frontiers in Neuroscience

ORIGINAL RESEARCH
published: 10 July 2018
doi: 10.3389/fnins.2018.00437

Modeling Semantic Encoding in a Common Neural Representational Space

Cara E. Van Uden[1†], Samuel A. Nastase[1,2†], Andrew C. Connolly[3], Ma Feilong[1], Isabella Hansen[1], M. Ida Gobbini[1,4] and James V. Haxby[1]*

[1] Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, United States, [2] Princeton Neuroscience Institute, Princeton University, Princeton, NJ, United States, [3] Department of Neurology, Geisel School of Medicine, Dartmouth College, Hanover, NH, United States, [4] Dipartimento di Medicina Specialistica, Diagnostica e Sperimentale (DIMES), Medical School, University of Bologna, Bologna, Italy

Encoding models for mapping voxelwise semantic tuning are typically estimated separately for each individual, limiting their generalizability. In the current report, we develop a method for estimating semantic encoding models that generalize across individuals. Functional MRI was used to measure brain responses while participants freely viewed a naturalistic audiovisual movie. Word embeddings capturing agent-, action-, object-, and scene-related semantic content were assigned to each imaging volume based on an annotation of the film. We constructed both conventional within-subject semantic encoding models and between-subject models where the model was trained on a subset of participants and validated on a left-out participant. Between-subject models were trained using cortical surface-based anatomical normalization or surface-based whole-cortex hyperalignment. We used hyperalignment to project group data into an individual's unique anatomical space via a common representational space, thus leveraging a larger volume of data for out-of-sample prediction while preserving the individual's fine-grained functional–anatomical idiosyncrasies. Our findings demonstrate that anatomical normalization degrades the spatial specificity of between-subject encoding models relative to within-subject models. Hyperalignment, on the other hand, recovers the spatial specificity of semantic tuning lost during anatomical normalization, and yields model performance exceeding that of within-subject models.

Keywords: fMRI, forward encoding models, functional alignment, hyperalignment, individual variability, natural vision, semantic representation

OPEN ACCESS

Edited by:
Michael Hanke, Universitätsklinikum Magdeburg, Germany

Reviewed by:
Marcel van Gerven, Radboud University Nijmegen, Netherlands
Shinji Nishimoto, CiNet, National Institute of Information and Communications Technology, Japan

****Correspondence:***
Samuel A. Nastase
sam.nastase@gmail.com

[†] These authors have contributed equally to this work.

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *25 March 2018*
Accepted: *11 June 2018*
Published: *10 July 2018*

Citation:
Van Uden CE, Nastase SA, Connolly AC, Feilong M, Hansen I, Gobbini MI and Haxby JV (2018) Modeling Semantic Encoding in a Common Neural Representational Space. Front. Neurosci. 12:437. doi: 10.3389/fnins.2018.00437

INTRODUCTION

Recent neuroimaging work has revealed widespread cortical representation of semantic content conveyed by visual and linguistic stimuli (Huth et al., 2012, 2016; Wehbe et al., 2014; Pereira et al., 2018). These findings hinge on the development of forward encoding models, which find a mapping from stimuli to voxelwise responses via a complex intermediate feature space (Naselaris et al., 2011). These feature spaces may capture distributional properties of large corpora of text (e.g., word co-occurrence) in the case of semantic representation (e.g., Mitchell et al., 2008; Huth et al., 2016), or comprise neurally inspired models of vision (e.g., Kay et al., 2008; Nishimoto et al., 2011; Güçlü and van Gerven, 2015) or audition (e.g., Santoro et al., 2014; de Heer et al., 2017).

If the intermediate feature space adequately captures stimulus qualities of interest and the model is trained on a sufficiently diverse sample of stimuli, the estimated model will generalize well to novel stimuli. Naturalistic stimuli and tasks (such as watching movies, listening to stories) enhance this approach by evoking reliable neural responses (Hasson et al., 2010) and broadly sampling stimulus space (Haxby et al., 2014), as well as increasing ecological validity (Felsen and Dan, 2005) and participant engagement (Vanderwal et al., 2017).

Although encoding models provide a fine-grained voxel-specific measure of functional tuning, they are typically estimated independently for each participant (e.g., Huth et al., 2012, 2016). This is problematic because we can collect only a limited volume of data in any one participant, and each participant's model has limited generalizability across individuals (cf. Yamada et al., 2015; Güçlü and van Gerven, 2017; Vodrahalli et al., 2017). Recent work has demonstrated that group-level estimates of functional organization obscure marked individual-specific idiosyncrasies (Laumann et al., 2015; Braga and Buckner, 2017; Gordon et al., 2017). This is because functional–anatomical correspondence—the mapping between functional tuning and macroanatomical structure—varies considerably across individuals (Watson et al., 1993; Riddle and Purves, 1995; Aine et al., 1996; Frost and Goebel, 2012; Zhen et al., 2015, 2017). While macroanatomical normalization (i.e., nonlinear volumetric or cortical surface-based alignment) may be sufficient for capturing commonalities in coarse-grained functional areas, it cannot in principle align fine-grained functional topographies across individuals (cf. Sabuncu et al., 2010; Conroy et al., 2013). If we hope to predict functional tuning across individuals at the specificity of individual cortical vertices, we need to circumvent the correspondence problem between function and anatomy (Dubois and Adolphs, 2016; Poldrack, 2017).

In the following, we outline an approach for estimating encoding models that can make detailed predictions of responses to novel stimuli in novel individuals at the specificity of cortical vertices. To accommodate idiosyncratic functional topographies, we use hyperalignment to derive transformations to map each individual's responses into a common representational space (Haxby et al., 2011; Guntupalli et al., 2016). The searchlight hyperalignment algorithm learns a locally constrained whole-cortex transformation rotating each individual's anatomical coordinate space into a common space that optimizes the correspondence of representational geometry (in this case, the response patterns to the movie stimulus at each time point) across brains. We use a dynamic, naturalistic stimulus – the *Life* nature documentary narrated by David Attenborough – for the dual purpose of deriving hyperalignment transformations and fitting the encoding model. Using a naturalistic paradigm that thoroughly samples both stimulus space and neural response space is critical for robustly fitting the encoding model and ensuring the hyperalignment transformations generalize to novel experimental contexts (Haxby et al., 2011; Guntupalli et al., 2016).

Although hyperalignment dramatically improves between-subject decoding (Haxby et al., 2011; Guntupalli et al., 2016), relatively few attempts have been made to integrate hyperalignment and voxelwise encoding models. Yamada et al. (2015) used a many-to-one sparse regression to predict voxel responses to simple visual stimuli across pairs of participants. Bilenko and Gallant (2016) implemented hyperalignment using regularized kernel canonical correlation analysis (Xu et al., 2012) to compare encoding models across subjects. Recent work (Vodrahalli et al., 2017) using a probabilistic, reduced-dimension variant of hyperalignment (Chen et al., 2015) has suggested that encoding models perform better in a lower-dimensional shared response space. Finally, Güçlü and van Gerven (2017) and Wen et al. (2018) have employed hyperalignment in conjunction with a deep convolutional neural network (e.g., Tran et al., 2015) to predict responses to video clips visual areas. They demonstrated that estimating an encoding model in a common representational space does not diminish model performance, and that aggregating additional subjects in the common spaces can improve performance.

To evaluate hyperalignment in the context of encoding models, we compared within-subject encoding models and between-subject encoding models where a model trained on three-fourths of the movie in a subset of participants is used to predict responses at each cortical vertex for the left-out fourth of the movie in a left-out participant. We compared between-subject models using high-performing surface-based anatomical normalization (Klein et al., 2010; Fischl, 2012) and surface-based searchlight whole-cortex hyperalignment (Guntupalli et al., 2016). We model semantic tuning at each cortical vertex based on distributed word embeddings (word2vec; Mikolov et al., 2013) assigned to each imaging volume based on an annotation of the documentary. We first show that constructing between-subject models using anatomical alignment reduces the spatial specificity of vertex-wise semantic tuning relative to within-subject models. Next, we demonstrate that hyperalignment generally leads to improved between-subject model performance, exceeding within-subject models. Hyperalignment effectively recovers the specificity of within-subject models, allowing us to leverage a large volume of group data for individualized prediction at the specificity of individual voxels or cortical vertices.

MATERIALS AND METHODS

Participants

Eighteen right-handed adults (10 female) with normal or corrected-to-normal vision participated in the experiment. Participants reported no neurological conditions. All participants gave written, informed consent prior to participating in the study, and the study was approved by the Institutional Review Board of Dartmouth College. These data have been previously used for the purpose of hyperalignment in a published report by Nastase et al. (2017).

Stimuli and Design

Participants freely viewed four segments of the *Life* nature documentary narrated by David Attenborough. The four runs were of similar duration (15.3, 14, 15.4, and 16.5 min), totaling 63 min. The movie stimulus included both the visual and

auditory tracks, and sound was adjusted to a comfortable level for each participant. The video was back-projected on a screen placed at the rear of the scanner bore, and was viewed with a mirror attached to the head coil. Audio was delivered using MRI-compatible fiber-optic electrodynamic headphones (MR confon GmbH, Magdeburg, Germany). Participants were instructed to remain still and watch the documentary as though they were watching a movie at home. Note that this free viewing task contrasts with prior forward-encoding studies that enforced central fixation while viewing videos (e.g., Nishimoto et al., 2011; Huth et al., 2012), which we expect to affect the comparative performance of forward encoding models, especially in early visual cortex; however, a full treatment of the magnitude of such effects is beyond the scope of this paper. Stimuli were presented using PsychoPy (Peirce, 2007).

