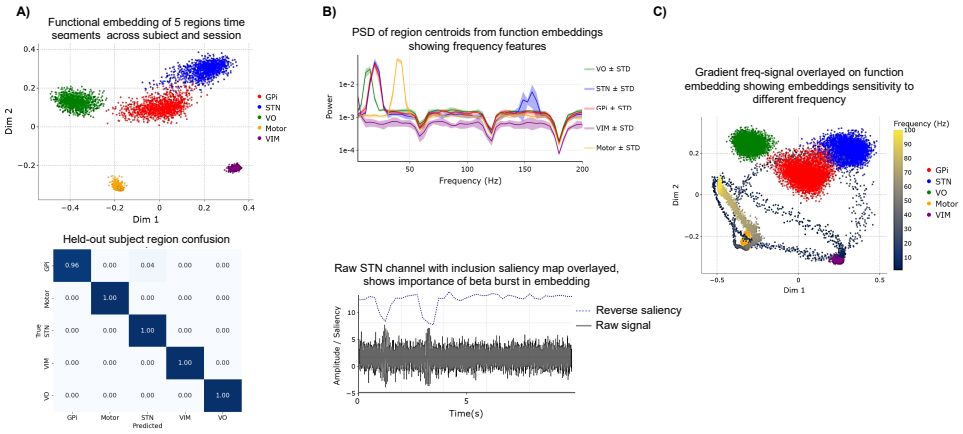


Figure 2: *Simulation-based validation of the functional embedding.* **(A)** Contrastive encoder trained on simulated data clusters test segments by region and generalizes to held-out subjects (top) with high accuracy (bottom). **(B)** Top: Power spectra of signals nearest each region centroid match the ground-truth signatures; Bottom: Perturbation saliency shows largest embedding shifts when disrupting signature bursts. **(C)** Frequency sweeps trace smooth trajectories through the embedding space, indicating a locality-sensitive functional map.

the signal were perturbed, and the resulting displacement in embedding space was quantified. Freezing segments overlapping the bursts that defined the region's neural signature produced the smallest embedding shifts (Fig. 2B bottom), confirming that the embeddings are sensitive to physiologically relevant signal components.

Spectral properties and smoothness of the embedding space. To further link the embedding space to interpretable features, we computed centroids for each region cluster and examined the power spectra of signals nearest to each centroid. These spectra matched the corresponding ground-truth region signatures (Fig. 2B top), showing that embedding clusters reflect distinct frequency-domain patterns. Finally, presenting pure sinusoidal inputs with gradually increasing frequency produced smooth trajectories through the embedding space, passing near centroids of regions tuned to corresponding frequency bands (Fig. 2C). This demonstrates that the functional embedding defines a continuous, locality-sensitive map where novel signals can be meaningfully placed relative to known regions.

Single-subject validation in human recordings. Across 20 subjects (most with multiple sessions), per-subject functional encoders trained on 10 s segments using PSC achieved a mean test accuracy of $75.78\% \pm 17.90\%$ (M±SD) on held-out time segments from electrodes seen during training (Fig. 3A, top). Under the stricter held-out channel evaluation (regions with > 3 channels), accuracy remained above chance at $45.79\% \pm 18.44\%$, with some subjects reaching $\sim 70\%$ (Fig. 3A, bottom). For an example subject (S4), the 32-D embeddings projected via PCA exhibit clear region-specific separation for held-out time segments, reflecting a locality-preserving organization of the learned space (Fig. 3B, bottom). Applying k-NN on these embeddings yields confusion matrices that are strongly diagonal for held-out time segments and weaker, but still structured, for held-out channels (Fig. 3B, top). The corresponding per-subject visualizations and confusion matrices are provided in Appendix A.6.

Overall, these results show that functional embeddings capture region-sensitive neural structure in both controlled simulations and real human recordings. Channel-level generalization is lower than time-segment generalization, likely due to the limited number of electrodes per region within a subject ($\sim$ 6–12), yet remains above chance on average.

4.2 Pooled functional embeddings form a subject-agnostic coordinate system

Functional encoder successfully learns combined dataset. When the functional embedding was trained jointly on the entire dataset-sampling similar/dissimilar pairs across subjects and sessions (using PSC), it learned a global, subject-independent mapping with region-specific neural signatures. On unseen time segments, it achieved $80.71\% \pm 11.41\%$ (M±SD) accuracy. Same network generalized zero-shot to held-out channels, yielding $49.18\% \pm 12.11\%$ accuracy (above per-subject

chance). Compared to training separate CNNs per subject, the pooled encoder improved held-out time and held-out channel accuracy by approximately 5% on average, suggesting that cross-subject training stabilizes functional coordinates; the full head-to-head comparison is shown in Appendix Fig. A.5.

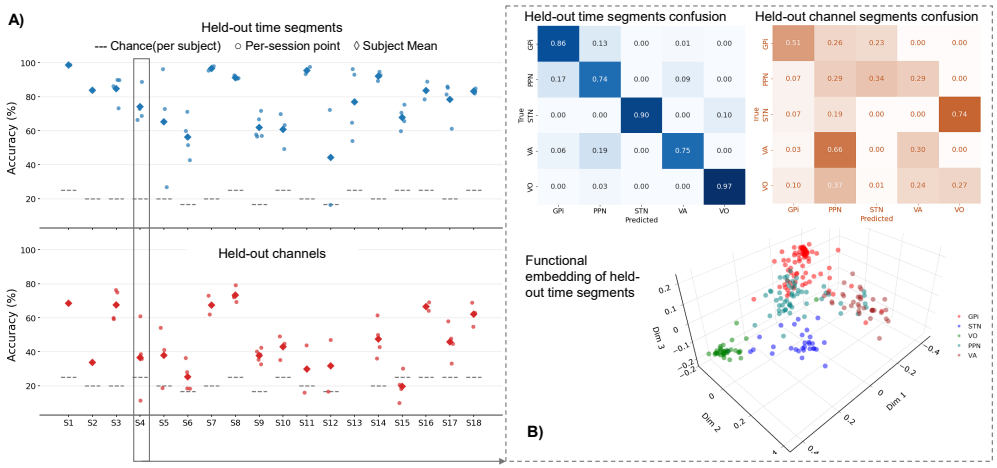


Figure 3: *Single-subject functional embeddings.* **(A)** Per-subject accuracy for held-out time segments and held-out channels. Diamonds denote subject means, circles denote session-level points, and dashed lines indicate chance. **(B)** Example subject (S4): k-NN confusion matrices and 3D PCA projection of the 32-D embedding space, showing region-sensitive structure.

Contrastive objective comparison. We compared PSC and MSC under the same multi-subject regime. Both objectives produced region-structured embedding spaces (Fig. 4), but MSC yielded stronger held-out-channel generalization, indicating better robustness to unseen electrodes. Because the transformer must query held-out target channels, we used MSC functional coordinates for all transformer experiments. Full results and per-subject visualizations appear in Appendices A.9 and A.6. Across subjects, functional and MNI distances showed moderate correspondence across nuclei but little consistent correspondence within individual nuclei (Appendix A.10).

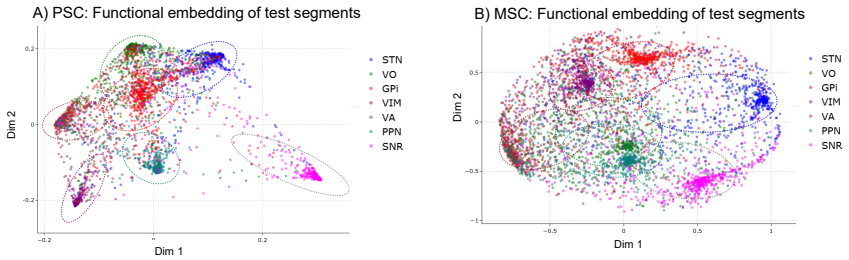


Figure 4: *Comparison of contrastive methods (PSC vs. MSC)* 2D PCA view of test-segment embeddings for PSC (A) and MSC (B) methods.

4.3 Functional embeddings enable aggregation of data over subjects using transformers

We performed an ablation study on channel coordinate systems for cross-region LFP prediction, comparing otherwise-identical transformers using region labels (one-hot encoding), standard MNI coordinates (3-d vector), or functional coordinates (MSC) for both source tokenization and target channel queries (Fig. 5A). Region labels and MNI coordinates provide controls for whether functional embeddings capture information beyond coarse anatomical identity or electrode location. In each evaluation window, all channels from a target region (here, VO) were withheld, while signals from the remaining regions (i.e., GPi, STN, and VIM) were provided as source tokens. Performance was

quantified using the Pearson correlation (r) between predicted and true waveforms, aggregated per subject using Fisher-z averaging.

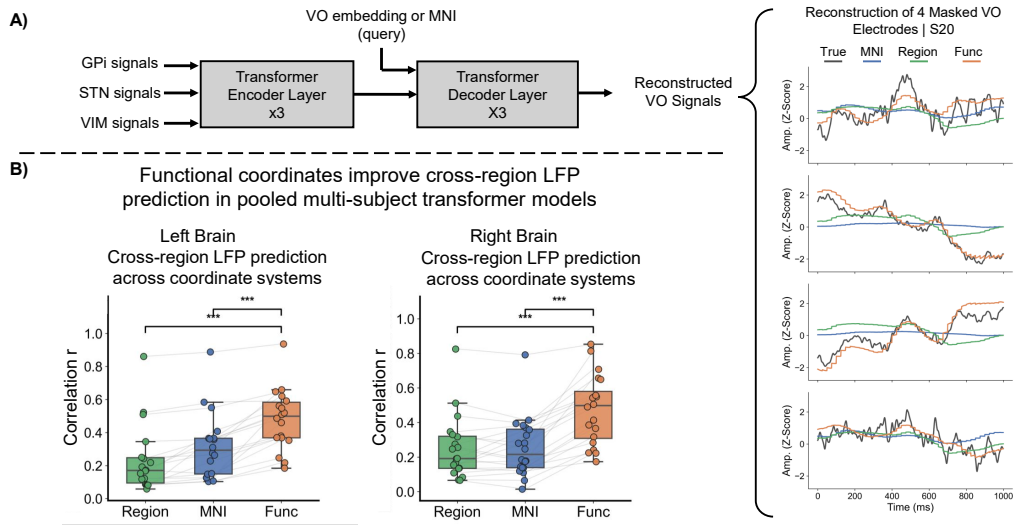


Figure 5: *Functional coordinates improve cross-region LFP prediction across pooled subjects and sessions.* **(A)** Functional Transformer schematic and example masked-VO prediction using MNI, region-label, or functional coordinates. **(B)** Subject-level prediction correlations show consistent gains for functional coordinates over MNI and region labels in both hemispheres. All models use a single pooled transformer without subject IDs or fine-tuning. **p < 0.01, ***p < 0.001.

Across subjects, functional coordinates significantly improved cross-region LFP prediction accuracy relative to both anatomical coordinate systems for both hemispheres (Fig. 5B). Across-subject box plots and paired subject comparisons show consistent upward shifts from region/MNI coordinates to functional coordinates across the 20-subject cohort, with functional coordinates significantly outperforming both controls ($p < 0.0001$). Importantly, all results were obtained using a single shared transformer model across all subjects, without subject-ID tokens, subject-specific heads, or fine-tuning. Detailed subject-wise prediction results and session-level variability are provided in Appendix Fig. A.6. A stricter capacity-matched comparison also gave MNI and region descriptors nonlinear learned 32-D representations with slightly more trainable parameters; the functional-coordinate advantage persisted (Appendix A.11). Qualitatively, functional coordinates also produced more differentiated reconstructions across simultaneously recorded VO channels (Fig. 5B). Because VO electrodes often shared similar coordinates, region- and MNI-based models generated similar target waveforms, whereas functional coordinates produced more channel-specific predictions. In a 14-subject publicly available movie-viewing dataset (Keles et al., 2024b), separately trained models for the same masked-region task preserved the functional-coordinate advantage (Appendix A.12).

4.4 Functional transformers outperform subject-specific baselines and scale with dataset size

We next asked whether the pooled Functional Transformer provides benefits beyond standard subject-specific time-series predictors. We compared against linear and nonlinear baselines trained separately within each subject, including a causal multi-output FIR model, TCN, GRU, CopyBest correlation baseline, and a single-subject Functional Transformer. Further details on baseline architectures, training, and statistics are provided in Appendix A.14.

Table 1: **Cross-region LFP prediction baselines.** Values report mean ± SEM subject-level Fisher-mean prediction correlation r ($n = 18$ left, $n = 20$ right). Pooled-FUNC significantly outperforms all subject-specific baselines after Holm and BH/FDR correction ($p_{\text{corr}} < 0.001$).

Hemisphere	CopyBest	FIR	TCN	GRU	Single-FUNC	Pooled-FUNC
Left	0.119 ± 0.055	0.235 ± 0.057	0.250 ± 0.057	0.239 ± 0.060	0.266 ± 0.055	$\mathbf{0.481 \pm 0.043}$
Right	0.117 ± 0.048	0.257 ± 0.048	0.279 ± 0.048	0.255 ± 0.050	0.290 ± 0.053	$\mathbf{0.475 \pm 0.045}$

The pooled Functional Transformer achieved the highest reconstruction accuracy across both hemispheres, indicating that functional coordinates allow a single shared model to aggregate heterogeneous cross-subject recordings and learn stronger inter-regional relationships than models trained separately within each subject.

To test whether performance improves with increasing cross-subject data, we trained Functional Transformer models using progressively larger subject subsets while keeping the functional encoder and test set fixed.

Performance increased with dataset size, with strong initial gains at small cohort sizes and continued improvement up to the full 20-subject dataset (Fig. 6). These gains reflect the benefit of scaling pooled transformer training rather than changes in the coordinate system. Additional analyses show that functional embeddings remain effective for subjects excluded from embedding training, indicating that the learned coordinate system can generalize to new subjects rather than requiring retraining for every individual (Appendix A.15). Functional coordinates also improve spectral band-power prediction over anatomical coordinates, suggesting that the learned coordinate system benefits not only waveform reconstruction but also clinically interpretable neural features (Appendix A.16).

