SPIRE: SHARED-PRIVATE REPRESENTATION LEARNING FOR CONTINUOUS MULTI-REGION LFP UNDER PERTURBATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning shared and region-specific representations from continuous multi-region neural signals is challenging when observations are nonlinear, autocorrelated, spectrally structured, and temporally misaligned. These challenges are particularly pronounced in local field potentials (LFPs), where separating activity that is coordinated across regions from activity specific to individual regions is important for studying distributed neural systems and their response to intervention. We introduce **SPIRE** (Shared-Private Inter-Regional Encoder), a nonlinear, time-resolved framework for learning shared and private representations from continuous multi-region LFP. SPIRE combines region-specific sequence encoders with constrained temporal correspondence, shared-only reconstruction, and variance regularization for shared/private separation. On LFP-oriented synthetic benchmarks with known ground-truth factors, SPIRE preferentially assigns shared and private information to the intended latent spaces under nonlinear observation mappings, time-varying delays, and spectral confounding. On held-out human GPi-STN LFP, nonlinear prediction reveals strong transfer of shared representations across regions, whereas cross-region prediction from private representations remains at a structured temporal null, supporting functional rather than strict statistical separation. We then freeze representations learned from off-stimulation recordings and apply them to intracranial deep brain stimulation (DBS) data. Stimulation induces larger changes and carries more condition information in shared than private representations. A matched CTAE analysis reproduces the qualitative shared-over-private effect while retaining substantially less stimulation information. Together, these results establish SPIRE as a shared/private representation-learning framework for continuous multi-region LFP and show that DBS-related changes are preferentially expressed in representations shared across the recorded regions.

1 INTRODUCTION

Learning low-dimensional representations from multiple neural populations requires separating structure that is shared across regions from structure that is specific to each region. This becomes difficult for continuous neural signals, where observations can be strongly autocorrelated, nonlinear, spectrally structured, and temporally misaligned. In such settings, high reconstruction quality alone does not establish that learned "shared" and "private" representations actually capture transferable and region-specific information.

Local field potentials (LFPs) provide an important instance of this problem. LFPs contain continuous oscillatory and broadband activity, strong temporal autocorrelation, nonstationarity, and substantial $1/f$ structure (Bedard et al., 2006; He, 2014; Dupré la Tour et al., 2017). Related processes may also appear differently across recording sites because of region-specific nonlinear mixing and temporal offsets. These properties can confound shared/private decomposition, for example by assigning dominant low-frequency or high-variance activity to the shared space even when it is not truly cross-regional.

A reliable decomposition is particularly useful for studying deep brain stimulation (DBS). DBS acts on distributed basal ganglia-thalamo-cortical circuits (Benabid, 2003; Lozano et al., 2019; Larsh et al.,

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.

2021; McIntyre & Hahn, 2010; Horn & Fox, 2020; Yang et al., 2021), yet many intracranial analyses characterize stimulation through local quantities such as spectral power or evoked responses (Dale et al., 2022; Milosevic et al., 2018; Schmidt et al., 2020; Anonymous, 2025; 2024b). During ongoing stimulation, population-level spike measurements can be difficult to obtain reliably because stimulation artifacts, amplifier blanking or saturation, and disrupted spike detection and sorting can obscure spiking activity. Multi-region LFP provides a complementary population-level signal that is particularly useful for asking whether DBS-related changes are expressed primarily in activity specific to individual regions or in activity coordinated across them.

Existing latent-variable methods address parts of this problem. Linear approaches such as CCA and DLAG can identify shared structure across neural populations (Bach & Jordan, 2005; Gokcen et al., 2022), but their linear or linear-Gaussian assumptions can be restrictive for continuous LFP. More recent nonlinear approaches, particularly CTAE, learn time-resolved shared and private neural representations using coupled autoencoders, latent alignment, and cross-region reconstruction (Sristi et al., 2025). However, it remains unclear whether shared/private representations retain their intended functional roles under the nonlinear mixing, temporal variability, and spectral confounding characteristic of continuous multi-region LFP. Evaluating only reconstruction or intended-factor recovery is insufficient to establish that information has been preferentially assigned to the correct latent space.

We introduce **SPIRE** (Shared-Private Inter-Regional Encoder), a nonlinear, time-resolved framework for shared/private representation learning in continuous multi-region LFP. SPIRE uses region-specific sequence encoders and decoders to produce shared and private latent trajectories, with constrained temporal correspondence and shared-only reconstruction encouraging transferable cross-regional structure, and cross-covariance and variance regularization promoting functional separation between the two spaces.

We evaluate SPIRE on LFP-oriented synthetic benchmarks spanning linear, nonlinear, temporally misaligned, and spectrally confounded regimes, using complete intended-versus-crossed recovery to test whether shared and private information is preferentially assigned to the correct latent spaces. We then validate this distinction on held-out human GPi-STN LFP using nonlinear cross-region prediction against a structured temporal null. Finally, we freeze representations learned from off-stimulation recordings and use them to test whether DBS preferentially reorganizes shared or private activity, with a matched CTAE analysis assessing whether the resulting stimulation effects generalize across shared/private models.

Our main contributions are:

1. **Shared/private representation learning for continuous multi-region LFP.** We introduce SPIRE, combining nonlinear region-specific sequence models, constrained temporal correspondence, shared-only reconstruction, and variance constraints for shared/private separation.
2. **Direct validation of functional shared/private separation.** We evaluate complete intended-versus-crossed recovery under synthetic ground truth and nonlinear cross-region predictability in held-out human LFP, showing that shared representations preferentially retain transferable structure while private representations retain predominantly region-specific structure.
3. **Application to distributed effects of human DBS.** DBS produces larger representational changes and more strongly encodes stimulation condition in shared than private latents. A matched CTAE analysis reproduces the direction of this result, while SPIRE retains substantially more stimulation information.

2 Related work

Multi-region latent-variable models. Classical latent models such as GPFA (Yu et al., 2008), CCA (Bach & Jordan, 2005), and DLAG (Gokcen et al., 2022) provide principled low-dimensional descriptions of neural population activity, with DLAG explicitly separating across-population and within-population latent structure. Their linear or linear-Gaussian observation assumptions are well matched to some neural population datasets but can be restrictive for spectrally rich LFP. Nonlinear and dynamical approaches including LFADS (Pandarinath et al., 2018), MARBLE (Gosztolai et al., 2025), and MRDS-IR (Dowling & Savin, 2025) address complementary questions such as latent trajectory modeling or inter-region dynamics. SPIRE instead targets descriptive time-resolved

shared/private representation learning; it does not estimate a state-transition law, directed communication, unobserved inputs, or future responses to unseen perturbations. Other related approaches include MR-SDS (Karniol-Tambour et al., 2024), and MR-LFADS (Liu et al., 2025), which target switching multi-region dynamics, and inter-regional communication with latent inputs, respectively.

Shared/private nonlinear representation learning. Shared/private factorization is common in multimodal representation learning (Yi et al., 2025; Shi et al., 2019; Lee & Pavlovic, 2021). CTAE (Sristi et al., 2025) is the closest neural baseline because it learns time-resolved shared and region-specific latent trajectories with transformer-based coupling, cross-reconstruction, alignment, and orthogonality. SPIRE differs in using region-specific recurrent encoders/decoders; near-identity temporal alignment followed by controlled linear mappings; variance guards for both branches; shared-self and aligned cross-shared reconstruction; and staged private activation/freezing. The contribution is not any single module in isolation, but the combined formulation and its validation under nonlinear, delayed, and spectrally confounded LFP conditions.

LFP and perturbation analysis. LFP reflects complementary aspects of neural population activity and poses signal-processing challenges distinct from spikes (Bedard et al., 2006; He, 2014; Erturk et al., 2025). During DBS, simultaneous multi-region LFP is particularly useful for examining distributed network effects despite stimulation-related recording constraints (Alarie et al., 2022). Recent work further highlights that spiking and LFP signals capture complementary aspects of neural activity and require distinct modeling approaches (Erturk et al., 2025). Our perturbation analysis therefore learns the representation without stimulation recordings or labels, freezes it, and measures off-to-on changes in the resulting coordinate system. This design avoids circularly shaping the representation toward the downstream stimulation labels.

3 SPIRE FOR CONTINUOUS MULTI-REGION LFP

3.1 MODEL OVERVIEW

We introduce **SPIRE** (Shared-Private Inter-Regional Encoder), a nonlinear, time-resolved framework that learns shared (cross-regional) and private (region-specific) latent representations from multi-region neural recordings. Figure 1 shows a schematic of SPIRE for a two-region example (GPi and STN).

For each region $r \in \mathcal{R}$, SPIRE instantiates an encoder–decoder pair. Given multichannel input $x^{(r)} \in \mathbb{R}^{B\times T\times C_r}$ (batch B, length T, C_r channels), a recurrent encoder produces hidden states $h^{(r)}$, which are linearly projected into shared and private latent sequences ($d_{\text{sh}}, d_{\text{pr}}$ are shared and private latent dimensionalities):

$$z_{\text{sh}}^{(r)} = W_{\text{sh}}^{(r)} h^{(r)} \in \mathbb{R}^{B\times T\times d_{\text{sh}}}, \qquad z_{\text{pr}}^{(r)} = W_{\text{pr}}^{(r)} h^{(r)} \in \mathbb{R}^{B\times T\times d_{\text{pr}}}, \tag{1}$$

Each region is reconstructed from its own latents:

$$\hat{x}^{(r)} = f_{\text{dec}}^{(r)}\Big([\, z_{\text{sh}}^{(r)},\, z_{\text{pr}}^{(r)}\,]\Big) \in \mathbb{R}^{B\times T\times C_r}. \tag{2}$$

The terms "shared" and "private" are operational: shared structure should preferentially transfer across the modeled regions, whereas private structure should preferentially remain region specific. They are not assumed to be statistically independent biological sources.

Constrained temporal correspondence. Shared trajectories from region s are mapped into region r as

$$\tilde{z}_{\text{sh}}^{(s\to r)} = M^{(s\to r)}\,\text{ConvAlign}^{(s\to r)}(z_{\text{sh}}^{(s)}), \tag{3}$$

where the depthwise temporal convolution and linear mapper are initialized near identity and regularized to remain close to identity. This permits local temporal/subspace adjustment without allowing an unconstrained transformation to define the shared space.

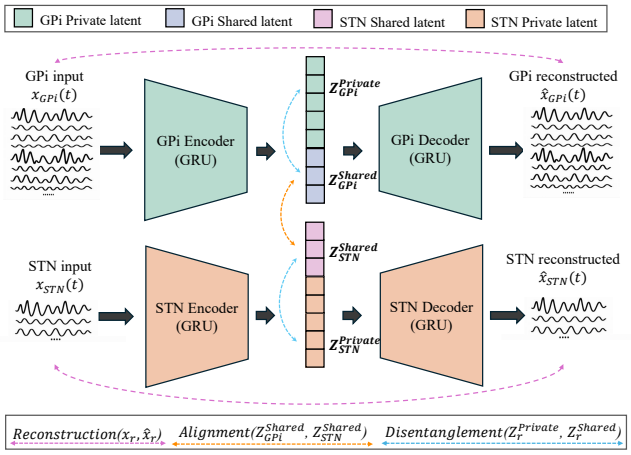


Figure 1: SPIRE separates each region into time-resolved shared and private latent trajectories. Shared trajectories are aligned across regions through constrained temporal and linear mappings, while shared/private separation and reconstruction objectives preserve complementary information.

3.2 Training objective

The objective has three conceptual families,

$$\mathcal{L} = \lambda_{\text{rec}}\mathcal{L}_{\text{rec}}^{\text{family}} + \lambda_{\text{align}}\mathcal{L}_{\text{align}}^{\text{family}} + \lambda_{\text{dis}}\mathcal{L}_{\text{dis}}^{\text{family}}. \tag{4}$$

The reconstruction family contains full reconstruction, shared-self reconstruction, and aligned cross-region shared reconstruction. The alignment family uses variance–invariance–covariance regularization objective (VICReg, Bardes et al., 2022) between native and aligned shared trajectories together with near-identity mapper/ConvAlign regularization. Explicitly,

$$\mathcal{L}_{\text{VICReg}}(A, B) = \lambda_{\text{inv}}\mathcal{L}_{\text{inv}} + \lambda_{\text{var}}\mathcal{L}_{\text{var}} + \lambda_{\text{cov}}\mathcal{L}_{\text{cov}}, \tag{5}$$

where invariance aligns corresponding trajectories, the variance term prevents collapse, and the covariance term discourages redundant dimensions. The disentanglement family penalizes standardized shared-private cross-covariance and includes shared/private variance guards. Full equations, coefficient definitions, and optimization schedules are provided in Appendix A.1.1. The objective contains a fixed set of coefficient types: region-wise terms are averaged across regions and pairwise terms across directed region pairs, so adding a region adds modules and computation but not pair-specific loss-weight types.

3.3 Implementation and evaluation protocol

Encoders/decoders are GRU sequence models trained with Adam, gradient clipping, learning-rate scheduling, and early stopping. Short lag-stacked inputs provide local within-region context, while ConvAlign acts on encoded shared trajectories across regions; all compared real-data models receive identical lag augmentation. Detailed preprocessing, schedules, loss ablations, and a no-lag ablation are given in Appendix A.1.2, Appendix A.2.5 and Appendix A.4.

For human DBS analyses, SPIRE is trained and selected per hemisphere using off-stimulation data only, then frozen. Stimulation recordings are used only for downstream displacement and decoding analyses. Appendix A.6 verifies that the frozen representation remains reconstructive and functionally organized under stimulation.

SPIRE is implemented and evaluated in PyTorch (Paszke, 2019), statistical analyses were performed in R (R Core Team, 2025) using RStudio (Posit team, 2025).

4 Controlled LFP-oriented validation

A key advantage of synthetic data is that the true shared and private latent processes are known. This allows us to test directly both whether a model recovers the underlying factors and whether it

Table 1: Nonlinear recovery of known ground-truth factors. Each cell reports intended shared/private recovery, $R^2(Z_{\mathrm{sh}} \to S_{\mathrm{GT}})$ / $R^2(Z_{\mathrm{pr}} \to P_{\mathrm{GT}})$, averaged across four matched seeds. Higher values indicate stronger recovery of the corresponding ground-truth factors. Full means and standard deviations are reported in Appendix A.4.1.

Regime	SPIRE	DLAG	CTAE
D0 linear	0.72 / 0.64	**0.89 / 0.83**	0.29 / 0.31
D1 nonlinear	**0.76 / 0.60**	0.64 / 0.48	0.14 / 0.17
D2 + delay	**0.75 / 0.58**	0.57 / 0.48	0.12 / 0.16
D3 spectral confound	**0.54 / 0.55**	0.42 / 0.46	0.17 / 0.35

assigns them to the intended shared or private representation. We generated four regimes with three shared and three private latent factors per region. D0 is an approximately linear positive control; D1 adds nonlinear region-specific warping and nonlinear observation mixing; D2 additionally introduces time-varying inter-region delay; and D3 deliberately makes strong private structure low-frequency while shared structure occupies higher-frequency bands, creating an LFP-like spectral confound. The generator additionally includes colored noise, autocorrelation, common-mode structure, and bipolar rereferencing; full generation details and parameter settings are provided in Appendix A.2.1. We compare SPIRE with DLAG (Gokcen et al., 2022) and CTAE (Sristi et al., 2025) using four matched seeds (implementation details mentioned in Appendix A.2.2, A.2.3 and A.2.4).

Ground-truth recovery. Let S_{GT} denote the ground-truth shared latent factors used to generate activity across regions, and let P_{GT} denote the ground-truth private factors specific to the evaluated region. Similarly, Z_{sh} and Z_{pr} denote the corresponding learned shared and private representations. We measure nonlinear recovery using Extra Trees regression (Geurts et al., 2006) with five-fold trial-wise cross-validation and report the held-out coefficient of determination, R^2.

We first evaluate *intended recovery*: shared recovery, $R^2(Z_{\mathrm{sh}} \to S_{\mathrm{GT}})$, and private recovery, $R^2(Z_{\mathrm{pr}} \to P_{\mathrm{GT}})$. Table 1 shows that DLAG performs best in the linear D0 positive control, whereas SPIRE achieves the strongest intended recovery in each nonlinear regime.

Figure 2 provides a representative visualization from D3, the most spectrally confounded regime. SPIRE tracks both shared and private ground-truth trajectories closely, whereas the baseline representations show larger deviations.

Intended-versus-crossed recovery. High intended recovery alone is not sufficient to establish a meaningful shared/private decomposition: for example, a learned shared representation could recover the true shared factors while also retaining substantial private information. We therefore evaluate the complete 2×2 recovery matrix.