Image Acquisition

Structural and functional images were acquired using a 3T Philips Intera Achieva MRI scanner (Philips Medical Systems, Bothell, WA, United States) with a 32-channel phased-array SENSE (SENSitivity Encoding) head coil. Functional, blood-oxygenation-level-dependent (BOLD) images were acquired in an interleaved fashion using single-shot gradient-echo echo-planar imaging with fat suppression and a SENSE parallel acceleration factor of 2: TR/TE = 2500/35 ms, flip angle = 90°, resolution = 3 mm^3 isotropic voxels, matrix size = 80 × 80, FOV = 240 × 240 mm^2, 42 transverse slices with full brain coverage and no gap, anterior–posterior phase encoding. Four runs were collected for each participant, consisting of 374, 346, 377, and 412 dynamic scans, or 935, 865, 942.5, and 1030 s, respectively. A T1-weighted structural scan was obtained using a high-resolution 3D turbo field echo sequence: TR/TE = 8.2/3.7 ms, flip angle = 8°, resolution = 0.9375 × 0.9375 × 1.0 mm^3, matrix size = 256 × 256 × 220, and FOV = 240 × 240 × 220 mm^3.

Preprocessing

Raw data were organized to conform to the Brain Imaging Data Structure (BIDS; Gorgolewski et al., 2016) specifications and were preprocessed using fmriprep (Gorgolewski et al., 2011, 2017; Esteban et al., 2017), which provides a streamlined, state-of-the-art preprocessing pipeline that incorporates various software packages. Within the fmriprep framework, cortical surfaces were reconstructed from the T1-weighted structural images using FreeSurfer (Dale et al., 1999) and spatially normalized to the fsaverage6 template based on sulcal curvature (Fischl et al., 1999). Prior to spatial normalization, T2*-weighted functional volumes were slice-time corrected (Cox, 1996), realigned for head motion (Jenkinson et al., 2002), aligned to the anatomical image (Greve and Fischl, 2009), and sampled to the cortical surface. Time-series data were detrended using AFNI's 3dTproject (Cox, 1996), which removes nuisance variables and trends via a single linear regression model. The regression model included a framewise displacement regressor (Power et al., 2012), the first six principal components from cerebrospinal fluid (Behzadi et al., 2007), head motion parameters, first- and second-order polynomial trends, and a band-pass filter (0.00667–0.1 Hz). We did not explicitly spatially smooth the functional data during preprocessing. All surface data were visualized in SUMA (Saad et al., 2004).

Whole-Brain Hyperalignment

Surface-based searchlight whole-cortex hyperalignment (Haxby et al., 2011; Guntupalli et al., 2016) was performed based on the data collected while participants viewed the *Life* nature documentary using leave-one-run-out cross-validation: three of four runs were used to estimate the hyperalignment transformations for all participants; these transformations were then applied to the left-out run for model evaluation. The hyperalignment algorithm, described in detail by Guntupalli et al. (2016, 2018), uses iterative pairwise applications of the Procrustes transformation (Gower, 1975), effectively rotating a given subject's multivariate response space to best align their patterns of response to time-points in the movie with a reference time series of response patterns.

In the first iteration, the response trajectory of an arbitrarily chosen subject serves as the reference, and a second subject's data are rotated via the Procrustes transformation into alignment with that reference. For each additional subject, a new reference trajectory is computed by averaging the previously aligned subject's data and the reference, and the new subject is aligned to this reference. Aligning and averaging all subjects' data in this way results in an intermediate template. In the second iteration, each subject's data are again aligned to this intermediate reference, and the average of all subjects' aligned response vectors are recomputed. This average response trajectory serves as the final functional template in a common representational space. For each subject, we calculate a final transformation to this functional template. These hyperalignment transformations can then be used to project data from a left-out run into the common representational space, or the transpose of a given subject's transformation matrix can be used to project from the common space into a particular subject's response space.

To locally constrain hyperalignment, we compute these transformations separately within large 20 mm radius surface-based searchlight disks centered on each cortical vertex (Kriegeskorte et al., 2006; Oosterhof et al., 2011; Guntupalli et al., 2016). Each searchlight comprised on average 610 vertices (*SD* = 162, median = 594, range: 237–1,238 vertices). The resulting rotation parameters are only defined for vertices within a given searchlight; however, searchlights are heavily overlapping. These local transformations are aggregated by summing overlapping searchlight transformation parameters to construct a single sparse transformation for each cortical hemisphere. For each subject, this results in two $N \times N$ transformation matrices, one for each cortical hemisphere, where N is the number of vertices in a hemisphere (40,962 for the fsaverage6 template). These matrices contain non-zero values only for vertices within the radius of a searchlight. Because each vertex is a constituent of many overlapping searchlights, the final rotation parameters for a given vertex will reflect transformations for all searchlights to which it contributes. Response time series for each vertex are z-scored before and after each application of the Procrustes transformation. All functional data were anatomically normalized to the fsaverage6 template

prior to hyperalignment. This procedure is not strictly necessary for hyperalignment (and is not optimal due to interpolation during surface projection), but is used here for simplicity and to facilitate comparison between anatomically normalized and hyperaligned data. Note that hyperalignment does not yield a one-to-one mapping between voxels or vertices across subjects, but rather models each voxel's or vertex's response profile in a given subject as a weighted sum of local response profiles in the common space.

The searchlight hyperalignment algorithm generates an abstract feature space that does not directly map onto the anatomical space of any particular subject. This means we cannot directly compare, vertex by vertex, data in the common space generated by hyperalignment to data in any individual subject's anatomical space. To directly compare the hyperaligned between-subject model to the other two types of models, for each leave-one-subject-out cross-validation fold we first transformed the training data – 3 runs for each of 17 subjects – into the common space using each subject's unique hyperalignment transformation, and then mapped all 17 training subjects' response vectors from the common space into the left-out test subject's anatomical space (normalized to the fsaverage6 template) using the transpose of the left-out test subject's hyperalignment transformation matrix. That is, for each left-out test subject, we mapped responses for the 17 training subjects into the left-out subject's space *via* the common space. We then averaged response time series across training subjects. We did not apply any hyperalignment transformations to the validation data (the left-out test subject's left-out test run). Note, however, that the whole-cortex matrix of local transformations learned by the searchlight hyperalignment algorithm is not orthogonal. This approach allows us to directly compare the three types of vertex-wise models on a subject-by-subject basis (i.e., when performing paired statistical tests). Whole-brain hyperalignment and several of the subsequent analyses were implemented using PyMVPA (Hanke et al., 2009).

Semantic Features

The *Life* documentary was annotated with a list of words describing the agents (i.e., animals), actions, objects, and scene for each camera angle of the movie. For example, if one camera angle depicted a giraffe eating grass on the savannah, the corresponding annotation would be the list of words "giraffe," "eating," "grass," and "savannah." Then, the camera angle annotations were interpolated for every 2.5 s of the movie, so that every imaging volume was assigned semantic labels. On average, a single camera angle's annotations covered 1.97 TRs (SD = 1.20). The annotation contained 277 unique words in total, and each imaging volume was assigned on average 5.28 words (SD = 1.91).

Next, we assigned a 300-dimensional word2vec semantic feature vector to each label in the annotation. We used pre-trained word2vec embeddings comprising a vocabulary of 3 million words and phrases trained on roughly 100 billion words from the Google News text corpus using the skip-gram architecture (Mikolov et al., 2013). Semantic vectors for all labels assigned to a given imaging volume were averaged to create a single 300-dimensional semantic vector per volume (cf. Vodrahalli et al., 2017). Mikolov et al. (2013) demonstrated that the word representations learned by the skip-gram model exhibit a linear structure that makes it possible to meaningfully combine words by an element-wise addition of their vector representations.

To accommodate the delayed hemodynamic response, we concatenated semantic vectors from the previous TRs (2.5, 5.0, 7.5, and 10.0 s; similarly to, e.g., Huth et al., 2012). The final vector assigned to each imaging volume for training and testing the encoding model comprised a concatenated 1,200-dimensional vector capturing the semantic content of the four preceding time points.

Regularized Regression

We estimated vertex-wise forward encoding models using the L2-penalized linear least squares regression (i.e., ridge regression) in three different ways: (a) within-subject models; (b) between-subject models using anatomical normalization; and (c) between-subject models using hyperalignment following anatomical normalization. All models were evaluated using leave-one-run-out cross-validation. In each of these leave-one-out runs, the within-subject and between-subject models were trained as follows: within-subject models were trained on three of the four imaging runs, then tested on the left-out fourth run separately for each subject. Between-subject models were trained on the averaged time series of 17 of the 18 participants over three of the four runs. The estimated between-subject models were then tested on the left-out fourth run in the left-out 18th participant. This yielded 72 total data folds for each of the three types of models. **Figure 1** schematically depicts our approach for constructing between-subject semantic encoding models using hyperalignment.

