Temporal PHATE: A multi-view manifold learning method for brain state trajectories

Erica L. Busch[1], Jessie Huang[2], Andrew Benz[3], Tom Wallenstein[2], Guillaume Lajoie[7,8], Guy Wolf[7,8], Smita Krishnaswamy[2,4,5,†,*], and Nicholas B. Turk-Browne[1,6,†]

[1] Department of Psychology, Yale University
[2] Department of Computer Science, Yale University
[3] Department of Mathematics, Yale University
[4] Department of Genetics, Yale University
[5] Program in Applied Mathematics, Yale University
[6] Wu Tsai Institute, Yale University
[7] Department of Mathematics and Statistics, Université de Montréal
[8]Mila – Quebec AI Institute
[†]Equal contribution
[*]Corresponding author: Smita.Krishnaswamy@Yale.edu

1 Abstract

Brain activity as measured with functional magnetic resonance imaging (fMRI) gives the illusion of intractably high dimensionality, rife with collection and biological noise. Nonlinear dimensionality reductions like PCA, UMAP, tSNE, and PHATE have proven useful for high-throughput biomedical data, but have not been extensively used in fMRI, which is known to reflect the redundancy and co-modulation of neural population activity. Here we take the manifold-geometry preserving method PHATE and extend it for use in brain activity timeseries data in a method we call temporal PHATE (T-PHATE). We observe that in addition to the intrinsically lower dimensionality of fMRI data, it also has significant autocorrelative structure that we can exploit to faithfully denoise the signal and learn brain activation manifolds. We empirically validate T-PHATE on three fMRI tasks and show that T-PHATE manifolds improve visualization fidelity, stimulus feature classification, and neural event segmentation. T-PHATE demonstrates impressive improvements over previous cutting-edge approaches to understanding the nature of cognition from fMRI and bodes potential applications broadly for high-dimensional datasets of temporally-diffuse processes.

2 Introduction

As we walk through the world, our brains perform innumerable operations to process, represent, and integrate information. How does a system composed of limited computational units represent so much while handling a constant barrage of incoming information? How does the way the brain processes this information support our psychological experience? One possible explanation is that individual neural units operate in tandem to encode and update information in population codes, which exponentially expands the distinct ways information can be represented by a fixed number of units. Analysis of neural population codes better predicts behavior than analysis of single neurons, particularly as we consider how representations change over time in response to new information [1]. For example, populations of neurons in the olfactory system represent odor codes using dynamic processes, which afford rapid classification and learning of scents [2]. Motor population codes underlie the ability to plan, prepare, and execute movements [3], and populations in the face processing network represent face identity in both static [4] and viewpoint-invariant representations [5]. More recent studies have shown that though neural population codes exist in high dimensions, where each neural unit represents its own dimension, these dimensions are not orthogonal: they are redundant [6] and from them emerge dominant latent signals that code for behavior and information [7, 8].

Principles of neural population codes were defined using direct neural recordings in nonhuman primates but extend to noninvasive, indirect measurements of brain activity using functional magnetic resonance imaging (fMRI) in humans. Over the recent decades, multivariate pattern analysis of fMRI data has afforded valuable insights into the structure and content of cognitive representations by looking at high-dimensional patterns of activity across many voxels instead of a singular voxel or averaged regional signal alone [9, 10]. By considering patterns of brain activity, fMRI has provided insight into how the human brain represents low-level [11] and high-level [10, 12, 13] sensory stimuli, and higher-order processes like memory [14], decision making, and theory of mind [15]. These insights, however, have largely come from group-level, static analyses, which require aggregation over multiple subjects and timepoints to draw conclusions, due to the high levels of spatio-temporal noise inherent to fMRI. Noise in fMRI originates at the collection level, from subject movement and scanner conditions, to the BOLD signal being measured and the low temporal resolution (typical brain volume acquisition time is around 1.5 seconds, far slower than most cognitive processes) [16]. Further, the BOLD signal is a slow, secondary measurement of brain activity, peaking approximately 4-5 seconds after stimulation before returning to baseline. Finally, complex cognitive processes like memory [17], narrative comprehension [18–20], or learning are dynamic: they do not exist at a singular timepoint but unfurl over longer periods of time [21]. Such complex yet fundamental processes require integration across modalities and time, so the brain likely performs computations on these representations at varying timescales to support our conscious experience [22, 23].

In sum, many brain representations of interest are coded in high-dimensional patterns of activation [2–4, 9, 10] which are often characterized by low-dimensional latent signals [1, 8, 24]. fMRI affords unique insight into the healthy, behaving human brain, but

the data is noisy, high-dimensional, of low temporal resolution, and the measured signal is extended in time, giving rise to high signal autocorrelation [16]. Further, complex cognitive processes are extended in time. To account for these issues, we need to consider the various ways in which both time and dimensionality interact with the measured BOLD signal in order to optimally characterize fMRI activity at the single-subject level.

We introduce a nonlinear, temporal manifold learning approach designed for the high noise and temporal properties of both fMRI data and complex cognitive processing. Nonlinear dimensionality reduction algorithms have been applied in past studies to characterize the geometry and underlying temporal dynamics in neural recordings [7, 24, 25] by integrating local similarities among data points into a global representation [8, 26–29]. We build upon a prior manifold learning algorithm PHATE [26] designed for high-throughput, high-noise biomedical data, with a second, explicit model of the activity's temporal properties. This second view, learned from the data, models temporal affinity autocorrelation of the BOLD signal and the temporal dynamics that may be specific to a given stimulus or brain region (Fig. 1A). Our algorithm, temporal-PHATE or T-PHATE, better captures signal in simulated data under a range of noise conditions than other well-known manifold learning methods (including principal components analysis (PCA), uniform manifold approximation (UMAP) [30], and PHATE [26]) using a measurement of manifold preservation called DeMAP [26] (Fig. 1B). Unlike prior studies [24], we are interested in learning a manifold that captures temporal dynamics within a given naturalistic task that could later be extended to new subjects [31] or predict future brain states. Thus, we test T-PHATE on two movie-viewing fMRI datasets and find that it both denoises the data and affords enhanced access to brain-state trajectories than either voxel resolution data or a variety of other dimensionality reduction approaches. By mitigating noise and voxel covariance, this subspace yields clearer access to regional dynamics of brain signals, which we can then relate to time-dependent psychological phenomena. In all, we show that information about cortical dynamics during naturalistic stimuli lies in a low-dimensional latent space that can be best modeled with a nonlinear, temporal manifold.

3 Results

3.1 Evaluating manifold quality

Functional magnetic resonance imaging (fMRI) is a powerful tool for studying the neural bases of higher-order cognitive abilities in healthy adults. fMRI data is highly noisy in both space and time, with the measured BOLD signal canonically peaking 4-5 s after stimulus onset before slowly returning to baseline. Naturalistic, complex stimuli like movies are increasingly used to probe real-world type cognition and reasoning in the brain. These stimuli are dynamic, with conversations and plot lines playing out over different timescales, which cause the autocorrelation of brain responses to extend beyond the curve of the BOLD signal and vary according to the temporal integrative properties of individual brain regions. Thus, we designed temporal PHATE (T-PHATE) as a variant of the PHATE algorithm that combines the robust manifold-geometry preserving prop-

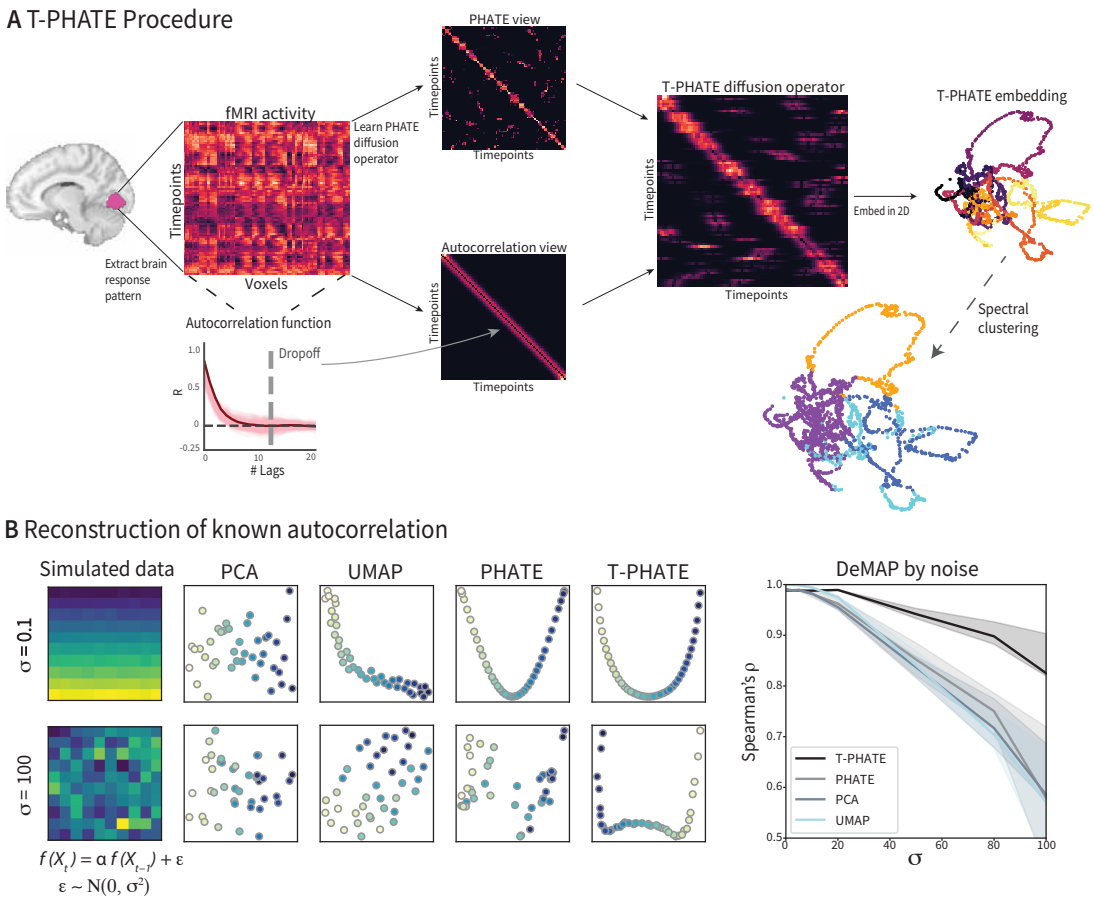

Figure 1: **A.** Temporal PHATE (T-PHATE) procedure. A multivariate pattern of fMRI activation over time (timepoints by voxels) is extracted from a region of interest (ROI). PHATE is then run to learn a PHATE-based affinity matrix between timepoints. The autocorrelation function is also estimated for the data separately to make an autocorrelation-based affinity matrix. These two views are combined into the T-PHATE diffusion operator then embedded into lower dimensions. **B.** We tested this approach on simulated data with a known autocorrelative function and varied the error added to the signal. T-PHATE both visually better reconstructs the autocorrelative signal from the noise and quantitatively better preserves manifold distances, compared with PCA, UMAP, and PHATE.

erties of PHATE with an explicit model of the data's autocorrelation in a dual-diffusion operator. We compared manifolds learned with T-PHATE to those from PCA, PHATE, and UMAP.