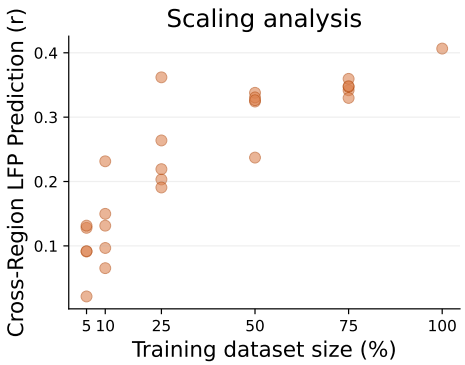


Figure 6: **Scaling analysis.** Cross-region LFP prediction improves as the number of training subjects increases.

Cross-species behavioral evaluation. On public Allen Visual Behavior Neuropixels LFPs (Bennett et al., 2026; Bennett, 2024), an MSC encoder trained on eight mice identified six visual areas in two unseen mice (192 channels; balanced accuracy 0.609 versus six-class chance 0.167). With that encoder frozen, a separately trained pooled transformer predicted hit versus miss without test-mouse fine-tuning in two further unseen mice (three eligible sessions, 724 trials): mouse-mean AUROC was 0.725 with functional coordinates versus 0.712 with CCF coordinates and 0.685 with region labels (Appendix A.13).

5 Discussion

FunctionalMap learns region-structured, subject-agnostic coordinates from LFP dynamics. In our 20-subject clinical cohort, one pooled transformer using these coordinates improves cross-region waveform prediction over MNI and region-label controls across heterogeneous electrode sets. Capacity-matched comparisons, spectral band-power prediction, and increasing-subject experiments support this result. A separate public 14-person movie-viewing cohort reproduces the masked-region prediction advantage with new models trained for its different regions (Appendix A.12); human-trained weights were not transferred between datasets.

The Allen mouse experiment tests a different species and downstream task. A frozen, mouse-trained MSC encoder classified six visual areas in unseen animals; a separately trained pooled decoder then predicted hit versus miss on two further unseen mice without tuning to their behavioral labels or fitting mouse- or session-specific parameters (Appendix A.13). Functional coordinates gave the highest mouse-mean AUROC among equal-capacity decoders using functional, CCF, or region inputs. This comparison is descriptive because it involves two final mice. The Functional decoder uses neither CCF coordinates nor region labels as inputs, although atlas annotations selected the six-area channel set and early unlabeled recordings supplied each new electrode's coordinate and amplitude scale.

This learned channel identity complements coordinate-conditioned decoding (Mentzelopoulos et al., 2024), population-level aggregation (Chau et al., 2025), and work on anatomical tokenization scale (Oganesian et al., 2025): our coordinates can condition a pooled predictor while preserving channel-specific dynamics, even when neighboring micro-contacts share similar anatomical positions.

Limitations and future work. The encoder uses region labels during contrastive training; the mouse behavior result rests on two held-out animals; and fixed-encoder or predictor transfer between datasets remains untested. Self- or weakly supervised coordinates, behavioral evaluation in larger cohorts, forecasting, and imputation remain future directions.

AI USE STATEMENT

We used ChatGPT as a general-purpose assistant during the preparation of this work. The model was employed to help with editing for clarity, conciseness, and formatting. All scientific ideas, analyses, experiments, and interpretations presented in this paper were conceived, implemented, and validated by the authors. The authors take full responsibility for the content of this manuscript.

ETHICS STATEMENT

All human neural data analyzed in this study were collected under protocols approved by the Institutional Review Boards (IRBs) of the participating medical centers. Written informed consent was obtained from all participants or their legal guardians, with child assent obtained when appropriate. Data handling complied with HIPAA and institutional privacy policies: recordings were de-identified prior to analysis, stored on secure access-controlled systems, and used solely for research purposes. This work involved retrospective analysis of clinical electrophysiological recordings and did not involve new interventions, clinical decision-making, or attempts to re-identify participants. Due to ethical and regulatory constraints, raw neural recordings cannot be publicly released; de-identified data may be shared upon reasonable request and execution of a data use agreement consistent with IRB approvals. The authors affirm compliance with the ICLR Code of Ethics and disclose no conflicts of interest or external sponsorship that could influence the work.

REPRODUCIBILITY STATEMENT

All code used in this study is released to support reproducibility at `https://github.com/Functional-Embedding/ICLR2027_Functional_Map`. Complete and detailed explanation of datasets used, data preprocessing, training details and training resources and are provided in Appendix.

REFERENCES

Miguel Angrick, Maarten C Ottenhoff, Lorenz Diener, Darius Ivucic, Gabriel Ivucic, Sophocles Goulis, Jeremy Saal, Albert J Colon, Louis Wagner, Dean J Krusienski, et al. Real-time synthesis of imagined speech processes from minimally invasive recordings of neural activity. *Communications biology*, 4(1):1055, 2021.

Anonymous. details omitted for double blind review. 2003.

Anonymous. details omitted for double blind review. 2019.

Anonymous. details omitted for double blind review. 2024.

K Ashkan, P Blomstedt, L Zrinzo, S Tisch, T Yousry, P Limousin-Dowsey, and MI Hariz. Variability of the subthalamic nucleus: the case for direct mri guided targeting. *British journal of neurosurgery*, 21(2):197–200, 2007.

Brian B Avants, Charles L Epstein, Murray Grossman, and James C Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41, 2008.

Mehdi Azabou, Vinam Arora, Venkataramana Ganesh, Ximeng Mao, Santosh Nachimuthu, Michael Mendelson, Blake Richards, Matthew Perich, Guillaume Lajoie, and Eva Dyer. A unified, scalable framework for neural population decoding. *Advances in Neural Information Processing Systems*, 36:44937–44956, 2023.

Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. In *International Conference on Learning Representations (ICLR)*, 2018. URL `https://arxiv.org/abs/1803.01271`.

F Bartel, Martin Visser, Michiel de Ruiter, José Belderbos, F Barkhof, H Vrenken, JC de Munck, Marcel van Herk, Alzheimer's Disease Neuroimaging Initiative, et al. Non-linear registration improves statistical power to detect hippocampal atrophy in aging and dementia. *NeuroImage: Clinical*, 23:101902, 2019.

Corbett Bennett. Allen institute - visual behavior - neuropixels, 2024. URL https://doi.org/10.48324/dandi.000713/0.240702.1725.

Corbett Bennett, Samuel D. Gale, Greggory Heller, et al. Map of spiking activity underlying change detection in the mouse visual system. *Cell*, 189(15):4775–4791.e13, 2026. doi: 10.1016/j.cell.2026.06.025.

Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. Signature verification using a" siamese" time delay neural network. *Advances in neural information processing systems*, 6, 1993.

Geeling Chau, Christopher Wang, Sabera Talukder, Vighnesh Subramaniam, Saraswati Soedarmadji, Yisong Yue, Boris Katz, and Andrei Barbu. Population transformer: Learning population-level representations of neural activity. *ArXiv*, pp. arXiv–2406, 2025.

Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

Jan Cimbálník, JS Garcia-Salinas, N Hamedi, S Prathapagiri, LF Sarmiento Rivera, M Galanina, J Dolezal, L Jurkovicova, Pavel Daniel, Martin Kojan, et al. Human brain local field potential recordings during a battery of multilingual cognitive and eye-tracking tasks. *Scientific Data*, 12(1): 899, 2025.

Wenhui Cui, Woojae Jeong, Philipp Thölke, Takfarinas Medani, Karim Jerbi, Anand A Joshi, and Richard M Leahy. Neuro-gpt: Towards a foundation model for eeg. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5. IEEE, 2024.

Slawomir Daniluk, Keith G. Davies, Samuel A Ellias, Peter Novak, and Jules M Nazzaro. Assessment of the variability in the anatomical position and size of the subthalamic nucleus among patients with advanced parkinson's disease using magnetic resonance imaging. *Acta neurochirurgica*, 152 (2):201–210, 2010.

Alan D Degenhart, William E Bishop, Emily R Oby, Elizabeth C Tyler-Kabara, Steven M Chase, Aaron P Batista, and Byron M Yu. Stabilization of a brain–computer interface via the alignment of low-dimensional spaces of neural activity. *Nature biomedical engineering*, 4(7):672–685, 2020.

Yi Ding, Chengxuan Tong, Shuailei Zhang, Muyun Jiang, Yong Li, Kevin JunLiang Lim, and Cuntai Guan. Emt: A novel transformer for generalized cross-subject eeg emotion recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 2025.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Ali Farshchian, Juan A Gallego, Joseph P Cohen, Yoshua Bengio, Lee E Miller, and Sara A Solla. Adversarial domain adaptation for stable brain-machine interfaces. *arXiv preprint arXiv:1810.00045*, 2018.

Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, volume 2, pp. 1735–1742. IEEE, 2006.

Brendan Harris and Pulin Gong. Nested spatiotemporal theta–gamma waves organize hierarchical processing across the mouse visual cortex. *Nature Communications*, 17:2629, 2026. doi: 10.1038/s41467-026-68893-4.

Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016.

Estefania Hernandez-Martin, Maral Kasiri, Sumiko Abe, Jennifer MacLean, Joffre Olaya, Mark Liker, Jason Chu, and Terence D Sanger. Globus pallidus internus activity increases during voluntary movement in children with dystonia. *Iscience*, 26(7), 2023a.

Estefania Hernandez-Martin, Jessica SL Vidmark, Jennifer A MacLean, and Terence D Sanger. What is the effect of benzodiazepines on deep brain activity? a study in pediatric patients with dystonia. *Frontiers in Neurology*, 14:1215572, 2023b.

Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*, 2020.

Brianna M Karpowicz, Yahia H Ali, Lahiru N Wimalasena, Andrew R Sedler, Mohammad Reza Keshtkaran, Kevin Bodkin, Xuan Ma, Daniel B Rubin, Ziv M Williams, Sydney S Cash, et al. Stabilizing brain-computer interfaces through alignment of latent dynamics. *Nature Communications*, 16(1):4662, 2025.

Umit Keles, Julien Dubois, Kevin J. M. Le, J. Michael Tyszka, David A. Kahn, Chrystal M. Reed, Jeffrey M. Chung, Adam N. Mamelak, Ralph Adolphs, and Ueli Rutishauser. Data for: Multimodal single-neuron, intracranial eeg, and fmri brain responses during movie watching in human patients, 2024a. URL https://dandiarchive.org/dandiset/000623/0.240130.2108. DANDI:DANDI:000623/0.240130.2108.

Umit Keles, Julien Dubois, Kevin JM Le, J Michael Tyszka, David A Kahn, Chrystal M Reed, Jeffrey M Chung, Adam N Mamelak, Ralph Adolphs, and Ueli Rutishauser. Multimodal single-neuron, intracranial eeg, and fmri brain responses during movie watching in human patients. *Scientific data*, 11(1):214, 2024b.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673, 2020.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015. URL https://arxiv.org/abs/1412.6980. arXiv:1412.6980.

Trung Le and Eli Shlizerman. Stndt: Modeling neural population activity with spatiotemporal transformers. *Advances in Neural Information Processing Systems*, 35:17926–17939, 2022.

Mark A Liker, Terence D Sanger, Jennifer A MacLean, Jaya Nataraj, Enrique Arguelles, Mark Krieger, Aaron Robison, and Joffre Olaya. Stereotactic awake basal ganglia electrophysiological recording and stimulation (sabers): A novel staged procedure for personalized targeting of deep brain stimulation in pediatric movement and neuropsychiatric disorders. *Journal of Child Neurology*, pp. 08830738231224057, 2024.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR)*, 2019. URL https://arxiv.org/abs/1711.05101. arXiv:1711.05101.

Xuan Ma, Fabio Rizzoglio, Kevin L Bodkin, Eric Perreault, Lee E Miller, and Ann Kennedy. Using adversarial networks to extend brain computer interface decoding accuracy over time. *elife*, 12: e84296, 2023.

Kevin Meng, David B Grayden, Mark J Cook, Simon Vogrin, and Farhad Goodarzy. Identification of discriminative features for decoding overt and imagined speech using stereotactic electroencephalography. In *2021 9th International Winter Conference on Brain-Computer Interface (BCI)*, pp. 1–6. IEEE, 2021.

Georgios Mentzelopoulos, Evangelos Chatzipantazis, Ashwin G Ramayya, Michelle J Hedlund, Vivek P Buch, Kostas Daniilidis, Konrad P Kording, and Flavia Vitale. Neural decoding from stereotactic eeg: accounting for electrode variability across subjects. *Advances in Neural Information Processing Systems*, 37:108600–108624, 2024.

Kelsey A Nestor, Jacob D Jones, Christopher R Butson, Takashi Morishita, Charles E Jacobson IV, David A Peace, Dennis Chen, Kelly D Foote, and Michael S Okun. Coordinate-based lead location does not predict parkinson's disease deep brain stimulation outcome. *PloS one*, 9(4):e93524, 2014.

Lucine L Oganesian, Saba Hashemi, and Maryam M Shanechi. Barista: Brain scale informed spatiotemporal representation of human intracranial neural activity. *arXiv preprint arXiv:2512.12135*, 2025.

Chethan Pandarinath, Daniel J O'Shea, Jasmine Collins, Rafal Jozefowicz, Sergey D Stavisky, Jonathan C Kao, Eric M Trautmann, Matthew T Kaufman, Stephen I Ryu, Leigh R Hochberg, et al. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature methods*, 15(10):805–815, 2018.

S Parmigiani, E Mikulan, S Russo, S Sarasso, FM Zauli, A Rubino, A Cattani, M Fecchio, D Giampiccolo, J Lanzone, et al. Simultaneous stereo-eeg and high-density scalp eeg recordings to study the effects of intracerebral stimulation parameters. *Brain Stimulation*, 15(3):664–675, 2022.

Artur Petrosyan, Alexey Voskoboinikov, Dmitrii Sukhinin, Anna Makarova, Anastasia Skalnaya, Nastasia Arkhipova, Mikhail Sinkin, and Alexei Ossadtchi. Speech decoding from a small set of spatially segregated minimally invasive intracranial eeg electrodes with a compact and interpretable neural network. *Journal of Neural Engineering*, 19(6):066016, 2022.

Mostafa Safaie, Joanna C Chang, Junchol Park, Lee E Miller, Joshua T Dudman, Matthew G Perich, and Juan A Gallego. Preserved neural dynamics across animals performing similar behaviour. *Nature*, 623(7988):765–771, 2023.

