NeuroImage 216 (2020) 116458

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Predicting individual face-selective topography using naturalistic stimuli

Guo Jiahui [a], Ma Feilong [a], Matteo Visconti di Oleggio Castello [b], J. Swaroop Guntupalli [c], Vassiki Chauhan [a], James V. Haxby [a,**], M. Ida Gobbini [d,e,*]

[a] Center for Cognitive Neuroscience, Dartmouth College, NH, USA
[b] Helen Wills Neuroscience Institute, University of California, Berkeley, CA, USA
[c] Vicarious AI, Union City, CA, USA
[d] Cognitive Science, Dartmouth College, NH, USA
[e] Dipartimento di Medicina Specialistica, Diagnostica e Sperimentale, Università di Bologna, Bologna, 40126, Italy

ARTICLE INFO

Keywords:
Hyperalignment
Functional topography
Faces
Naturalistic stimuli
Localizer

ABSTRACT

Subject-specific, functionally defined areas are conventionally estimated with functional localizers and a simple contrast analysis between responses to different stimulus categories. Compared with functional localizers, naturalistic stimuli provide several advantages such as stronger and widespread brain activation, greater engagement, and increased subject compliance. In this study we demonstrate that a subject's idiosyncratic functional topography can be estimated with high fidelity from that subject's fMRI data obtained while watching a naturalistic movie using hyperalignment to project other subjects' localizer data into that subject's idiosyncratic cortical anatomy. These findings lay the foundation for developing an efficient tool for mapping functional topographies for a wide range of perceptual and cognitive functions in new subjects based only on fMRI data collected while watching an engaging, naturalistic stimulus and other subjects' localizer data from a normative sample.

1. Introduction

The topographies of category-selective areas are mostly distributed similarly across individuals, but great individual variability exists in the locus, the size, and the shape of the category-selective areas (Zhen et al., 2015, 2017). For example, faces selectively activate areas in the lateral occipital cortex (the occipital face area, OFA), ventral temporal cortex (the fusiform face area, FFA, and the anterior temporal face area, ATFA), along the superior temporal sulcus (the posterior and anterior superior temporal face areas, pSTS and aSTS), and in lateral prefrontal cortex (the right inferior frontal face area, rIFFA) (Guntupalli et al., 2017; Haxby and Gobbini, 2011; Haxby et al., 1994, 2000; Visconti di Oleggio Castello et al., 2017). To deal with the idiosyncratic topography of functional areas, category-selective areas are identified separately in each individual using a "functional localizer" fMRI scan. Functional localizers use a simple contrast between responses to different categories of stimuli, such as responses to faces versus responses to objects, to identify category-selective areas or to map a category-selectivity topography (Saxe et al., 2006).

We have shown that idiosyncratic topographies for category-selectivity and retinotopy can be estimated in individual brains with high fidelity using hyperalignment to project other subjects' functional localizer data into a target subject's ventral temporal and occipital cortical anatomy (Guntupalli et al., 2016; Haxby et al., 2011). Hyperalignment derives individual transformation matrices to project information encoded in idiosyncratic topographies into a common model information space. These matrices are derived based either on responses to a naturalistic stimulus, such as a movie, or on functional connectivity (Guntupalli et al., 2018). Our findings show that individually-tailored maps estimated from other subjects' data after hyperalignment correlate much more highly with maps estimated from that subject's own localizer data than does a group average map based on anatomical normalization.

There are many potential advantages to using data collected during movie-viewing for estimating category-selective topographies. Movies are more engaging and result in better compliance (Vanderwal et al., 2015). Movie viewing can also be used in subject populations, such as children (Richardson et al., 2018) or patients, that may have trouble maintaining attention during repetitions of a tedious localizer task. Movies engage multiple brain systems in parallel. From a single movie

* Corresponding author. Cognitive Science, Dartmouth College, NH, USA.
** Corresponding author.
E-mail address: mariaida.gobbini@unibo.it (M.I. Gobbini).

https://doi.org/10.1016/j.neuroimage.2019.116458
Received 10 August 2019; Received in revised form 16 November 2019; Accepted 9 December 2019
Available online 13 December 2019

dataset multiple functional topographies can be estimated (Guntupalli et al., 2016), whereas different localizers are typically required to map different functional topographies, making a thorough mapping of selective topographies time-consuming and inefficient. Movies also simulate better the statistics of natural viewing and listening and may provide more ecologically valid maps. Analogously, the introduction of dynamic videos of faces and control categories to localize face-selective topographies provides more reliable maps and better estimate the extent of face-selective regions than do localizers with still image stimuli (Fox et al., 2009; Pitcher et al., 2011). Similarly, naturalistic stimuli may better sample the full range of responses to faces and other stimuli that contribute to face-selective topographies.