We validated the quality of manifolds learned from data with a simulated high – autocorrelative signal using Denoised Manifold Affinity Preservation (DeMAP) [26]. DeMAP takes pristine, noiseless data (here, a simulated multivariate timeseries where $f(X_t) = \alpha f(X_{t-1})$) and computes the Spearman correlation between the geodesic distances of the noiseless data and the Euclidean distances on the embeddings learned from noisy data. Noisy data was generated by adding an error term ϵ to the pristine data, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$. We tested robustness to noise by varying $\sigma = (0, 100)$. Higher DeMAP

scores suggest that an embedding method performs effective denoising and preserves the geometric relationships in the original data. DeMAP results for PCA, UMAP, PHATE, and T-PHATE are shown in Fig. 1B. At low levels of noise, all methods achieve high DeMAP scores, though PCA visualizations do not appear to reflect much structure. With increasing noise, T-PHATE outperforms the other methods at denoising the simulated data, providing visualizations that most closely resemble that of lower noise conditions.

Having validated that T-PHATE learns meaningful manifolds from simulated data better than previous methods particularly under noisy conditions, we next extended this analysis to our fMRI data. First, we embedded the movie viewing data for the *Sherlock* and *StudyForrest* datasets with PCA, UMAP, PHATE, and T-PHATE in two dimensions to visually inspect the properties of the data highlighted by the manifold. Visually, T-PHATE embeddings denoise the timeseries data and capture stimulus-related trajectory structure better than the other methods, as shown in all brain regions for a sample subject (Fig. 2A) as colored by timepoint. PCA visualizations show no apparent structure or clustering, whereas UMAP shows at least slight clustering of nearby timepoints in high visual (HV) and early auditory (EA) cortices but often creates small, shattered clusters with local structure placed through the latent space. PHATE visually yields slight improvements over UMAP, notably in HV and EA. Clusters in these plots appear larger and less disjointed, connected by tendrils that flow from a hub-like center space where the majority of clusters exist in the embedding. T-PHATE, however, reveals remarkable trajectories through the latent space that are clearly reflective of temporal structure but also shows a hub in the center of the space. T-PHATE manifolds also better reflect non-temporal attributes of the movie, as shown by classification of stimulus features at given timepoints. We ran a support vector classification to predict whether music was playing at a given timepoint from the embeddings of neural data (or the full voxel-resolution data). Significance of these predictions was tested by shifting the labels over 1 TR for every TR in the timeseries to establish a null distribution; results are presented as the z-statistic of the true prediction accuracy between brain data and movie labels versus the shifted versions.

The clarity of the dynamic structure in T-PHATE manifolds lead us to ask two additional questions. First, does T-PHATE require *autocorrelation* information or would other types of temporal information provide a similar advantage to the informativeness of the manifold? To address this, we designed two additional temporal algorithms, Smooth PHATE and PHATE+Time, that used distinct models of built temporal structure in the manifold calculation. Smooth PHATE performs temporal smoothing along each voxel with a double gamma function (the canonical hemodynamic response function), which is a common preprocessing step for fMRI data generally, before embedding with PHATE. PHATE+Time adds the TR label for each timepoint as an additional feature vector to the voxel data before embedding with PHATE. PHATE+Time manifolds show the clearest temporal trajectory through the embedding space, unraveling the brain data in time due to the disproportionate weighting of time in the input data (Fig. 3A) most clearly unravel the timeseries data, with less of the hub structure shown T-PHATE. Smooth PHATE shows more comparable structure to PHATE or UMAP, with more shattered trajectories and mild clustering. This lead us to ask whether adding a time vector to the data provides

as valuable information as autocorrelation to manifold learning when the stimulus is not intrinsically linked to time, as it is with a movie stimulus. We next embedded the *Study-Forrest* object category localizer data with T-PHATE, PHATE+Time, and Smooth PHATE for early visual (EV) and high visual (HV) ROIs. In this task, subjects were shown static images of faces, bodies, small objects, houses, scenes, and scrambled images randomly with no temporal structure. A meaningful manifold of brain activity in visual regions during this task would show clustering of stimulus types, which would be independent of time as the stimulus presentation was random. T-PHATE manifolds show the best clustering of stimulus categories in the embedding space, compared with PHATE+Time and Smooth PHATE. Smooth PHATE showed some clustering but less pronounced, whereas PHATE+Time showed very minimal clustering by category but instead still retained temporal structure.

We quantified ability to distinguish object categories within the latent space using a within-subject support vector classifier with leave-one-run-out cross-validation (6-way classification; chance = $1/6$) (Fig. 3C). In EV, classification accuracy on T-PHATE embeddings surpassed all other methods by $> 5\%$ on average (T-PHATE M=0.51, CI=[0.42, 0.6]; voxel M=0.46, CI=[0.38, 0.55]; PCA M=0.43, CI=[0.34, 0.52]; UMAP M=0.4, CI=[0.39, 0.42]; PHATE M=0.39, CI=[0.37, 0.42]; PHATE+Time M=0.38, CI=[0.37, 0.39]; Smooth PHATE M=0.42, CI=[0.38, 0.47]). In HV, T-PHATE outperformed all dimensionality reduction methods by $> 7\%$ on average, and the full voxel resolution data by 2% (T-PHATE M=0.55, CI=[0.47, 0.62]; voxel M=0.53, CI=[0.46, 0.61]; PCA M=0.48, CI=[0.46, 0.51]; UMAP M=0.42, CI=[0.41, 0.44]; PHATE M=0.43, CI=[0.42, 0.45]; PHATE+Time M=0.38, CI=[0.37, 0.39]; Smooth PHATE M=0.48, CI=[0.41, 0.56]). This analysis showed that T-PHATE captures structure in brain activity via the autocorrelative denoising, whereas adding time explicitly into the manifold does little for capturing the data geometry aside from unrolling the signal in time. Temporal smoothing provides an advantage over time-agnostic dimensionality reduction (PCA, PHATE, UMAP), but less of an advantage than the autocorrelation kernel in either the non-temporal localizer task or the dynamic movie task.

3.2 Event segmentation

The previous analysis established that incorporating temporal dynamics (namely autocorrelation) into a manifold learning algorithm helps denoise fMRI data given its spatiotemporal noise, regardless of stimulation type. Our second hypothesis was that T-PHATE would recover cognitive signals that are represented over time, such as neural event segmentation. To quantify the degree to which a manifold captures neural dynamics associated with the stimulus, we ran a Hidden Markov Model (HMM) to segment the data into discrete neural events and measure the separability of timepoints representing the same neural event versus distinct ones. We started by learning the optimal number of events experienced by each ROI using a leave-one-subject-out cross validation procedure on the voxel resolution data (Fig. 4). We find a pattern of results consistent across datasets where the number of events K represented by a brain region is higher for early sensory cortices (EV) than late sensory cortices (HV) and lower for integrative cortices

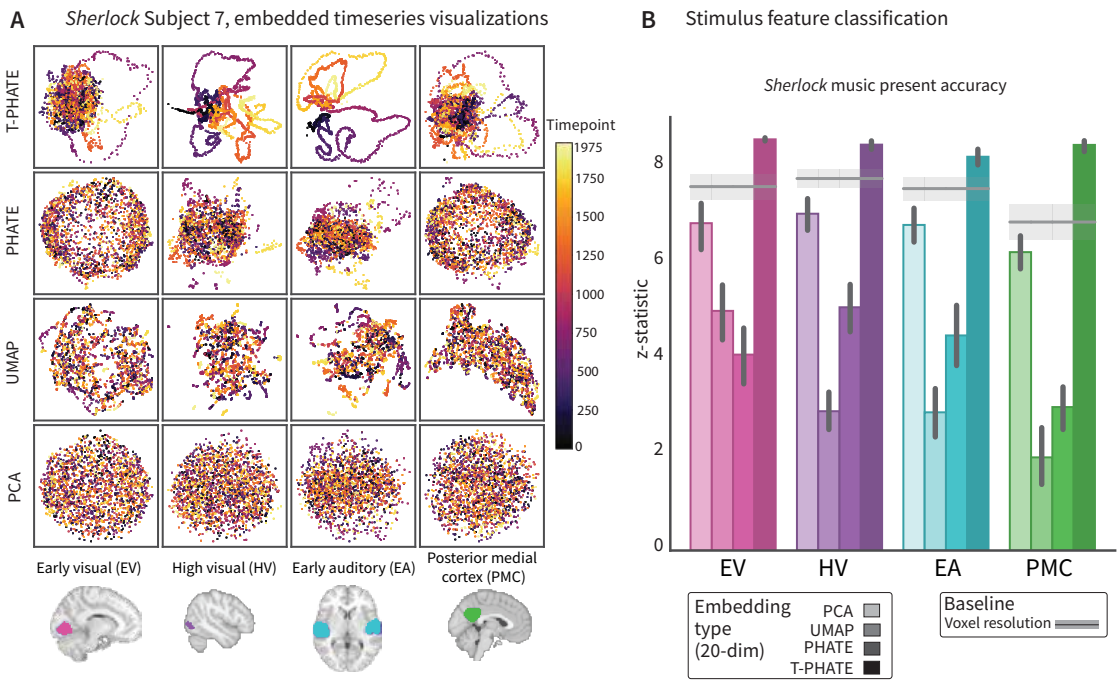

Figure 2: **A**. Manifold embeddings using PCA, UMAP, PHATE, and T-PHATE for the *Sherlock* dataset subject 7 in 4 regions of interest. Points are colored by time. **B**. Support vector classification accuracy predicting music presence for each timepoint. Results presented as z-statistic of true prediction versus shifted timeseries to simulate a null distribution with the same temporal properties. Extended SVC results included in Figure 8.

(PMC) (values presented as mean number of events $\pm$ s.d. across subjects): *Sherlock* EV $= 53 \pm 12$, HV $= 21 \pm 7$, AE $= 18 \pm 5$, PMC $= 22 \pm 6$; *StudyForrest* EV $= 111 \pm 43$, HV $= 69 \pm 17$, AE $= 74 \pm 18$, PMC $= 96 \pm 24$. after fixing the K parameter for each ROI based on the voxel data, we learned the number of manifold dimensions that would best reveal temporal structure for each region, subject, and method with the same cross-validation procedure. Number of dimensions is low (between 3-6 dimensions) with small variance across subjects and method. This indicates substantial redundancy, covariance, and noise among voxels in the ROIs, obscuring meaningful, dominant signals (Fig. 7).