Terence D Sanger, Aaron Robison, Enrique Arguelles, Diana Ferman, and Mark Liker. Case report: targeting for deep brain stimulation surgery using chronic recording and stimulation in an inpatient neuromodulation monitoring unit, with implantation of electrodes in gpi and vim in a 7-year-old child with progressive generalized dystonia. *Journal of child neurology*, 33(12):776–783, 2018.

Thomas Schönecker, Doreen Gruber, Anatol Kivi, Bianca Müller, Elmar Lobsien, Gerd-Helge Schneider, Andrea A Kühn, Karl-Titus Hoffmann, and Andreas R Kupsch. Postoperative mri localisation of electrodes and clinical efficacy of pallidal deep brain stimulation in cervical dystonia. *Journal of Neurology, Neurosurgery & Psychiatry*, 86(8):833–839, 2015.

Hossein Shahabi, Dileep R Nair, and Richard M Leahy. Multilayer brain networks can identify the epileptogenic zone and seizure dynamics. *elife*, 12:e68531, 2023.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

Philip A Starr, Robert S Turner, Geoff Rau, Nadja Lindsey, Susan Heath, Monica Volz, Jill L Ostrem, and William J Marks. Microelectrode-guided implantation of deep brain stimulators into the globus pallidus internus for dystonia: techniques, electrode locations, and outcomes. *Journal of neurosurgery*, 104(4):488–501, 2006.

PT Jean Talairach. Co-planar stereotaxic atlas of the human brain. *(No Title)*, 1988.

Nicholas J Tustison, Andrew J Holbrook, Brian B Avants, Jared M Roberts, Philip A Cook, Zachariah M Reagh, Jeffrey T Duda, James R Stone, Daniel L Gillen, Michael A Yassa, et al. Longitudinal mapping of cortical thickness measurements: An alzheimer's disease neuroimaging initiative-based evaluation study. *Journal of Alzheimer's Disease*, 71(1):165–183, 2019.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Zhijiang Wan, Manyu Li, Shichang Liu, Jiajin Huang, Hai Tan, and Wenfeng Duan. Eegformer: A transformer–based brain activity classification method using eeg signal. *Frontiers in neuroscience*, 17:1148855, 2023.

Christopher Wang, Vighnesh Subramaniam, Adam Uri Yaari, Gabriel Kreiman, Boris Katz, Ignacio Cases, and Andrei Barbu. Brainbert: Self-supervised representation learning for intracranial recordings. *arXiv preprint arXiv:2302.14367*, 2023.

Christopher Wang, Adam Yaari, Aaditya K Singh, Vighnesh Subramaniam, Dana Rosenfarb, Jan DeWitt, Pranav Misra, Joseph R Madsen, Scellig Stone, Gabriel Kreiman, et al. Brain treebank: Large-scale intracranial recordings from naturalistic language stimuli. *Advances in Neural Information Processing Systems*, 37:96505–96540, 2024.

Walter F Wiggins and Ali S Tejani. On the opportunities and risks of foundation models for natural language processing in radiology. *Radiology: Artificial Intelligence*, 4(4):e220119, 2022.

Xiaolong Wu, Shize Jiang, Guangye Li, Shengjie Liu, Benjamin Metcalfe, Liang Chen, and Dingguo Zhang. Deep learning with convolutional neural networks for motor brain-computer interfaces based on stereo-electroencephalography (seeg). *IEEE journal of biomedical and health informatics*, 27(5):2387–2398, 2023.

Xiaolong Wu, Guangye Li, Xin Gao, Benjamin Metcalfe, and Dingguo Zhang. Channel selection for stereo-electroencephalography (seeg)-based invasive brain-computer interfaces using deep learning methods. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 32:800–811, 2024.

Joel Ye and Chethan Pandarinath. Representation learning for neural population activity with neural data transformers. *arXiv preprint arXiv:2108.01210*, 2021.

Joel Ye, Jennifer Collinger, Leila Wehbe, and Robert Gaunt. Neural data transformer 2: multi-context pretraining for neural spiking activity. biorxiv, 2023.

A Appendix

A.1 Additional Details on Datasets: Electrophysiological Dataset and Clinical Metadata

Subjects. Twenty subjects who underwent DBS surgery for treatment of dystonia participated in this study (Table A.1). Dystonia diagnosis was established by a pediatric movement disorder specialist (Anonymous) using standard criteria (Anonymous, 2003). Subjects were recruited from participating hospitals anonymized for review. All participants, or legal guardians for minors, provided consent for research use of electrophysiological recordings under IRB-approved protocols.

Table A.1: *Demographic characteristics:* 20 male subjects, between 5 and 24 years old. H-ABC: Hypomyelination with Atrophy of the Basal Ganglia and Cerebellum; CP: Cerebral palsy; GA1: Glutaric aciduria type 1; MEPAN: Mitochondrial Enoyl CoA Reductase Protein-Associated Neurodegeneration; KMT2B: K-specific Methyltransferase 2B; PKAN: Pantothenate Kinase-Associated Neurodegeneration; MYH2: Myosin Heavy Chain 2; GPi: globus pallidus internus; VoaVop: ventral oralis anterior-posterior; STN: subthalamic nucleus; VA: ventral anterior; VPL: ventral posterolateral; VIM: ventral intermediate; PPN: Pedunculopontine nucleus; SNR: Substantia nigra pars reticulata; CM: Centromedian nucleus; CL: Centrolateral nucleus.

Subject	Etiology	Implanted Regions
1	Tourette	NA, VoaVop, VIM, STN, CMP, GPi
2	CP	VA, VoaVop, VIM, STN, GPi
3	GA1	VA, VoaVop, VIM, STN, GPi
4	MEPAN	VA, VoaVop, VIM, PPN, STN, GPi
5	MEPAN	VA, VoaVop, VIM, PPN, STN, GPi
6	Unknown	VA, VoaVop, PPN, STN, GPi
7	KMT2B	VA, VoaVop, VIM, PPN, STN, GPi
8	CP	VA, VoaVop, VIM, PPN, STN, GPi
9	GA1	VA, VoaVop,VIM, PPN, STN, GPi
10	PKAN	VoaVop, VIM, PPN, SNR, STN, GPi
11	CP	VA, VoaVop, VIM, PPN, SNR, STN, GPi
12	MYH2	VoaVop, VIM, PPN, SNR, STN, GPi
13	CP	VoaVop, VIM, PPN, SNR, STN, GPi
14	GA1	VoaVop, VIM, PPN, SNR, STN, GPi
15	CP	VoaVop, VIM, STN, GPi
16	CP	VA, VoaVop, VIM, PPN, STN, GPi
17	KMT2B	VoaVop, VIM, CM, STN, GPi
18	CP	VoaVop, VIM, CM, STN, GPi
19	CP	VoaVop, VIM, CM, SNR, STN, GPi
20	KMT2B	VoaVop, VIM, CM, CL, STN, GPi

Experimental protocol and recording sessions. Participants performed self-paced finger-to-nose reaching or hand-squeeze movements, depending on motor ability. The clinical team administered short blocks consisting of $\sim$60 s of active movement followed by $\sim$30 s of rest, repeated at least four times per hand. Trajectory and speed were unconstrained, and block durations were adjusted as needed to accommodate each participant. For all analyses, recordings were not segmented by trials or active/rest windows; instead, we used the entire continuous LFP recordings from each session. This yields behaviorally diverse, non-stationary signals, underscoring the advantage of our representation-learning approach, which remains robust to noise and variability in naturalistic clinical settings.

Recording electrodes and spatial coverage. Local field potentials (LFPs) were recorded from high-impedance microwire ("micro-contact") electrodes on temporary Stereo-EEG (SEEG) depth leads (AdTech MM16C; AdTech Medical Instrument Corp., Oak Creek, WI, USA) implanted with standard stereotactic procedures Anonymous (2024; 2019). Up to 12 SEEG leads were placed per subject to sample candidate DBS targets. Each lead includes six low-impedance ring macro-contacts (1–2 kΩ; 2 mm height) and ten high-impedance 50-μm micro-contacts (70–90 kΩ). Micro-contact signals were sampled at 24,414 Hz using a PZ5M 256-channel digitizer and RZ2 processor, with data stored on an RS4 system (Tucker-Davis Technologies, Alachua, FL, USA). All analyses in this paper use LFPs from the micro-contacts; macro-contact recordings were not used.

Table A.2 summarizes electrode spatial coverage per subject (S1–S20), listing the number of unique recording channels in each brain region alongside the total recording time and electrode-hours. Brain-region assignments were performed by an expert via manual review of co-registered MRI–CT volumes. Recording time is computed as the sum of session lengths for each subject, while electrode-hours equal each subject's total recording time multiplied by their total channel count. Across all subjects, the dataset comprises 9.96 recording hours and 442.86 electrode-hours.

Table A.2: *Recording coverage by subject and brain region:* "Recording time" sums recording lengths over all sessions for each subject. "Electrode hours" multiply each subject's total recording time by that subject's number of recorded channels across regions.

Subject (sessions)	GPi	PPN	SNR	STN	VA	VIM	VO	Recording time (hrs)	Electrode hours
S1 (2 sessions)	19	0	0	6	0	9	6	1.60	63.81
S2 (3 sessions)	12	0	0	6	6	6	6	0.36	12.97
S3 (3 sessions)	12	0	0	6	6	6	6	0.37	13.40
S4 (2 sessions)	18	12	0	6	8	0	8	0.23	12.06
S5 (4 sessions)	18	6	0	6	8	0	8	0.62	28.36
S6 (4 sessions)	8	3	0	6	8	0	6	0.47	14.46
S7 (3 sessions)	12	6	0	6	12	0	6	0.53	22.29
S8 (4 sessions)	12	6	0	6	6	6	6	0.46	19.42
S9 (2 sessions)	21	6	0	8	11	0	4	0.24	11.94
S10 (4 sessions)	29	9	0	10	0	0	4	0.44	23.07
S11 (5 sessions)	14	9	14	6	15	0	8	0.53	35.24
S12 (3 sessions)	25	9	0	10	0	0	4	0.30	14.64
S13 (2 sessions)	8	9	12	4	0	0	4	0.55	20.27
S14 (2 sessions)	15	8	6	6	6	0	6	0.25	11.74
S15 (4 sessions)	19	0	0	8	0	9	2	0.61	23.27
S16 (4 sessions)	15	10	0	6	8	0	8	0.63	29.51
S17 (4 sessions)	18	0	0	10	0	8	4	0.50	20.18
S18 (2 sessions)	20	0	0	11	0	10	6	0.25	11.84
S19 (4 sessions)	26	12	9	9	0	0	2	0.51	29.80
S20 (3 sessions)	20	0	0	12	0	10	8	0.49	24.62
Total	**341**	**105**	**41**	**148**	**94**	**64**	**112**	**9.96**	**442.86**

MNI coordinate extraction. MNI coordinates for a total of 20 subjects yielding approximately 1200 macro-contact were extracted. Structural T1-weighted (T1W) MRI scans for all 20 subjects were nonlinearly normalized to MNI space using the Advanced Normalization Tools (ANTs) framework (Avants et al., 2008), a widely adopted tool for high-accuracy nonlinear image registration (Tustison et al., 2019; Bartel et al., 2019). Postoperative CT scans were rigidly co-registered to each subject's preoperative T1W MRI using the FLIRT (FMRIB's Linear Image Registration Tool) module of the FMRIB Software Library (FSL). The nonlinear warping transformations obtained from ANTs normalization of the T1W images were subsequently applied to the co-registered CT scans, thereby aligning them to MNI space. Following spatial normalization, approximately 600 macro-contacts were localized across all patients. Contact identification was performed using histogram analysis of CT metal artifacts, with the voxel exhibiting the maximum CT intensity within each artifact region selected as the contact center coordinate. These standardized MNI coordinates of macro-contacts were then used to generate known relative positions for micro-contacts which recorded the LFP signals. micro-contact locations were then used as inputs for the machine learning models. Figures A.1 shows all the micro-electrode locations from 20 subjects relative to 4 deep brain regions.

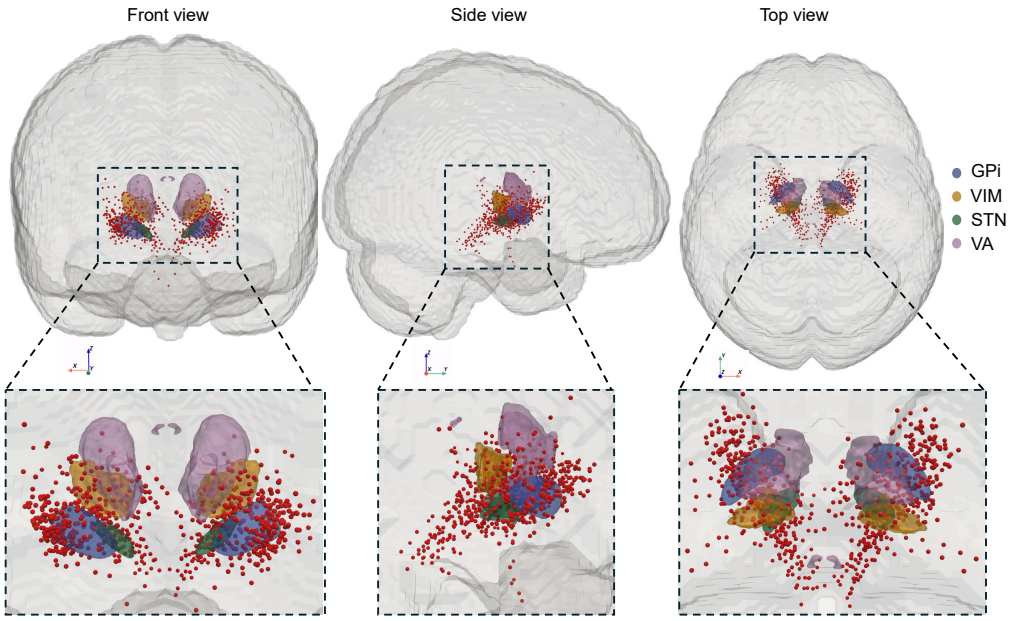


Figure A.1: *3 views of aggregated micro-contacts locations over 20 subjects relative to 4 deep brain regions.*

A.2 Additional Details on Datasets: Signal Preprocessing Details

All preprocessing was performed per channel with zero-phase filtering. We describe the common steps followed by branch-specific steps for the functional encoder and transformer experiments.