We define

$$SS = R^2(Z_{\mathrm{sh}} \to S_{\mathrm{GT}}), \qquad SP = R^2(Z_{\mathrm{sh}} \to P_{\mathrm{GT}}), \tag{6}$$

$$PP = R^2(Z_{\mathrm{pr}} \to P_{\mathrm{GT}}), \qquad PS = R^2(Z_{\mathrm{pr}} \to S_{\mathrm{GT}}). \tag{7}$$

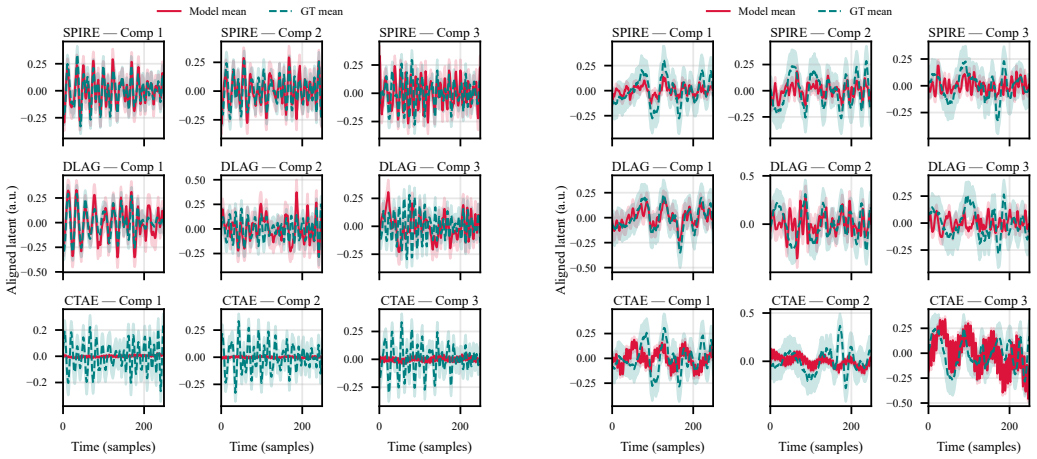
Here, SS and PP measure recovery of the factors intended for each branch, whereas SP and PS quantify crossed recovery: private information recoverable from the learned shared space and shared information recoverable from the learned private space, respectively.

We summarize branch specificity as

$$\Delta_{\mathrm{sh}} = SS - SP, \qquad \Delta_{\mathrm{pr}} = PP - PS. \tag{8}$$

Positive Δ_{sh} therefore means that the learned shared representation predicts true shared factors better than true private factors, while positive Δ_{pr} means that the learned private representation preferentially predicts true private factors. Table 2 reports these specificity margins; the full 2×2 recovery matrices with standard deviations are given in Appendix A.4.1.

The two analyses provide complementary evidence. D0 acts as a positive control: in the approximately linear regime, which is closest to DLAG's linear-Gaussian assumptions, DLAG achieves both the strongest ground-truth recovery and the strongest shared/private specificity. In the nonlinear D1 and



(a) Shared latent recovery.

(b) Private latent recovery.

Figure 2: Representative latent recovery in the challenging D3 synthetic regime. Recovered trajectories (solid red) are compared with ground-truth trajectories (dashed teal) after CCA-based alignment for visualization. D3 combines nonlinear observation mappings with strong low-frequency private structure and higher-frequency shared factors, producing substantial spectral confounding. Shaded regions denote SEM across trials. Quantitative nonlinear recovery is reported in Table 1, and branch specificity in Table 2.

Table 2: Nonlinear shared/private specificity. Each cell reports $\Delta_{\mathrm{sh}}/\Delta_{\mathrm{pr}}$ (mean across four matched seeds), where $\Delta_{\mathrm{sh}} = SS - SP$ and $\Delta_{\mathrm{pr}} = PP - PS$. Positive values indicate preferential assignment of information to the intended latent branch.

Regime	SPIRE	DLAG	CTAE
D0 linear	0.48 / 0.40	**0.81 / 0.68**	0.07 / -0.01
D1 nonlinear	**0.44 / 0.39**	0.43 / 0.20	0.02 / -0.01
D2 + delay	**0.42 / 0.38**	0.37 / 0.16	0.04 / -0.01
D3 spectral confound	**0.42 / 0.27**	0.15 / 0.13	-0.06 / 0.14

delayed D2 regimes, SPIRE achieves the strongest intended recovery while maintaining positive specificity in both branches; its private specificity is substantially larger than DLAG's in both settings. In the challenging D3 regime, SPIRE achieves shared/private specificity margins of $0.42/0.27$, compared with $0.15/0.13$ for DLAG and $-0.06/0.14$ for CTAE. Because D3 deliberately places the shared factors at higher frequencies than the dominant private factors, this result argues against the explanation that SPIRE obtains its decomposition simply by assigning slower LFP structure to the shared branch. Remaining recovered trajectories of D0-D2 regimes, together with complementary CCA-based recovery analyses, are shown in Appendix A.2.6.

5 Functional shared/private organization in human LFP

We analyzed intracranial LFP from ten pediatric dystonia patients with recordings from GPi, STN, and, in a subset, thalamus (Anonymous, 2019; 2024a; 2018). The principal analyses use paired GPi-STN recordings from 17 hemispheres; stimulation analyses use 17 GPi-stimulation hemispheres and 13 STN-stimulation hemispheres. Signals were band-limited to suppress stimulation artifacts, bipolar rereferenced, and segmented into 0.5-s windows. Models were fit separately by hemisphere because channel count, electrode placement, anatomy, and recording duration differ across subjects and hemispheres. Shared and private latent dimensions were selected from a predefined grid using reconstruction performance and variance partitioning criteria. Dataset details, latent-dimension selection, and training details are provided in Appendix A.3.1-A.3.3 and Appendix A.3.5.

Table 3: Held-out nonlinear organization of human GPi-STN LFP. Cross-region shared prediction is strongly above the structured temporal null, whereas cross-region private prediction is approximately at the null. Values are hemisphere-level mean excess R^2 with 95% confidence intervals.

Prediction category	Excess R^2	95% CI
Shared across regions	**0.636**	[0.507, 0.764]
Shared → private, within region	0.504	[0.376, 0.633]
Private → shared, within region	0.312	[0.183, 0.440]
Private across regions	**0.009**	[-0.120, 0.138]

5.1 Nonlinear cross-region predictability defines functional segregation

True biological shared/private factors are unobserved in human data, so we use held-out nonlinear predictability to test the operational role of each branch. Extra Trees regressors were evaluated with five-fold cross-validation holding out complete 0.5-s windows. A structured temporal null reassigns complete target windows and circularly shifts trajectories, preserving marginal distributions and temporal autocorrelation. We report excess R^2 relative to this null.

As shown in Table 3, shared representations transfer strongly across GPi and STN, with mean excess $R^2 = 0.636$, whereas cross-region prediction from private representations remains near the structured null ($R^2 = 0.009$, $p = 0.443$). Within each region, shared and private representations retain substantial nonlinear predictability in both directions, indicating that the learned branches are not statistically independent.

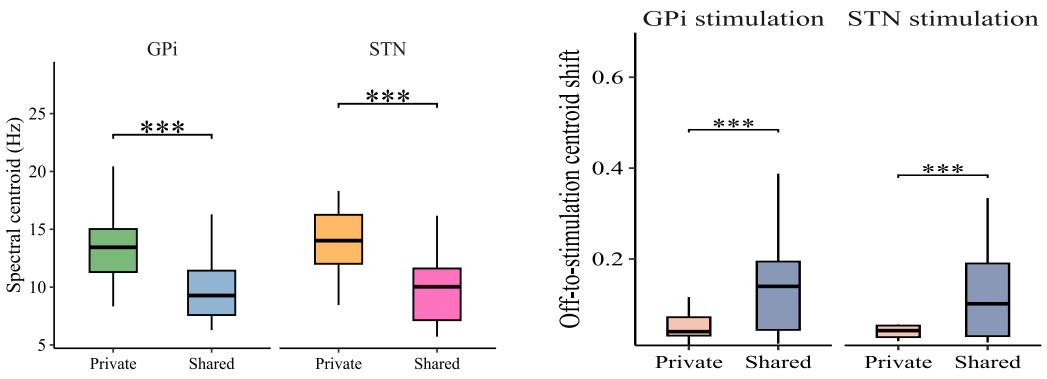
Together with the synthetic ground-truth recovery and specificity results in Tables 1 and 2, these findings support a narrower interpretation of *functional segregation*: shared representations preferentially capture information that transfers between recorded regions, whereas private representations remain predominantly region specific. Complementary human-data comparisons with CTAE using held-out reconstruction and CCA-based latent organization are reported in Appendix A.5.

Shared and private representations exhibit distinct spectral structure. Beyond cross-region transferability, the learned representations also differ in their temporal organization. Shared latents exhibit significantly lower spectral centroids than private latents in both GPi and STN (Fig. 3a), indicating that the latent structure that transfers across regions in these recordings is associated with slower spectral content than the predominantly region-specific structure. Importantly, this separation is not imposed by the model: in the D3 synthetic benchmark, shared ground-truth factors were deliberately placed at higher frequencies than the dominant private factors, yet SPIRE still preferentially recovered the correct shared and private factors. Thus, the spectral difference observed in human LFP reflects structure in the data rather than a generic low-frequency bias of the learned decomposition.

6 DBS-related changes are preferentially expressed in the shared representation

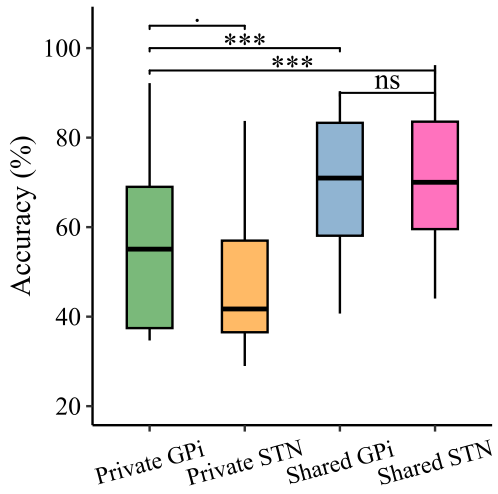
Having established the functional roles of the two latent spaces on held-out off-stimulation data, we freeze each selected model and analyze unseen stimulation recordings. Additional robustness analyses, including train-off/test-off/test-on evaluation, are reported in Appendix A.6. We quantify off-to-on centroid displacement and decode stimulation frequency from each latent subset with Random Forest classifiers (Breiman, 2001).

Shared latents exhibit significantly larger off-to-on displacement than private latents for both GPi and STN stimulation (Figure 3b). For GPi stimulation, classifiers distinguish Off, 85, 185, and 250 Hz across 17 hemispheres; for STN stimulation, they distinguish Off, 85, and 185 Hz across 13 hemispheres. Shared latents provide consistently higher decoding accuracy than private latents across regions and stimulation sites (Figures 3c and 3d). Thus stimulation-related information is preferentially expressed in the learned shared representation rather than only in the learned region-specific residual representation.

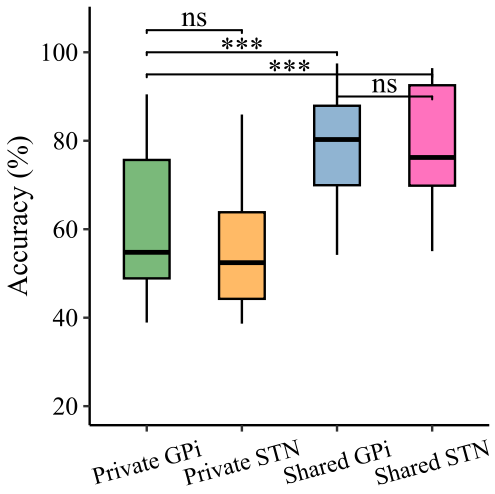


(a) Spectral organization of latent representations.

(b) Off-to-on latent displacement.



(c) GPi stimulation decoding.



(d) STN stimulation decoding.

Figure 3: Functional organization and DBS-induced changes in SPIRE representations. (a) Shared and private latent trajectories exhibit distinct spectral structure, with shared latents showing lower spectral centroids in human GPi-STN LFP. (b) DBS produces larger off-to-on displacement in shared than private representations. (c,d) Shared representations encode stimulation condition more strongly than private representations under both GPi and STN stimulation. Statistical significance: $*: p < 0.05$, $**: p < 0.01$, $***: p < 0.001$; ns, not significant.

6.1 Matched CTAE analysis

To test whether the qualitative neuroscience conclusion depends on SPIRE, we repeated the GPi-stimulation decoding analysis with CTAE using the same 17 hemispheres, preprocessing, lag-augmented inputs, matched latent dimensionalities, stimulation tensors, and classifier procedure. CTAE independently reproduces shared $>$ private decoding: its mean shared accuracy is 38.4% versus 31.0% private, a 7.41-point difference ($F_{1,48} = 44.53$, $p = 2.34 \times 10^{-8}$). However, SPIRE retains substantially more stimulation information in both spaces (Table 4, details in A.7).

This agreement is useful: the shared-over-private stimulation effect is not unique to SPIRE. At the same time, SPIRE retains substantially more stimulation-related information under this decoding evaluation. The difference-of-differences between the models' shared-minus-private gaps is not significant ($p = 0.076$), so we do not claim that SPIRE changes the qualitative conclusion; rather, it preserves considerably more stimulation information while recovering the intended shared/private roles more reliably in the LFP-oriented validation.

Table 4: Matched GPi-stimulation decoding comparison. Accuracies are averaged across regions after first averaging stimulation-contact settings within hemisphere. Both SPIRE-CTAE differences have Holm-adjusted $p < 0.0001$.

Latent class	CTAE	SPIRE	Difference
Private	31.0%	53.6%	+22.6 points
Shared	38.4%	67.7%	+29.3 points

6.2 Extension to three recorded regions.

Although the primary analyses focus on GPi-STN recordings, we also instantiated SPIRE jointly on GPi, STN, and thalamic LFP. The three-region model preserved the expected latent organization, with shared representations showing the strongest cross-region CCA similarity and private representations the weakest. Shared latents from all three regions also encoded GPi stimulation frequency more strongly than their corresponding private latents. These results demonstrate practical extension beyond two regions, while not implying unrestricted many-region scalability. Implementation details and full three-region analyses are reported in Appendix A.1.3 and Appendix A.3.4, respectively.

7 Discussion

SPIRE is a time-resolved shared/private representation model for continuous multi-region LFP. The strongest evidence for its decomposition comes from two complementary tests: complete intended-versus-crossed recovery when ground truth is known, and nonlinear cross-region transfer in held-out human LFP. These tests, together with the distinct spectral organization observed in human LFP, support a narrower and more defensible interpretation than strict disentanglement: the shared branch preferentially represents information transferable between recorded regions, whereas the private branch preferentially represents nontransferable regional structure.

Applied to human DBS recordings, this organization reveals a consistent network-level pattern. Shared trajectories shift more strongly under stimulation and encode stimulation condition more accurately than private trajectories. The matched CTAE analysis reaches the same qualitative conclusion, which suggests that the qualitative shared-over-private pattern is not unique to SPIRE, while SPIRE preserves substantially more stimulation information. These observations are consistent with network perspectives on DBS (McIntyre & Hahn, 2010; Horn & Fox, 2020) and with theories emphasizing distributed coordination in basal ganglia-thalamo-cortical circuits (Galvan et al., 2015; Jinnah & Hess, 2006; Obeso et al., 2008; Zhuang et al., 2004).

Several limitations define the scope of the claims. First, "shared" means shared among the *recorded* regions. Unobserved regions or common inputs may influence both GPi and STN, so shared activity cannot be interpreted as direct causal communication. ConvAlign establishes temporal correspondence, not causal direction or information flow. Second, SPIRE learns latent trajectories rather than an explicit dynamical transition law and is not intended for future prediction or simulation of unseen perturbations. Third, the current formulation uses one shared and one private branch per region and does not explicitly separate globally shared from subnetwork-specific factors; with heterogeneous larger recordings, subnetwork structure may be split across the two spaces. Fourth, directed pairwise aligners scale as $O(|\mathcal{R}|^2)$, making the present implementation best suited to a modest number of regions (A.9). Fifth, models are hemisphere specific; a common cross-subject latent space remains an unvalidated extension. Finally, SPIRE requires dataset-level choices for latent dimensionality, regularization strengths, and optimization settings. Appendix A.8 provides a preliminary application to a four-animal CA1-PFC dataset after dataset-specific adaptation, while Appendix A.4 reports no-lag, dimensionality, and alternative-objective robustness analyses.

Overall, the results position SPIRE as a practical shared/private representation tool for continuous multiregion LFP rather than a causal or generative dynamical system. The DBS application illustrates how such a decomposition can separate perturbation-related changes expressed across recorded regions from changes that remain region specific.

AI USE STATEMENT

We used OpenAI's ChatGPT as an assistive tool during implementation, analysis, and manuscript preparation. The research question, model design, objectives, experimental analyses, and scientific conclusions were developed and determined by the authors. Generative AI assisted with implementing author-specified model and analysis code; coding the synthetic-data generator from LFP characteristics specified by the authors; generating plotting and analysis scripts; suggesting additional robustness checks; and providing secondary feedback on the interpretation and presentation of results. It was also used to summarize selected literature identified by the authors and to assist with manuscript drafting, editing, organization, and phrasing.

Generative AI was not used for collection or preprocessing of the human neural recordings. All AI-assisted code and text incorporated into the work were reviewed and verified by the authors, who made the final methodological and interpretive decisions and take full responsibility for the content of the paper.

ETHICS STATEMENT

The human intracranial recordings analyzed in this study were collected under protocols approved by the institutional review boards of the participating medical centers, with informed consent obtained in accordance with applicable regulations. The study analyzes recordings acquired from pediatric patients undergoing clinically indicated intracranial monitoring, with research stimulation performed under approved protocols. Participant information is anonymized in the submission, and the analyses in this work do not alter clinical treatment decisions. We report results at the subject/hemisphere or aggregate level and limit our claims to properties of the recorded signals and learned representations.

REPRODUCIBILITY STATEMENT

We provide the information required to reproduce the model and experiments throughout the paper and supplementary material. The complete SPIRE objective, architecture, optimization schedules, and multi-region formulation are described in Appendix A.1.1-A.1.3. Synthetic-data generation and baseline fitting procedures are detailed in Appendix A.2.1-A.2.4. Human-data preprocessing, latent-dimension selection, and training procedures are provided in Appendix A.3.1, Appendix A.3.2, and Appendix A.3.3. Additional robustness analyses and complete recovery results are reported in Appendix A.4. The human intracranial recordings contain sensitive clinical data and are not publicly distributed. De-identified data may be made available upon reasonable request, subject to completion of an appropriate data-use agreement and any applicable institutional and ethical approvals. Code and experiment scripts are available in the anonymous repository: `https://github.com/GitAnonymSubmission/spire-ICLR2027`.