Here, we show that precise mapping of functional topographies in a new subject can be achieved using hyperalignment and a database of movie and localizer data from other subjects. We present a proof-of-concept analysis of two different data sets with different movies and different face-selectivity localizers. We used an optimized hyperalignment procedure for this application that directly projects all other subjects' data into a target subject's cortical anatomy (one-step algorithm) without the intermediate step of projecting individual data into a group common information space and then projecting data from the common space into a new subject's cortical anatomy (two-step algorithm). The results replicate our earlier findings with both datasets, expanding previous analysis from ventral temporal cortex to the whole cortex, showing strong correlations of face-selectivity topographic maps derived from a subject's own localizer data with maps derived from other subjects' localizer data projected into that subject's cortical anatomy. Both the two-step algorithm and the new one-step algorithm, which we introduce here, produce high-fidelity, individualized topographic maps, but the new one-step algorithm maps were superior.

These results lay a foundation for building a computational tool with a database that could allow others to map multiple functional topographies in new subjects using only data collected during movie viewing. Functional localizers are inefficient because they only estimate one or a few topographies for each localizer. Movies, by contrast, engage in parallel multiple neural systems for vision, audition, language, person perception, social cognition, and other functions. Consequently, movies have the potential to estimate selective topographies in all of these domains. Such a tool would require a database of data for movies and a range of functional localizers in a normative group of subjects. A new subject's functional topographies could be estimated based only on that subject's movie data and other subjects' localizer data from the normative database that could be projected into that subject's cortical anatomy using hyperalignment transformation matrices derived from movie data. Such a resource would be more efficient and replace tedious functional localizers with an engaging movie and could enable mapping of multiple functional topographies with data from a single fMRI using a naturalistic stimulus.

2. Materials and methods

2.1. Participants

2.1.1. StudyForrest

This dataset is publicly available at http://www.studyforrest.org/(Hanke et al., 2014). Fifteen adults (mean age 29.4 years, range 21–39, 6 females) were recruited. Data were collected at the Otto-von-Guericke University in Germany and the native language of all participants was German (Hanke et al., 2016; Sengupta et al., 2016).

2.1.2. Grand Budapest Hotel

Twenty-one student participants (mean age 27.3 years, range 22–31, 11 females) were recruited at Dartmouth. All participants had normal hearing and normal or corrected-to-normal vision, and no known history of neurological illness. The study was approved by the Dartmouth Committee for the Protection of Human Subjects.

2.2. Naturalistic movie watching and localizers

2.2.1. StudyForrest

Participants watched the audio-visual feature movie Forrest Gump while fMRI data were collected (Hanke et al., 2016). The Forrest Gump movie was divided into eight parts with each part lasting approximately 15 min.

A four-run block-design static localizer was included in the study (Sengupta et al., 2016). There were six stimulus categories: human faces, human bodies without heads, small objects, houses and outdoor scenes comprised of nature and street scenes, and phase scrambled images. Each category had 24 Gray-scale images and their mirrored views. Each block had 16 images (900 ms display + 100 ms intertrial interval each) and lasted 16 s. Participants were asked to press a button when they saw a repetition of an image to maintain attention. Each category was repeated twice in each of four 312 s runs, for a total of 20′48″ of scan time.

2.2.2. Grand Budapest Hotel

The full-length of the Grand Budapest Hotel movie was divided into six parts. Parts were divided at scene changes to keep the narrative of the movie intact. Participants watched the first part of the movie (~45 min) outside the scanner. Immediately thereafter, participants watched the remaining five parts of the movie in the scanner (~50 min, each part lasting 9–13 min each) with audio.

In a separate scanning section, participants performed a dynamic localizer task (Pitcher et al., 2011). Participants watched 3 s clips of faces, bodies, scenes, objects, and scrambled objects. The clips were presented continuously in 18 s blocks of each category, without blank periods between blocks. Participants were required to press a button whenever they saw a repetition of a clip. The blocks followed this order: an 18 s fixation block, five blocks of different categories (each lasting 18 s) in random order, an 18 s fixation block, five blocks of the categories in reversed order, and a final 18 s fixation block. Thus, if the order in the first part of the run was B-S-F-O-Sc, in the second part would be Sc-O-F-S-B (Pitcher et al., 2011). Four 234 s runs were collected for a total of 15′44″ of scan time.