In these models, the number of predictor variables (1,200) exceeds the number of observations (ranging from 1,097 to 1,163 imaging volumes). We used ridge regression to estimate regression coefficients (weights) for the semantic predictor variables so as to best predict the response time series at each vertex. We used a modified implementation of ridge regression authored by Huth et al. (2012). Ridge regression uses a regularization hyperparameter to control the magnitude of the regression coefficients, where a larger regularization parameter yields greater shrinkage and reduces the effect of collinearity among predictor variables. The regularization parameter was chosen using leave-one-run-out cross-validation nested within each set of training runs. We estimated regression coefficients for a grid of 20 regularization parameters log-spaced from 1 to 1,000 at each vertex within each set of two runs in the training set of three runs. We then predicted the responses for the held-out third run (within the training set) and evaluated model prediction performance by computing the correlation between the predicted and actual responses. These correlations were averaged over the three cross-validation folds nested within the training set, then averaged across all vertices. We then selected the regularization parameter with the maximal model performance across runs and vertices. Selecting a single regularization parameter across all vertices ensures that estimated regression coefficients are comparable across vertices. This regularization parameter

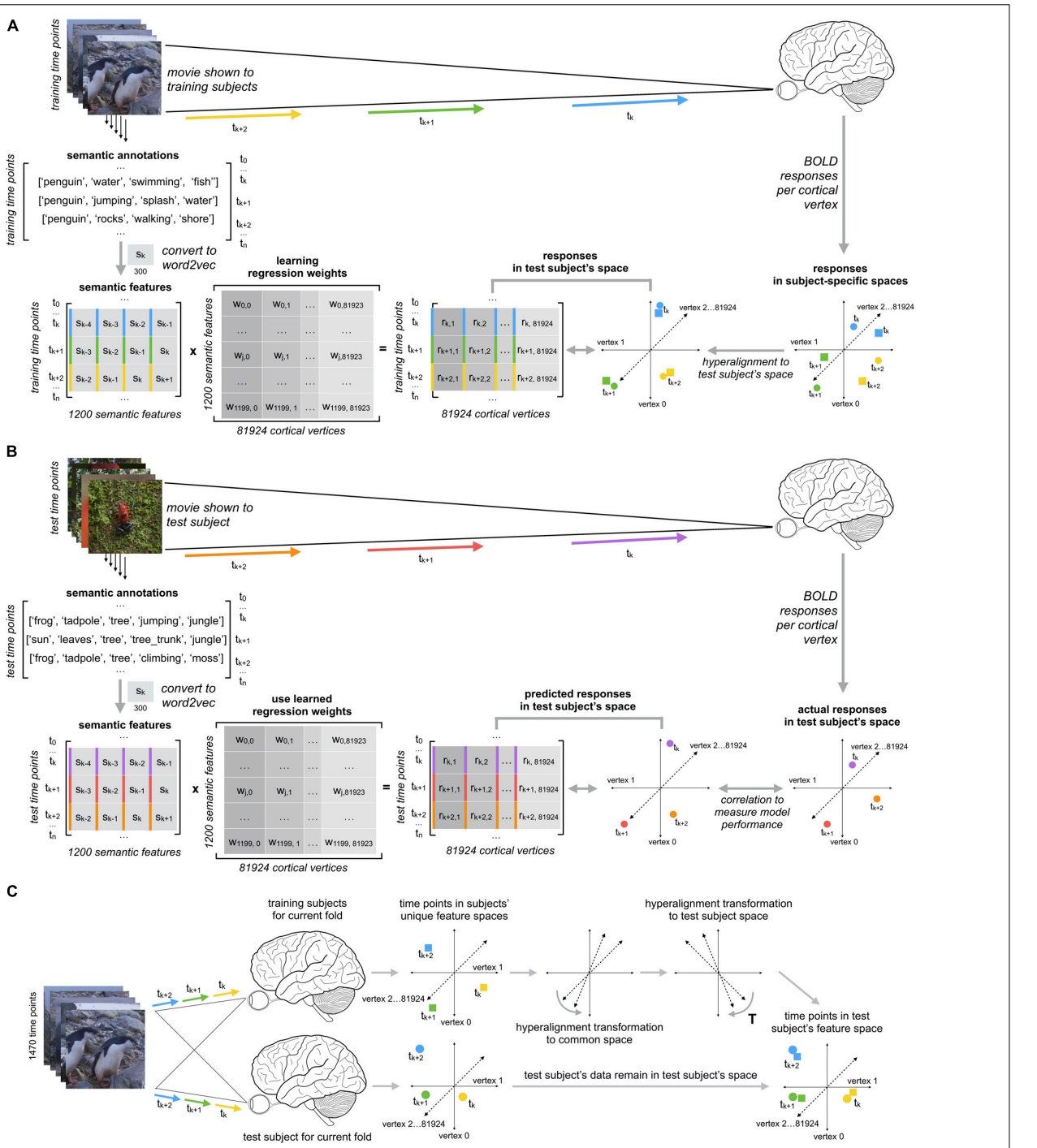

FIGURE 1 | Schematic for constructing between-subject semantic encoding models using hyperalignment. The schematic depicts onefold of the nested leave-one-out cross-validation procedure repeated for 4 test runs and 18 test participants (72 cross-validation folds in total). **(A)** Training between-subject semantic encoding models using ridge regression. Regression coefficients (weights) are estimated to predict response time series per vertex based on three training runs. **(B)** Testing semantic encoding models. Regression weights estimated on training data are used to predict response time series for a fourth test run. Model prediction performance is evaluated by computing the Pearson correlation between the predicted responses and are the actual response time series per vertex. **(C)** Hyperalignment for between-subject semantic encoding models. For each test subject in the leave-one-subject-out cross-validation procedure, we first projected each training subject's data into the common space using their subject-specific hyperalignment transformations. We then use the transpose of the test subject's hyperalignment transformation to project all training subjects' data into the test subject's space. We averaged response vectors for all training subjects in the test subject's space, then trained the encoding model on this averaged response trajectory. Finally, we evaluated between-subject model performance by predicting vertex-wise response time series for the left-out test run in the left-out test participant, and computed the Pearson correlation between the predicted time series and the actual time series per vertex.

was then used at the final stage when estimating the encoding model across all three training runs for evaluation on the left-out fourth run. Note, however, that different regularization parameters were chosen for each of the four leave-one-run-out cross-validation folds (where both stimuli and hyperalignment transformation differed for each set of training runs), and for each of the 18 leave-one-subject-out cross-validation folds used for between-subject models. For the two between-subject models, the optimal regularization parameter was either 12.74 or 18.33 for every test subject (due to averaging response time series across training subjects). Note that these regularization parameters are considerably lower than those reported by Huth et al. (2016). This may be due to several factors, including our procedure for averaging time series across subjects during training, having fewer time points in the training set, and our use of the relatively dense lower-dimensional word2vec embeddings. However, for within-subject models, the optimal regularization parameter was more variable, likely due to increased noise.

To evaluate the vertex-wise forward encoding models, we used the regression coefficients from the model trained on three training runs to predict the response time series for the left-out fourth run. For between-subject models, we used the regression coefficients estimated on the training runs in the training subjects (transformed into the test subject's space via the common space estimated using hyperalignment) to predict responses for the left-out run in the left-out subject. For between-subject models, both the hyperalignment transformations and encoding models were cross-validated to previously unseen data; the test run in the test subject played no role in estimating the hyperalignment transformations or the regression weights of the encoding model. For each vertex, we then computed the Pearson correlation between the predicted time series and actual time series for that run to measure model prediction performance (as in, e.g., Huth et al., 2012). Pearson correlations were Fisher z-transformed prior to statistical tests. We then averaged together the Pearson correlations for each of the four held-out test runs for visualization.

RESULTS

Inter-Subject Correlations

To ensure that the common space learned by hyperalignment finds common bases for fine-grained functional topographies across subjects, we computed ISCs for both vertex-wise response time series and searchlight representational geometries using anatomical normalization and hyperalignment. To assess how hyperalignment impacts time series ISCs (Hasson et al., 2004), for each run of the movie we computed the correlations between each subject's time series per surface vertex and the average of all other subjects before and after hyperalignment (Guntupalli et al., 2016). We then averaged these ISCs across all four movie runs; this results in a correlation for each subject for each vertex. We visualized the mean correlation across subjects for each vertex. Hyperalignment improved inter-subject correlations of time series throughout cortex, particularly in posterior perceptual regions, but also in lateral prefrontal regions presumably supporting processes engaged by the movie stimulus, such as attention and working memory. Some regions, such as central sulcus and medial prefrontal cortex, have low inter-subject correlations: these regions may primarily encode information (e.g., motor behaviors) unrelated to the experimental paradigm and thus would not be engaged by the stimulus consistently across subjects or at all. **Figure 2A** shows cortical maps of time series ISCs before and after hyperalignment. Hyperalignment increased the mean ISC of time series across vertices from 0.077 to 0.151.

We next analyzed the ISC of local representational geometries by calculating representational dissimilarity matrices (RDMs) comprising pairwise correlations between response vectors for all time points (Kriegeskorte et al., 2006; Oosterhof et al., 2011) in the test run of the movie using 9 mm radius surface-based searchlight disks (Kriegeskorte et al., 2006; Oosterhof et al., 2011). This procedure was repeated for each of the four runs. We averaged all pairwise correlations in the upper triangle of this matrix as well as averaging across runs. All operations involving correlations were performed after Fisher z-transformation and the results were inverse Fisher transformed for visualization. Hyperalignment also improved inter-subject correlations of searchlight representational geometries throughout much of cortex. **Figure 2B** shows cortical maps of ISCs of representational geometries before and after hyperalignment. Hyperalignment increased mean ISC of representational geometries across vertices from 0.157 to 0.230.

Differences in Model Performance

We formally compared three types of semantic encoding models: within-subject models, between-subject models using anatomical normalization, and between-subject models using hyperalignment. For each subject, the within-subject model was compared to the between-subject models where that subject served as the test subject. For the hyperaligned between-subject model, group data were projected into the test subject's space prior to model estimation. **Figure 3** depicts model prediction performance for the three model types in two representative subjects, while **Figure 4** depicts average model performance across subjects.