Common steps.

1. **Anti-aliasing and resampling.** A 4th-order Butterworth low-pass filter with cutoff 500 Hz was applied, then signals were resampled to 1,000 Hz .
2. **Line-noise removal.** Narrowband notches were applied at 60, 120, and 180 Hz using an IIR notch filter with quality factor Q=30.
3. **Session-wise normalization.** For each session, the signal was standardized by subtracting the session mean and dividing by the session standard deviation.

Functional encoder (contrastive) pipeline.

1. **Windowing.** Continuous recordings were segmented into non-overlapping 10 s windows.
2. **Robust artifact rejection.** Let x denote the full-session signal values for a channel. Define
$$\mathrm{med} = \mathrm{median}(x), \quad \widehat{\sigma} = 1.4826 \cdot \mathrm{median}(|x - \mathrm{med}|) \quad \text{(MAD-based robust std)}.$$
Windows were discarded if any sample fell outside $[\mathrm{med} - \tau\widehat{\sigma},\ \mathrm{med} + \tau\widehat{\sigma}]$, with threshold $\tau = 10$ in our implementation.
3. **Splits.** Non-overlapping 10 s windows were assigned to splits without shuffling in chronological order: the first 70% of windows to training, the next 15% to validation, and the final 15% to test. Entire windows were assigned to a single split (no window was divided across sets).

Transformer pipeline.

1. **Low-pass filtering.** After the notch filters, a 2nd-order Butterworth low-pass filter with 50 Hz cutoff was applied.
2. **Chronological splits.** The first 70% of each session (in time) was used to construct training examples; the next 15% and final 15% were used for validation and test, respectively.

A.3 Additional Details on Datasets: Simulation Study

To evaluate our embedding methods in a fully controlled setting, we constructed a synthetic dataset designed to reflect known region-specific neural signatures while allowing precise control over subject-, session-, and electrode-level variability.

The dataset includes:

- $N = 10$ subjects
- $S = 2$ sessions per subject
- $R = 5$ brain regions per session: GPi, STN, VO, Motor Cortex, and VIM
- $C = 4$ electrodes (channels) per region
- $f_s = 1000$ Hz sampling rate

Each recording channel was synthesized by combining a region-specific neural signal generator with additive Gaussian noise, modulated by three hierarchical levels of parameters: subject-level, session-level, and electrode-level.

1. Region-Specific Neural Signatures

Each brain region was associated with one or more characteristic oscillatory components:

Each oscillatory component was implemented using sine waves modulated by windowed envelopes (e.g., Hann windows for bursts), with frequency jitter and amplitude variation across subjects and channels.

Region	Signature Components
GPi	Beta bursts (∼20 Hz)
STN	Beta bursts + HFOs (∼150 Hz)
VO	Spindle bursts (∼12–16 Hz)
Motor	Continuous gamma activity (∼40 Hz)
VIM	Slow oscillations (∼1 Hz)

2. Parameter Hierarchy

We define the following hierarchical parameter sets:

Subject-Level Parameters: θ_{subject} Sampled once per subject, controlling global oscillation strength and structure:

$$\theta_{\text{subject}} = \{\beta_{\text{gain}},\ \beta_{\text{freq_offset}},\ \beta_{\text{burst_prob}},\ \text{spindle}_{\text{gain}},\ \text{spindle}_{\text{freq_offset}},\ \text{spindle}_{\text{burst_prob}},\ \gamma_{\text{gain}},\ \gamma_{\text{freq_offset}},\ \text{HFO}_{\text{gain}},\ \text{HFO}_{\text{freq_offset}},\ \text{slow}_{\text{gain}},\ \text{slow}_{\text{freq_offset}},\ \text{noise}_{\text{level}}\}$$

Sampled from the following distributions:

$$\begin{aligned}
\beta_{\text{gain}} &\sim \mathcal{U}(0.8,\ 1.5) & \beta_{\text{freq_offset}} &\sim \mathcal{U}(-2.0,\ 2.0) & \beta_{\text{burst_prob}} &\sim \mathcal{U}(0.3,\ 0.7)\\
\text{spindle}_{\text{gain}} &\sim \mathcal{U}(0.8,\ 1.3) & \text{spindle}_{\text{freq_offset}} &\sim \mathcal{U}(-1.0,\ 1.0) & \text{spindle}_{\text{burst_prob}} &\sim \mathcal{U}(0.2,\ 0.5)\\
\gamma_{\text{gain}} &\sim \mathcal{U}(0.3,\ 0.7) & \gamma_{\text{freq_offset}} &\sim \mathcal{U}(-5.0,\ 5.0)\\
\text{HFO}_{\text{gain}} &\sim \mathcal{U}(0.1,\ 0.3) & \text{HFO}_{\text{freq_offset}} &\sim \mathcal{U}(-10.0,\ 10.0)\\
\text{slow}_{\text{gain}} &\sim \mathcal{U}(0.8,\ 1.2) & \text{slow}_{\text{freq_offset}} &\sim \mathcal{U}(-0.3,\ 0.3)\\
\text{noise}_{\text{level}} &\sim \mathcal{U}(0.3,\ 0.4)
\end{aligned}$$

Session-Level Parameters: θ_{session} Sampled per session:

$$\theta_{\text{session}} = \{\text{session_noise} \sim \mathcal{U}(0.05,\ 0.15)\}$$

Electrode-Level Parameters: $\theta_{\text{electrode}}$ Sampled per electrode:

$$\theta_{\text{electrode}} = \{\text{gain} \sim \mathcal{U}(0.9,\ 1.1),\ \text{electrode_noise} \sim \mathcal{U}(0.03,\ 0.1)\}$$

3. Signal Model

For a given region r, session s, and electrode e, the generated signal is defined as:

$$\text{Signal}_{r,s,e}(t) = \text{Signature}_r(t;\ \theta_{\text{subject}},\ \theta_{\text{electrode}}) + \mathcal{N}\big(0,\ \sigma^2_{\text{subject}} + \sigma^2_{\text{session}} + \sigma^2_{\text{electrode}}\big)$$

Each component of the signal is constructed from filtered oscillatory patterns with gain-scaled amplitudes, frequency offsets, and additive Gaussian noise aggregated from all three levels. The pipeline for generating simulated dataset is depicted in fig. A.2.

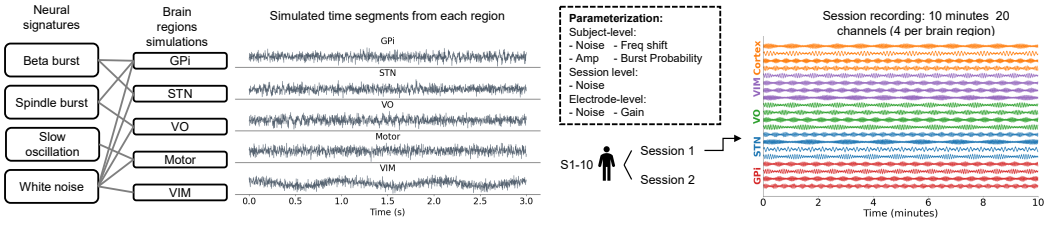


Figure A.2: *Simulation dataset.* pipeline: each region is assigned a parameterized neural signature; recordings are generated across subjects/sessions with gain, frequency shifts, and noise parameters.

A.4 Additional Details on Methods: CNN Functional Encoder-Architecture and Training

Setup. We learn a subject-agnostic 32D embedding of 1D neural time series using convolutional neural network.

Single-session encoder. The encoder f_θ is a 1D CNN with following blocks:

$$\begin{aligned}
&\text{Conv1d}(1, 64, k{=}15, s{=}4) \to \text{BN}(64) \to \text{GELU} \to \text{MaxPool}(3, s{=}2) \to \text{Dropout} \\
&\text{Conv1d}(64, 64, k{=}11, s{=}2) \to \text{BN}(64) \to \text{GELU} \to \text{MaxPool}(3, s{=}2) \to \text{Dropout} \\
&\text{Conv1d}(64, 64, k{=}7, s{=}2) \to \text{BN}(64) \to \text{GELU} \to \text{MaxPool}(3, s{=}2) \to \text{Dropout} \\
&\text{Conv1d}(64, 64, k{=}5, s{=}2) \to \text{BN}(64) \to \text{GELU} \to \text{MaxPool}(3, s{=}2) \to \text{Dropout} \\
&\text{AdaptiveAvgPool1d}(10) \;\to\; \text{Flatten } (10{\times}64) \;\to\; \text{MLP: } \mathbb{R}^{640} \to \mathbb{R}^{32} \to \mathbb{R}^{32}.
\end{aligned}$$

The output embedding is ℓ_2-normalized. This configuration has $117{,}504$ trainable parameters.

Training. We train for 100 epochs using AdamW (Loshchilov & Hutter, 2019) (weight decay 10^{-4}, $\beta_1{=}0.9$, $\beta_2{=}0.99$) with initial learning rate $\text{lr}{=}0.01$, batch size 128, and a ReduceLROnPlateau scheduler (mode=`min`, patience=5, factor=0.5, $\text{lr}_{\min}{=}10^{-5}$). Dropout in conv blocks is 0.5. The embedding head applies a small dropout before the final linear layer and outputs unit-norm vectors.

Multi-subject / multi-session encoder For pooled training across subjects and sessions we use a deeper/wider encoder with the same design principles, increasing channels and depth before the global pooling step:

$$\begin{aligned}
&\text{Conv1d}(1, 64, k{=}15, s{=}4) \to \text{BN}(64) \to \text{GELU} \to \text{MaxPool}(3, s{=}2) \to \text{Dropout} \\
&\text{Conv1d}(64, 128, k{=}11, s{=}2) \to \text{BN}(128) \to \text{GELU} \to \text{MaxPool}(3, s{=}2) \to \text{Dropout} \\
&\text{Conv1d}(128, 128, k{=}11, s{=}2) \to \text{BN}(128) \to \text{GELU} \to \text{MaxPool}(3, s{=}2) \to \text{Dropout} \\
&\text{Conv1d}(128, 128, k{=}11, s{=}2) \to \text{BN}(128) \to \text{GELU} \to \text{MaxPool}(3, s{=}2) \to \text{Dropout} \\
&\text{Conv1d}(128, 128, k{=}7, s{=}2) \to \text{BN}(128) \to \text{GELU} \to \text{Dropout} \\
&\text{Conv1d}(128, 128, k{=}5, s{=}2) \to \text{BN}(128) \to \text{GELU} \to \text{Dropout} \\
&\text{AdaptiveAvgPool1d}(10) \;\to\; \text{Flatten } (10{\times}128) \;\to\; \text{MLP: } \mathbb{R}^{1280} \to \mathbb{R}^{32} \to \mathbb{R}^{32}.
\end{aligned}$$

The output embedding remains $F{=}32$ and is ℓ_2-normalized. This configuration has $694{,}464$ trainable parameters. Dropout in conv blocks is set to 0.3.

Training. We use the same optimizer family and regularization as above (AdamW with weight decay 10^{-4}, $\beta_1{=}0.9$, $\beta_2{=}0.99$), increasing the batch size to 1024 to leverage cross-subject aggregation. Learning-rate scheduling follows the same ReduceLROnPlateau policy; the initial learning rate matches the single-session setup unless otherwise stated. The embedding head outputs unit-norm vectors used by the contrastive criteria in the main text.

A.5 Additional Details on Methods: Functional Transformer-Architecture and Training

The functional embedding of each electrode serves as a subject-agnostic channel identity. For every channel, the first 10 s snippet (from training dataset) is fed to the embedding encoder to produce its functional vector; this vector is then projected to the model width and reused unchanged for all tokens originating from that channel (across time patches and windows). We fuse the projected embedding with time–patch features and a time-only positional code to form the tokens consumed by the transformer. Variable channel counts are handled with padding masks.

Cross-Region LFP Prediction Setup. For each window we have multichannel LFP signals $\mathbf{X} \in \mathbb{R}^{C\times T}$ and a per–channel *functional embedding* matrix $\mathbf{E} \in \mathbb{R}^{C\times F}$ with $F{=}32$. At evaluation, a region $\mathcal{R}$ (e.g., VO) is fully masked: its channels provide no signal. Let $\mathcal{S} = \{c : c \notin \mathcal{R}\}$ be *sources* and $\mathcal{T} = \{k : k \in \mathcal{R}\}$ be *targets*. Given source signals $\{\mathbf{x}_c\}_{c\in\mathcal{S}}$ and source and target embeddings $\{\mathbf{e}_c\}_{c\in\mathcal{S}\cup\mathcal{T}}$, the model reconstructs target signals $\{\widehat{\mathbf{y}}_k\}_{k\in\mathcal{T}}$.

Functional tokens. For each channel, a single 10 s snippet (from training data) is fed to the embedding encoder to produce its functional vector $\mathbf{e}_c \in \mathbb{R}^{32}$. This vector is linearly mapped to model width $\mathbf{f}_c \in \mathbb{R}^d$ and then reused unchanged for every token originating from that channel (across patches and windows). Layer normalization and learned type embeddings follow standard practice and are omitted from equations for clarity.