2.3. MRI data acquisition

2.3.1. StudyForrest

Scanning was done with a whole-body 3 T Philips Achieva dStream MRI scanner equipped with a 32 channel head coil. Data were collected with gradient-echo, 2 s repetition time (TR), 30 ms echo time, 90° flip angle, 1943 Hz/px bandwidth, and parallel acquisition with sensitivity encoding (SENSE) reduction factor 2. Each volume comprised 35 axial slices with anterior-to-posterior phase-encoding direction that were collected in ascending order, which mostly covered the entire brain. Each slice was 3.0 mm thick with a 10% inter-slice gap, and had a 240 mm × 240 mm field-of-view comprising 80 × 80 3 mm isotropic voxels. Please see Hanke et al. (2016) and Sengupta et al. (2016) for more detailed MRI data acquisition parameters.

2.3.2. Grand Budapest Hotel

All scans in the Grand Budapest Hotel dataset were acquired using a 3 T S Magnetom Prisma MRI scanner with a 32 channel head coil at the Dartmouth Brain Imaging Center. CaseForge headcases (https://caseforge.co/) were used to minimize head motion. BOLD images were acquired in an interleaved fashion using gradient-echo echo-planar imaging with pre-scan normalization, fat suppression, multiband (i.e., simultaneous multi-slice; SMS) acceleration factor of 4 (using blipped CAIPIRINHA), and no in-plane acceleration (i.e., GRAPPA acceleration factor of one): TR/TE = 1000/33 ms, flip angle = 59°, resolution = 2.5 mm^3 isotropic voxels, matrix size = 96 × 96, FoV = 240 × 240 mm, 52 axial slices with full brain coverage and no gap, anterior–posterior phase encoding. At the beginning of each run, three dummy scans were acquired to allow for signal stabilization. The T1-weighted structural scan

was acquired using a high-resolution single-shot MPRAGE sequence with an in-plane acceleration factor of 2 using GRAPPA: TR/TE/TI = 2300/2.32/933 ms, flip angle = 8°, resolution = 0.9375 × 0.9375 × 0.9 mm voxels, matrix size = 256 × 256, FoV = 240 × 240 × 172.8 mm, 192 sagittal slices, ascending acquisition, anterior–posterior phase encoding, no fat suppression, and with 5 min 21 s total acquisition time. A T2-weighted structural scan was acquired with an in-plane acceleration factor of 2 using GRAPPA: TR/TE = 3200/563 ms, flip angle = 120°, resolution = 0.9375 × 0.9375 × 0.9 mm voxels, matrix size = 256 × 256, FoV = 240 × 240 × 172.8 mm, 192 sagittal slices, ascending acquisition, anterior–posterior phase encoding, no fat suppression, and lasted for 3 min 21 s. At the beginning of each session (movie and localizer), a fieldmap scan was collected for distortion correction.

2.4. Data analysis

2.4.1. Preprocessing

2.4.1.1. StudyForrest. We started with cortical surfaces reconstructed by FreeSurfer, available at https://github.com/psychoinformatics-de/studyforrest-data-freesurfer, and fMRI data pre-aligned to a subject-specific template by FSL MCFLIRT available at https://github.com/psychoinformatics-de/studyforrest-data-aligned, which were provided by the studyforrest group (Hanke et al., 2014, 2016; Sengupta et al., 2016). We computed mapping between subject-specific templates and cortical surfaces using boundary-based registration (Greve and Fischl, 2009), which was used to project these pre-aligned fMRI data to the fsaverage template (Fischl et al., 1999), such that these data were aligned across individuals anatomically based on cortical folding patterns.

Further preprocessing were performed using PyMVPA (Hanke et al., 2009; http://www.pymvpa.org/) to resample data to a standard cortical mesh and to remove noise from fMRI data using linear regression. The cortical mesh had 18,742 vertices across both hemispheres (approximately 3 mm vertex spacing; 20,484 vertices before removing non-cortical vertices). The nuisance regressors included 6 motion parameters and their derivatives, 5 principal components from cerebrospinal fluid and white matter (https://www.zotero.org/google-docs/?q9b9WEaCompCor; Behzadi et al., 2007), and polynomial trends up to second order.