We summarized differences in model performance across the entire cortex in two ways. To constrain our analysis to well-predicted vertices, for each subject we selected the 10,000 vertices with highest model performance separately for each model. We then considered only the union of well-predicted vertices across all three models (on average 15,724 vertices per subject, SD = 1,293 across subjects). First, for each pair of models, we computed the proportion of vertices with greater model prediction performance (i.e., correlation between predicted and actual time series for the test data) for one model relative to the other. We calculated these proportions per subject, then computed a paired t-test to assess statistical significance per model pair. When comparing the model performance for the within-subject and the between-subject models, the between-subject model using anatomical alignment yielded higher correlations in 50.7% of selected cortical vertices [$t(17) = 0.717$, $p = 0.483$]. The between-subject model using hyperalignment

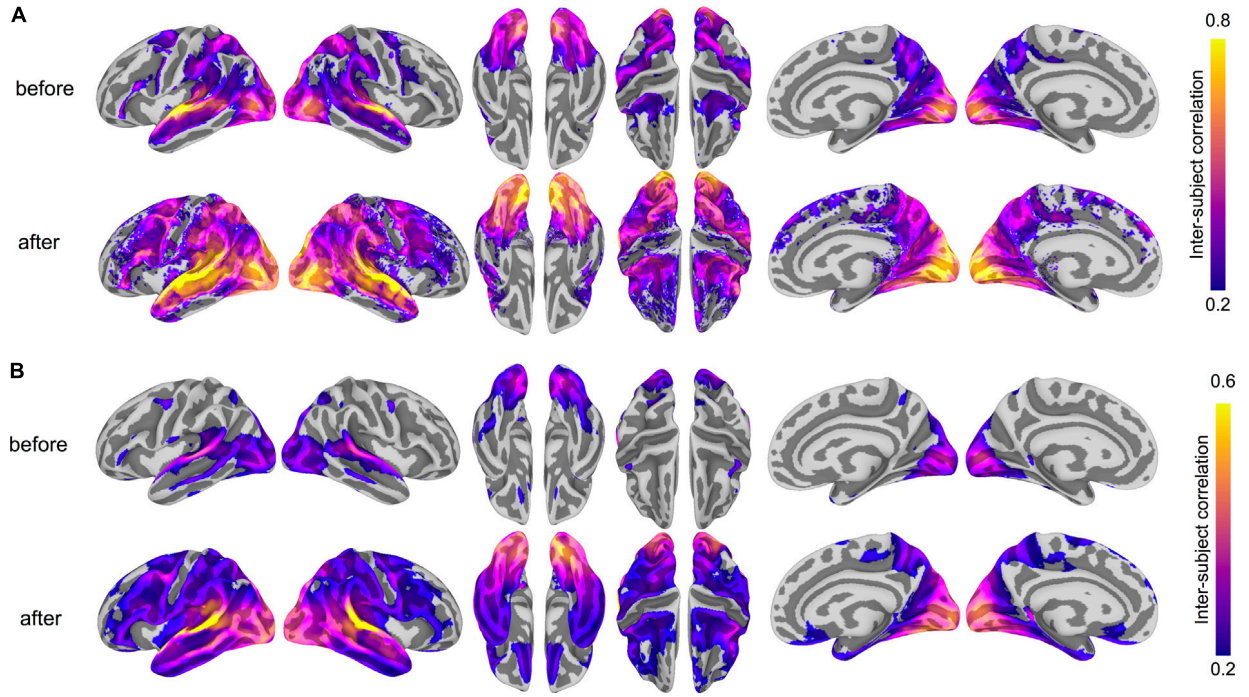

FIGURE 2 | Hyperalignment improves inter-subject correlation (ISC) of response profiles and representational geometry. **(A)** ISC of vertex-wise response time series before and after hyperalignment. Colored vertices reflect the mean ISC across subjects, thresholded at a mean correlation of 0.2. ISCs are highest in the superior temporal gyrus (in the vicinity of auditory cortex), as well as the dorsal and ventral visual pathways, comprising early visual, lateral occipitotemporal, ventral temporal, posterior parietal, and intraparietal cortices. **(B)** ISC of searchlight representational geometries (time-point RDMs in 9 mm radius searchlights) before and after hyperalignment. Colored vertices reflect the mean pairwise correlation across subjects for each searchlight, thresholded at a mean correlation of 0.2, revealing a broader extent of cortex with improved alignment of functional topography after hyperalignment. ISCs were Fisher *z*-transformed before averaging across all subjects and inverse Fisher transformed before mapping onto the cortical surface for visualization. All maps are rendered on the fsaverage6 surface template.

yielded better performance than the within-subject model in 58.9% of selected cortical vertices [$t(17) = 8.539$, $p < 0.001$]. The between-subject model using hyperalignment also yielded better performance than the between-subject model using anatomical alignment [58.7% of cortical vertices; $t(17) = 20.736$, $p < 0.001$].

Second, we assessed the difference in model prediction performance averaged across the same subset of well-predicted vertices. The between-subject model using anatomical alignment performed similarly to the within-subject model [0.120 and 0.124, respectively; $t(17) = 1.866$, $p = 0.079$]. The between-subject model using hyperalignment performed better than the within-subject model [0.135 and 0.124, respectively; $t(17) = 8.547$, $p < 0.001$]. Additionally, hyperalignment exceeded anatomical alignment when comparing the performance of between-subject models [0.135 and 0.120, respectively; $t(17) = 15.800$, $p < 0.001$].

To visualize differences in model performance, we compared model performance maps on the cortical surface (**Figure 5**). We computed vertex-wise paired *t*-tests for each of the three model comparisons. For visualization, we thresholded maps at a *t*-value of 2.11 ($p < 0.05$, two-tailed test, uncorrected for multiple tests).

Spatial Specificity of Semantic Tuning

To compare the spatial specificity of semantic tuning across model types, we computed the spatial point spread function (PSF) of the semantic model predictions (**Figure 6**). To constrain our analysis to well-predicted vertices, for each subject we again selected the 10,000 vertices with highest model performance separately for each model and considered only the union of these well-predicted vertices across all three models.

For each well-predicted vertex, we computed the model prediction performance (Pearson correlation between predicted and actual time series) for that vertex, and for neighboring vertices using the same prediction equation at 2 mm intervals up to 12 mm. That is, we used the encoding model at each vertex to predict the actual time series at neighboring, increasingly distant vertices. Each "ring" of vertices (e.g., the ring of vertices at a radius 10–12 mm from the central vertex of interest) was 2 mm wide and excluded vertices sampled at smaller radii. For a given ring of vertices, model performance was computed at each vertex in the ring and averaged across those vertices. Model performances at each radius per vertex were then averaged across the set of selected well-predicted vertices. To statistically assess PSFs, we computed bootstrapped confidence intervals around the model performance estimates at each radius by resampling subjects with replacement. To quantify the decline in spatial specificity of model performance over radii, we fit a logarithmic function to the PSF for each model at the midpoint of each ring (i.e., the vertex of interest, 1 mm, 3 mm, etc.) and reported the slope of this fit. The spatial point-spread function of the model predictions for the between-subject model using

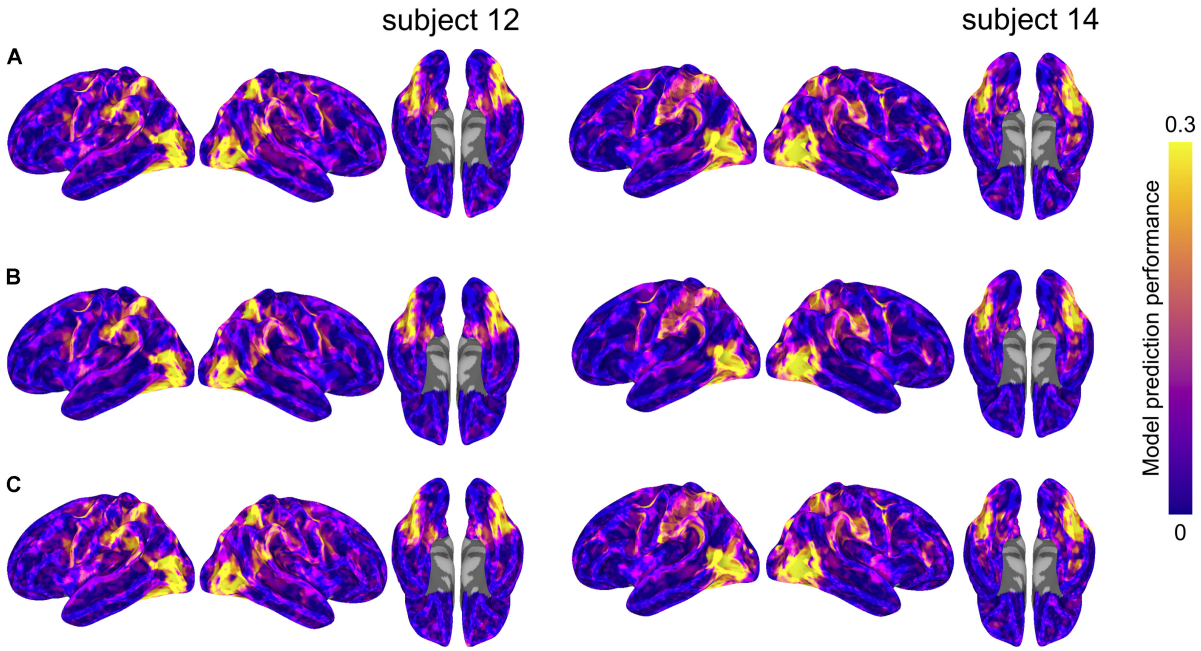

FIGURE 3 | Model prediction performance maps for two example subjects (left and right). Colored vertices reflect the Pearson correlation between the predicted and actual time series averaged across the four test runs. Three types of models are presented: **(A)** within-subject model performance maps; **(B)** between-subject anatomically aligned model performance maps; and **(C)** between-subject hyperaligned model performance maps. Maximum correlations are 0.37 (both subjects 12 and 14, area LO, all three maps) and correlations at the 95% level for vertices within the mask ranged from 0.29 to 0.31. All maps are unthresholded and uncorrected for multiple tests.