Tokenization of source signals. A 1D conv "tokenizer" converts each source $\mathbf{x}_c$ into P time patches. Let $t \in \{1, \dots, P\}$ index the patch. Formally,

$$\mathbf{Z}_c \;=\; \mathrm{Conv1d}(\mathbf{x}_c;\, L_{\text{patch}}, S) \in \mathbb{R}^{P \times d}, \qquad P \approx \left\lfloor \tfrac{T - L_{\text{patch}}}{S} \right\rfloor + 1,$$

where d is both the number of conv filters and the Transformer width. $\mathbf{Z}_{ct} \in \mathbb{R}^d$ is the d-dim temporal feature for channel c at patch t. We *broadcast* the functional identity $\mathbf{f}_c \in \mathbb{R}^d$ to all patches of channel c and *concatenate* it with the temporal features, $[\mathbf{Z}_{ct} \| \mathbf{f}_c] \in \mathbb{R}^{2d}$. A dense layer *reduces back to width* d and we add a *time-only* positional code $\mathbf{P}_t^{(t)} \in \mathbb{R}^d$, yielding a source token of shape:

$$\widetilde{\mathbf{Z}}_{ct} \;\in\; \mathbb{R}^d \qquad (c \in \mathcal{S},\; t = 1, \dots, P).$$

Stacking tokens across channels and patches produces the encoder input sequence

$$\mathbf{X}_{\text{src}} \;=\; [\,\widetilde{\mathbf{Z}}_{ct}\,]_{c \in \mathcal{S},\; t=1..P} \;\in\; \mathbb{R}^{S_{\text{tok}} \times d}, \quad S_{\text{tok}} = |\mathcal{S}|\, P,$$

with a key-padding mask $\mathbf{m}_{\text{src}} \in \{0,1\}^{S_{\text{tok}}}$ to ignore padded (absent) sources within a batch (due to variable number of channels present in the dataset).

Tokenization of query signals (cross-region reconstruction targets). Targets have no signal, so temporal features are replaced with a *learned query base* $\mathbf{U}_t \in \mathbb{R}^d$ (one d-vector per patch index t). For each target channel $k \in \mathcal{T}$ and patch t, we concatenate the base with the functional embedding of that channel $[\mathbf{U}_t \| \mathbf{f}_k] \in \mathbb{R}^{2d}$, reduce back to width d, and add the same time-only positional code, producing query tokens of shape:

$$\mathbf{Q}_{kt} \;\in\; \mathbb{R}^d \qquad (k \in \mathcal{T},\; t = 1, \dots, P).$$

Stacking all targets and patches forms the decoder input sequence

$$\mathbf{Q} \;=\; [\,\mathbf{Q}_{kt}\,]_{k \in \mathcal{T},\; t=1..P} \;\in\; \mathbb{R}^{T_{\text{tok}} \times d}, \quad T_{\text{tok}} = |\mathcal{T}|\, P,$$

with a query-padding mask $\mathbf{m}_{\text{tgt}} \in \{0,1\}^{T_{\text{tok}}}$ to ignore padded (absent) targets.

Encoder-decoder attention. We use pre-LN Transformer blocks with h heads and per–head size $d_h = d/h$. In particular multi-head self-attention (MHSA) and multi-head attention (MHA) blocks are used where :

$$\mathrm{MHSA}(\mathbf{X}) = \mathrm{Concat}_h\big[\mathrm{Attn}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i)\big]\mathbf{W}^O, \quad \mathbf{Q}_i = \mathbf{X}\mathbf{W}_i^Q,\; \mathbf{K}_i = \mathbf{X}\mathbf{W}_i^K,\; \mathbf{V}_i = \mathbf{X}\mathbf{W}_i^V,$$

where $\mathbf{W}^O$ projects h concatenated outputs to model width (d), i denotes a specific head from h heads, and Attn defined as:

$$\mathrm{Attn}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i) = \mathrm{softmax}\left(\frac{\mathbf{Q}_i \mathbf{K}_i^\top}{\sqrt{d_h}}\right)\mathbf{V}_i.$$

MHA follows identical formulation where query ($\mathbf{Q}_i$) comes from a source other than $\mathbf{X}$.

Encoder (self-attn on sources).
With source tokens $\mathbf{X}_{\text{src}} \in \mathbb{R}^{S_{\text{tok}} \times d}$,

$$\underbrace{\mathbf{Q}^{\text{enc}} = \mathbf{X}_{\text{src}}\,\mathbf{W}_{\text{enc}}^Q, \quad \mathbf{K}^{\text{enc}} = \mathbf{X}_{\text{src}}\,\mathbf{W}_{\text{enc}}^K, \quad \mathbf{V}^{\text{enc}} = \mathbf{X}_{\text{src}}\,\mathbf{W}_{\text{enc}}^V}_{\in\, \mathbb{R}^{S_{\text{tok}} \times d}} \;\Rightarrow\; \mathbf{M} = \mathrm{MHSA}(\mathbf{X}_{\text{src}}) \in \mathbb{R}^{S_{\text{tok}} \times d},$$

with key-padding mask $\mathbf{m}_{\text{src}}$ applied inside the softmax.

Decoder (self-attn on queries, then cross-attn to memory).
First, self-attention over query tokens $\mathbf{Q} \in \mathbb{R}^{T_{\text{tok}} \times d}$:

$$\underbrace{\mathbf{Q}^{\text{dec}} = \mathbf{Q}\,\mathbf{W}_{\text{dec}}^Q,\; \mathbf{K}^{\text{dec}} = \mathbf{Q}\,\mathbf{W}_{\text{dec}}^K,\; \mathbf{V}^{\text{dec}} = \mathbf{Q}\,\mathbf{W}_{\text{dec}}^V}_{\in\, \mathbb{R}^{T_{\text{tok}} \times d}} \;\Rightarrow\; \widetilde{\mathbf{Q}} = \mathrm{MHSA}(\mathbf{Q}) \in \mathbb{R}^{T_{\text{tok}} \times d},$$

masking padded query tokens via $\mathbf{m}_{\text{tgt}}$. Next, *cross-attention* where queries attend to encoder memory:

$$\underbrace{\mathbf{Q}^{\text{qry}} = \widetilde{\mathbf{Q}}\,\mathbf{W}^{Q}_{\text{qry}}}_{\in\,\mathbb{R}^{T_{\text{tok}}\times d}},\quad \underbrace{\mathbf{K}^{\text{mem}} = \mathbf{M}\,\mathbf{W}^{K}_{\text{mem}},\ \mathbf{V}^{\text{mem}} = \mathbf{M}\,\mathbf{W}^{V}_{\text{mem}}}_{\in\,\mathbb{R}^{S_{\text{tok}}\times d}} \Rightarrow\ \mathbf{H} = \mathrm{MHA}(\widetilde{\mathbf{Q}}, \mathbf{M}, \mathbf{M}) \in \mathbb{R}^{T_{\text{tok}}\times d},$$

with key-padding mask $\mathbf{m}_{\text{src}}$ applied to the memory keys.

Stacking L_{enc} and L_{dec} layers yields the final decoder states $\mathbf{H}$.

Waveform head and outputs. A linear head maps each decoder token to a patch of L_{patch} samples, then patches are reshaped and concatenated along time:

$$\mathbf{H} \in \mathbb{R}^{T_{\text{tok}}\times d} \xrightarrow{\text{linear}} \mathbf{R} \in \mathbb{R}^{T_{\text{tok}}\times L_{\text{patch}}} \xrightarrow{\text{reshape}} \widehat{\mathbf{Y}} \in \mathbb{R}^{K\times P\times L_{\text{patch}}} \xrightarrow{\text{concat over } t} \widehat{\mathbf{Y}} \in \mathbb{R}^{K\times T_m}$$

$$, T_m = P\,L_{\text{patch}}$$

(An overlap–add variant with stride $S{<}L_{\text{patch}}$ is compatible but not used here.)

Masking and batching. For each window, sources exclude all channels in $\mathcal{R}$ and targets include all channels in $\mathcal{R}$. Across windows/subjects the numbers of sources and targets vary; we pad to the batch maxima and use $\mathbf{m}_{\text{src}}$ and $\mathbf{m}_{\text{tgt}}$ so attention and losses ignore padded tokens.

Objective and optimization. Given ground truth $\mathbf{Y} \in \mathbb{R}^{K\times T_m}$ and predictions $\widehat{\mathbf{Y}}$,

$$\mathcal{L} \;=\; \underbrace{\tfrac{1}{T_m}\|\widehat{\mathbf{Y}} - \mathbf{Y}\|_2^2}_{\text{MSE}} + \lambda\Big(1 - \rho(\widehat{\mathbf{Y}}, \mathbf{Y})\Big), \qquad \lambda = 0.05,$$

where ρ is mean Pearson correlation across targets (after per-sequence standardization). The correlation term enforces *shape fidelity* (phase/temporal structure) and mitigates the tendency of pure MSE to prefer amplitude shrinkage or flat predictions under SNR or scale mismatch. No subject IDs, subject-specific heads, or fine-tuning are used; a single model is trained and evaluated across all subjects for multi-subject experiments.

Model parameters and training. For training on multi-session multi-subject data, we use $L_{\text{enc}} = 3$ encoder layers and $L_{\text{dec}} = 3$ decoder layers with model width $d = 128$ and $h = 4$ attention heads ($d_h = 32$). The Transformer feed-forward dimension is 384 with dropout set to 0.2. The tokenizer patch length is $L_{\text{patch}} = 25$ ms. Each model input window is 1 s of LFP, drawn from a variable number of channels (subject-dependent). Windows are extracted with a 500 ms overlap after the train/validation/test split to avoid leakage. The resulting model has approximately 1.37×10^6 trainable parameters.

Optimization uses AdamW with $\beta_1 = 0.9$, $\beta_2 = 0.99$, and $\varepsilon = 10^{-8}$. We employ per-parameter learning rates: 2×10^{-4} for all layers except the waveform head, which uses 3×10^{-4}. Weight decay is 10^{-4}, except it is disabled for parameters that should not be regularized (e.g., LayerNorm scales/biases, embedding and type-embedding vectors, and the learned query base), following common practice. 150 epochs and batch size of 128 was used for training.

For experiments trained on a single session (500–800 windows), we use a smaller Transformer with model width $d = 32$ and feed-forward dimension $= 32$, yielding $\sim 9.0 \times 10^4$ trainable parameters. All other architectural and optimization settings remain the same, except that we increase the learning rates by $10\times$ (i.e., 2×10^{-3} for all layers and 3×10^{-3} for the waveform head). Batch size and number of epochs are identical to the multi-subject setting.

A.6 Additional Figures: Single-Subject and Pooled Functional Embedding Spaces

Figure A.3 shows per-subject held-out time-segment embeddings and confusion matrices for PSC encoders trained independently within each subject. Figure A.4 shows the corresponding embeddings and confusion matrices for the pooled multi-subject MSC encoder, using a shared PCA space across subjects. These visualizations complement the main single-subject and pooled-encoder results in Figures 3 and 5.

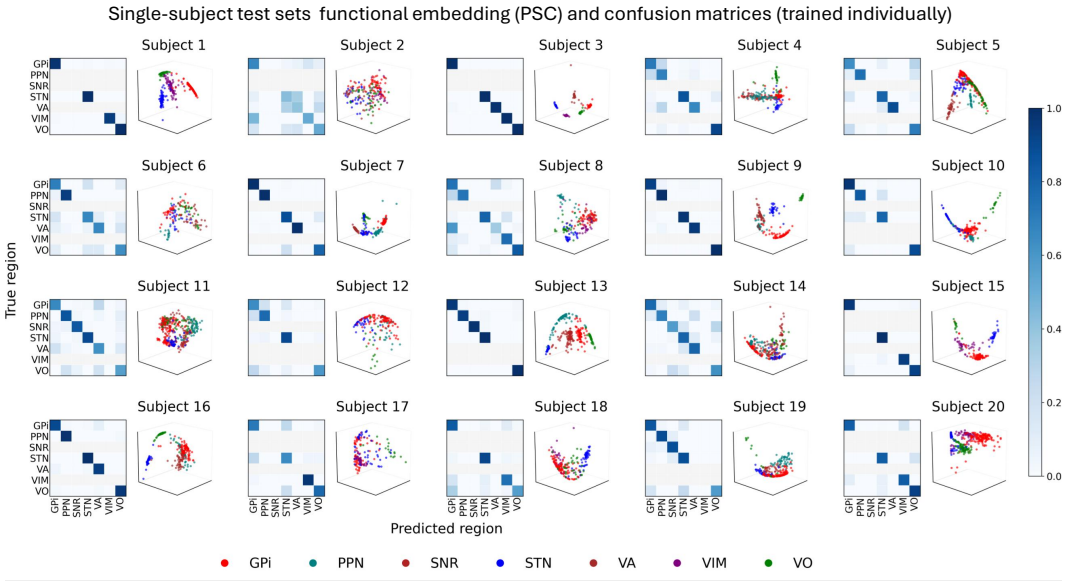


Figure A.3: **Single-subject PSC functional embeddings across subjects.** For each subject, the left panel shows the row-normalized k-NN confusion matrix for held-out time segments, and the right panel shows the 3D PCA projection of the 32-D functional embeddings. Each encoder was trained independently using PSC on that subject's data. Confusion matrices use a fixed region order and shared color scale across subjects.

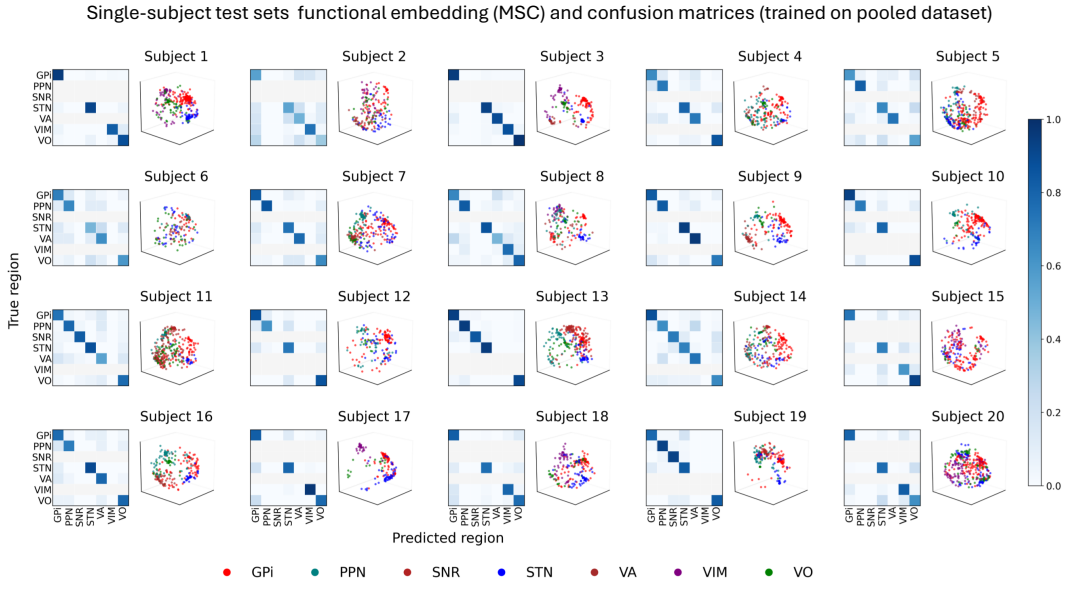


Figure A.4: **Pooled MSC functional embeddings across subjects.** For each subject, the left panel shows the row-normalized k-NN confusion matrix for held-out time segments, and the right panel shows the 3D PCA projection of embeddings from the pooled multi-subject MSC encoder. A single PCA basis is fit across all displayed subject test embeddings, allowing visualization in a shared low-dimensional coordinate system. Confusion matrices use a fixed region order and shared color scale across subjects.