A goal of T-PHATE is to recover dynamic cognitive signals that may have been previously obscured by the noise of redundancy and noise of the data. If dynamical structure during movie watching is unveiled in the T-PHATE latent space, we would see better success of the HMM on data embedded with T-PHATE than to the other methods or the voxel resolution data. This can be framed as 1) how well does the model fit the data (quantified using log-likelihood) and 2) how well do boundaries learned by the model capture structure of the neural data. For each subject, ROI, and embedding method, we fit a HMM with K hidden states on a M-dimensional embedding (where hyper-parameters K and M were cross-validated at the subject level for ROIs and methods, respectively). Model fit is quantified after hyper-parameter tuning using the log-likelihood, and T-PHATE shows significantly better model fit across both the *Forrest* and *Sherlock* datasets in all regions

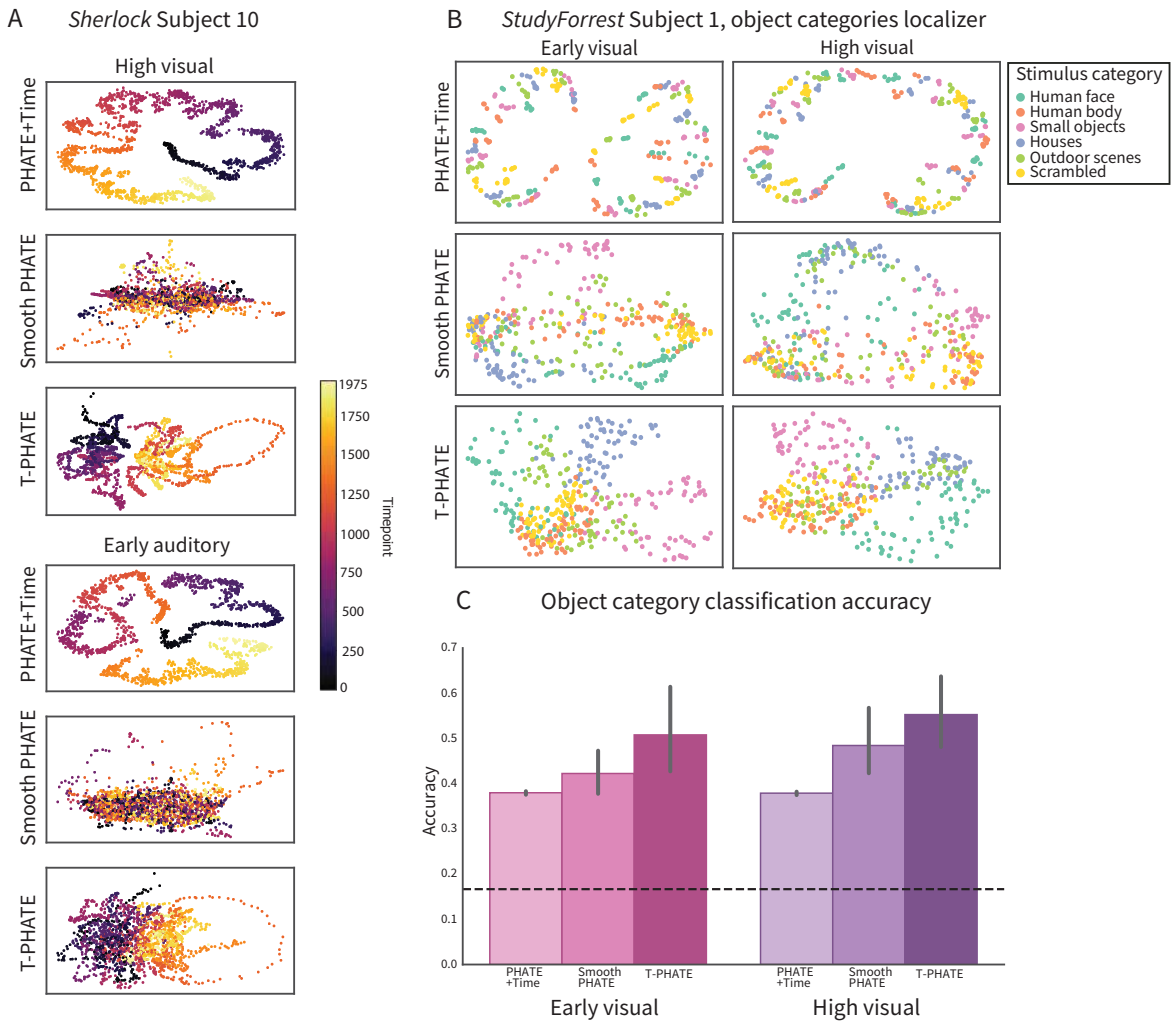

Figure 3: **A**. Manifold embedding of *Shelock* data in high visual (HV) and early auditory (EA) ROIs with T-PHATE, PHATE+Time, and Smooth PHATE, colored by time. **B** *StudyForrest* localizer data embeddings using T-PHATE, PHATE + Time, and Smooth PHATE in EV and HV. T-PHATE reveals the best clustering according to stimulus category, showing that the manifold successfully denoises data to learn a more accurate manifold. **C** T-PHATE manifolds also afford better stimulus category classification accuracy. Bars represent average accuracy across subjects; error bars represent 95% CIs of the mean from bootstrap resampling; dashed line represents chance.

(aside from the *Sherlock* PMC), visualized as negative log-likelihood, where closer to zero means a better model fit (Fig. 5A). This indicates a higher probability of the HMM's learned segmentation fitting the test data. To validate whether the boundaries identified actual differences in the data, we then applied the boundaries learned during the HMM fitting procedure to the data and asked whether pairs of timepoints that fall within the same event are more similar than pairs that fall across different events. To make sure this result was not driven by the inherent correlation of proximal timepoints given BOLD signal dynamics, we restricted the pairs of timepoints that could go into this calculation

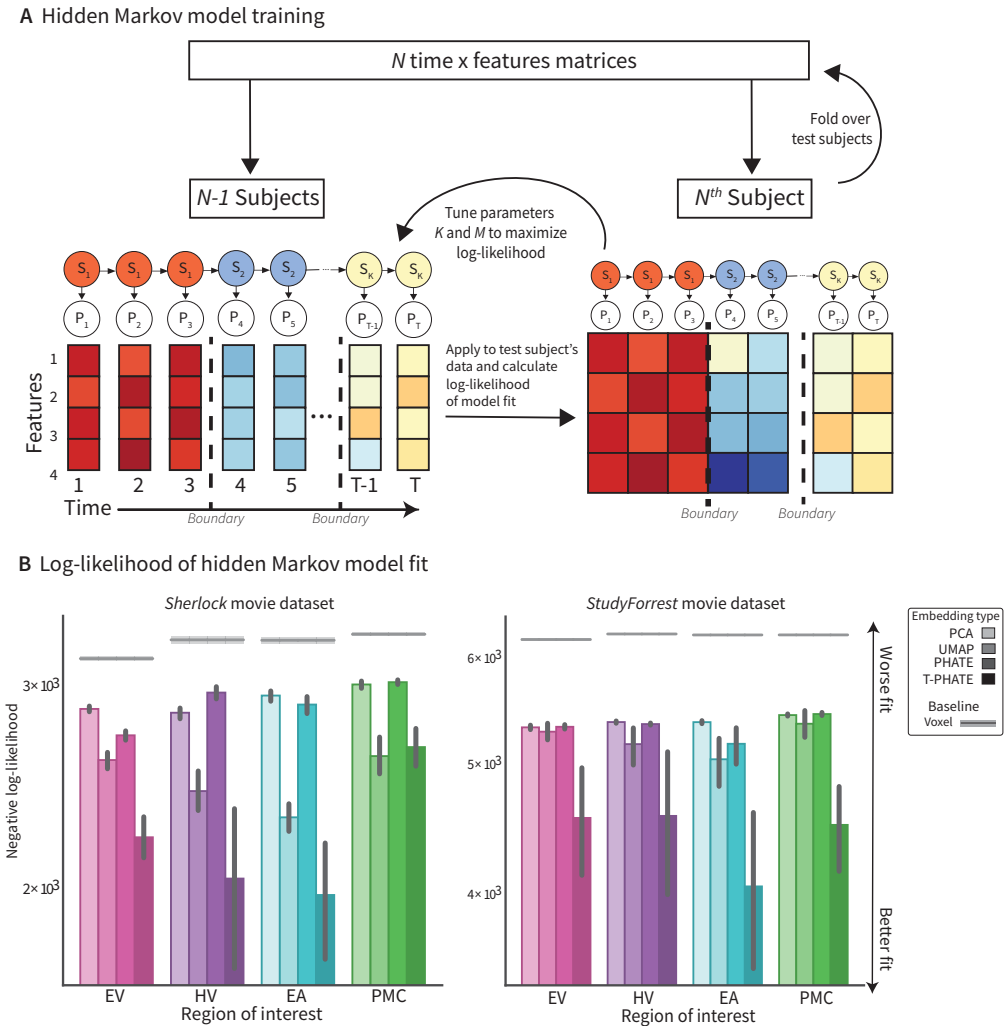

Figure 4: **A.** The HMM is trained on the average data for $N-1$ subjects to find K stable patterns of voxel activation, where the voxel patterns P for each timepoint belong to a given hidden state S. Model fit is evaluated on a test subject's data then repeated with leave-one-subject-out cross-validation. Event boundaries (dashed lines) are identified by the model as timepoints at a transition between hidden states. **B.** Log-likelihood of model fit is tested on a given subject using a HMM with hyper-parameters K and M tuned via leave-one-subject-out cross validation. Model fit here is presented as the negative log-likelihood, where a better model fit results in a lower negative log-likelihood. Dimensionality-reduced approaches are shown as bars, and full voxel-resolution benchmark shown as line with shading representing mean and 95% confidence interval.

by limiting the temporal distance between the timepoints and assuring that each pair of timepoints was matched as one within-event and one between-event comparison (Fig. 4). T-PHATE embeddings result in the significantly better within-versus-between event boundary distinctions across all regions (aside from *StudyForrest* early visual, where T-PHATE is marginally greater than PCA) (Fig. 3B). In the *Sherlock* dataset (and 2/4 ROIs of the *StudyForrest* data), T-PHATE at least doubles the performance of the second-best model. Notably, all dimensionality reduction outperform the original voxel resolution

data. This indicates that, though there is an inherent advantage in revealing dominant signal structure in data in lower dimensions, the temporal diffusion step of T-PHATE captures dynamic signals beyond a temporally-agnostic manifold.

In the previous analysis, event boundaries are learned and applied from a HMM fitted within-subject. To ensure the high performance in this analysis was not driven by overfitting by training/testing the boundaries within-subject, we performed cross-validation by applying the event boundaries identified for $N-1$ subjects to subject N's data individually. In other words, we ran the same analysis $N-1$ times for each subject with each other subject's boundaries. Again, we found the strongest effect for T-PHATE relative to the other methods (Fig. 5A), albeit weaker than when the event boundaries were learned within-subject, as expected. Interestingly, the voxel resolution data outperformed in a few cases (voxel $>$ others in *Sherlock* HV UMAP and PHATE, EA UMAP, PMC PHATE; *Study-Forrest* EV PHATE, HV PCA and PHATE, PMC PCA, PHATE and UMAP). This was not seen in the within-subject version, which suggesting that overfitting to subject-specific latent spaces occurred with some methods. These results gives us confidence that the previous, stronger within-subject effect was not driven by overfitting.