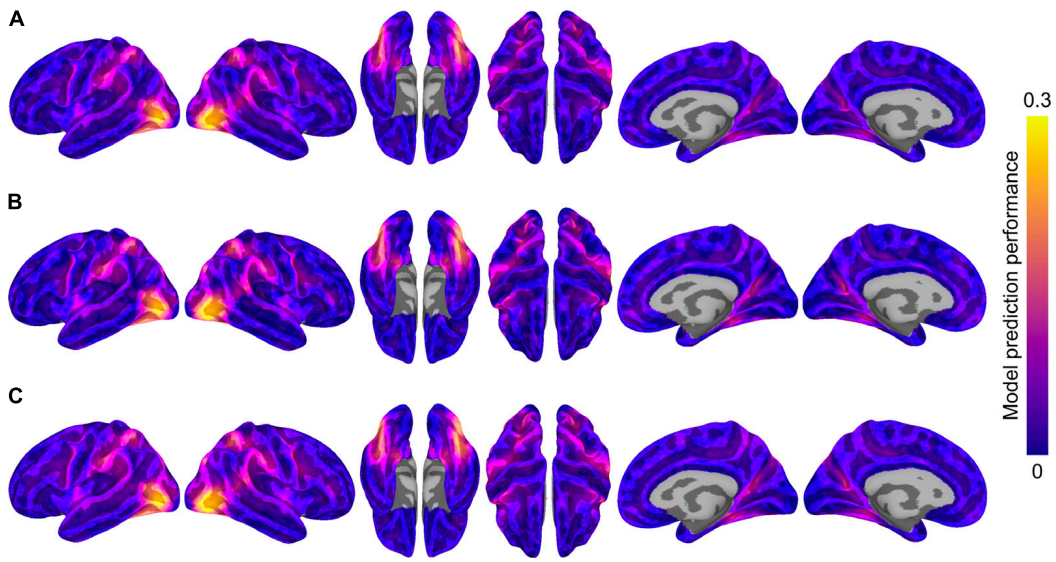

FIGURE 4 | Model prediction performance maps averaged across subjects. Colored vertices reflect the Pearson correlation between the predicted and actual time series averaged across the four test runs and averaged across subjects. Three types of models are presented: **(A)** within-subject model performance maps; **(B)** between-subject anatomically aligned model performance maps; and **(C)** between-subject hyperaligned model performance maps. Maximum correlations are 0.27 (in area LO, all three maps) and correlations at the 95% level for vertices within the mask ranged from 0.11 to 0.12. All maps are unthresholded and uncorrected for multiple tests.

anatomical alignment was relatively flat [negative slope of the logarithmic fit = 0.0172 (0.0166, 0.0178)]. The within-subject and hyperaligned between-subject models had steeper slopes [0.0447 (0.0409, 0.0488) and 0.0396 (0.0375, 0.0418), respectively; both $p < 0.001$], indicating greater spatial specificity in semantic tuning.

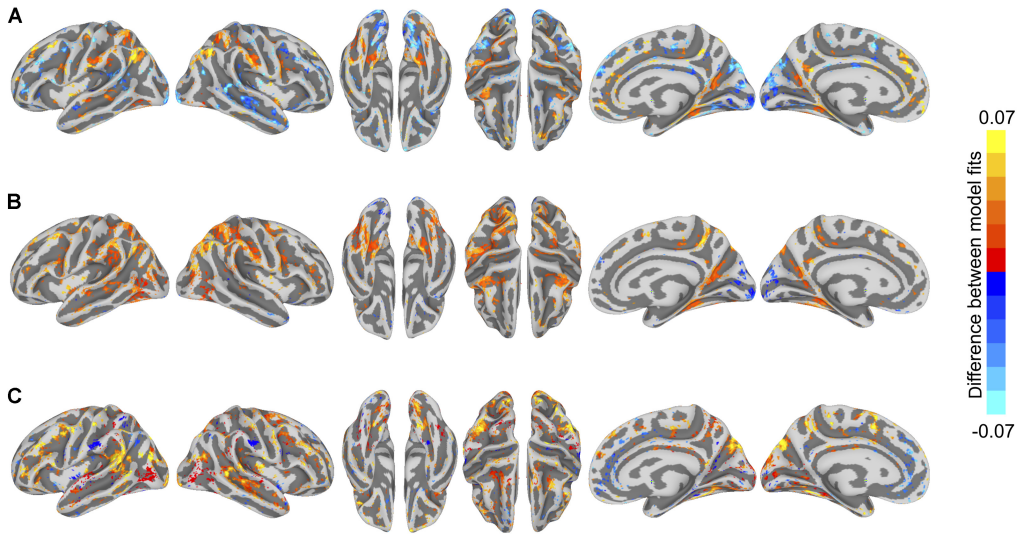

FIGURE 5 | Differences in semantic encoding model performance maps. **(A)** Paired differences in model performance between between-subject models using anatomical normalization and within-subject models. Warm colors indicate vertices where the anatomically aligned between-subject model performance exceeds within-subject model performance, and cool colors indicate where within-subject model performance exceeds anatomically aligned between-subject model performance. **(B)** Paired differences in model performance between between-subject models using hyperalignment and within-subject models. Warm colors indicate vertices where the hyperaligned between-subject model performance exceeds within-subject model performance, and cool colors indicate where within-subject model performance exceeds hyperaligned between-subject model performance. **(C)** Paired differences in model performance for between-subject models using hyperalignment and anatomical normalization. Warm colors indicate vertices where the hyperaligned between-subject model performance exceeds anatomically aligned between-subject model performance, and cool colors indicate where anatomically aligned between-subject model performance exceeds hyperaligned between-subject model performance. Colored vertices reflect mean paired differences in model performance, thresholded at an absolute t-value of $t(17) = 2.11$, $p < 0.05$, uncorrected for multiple comparisons.

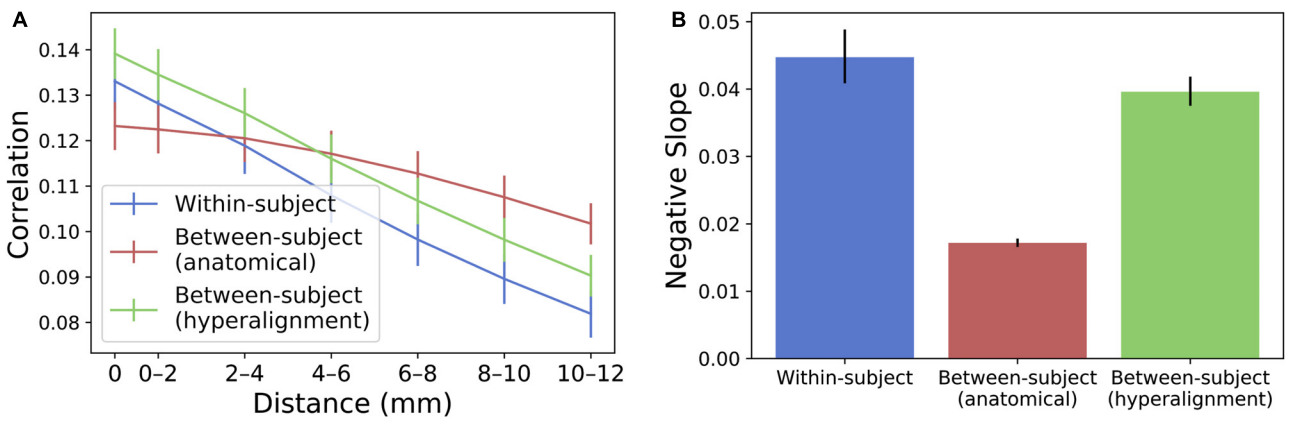

FIGURE 6 | Spatial PSF of semantic tuning. **(A)** Correlation between the predicted time series of one vertex and the actual time-series of its neighboring vertices up to 12 mm away. Correlations were aggregated based on distance from the central vertex of interest and averaged across vertices and subjects. Error bars denote 68% confidence intervals (standard error of the mean). **(B)** The within-subject and hyperaligned between-subject models have the steepest slopes (negative of the slope based on logarithmic curve fitting). Error bars denote 95% confidence intervals obtained by bootstrapping subjects 20,000 times with replacement.

The prediction performance maps for each model varied in their spatial smoothness (**Figure 7**). We computed the full width at half maximum (FWHM) of the model performance maps using SUMA's SurfFWHM. Spatial smoothness was computed per run in each hemisphere in each participant and averaged across hemispheres. Model performance maps for the between-subject model using anatomical alignment were significantly more spatially blurred than for the within-subject model [5.034 and 4.428 mm FWHM, respectively; $t(17) = 27.617$, $p < 0.001$]. The between-subject model using hyperalignment recovered the spatial specificity of the within-subject maps, and in fact yielded less smooth model performance maps (3.697 mm FWHM) than the within-subject model [$t(17) = 24.650$, $p < 0.001$].