REFERENCES

Lucia Abela and Manju Ann Kurian. Kmt2b-related dystonia synonyms: Dyt28, dyt-kmt2b, 1993. URL `https://www.ncbi.nlm.nih.gov/books/`.

Michaela E Alarie, Nicole R Provenza, Michelle Avendano-Ortega, Sarah A McKay, Ayan S Waite, Raissa K Mathura, Jeffrey A Herron, Sameer A Sheth, David A Borton, and Wayne K Goodman. Artifact characterization and mitigation techniques during concurrent sensing and stimulation using bidirectional deep brain stimulation platforms. *Frontiers in Human Neuroscience*, 16:1016379, 2022.

Anonymous. Details omitted for double-blind review, 2018.

Anonymous. Details omitted for double-blind review. 2019.

Anonymous. Details omitted for double-blind review, 2023.

Anonymous. Details omitted for double-blind review. 2024a.

Anonymous. Details omitted for double-blind review. 2024b.

Anonymous. Details omitted for double-blind review. 2025.

Francis R Bach and Michael I Jordan. A probabilistic interpretation of canonical correlation analysis. 2005.

Adrien Bardes, Jean Ponce, and Yann LeCun. VICReg: Variance-invariance-covariance regularization for self-supervised learning. In *International Conference on Learning Representations (ICLR)*, 2022. URL https://openreview.net/forum?id=xm6YD62D1Ub.

Claude Bedard, Helmut Kroeger, and Alain Destexhe. Does the 1/f frequency scaling of brain signals reflect self-organized critical states? *Physical review letters*, 97(11):118102, 2006.

Alim Louis Benabid. Deep brain stimulation for parkinson's disease. *Current opinion in neurobiology*, 13(6):696–706, 2003.

Nikolas et al. Boy. Proposed recommendations for diagnosing and managing individuals with glutaric aciduria type i: second revision. *Journal of inherited metabolic disease*, 40(1):75–101, 2017. ISSN 1573-2665 (Electronic). doi: 10.1007/s10545-016-9999-9.

Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

Jahrane Dale, Stephen L Schmidt, Kyle Mitchell, Dennis A Turner, and Warren M Grill. Evoked potentials generated by deep brain stimulation for parkinson's disease. *Brain stimulation*, 15(5): 1040–1047, 2022.

Matthew Dowling and Cristina Savin. Nonlinear multiregion neural dynamics with parametric impulse response communication channels. In *International Conference on Learning Representations*, volume 2025, pp. 90513–90531, 2025.

Tom Dupré la Tour, Lucille Tallot, Laetitia Grabot, Valérie Doyère, Virginie Van Wassenhove, Yves Grenier, and Alexandre Gramfort. Non-linear auto-regressive models for cross-frequency coupling in neural time series. *PLoS computational biology*, 13(12):e1005893, 2017.

Eray Erturk, Saba Hashemi, and Maryam M Shanechi. Cross-modal representational knowledge distillation for enhanced spike-informed lfp modeling. *arXiv preprint arXiv:2512.12461*, 2025.

Adriana Galvan, Annaelle Devergnas, and Thomas Wichmann. Alterations in neuronal activity in basal ganglia-thalamocortical circuits in the parkinsonian state. *Frontiers in neuroanatomy*, 9:5, 2015.

Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine learning*, 63(1):3–42, 2006.

Evren Gokcen, Anna I Jasper, João D Semedo, Amin Zandvakili, Adam Kohn, Christian K Machens, and Byron M Yu. Disentangling the flow of signals between populations of neurons. *Nature Computational Science*, 2(8):512–525, 2022.

Adam Gosztolai, Robert L Peach, Alexis Arnaudon, Mauricio Barahona, and Pierre Vandergheynst. Marble: interpretable representations of neural population dynamics using geometric deep learning. *Nature Methods*, pp. 1–9, 2025.

Allison Gregory and Susan J Hayflick. Pantothenate kinase-associated neurodegeneration. 2017.

Ameer Hamza. Kernicterus. *Autopsy and Case Reports*, 9(1):0–0, 2019.

Biyu J He. Scale-free brain activity: past, present, and future. *Trends in cognitive sciences*, 18(9): 480–487, 2014.

Gali Heimer, Allison Gregory, Penelope Hogarth, Susan Hayflick, et al. Mecr-related neurologic disorder. 2019.

Andreas Horn and Michael D Fox. Opportunities of connectomic neuromodulation. *Neuroimage*, 221:117180, 2020.

HA Jinnah and Ellen J Hess. A new twist on the anatomy of dystonia. *Neurology*, 67(10):1740–1741, 2006.

Orren Karniol-Tambour, David Zoltowski, E Mika Diamanti, Lucas Pinto, Carlos Brody, David Tank, and Jonathan Pillow. Modeling state-dependent communication between brain regions with switching nonlinear dynamical systems. In *International Conference on Learning Representations*, volume 2024, pp. 27010–27030, 2024.

Karen W Krigger. Cerebral palsy: an overview. *American family physician*, 73(1):91–100, 2006.

Travis Larsh, Steve W Wu, Sudhakar Vadivelu, Gerald A Grant, and Jennifer A O'Malley. Deep brain stimulation for pediatric dystonia. In *Seminars in pediatric neurology*, volume 38, pp. 100896. Elsevier, 2021.

Mihee Lee and Vladimir Pavlovic. Private-shared disentangled multimodal vae for learning of latent representations. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, pp. 1692–1700, 2021.

Belle Liu, Jacob Sacks, and Matthew D Golub. Accurate identification of communication between multiple interacting neural populations. *Proceedings of machine learning research*, 267:39381, 2025.

Andres M Lozano, Nir Lipsman, Hagai Bergman, Peter Brown, Stephan Chabardes, Jin Woo Chang, Keith Matthews, Cameron C McIntyre, Thomas E Schlaepfer, Michael Schulder, et al. Deep brain stimulation: current challenges and future directions. *Nature Reviews Neurology*, 15(3):148–160, 2019.

Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

Cameron C McIntyre and Philip J Hahn. Network perspectives on the mechanisms of deep brain stimulation. *Neurobiology of disease*, 38(3):329–337, 2010.

Manuel R Mercier, Anne-Sophie Dubarry, François Tadel, Pietro Avanzini, Nikolai Axmacher, Dillan Cellier, Maria Del Vecchio, Liberty S Hamilton, Dora Hermes, Michael J Kahana, et al. Advances in human intracranial electroencephalography research, guidelines and good practices. *Neuroimage*, 260:119438, 2022.

Luka Milosevic, Suneil K Kalia, Mojgan Hodaie, Andres M Lozano, Milos R Popovic, and William D Hutchison. Physiological mechanisms of thalamic ventral intermediate nucleus stimulation for tremor suppression. *Brain*, 141(7):2142–2155, 2018.

Jose A Obeso, Concepcio Marin, C Rodriguez-Oroz, Javier Blesa, Beatriz Benitez-Temiño, Juan Mena-Segovia, Manuel Rodríguez, and C Warren Olanow. The basal ganglia in parkinson's disease: current concepts and unexplained observations. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 64(S2):S30–S46, 2008.

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.

A Paszke. Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703*, 2019.

Posit team. *RStudio: Integrated Development Environment for R*. Posit Software, PBC, Boston, MA, 2025. URL `https://www.posit.co/`.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2025. URL `https://www.R-project.org/`.

Stephen L Schmidt, David T Brocker, Brandon D Swan, Dennis A Turner, and Warren M Grill. Evoked potentials reveal neural circuits engaged by human deep brain stimulation. *Brain stimulation*, 13 (6):1706–1718, 2020.

Yuge Shi, Brooks Paige, Philip Torr, et al. Variational mixture-of-experts autoencoders for multi-modal deep generative models. *Advances in neural information processing systems*, 32, 2019.

Ram Dyuthi Sristi, Sowmya Manojna Narasimha, Jingya Huang, Alice Despatin, Simon Musall, Vikash Gilja, and Gal Mishne. Coupled transformer autoencoder for disentangling multi-region neural latent dynamics. *arXiv preprint arXiv:2510.20068*, 2025.

Homa Tajsharghi, Niklas Darin, Elham Rekabdar, Mårten Kyllerman, Jan Wahlström, Tommy Martinsson, and Anders Oldfors. Mutations and sequence variation in the human myosin heavy chain iia gene (myh2). *European journal of human genetics*, 13(5):617–622, 2005.

Robert C Vannucci. Hypoxic-ischemic encephalopathy. *American journal of perinatology*, 17(03): 113–120, 2000.

Yuxiao Yang, Shaoyu Qiao, Omid G Sani, J Isaac Sedillo, Breonna Ferrentino, Bijan Pesaran, and Maryam M Shanechi. Modelling and prediction of the dynamic responses of large-scale brain networks during direct electrical stimulation. *Nature biomedical engineering*, 5(4):324–345, 2021.

Daiyao Yi, Hao Dong, Michael Higley, Anne Churchland, and Shreya Saxena. Shared-ae: Automatic identification of shared subspaces in high-dimensional neural and behavioral activity. In *International Conference on Learning Representations*, volume 2025, pp. 68039–68066, 2025.

Byron M Yu, John P Cunningham, Gopal Santhanam, Stephen Ryu, Krishna V Shenoy, and Maneesh Sahani. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *Advances in neural information processing systems*, 21, 2008.

Ping Zhuang, Yongjie Li, and Mark Hallett. Neuronal activity in the basal ganglia and thalamus in patients with dystonia. *Clinical neurophysiology*, 115(11):2542–2557, 2004.

A Technical appendices and supplementary material

A.1 SPIRE details

A.1.1 Detailed Training Objectives

Let $\mathcal{R}$ denote the set of regions. For any region $r \in \mathcal{R}$, let $x^{(r)} \in \mathbb{R}^{B\times T\times C_r}$ be the input and $\hat{x}^{(r)}$ its reconstruction. Let $z_{\text{sh}}^{(r)} \in \mathbb{R}^{B\times T\times d_{\text{sh}}}$ and $z_{\text{pr}}^{(r)} \in \mathbb{R}^{B\times T\times d_{\text{pr}}}$ be shared and private latents, and $\tilde{z}_{\text{sh}}^{(s\to r)} = M^{(s\to r)}\,\mathrm{ConvAlign}(z_{\text{sh}}^{(s)})$ the temporally aligned and linearly mapped shared latents from s into the space of r (see main text). When needed, we flatten time and batch so that $Z_{\bullet}^{(r)} \in \mathbb{R}^{(BT)\times d_{\bullet}}$, write $N{=}BT$, and use

$$\mathrm{Cov}(A,B) \;=\; \tfrac{1}{N-1}\,A^{\top}B, \qquad \mathrm{std}(\cdot) \text{ computed per feature over the } N \text{ rows.}$$

Reconstruction objective. The primary reconstruction term ensures that each region can be reconstructed from its own shared and private latent representations:

$$\mathcal{L}_{\text{rec}} = \sum_{r\in\mathcal{R}} \left\| x^{(r)} - \hat{x}^{(r)} \right\|_2^2. \tag{A.1}$$

To encourage shared latents to encode transferable cross-regional structure, we additionally include shared-only reconstruction losses.

Cross-reconstruction uses only aligned shared latents from another region:

$$\mathcal{L}_{\text{cross}} = \sum_{\substack{r,s\in\mathcal{R}\\ r\neq s}} \left\| x^{(r)} - f_{\text{dec}}^{(r)}(\tilde{z}_{\text{sh}}^{(s\to r)}, 0) \right\|_2^2. \tag{A.2}$$

Self-reconstruction uses only the region's own shared latents:

$$\mathcal{L}_{\text{self}} = \sum_{r\in\mathcal{R}} \left\| x^{(r)} - f_{\text{dec}}^{(r)}(z_{\text{sh}}^{(r)}, 0) \right\|_2^2. \tag{A.3}$$

Alignment objective. Shared latent spaces are aligned across regions using VICReg:

$$\mathcal{L}_{\text{align}} = \frac{1}{2} \sum_{\substack{r,s \in \mathcal{R} \\ r \neq s}} \Big(\text{VICReg}(z_{\text{sh}}^{(r)}, \tilde{z}_{\text{sh}}^{(s \to r)}) + \text{VICReg}(z_{\text{sh}}^{(s)}, \tilde{z}_{\text{sh}}^{(r \to s)}) \Big). \tag{A.4}$$

To preserve interpretability and prevent excessive deformation, alignment modules are softly regularized toward identity mappings.

Linear mappers are penalized via

$$\mathcal{L}_{\text{mapid}} = \sum_{s \to r} \left\| M^{(s \to r)} - I \right\|_F^2. \tag{A.5}$$

ConvAlign filters are regularized toward impulse-like unit-sum kernels:

$$\mathcal{L}_{\text{align-reg}} = \sum_{s \to r} \sum_{j=1}^{d_{\text{sh}}} \Big(\|k_{s \to r, j} - \delta\|_2^2 + \|\mathbf{1}^\top k_{s \to r, j} - 1\|_2^2 \Big). \tag{A.6}$$

Disentanglement objective. To reduce redundancy between shared and private subspaces, we penalize the cross-covariance between per-feature standardized latents:

$$\mathcal{L}_{\text{orth}} = \sum_{r \in \mathcal{R}} \|\text{Cov}(\bar{Z}_{\text{sh}}^{(r)}, \bar{Z}_{\text{pr}}^{(r)})\|_F^2, \tag{A.7}$$

where $\bar{Z}$ denotes zero-mean, unit-variance normalization.

To avoid degenerate solutions, we additionally include variance guards:

$$\mathcal{L}_{\text{var-sh}} = \sum_{r \in \mathcal{R}} \|\text{std}(z_{\text{sh}}^{(r)}) - 1\|_2^2, \tag{A.8}$$

$$\mathcal{L}_{\text{var-pr}} = \sum_{r \in \mathcal{R}} \sum_{j=1}^{d_{\text{pr}}} \max\Big(0, \tau - \text{std}(z_{\text{pr},j}^{(r)})\Big)^2. \tag{A.9}$$

Total objective. The full SPIRE objective is:

$$\mathcal{L} = \lambda_{\text{rec}} \mathcal{L}_{\text{rec}}^{\text{family}} + \lambda_{\text{align}} \mathcal{L}_{\text{align}}^{\text{family}} + \lambda_{\text{dis}} \mathcal{L}_{\text{dis}}^{\text{family}}, \tag{A.10}$$

where

$$\mathcal{L}_{\text{rec}}^{\text{family}} = \mathcal{L}_{\text{rec}} + \lambda_{\text{self}} \mathcal{L}_{\text{self}} + \lambda_{\text{cross}} \mathcal{L}_{\text{cross}}, \tag{A.11}$$

$$\mathcal{L}_{\text{align}}^{\text{family}} = \mathcal{L}_{\text{align}} + \lambda_{\text{map}} \mathcal{L}_{\text{mapid}} + \lambda_{\text{conv}} \mathcal{L}_{\text{align-reg}}, \tag{A.12}$$

$$\mathcal{L}_{\text{dis}}^{\text{family}} = \mathcal{L}_{\text{orth}} + \lambda_{\text{sh}} \mathcal{L}_{\text{var-sh}} + \lambda_{\text{pr}} \mathcal{L}_{\text{var-pr}}. \tag{A.13}$$

The reconstruction term was fixed to weight 1 throughout all experiments, and variance-guard coefficients were fixed across datasets. In practice, only a subset of loss coefficients were scheduled during training; variance guards remained fixed across datasets.

A.1.2 Implementation Details

Lag augmentation. To improve robustness to temporal misalignment, SPIRE receives lag-augmented features. For region 1 (GPi) with C channels and sequence length T, we stack the signal and its lagged versions up to lag L:

$$X^{(1)} = \begin{bmatrix} x_{:,\,0:T-L}^{(1)} \\ x_{:,\,1:T-L+1}^{(1)} \\ \vdots \\ x_{:,\,L:T}^{(1)} \end{bmatrix} \in \mathbb{R}^{C(L+1) \times (T-L)}, \tag{A.14}$$

and analogously for region 2 (STN). This expands each input into $(L+1)$ lagged copies, enabling SPIRE to model short-range delays directly.

Training setup. Data are split 80/20 into training and validation sets, with mini-batches of size 8 for training and batch size 1 for validation. Learning rate is initialized at 10^{-3} and halved on plateaus (patience 10 epochs). Gradients are scaled and clipped to an L2 norm of 1.0. Validation loss is monitored each epoch with early stopping (patience 20, ignoring the first 140 epochs).

A.1.3 MULTI-REGION EXTENSION

Model overview. SPIRE generalizes naturally from two-region to multi-region recordings. Given a set of regions $\mathcal{R} = \{r_1, \ldots, r_n\}$, the model instantiates one encoder and one decoder per region. For every ordered pair of distinct regions (s, r), SPIRE additionally instantiates one temporal alignment module $\mathrm{ConvAlign}^{(s\to r)}$ and one linear mapper $M^{(s\to r)}$. Thus, the number of directed alignment pathways scales as $|\mathcal{R}|(|\mathcal{R}| - 1)$.