Identifying neural event boundaries with HMM is a data-driven approach to understanding how the brain makes sense of continuous stimulation. However, the identified boundaries have no ground-truth. Another way to assess whether learned neural boundaries are meaningful is to use a separate behavioral task where individuals watch the same stimulus and indicate where they believe there was a shift in events in the stimulus. Ample literature supports the use of this type of parsing task for studying human event segmentation [32–35], allowing us to treat human-rated boundaries as a ground-truth. A separate cohort of human subjects watched the *Sherlock* movie and were asked to indicate where they thought an event boundary occurred, which is a common behavioral task for measuring event segmentation. We applied these boundaries to the data in the same fashion as we did the neural events and measured the distance across event boundaries. The ground-truth event boundaries resulted in significantly better within-versus-between event differences when applied to T-PHATE than to any of the other methods (Figure 6B). The PMC is a higher-order integrative region, and we expected it to show neural event structure most similar to how people would consciously rate judge event boundaries. PMC showed a significant effect in the voxel resolution data (as with other methods), but the effect was an order of magnitude greater for T-PHATE relative to the voxel resolution data.

4 Discussion

In this study, we investigated the latent space of regional brain activity during naturalistic viewing from single-subject fMRI data. We found that the latent space of this activity best captured dynamic information about the experience of the stimulus using a low-dimensional subspace identified with T-PHATE, our novel method that incorporates the temporal dynamics of a signal to learn a more robust manifold. As previously shown with calcium imaging data from mice [7, 36], we found that a nonlinear dimensionality

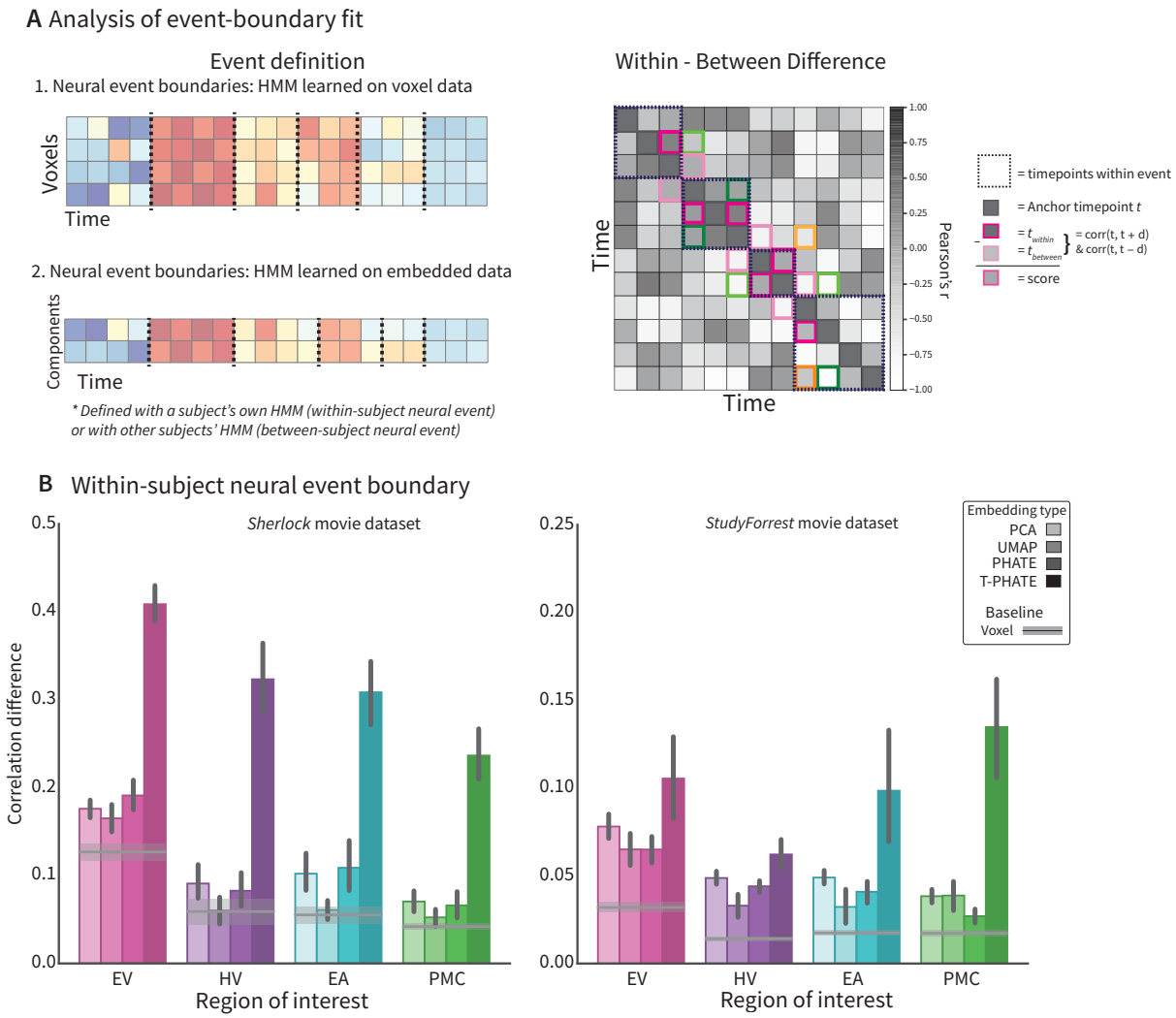

Figure 5: **A.** Event boundaries are learned both from the data (by applying the HMM on voxel resolution or embedded timeseries data and applying those boundaries either within or across subjects) and from independent sources including human behavioral ratings. (Right) Representation of dynamic information in a latent space is quantified by applying event boundaries to embedded data. Data emphasizes brain state transitions if the boundaries maximize the difference between brain activity at timepoints across events versus within events. For each TR t_0 in the timeseries, we correlate the activation pattern across features at t_0 with that of t_1 and t_2 such that t_1 is in the same event as t_0 (within a dashed box) and t_2 is in a different event. The within-between score is the subtraction of $corr(t_2, t_0)$ from $corr(t_0, t_1)$. To account for intrinsically higher correlation of two temporally proximal points, we restrict comparisons to where $|t_0 - t_1| = |t_0 - t_2|$. Here, $corr(t_0, t_1)$ is denoted by a darker colored box and its corresponding $corr(t_2, t_0)$ a lighter shade. **B.** Neural event boundaries are then identified using the fitted HMM and applied to the subjects' own data, where fit is evaluated. Dimensionality-reduced approaches are shown as bars, and full voxel-resolution benchmark shown as line with shading representing mean and 95% confidence interval.

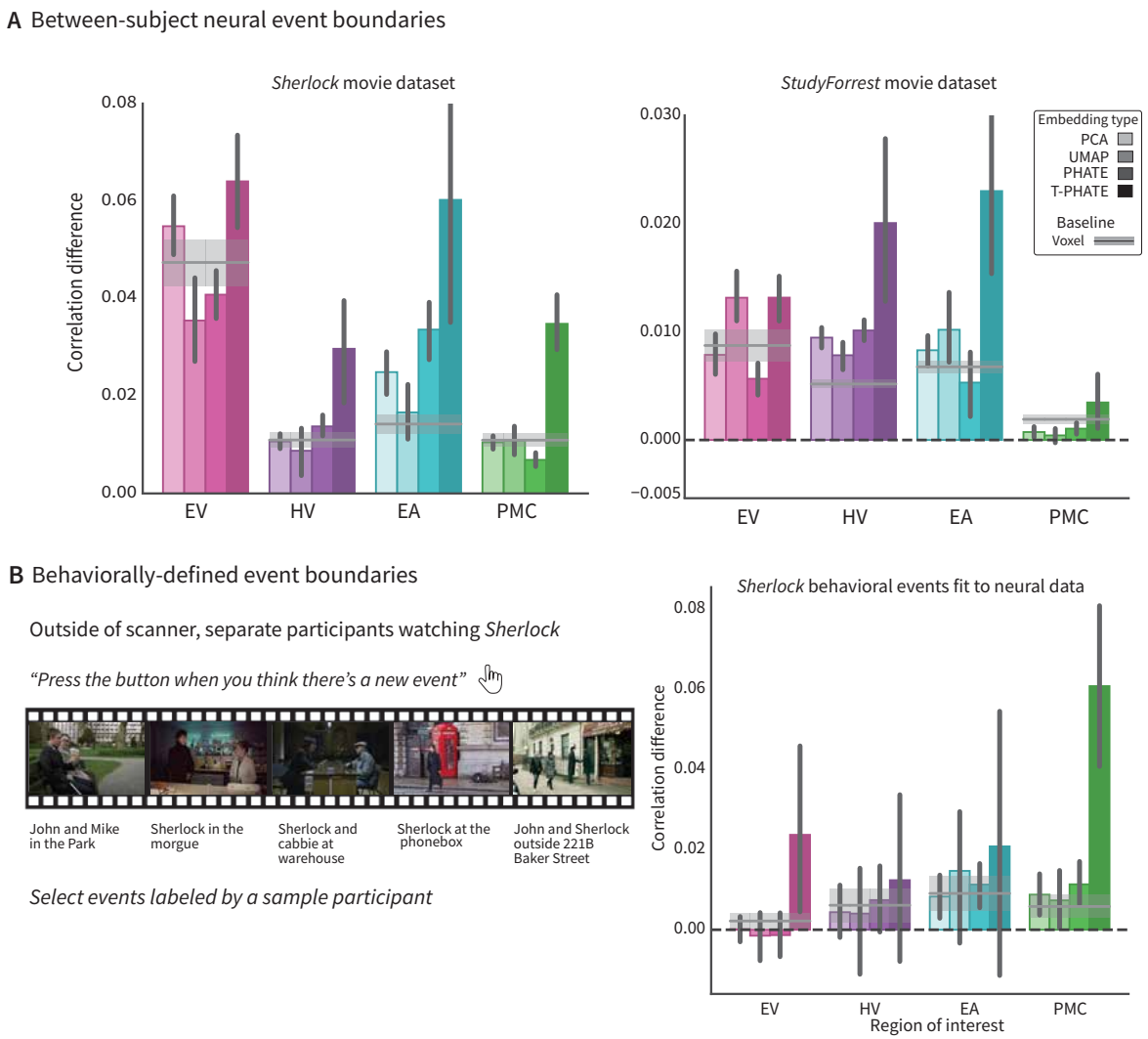

Figure 6: **A.** Event boundaries were rotated between subjects by testing subject *N* iteratively with the event boundaries learned from HMMs from each of the remaining subjects *N-1* subjects (within ROI and embedding method). T-PHATE embeddings reveal more robust event-related structure even when event boundaries are learned on different subjects. **B.** Event boundaries were also identified by behavioral ratings in a separate experiment on the *Sherlock* dataset. These boundaries best reflect structure in the posterior medial cortex and are best revealed with T-PHATE. Dimensionality-reduced approaches are shown as bars, and full voxel-resolution benchmark shown as line with shading representing mean and 95% confidence interval.

reduction method that focused on not just the geometry of multivoxel activity patterns but also their temporal dynamics captures trajectories through cognitive states that were previously obscured by considering the data in its full resolution. T-PHATE can capture meaningful cognitive trajectories in as few as three dimensions from single-subject fMRI data, which is inherently noisy in both space and time.