A.7 Additional Figures: Pooled vs. Single-Subject Functional Encoders

To assess whether functional embeddings benefit from cross-subject aggregation, we compared the pooled multi-subject encoder against encoders trained separately within each subject. Both models were evaluated on held-out time segments and held-out channels from the same recording sessions.

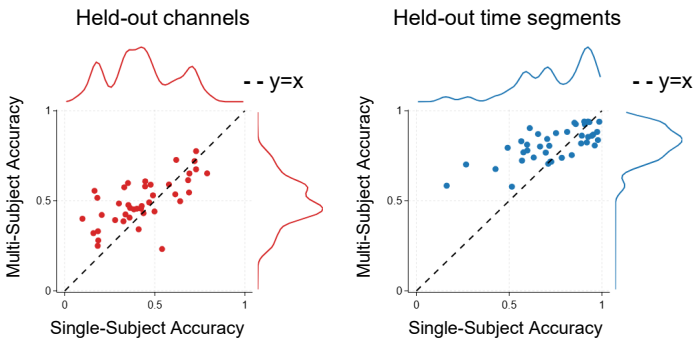


Figure A.5: **Pooled vs. single-subject functional encoder performance.** Each point represents one recording session. The x-axis shows accuracy from a functional encoder trained only on that subject, while the y-axis shows accuracy from the pooled multi-subject encoder. Results are shown separately for held-out channels and held-out time segments. The dashed line denotes equal performance. Points above the diagonal indicate sessions where pooled training improves accuracy.

The pooled encoder performs better than the corresponding single-subject encoder in most sessions, especially for sessions where single-subject performance is low. This suggests that cross-subject training stabilizes the functional embedding space and helps compensate for the limited number of channels available within individual subjects. In the main text, we summarize this result quantitatively: the pooled encoder improves held-out time and held-out channel accuracy by approximately 5% on average, while reducing variability across subjects.

A.8 Additional Figures: Per-subject cross-region prediction performance

Figure A.6 provides subject-wise prediction performance corresponding to the pooled cross-region LFP prediction experiment shown in the main text.

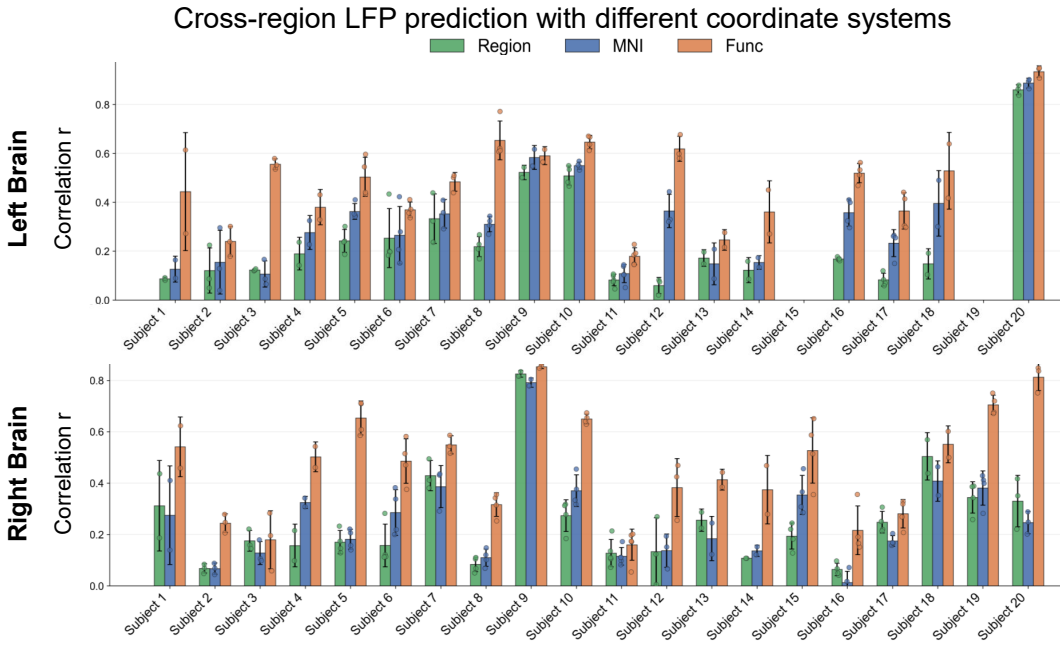


Figure A.6: *Subject-wise cross-region LFP prediction across coordinate systems.* Bar plots show the average Fisher-transformed prediction correlation (r) for each subject, separately for left and right hemispheres, using one-hot region-label, MNI coordinates, or functional coordinates. Error bars indicate variability across recording sessions, while faint circles denote individual session-level results. Functional coordinates consistently improve prediction performance relative to anatomical coordinate systems across most subjects and both hemispheres. This figure corresponds to the pooled multi-subject multi-session transformer experiment (cross-region LFP prediction) presented in the main text.

A.9 Additional Figures: Comparison of PSC and MSC Functional Embeddings

To further compare the two contrastive objectives used for pooled functional-embedding training, we evaluated PSC and MSC on held-out time segments and held-out channels from the multi-subject test set. PSC achieved $80.71\% \pm 11.41\%$ accuracy on held-out time segments and $49.18\% \pm 12.11\%$ on held-out channels. MSC achieved 79.99% accuracy on held-out time segments and 52.84% on held-out channels. Thus, while PSC performed slightly better on held-out time segments, MSC showed stronger held-out-channel generalization, motivating its use as the functional coordinate system for the transformer experiments. The embedding geometries also differed qualitatively: PSC produced more compact, centroid-like clusters in Euclidean feature space, whereas MSC, optimized with cosine similarity on ℓ_2-normalized embeddings, formed broader angle-separated region clouds on the hypersphere.

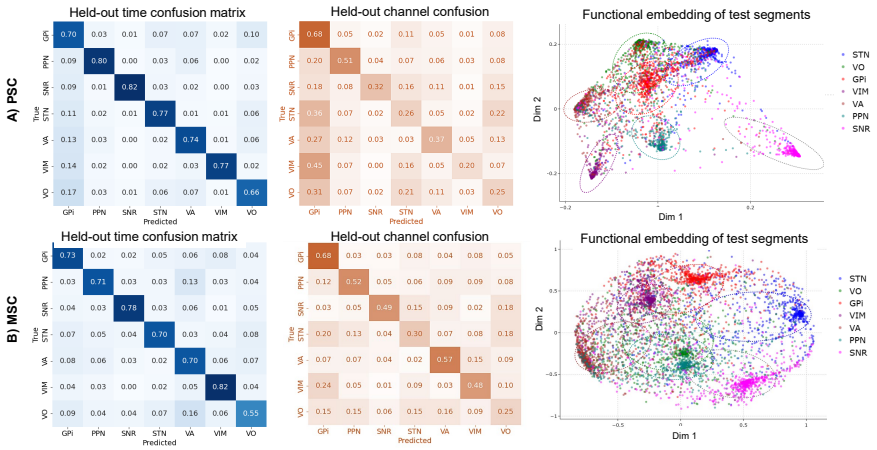


Figure A.7: **PSC vs. MSC held-out evaluation.** Row-normalized k-NN confusion matrices for pooled functional embeddings evaluated on held-out time segments and held-out channels. MSC improves held-out-channel accuracy, indicating better generalization to unseen electrodes while preserving region-structured embeddings.

A.10 Additional Analysis: Relationship Between Functional and MNI Geometry

We examined whether the learned MSC functional coordinates reproduce anatomical distances. Using the chronologically held-out 10 s segments used for the MSC test visualization, we averaged each physical channel's normalized 32-D embeddings across its available test segments. We compared cosine distances between these channel coordinates with Euclidean distances between their MNI coordinates. All pairs comprised channels from *different subjects*; analyses were conducted separately by hemisphere. We used Spearman rank correlations, permutation p-values for the across-region analyses, and Holm correction for the within-region comparisons.

Table A.3: *Across-subject distance correspondence across all sampled nuclei.* Spearman correlations relate functional cosine distance to MNI Euclidean distance for channel pairs belonging to different subjects.

Hemisphere	Channels	Cross-subject pairs	Spearman ρ	Permutation p
Left	293	39,960	0.346	9.99×10^{-4}
Right	247	28,359	0.398	9.99×10^{-4}

The moderate across-region correlations indicate that channels separated in MNI space tend also to be separated in the functional embedding. We repeated the distance analysis within each of the four most sampled nuclei; none of the within-nucleus correlations survived Holm correction (Table A.4). To ask whether the *full* embedding predicts absolute position, we fitted ridge regression separately for MNI x, y, and z within each nucleus and hemisphere, with leave-one-subject-out evaluation. Most cross-validated R^2 values were negative relative to the training-set mean-coordinate baseline (Table A.4).

Table A.4: *Within-nucleus functional–MNI correspondence across subjects.* ρ compares cross-subject functional and anatomical pairwise distances; p_{Holm} corrects the within-nucleus distance tests. The three R^2 columns evaluate ridge prediction of absolute MNI coordinates from the full functional embedding under leave-one-subject-out cross-validation.

Hemisphere	Nucleus	Channels	Spearman ρ	p_{Holm}	R^2_x	R^2_y	R^2_z
Left	GPi	103	-0.087	0.396	-0.239	-0.554	-0.476
Left	STN	50	0.165	0.120	0.152	-0.142	0.158
Left	VO	35	0.051	0.959	-0.098	-0.519	-0.256
Left	VIM	22	0.018	0.959	-0.356	-1.112	-0.744
Right	GPi	97	-0.009	1.000	-0.206	-0.190	-0.213
Right	STN	39	-0.017	1.000	-0.503	-0.610	-0.634
Right	VO	29	0.155	0.427	-0.185	-0.509	-1.273
Right	VIM	14	-0.089	1.000	-1.807	-0.908	-1.114

Left STN was a limited exception in the absolute-coordinate analysis: its x and z axes gave $R^2 = 0.152$ and 0.158, respectively (observed–predicted $\rho = 0.356$, $p = 0.011$ for x; $\rho = 0.463$, $p = 7.12 \times 10^{-4}$ for z). Correlations between individual functional principal components and MNI axes did not survive multiple-comparison correction, and analogous right-STN effects were absent. Thus, the global distance relationship is consistent with coarse organization across nuclei; these tests do not establish that the embedding reproduces continuous MNI geometry within nuclei.

A.11 Additional Analysis: Capacity-Matched Coordinate Controls

In the original pooled prediction comparison, all coordinate descriptors entered the same token pathway. A trainable projection mapped a functional (32-D), MNI (3-D), or one-hot region (4-D) descriptor to width $d = 128$. This output was concatenated with a 128-D temporal feature or target-query representation and fused back to a 128-D token before the same encoder–decoder Transformer.

The three models therefore shared the temporal tokenizer, query representation, attention and feed-forward layers, masking, output head, and attention-token width. Only the input coordinate projection varied with descriptor dimension (Table A.5).

We also trained two nonlinear controls that first formed learned 32-D anatomical or categorical representations:

$$\text{MNI-MLP32:} \quad \mathbb{R}^3 \xrightarrow{\text{Linear}} \mathbb{R}^{32} \xrightarrow{\text{GELU}} \xrightarrow{\ell_2\text{-normalize}} \xrightarrow{\text{Linear}} \mathbb{R}^{128},$$

$$\text{Region-MLP32:} \quad \mathbb{R}^4 \xrightarrow{\text{Linear}} \mathbb{R}^{32} \xrightarrow{\text{GELU}} \xrightarrow{\ell_2\text{-normalize}} \xrightarrow{\text{Linear}} \mathbb{R}^{128}.$$

Two linear layers alone would collapse to one affine map, so the GELU and normalization make these genuine nonlinear controls. The downstream Transformer, data splits, source/target masking, optimization, and evaluation protocol were the same as in the original comparison. The nonlinear controls have slightly more trainable prediction-model parameters than Functional32.

Table A.5: *Trainable parameters in the coordinate-conditioned prediction models.* The original comparisons already matched token width; the additional MLP32 controls match the intermediate coordinate dimension and slightly exceed Functional32 in trainable parameter count.

Coordinate descriptor	Trainable parameters	Difference from Functional32
Functional32	1,366,555	—
MNI	1,362,843	$-3,712$ (-0.272%)
Region	1,362,971	$-3,584$ (-0.262%)
MNI-MLP32	1,366,683	$+128$
Region-MLP32	1,366,715	$+160$

Table A.6: *Subject-level Fisher-averaged reconstruction correlations (r; mean $\pm$ SD).* Left and right columns include $n = 18$ and $n = 20$ subjects, respectively.

Coordinate descriptor	Left ($n = 18$)	Right ($n = 20$)
Region	0.2373 ± 0.2074	0.2505 ± 0.1823
Region-MLP32	0.2416 ± 0.2023	0.2553 ± 0.1838
MNI	0.3183 ± 0.2031	0.2561 ± 0.1734
MNI-MLP32	0.3186 ± 0.2071	0.3413 ± 0.1837
Functional32	$\mathbf{0.4817 \pm 0.1844}$	$\mathbf{0.4759 \pm 0.1999}$

Against the strongest capacity-matched anatomical control, MNI-MLP32, Functional32 improved mean correlation by $\Delta r = 0.1631$ on the left (paired Wilcoxon test, Holm-corrected $p = 3.1 \times 10^{-5}$) and $\Delta r = 0.1346$ on the right (Holm-corrected $p = 2.7 \times 10^{-5}$). The advantage thus persists when the anatomical and categorical descriptors receive nonlinear learned 32-D representations and the prediction models use identical 128-D Transformer tokens.