Let

$$\mathcal{P} = \{(s, r) : s, r \in \mathcal{R},\ s \neq r\}$$

denote the set of ordered region pairs. For each pair, the aligned shared latent is

$$\tilde{z}_{\mathrm{sh}}^{(s\to r)} = M^{(s\to r)} \mathrm{ConvAlign}^{(s\to r)} \left(z_{\mathrm{sh}}^{(s)} \right).$$

Multi-region objective. The training objective extends directly by averaging region-wise losses over regions and pairwise losses over ordered region pairs. Specifically, reconstruction, self shared-only reconstruction, orthogonality, and variance guards are averaged over $r \in \mathcal{R}$, while cross-reconstruction, shared alignment, mapper identity, and ConvAlign regularization are averaged over $(s, r) \in \mathcal{P}$:

$$\mathcal{L}_{\mathrm{rec}} = \frac{1}{|\mathcal{R}|} \sum_{r\in\mathcal{R}} \left\| x^{(r)} - \hat{x}^{(r)} \right\|_2^2, \tag{A.15}$$

$$\mathcal{L}_{\mathrm{self}} = \frac{1}{|\mathcal{R}|} \sum_{r\in\mathcal{R}} \left\| x^{(r)} - f_{\mathrm{dec}}^{(r)}(z_{\mathrm{sh}}^{(r)}, 0) \right\|_2^2, \tag{A.16}$$

$$\mathcal{L}_{\mathrm{cross}} = \frac{1}{|\mathcal{P}|} \sum_{(s,r)\in\mathcal{P}} \left\| x^{(r)} - f_{\mathrm{dec}}^{(r)}(\tilde{z}_{\mathrm{sh}}^{(s\to r)}, 0) \right\|_2^2, \tag{A.17}$$

$$\mathcal{L}_{\mathrm{align}} = \frac{1}{|\mathcal{P}|} \sum_{(s,r)\in\mathcal{P}} \mathrm{VICReg} \left(z_{\mathrm{sh}}^{(r)}, \tilde{z}_{\mathrm{sh}}^{(s\to r)} \right), \tag{A.18}$$

$$\mathcal{L}_{\mathrm{orth}} = \frac{1}{|\mathcal{R}|} \sum_{r\in\mathcal{R}} \left\| \mathrm{Cov}(\bar{Z}_{\mathrm{sh}}^{(r)}, \bar{Z}_{\mathrm{pr}}^{(r)}) \right\|_F^2. \tag{A.19}$$

The variance guards are similarly averaged over regions:

$$\mathcal{L}_{\mathrm{var-sh}} = \frac{1}{|\mathcal{R}|} \sum_{r\in\mathcal{R}} \left\| \mathrm{std}(z_{\mathrm{sh}}^{(r)}) - 1 \right\|_2^2, \tag{A.20}$$

$$\mathcal{L}_{\mathrm{var-pr}} = \frac{1}{|\mathcal{R}|} \sum_{r\in\mathcal{R}} \sum_{j=1}^{d_{\mathrm{pr}}} \max \left(0, \tau - \mathrm{std}(z_{\mathrm{pr},j}^{(r)}) \right)^2. \tag{A.21}$$

Finally, alignment-module regularizers are averaged over ordered region pairs:

$$\mathcal{L}_{\mathrm{mapid}} = \frac{1}{|\mathcal{P}|} \sum_{(s,r)\in\mathcal{P}} \left\| M^{(s\to r)} - I \right\|_F^2, \tag{A.22}$$

$$\mathcal{L}_{\mathrm{align-reg}} = \frac{1}{|\mathcal{P}|} \sum_{(s,r)\in\mathcal{P}} \mathrm{Reg} \left(\mathrm{ConvAlign}^{(s\to r)} \right). \tag{A.23}$$

The total loss is the same weighted combination used in the two-region case:

$$\mathcal{L} = w_{\mathrm{rec}}\mathcal{L}_{\mathrm{rec}} + w_{\mathrm{align}}\mathcal{L}_{\mathrm{align}} + w_{\mathrm{orth}}\mathcal{L}_{\mathrm{orth}} + w_{\mathrm{self}}\mathcal{L}_{\mathrm{self}} + w_{\mathrm{cross}}\mathcal{L}_{\mathrm{cross}} \tag{A.24}$$

$$+ w_{\mathrm{mapid}}\mathcal{L}_{\mathrm{mapid}} + w_{\mathrm{align-reg}}\mathcal{L}_{\mathrm{align-reg}} + w_{\mathrm{var-sh}}\mathcal{L}_{\mathrm{var-sh}} + w_{\mathrm{var-pr}}\mathcal{L}_{\mathrm{var-pr}}. \tag{A.25}$$

Because each term is normalized by either $|\mathcal{R}|$ or $|\mathcal{P}|$, adding additional regions does not automatically increase the magnitude of the objective. This normalization allows the same loss schedules to be reused across two-region and three-region SPIRE models and adding a third region does not introduce additional loss-weight hyperparameters.

A.2 Synthetic data

A.2.1 Synthetic benchmark generator

Our synthetic generator builds multichannel observations from ground-truth shared and private latent processes.

Latent dynamics. Shared and private latent processes are generated as bursty oscillatory signals with controllable spectral structure. Each latent dimension is initialized from a base oscillation with variable frequency, occasional 30-40 Hz transient bursts, and stochastic frequency shifts over time. Optional AR(1) filtering introduces temporal autocorrelation and smooth latent evolution.

To model realistic cross-region nonlinearities, shared latents may undergo region-specific monotone nonlinear warps (`region_mismatch`) before observation generation. Optional sinusoidal time-varying delays are additionally applied to one region to simulate nonstationary inter-regional timing relationships. The generator also supports explicit spectral separation between shared and private processes through independently configurable base frequencies, amplitudes, and colored-noise structure.

Observation generation. Each region is assigned a synthetic electrode geometry with contacts arranged in rows. Shared and private latent sources are projected to contacts through randomized spatial mixing matrices, optionally incorporating smooth spatial kernels. Observed signals are then generated through linear or nonlinear mixing functions (`gain`, `bilinear`, or `gain+bilinear`), with optional multiplicative interactions between shared and private components.

To better approximate intracranial recordings, common-mode structure, row-wise drift, heteroscedastic noise, and sensor noise may be added before bipolar rereferencing is applied.

Noise structure. The generator supports multiple biologically inspired noise processes, including colored $1/f$ noise, AR(1) temporal structure, heteroscedastic amplitude-dependent noise, row-wise drift, and low-rank common-mode fluctuations. Different benchmark presets selectively enable subsets of these mechanisms to progressively increase realism and modeling difficulty.

Benchmark datasets. We evaluated models on four synthetic benchmark regimes of progressively increasing complexity:

- `D0`: approximately linear regime with identity shared mappings, minimal sensor noise, and no temporal misalignment. This setting is favorable to linear latent-variable models such as DLAG.
- `D1`: nonlinear regime with region-specific shared warping, nonlinear gain+bilinear mixing, randomized spatial mixing, colored $1/f$ noise, and AR(1) latent dynamics.
- `D2`: same as D1, but with an additional sinusoidal time-varying delay between regions to model nonstationary inter-regional timing relationships.
- `D3`: spectrally asymmetric regime in which shared latents occupy higher-frequency bands while private latents dominate lower-frequency structure and variance. This setting tests whether models can correctly disentangle cross-regional dynamics from strong region-specific activity.

All datasets contain 100 trials of 250 timepoints sampled at 500 Hz, with 3 shared and 3 private latent dimensions per region.

Figure A.1 illustrates representative observed channels, latent trajectories, nonlinear cross-region relationships, and time-varying delay structure across the benchmark regimes.

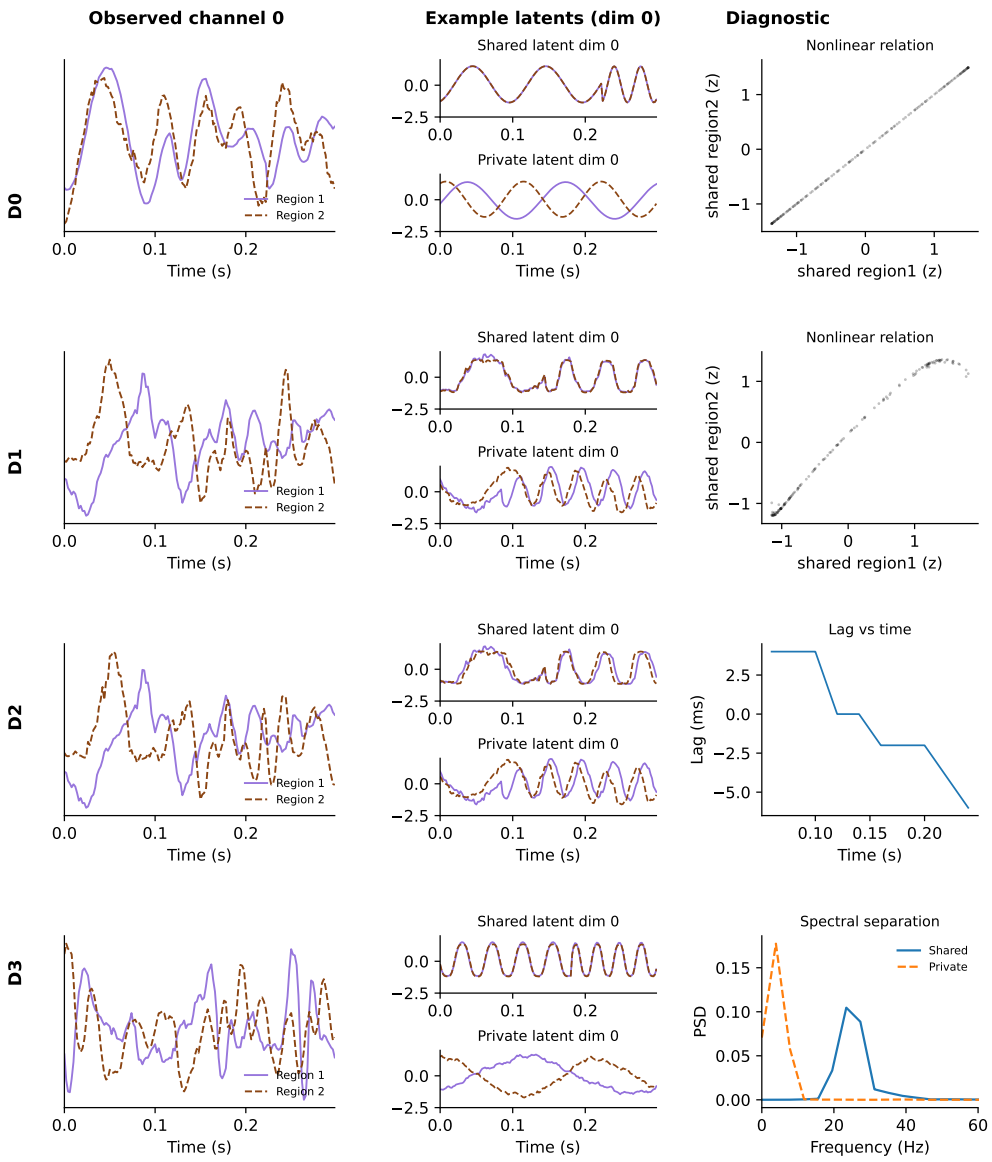


Figure A.1: Synthetic benchmark datasets. Each trial is 0.5 s, with a 0.3 s segment shown for visualization. Left: example observed channels from region 1 and region 2. Middle: example shared and private latent trajectories. Right: diagnostic plots showing nonlinear cross-region relationships, time-varying delay structure, and spectral separation between shared and private dynamics.

A.2.2 SPIRE TRAINING DETAILS

Model architecture. Each region used a GRU encoder/decoder (hidden size 64, dropout 0.3). Encoders projected into shared ($d_s = 3$) and private ($d_p = 3$) latents. Cross-regional alignment was performed with depthwise ConvAlign filters (kernel size 9) followed by lightweight linear mappers initialized as identity. ConvAlign filters were regularized toward impulse responses.

Training regime. Models were trained for up to 500 epochs. A private-gate α_p gradually opened between epochs 80-140, phasing in private latents. During epochs 90-110, shared projections were frozen while aligners remained trainable.

Table A.1: Loss weights for synthetic training across phases. α_p ramps from 0 to 1 between epochs 80-140.

Loss component	Pre (0-80)	Ramp (80-140)	Post (140-500)
$\mathcal{L}_{\text{rec}}$ (reconstruction)	1.00	1.00	1.00
$\mathcal{L}_{\text{align}}$ (shared alignment)	0.22	0.10	0.08
$\mathcal{L}_{\text{cross}}$ (cross recon)	0.03	0.05	0.07
$\mathcal{L}_{\text{self}}$ (self recon)	0.02	0.04	0.03
$\mathcal{L}_{\text{orth}}$ (shared$\perp$private)	0.008	0.015	0.025
$\mathcal{L}_{\text{mapid}}$ (mapper identity)	0.010	0.005	0.000
$\mathcal{L}_{\text{align-reg}}$ (ConvAlign reg.)	1×10^{-4}	5×10^{-4}	1×10^{-4}
Fixed variance guards			
$\mathcal{L}_{\text{var}}^{\text{shared}}$ (unit var)	0.005	0.005	0.005
$\mathcal{L}_{\text{var}}^{\text{private}}$ (floor)	0.002	0.002	0.002

Freeze window: epochs 90-110 (shared projections frozen, aligners remain trainable).

Loss scheduling. Loss weights were scheduled in three phases (pre, ramp, post) aligned to the private-gate ramp. Table A.1 summarizes the schedules for all components.

A.2.3 DLAG FITTING DETAILS

Data formatting. Following the public DLAG demo, we represent each trial as a struct with fields `trialId`, `T`, and `y`. For our two regions (GPi-like = group 1; STN-like = group 2), we construct `y` by stacking the trial's multichannel recordings from both regions along the channel dimension, yielding a matrix of shape $[C_1{+}C_2] \times T$ per trial. A small i.i.d. Gaussian jitter ($\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$, $\sigma \approx 10^{-6}$) is added to avoid singular covariances during initialization. We set the across-group (shared) and within-group (private) latent dimensionalities in DLAG to match the synthetic ground-truth for each preset: 3 shared and 3 private latents.

Temporal and kernel settings. We follow the demo configuration and use the RBF Gaussian-process kernel (`covType='rbf'`) for latent dynamics. The GP timescale initialization is `startTau` $= 2 \times$ `binWidth`. The `binWidth` parameter (ms) is set to the sample period used to bin the synthetic trials, and `segLength` $= T$ matches the trial length (in time bins).

Optimization and stopping. We fit DLAG with EM up to `maxIters` $= 500$ (demo-style runs may use higher values), log-likelihood tolerance `tolLL` $= 10^{-8}$, and progress diagnostics saved every `freqParam` $= 10$ iterations. Private noise floors are enforced with `minVarFrac` $= 10^{-3}$. Unless otherwise noted, we do not perform cross-validation (`numFolds` $= 0$). We fix a random seed for reproducibility.

Evaluation protocol. For each synthetic preset, we train four seeds and align the recovered DLAG latents to ground truth via CCA (performed separately for across-group vs. within-group latents). We then report canonical correlations for shared and private components, averaged across regions, trials, and seeds (Appendix Table A.2). This evaluation protocol was shared across SPIRE, DLAG, and CTAE to enable consistent comparison of recovered shared and private latent structure across models.

A.2.4 CTAE TRAINING DETAILS

Data formatting. CTAE was trained directly on the synthetic multichannel time-series recordings without modifying the original model architecture. For each synthetic preset, region 1 and region 2 recordings were represented as tensors of shape $[N \times T \times C]$, where N denotes trials, T timepoints, and C channels. Following the original CTAE implementation, the two regions were concatenated along the channel dimension to form a joint multiregion input tensor. All channels were z-scored across trials and time prior to training.

To ensure comparability with SPIRE and DLAG, the latent dimensionalities were matched to the synthetic ground truth, using 3 shared and 3 private latent dimensions per region.

Training protocol. We used the original CTAE architecture and loss formulation without architectural modifications. Training employed the same optimization pipeline used for the real-data experiments, including reconstruction, cross-reconstruction, alignment, and orthogonality objectives. Unless otherwise noted, models used 2 transformer layers, 1 attention head, latent dimensionalities matched to ground truth (3 shared and 3 private dimensions), learning rate 10^{-4}, and batch size 8. Positional encodings were enabled using fixed sinusoidal embeddings.

The orthogonality penalty was gradually introduced using a warm-up schedule during early training. We additionally explored longer training durations, larger synthetic datasets, and multiple alignment regularization strengths ($\lambda_{\text{align}} \in \{0.01, 0.05\}$); however, these changes did not substantially alter the qualitative performance trends reported in the main text. Early stopping based on validation reconstruction performance was used to prevent overfitting.

Losses and regularization. The CTAE objective combines within-region reconstruction, cross-region reconstruction, shared-latent alignment, and latent orthogonality regularization. Orthogonality regularization was gradually increased after an initial warm-up period to stabilize optimization and avoid early collapse of latent representations.

Evaluation protocol. For each synthetic benchmark regime, CTAE was trained using four random seeds. Recovered shared and private latent trajectories were aligned to the corresponding ground-truth latent processes using canonical correlation analysis (CCA), performed separately for shared and private components. Canonical correlations were then averaged across latent dimensions, trials, regions, and seeds. This evaluation exactly mirrors the SPIRE protocol to enable direct comparison of disentanglement quality under identical synthetic conditions.

A.2.5 Ablation analysis

To evaluate the contribution of SPIRE's individual objective families and training-control mechanisms, we performed two categories of ablations on the synthetic benchmarks.

Family ablations. We first removed entire groups of loss terms: (i) `SPIRE_no_shared_only`, which removes shared-only and cross-region reconstruction objectives, (ii) `SPIRE_no_alignment`, which removes shared-latent alignment and alignment regularization terms, and (iii) `SPIRE_no_disentangle`, which removes disentanglement-related regularization, including orthogonality and mapper-identity constraints.

Control ablations. We additionally evaluated training-control variants: (i) `ctrl_identity_aligner`, which bypasses convolutional temporal aligners and uses identity alignment, (ii) `ctrl_no_private_ramp`, which disables gradual introduction of private latents, (iii) `ctrl_no_freeze`, which removes the early freeze window during optimization, and (iv) `ctrl_no_var_guard`, which disables variance stabilization constraints on latent representations.