We quantified the degree to which manifold representations of brain activity reflect the dynamics of the stimulus and cognitive processing using the neural event segmentation

framework. Using a Hidden Markov Model, we learned from the data where there are transitions between stable neural events. This revealed clearest transitions in T-PHATE embeddings, meaning that T-PHATE most clearly highlights discrete neural events that correlate best with events in the stimuli. To our knowledge, previous investigations of neural event segmentation have not used dimensionality reduction to either denoise neural data or model the structure of neural dynamics. Consistent with our predictions, all dimensionality reduction methods we presented did a better job at modeling event dynamics than the voxel-dimensional data does (Fig. 3), likely because they denoised the data in low dimensions. T-PHATE vastly outperforms PHATE, UMAP, and PCA at capturing event structure with event boundaries learned from neural data within-subject (Fig. 3), from other subjects(Fig. 5A), and from behavioral ratings(Fig. 5B). T-PHATE manifolds reveal neural dynamics previously obscured by the high-dimensionality, noise, and redundancy of both fMRI data and the neural population activity underlying it.

Importantly, the advantage T-PHATE gains in dynamic modeling is not at the expense of static decoding. Mutivariate pattern classification of stimulus features on these low-dimensional embeddings remains as good as the high-dimensional voxel data with T-PHATE embeddings and better than other methods, showing that T-PHATE performs valuable denoising (Fig. 8). This shows that while both time-considerate and time-agnostic manifolds can encode static information, only the former unveil dynamic information.

Cognitive processes like learning, language comprehension [18], theory of mind [37, 38], narrative understanding [18–20], and memory [17] do not occur at a single point in time – they require the accumulation of information over longer temporal windows [21]. These processes occur rapidly to define our conscious experience of the world, and it is suggested that event segmentation supports our perception and memory of the world. [32–34, 39–41]. To understand how such complex, integrative processing can happen so quickly, we must consider the geometry of its representation in the brain. A wide body of fMRI literature shows that flexible population codes underlie complex cognitive processing [12, 15, 41–43] among many others. Further, the animal literature shows that population representations underlying cognitive processing can be modeled in low-dimensional manifolds [3,7,8], and the temporal unfolding of behavior are best predicted with nonlinear, temporally-dependent manifolds [7,36]. In the fMRI literature, linear manifolds have been used to track brain states across discrete neurocognition tasks [44, 45] and nonlinear manifolds have been used to model the dynamics of cognitive load across these tasks [24]. These designs vary from ours along two critical dimensions: the aggregation of whole-brain data across 100+ subjects to define the manifold, and the block-design task stimulus used [24, 44, 45]. To our knowledge, no one has yet investigated manifolds learned from single-subject fMRI data collected with naturalistic stimuli to investigate temporal dynamics of individual brain regions. This is likely due to the noise of single-subject, naturalistic fMRI data. With the proper choice of dimensionality reduction algorithm [8, 25], we accounted for these challenges to unveil latent dynamics of the cognitive processes related to movie watching.

In conclusion, we found that nonlinear, temporal manifold learning can identify a low-dimensional subspace of brain activity from continuous, naturalistic stimuli that captures

signatures of higher-order cognitive integration in the brain. Critically, latent dimensions learned with this combination of nonlinear dimensionality reduction and temporal diffusion outperforms methods lacking either of those components. This suggests that the relation between the dimensions of individual voxels and the dominant signals that emerge from them are tied by nonlinear relations. By modeling a subject's BOLD timeseries in these dimensions, we can gauge the trajectory of their brain states throughout the course of the experiment, and we can relate this trajectory to theoretical frameworks like event segmentation theory that facilitate higher-order cognitive processes in the brain. Since this is performed at the level of individual subjects, future investigations could use this approach to probe individual difference questions within the native latent space. In sum, the characteristics of these manifolds could, in turn, inform theories about the nature and dynamics of human cognitive representation.

5 Methods

5.1 Manifold embeddings

Different dimensionality reduction algorithms focus on different aspects of a dataset, thus amplify distinct signals depending upon the structure targeted by the algorithm. We hypothesized that a nonlinear manifold learning algorithm designed to handle noisy, high-dimensional biological data, would be best suited for the signal and intrinsic noise of fMRI data. Potential of heat diffusion for affinity-based transition embedding (PHATE) [46] is a diffusion-based manifold learning method that models local and global structures simultaneously in nonlinear dimensions. Brain activity measured with fMRI is highly noisy in both space and time, with the BOLD signal canonically peaking 4-5 s after stimulus onset before slowly returning to baseline. With a temporally-dependent stimulus such as a movie, where conversations and plot lines play out over different timescales, the autocorrelation of nearby timepoints will likely extend beyond the curve of the BOLD signal and vary by ROI along the hierarchy of temporal integration in the brain [23]. We estimate an autocorrelation kernel for each ROI by correlating each voxel timeseries with lagged versions of itself until the correlation drops and stays below zero, then averaging the functions across voxels to get a regional autocorrelation function. This function is then expanded to calculate the transitional probability between all pairs of timepoints based solely on their estimated autocorrelation and then combined with the PHATE-based transitional probability matrix and embedded in lower dimensions (Fig. 1A).

5.1.1 PHATE

Given a dataset of voxel time-series data, $\mathcal{X} = x_1, x_2, \ldots, x_T$, where $x_t \in \mathbb{R}^n$ at time t is a n-dimensional vector and n is the number of voxels. Construction of the PHATE diffusion geometry starts by computing the Euclidean distance matrix D between data pairs, where $D(i,j) = ||x_i - x_j||^2$. D is then converted into a local affinity matrix K using an adaptive bandwidth Gaussian kernel. The affinity matrix K is then row-normalized to get the initial probabilities P_D, which are used for the Markovian random-walk diffusion process.

The diffusion operator P_D is then raised to the t_Dth power, where $t_D$is the PHATE optimal diffusion time scale, to perform a t_D-step random walk over the data graph. This specific diffusion process infers more global relations beyond the local affinities. To further preserve both local and manifold-intrinsic global structure of data, PHATE computes the diffusion potential U by taking the logarithm of $P_D^{t_D}$ and the diffusion potential distance D_U using distance between row pairs of U. This diffusion potential distance can be reduced to 2-3 dimensions for visualization or to any other dimensionality (such as that obtained by cross-validation) by performing metric MDS (MMDS).

5.1.2 T-PHATE

We designed T-PHATE as a variant of PHATE that uses a dual-view diffusion operator to embed timeseries data in a low-dimensional space. The first view, P_D, of the diffusion operator uses the same process as PHATE to build an affinity matrix based on the Euclidean distance between data samples (here, timepoints) and then row-normalize it to obtain the transition probability matrix, where P_D is a TxT matrix. The second view P_T is based on an affinity matrix that summarizes the autocorrelation function of the data [47]. First we calculate a direct estimator of the autocovariance of the timeseries vector V from each voxel using $|V| - 1$ lags. Then the functions are averaged across voxels to obtain a single autocorrelation function (acf) and smoothed with a rolling average over w timepoints, to account for possible jittering around where $acf = 0$. Next, we find the first lag lag_{max} where $acf = 0$, which defines the maximum width of smoothing for the temporal affinity matrix A. The temporal affinity matrix A is calculated as $A(i,j) = acf(i,j), 0 < |i-j| <= lag_{max}$ and zero elsewhere. The autocorrelative probability P_t is obtained by row-normalizing the affinity matrix A. These views are combined with alternating diffusion into a single operator $P = P_D P_T$, which is used as the initial probabilities for diffusion. The rest of constructing diffusion geometry and data visualization is performed the same as in PHATE. This dual-view diffusion step allows T-PHATE to learn data geometry and latent signals that represent cognitive processes that play out over longer temporal windows.

We compare T-PHATE's performance at learning dynamic neural manifolds with common dimensionality reduction algorithms that are agnostic to time including principal components analysis (PCA), uniform manifold approximation and projection (UMAP) [30], and vanilla PHATE [46]. To test whether our autocorrelation kernel was the best approach to modeling temporal dynamics in this manifold, we tested two additional versions of PHATE: incorporating time as a feature in the input data (PHATE+Time) and smoothing the data temporally (Smooth PHATE). fMRI data were extracted from regions of interest (ROIs) and normalized before embedding.

5.2 Event segmentation modeling

Human experience of real life is continuous, yet perception and conception of this experience is typically divided into discrete moments or events. Event segmentation theory explains that humans automatically divide continuous streams of information into dis-

crete events [32, 33] in order to form, organize, and recollect memories, make decisions, and predict the future [48]. Participants show high consistency in explicitly segmenting continuous stimuli [34] and also in how their brains represent these event boundaries [40, 41]. Event boundaries can be represented, both behaviorally and in cortical activity, at different timescales depending on the information being used to draw event boundaries [39]. In the brain, event boundaries are reflected by shifts in the stability of patterns of cortical activity, and we can learn the structure of cortical events with a variant of a hidden Markov model [41]. During a continuous stimulus such as a movie, different brain regions represent events along different time scales, which reflect the dynamics of the information being represented by a given region. For example, early sensory regions represent low-level sensory information about a stimulus, and would reflect this in more frequent event shifts, because low-level features shift quickly in most movie stimuli as the cameras change angles or characters speak. Later sensory regions represent longer timescale events, as features like physical environments change less frequently than features like camera angles. Regions associated with memory or narrative processing reflect event boundaries on longer timescales, and these boundaries best correspond with scene changes marked by human raters [49].

We used the Hidden Markov Model (HMM) variant presented in [41] and implemented in BrainIAK [50] to learn from different representations of BOLD timeseries data where a brain region experiences event boundaries, or transitions between stable states of activation. Given an activation timeseries of a brain region during a continuous stimulus, the HMM identifies stable activity patterns or "events" that are divided by boundaries, where the timeseries transitions between two stable patterns of activation. This is done iteratively and the log-likelihood of the model fit is evaluated at each iteration, and model fitting stops when the log-likelihood begins to decrease. The first step of this process is to use the HMM to learn from the data the number of events a brain region represents for a given stimulus. Past studies have used HMMs for event segmentation on multivoxel activity patterns and have validated this approach against behavioral segmentation [40, 41, 49]. This shows that voxel resolution data reflects meaningful meaningful event segmentation, so we chose to estimate the optimal number of events (K) for each brain region using the voxel resolution data(Fig. 4).

To run an HMM for event segmentation on our manifold embedding data, we needed to tune two parameters for each subject: K, or the number of hidden states through which a brain region passes during the stimulus, and M, the dimensionality of the latent space that captures this signal.

5.2.1 Optimizing number of neural events.

We estimated K and M for each brain region and embedding method with a leave-one-subject-out cross-validation scheme. Using BOLD data from $N-1$ training subjects, we created an average timeseries matrix by averaging response profiles for each voxel across the training subjects. We define K as the number of hidden states in an HMM and optimize K by fitting a model with K states on the training data and applying it on the BOLD data of the held-out test subject and calculating the log probability of the test data accord-

ing to the learned event segmentation of the model (hence referred to as the model log-likelihood). We tested a range of K from 2 to 200 for *Sherlock* and 2 to 400 for *StudyForrest*. For each training fold, we selected the K value that maximized the model log-likelihood.

5.2.2 Optimizing manifold dimensionality.

We used the same leave-one-subject-out cross-validation scheme to select the dimensionality of manifold embeddings for each subject. We embedded each subject's timeseries data into an array of possible M values then fit an HMM on each M-dimensional embedding with the subject's cross-validated K value, which had been learned in the previous step from all the other subjects' voxel resolution data. Event boundary fit evaluation is described in more detail below. The M parameter that maximizes the event boundary fit was recorded for each subject. We then set the M value to be used for a subject N as the average of all $N-1$ subjects' best M values.