A.12 External Validation on a Public Movie-Viewing Dataset

We repeated the cross-region LFP prediction task on the independent public dataset described by Keles et al. (2024b) (DANDI:000623; Keles et al., 2024a). The analyzed subset comprised 14 people with epilepsy and 26 movie-viewing runs with consistent coverage of five regions: anterior cingulate cortex (ACC), amygdala, hippocampus, pre-supplementary motor area (preSMA), and ventromedial prefrontal cortex (vmPFC). It provided approximately 3.5 hours of simultaneous intracranial LFP recordings from hybrid Behnke–Fried depth electrodes, with eight microwires per bundle, along with region labels and MNI coordinates. The public LFPs were sampled at 1,000 Hz. Relative to the primary dystonia cohort, these recordings came from a different institution, disease population, behavioral paradigm, and cortical/limbic anatomy. They also sampled each region and hemisphere more locally, typically through one compact microwire bundle, and had a shorter total duration.

We trained a *new* MSC functional encoder and a *new* encoder–decoder Transformer on this external dataset as the regions recorded were different from the SEEG cohort reported in this paper. The architectures and masked-region prediction task followed the primary experiment, but region identities and all fitted weights were specific to this cohort; no encoder or predictor parameters were transferred from the dystonia cohort. Preprocessing followed the primary protocol, including 1,000-Hz sampling and low-pass filtering. Within each recording session, the first 70% of time was assigned to training, the next 15% to validation and checkpoint selection, and the final 15% to testing; the two hemispheres were pooled. We withheld *all* vmPFC channels as targets and predicted their waveforms from ACC, amygdala, hippocampus, and preSMA channels. The training protocol ran for 150 epochs.

We compared the 32-D signal-derived coordinates (Functional32) with nonlinear, capacity-matched coordinate controls built and trained separately for this dataset, following Appendix A.11. MNI-MLP32 maps three MNI coordinates through a 32-D GELU and ℓ_2-normalized representation to the common 128-D token pathway. Region-MLP32 does the same using a five-element one-hot region descriptor. Their prediction models had 1,366,683 and 1,366,747 trainable parameters, respectively, compared with 1,366,555 for Functional32. Transformer architecture, masking, optimization, and evaluation procedures were matched across these three external models.

Table A.7: *External public dataset: subject-level masked-vmPFC prediction.* Fisher-averaged test correlation r is reported as mean $\pm$ SD across 14 subjects; trainable parameter counts refer to the coordinate-conditioned prediction models. The functional encoder and each predictor were trained anew on the external dataset.

Coordinate descriptor	Trainable parameters	Test correlation ($n = 14$)
Region-MLP32	1,366,747	0.3089 ± 0.2433
MNI-MLP32	1,366,683	0.3234 ± 0.2538
Functional32	1,366,555	$\mathbf{0.3772 \pm 0.2573}$

Functional32 outperformed MNI-MLP32 in all 14 subjects (median paired $\Delta r = 0.034$; paired Wilcoxon test, Holm-corrected $p = 1.22 \times 10^{-4}$) and Region-MLP32 in 13 of 14 subjects (median paired $\Delta r = 0.063$; Holm-corrected $p = 1.22 \times 10^{-4}$). This reproduces the masked-region prediction advantage with new subjects, recording sites, regions, and behavioral context under a comparable LFP acquisition and modeling framework. Because both stages were trained from scratch on the public dataset, this is *replication across datasets*, not transfer of a previously learned functional coordinate system or predictor.

A.13 Cross-Species Validation on Allen Visual Behavior Neuropixels

Dataset, signals, and splits. We used public Allen Visual Behavior Neuropixels LFPs from an image-change detection task (Bennett et al., 2026; Bennett, 2024), sampling eight valid electrodes in each of six visual cortical areas (VISp, VISl, VISal, VISrl, VISam, VISpm) per session. Sessions included Familiar and Novel image experience. LFPs were low-pass filtered at 500 Hz before resampling from 1,250 to 1,000 Hz and notched at 60/120/180 Hz. The 692,288-parameter MSC coordinate encoder used clean, nonoverlapping 10-s segments after a 120-s warm-up; its training, checkpoint-validation, and held-out test groups comprised eight, two, and two distinct mice, respec-

tively. The cosine 5-nearest-neighbor region readout used only training-channel embeddings; its test on 192 physical channels from the two held-out mice yielded six-area balanced accuracy 0.609 (macro-F1 0.607; chance 0.167). The 32-D channel coordinate was the normalized mean of up to five clean segments from early unlabeled recordings. Region labels supervised encoder training and atlas annotations selected the six-area electrode set, but the Functional decoder received no region or CCF input at test time.

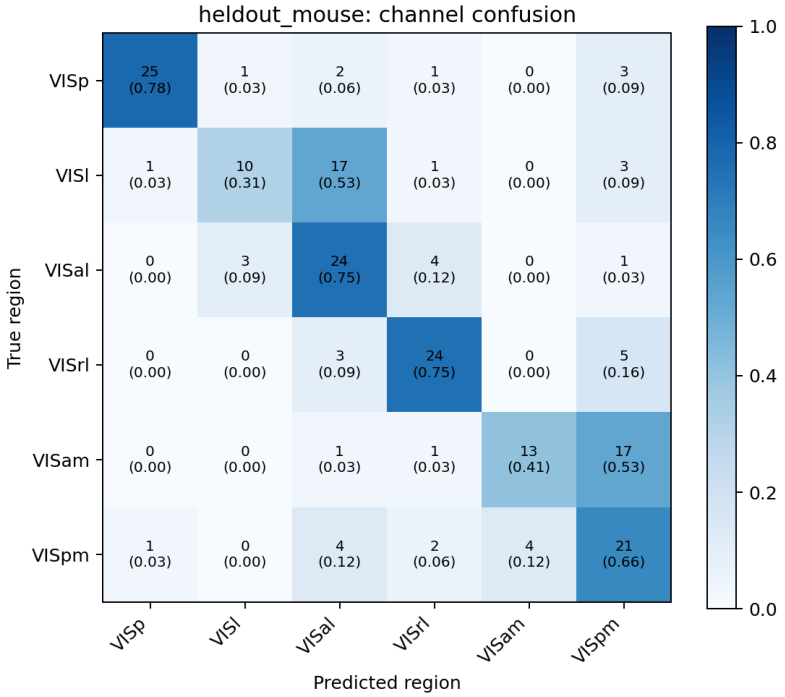


Figure A.8: *confusion matrix of the held-out unseen test mice.* held-out mice confusion matrix shows generalizability of functional coordinates across mice and its region prediction power

Behavioral prediction. We froze the mouse-trained MSC encoder and separately trained pooled Functional, CCF, and region-label classifiers on 12 training mice (24 sessions; 4,656 unique active-block go trials). Three different validation mice (six sessions; 1,468 trials) selected weights and decision thresholds. Two further mice, absent from both stages of training, were reserved for final evaluation. Hits and misses were labeled at the display-corrected onset of the change image; aborted, auto-rewarded, or incompletely labeled trials and crops containing a lick or reward in $[-1.00, +0.25)$ s were removed. The classifier used 1-s LFP windows spanning $[-0.75, +0.25)$ s about change onset, low-pass filtered at 50 Hz, normalized using robust per-channel statistics from an early unlabeled calibration minute, and masked where channels failed quality control. The encoder and amplitude calibration used no test-mouse behavioral labels. The excluded Familiar session of mouse 544838 (ID 1065449881) had 151 hits and only 10 misses; the other three sessions contained 724 eligible trials (449 hits, 275 misses). Table A.8 gives the session IDs and mouse-disjoint evaluation roles.

Each behavior model used the same 48,049-trainable-parameter transformer classifier and 32-D nonlinear coordinate adapter, with a shared 48-D attention-token width; the three conditions were trained *separately*, not with shared weights. CCF used three spatial coordinates and Region a six-area one-hot vector, both padded to 32 dimensions before their matched adapters. Each electrode supplied four 250-ms signal tokens to a two-layer, four-head transformer; a pooled [CLS] token predicted the outcome without mouse, session, or electrode ID embeddings. The Functional encoder was frozen throughout decoder training and final testing. Training batches balanced hit and miss trials; epoch and operating threshold were chosen on validation mice. No parameter or threshold was fitted on the two final mice. This is zero-shot transfer *in fitted weights and behavioral labels*, with unlabeled per-electrode coordinate and amplitude calibration.

Table A.8: Allen `ecephys_session_id` values by mouse and role. The first eight decoder-training mice also trained the MSC encoder; two validation mice likewise validated the encoder. Encoder-test mice were examined separately and are not part of the final behavior result. The marked session was excluded for too few misses.

Role	Mouse	Familiar session	Novel session
MSC encoder and behavior-decoder training			
	533537	1055221968	1055403683
	536480	1064415305	1064639378
	544358	1065437523	1065905010
	544456	1064400234	1064644573
	544836	1067588044	1067781390
	570301	1108334384	1108531612
	574078	1115086689	1115368723
	576323	1116941914	1117148442
Additional behavior-decoder training			
	576324	1118324999	1118512505
	577287	1124507277	1124285719
	579993	1130113579	1130349290
	585326	1128520325	1128719842
Validation (527749 and 574082 also validated the encoder)			
	527749	1053718935	1053941483
	571520	1108335514	1108528422
	574082	1118327332	1118508667
Encoder test; behavioral exploratory only			
	541234	1062755779	1063010385
	568963	1111013640	1111216934
Final behavioral test: two further unseen mice			
	530862	1052342277	1052533639
	544838	1065449881*	1065908084

Table A.9: Held-out behavioral classification across two new mice and three eligible sessions. AUROC and balanced accuracy (BA) are equal-weight means of mouse-level values; BA uses thresholds set on validation mice. All three decoders have 48,049 trainable parameters.

Coordinate input	Mouse-mean AUROC	Mouse-mean BA
Functional (frozen MSC)	**0.725**	**0.664**
CCF location	0.712	0.654
Region label	0.685	0.640

Functional coordinates gave the highest AUROC in all three eligible sessions; the mouse-mean gain over CCF was 0.013, so the comparison is descriptive with only two final mice. A separately trained Functional classifier restricted to the preceding $[-1, 0)$ s window scored AUROC 0.590 and BA 0.549 on these same mice, compared with 0.725 and 0.664 for the main crop. Prior hit/miss classification on these recordings used within-session cross-validation (Harris & Gong, 2026); its scores are not a matched comparison to our held-out-mouse evaluation. Mouse weights were trained anew rather than transferred from the human datasets.

A.14 Additional Analysis: Baselines and Associated Training Protocol

We compare our cross-region LFP predictor transformer against a set of standard time–series baselines that map *source* channels to *target* channels within each window for our 20-subject dataset. For every subject and sessions, we use identical train/validation/test splits and the same masking pipeline: all channels whose region label matches a pre-specified exclusion predicate (e.g., `VO*`) are designated as targets, while the remaining channels are sources.

Baselines.

- **CopyBest** (nonparametric). For each target channel, we select the single source channel with the highest training correlation (Z-scored across time) and fit a scalar gain by least squares; the prediction is a scaled copy of that source. This yields a strong correlation-based lower bound without temporal modeling.
- **FIR (linear, causal MIMO).** A multi-input multi-output finite impulse response filter that predicts targets from sources using a causal 1-D convolution with kernel size $L{+}1$. We set $L{=}64$ (thus receptive field 65 samples equivalent to 65ms), implemented as a single causal Conv1d from C_{src} to C_{tgt} and trained end-to-end.
- **TCN (temporal CNN).** (Bai et al., 2018) A temporal convolutional network with pointwise input/output projections and $D{=}5$ residual dilated blocks (dilations $1, 2, 4, 8, 16$), kernel size $k{=}5$, width 128, GroupNorm(1) in each block, GELU activations, and dropout 0.3. The effective receptive field approximately covers the FIR context.
- **GRU (recurrent).** A 2-layer Gated Recurrent Unit (GRU) (Cho et al., 2014) with hidden size 128 (dropout 0.3 between layers), followed by a linear projection to C_{tgt} at each time step.

Optimization and early stopping. All trainable baselines (FIR, TCN, GRU) are trained with Adam (Kingma & Ba, 2015) (learning rate $3{\times}10^{-4}$, weight decay 10^{-4}), batch size 8, for up to 150 epochs with early stopping (patience 10) based on validation MSE.

Evaluation metric. We report Fisher-$\bar{r}$ over all valid window$\times$target pairs:

$$z_i = \operatorname{arctanh}(r_i), \quad \bar{z} = \frac{1}{N}\sum_i z_i, \quad \bar{r} = \tanh(\bar{z}),$$

where r_i is the per-sample Pearson correlation (over time) between prediction and ground truth for a target channel.

Statistics. For across-subject comparisons we used paired two-sided t-tests between each baseline and our method (FUNC), applying both Holm–Bonferroni and Benjamini–Hochberg (FDR) corrections for multiple comparisons; all corrected p-values were <0.001, indicating statistically significant improvements.

Baseline comparison. Figure A.9 summarizes subject-level performance (Fisher-$\bar{r}$) for our proposed **Transformer+Functional Embedding(FUNC)** against standard time–series baselines: **FIR** (linear causal MIMO), **TCN**, **GRU**, and **CopyBest** (correlation-selected source with scalar gain). Each dot is a subject and gray lines connect the same subject across models (paired comparison). The pooled (multi-subject) **FUNC** model clearly outperforms all baselines, indicating that the transformer with functional coordinates can effectively *aggregate* cross-subject data to learn stronger source$\rightarrow$target mappings. We also include a **Single-FUNC** control (the same transformer+functional embedding trained *per subject*); as expected, it underperforms the pooled model because transformers are data-hungry and a single session in this dataset contributes only $\sim$500–900 windows—insufficient for reliable convergence. Statistical tests (paired two-sided t-tests vs FUNC with Holm and BH/FDR corrections) show all comparisons are significant ($p_{\mathrm{corr}} < 0.001$).