All variants were trained using the same protocol as the full SPIRE model across all four synthetic benchmark regimes (D0-D3). Recovered shared and private latent trajectories were aligned to ground truth using canonical correlation analysis (CCA), separately for shared and private components. Results were averaged across regions, trials, and random seeds. Statistical comparisons were performed relative to the full model (`SPIRE_synth`) using linear models with Dunnett-adjusted contrasts.

Figures A.2a and A.2b summarize shared and private latent recovery, respectively. Across benchmark regimes, no ablation variant consistently outperformed the full SPIRE model. Instead, removing any major objective family or control mechanism degraded recovery performance in at least one dataset or latent type.

The impact of ablations varied across synthetic regimes. Removing alignment-related terms most strongly degraded shared latent recovery in the nonlinear and temporally distorted settings (D1-D2), highlighting the importance of explicit cross-region alignment under nonlinear warping and time-varying delays. Disentanglement-related regularization additionally became critical in the spectrally

Table A.2: CCA-based synthetic latent recovery across benchmark regimes. Values report mean canonical correlation (SD) between learned and corresponding ground-truth latent subspaces across four matched random seeds. Bold indicates the highest recovery within each regime and latent class. Significance was assessed using regime-wise Dunnett-adjusted contrasts comparing each baseline against SPIRE ($^{\cdot}p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$).

Dataset	Metric	SPIRE	DLAG	CTAE
D0	Shared latent recovery	0.72 (0.03)	**0.91 (0.03)*****	0.26 (0.04)***
	Private latent recovery	0.71 (0.08)	**0.87 (0.05)*****	0.38 (0.02)***
D1	Shared latent recovery	**0.82 (0.02)**	0.75 (0.01)$^{\cdot}$	0.25 (0.09)***
	Private latent recovery	**0.63 (0.03)**	0.55 (0.02)*	0.27 (0.06)***
D2	Shared latent recovery	**0.82 (0.00)**	0.71 (0.01)**	0.22 (0.09)***
	Private latent recovery	**0.61 (0.01)**	0.55 (0.01)*	0.28 (0.05)***
D3	Shared latent recovery	**0.63 (0.02)**	0.47 (0.04)***	0.21 (0.06)***
	Private latent recovery	**0.66 (0.05)**	0.48 (0.03)***	0.44 (0.03)***

asymmetric benchmark (D3), where removing orthogonality and mapper constraints substantially reduced both shared and private latent recovery.

Training-control mechanisms also contributed meaningfully to stable disentanglement. Disabling the variance guard produced unstable latent recovery in several regimes, while removing the private latent ramp or optimization freeze window reduced performance selectively depending on dataset complexity. Notably, identity alignment occasionally performed competitively in the approximately linear regime (D0), but degraded performance in more realistic nonlinear settings.

Overall, these results demonstrate that SPIRE's performance depends on the complementary interaction of reconstruction, alignment, disentanglement, and training-control mechanisms. Different synthetic regimes stress different aspects of the model, and no reduced variant matched the robustness of the full SPIRE objective across all benchmark conditions.

A.2.6 Additional CCA recovery and trajectory results

The main text evaluates ground-truth recovery using nonlinear, cross-validated R^2, which can capture nonlinear relationships between learned representations and the generating latent factors. As a complementary linear measure, we also quantified recovery using canonical correlation analysis (CCA). CCA measures similarity between the learned latent subspace and the corresponding ground-truth latent subspace while being invariant to linear rotations of the latent coordinates. We additionally visualize representative trajectories after CCA-based alignment to facilitate qualitative comparison of the recovered and ground-truth dynamics.

The CCA analysis is consistent with the nonlinear recovery results in the main text. DLAG achieves the strongest recovery in the approximately linear D0 positive-control regime, whereas SPIRE achieves the strongest shared and private recovery in D1-D3 as nonlinear observation structure, temporal misalignment, and spectral confounding are introduced.

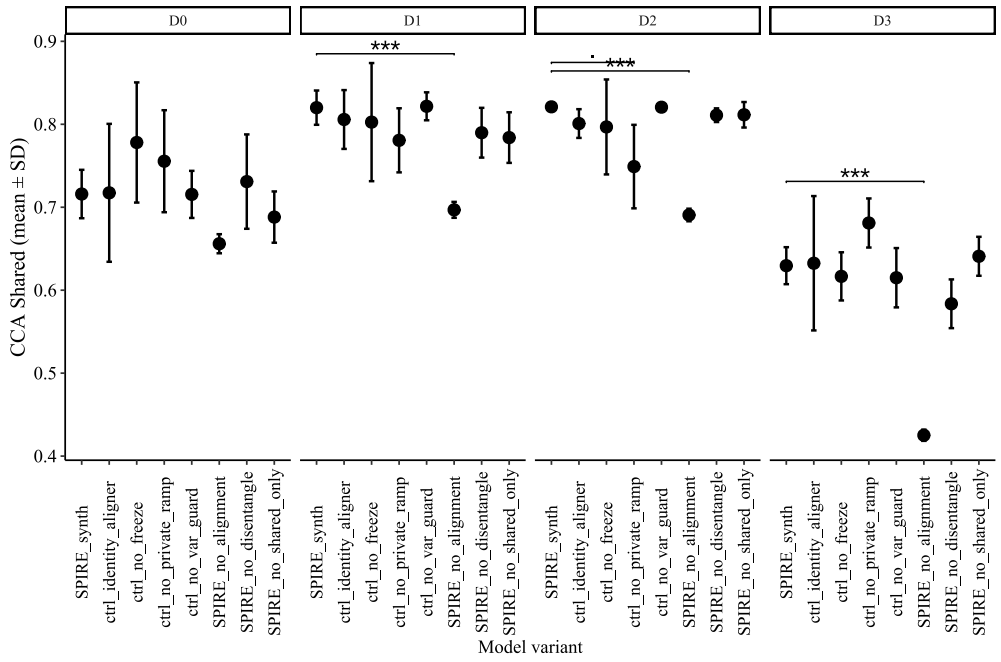
Representative trajectory recovery for the remaining D0-D2 regimes is shown in Figs. A.3-A.5.

A.3 Real intracranial DBS recordings

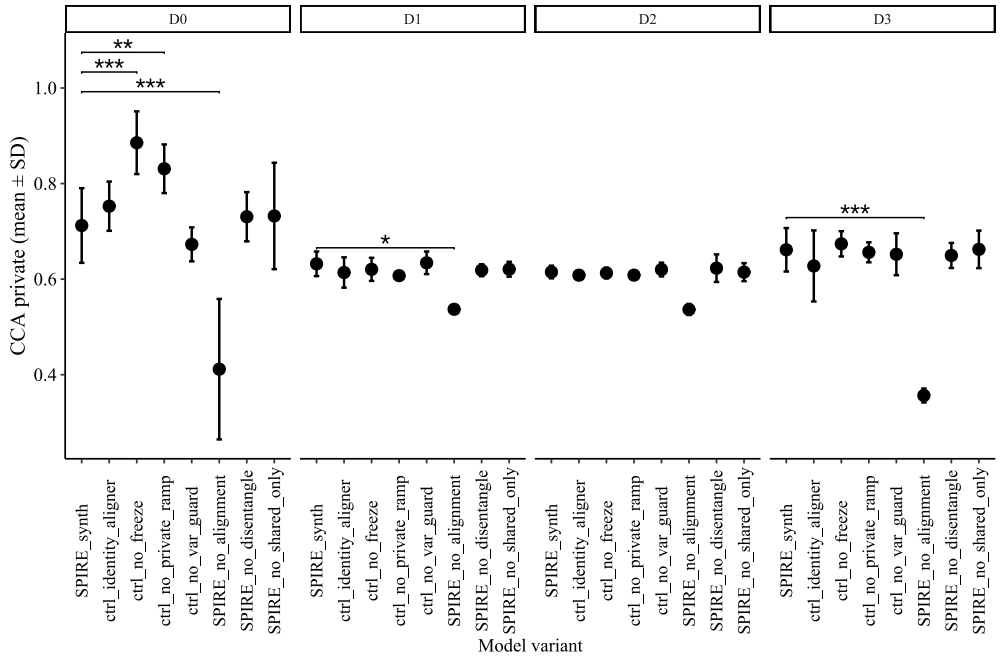
A.3.1 Dataset information

Participant information Table A.3 summarizes demographics, etiologies, and the number of usable recording and stimulation contacts per hemisphere. Etiologies spanned genetic, metabolic, and acquired forms of dystonia. All subjects were male and between 5 and 23 years of age.

Analyses were performed at the hemisphere level. Based on channel quality thresholds (at least three usable microelectrode contacts per region), 17 hemispheres were included for GPi stimulation experiments and 13 for STN stimulation. Due to clinical time constraints, the 250 Hz condition was



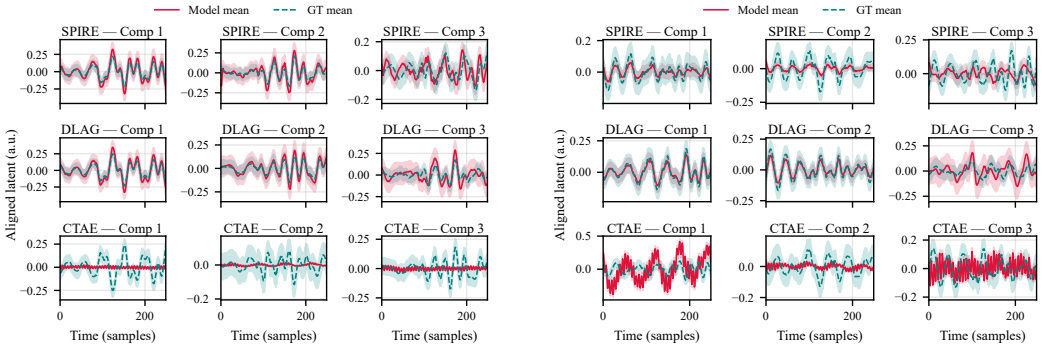
(a) Shared latents



(b) Private latents

Figure A.2: Ablation results for shared latents with statistical comparison. Mean ± SD CCA across four seeds, faceted by dataset regime. Dunnett-adjusted contrasts are shown relative to the full SPIRE model (`SPIRE_synth`).

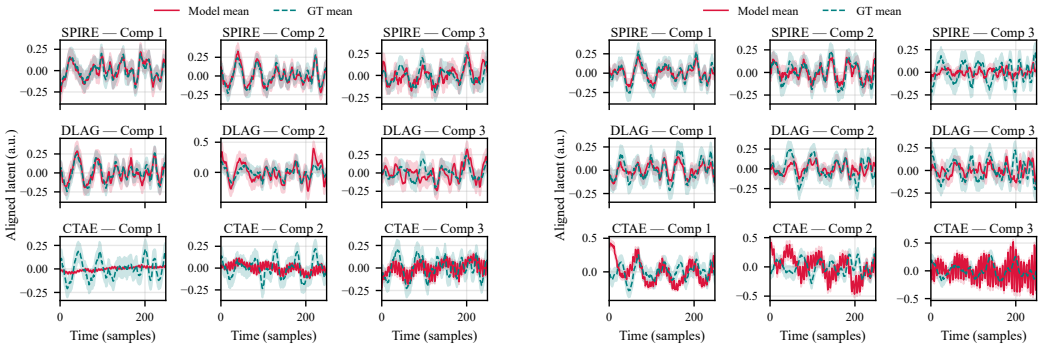
not administered in 8 hemispheres with STN stimulation hence we only consider 85 and 185 Hz stimulation for STN stimulation analysis.



(a) Shared latent recovery.

(b) Private latent recovery.

Figure A.3: Representative latent recovery in the D0 synthetic regime. Recovered trajectories (solid red) are compared with ground-truth trajectories (dashed teal) after CCA-based alignment for visualization. Shaded regions denote SEM across trials.



(a) Shared latent recovery.

(b) Private latent recovery.

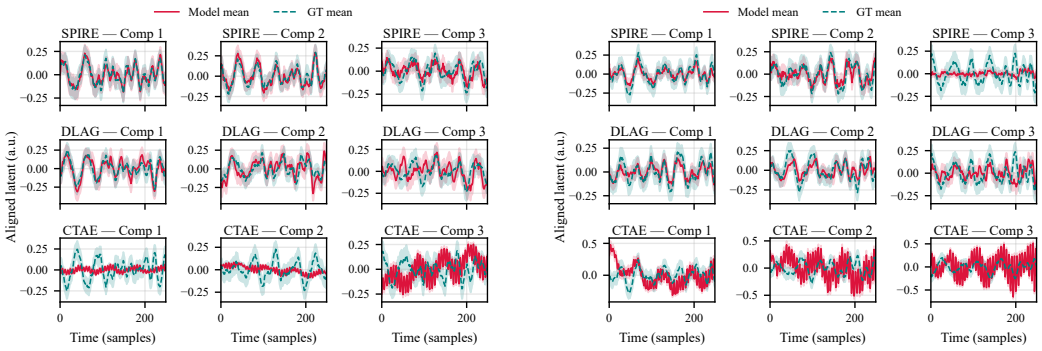
Figure A.4: Representative latent recovery in the D1 synthetic regime. Recovered trajectories (solid red) are compared with ground-truth trajectories (dashed teal) after CCA-based alignment for visualization. Shaded regions denote SEM across trials.

Stimulation paradigm All procedures were approved by institutional review boards at participating medical centers (Anonymous), and informed consent was obtained in accordance with HIPAA regulations.

Stimulation experiments were conducted 2–3 days post-implantation, while temporary stereoelectroencephalography (sEEG) leads remained in place. Our stimulation protocol (adapted from prior work (Anonymous, 2023)) was designed to probe the frequency-dependent effects of DBS using clinical stimulation ranges commonly used for dystonia. Trains of biphasic pulses were delivered at 85, 185, and 250 Hz through adjacent macro-contacts on each DBS lead. Each frequency–contact pair was tested under two configurations (cathode–anode and reversed polarity), with approximately 1000 pulses delivered per condition (pulse width: 90 µs).

Amplitude was set to 3 V by default but reduced if subjects experienced discomfort. For frequencies above 100 Hz, a ramping protocol gradually increased voltage from 0 to 3 V over 12 seconds to improve tolerance. Intracranial signals from all implanted leads were recorded simultaneously, with recording duration adjusted to ensure that 1000 pulses were captured in each case.

Signal pre-processing (extended details) Recordings were acquired at 24,414 Hz using a Tucker-Davis Technologies (TDT) system and preprocessed initially in MATLAB (R2023b). Bipolar derivation (subtracting adjacent contacts) was applied to enhance spatial specificity and reduce common-mode noise (Mercier et al., 2022). To reduce dimensionality while retaining relevant neural



(a) Shared latent recovery.

(b) Private latent recovery.

Figure A.5: Representative latent recovery in the D2 synthetic regime. Recovered trajectories (solid red) are compared with ground-truth trajectories (dashed teal) after CCA-based alignment for visualization. Shaded regions denote SEM across trials.

Table A.3: Participant demographics and number of stimulation and recording contacts per hemisphere. The table includes each subject's etiology, the number of usable microelectrode recording contacts per region (GPi, STN), and whether GPi or STN stimulation was applied in each hemisphere. Abbreviations: MEPAN = Mitochondrial Enoyl CoA Reductase Protein-Associated Neurodegeneration Heimer et al. (2019), KMT2B = Lysine Methyltransferase-2B Abela & Kurian (1993), GA1 = Glutaric aciduria type 1 Boy (2017), PKAN = Pantothenate Kinase-Associated Neurodegeneration Gregory & Hayflick (2017), HIE = Hypoxic-ischemic encephalopathy Vannucci (2000), MYH2 = Myosin heavy chain IIa mutation Tajsharghi et al. (2005), CP = Cerebral palsy Krigger (2006), BG = Basal ganglia.

Subject	Etiology	Hemisphere	GPi rec.	STN rec.	VO rec.	VA rec.	VIM rec.	GPi stim.	STN stim.
S1	MEPAN	Left	18	3	4	4	0	3	1
		Right	12	6	2	3	0	2	1
S2	GA1	Left	6	6	1	6	0	2	1
		Right	9	1	1	4	0	2	1
S3	Atypical PKAN	Left	11	4	1	0	0	4	1
		Right	8	4	1	0	0	3	1
S4	HIE	Left	6	3	1	5	0	2	0
		Right	3	3	2	7	0	2	0
S5	CP (Prematurity)	Left	3	1	1	0	0	1	1
		Right	3	4	1	0	0	1	2
S6	GA1	Left	3	4	1	2	0	2	1
		Right	6	0	2	2	0	2	1
S7	Kernicterus Hamza (2019)	Left	9	3	0	0	3	2	1
		Right	8	4	1	0	3	2	2
S8	BG hemorrhage	Left	6	3	4	2	0	1	0
		Right	9	3	3	2	0	1	0
S9	KMT2B	Left	7	4	1	0	4	3	2
		Right	7	4	1	0	3	3	2
S10	HIE	Left	8	4	1	0	4	2	2
		Right	8	6	2	0	4	2	1

dynamics, data were downsampled to 500 Hz, notch filtered and band-limited with a 50 Hz low-pass Butterworth filter. This preprocessing removed high-frequency stimulation artifacts (Alarie et al., 2022) while preserving oscillatory components of interest.

Then signals were segmented into non-overlapping 0.5-s windows. To capture temporal dependencies, lag-augmented features with 0–3 sample delays were appended to each channel, expanding the feature dimensionality four-fold. This choice balanced computational efficiency with the ability to represent short-timescale spatiotemporal dynamics.