2.4.1.2. Grand Budapest Hotel. MRI data were preprocessed using the fMRIPrep software version 1.4.1 (Esteban et al., 2019). T1-weighted images were corrected for intensity non-uniformity (Tustison et al., 2010) and skullstripped using antsBrainExtraction.sh. High resolution cortical surfaces were reconstructed with FreeSurfer (Fischl, 2012) using both T1-weighted and T2-weighted images, and then registered to the fsaverage template (Fischl et al., 1999). Functional data was slice-time corrected using 3dTshift (Cox, 1996), motion corrected using MCFLIRT (Jenkinson et al., 2002), distortion corrected using fieldmap estimate scans (one for each session), and then resampled to the fsaverage template based on boundary-based registration (Greve and Fischl, 2009). After these steps, functional data were in alignment with the fsaverage template based on cortical folding patterns.

The data were further preprocessed with Python scripts. Six motion parameters and their derivatives, global signal, framewise displacement (Power et al., 2014), 6 principal components from cerebrospinal fluid and white matter (https://www.zotero.org/google-docs/?ORzJWDaCompCor; Behzadi et al., 2007), and polynomial trends up to second order were regressed out from both movie and localizer data for each run independently.

2.5. Searchlight hyperalignment

We optimized the hyperalignment algorithm (Feilong et al., 2018; Guntupalli et al., 2016, 2018; Haxby et al., 2011) to predict functional topographies in new subjects. Our new 1-step algorithm, unlike our previous 2-step method, uses searchlight hyperalignment to directly transform data from one participant's cortical anatomy into another participant's cortical anatomy, without projecting it through a common space. In each 15 mm searchlight a given participant's response pattern vectors for all time points (TRs) are aligned to the target participant response patterns using the Procrustes transformation. The resulting searchlight transformation matrices are then aggregated into a single transformation matrix for each hemisphere for each pair of participants (Guntupalli et al., 2016). Our new 1-step algorithm requires the estimation of a transformation matrix for each pair of participants, thus we computed 105 (15 × 14 ÷ 2) and 210 (21 × 20 ÷ 2) transformation matrices for the fifteen and the twenty-one participants in each study, respectively. To estimate one participant's face-selectivity contrast map (participant A), all the other participants' localizer data were projected directly into participant A's cortical anatomy using transformation matrices calculated based only on the movie-viewing data for all pairings of participant A with each of the other 14 or 20 participants.

Our previously described hyperalignment algorithm (2-step method, Feilong et al., 2018; Guntupalli et al., 2016, 2018; Haxby et al., 2011) builds a common model information space where patterns of fMRI responses to a movie are aligned across subjects. Individual, whole-cortex transformation matrices are calculated using a searchlight-based algorithm to project each participant's cortical space into the common information space. Transformation matrices are calculated for all 15 mm searchlights in each brain using an iterative procedure and Procrustes alignment, and then aggregated into a single matrix for each hemisphere. New data can be projected into any individual brain space by first projecting data from other brains into the common model space, and then projecting those data from the common model space into that participant's cortical anatomy using the transpose of his or her transformation matrix. The reader is referred to the original papers for details (Feilong et al., 2018; Guntupalli et al., 2016, 2018; Haxby et al., 2011). To estimate one participant's (participant A) face-selectivity map using this algorithm, all the other participants' localizer data were projected into the common model space, and then mapped to participant A's space using the transpose of his or her hyperalignment transformation matrix.

2.6. Predicting individual contrast map

We estimated each participant's face-selectivity map based on that participant's localizer data and based on other participants' localizer data projected into that participant's cortical anatomy using hyperalignment and anatomical surface alignment (see Fig. 1). We then calculated correlations between the map based on a participant's own data and the maps estimated from other participants' data and compared these correlations to the reliability of the participant's map, indexed with Cronbach's Alpha. We estimated the face-selectivity map for a participant from his or her own localizer data by calculating the GLM univariate contrast map of faces vs. all the other categories (e.g., body, place) for each run and averaging t-values across the four maps. We estimated a participant's map from other participants' data by first projecting all other participants' localizer data into that participant's cortical anatomy and calculating the GLM univariate contrast map of faces versus all other categories for each run in each other participant then averaging t-values across the maps (56 maps for StudyForrest, 14 subjects x 4 runs for each; 80 maps for Grand Budapest, 20 subjects x 4 runs each). We projected other participants' localizer data into each individual participant's cortical anatomy using hyperalignment with either direct projection or via the common model space, as described above. We also estimated each participant's face-selectivity map based on other participants' anatomically aligned localizer data by simply averaging across the maps based on anatomically-normalized data for each N-1 set of 14 (StudyForrest) or 20 (Grand Budapest) participants. Thus, the map estimated from other participants' localizer data was a map of average t-values across 56 or 80 maps (four runs per participant) for the contrast faces vs all. The analysis