We also assessed how well the between-subject model performance maps approximated the spatial organization of the within-subject model performance maps by computing

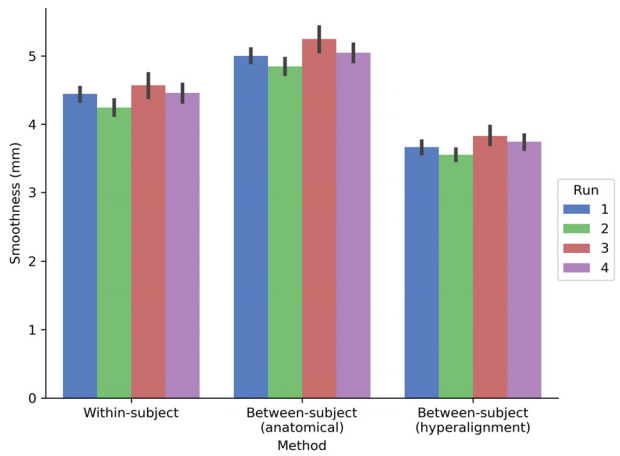

FIGURE 7 | Spatial smoothness (FWHM) of model prediction performance maps on the cortical surface. The between-subject model performance maps using anatomical alignment are blurred relative to the within-subject model performance maps, while the hyperaligned between-subject model recovers the spatial specificity of the within-subject model. The height of each bar indicates spatial smoothness averaged across hemispheres and participants for each run. Error bars indicate bootstrapped 95% confidence intervals estimated by resampling participants (1,000 bootstrap samples).

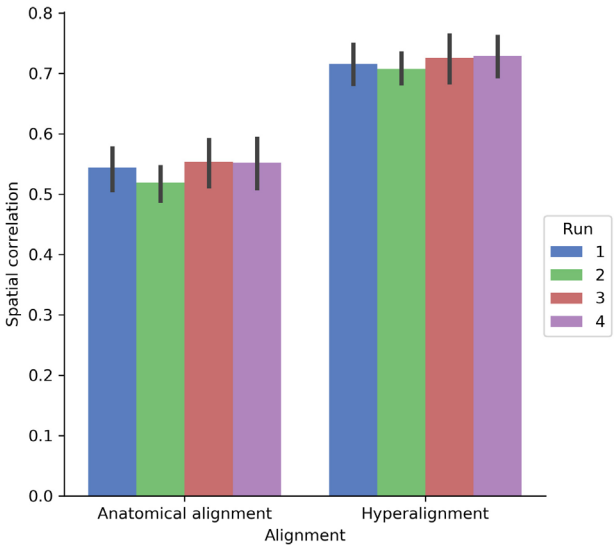

FIGURE 8 | Spatial correlation between within-subject model prediction performance maps and between-subject model prediction performance maps using either anatomical alignment or hyperalignment. The between-subject model using hyperalignment yielded a model performance map that is more similar to the within-subject model than the model performance map of the between-subject model using anatomical alignment. The height of each bar indicates mean spatial correlation across participants for each run. Error bars indicate bootstrapped 95% confidence intervals estimated by resampling participants (1,000 bootstrap samples).

the Pearson correlation between model performance maps (**Figure 8**). Correlations were computed across both cortical hemispheres within each participant and run. The spatial correlation between the model performance maps for the within-subject and between-subject models was 0.542 using anatomical normalization and 0.719 when using hyperalignment. That is, the spatial correlation between the map of within-subject model fits and the map of between-subject model fits increased by 0.177 after hyperalignment [a 33% increase; $t(17) = 22.432$, $p < 0.001$].

DISCUSSION

We developed a framework for constructing between-subject semantic encoding models that generalize to both novel stimuli and novel subjects. Vertex-wise forward encoding models were used in conjunction with hyperalignment to translate fine-grained functional topographies across individuals. Naturalistic experimental paradigms that broadly sample neural representational space play a critical role in this procedure, effectively enhancing the generalizability of both the encoding model and the hyperalignment transformations (Haxby et al., 2011, 2014).

Typically, encoding models are estimated separately for each subject using a relatively large volume of data (e.g., Mitchell et al., 2008; Huth et al., 2012; Pereira et al., 2018). Mirroring recent reports on resting-state functional connectivity in highly sampled individuals (Laumann et al., 2015; Gordon et al., 2017), these within-subject models can reveal highly detailed, idiosyncratic functional organization. However, there is a trade-off: we can only acquire large volumes of data in relatively few individuals (often the authors themselves, e.g., Nishimoto et al., 2011; Huth et al., 2012, 2016; Laumann et al., 2015; Gordon et al., 2017). This inherently limits the generality of conclusions drawn from within-subject models and undercuts efforts to relate the acquired data to between-subject variables. Constructing between-subject models that make individualized predictions in novel subjects is a critical step toward increasing the utility of cognitive neuroscience (Gabrieli et al., 2015). Although between-subject models can be constructed using anatomical normalization, this obscures considerable heterogeneity in functional organization because fine-scale variations in functional tuning are not tightly tethered to macroanatomical features (Guntupalli et al., 2016, 2018). Hyperalignment affords aggregation of data across individuals that aligns these fine-scale variations, thus alleviating this tension. In constructing a common representational space, we decouple functional tuning from anatomy, registering representational geometries rather than anatomical features. Unlike anatomical normalization, averaging across subjects in this space does not collapse responses that map onto topographies that are idiosyncratic to individual brains. Critically, we also preserve each individual's idiosyncratic functional–anatomical mapping in their respective transformation matrix, allowing us to project group data into any individual subject's anatomical space with high fidelity. This precision mapping enables out-of-sample prediction on the scale of individual voxels (Dubois and Adolphs, 2016; Poldrack, 2017).

Overall, our findings demonstrate that between-subject models estimated using hyperalignment outperform within-subject models. Between-subject models estimated using anatomical normalization yield artificially smooth maps of semantic tuning. Hyperalignment, on the other hand, retained the spatial specificity of within-subject models. The semantic encoding model used here best predicted responses in a network of areas previously implicated in representing animal taxonomy (Connolly et al., 2012, 2016; Sha et al., 2015) and observed action (Oosterhof et al., 2013; Wurm and Lingnau, 2015; Wurm et al., 2016; Nastase et al., 2017), including ventral temporal, lateral occipitotemporal, anterior intraparietal, and premotor cortices. Interestingly, inter-subject correlations were highest in superior temporal cortex, encompassing auditory areas. Although this suggests that the auditory narrative evoked highly reliable neural responses, in the present analyses the linguistic content of the narrative was not explicitly included in the semantic annotation.

The current approach for constructing between-subject encoding models using hyperalignment differs in several ways from related reports. Yamada et al. (2015) introduced a sparse regression algorithm for predicting voxel responses across pairs of subjects. This algorithm estimates a more flexible mapping than the Procrustes transformation between pairs of subjects and does not yield a common representational space across all subjects. Their approach was evaluated in early visual cortex using 10 pixel × 10 pixel black-and-white geometric images. While more recent work by Güçlü and van Gerven (2017) and Wen et al. (2018) used naturalistic visual stimuli, subjects in these studies were required to perform a highly non-naturalistic central fixation task (Nishimoto et al., 2011). Yamada et al. (2015), Güçlü and van Gerven (2017), and Wen et al. (2018) validated their models in a limited cohort of three subjects. Vodrahalli et al. (2017) used a variant of hyperalignment to estimate encoding models in a lower-dimensional (20 dimensions) common space. Models were evaluated in this low-dimensional shared space using a scene classification analysis. Their findings suggest that using a weighted averaging scheme for aggregating word embeddings assigned to a given imaging volume can improve model performance. However, in that experiment, annotators provided natural-language descriptions of the film (including many uninformative words). In the current study, the annotation included only the most salient or descriptive labels, effectively filtering out stop words and otherwise uninformative labels. Unlike previous reports, which limited their analyses to one or several regions of interest, we used searchlight hyperalignment to derive a locally constrained common space for each cortical hemisphere and estimated between-subject encoding models across the entire cortex. Searchlights were relatively large (20 mm radius), overlapping, and centered on every cortical vertex, yielding a spatially contiguous and smooth transformation. The motivation for using searchlights is to impose a spatial locality constraint on the transformation so that functional responses are not mixed across distant brain areas or cortical hemispheres. An alternative approach would be to perform hyperalignment within anatomically or functionally defined cortical parcels (e.g., Haxby et al., 2011; Chen et al., 2015). However, imposing *a priori* anatomical or functional boundaries may defeat the purpose, as the boundaries of parcels are inexact and highly variable across individuals (e.g., Laumann et al., 2015; Gordon et al., 2017). Alternative implementations of hyperalignment (e.g., Xu et al., 2012; Chen et al., 2015, 2016; Yousefnezhad and Zhang, 2017) may improve the prediction performance of voxel-wise encoding models, but carry with them different neuroscientific assumptions (e.g., the Procrustes transformation preserves representational geometry in high-dimensional representational spaces, while other transformations may not). Finally, none of the previously mentioned studies projected group data (via the common space) into each test subject's idiosyncratic response space prior to model estimation. Here, we used data from a naturalistic stimulus and task, transformed through a high-dimensional common space into each subject's idiosyncratic anatomical space, to estimate between-subject models across all cortical vertices.