Finally, data were split into training (80%) and test (20%) sets for each subject and hemisphere before model fitting.

A.3.2 VARIANCE PARTITION AND MODEL SELECTION

Motivation. Subjects varied widely in etiology, anatomy, and recording coverage, so we allowed latent dimensionalities to adapt per subject–hemisphere. We trained SPIRE across a grid of shared (3–5) and private (2–4) latent dimensions. To compare candidate models, we quantified how much observed variance in GPi and STN signals was uniquely explained by shared latents, uniquely explained by private latents, or redundantly captured by both. This partitioning ensured that both subspaces contributed meaningfully, avoiding degenerate solutions where all variance collapses into one.

Fraction of variance explained (FVE). For each region $r \in \{\text{GPi}, \text{STN}\}$, we computed cross-validated fraction of variance explained (FVE) by regressing observed signals Y^r on latent trajectories Z:

$$\text{FVE}(Z \to Y^r) = 1 - \frac{\sum \|Y^r - \hat{Y}^r(Z)\|^2}{\sum \|Y^r - \bar{Y}^r\|^2},$$

where $\hat{Y}^r(Z)$ is the ridge-regression prediction and $\bar{Y}^r$ the global mean. We evaluated three cases:

$$\text{FVE}_S^r = \text{FVE}(Z_S^r \to Y^r), \quad \text{FVE}_P^r = \text{FVE}(Z_P^r \to Y^r), \quad \text{FVE}_{SP}^r = \text{FVE}([Z_S^r, Z_P^r] \to Y^r).$$

Order-free partition. Using these, we derived an order-invariant variance partition:

$$\text{UniqueShared}^r = \max\big(0,\ \text{FVE}_{SP}^r - \text{FVE}_P^r\big),$$
$$\text{UniquePrivate}^r = \max\big(0,\ \text{FVE}_{SP}^r - \text{FVE}_S^r\big).$$
$$\text{Redundant}^r = \max\big(0,\ \text{FVE}_S^r + \text{FVE}_P^r - \text{FVE}_{SP}^r\big).$$

Model selection rule. We trained SPIRE across $(d_s, d_p) \in \{3,4,5\} \times \{2,3,4\}$ per subject–hemisphere. Variants were filtered by: (i) redundancy threshold $\max_r \text{Redundant}^r \le \tau_{\text{red}} = 0.20$, (ii) collapse guard $\min_r(\text{UniqueShared}^r, \text{UniquePrivate}^r) \ge \tau_{\text{uni}} = 0.01$. Among survivors, the variant with lowest validation reconstruction loss was chosen, preferring smaller (d_s, d_p) in ties.

Results. Across subjects with chosen dimensions, GPi latents explained $0.302(\pm 0.270)$ unique shared and $0.387(\pm 0.215)$ unique private variance; STN latents explained $0.401(\pm 0.273)$ and $0.403(\pm 0.293)$, respectively. Figure A.6 summarizes the chosen models: each subject–hemisphere shows GPi and STN variance partitioned into UniqueShared, UniquePrivate, and Redundant components. The cohort was heterogeneous: some hemispheres (e.g., S8_R) were private-dominated, while others (e.g., S3) exhibited balanced shared and private contributions. Example UMAP (McInnes et al., 2018) and time-domain traces for balanced case is shown in Appendix Fig. A.7 and an example of private dominant case is illustrated in Fig. A.8. Full dimensionality sweeps appear in Fig. A.9.

A.3.3 TRAINING DETAILS

Architecture. Same GRU encoder–decoder backbone as in the synthetic case, but with dropout 0.2. Aligners were initialized as identity mappers with convolutional filters set to impulses.

Training regime. Models were trained for up to 200 epochs. For the first 60 epochs, aligners remained in identity mode (no convolutional shifts); convolutional alignment was enabled thereafter. Unlike the synthetic setting, no explicit freeze window was used beyond this identity–to–aligner transition.

Loss scheduling. Real-data training followed a four-stage schedule tuned for stability of intracranial recordings. - Epochs < 60: warm-up with mapper-only alignment, no cross-reconstruction. - Epochs 60–100: enable convolutional alignment, modest self/cross reconstruction. - Epochs 100–140: increase alignment and cross/self terms. - Epochs ≥ 140: final tightening of alignment, mapper penalty removed. Loss weights are reported in Table A.4

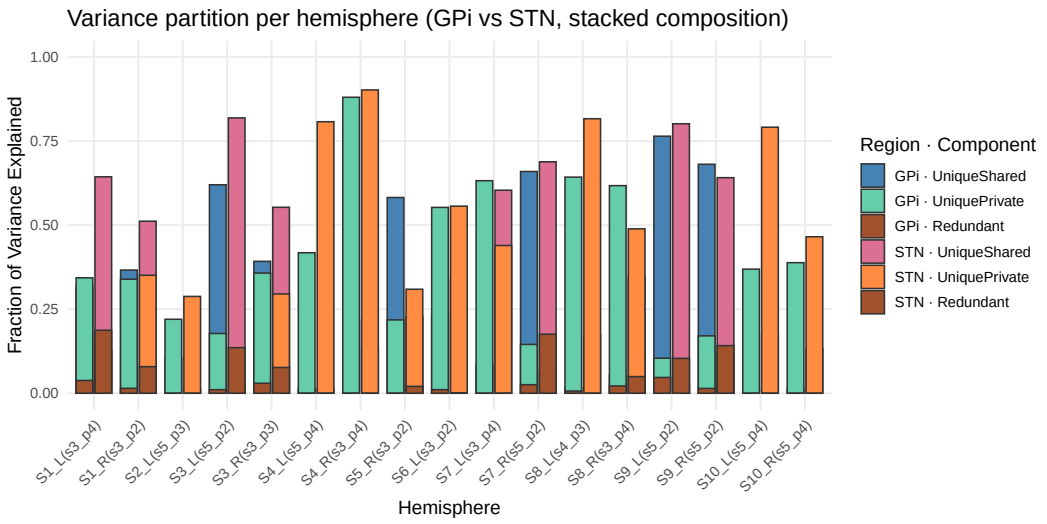
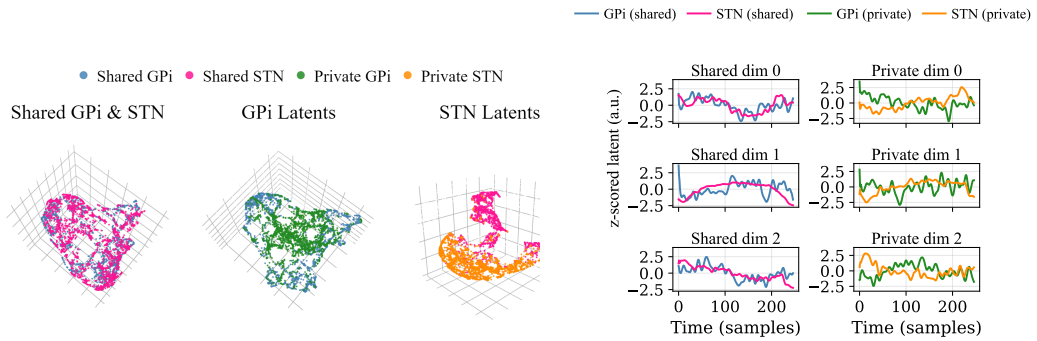


Figure A.6: Subject-specific variance partition of GPi vs. STN. For each subject-hemisphere, two adjacent bars (GPi, STN) show the fraction of variance explained by the *unique shared*, *unique private*, and *redundant* components. Labels include the selected latent sizes (d_s, d_p) per hemisphere chosen via the appendix rule A.3.2). Cohort heterogeneity is evident: several hemispheres are dominated by private variance (weak GPi–STN shared structure), whereas others show a more balanced decomposition consistent with stronger cross-regional coordination. Full dimension sweeps (per hemisphere and region) are in Appendix Fig. A.9.



(a) UMAP of shared vs. private latents.

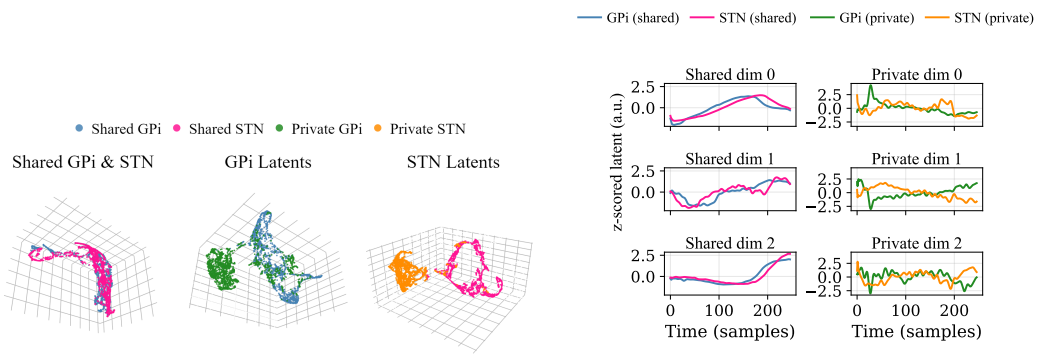
(b) Time-domain latents.

Figure A.7: Representative held-out shared/private latent organization. Shared GPi and STN trajectories overlap more strongly than their private counterparts.

A.3.4 Three-region SPIRE implementation

For the three-region experiments, SPIRE was instantiated with GPi, STN, and thalamic inputs. The model used one GRU encoder-decoder pair per region and one directed alignment pathway for each ordered region pair: GPi→STN, STN→GPi, GPi→Thal, Thal→GPi, STN→Thal, and Thal→STN. Each directed pathway contained a ConvAlign module followed by an identity-initialized linear mapper. As in the two-region setting, all region-wise losses were averaged over regions and all pairwise losses were averaged over ordered pairs, preventing the total objective scale from increasing with the number of regions.

Architecture. The three-region model used the same GRU encoder-decoder backbone as the two-region real-data model, with hidden size 64, dropout 0.2, ConvAlign kernel size 9, and two decoder layers. Shared dimensionality was selected from $d_{\text{sh}} \in \{3, 4, 5\}$, and private dimensionality was



(a) UMAP of shared vs. private latents.

(b) Time-domain latents (shared low-amplitude with lag; private higher variance).

Figure A.8: Representative shared and private latent organization. (a) 3D UMAPs of shared and private GPi/STN latent spaces. (b) Corresponding latent trajectories across time. Shared GPi/STN trajectories are *phase-aligned* but have *very low amplitude* and primarily reflect *slow baseline co-modulation*, whereas private trajectories carry larger, region-specific dynamics that dominate the variance.

Table A.4: Loss weight schedule for real DBS training. Values are weight multipliers in the objective.

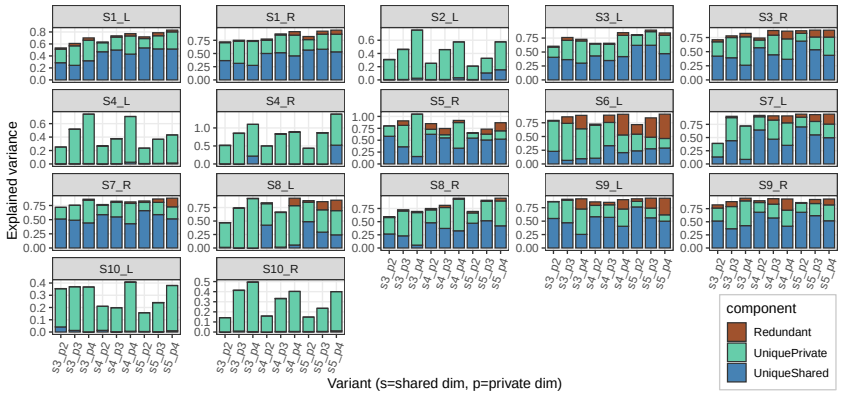
Loss component	**Epoch** (< 60)	**(60-100)**	**(100-140)**	(≥ 140)
Reconstruction ($\mathcal{L}_{\text{rec}}$)	1.00	1.00	1.00	1.00
Alignment ($\mathcal{L}_{\text{align}}$)	0.30	0.30	0.38	0.45
Orthogonality ($\mathcal{L}_{\text{orth}}$)	0.012	0.012	0.015	0.015
Self recon. ($\mathcal{L}_{\text{self}}$)	0.03	0.05	0.05	0.05
Cross recon. ($\mathcal{L}_{\text{cross}}$)	0.00	0.05	0.06	0.06
Mapper identity ($\mathcal{L}_{\text{mapid}}$)	0.005	0.005	0.003	0.000
Aligner regularizer ($\mathcal{L}_{\text{align-reg}}$)	0.000	5×10^{-5}	7.5×10^{-5}	1×10^{-4}
Fixed variance guards				
Shared variance guard ($\mathcal{L}_{\text{var}}^{\text{sh}}$)	0.005	0.005	0.005	0.005
Private variance floor ($\mathcal{L}_{\text{var}}^{\text{pr}}$)	0.002	0.002	0.002	0.002

selected from $d_{\text{pr}} \in \{2, 3, 4\}$ using the same validation-loss and variance-partition criteria used for the two-region models.

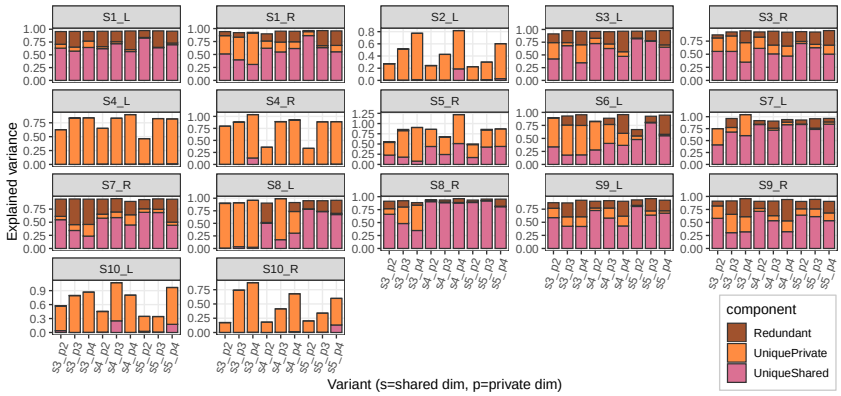
Training regime. Models were trained for up to 200 epochs using Adam with learning rate 10^{-3}, weight decay 10^{-5}, batch size 8, validation batch size 1, mixed precision, gradient clipping at norm 1.0, and ReduceLROnPlateau scheduling. Early stopping used patience 20, ignoring the first 140 epochs. Unlike the synthetic setting, no explicit shared-latent freeze window was used. For the first 60 epochs, alignment was restricted to identity mapping; after epoch 60, ConvAlign modules were enabled.

Loss scheduling. Three-region real-data training followed a four-stage schedule. Epochs < 60 used identity alignment and no cross-reconstruction. Epochs 60-110 enabled ConvAlign with modest cross-reconstruction. Epochs 110-150 increased alignment and cross/self reconstruction. Epochs ≥ 150 further tightened alignment while removing the mapper identity penalty. Loss weights are reported in Table A.5.

Model selection. For each subject, we trained candidate models over private dimensions $d_{\text{pr}} \in \{2, 3, 4\}$ and shared dimension $d_{\text{sh}} \in \{3, 4, 5\}$. The final configuration was selected using the same criteria as the two-region analysis: low validation loss together with a non-degenerate variance partition between shared and private components. This ensured that adding thalamus did not collapse the representation into either a purely shared or purely private solution. The shared and private



(a) GPi



(b) STN

Figure A.9: Variance partition across latent dimensionalities. For each subject–hemisphere, we trained SPIRE with shared dimension $d_s \in \{3, 4, 5\}$ and private dimension $d_p \in \{2, 3, 4\}$. Each bar shows the fraction of explained variance in the observed signal that is uniquely attributable to shared latents, uniquely attributable to private latents, or redundantly captured by both. (a) GPi recordings. (b) STN recordings. These sweeps were used to select the final (d_s, d_p) configuration per subject–hemisphere, as reported in the main text (Fig. A.6).

variance partitions for each selected subject-hemisphere are shown in Fig. A.10, with the full dimensionality sweep shown in Fig. A.11.

Latent alignment analysis . We also evaluated whether the three-region SPIRE model preserved meaningful shared/private structure by computing CCA similarity among latent subspaces. Shared latents aligned across different regions showed the highest similarity, indicating that SPIRE learned a common cross-regional manifold spanning GPi, STN, and thalamus. Shared-private similarities within the same region were lower, suggesting partial but limited leakage between shared and private components. Private-private similarities across regions were lowest, consistent with private latents capturing region-specific residual dynamics rather than shared network structure (Figure A.12).

Three-region stimulation decoding. We next tested whether the three-region SPIRE model captured stimulation-dependent structure during GPi stimulation. Random Forest classifiers were trained to decode GPi stimulation frequency from each latent subset, using four conditions (off, 85, 185, and 250 Hz). Across GPi, STN, and thalamus, shared latents achieved higher decoding accuracy than private latents, with all shared representations performing well above chance. Notably, shared latents from STN and thalamus also encoded GPi stimulation frequency, indicating that stimulation effects were not restricted to the stimulated region but were expressed in the shared multi-region latent manifold. Private latents showed lower and more variable decoding accuracy, consistent with

Table A.5: Loss weight schedule for three-region SPIRE training on real DBS recordings. Values are weight multipliers in the objective.