5.2.3 Evaluating the fit of event boundaries.

To quantify how well a learned manifold embedding amplifies dynamic structure in the data, we calculated difference in correlation across latent dimensions for pairs of timepoints within and across event boundaries. For this calculation, we restricted the timepoints that went into the calculation as follows. We calculated the length of the longest event, and only considered timepoints that were less than that temporal distance apart. We anchored a comparison on each timepoint t in the timeseries, and for each temporal distance n ranging from 1 timepoint to the length of the longest event, we considered a pair of timepoints $t-n$ and $t+n$ if one timepoint fell in the same event as the anchor and the other a different event. We then took the correlation between the anchor and each timepoint, binning them according to whether they were a within-event comparison or a between-event comparison. This process assures that there are equal numbers of timepoints going into each comparison type (Fig. 4B).

One issue of defining and testing event boundaries within-subject is overfitting. To circumvent this, we rotated event boundaries between subjects, testing $N-1$ sets of event boundaries on each subject separately using the same procedure outlined above. The *Sherlock* dataset has a separate set of human subjects rate event boundaries behaviorally as they watch the film, and we applied those event boundaries as well to the data.

5.3 Data

5.3.1 Sherlock dataset

For full details on the *Sherlock* data, see the original publication of this dataset [43]. Here, we used data from the 16 participants who viewed a 48-minute clip of the BBC television series "Sherlock." Data were collected in two fMRI runs of 946 and 1030 timepoints (repetition time; TRs) and was downloaded from the DataSpace public repository (http://arks.princeton.edu/ark:/88435/dsp01nz8062179). The data were collected on a Siemens Skyra 3-T scanner with a 20-channel head coil. Functional images were acquired

with a T2*-weighted echo-planar imaging sequence (TE 28ms, TR 1.5s, 64-degree flip angle, whole brain coverage with 27 slices of 4mm thickness and 3x3 mm^2 in-plane resolution, 192x192mm^2 FOV). Anatomical images were acquired with a T1-weighted MPRAGE pulse sequence with 0.89mm^3 resolution. Slice-time correction, motion correction, linear detrending, high-pass filtering (140s cutoff), and co-registration and affine transformation of functional volumes to the Montreal Neurological Institute (MNI) template were all performed with FSL. Functional images were then resampled from native resolution to 3mm isotropic voxels for all analyses, z-scored across time at every voxel, and smoothed with a 6mm kernel.

5.3.2 StudyForrest dataset

For full details on the *StudyForrest* data, see the original publication of the movie dataset [51] and the localizer extension dataset [52]. Here, we included data from 15 participants who completed both tasks. All participants were native German speakers. In the movie-viewing task, participants watched a 2-hour version of the movie "Forrest Gump." These data were collected in eight fMRI runs resulting in a full timeseries of 3599 TRs (451, 441, 438, 488, 462, 439, 542, and 338 per run). Movie data were collected on a 7-T Siemens scanner with a 32-channel headcoil. Functional images were acquired with a T2*-weighted echo-planar imaging sequence (TE 22ms, TR 2s, whole brain coverage with 36 slices of 1.4 mm thickness and 1.4 x 1.4 mm^2 in-plane resolution, 224mm FOV). In the localizer task, the same fifteen participants viewed 24 unique grayscale images from 6 categories (human faces, human bodies without heads, houses, small objects, outdoor scenes, and scrambled images) in four (156 TR) block-design runs with two 16s blocks per stimulus category per run. Localizer data were collected on a 3-T Philips Achieva scanner with a 32-channel headcoil. Functional images for the localizer task were acquired with a T2*-weighted echo-planar imaging sequence (TE 30ms, TR 2s, 90-degree flip angle, 35 slices of 3mm thickness and 3x3 mm^2 in-plane resolution, 240mm FOV). Structural images were acquired with a 3-T Philips Achieva using a 32-channel headcoil. T1-weighted anatomical images were acquired with a 3D turbo field echo sequence with 0.67mm isotropic resolution. Slice-time correction, co-registration and affine transformation to MNI template were performed with FSL where functional images were resampled to 3mm isotropic voxels. Additional preprocessing included linear detrending, high-pass filtering (100s cutoff), spatial smoothing with a 6mm kernel, and nuisance regression (including 6 motion parameters, global signal, white matter, and cerebrospinal fluid), and z-scoring within voxel to match the *Sherlock* preprocessing.

5.4 Region of interest (ROI) selection

We selected four regions of interest (ROIs) based on the original publication of the *Sherlock* dataset [43], subsequent publications [41], and regions known to have reliably strong signals in response to audiovisual movie stimuli [53]. The early visual (EV), early auditory (EA), and high visual (HV) region masks were based on a functional atlas defined with resting-state connectivity [54]. As in the original *Sherlock* publication, we defined a poste-

rior medial cortex (PMC) ROI as the posterior medial cluster of voxels within the dorsal default mode network [43]. In the *Sherlock* data, the dimensionality of voxel-resolution ROIs are as follows: EV = 307, HV = 571, EA = 1018, PMC = 481. In the *StudyForrest* data, the dimensionality of voxel-resolution ROIs are as follows: EV = 166, HV = 456, EA = 657, PMC = 309.

6 Supplemental methods

6.1 Movie annotations

***Sherlock* annotations** We used annotations for the *Sherlock* movie stimulus released with [55] and available freely on Github. These annotations were created by human annotators who viewed the film and labeled each timepoint on a variety of metrics. Specifically here we used two sets of binary labels that correspond with visual and auditory aspects of the stimulus: whether a given TR featured an indoor or outdoor scene and whether or not the TR featured music.

***StudyForrest* annotations** Annotations of scene cuts and locations used in later analysis of the Forrest Gump movie viewing data were downloaded from a subsequent *StudyForrest* publication [56]. These provide additional information about a number of shots from the film, including the setting of the shot, whether a setting is interior or exterior, the temporal progression of a shot with reference to the previous shot, and the time of day. These are used in the movie feature decoding analysis for the *StudyForrest* dataset.

7 Supplemental results

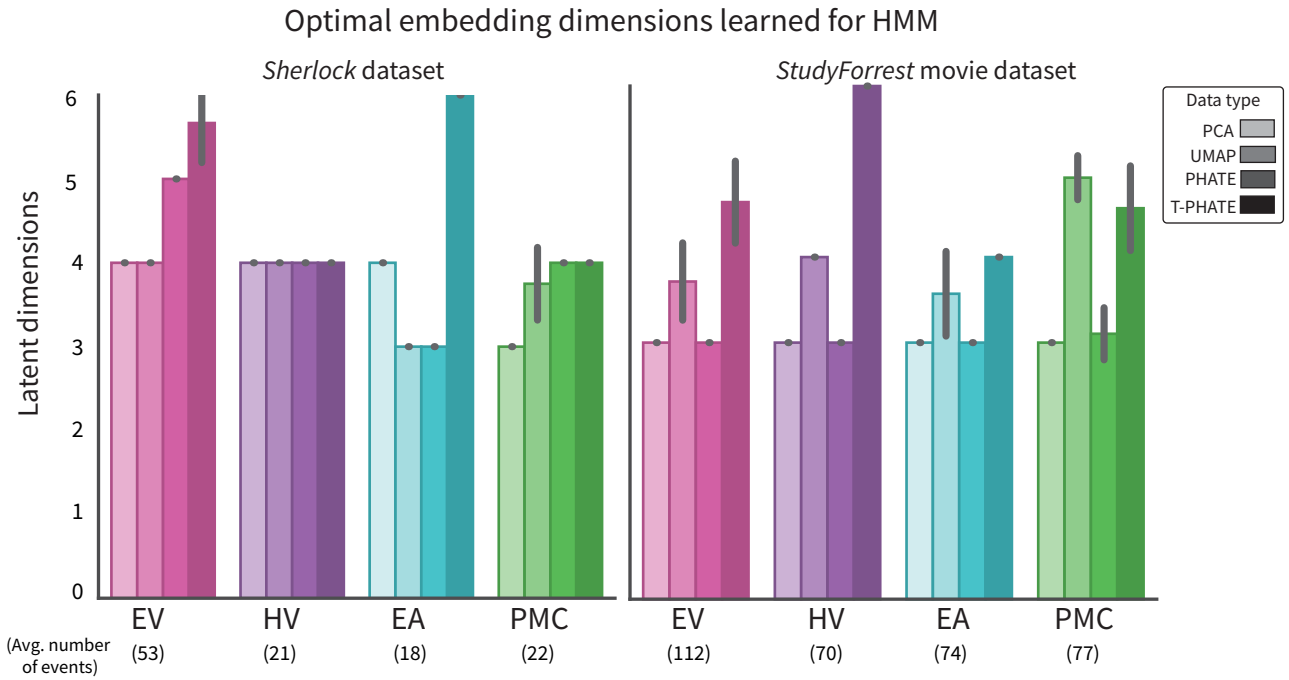

Figure 7: HMM parameters K and M learned with leave-one-subject-out cross validation. Dimensionality reduction methods are able to optimize event boundaries in relatively few dimensions. K values follow the predicted event hierarchy in the brain, with lower-level sensory regions experiencing more events than higher-order cognitive regions.