We expected between-subject models with anatomical normalization alone to perform more poorly than they did. There are several possible reasons for this. First, FreeSurfer's surface-based anatomical normalization based on sulcal curvature is a high-performing, nonlinear normalization algorithm and outperforms commonly used linear volumetric normalization algorithms (as used in, e.g., Haxby et al., 2011). Second, our model is trained with a significantly smaller volume of data than comparable reports by, e.g., Huth et al., 2012 (∼45 min vs. ∼2 h). Third, our semantic labels were assigned at the temporal resolution of camera cuts (rather than, e.g., frames or TRs), then resampled to TRs. Note that model prediction performance (Pearson correlation) values reported here were not normalized by a noise ceiling estimate (cf. Huth et al., 2012).

Although the current findings demonstrate the utility of hyperalignment in constructing between-subject encoding models, there are several open questions. Under what circumstances will a between-subject model outperform within-subject models? Averaging group data in a common representational space provides a more robust estimate of response trajectories without sacrificing anatomical specificity. This should provide an advantage when predicting semantic tuning for noisy features in a given test subject, as the group estimate will be more robust. In addition to leveraging a larger volume of group data with the precision of within-subject models, hyperalignment effectively filters response profiles, suppressing variance not shared across subjects (Guntupalli et al., 2018). More generally, between-subject models can improve performance in areas where responses are highly stereotyped across individuals. For example, in the current study, both types of between-subject models improved model performance in anterior intraparietal areas, which are implicated in observed action representation during natural vision (Nastase et al., 2017).

When will hyperalignment fall short of within-subject performance? First, within-subject performance should be superior by virtue of capturing idiosyncratic functional

tuning, but it is usually impractical or impossible to collect sufficient data in each subject. Because hyperalignment largely preserves each subject's representational geometry, we expect any advantage will be attenuated when the test subject's representational geometry is idiosyncratic, irrespective of functional–anatomical correspondence (Kriegeskorte and Kievit, 2013; Charest et al., 2014). Note, however, that hyperalignment may serve to disentangle idiosyncrasies in representation from idiosyncrasies in functional–anatomical correspondence. Furthermore, we would not expect an advantage from hyperalignment if the stimulus or experimental paradigm used to derive the hyperalignment transformations did not adequately sample the neural representational subspaces important for estimating the encoding model. For example, hyperalignment may perform worse than within-subject models in early visual cortex because subjects freely viewed the movie stimulus, allowing for idiosyncratic gaze trajectories (note that the semantic model does not explicitly capture low-level visual features). Hyperalignment may perform worse than anatomical alignment in "task-negative" areas such as angular gyrus and medial prefrontal cortex by filtering the data without capturing any meaningful shared signal. This concern becomes particularly relevant if, in contrast to the current study, hyperalignment parameters and the encoding model are estimated on data derived from experimental paradigms that are more restricted and non-naturalistic. Finally, representations that are encoded in a coarse-grained or anatomically stereotyped manner will benefit less from hyperalignment, and anatomical normalization may be sufficient. However, as the resolution and sensitivity of functional measurements improves, and as more sophisticated encoding models begin to make finer-grained predictions, hyperalignment will become increasingly necessary.

DATA AVAILABILITY

The datasets generated and analyzed for this study can be found in the DataLad repository (http://datasets.datalad.org/?dir=/labs/haxby/life).

AUTHOR CONTRIBUTIONS

CVU, SN, AC, MF, MG, and JH designed the experiment. SN and AC collected the data. CVU, SN, MF, and IH analyzed the data. CVU, SN, AC, MF, and JH wrote the manuscript.

FUNDING

CVU was supported by the Neukom Scholarship in Computational Science, the Paul K. Richter and Evalyn E. Cook Richter Memorial Fund, and the David C. Hodgson Endowment for Undergraduate Research Award. This work was also supported by the National Institute of Mental Health at the National Institutes of Health (grant numbers F32MH085433-01A1 to AC and 5R01MH075706 to JH) and the National Science Foundation (grant numbers NSF1129764 and NSF1607845 to JH).

ACKNOWLEDGMENTS

We thank Matteo Visconti di Oleggio Castello, Yaroslav O. Halchenko, Vassiki Chauhan, Easha Narayan, Kelsey G. Wheeler, Courtney Rogers, and Terry Sackett for helpful suggestions and administrative support. The computations in this work were performed on the Discovery cluster supported by the Research Computing group at Dartmouth College.

REFERENCES

Aine, C. J., Supek, S., George, J. S., Ranken, D., Lewine, J., Sanders, J., et al. (1996). Retinotopic organization of human visual cortex: departures from the classical model. *Cereb. Cortex* 6, 354–361. doi:10.1093/cercor/6.3.354

Behzadi, Y., Restom, K., Liau, J., and Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* 37, 90–101. doi:10.1016/j.neuroimage.2007.04.042

Bilenko, N. Y., and Gallant, J. L. (2016). Pyrcca: regularized kernel canonical correlation analysis in python and its applications to neuroimaging. *Front. Neuroinform.* 10:49. doi: 10.3389/fninf.2016.00049

Braga, R. M., and Buckner, R. L. (2017). Parallel interdigitated distributed networks within the individual estimated by intrinsic functional connectivity. *Neuron* 95, 457–471.e5. doi: 10.1016/j.neuron.2017.06.038

Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D., and Kriegeskorte, N. (2014). Unique semantic space in the brain of each beholder predicts perceived similarity. *Proc. Natl. Acad. Sci. U.S.A.* 111, 14565–14570. doi: 10.1073/pnas.1402594111

Chen, P.-H., Chen, J., Yeshurun, Y., Hasson, U., Haxby, J. V., and Ramadge, P. J. (2015). "A reduced-dimension fMRI shared response model," in *Advances in Neural Information Processing Systems 28*, eds C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett (Red Hook, NY: Curran Associates, Inc.), 460–468.

Chen, P.-H., Zhu, X., Zhang, H., Turek, J. S., Chen, J., Willke, T. L., et al. (2016). A convolutional autoencoder for multi-subject fMRI data aggregation. arXiv:1608.04846 [Preprint].

Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y.-C., et al. (2012). The representation of biological classes in the human brain. *J. Neurosci.* 32, 2608–2618. doi: 10.1523/JNEUROSCI.5547-11.2012

Connolly, A. C., Sha, L., Guntupalli, J. S., Oosterhof, N., Halchenko, Y. O., Nastase, S. A., et al. (2016). How the human brain represents perceived dangerousness or "predacity" of animals. *J. Neurosci.* 36, 5373–5384. doi: 10.1523/JNEUROSCI.3395-15.2016

Conroy, B. R., Singer, B. D., Guntupalli, J. S., Ramadge, P. J., and Haxby, J. V. (2013). Inter-subject alignment of human cortical anatomy using functional connectivity. *Neuroimage* 81, 400–411. doi: 10.1016/j.neuroimage.2013.05.009

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162–173. doi: 10.1006/cbmr.1996.0014

Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9, 179–194. doi: 10.1006/nimg.1998.0395

de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L., and Theunissen, F. E. (2017). The hierarchical cortical organization of human speech processing. *J. Neurosci.* 37, 6539–6557. doi: 10.1523/JNEUROSCI.3267-16.2017

Dubois, J., and Adolphs, R. (2016). Building a science of individual differences from fMRI. *Trends Cogn. Sci.* 20, 425–443. doi: 10.1016/j.tics.2016.03.014

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., et al. (2017). *poldracklab/fmriprep: 1.0.0-rc5*. Available at: https://zenodo.org/record/996169#

Felsen, G., and Dan, Y. (2005). A natural approach to studying vision. *Nat. Neurosci.* 8, 1643–1646. doi: 10.1038/nn1608

Fischl, B. (2012). FreeSurfer. *Neuroimage* 62, 774–781. doi: 10.1016/j.neuroimage.2012.01.021

Fischl, B., Sereno, M. I., Tootell, R. B., and Dale, A. M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum. Brain Mapp.* 8, 272–284. doi: 10.1002/(SICI)1097-0193(1999)8:4<272::AID-HBM10>3.0.CO;2-4

Frost, M. A., and Goebel, R. (2012). Measuring structural–functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. *Neuroimage* 59, 1369–1381. doi: 10.1016/j.neuroimage.2011.08.035

Gabrieli, J. D. E., Ghosh, S. S., and Whitfield-Gabrieli, S. (2015). Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron* 85, 11–26. doi: 10.1016/j.neuron.2014.10.047

Gordon, E. M., Laumann, T. O., Gilmore, A. W., Newbold, D. J., Greene, D. J., Berg, J. J., et al. (2017). Precision functional mapping of individual human brains. *Neuron* 95, 791.e–807.e. doi: 10.1016/j.neuron.2017.07.011

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., et al. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. *Front. Neuroinform.* 5:13. doi: 10.3389/fninf.2011.00013

Gorgolewski, K. J., Alfaro-Almagro, F., Auer, T., Bellec, P., Capotã, M., Chakravarty, M. M., et al. (2017). BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLoS Comput. Biol.* 13:e1005209. doi: 10.1371/journal.pcbi.1005209

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., et al. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* 3:160044. doi: 10.1038/sdata.2016.44

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika* 40, 33–51. doi: 10.1007/BF02291478

Greve, D. N., and Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* 48, 63–72. doi: 10.1016/j.neuroimage.2009.06.060

Güçlü, U., and van Gerven, M. A. J. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *J. Neurosci.* 35, 10005–10014. doi: 10.1523/JNEUROSCI.5023-14.2015

Güçlü, U., and van Gerven, M. A. J. (2017). Increasingly complex representations of natural movies across the dorsal stream are shared between subjects. *Neuroimage* 145, 329–336. doi: 10.1016/j.neuroimage.2015.12.036