Loss component	**Epoch** (< 60)	**(60-110)**	**(110-150)**	(≥ 150)
Reconstruction ($\mathcal{L}_{\text{rec}}$)	1.00	1.00	1.00	1.00
Alignment ($\mathcal{L}_{\text{align}}$)	0.22	0.24	0.30	0.36
Orthogonality ($\mathcal{L}_{\text{orth}}$)	0.012	0.012	0.015	0.015
Self recon. ($\mathcal{L}_{\text{self}}$)	0.03	0.04	0.05	0.05
Cross recon. ($\mathcal{L}_{\text{cross}}$)	0.00	0.025	0.04	0.05
Mapper identity ($\mathcal{L}_{\text{mapid}}$)	0.005	0.005	0.003	0.000
Aligner regularizer ($\mathcal{L}_{\text{align-reg}}$)	0.000	5×10^{-5}	7.5×10^{-5}	1×10^{-4}
Fixed variance guards				
Shared variance guard ($\mathcal{L}_{\text{var}}^{\text{sh}}$)	0.005	0.005	0.005	0.005
Private variance floor ($\mathcal{L}_{\text{var}}^{\text{pr}}$)	0.002	0.002	0.002	0.002

The private standard-deviation floor was fixed to 0.20 in all stages. Alignment used identity mode for epochs < 60 and ConvAlign thereafter.

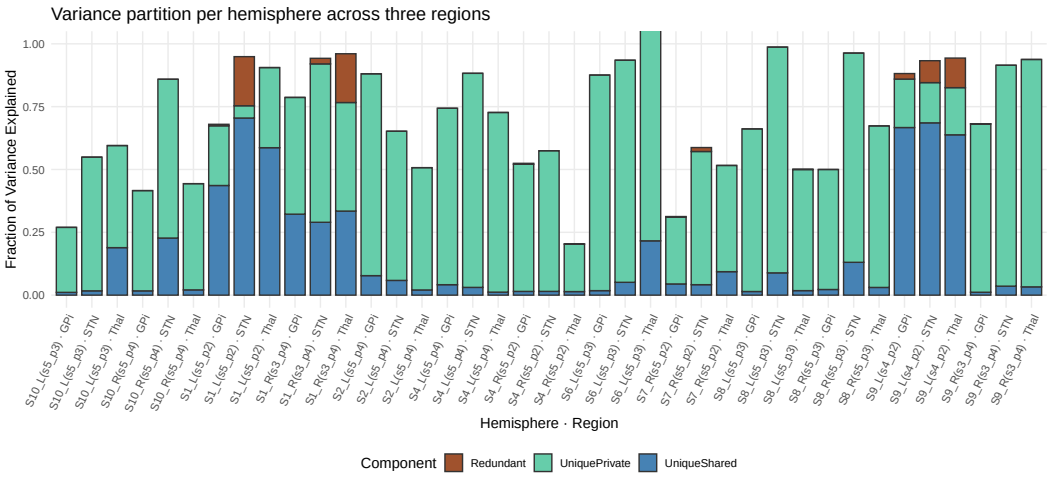


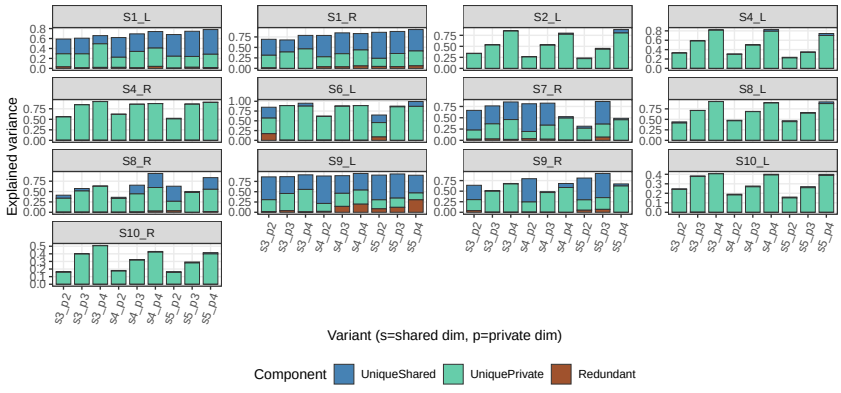
Figure A.10: Subject-specific variance partition of GPi vs. STN vs. Thalamus For each subject-hemisphere, three adjacent bars (GPi, STN, Thal) show the fraction of variance explained by the *unique shared*, *unique private*, and *redundant* components. Labels include the selected latent sizes (d_s, d_p) per hemisphere chosen via the appendix rule A.3.2). Cohort heterogeneity is evident. Full dimension sweeps (per hemisphere and region) are in Appendix Fig. A.11.

their role in capturing region-specific residual dynamics rather than distributed stimulation structure (Figure A.13).

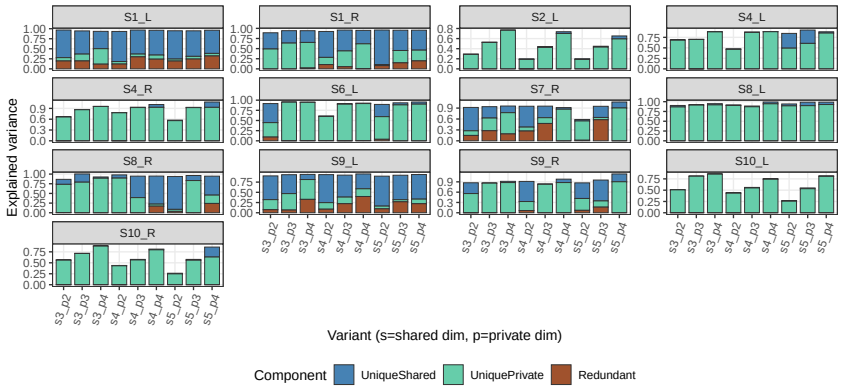
A.3.5 Baseline models

CTAE (Sristi et al., 2025) is a recently proposed transformer-based framework for learning coupled latent representations across multiple neural populations. CTAE employs shared and region-specific latent subspaces together with reconstruction, cross-reconstruction, alignment, and orthogonality objectives to model inter-regional neural structure. We selected CTAE as a strong nonlinear baseline due to its conceptual similarity to SPIRE, particularly its use of coupled latent spaces and explicit alignment objectives.

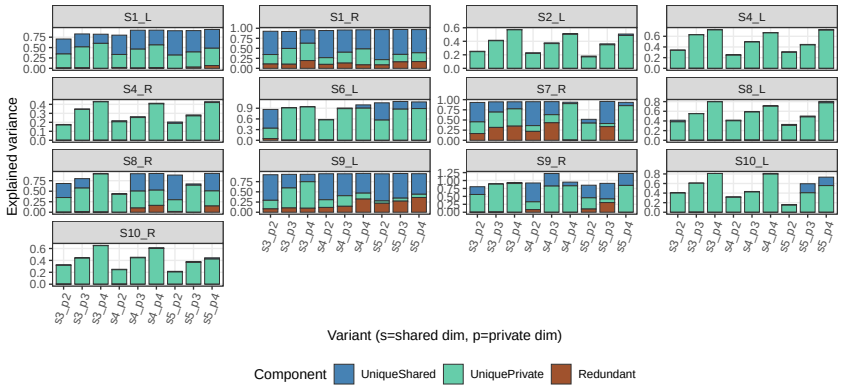
To adapt CTAE to intracranial LFP recordings from GPi and STN, we used the original two-region architecture without structural modifications. Following the original implementation, recordings from both regions were concatenated along the channel dimension and represented as multichannel temporal windows of shape $[N \times T \times C]$. All channels were z-scored prior to training. We trained



(a) GPi



(b) STN



(c) Thalamus

Figure A.11: Variance partition across latent dimensionalities. For each subject–hemisphere, we trained SPIRE with shared dimension $d_s \in \{3, 4, 5\}$ and private dimension $d_p \in \{2, 3, 4\}$. Each bar shows the fraction of explained variance in the observed signal that is uniquely attributable to shared latents, uniquely attributable to private latents, or redundantly captured by both. (a) GPi recordings. (b) STN recordings. (c) Thalamus recording. These sweeps were used to select the final (d_s, d_p) configuration per subject–hemisphere.

CTAE using the same train/validation splits as SPIRE and matched latent dimensionalities to the corresponding SPIRE configurations for each subject. Models were trained using the Adam optimizer

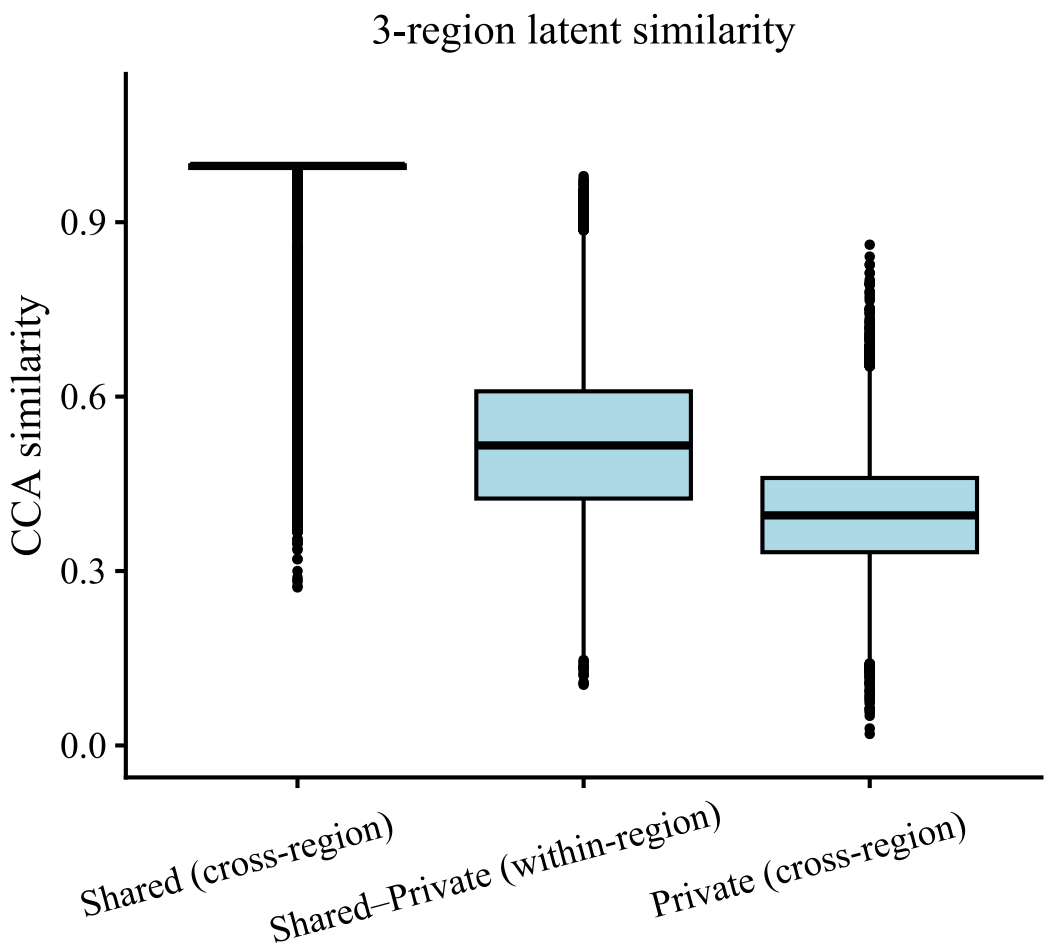


Figure A.12: Three-region latent similarity analysis. CCA similarity was computed among grouped latent pairs across GPi, STN, and thalamus. Shared cross-region latents showed the highest similarity, indicating recovery of a common multi-region latent manifold. Shared-private similarities were lower, reflecting limited within-region leakage, while private cross-region similarities were lowest, consistent with private latents capturing region-specific dynamics.

with learning rate 10^{-4}, two transformer layers, one attention head, fixed sinusoidal positional encodings, and early stopping based on validation reconstruction performance, and max 300 epochs.

CTAE produces time-resolved latent trajectories and therefore supports analyses requiring temporal latent dynamics, including latent similarity analysis and stimulation decoding. However, CTAE does not explicitly separate shared and private temporal dynamics through dedicated temporal alignment modules or variance-partitioning constraints.

Fundamental differences. CTAE was developed for time-resolved shared/private neural representation learning and is therefore the closest nonlinear baseline to SPIRE. It uses transformer-based latent representations together with reconstruction, cross-reconstruction, shared-latent alignment, and orthogonality objectives. SPIRE targets the same general shared/private representation problem but differs in several design choices motivated by continuous LFP: constrained temporal ConvAlign operators, near-identity mapping regularization, shared/private variance guards, shared-self reconstruction, and staged optimization of the shared and private branches. These differences motivate our comparison under nonlinear observation mappings, time-varying temporal offsets, and spectral confounding, where stable assignment of transferable and region-specific information is particularly challenging.

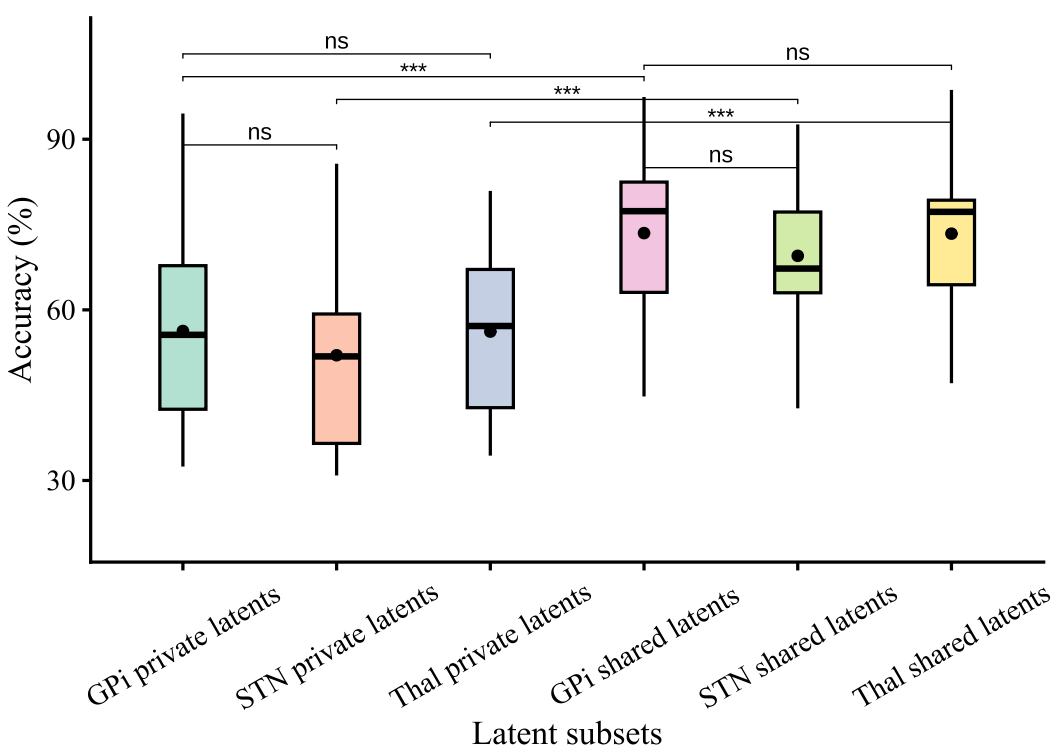


Figure A.13: Three-region decoding of GPi stimulation frequency from SPIRE latents. Random Forest classifiers were trained to decode stimulation frequency (off, 85, 185, and 250 Hz) from private and shared latent subsets of GPi, STN, and thalamus. Shared latents across all three regions achieved higher decoding accuracy than private latents, indicating that GPi stimulation induces structured changes in the shared multi-region latent manifold. Private latents showed lower and more variable accuracy, consistent with region-specific residual activity. Boxplots show median and interquartile range, whiskers extend to 1.5×IQR, and black dots denote means. Statistical comparisons are shown with conventional significance notation: ns, not significant; $^{***}p < 0.001$.

Attempted DLAG baseline. We sought to benchmark SPIRE against DLAG (Gokcen et al., 2022), which has been successfully applied to multi-area spiking data. While DLAG worked on our synthetic datasets (Section 4), it proved infeasible on the human intracranial recordings. Specifically:

- Ill-conditioned covariance matrices. After bipolar re-referencing and lag augmentation (Appendix A.3.1), the empirical covariance matrices often had deficient rank (e.g., eigenvalues ≈ 0), causing the Gaussian process kernel inversion in DLAG to fail despite added jitter.
- Model mismatch. DLAG assumes stationary Gaussian processes with smoothly varying latents, whereas intracranial LFPs exhibited strong nonstationarities that violate these assumptions.
- Data dimensionality. Practical channel counts (3–8 bipolar contacts per region) combined with thousands of short windows led to numerical instabilities in the Toeplitz inversion routines central to DLAG. Even under aggressive dimensionality reduction or simplified hyperparameters, the model either diverged or collapsed to degenerate solutions.

We tried multiple remedies—reducing latent dimensionality, adding jitter, trimming channels, and using minimal configurations—but none yielded stable training. We therefore conclude that DLAG, in its current form, is not well suited to our bipolarized intracranial recordings. This limitation highlights the need for alternative frameworks, like SPIRE, that explicitly handle bipolar data, short windows, and nonstationary dynamics.

Compute comparison All experiments were conducted on a local workstation equipped with an NVIDIA RTX A4000 GPU (17.17 GB VRAM), Intel Xeon Silver 4216 CPU (32 logical processors), and 192 GB RAM.

All measurements were performed in PyTorch using automatic mixed precision (AMP, fp16) and `torch.inference_mode()`, with batch size $B{=}8$ and inputs containing GPi/STN channels $(C_{\mathrm{gpi}}, C_{\mathrm{stn}}){=}(36, 12)$. Each latency value represents the mean over 50 forward passes following 10 warm-up passes, with device synchronization before and after timing. SPIRE was evaluated using `private_gate=1.0`. CTAE measurements used the original two-region transformer architecture with the same inference protocol as the real-data experiments.