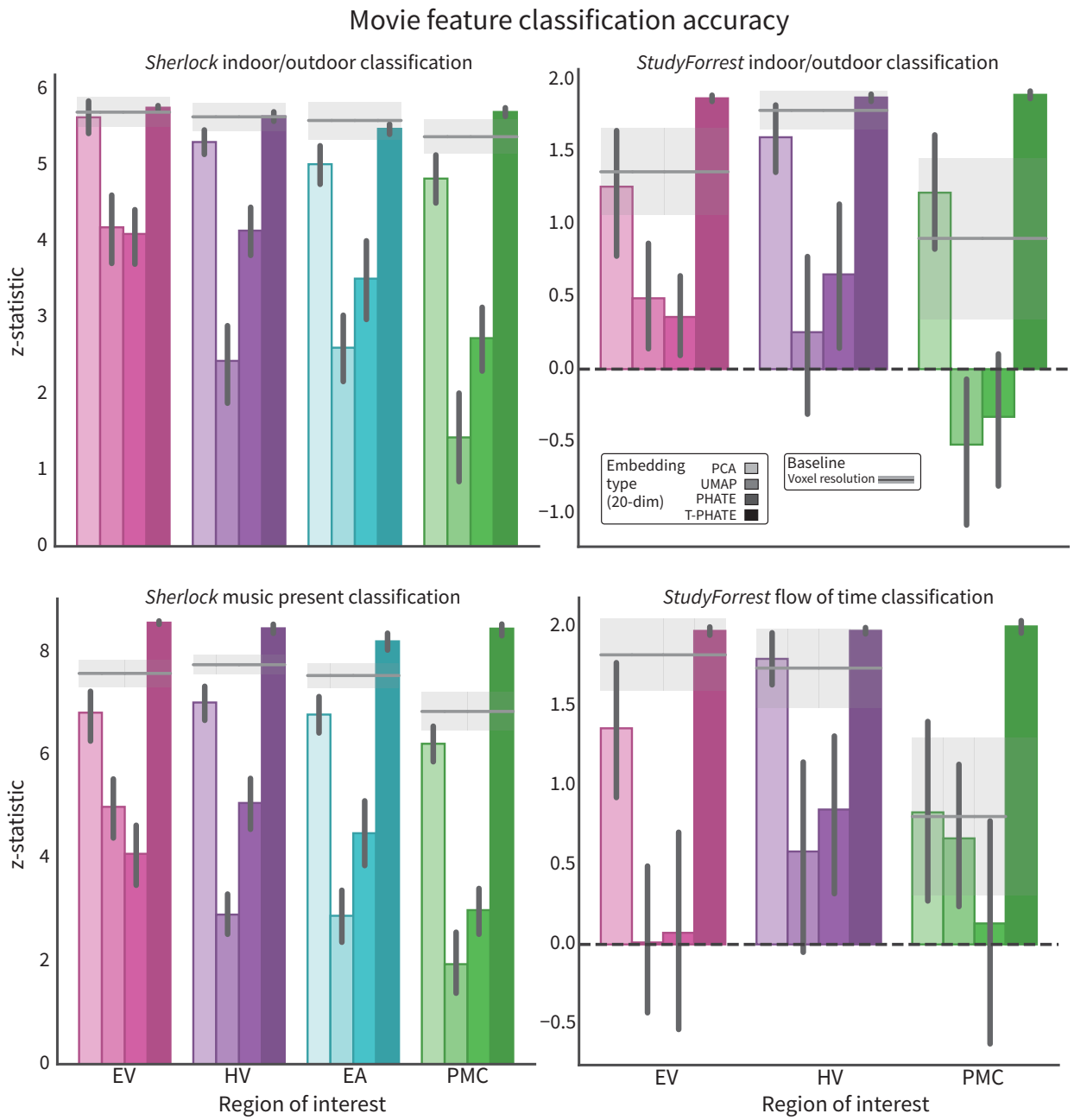

Figure 8: Support vector classification of movie features in the *Sherlock* and *StudyForrest* movie datasets. All labels were binary aside from the Flow of Time labels, which had four levels. Results presented are the z-statistic of true classification accuracy versus 1000 repetitions of time-shifting labels and retraining the model.

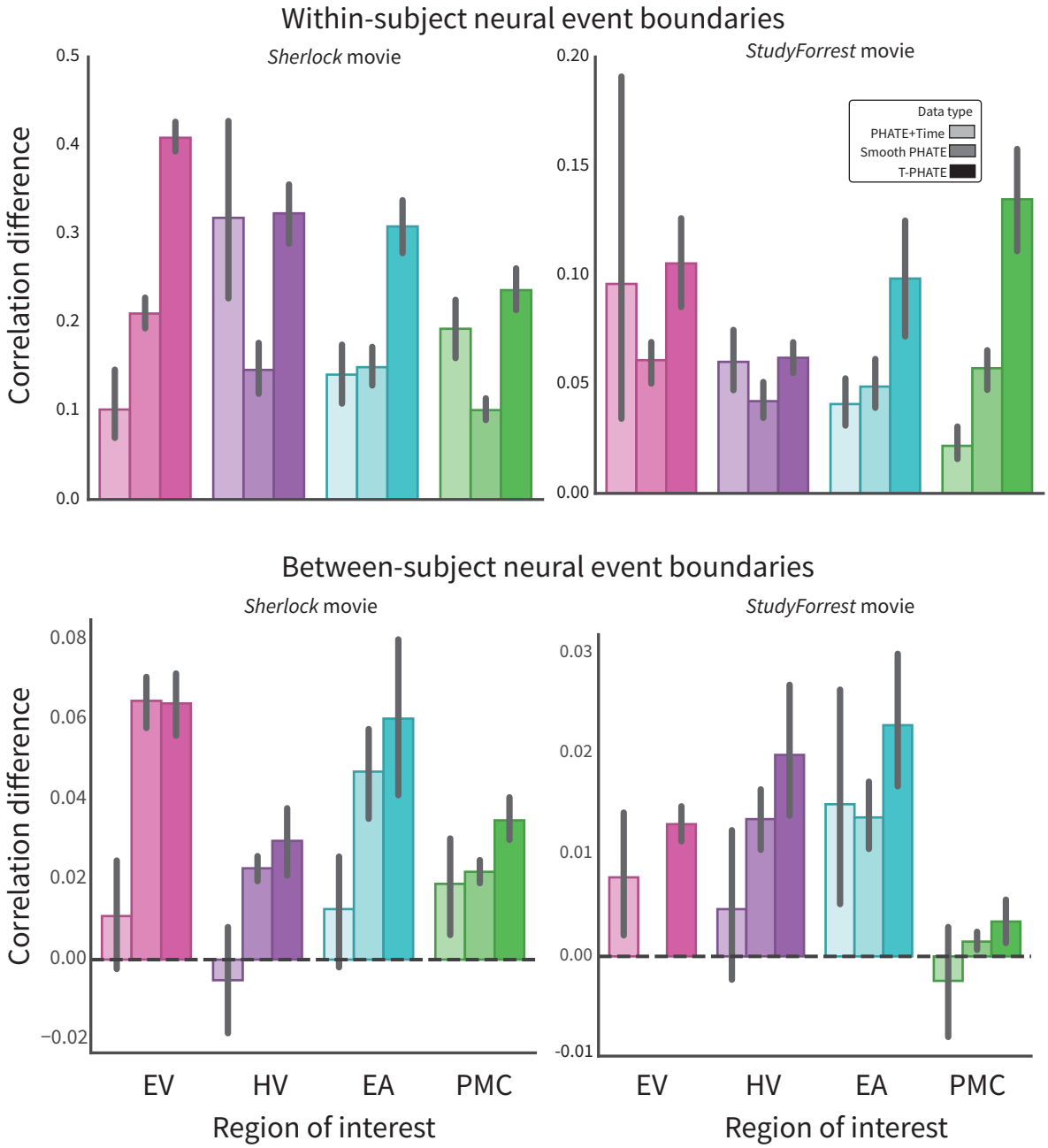

Figure 9: Within-vs-between neural event boundaries for the time alternative methods. **A.** Event boundaries defined as neural events within-subject. **B.** Event boundaries cross-validated with leave-one-subject-out boundaries.

References

[1] B. B. Averbeck, P. E. Latham, and A. Pouget, "Neural correlations, population coding and computation," *Nature reviews neuroscience*, vol. 7, no. 5, pp. 358–366, 2006.

[2] G. Laurent, "Olfactory network dynamics and the coding of multidimensional signals," *Nature Reviews Neuroscience*, vol. 3, no. 11, pp. 884–895, 2002. [Online]. Available: https://doi.org/10.1038/nrn964

[3] M. M. Churchland, J. P. Cunningham, M. T. Kaufman, S. I. Ryu, and K. V. Shenoy, "Cortical preparatory activity: representation of movement or first cog in a dynamical machine?" *Neuron*, vol. 68, no. 3, pp. 387–400, 2010.

[4] L. Chang and D. Y. Tsao, "The code for facial identity in the primate brain," *Cell*, vol. 169, no. 6, pp. 1013–1028.e14, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S009286741730538X

[5] W. A. Freiwald and D. Y. Tsao, "Functional compartmentalization and viewpoint generalization within the macaque face-processing system," *Science*, vol. 330, no. 6005, p. 845, 2010. [Online]. Available: http://science.sciencemag.org/content/330/6005/845.abstract

[6] J. A. Hennig, M. D. Golub, P. J. Lund, P. T. Sadtler, E. R. Oby, K. M. Quick, S. I. Ryu, E. C. Tyler-Kabara, A. P. Batista, B. M. Yu, and S. M. Chase, "Constraints on neural redundancy," *eLife*, vol. 7, p. e36774, 2018. [Online]. Available: https://doi.org/10.7554/eLife.36774

[7] E. H. Nieh, M. Schottdorf, N. W. Freeman, R. J. Low, S. Lewallen, S. A. Koay, L. Pinto, J. L. Gauthier, C. D. Brody, and D. W. Tank, "Geometry of abstract learned knowledge in the hippocampus," *Nature*, vol. 595, no. 7865, pp. 80–84, 2021. [Online]. Available: https://doi.org/10.1038/s41586-021-03652-7

[8] J. P. Cunningham and B. M. Yu, "Dimensionality reduction for large-scale neural recordings," *Nature Neuroscience*, vol. 17, no. 11, pp. 1500–1509, 2014. [Online]. Available: https://www.nature.com/articles/nn.3776files/202/nn.htmlhttps://www.nature.com/articles/nn.3776.pdf

[9] K. A. Norman, S. M. Polyn, G. J. Detre, and J. V. Haxby, "Beyond mind-reading: multi-voxel pattern analysis of fmri data," *Trends in Cognitive Sciences*, vol. 10, no. 9, pp. 424–430, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364661306001847files/124/S1364661306001847.html

[10] D. D. Cox and R. L. Savoy, "Functional magnetic resonance imaging (fmri) "brain reading": detecting and classifying distributed patterns of fmri activity in human visual cortex," *NeuroImage*, vol. 19, no. 2, pp. 261–270, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811903000491

[11] Y. Kamitani and F. Tong, "Decoding the visual and subjective contents of the human brain," *Nature Neuroscience*, vol. 8, no. 5, pp. 679–685, 2005. [Online]. Available: https://doi.org/10.1038/nn1444

[12] J. V. Haxby, "Distributed and overlapping representations of faces and objects in ventral temporal cortex," *Science*, vol. 293, no. 5539, pp. 2425–2430, 2001. [Online]. Available: https://www.sciencemag.org/lookup/doi/10.1126/science.1063736

[13] J.-D. Haynes and G. Rees, "Decoding mental states from brain activity in humans," *Nature Reviews Neuroscience*, vol. 7, no. 7, pp. 523–534, 2006. [Online]. Available: https://doi.org/10.1038/nrn1931

[14] S. M. Polyn, V. S. Natu, J. D. Cohen, and K. A. Norman, "Category-specific cortical activity precedes retrieval during memory search," *Science*, vol. 310, no. 5756, pp. 1963–1966, 2005. [Online]. Available: https://science.sciencemag.org/content/sci/310/5756/1963.full.pdf

[15] C. Davatzikos, K. Ruparel, Y. Fan, D. G. Shen, M. Acharyya, J. W. Loughead, R. C. Gur, and D. D. Langleben, "Classifying spatial patterns of brain activity with machine learning methods: Application to lie detection," *NeuroImage*, vol. 28, no. 3, pp. 663–668, 2005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811905005914

[16] N. B. Turk-Browne, "Functional interactions as big data in the human brain," *Science*, vol. 342, no. 6158, pp. 580–584, 2013. [Online]. Available: https://science.sciencemag.org/content/342/6158/580files/143/580.htmlhttp://www.ncbi.nlm.nih.gov/pubmed/24179218

[17] U. Hasson, J. Chen, and C. J. Honey, "Hierarchical process memory: memory as an integral component of information processing," *Trends in cognitive sciences*, vol. 19, no. 6, pp. 304–313, 2015. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/25980649https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4457571/

[18] J. Xu, S. Kemeny, G. Park, C. Frattali, and A. Braun, "Language in context: emergent features of word, sentence, and narrative comprehension," *Neuroimage*, vol. 25, no. 3, pp. 1002–1015, 2005.