Guntupalli, J. S., Hanke, M., Halchenko, Y. O., Connolly, A. C., Ramadge, P. J., and Haxby, J. V. (2016). A model of representational spaces in human cortex. *Cereb. Cortex* 26, 2919–2934. doi: 10.1093/cercor/bhw068

Guntupalli, J. S., Feilong, M., and Haxby, J. V. (2018). A computational model of shared fine-scale structure in the human connectome. *PLoS Comput. Biol.* 14:e1006120. doi: 10.1371/journal.pcbi.1006120

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V., and Pollmann, S. (2009). PyMVPA: a Python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics* 7, 37–53. doi: 10.1007/s12021-008-9041-y

Hasson, U., Malach, R., and Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends Cogn. Sci.* 14, 40–48. doi: 10.1016/j.tics.2009.10.011

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science* 303, 1634–1640. doi: 10.1126/science.1089506

Haxby, J. V., Connolly, A. C., and Guntupalli, J. S. (2014). Decoding neural representational spaces using multivariate pattern analysis. *Annu. Rev. Neurosci.* 37, 435–456. doi: 10.1146/annurev-neuro-062012-170325

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., et al. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* 72, 404–416. doi: 10.1016/j.neuron.2011.08.026

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., and Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* 532, 453–458. doi: 10.1038/nature17637

Huth, A. G., Nishimoto, S., Vu, A. T., and Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* 76, 1210–1224. doi: 10.1016/j.neuron.2012.10.014

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841. doi: 10.1006/nimg.2002.1132

Kay, K. N., Naselaris, T., Prenger, R. J., and Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature* 452, 352–355. doi: 10.1038/nature06713

Klein, A., Ghosh, S. S., Avants, B., Yeo, B. T. T., Fischl, B., Ardekani, B., et al. (2010). Evaluation of volume-based and surface-based brain image registration methods. *Neuroimage* 51, 214–220. doi: 10.1016/j.neuroimage.2010.01.091

Kriegeskorte, N., Goebel, R., and Bandettini, P. (2006). Information-based functional brain mapping. *Proc. Natl. Acad. Sci. U.S.A.* 103, 3863–3868. doi: 10.1073/pnas.0600244103

Kriegeskorte, N., and Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn. Sci.* 17, 401–412. doi: 10.1016/j.tics.2013.06.007

Laumann, T. O., Gordon, E. M., Adeyemo, B., Snyder, A. Z., Joo, S. J., Chen, M.-Y., et al. (2015). Functional system and areal organization of a highly sampled individual human brain. *Neuron* 87, 657–670. doi: 10.1016/j.neuron.2015.06.037

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems 26*, eds C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Red Hook, NY: Curran Associates), 3111–3119.

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M., Malave, V. L., Mason, R. A., et al. (2008). Predicting human brain activity associated with the meanings of nouns. *Science* 320, 1191–1195. doi: 10.1126/science.1152876

Naselaris, T., Kay, K. N., Nishimoto, S., and Gallant, J. L. (2011). Encoding and decoding in fMRI. *Neuroimage* 56, 400–410. doi: 10.1016/j.neuroimage.2010.07.073

Nastase, S. A., Connolly, A. C., Oosterhof, N. N., Halchenko, Y. O., Guntupalli, J. S., Visconti di Oleggio Castello, M., et al. (2017). Attention selectively reshapes the geometry of distributed semantic representation. *Cereb. Cortex* 27, 4277–4291. doi: 10.1093/cercor/bhx138

Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., and Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Curr. Biol.* 21, 1641–1646. doi: 10.1016/j.cub.2011.08.031

Oosterhof, N. N., Tipper, S. P., and Downing, P. E. (2013). Crossmodal and action-specific: neuroimaging the human mirror neuron system. *Trends Cogn. Sci.* 17, 311–318. doi: 10.1016/j.tics.2013.04.012

Oosterhof, N. N., Wiestler, T., Downing, P. E., and Diedrichsen, J. (2011). A comparison of volume-based and surface-based multi-voxel pattern analysis. *Neuroimage* 56, 593–600. doi: 10.1016/j.neuroimage.2010.04.270

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *J. Neurosci. Methods* 162, 8–13. doi: 10.1016/j.jneumeth.2006.11.017

Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S. J., Kanwisher, N., et al. (2018). Toward a universal decoder of linguistic meaning from brain activation. *Nat. Commun.* 9:963. doi: 10.1038/s41467-018-03068-4

Poldrack, R. A. (2017). Precision neuroscience: dense sampling of individual brains. *Neuron* 95, 727–729. doi: 10.1016/j.neuron.2017.08.002

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI

networks arise from subject motion. *Neuroimage* 59, 2142–2154. doi: 10.1016/j.neuroimage.2011.10.018

Riddle, D. R., and Purves, D. (1995). Individual variation and lateral asymmetry of the rat primary somatosensory cortex. *J. Neurosci.* 15, 4184–4195. doi: 10.1523/JNEUROSCI.15-06-04184.1995

Saad, Z. S., Reynolds, R. C., Argall, B., Japee, S., and Cox, R. W. (2004). "SUMA: an interface for surface-based intra-and inter-subject analysis with AFNI," in *Proceedings of the 2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE)*, Arlington, VA, 1510–1513. doi: 10.1109/ISBI.2004.1398837

Sabuncu, M. R., Singer, B. D., Conroy, B., Bryan, R. E., Ramadge, P. J., and Haxby, J. V. (2010). Function-based intersubject alignment of human cortical anatomy. *Cereb. Cortex* 20, 130–140. doi: 10.1093/cercor/bhp085

Santoro, R., Moerel, M., De Martino, F., Goebel, R., Ugurbil, K., Yacoub, E., et al. (2014). Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. *PLoS Comput. Biol.* 10:e1003412. doi: 10.1371/journal.pcbi.1003412

Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., et al. (2015). The animacy continuum in the human ventral vision pathway. *J. Cogn. Neurosci.* 27, 665–678. doi: 10.1162/jocn_a_00733

Tran, D., Bourdev, L., Fergus, R., Torresani, L., and Paluri, M. (2015). "Learning spatiotemporal features with 3D convolutional networks," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV) (IEEE)*, Santiago, Chile, 4489–4497. doi: 10.1109/ICCV.2015.510

Vanderwal, T., Eilbott, J., Finn, E. S., Craddock, R. C., Turnbull, A., and Castellanos, F. X. (2017). Individual differences in functional connectivity during naturalistic viewing conditions. *Neuroimage* 157, 521–530. doi: 10.1016/j.neuroimage.2017.06.027

Vodrahalli, K., Chen, P.-H., Liang, Y., Baldassano, C., Chen, J., Yong, E., et al. (2017). Mapping between fMRI responses to movies and their natural language annotations. *Neuroimage* doi: 10.1016/j.neuroimage.2017.06.042 [Epub ahead of print].

Watson, J. D., Myers, R., Frackowiak, R. S., Hajnal, J. V., Woods, R. P., Mazziotta, J. C., et al. (1993). Area V5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging. *Cereb. Cortex* 3, 79–94. doi: 10.1093/cercor/3.2.79

Wehbe, L., Murphy, B., Talukdar, P., Fyshe, A., Ramdas, A., and Mitchell, T. (2014). Simultaneously uncovering the patterns of brain regions involved in different story reading subprocesses. *PLoS One* 9:e112575. doi: 10.1371/journal.pone.0112575

Wen, H., Shi, J., Chen, W., and Liu, Z. (2018). Transferring and generalizing deep-learning-based neural encoding models across subjects. *Neuroimage* 176, 152–163. doi: 10.1016/j.neuroimage.2018.04.053

Wurm, M. F., Ariani, G., Greenlee, M. W., and Lingnau, A. (2016). Decoding concrete and abstract action representations during explicit and implicit conceptual processing. *Cereb. Cortex* 29, 3390–3401. doi: 10.1093/cercor/bhv169

Wurm, M. F., and Lingnau, A. (2015). Decoding actions at different levels of abstraction. *J. Neurosci.* 35, 7727–7735. doi: 10.1523/JNEUROSCI.0188-15.2015

Xu, H., Lorbert, A., Ramadge, P. J., Guntupalli, J. S., and Haxby, J. V. (2012). "Regularized hyperalignment of multi-set fMRI data," in *Proceedings of the 2012 IEEE Statistical Signal Processing Workshop (SSP) (IEEE)*, Ann Arbor, MI, 229–232. doi: 10.1109/SSP.2012.6319668

Yamada, K., Miyawaki, Y., and Kamitani, Y. (2015). Inter-subject neural code converter for visual image representation. *Neuroimage* 113, 289–297. doi: 10.1016/j.neuroimage.2015.03.059

Yousefnezhad, M., and Zhang, D. (2017). "Deep hyperalignment," in *Advances in Neural Information Processing Systems 30*, eds I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, et al. (Red Hook, NY: Curran Associates, Inc.), 1604–1612.

Zhen, Z., Kong, X.-Z., Huang, L., Yang, Z., Wang, X., Hao, X., et al. (2017). Quantifying the variability of scene-selective regions: interindividual, interhemispheric, and sex differences. *Hum. Brain Mapp.* 38, 2260–2275. doi: 10.1002/hbm.23519

Zhen, Z., Yang, Z., Huang, L., Kong, X.-Z., Wang, X., Dang, X., et al. (2015). Quantifying interindividual variability and asymmetry of face-selective regions: a probabilistic functional atlas. *Neuroimage* 113, 13–25. doi: 10.1016/j.neuroimage.2015.03.010

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.