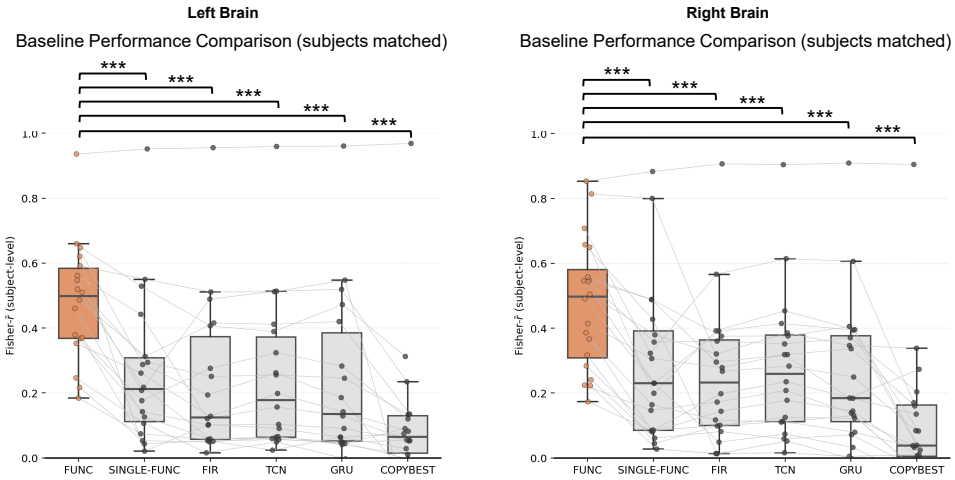


Figure A.9: *Baseline comparisons (subject-level Fisher-$\bar{r}$).* Boxes show median and IQR; dots are individual subjects; gray lines connect the same subject across models. Models: FIR (linear causal MIMO), TCN, GRU, CopyBest, and our Transformer with Functional Embedding (FUNC). The pooled (multi-subject) FUNC model outperforms all baselines. Paired t-tests vs the pooled Functional Transformer with Holm and BH/FDR corrections: all $p_{\text{corr}} < 0.001$. (***)

A.15 Additional Analysis: Do Functional Embeddings Need to Be Trained on Every Subject?

To evaluate whether the functional embedding must be trained on every subject in order to support cross-region LFP prediction, we performed a five-fold leave-four-subjects-out experiment separately for the right and left hemispheres. In each fold, the functional embedding was trained on 16 subjects while the remaining 4 subjects were excluded. The cross-region LFP prediction transformer was then trained on *all* subjects using that fold's embedding model. Thus, every subject was "held-out" exactly once and "held-in" during the remaining folds for each hemisphere.

Metric. For each subject and fold, reconstruction accuracy was measured using the Fisher-z corrected correlation coefficient r, averaged over all cross-region predictions across that subject's recording periods. Subject-level scores were aggregated across folds using an n-weighted Fisher mean, yielding one *held-in* and one *held-out* value per subject and hemisphere. We report the back-transformed correlation values r.

Statistical comparison by hemisphere. A paired t-test was used to compare held-in and held-out reconstruction accuracy within each hemisphere. For the **right hemisphere**, no significant difference was found ($p = 0.81$). The **left hemisphere** showed the same pattern, with no significant difference between held-in and held-out performance ($p = 0.703$). Across both hemispheres, held-in and held-out reconstruction scores were nearly identical, indicating that excluding individual subjects from functional-embedding training does not impair reconstruction quality for those subjects.

Results are reported in Tables A.10 and A.11 and Figure A.10.

Subject (right)	**Held-in**	**Held-out**	$\mathbf{\Delta}$ = held-in − held-out
1	0.5533	0.5369	0.0164
2	0.2362	0.2558	-0.0196
3	0.2320	0.2408	-0.0088
4	0.5383	0.5355	0.0028
5	0.6583	0.6613	-0.0030
6	0.4887	0.4964	-0.0077
7	0.5612	0.5657	-0.0045
8	0.3149	0.2934	0.0215
9	0.8520	0.8523	-0.0003
10	0.6434	0.5952	0.0482
11	0.1942	0.1886	0.0056
12	0.3861	0.4520	-0.0659
13	0.4103	0.4475	-0.0372
14	0.3671	0.3740	-0.0069
15	0.5889	0.5781	0.0108
16	0.2354	0.1874	0.0480
17	0.2890	0.2791	0.0098
18	0.5602	0.5520	0.0081
19	0.7217	0.7154	0.0064
20	0.8084	0.8049	0.0035

Table A.10: Right-hemisphere subject-level reconstruction performance when subjects were included (*held-in*) or excluded (*held-out*) from functional-embedding training. Differences are small and not systematically biased ($p = 0.81$).

Subject (left)	Held-in	Held-out	Δ = held-in − held-out
1	0.4120	0.4293	-0.0173
2	0.1816	0.1891	-0.0075
3	0.5660	0.5691	-0.0032
4	0.3568	0.3693	-0.0125
5	0.4879	0.4998	-0.0119
6	0.3660	0.3572	0.0088
7	0.4809	0.4921	-0.0111
8	0.6433	0.6591	-0.0158
9	0.5731	0.5761	-0.0029
10	0.6495	0.6575	-0.0080
11	0.1948	0.1901	0.0047
12	0.5762	0.5731	0.0031
13	0.2140	0.2326	-0.0185
14	0.3702	0.3533	0.0169
15	0.5218	0.4722	0.0496
16	0.3624	0.3192	0.0433
17	0.4737	0.5297	-0.0560
18	0.9316	0.9315	0.0001

Table A.11: Left-hemisphere subject-level reconstruction performance when subjects were included (*held-in*) or excluded (*held-out*) from functional-embedding training. As in the right hemisphere, differences are small and not systematically biased ($p = 0.703$).

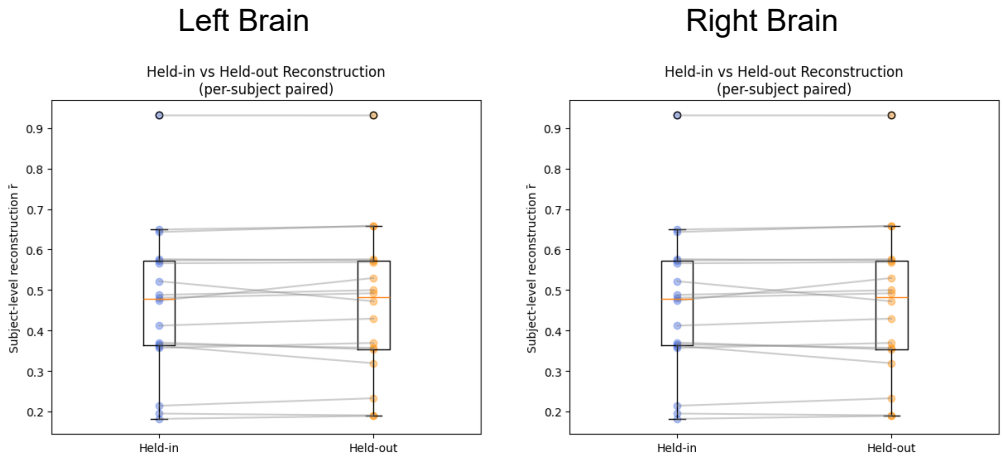


Figure A.10: *Held-in vs. held-out reconstruction performance for right and left hemispheres.* Each line connects the held-in and held-out value for a single subject. For both hemispheres, held-in and held-out distributions almost completely overlap, consistent with the non-significant differences reported in Tables A.10 and A.11.

A.16 Additional Analysis: Band–Power Prediction Task

Motivation To further evaluate whether functional coordinates provide improved decoding performance beyond the cross-region LFP prediction task, we introduce an additional decoding task with direct clinical relevance: *predicting per–window spectral band power*. Band–power is widely used in clinical neurophysiology and brain–computer interfaces because it reflects physiologically meaningful oscillatory activity (e.g., alpha, beta, and gamma rhythms). Prediction of band–power from distributed neural activity therefore represents a practical and interpretable downstream application for our learned coordinate representations.

Task Description For each 1-second input window and each held–out target channel, the model is trained to predict the *relative band–power* of the target channel in three clinically relevant frequency ranges:

- **Alpha**: 8–13 Hz
- **Beta**: 13–30 Hz
- **Low Gamma**: 30–50 Hz

Relative band–power for a given band b was defined as

$$p_b = \frac{P_b}{\sum_i P_i},$$

where P_b is the absolute power in band. This normalization constrains values to the interval $[0, 1]$, improves regression stability, and aligns with common clinical practice.

Lightweight Band–Power Transformer Because each target channel yields only a single scalar per band (rather than a full waveform), we use a lightweight version of our transformer architecture. Specifically:

- model dimension reduced from `d_model` $= 128$ to 64,
- number of attention heads reduced from 8 to 4,
- feed–forward hidden dimension reduced from 384 to 128,

which lowers the model size to $\sim$ 330k parameters. The prediction head is a single linear layer followed by a softmax over the three bands, ensuring non–negative outputs that sum to one.

All other components, including the functional coordinate input tokens, remain unchanged. As in the main experiments, we compare two coordinate systems:

1. **Region label coordinates**: regions labels associated with each channel.
2. **MNI coordinates**: standard stereotactic coordinates.
3. **Functional coordinates (Func)**: the learned functional embedding described in the main text.

Evaluation We evaluate the band–power transformer across subjects and periods separately in the *left* and *right* hemispheres. For each subject and each band, we compute the Pearson correlation coefficient r between the predicted and true relative band–power across all windows in the test set. Subject-level performance is obtained by averaging r across periods. Group-level statistics are computed across subjects.

Statistical analysis. For each frequency band, we perform paired comparisons across subjects between the functional-coordinate model (`Func`) and the two baseline coordinate systems (`MNI` and `Region`). Because subject-level decoding correlations are paired and not guaranteed to follow a normal distribution, we use a two-sided paired **Wilcoxon signed–rank test** as the primary statistical test. Positive differences (`Func`–baseline) indicate improved decoding performance using functional coordinates. Results are summarized in Figure A.11.

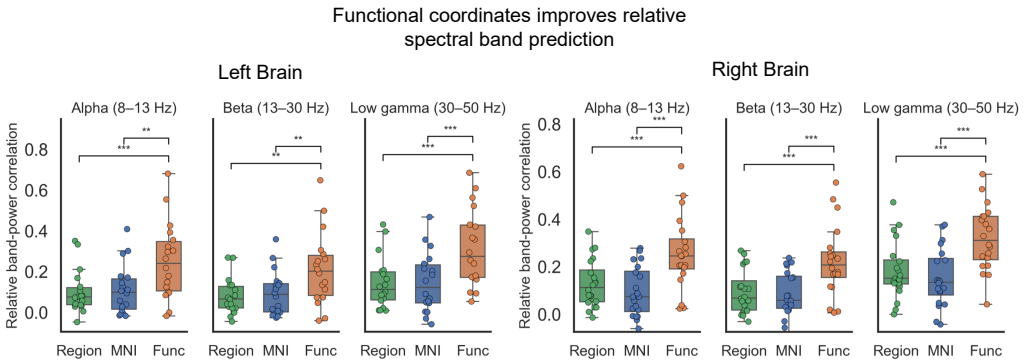


Figure A.11: **Relative spectral band prediction across coordinate systems.** Subject-level prediction correlation (r) for relative band-power decoding using three coordinate systems: region labels (`Region`), anatomical coordinates (`MNI`), and learned functional coordinates (`Func`). Columns correspond to alpha (8–13 Hz), beta (13–30 Hz), and low gamma (30–50 Hz) bands, while rows correspond to left and right hemispheres. Boxplots summarize subject-level mean correlations across recording segments, with individual points representing subjects. Across all frequency bands and both hemispheres, the functional coordinate model consistently achieves higher decoding performance than both region labels and MNI coordinates. Statistical significance was assessed using paired Wilcoxon signed-rank tests across subjects. Significance annotations indicate pairwise comparisons between `Func` and the corresponding baseline models (`Region` or `MNI`), where $^{**}p < 0.01$ and $^{***}p < 0.001$.

Results. Across both hemispheres, the functional coordinate model (`Func`) consistently outperformed both anatomical MNI coordinates and discrete region labels in relative band-power prediction (Figure A.11).

In the left hemisphere ($n = 18$ subjects), `Func` achieved substantially higher decoding correlations across all frequency bands, with mean improvements ranging from $\Delta r = 0.12$ to 0.17 relative to `MNI`, and from $\Delta r = 0.13$ to 0.17 relative to `Region`. All pairwise comparisons were significant under paired Wilcoxon signed-rank tests ($p < 0.003$).

Similarly, in the right hemisphere ($n = 20$ subjects), `Func` consistently yielded the strongest decoding performance across alpha, beta, and low-gamma bands. Mean improvements ranged from $\Delta r = 0.14$ to 0.17 relative to `MNI`, and from $\Delta r = 0.13$ to 0.14 relative to `Region`, with all comparisons remaining significant ($p < 1.1 \times 10^{-4}$).

Notably, improvements were observed consistently across all tested spectral bands and both hemispheres, suggesting that functional coordinates capture physiologically meaningful structure relevant for clinically interpretable spectral decoding tasks.

A.17 COMPUTE RESOURCES AND TRAINING TIMES

All models were trained on a single workstation equipped with an AMD Ryzen Threadripper PRO 7955WX (16 cores), 128 GiB RAM, and $1\times$ NVIDIA RTX 4500 Ada (24 GiB VRAM). Table A.12 summarizes training times and data scale for each model. Notably, the *Functional encoder (PSC, multi-subject)* is pair-based: given n input samples, it can form $\mathcal{O}(n^2)$ pairs, enabling large-scale training (here, $2{,}500{,}000$ pairs). In contrast, *Functional encoder (MSC, multi-subject)* computes its loss over full sample batches (not pairwise), so the effective sample count remains at $65{,}000$ despite the same source dataset.

Table A.12: *Training times and data scale.* Reported on a single workstation (AMD Ryzen Threadripper PRO 7955WX, 128 GiB RAM, $1\times$ NVIDIA RTX 4500 Ada 24 GiB).

Model	**Epochs**	**Samples**	**Time**
Functional encoder (single-subject)	100	20,000	$\sim$7 minutes
Functional encoder (PSC, multi-subject)	30	2,500,000	$\sim$5 hours
Functional encoder (MSC, multi-subject)	200	65,000	$\sim$1.5 hours
Cross-region predictor transformer	150	27,000	$\sim$2.5 hours