[19] C. Honey, T. Thesen, T. Donner, L. Silbert, C. Carlson, O. Devinsky, W. Doyle, N. Rubin, D. Heeger, and U. Hasson, "Slow cortical dynamics and the accumulation of information over long timescales," *Neuron*, vol. 76, no. 2, pp. 423–434, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0896627312007179

[20] Y. Yeshurun, M. Nguyen, and U. Hasson, "Amplification of local changes along the timescale processing hierarchy," *Proceedings of the National Academy of Sciences*, vol. 114, no. 35, pp. 9475–9480, 2017.

[21] U. Hasson, E. Yang, I. Vallines, D. J. Heeger, and N. Rubin, "A hierarchy of temporal receptive windows in human cortex," *Journal of Neuroscience*, vol. 28, no. 10, pp. 2539–2550, 2008. [Online]. Available: https://www.jneurosci.org/content/28/10/2539

[22] D. L. Barack and J. W. Krakauer, "Two views on the cognitive brain," *Nature Reviews Neuroscience*, vol. 22, no. 6, pp. 359–371, 2021. [Online]. Available: https://doi.org/10.1038/s41583-021-00448-6

[23] K. D. Himberger, H.-Y. Chien, and C. J. Honey, "Principles of temporal processing across the cortical hierarchy," *Neuroscience*, vol. 389, pp. 161–174, 2018, sensory Sequence Processing in the Brain. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306452218302951

[24] S. Gao, G. Mishne, and D. Scheinost, "Non-linear manifold learning in fmri uncovers a low-dimensional space of brain dynamics," *bioRxiv*, p. 2020.11.25.398693, 2020. [Online]. Available: https://www.biorxiv.org/content/biorxiv/early/2020/11/27/2020.11.25.398693.full.pdf

[25] M. Jazayeri and S. Ostojic, "Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity," *arXiv preprint arXiv:2107.04084*, 2021.

[26] K. R. Moon, D. van Dijk, Z. Wang, S. Gigante, D. B. Burkhardt, W. S. Chen, K. Yim, A. v. d. Elzen, M. J. Hirn, R. R. Coifman, N. B. Ivanova, G. Wolf, and S. Krishnaswamy, "Visualizing structure and transitions in high-dimensional biological data," *Nature Biotechnology*, vol. 37, no. 12, pp. 1482–1492, 2019. [Online]. Available: https://www.nature.com/articles/s41587-019-0336-3files/149/s41587-019-0336-3.html

[27] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[28] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000. [Online]. Available: https://science.sciencemag.org/content/sci/290/5500/2323.full.pdf

[29] M. Salhov, A. Bermanis, G. Wolf, and A. Averbuch, "Approximately-isometric diffusion maps," *Applied and Computational Harmonic Analysis*, vol. 38, no. 3, pp. 399–419, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1063520314000803

[30] E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I. W. Kwok, L. G. Ng, F. Ginhoux, and E. W. Newell, "Dimensionality reduction for visualizing single-cell data using umap," *Nature biotechnology*, vol. 37, no. 1, pp. 38–44, 2019.

[31] J. Huang, E. L. Busch, T. Wallenstein, M. Gerasimiuk, A. Benz, G. Lajoie, G. Wolf, N. B. Turk-Browne, and S. Krishnaswamy, "Learning shared neural manifolds from multi-subject fmri data," *arXiv preprint arXiv:2201.00622*, 2021.

[32] G. A. Radvansky and J. M. Zacks, "Event boundaries in memory and cognition," *Current Opinion in Behavioral Sciences*, vol. 17, pp. 133–140, 2017, memory in time and space. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352154617300037

[33] J. M. Zacks, N. K. Speer, K. M. Swallow, T. S. Braver, and J. R. Reynolds, "Event perception: a mind-brain perspective." *Psychological bulletin*, vol. 133, no. 2, p. 273, 2007.

[34] J. M. Zacks, N. K. Speer, K. M. Swallow, and C. J. Maley, "The brain's cutting-room floor: Segmentation of narrative cinema," *Frontiers in human neuroscience*, vol. 4, p. 168, 2010.

[35] N. K. Speer, J. M. Zacks, and J. R. Reynolds, "Human brain activity time-locked to narrative event boundaries," *Psychological science*, vol. 18, no. 5, pp. 449–455, 2007.

[36] R. J. Low, S. Lewallen, D. Aronov, R. Nevers, and D. W. Tank, "Probing variability in a cognitive map using manifold inference from neural dynamics," *bioRxiv*, 2018. [Online]. Available: https://www.biorxiv.org/content/early/2018/09/20/418939

[37] H. L. Gallagher and C. D. Frith, "Functional imaging of 'theory of mind'," *Trends in cognitive sciences*, vol. 7, no. 2, pp. 77–83, 2003.

[38] R. Saxe, S. Carey, and N. Kanwisher, "Understanding other minds: linking developmental psychology and functional neuroimaging," *Annu. Rev. Psychol.*, vol. 55, pp. 87–124, 2004.

[39] J. M. Zacks, T. S. Braver, M. A. Sheridan, D. I. Donaldson, A. Z. Snyder, J. M. Ollinger, R. L. Buckner, and M. E. Raichle, "Human brain activity time-locked to perceptual event boundaries," *Nature Neuroscience*, vol. 4, no. 6, pp. 651–655, 2001. [Online]. Available: https://doi.org/10.1038/88486

[40] T. S. Yates, L. J. Skalaban, C. T. Ellis, A. J. Bracher, C. Baldassano, and N. B. Turk-Browne, "Neural event segmentation of continuous experience in human infants," *bioRxiv*, p. 2021.06.16.448755, 2021. [Online]. Available: https://www.biorxiv.org/content/biorxiv/early/2021/06/16/2021.06.16.448755.full.pdf

[41] C. Baldassano, J. Chen, A. Zadbood, J. W. Pillow, U. Hasson, and K. A. Norman, "Discovering event structure in continuous narrative perception and memory," *Neuron*, vol. 95, no. 3, pp. 709–721.e5, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0896627317305937

[42] J. V. Haxby, A. C. Connolly, and J. S. Guntupalli, "Decoding neural representational spaces using multivariate pattern analysis," *Annual Review of Neuroscience*, vol. 37, no. 1, pp. 435–456, 2014, pMID: 25002277. [Online]. Available: https://doi.org/10.1146/annurev-neuro-062012-170325

[43] J. Chen, Y. C. Leong, C. J. Honey, C. H. Yong, K. A. Norman, and U. Hasson, "Shared memories reveal shared structure in neural activity across individuals," *Nature neuroscience*, vol. 20, no. 1, pp. 115–125, 2017. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/27918531https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5191958/

[44] J. M. Shine, P. G. Bissett, P. T. Bell, O. Koyejo, J. H. Balsters, K. J. Gorgolewski, C. A. Moodie, and R. A. Poldrack, "The dynamics of functional brain networks: integrated network states during cognitive task performance," *Neuron*, vol. 92, no. 2, pp. 544–554, 2016. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5073034/pdf/nihms-820230.pdf

[45] J. M. Shine, M. Breakspear, P. T. Bell, K. A. Ehgoetz Martens, R. Shine, O. Koyejo, O. Sporns, and R. A. Poldrack, "Human cognition involves the dynamic integration of neural activity and neuromodulatory systems," *Nature Neuroscience*, vol. 22, no. 2, pp. 289–296, 2019. [Online]. Available: http://www.nature.com/articles/s41593-018-0312-0

[46] K. R. Moon, J. S. Stanley, D. Burkhardt, D. van Dijk, G. Wolf, and S. Krishnaswamy, "Manifold learning-based methods for analyzing single-cell rna-sequencing data," *Current Opinion in Systems Biology*, vol. 7, pp. 36–46, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2452310017301877

[47] P. J. Brockwell and R. A. Davis, *Introduction to time series and forecasting*. Springer, 2002.

[48] Y. S. Shin and S. DuBrow, "Structuring memory through inference-based event segmentation," *Topics in Cognitive Science*, vol. 13, no. 1, pp. 106–127, 2021.

[49] C. S. Lee, M. Aly, and C. Baldassano, "Anticipation of temporally structured events in the brain," *eLife*, vol. 10, p. e64972, apr 2021. [Online]. Available: https://doi.org/10.7554/eLife.64972

[50] M. Kumar, M. J. Anderson, J. W. Antony, C. Baldassano, P. P. Brooks, M. B. Cai, P.-H. C. Chen, C. T. Ellis, G. Henselman-Petrusek, and D. Huberdeau, "Brainiak: The brain imaging analysis kit," 2020.

[51] M. Hanke, F. J. Baumgartner, P. Ibe, F. R. Kaule, S. Pollmann, O. Speck, W. Zinke, and J. Stadler, "A high-resolution 7-tesla fmri dataset from complex natural stimulation with an audio movie," *Scientific Data*, vol. 1, no. 1, p. 140003, 2014. [Online]. Available: https://doi.org/10.1038/sdata.2014.3https://www.nature.com/articles/sdata20143.pdf

[52] A. Sengupta, F. R. Kaule, J. S. Guntupalli, M. B. Hoffmann, C. Häusler, J. Stadler, and M. Hanke, "A studyforrest extension, retinotopic mapping and localization of higher visual areas," *Scientific Data*, vol. 3, no. 1, p. 160093, 2016. [Online]. Available: https://doi.org/10.1038/sdata.2016.93

[53] J. V. Haxby, J. S. Guntupalli, S. A. Nastase, and M. Feilong, "Hyperalignment: Modeling shared information encoded in idiosyncratic cortical topographies," *eLife*, vol. 9, 2021. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7266639/

[54] W. R. Shirer, S. Ryali, E. Rykhlevskaia, V. Menon, and M. D. Greicius, "Decoding subject-driven cognitive states with whole-brain connectivity patterns," *Cereb Cortex*, vol. 22, no. 1, pp. 158–65, 2012.

[55] K. Vodrahalli, P.-H. Chen, Y. Liang, C. Baldassano, J. Chen, E. Yong, C. Honey, U. Hasson, P. Ramadge, K. A. Norman, and S. Arora, "Mapping between fmri responses to movies and their natural language annotations," *NeuroImage*, vol. 180, no. Pt A, pp. 223–231, 2018. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/28648889https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5742073/

[56] C. Husler and M. Hanke, "An annotation of cuts, depicted locations, and temporal progression in the motion picture "forrest gump" [version 1; peer review: 3 approved]," *F1000Research*, vol. 5, no. 2273, 2016. [Online]. Available: https://f1000research.com/articles/5-2273/v1https://f1000researchdata.s3.amazonaws.com/manuscripts/10273/31755005-f43c-4206-b258-b0906f7e6c50_9536_-_michael_hanke.pdf?doi=10.12688/f1000research.9536.1&numberOfBrowsableCollections=26&numberOfBrowsableInstitutionalCollections=4&numberOfBrowsableGateways=29