We report two quantities per model (Table A.6): (i) trainable parameters (millions), and (ii) per-sequence inference latency (milliseconds). SPIRE was evaluated by a single forward pass on $(x_{\mathrm{gpi}}, x_{\mathrm{stn}})$. CTAE latency was measured using a single forward pass through the coupled transformer autoencoder without gradient computation.

Model	Params (M)	Seq length T	Inference (ms / seq)
SPIRE (ours)	0.116	247	4.45
CTAE	0.860	247	4.84

Table A.6: Compute comparison under matched batch size ($B{=}8$), channels $(36, 12)$, and AMP (fp16). Latencies are wall-clock forward-pass times averaged over 50 runs following 10 warm-up iterations.

Note. These measurements reflect inference only; training introduces additional backward-pass and optimizer overhead.

DLAG: Because DLAG is implemented in MATLAB and relies on an expectation–maximization (EM) procedure rather than a simple forward pass, we measured its compute cost separately. We used synthetic dataset D3 with 8 channels per region and prepared trial-wise sequences as in the DLAG demo code. To obtain an inference-like latency, we timed a single EM iteration (E–M step) with saving disabled using MATLAB's `timeit`, then normalized by the number of sequences. Under these conditions, DLAG required on average 155.85 ms per sequence. We do not report parameter counts, since DLAG's MATLAB implementation represents parameters via Gaussian process covariances and delay variables rather than explicit trainable tensors.

A.4 Additional functional validation and robustness

These analyses are secondary to the main narrative and evaluate robustness to lag augmentation, latent dimensionality, and alternative reconstruction objectives.

A.4.1 Complete nonlinear ground-truth recovery matrix

Using nonlinear Extra Trees regression with five-fold trial-wise cross-validation, we evaluated SS, SP, PP, and PS as defined in Section 4. Values are mean (SD) across four matched seeds. SPIRE satisfies $SS > SP$ and $PP > PS$ in every regime. DLAG shows the strongest factor segregation in linear D0, while its specificity degrades under nonlinear mappings, time-varying delay, and spectral confounding. CTAE shows approximately zero or negative specificity in several settings Table A.7.

A.4.2 No-lag retraining

All real-data baselines in the original comparison received the same 0-3 sample lag-augmented inputs. To test whether lag stacking itself generates the reported organization, we retrained SPIRE without lag augmentation while keeping architecture, objective, schedule, optimizer, training duration, and stimulation-independent selection criteria fixed. Valid decompositions were obtained in 16/17 no-lag hemispheres versus 17/17 with lagging. Across 16 matched hemispheres, a hemisphere-level model showed a strong latent-pair effect ($F_{3,120} = 33.80$, $p = 6.59 \times 10^{-16}$), but no augmentation effect ($p = 0.624$) and no augmentation-by-pair interaction ($p = 0.961$). The reconstruction hierarchy and stimulation conclusion were preserved. Under GPi stimulation, all four shared-versus-private centroid-shift contrasts remained significant without lagging ($p < 0.0001$). Under STN stimulation,

Table A.7: Complete nonlinear intended-versus-crossed recovery matrix. Specificity margins are $SS - SP$ and $PP - PS$.

Regime	Model	SS	SP	PP	PS	$SS - SP$	$PP - PS$
D0	SPIRE	0.72 (.02)	0.24 (.15)	0.64 (.11)	0.24 (.01)	0.48 (.16)	0.40 (.11)
	DLAG	0.89 (.03)	0.09 (.02)	0.83 (.06)	0.15 (.03)	0.81 (.04)	0.68 (.09)
	CTAE	0.29 (.10)	0.22 (.05)	0.31 (.02)	0.32 (.06)	0.07 (.12)	-0.01 (.07)
D1	SPIRE	0.76 (.01)	0.31 (.03)	0.60 (.01)	0.20 (.01)	0.44 (.02)	0.39 (.02)
	DLAG	0.64 (.01)	0.21 (.01)	0.48 (.03)	0.28 (.01)	0.43 (.01)	0.20 (.02)
	CTAE	0.14 (.05)	0.12 (.02)	0.17 (.04)	0.18 (.04)	0.02 (.05)	-0.01 (.02)
D2	SPIRE	0.75 (.00)	0.33 (.01)	0.58 (.01)	0.20 (.02)	0.42 (.01)	0.38 (.02)
	DLAG	0.57 (.01)	0.20 (.00)	0.48 (.00)	0.33 (.02)	0.37 (.01)	0.16 (.02)
	CTAE	0.12 (.06)	0.08 (.04)	0.16 (.05)	0.17 (.04)	0.04 (.03)	-0.01 (.03)
D3	SPIRE	0.54 (.02)	0.11 (.02)	0.55 (.04)	0.28 (.01)	0.42 (.00)	0.27 (.03)
	DLAG	0.42 (.03)	0.26 (.02)	0.46 (.01)	0.33 (.02)	0.15 (.04)	0.13 (.02)
	CTAE	0.17 (.05)	0.23 (.04)	0.35 (.03)	0.22 (.04)	-0.06 (.02)	0.14 (.05)

Table A.8: Median CCA organization with and without lag augmentation across 16 matched hemispheres.

Latent pair	Lag	No lag
Shared GPi-STN	0.987	0.750
Shared-private GPi	0.610	0.571
Shared-private STN	0.568	0.626
Private GPi-STN	0.390	0.406

all four retained the same direction and three remained significant after correction. Lag stacking therefore improves fitting robustness but is not required to produce the main shared-over-private stimulation result.

A.4.3 Latent-dimensionality sensitivity

We evaluated all nine combinations $d_{\mathrm{sh}} \in \{3, 4, 5\}$ and $d_{\mathrm{pr}} \in \{2, 3, 4\}$ using configurations selected only by off-stimulation variance criteria. Across 64 hemisphere-configuration models for GPi stimulation and 52 for STN stimulation, there was a significant overall shared-versus-private stimulation effect ($F = 16.15$, $p = 8.2 \times 10^{-5}$), but no interaction with shared dimension ($p = 0.410$) or private dimension ($p = 0.176$). The effect magnitude differed by stimulation site. A conservative analysis retaining partially collapsed configurations also preserved an overall shared/private effect with no dimension interaction. Thus, dimension dependence was not detected for the principal contrast.

A.4.4 Private-only reconstruction objective

We evaluated the performance adding another loss term requiring the private branch alone to reconstruct the complete same-region signal. At a relatively strong staged weight (maximum 0.03), only 2/17 hemispheres retained any configuration satisfying the predefined variance-partition criteria; most failures reflected excessive shared/private redundancy or insufficient unique shared variance. Under controlled synthetic ground truth, smaller maximum weights of 0.01 and 0.005 did not produce a uniform improvement in CCA or nonlinear intended recovery/specificity after correction Table A.9. We additionally retrained all 17 human hemispheres with maximum $\lambda_{\mathrm{private}} = 0.005$ while retain-

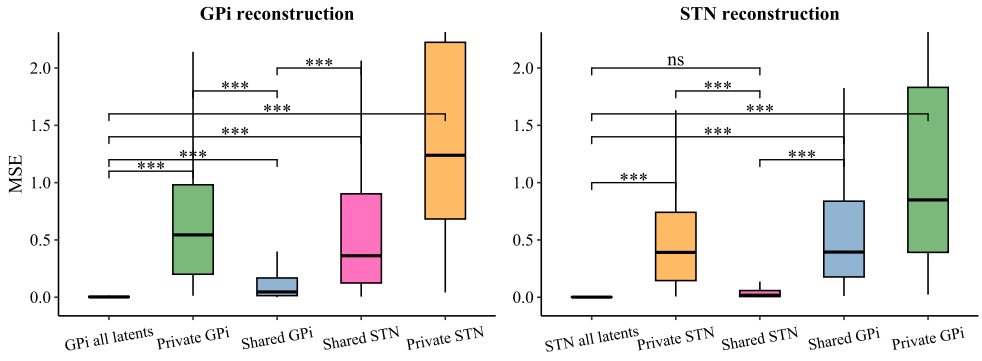
Table A.9: CCA shared/private specificity margins for the original objective and two lower-weight private-only reconstruction variants.

Regime	Original	Private 0.01	Private 0.005
D0	0.360 / 0.397	0.455 / 0.446	0.454 / 0.455
D1	0.644 / 0.548	0.647 / 0.526	0.645 / 0.548
D2	0.607 / 0.539	0.610 / 0.524	0.606 / 0.540
D3	0.472 / 0.463	0.482 / 0.475	0.478 / 0.480

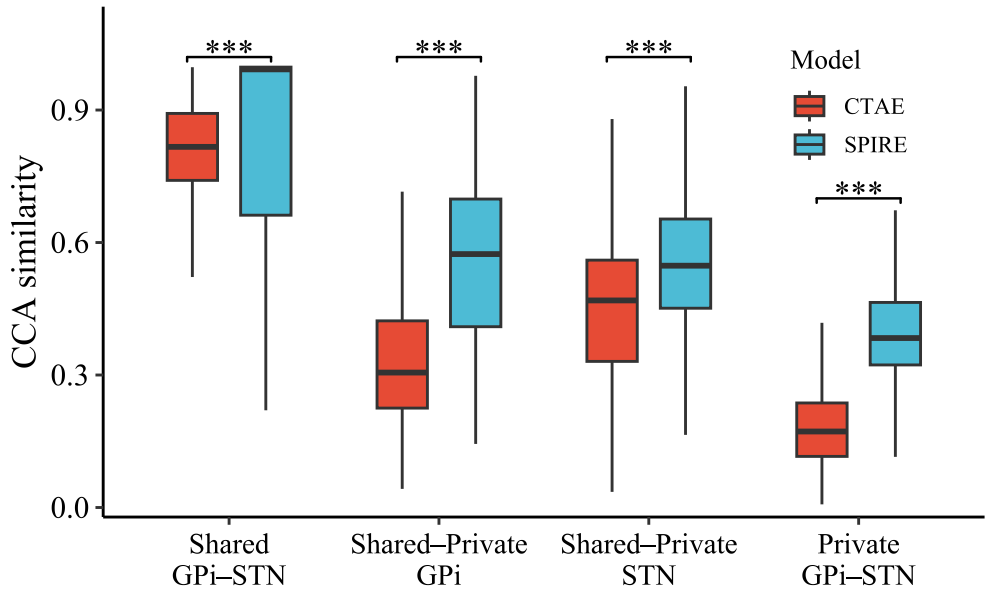
ing each hemisphere's previously selected dimensionality. Private-only reconstruction improved (GPi MSE 0.8659→0.482; STN 0.6110→0.376), while full and shared reconstruction remained similar and cross-region private reconstruction remained poor. However, nonlinear organization exposed a tradeoff: excess R^2 was 0.574 for shared-shared, 0.530 for shared-to-private, 0.315 for private-to-shared, and -0.001 for private-private cross-region prediction. The shared-shared minus shared-to-private margin therefore fell from 0.132 in the original model to 0.044 in the modified model. The main stimulation findings remained unchanged. This variant is viable, but it does not uniformly improve functional separation, motivating retention of the original objective.

A.5 Secondary human-LFP diagnostics

The main text uses nonlinear cross-region predictability as the primary functional validation of the shared/private decomposition. Here we provide additional held-out diagnostics that characterize complementary properties of the learned representation, including reconstruction quality, latent organization, and similarity between SPIRE and CTAE latent spaces Figure A.14. These analyses are not used to define the shared/private interpretation but provide additional characterization of the learned representations.



(a) Two-region reconstruction.



(b) CCA organization and CTAE comparison.

Figure A.14: Secondary held-out human-LFP diagnostics. (a) Reconstruction error when decoding GPi and STN activity from full, shared-only, and private-only representations. (b) CCA similarity between shared and private latent spaces for SPIRE and CTAE. These analyses provide complementary characterization of representation quality and organization; the primary evidence for functional segregation is the nonlinear cross-region prediction analysis in the main text.

A.6 Frozen-representation validation across Train-Off, Test-Off, and Test-On

This analysis verifies that models learned and selected on baseline data remain usable on stimulation recordings; it is not presented as a separate contribution. No model was retrained, fine-tuned, or reselected using stimulation data. We compared sampled baseline training windows (Train-Off), held-out baseline recordings (Test-Off), and held-out stimulation recordings (Test-On) in 17 GPi-stimulation and 13 STN-stimulation hemispheres.

Full reconstruction R^2 decreased under stimulation as expected for a distribution shift, but remained high Table A.10. The objective-related hierarchy full error $<$ shared-self error $<$ aligned cross-shared error was preserved for both regions in every hemisphere (17/17 under GPi stimulation and 13/13 under STN stimulation), and aligned cross-shared reconstruction did not significantly degrade from Test-Off to Test-On. Latent variance showed no significant overall stage effect ($p = 0.282$ for the GPi-stimulation cohort; $p = 0.201$ for STN stimulation), and shared SD changed by only -1.8% to

Table A.10: Mean Test-On full-reconstruction R^2.

Cohort	GPi target	STN target
GPi stimulation	0.888	0.952
STN stimulation	0.853	0.905

+3.2% from Test-Off to Test-On. Shared-private dependence increased modestly under stimulation but no region-specific Test-On versus Test-Off contrast survived correction.

Pointwise shared-alignment error increased under stimulation (48% for GPi stimulation and 60% for STN stimulation), which is compatible with the intended use of the representation to measure reorganization. As a secondary check not used during training, shared GPi-STN CCA remained larger than private GPi-STN CCA in every hemisphere A.11. Thus stimulation changes geometry within

Table A.11: Functional organization of the frozen representation on unseen stimulation data.

Cohort	Shared CCA	Private CCA	Shared $>$ private
GPi stimulation	0.588	0.189	17/17
STN stimulation	0.516	0.197	13/13

the learned coordinate system without causing collapse or reversal of the defining shared/private organization.

A.7 Matched CTAE stimulation details

The main text reports the class-averaged CTAE-SPIRE comparison. Region-specific CTAE mean accuracies were 31.0% for private GPi, 31.0% for private STN, 38.2% for shared GPi, and 38.5% for shared STN. Averaged across regions, CTAE shared accuracy exceeded private accuracy by 7.41 points ($F_{1,48} = 44.53$, $p = 2.34 \times 10^{-8}$), with no region effect ($p = 0.908$) and no class-by-region interaction ($p = 0.893$). SPIRE exceeded CTAE separately for private GPi, shared GPi, private STN, and shared STN; both class-averaged differences in Table 4 had Holm-adjusted $p < 0.0001$.

A.8 External CA1-PFC LFP transfer example

To evaluate applicability beyond the pediatric GPi-STN dataset, we applied the complete SPIRE architecture/objective to a distinct public rat CA1-PFC LFP dataset. This is a preliminary transfer example, not a main claim of universal or tuning-free generalization. The numerical optimization schedule was adapted once at the dataset level, then held fixed across all four animals. Every rat used three shared and three private dimensions with no rat-specific hyperparameter or dimension search. Models were trained separately using complete epochs split 70/15/15 into training, validation, and untouched test sets.

Table A.12: Held-out CA1-PFC latent organization in four rats.

Rat	Shared CCA	Private CCA	Within-region sh-pr CCA
T1	0.987	0.064	0.135-0.190
S2	0.995	0.022	0.038-0.122
Q1	0.997	0.067	0.033-0.052
R1	0.660	0.049	0.213-0.230

The intended ordering held in 4/4 animals; mean shared cross-region CCA was 0.910 versus 0.050 for private cross-region CCA A.12.

Cross-region private prediction remained at the structured temporal null while shared prediction remained strong in every animal A.13.

The reconstruction hierarchy full $<$ shared-only $<$ private-only held for both regions in all four rats. Four concordant animals give a one-sided exact sign-test value of $p = 0.0625$, so we emphasize

Table A.13: Nonlinear held-out cross-predictability in the CA1-PFC transfer example.

Rat	Shared-shared R^2	Shared-to-private R^2	Private-private R^2
T1	0.978	0.323	-0.008
S2	0.990	0.461	-0.018
Q1	0.997	0.345	-0.013
R1	0.620	0.154	-0.016

Table A.14: Held-out reconstruction MSE in the external LFP transfer example; each cell is full/shared-only/private-only.

Rat	CA1	PFC
T1	0.0044 / 0.158 / 0.863	0.0358 / 0.182 / 1.120
S2	0.0008 / 0.036 / 0.276	0.0245 / 0.267 / 1.074
Q1	0.0017 / 0.059 / 0.686	0.0260 / 0.196 / 2.566
R1	0.0025 / 0.175 / 0.809	0.0361 / 0.171 / 0.965

consistency rather than population-level inference A.14. This experiment shows practical dataset-level adaptation without animal-specific redesign; it does not establish universal transfer.

A.9 LIMITATIONS OF LARGER MULTI-REGION AND MULTI-SUBJECT EXTENSIONS

For R regions, SPIRE instantiates $R(R-1)$ directed alignment pathways. Region-wise losses are averaged across regions and pairwise losses across directed pairs, so the number of coefficient *types* does not increase, but memory and computation scale quadratically in the dense implementation. Sparse anatomical/data-driven alignment graphs or alignment to a learned reference space are natural extensions.

The present one-shared/one-private formulation also does not explicitly distinguish globally shared factors from factors shared only within a subset of regions. A subnetwork-specific process may be represented partly in shared space, partly in private space, or weakened by alignment to less-related regions. The current three-region analysis demonstrates recovery of a common GPi-STN-thalamic manifold, not arbitrary subnetwork discovery. Hierarchical global/subnetwork/private latents remain future work.

Finally, current models are fit separately by hemisphere. A multi-subject extension could combine subject-specific input projections/decoders with a common region-level shared space and subject-specific private components, informed by anatomical labels or electrode coordinates. Because electrodes sample different tissue across subjects, cross-subject correspondence cannot be assumed and is not established here.