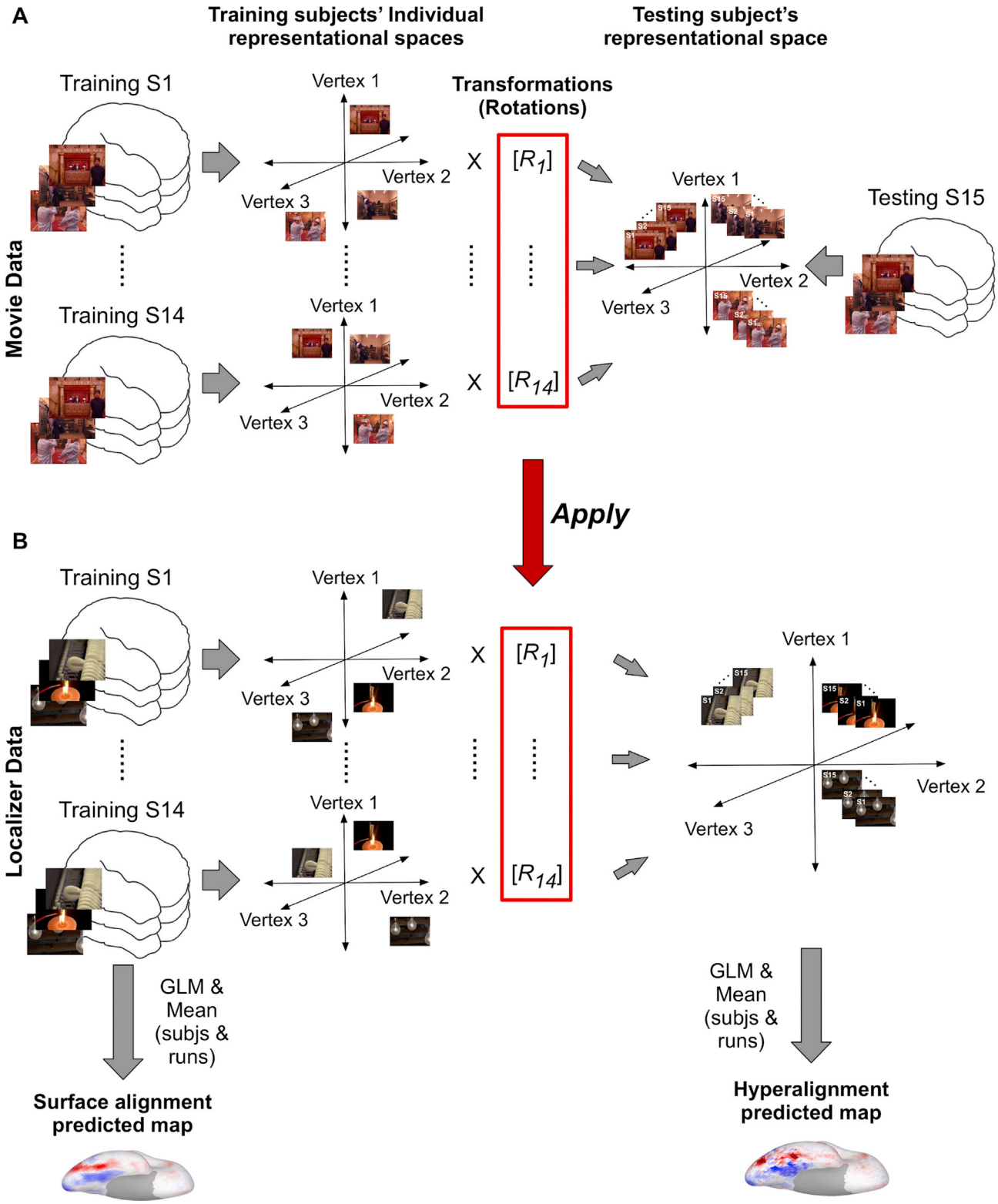

Fig. 1. Schematic of data analysis procedures. A. Transformation matrices were calculated by hyperaligning training participants' movie data to the left out testing participant's representational space. **B.** Transformation matrix for each training participant was applied to the localizer runs. The surface alignment predicted map was calculated by calculating the contrast map in each training participant and then averaging them across all the training participants. The hyperalignment predicted map was calculated by doing the same steps as above but using training participants' hyperaligned localizer time series.

pipeline was the same across conditions, and the only difference was whether the N-1 participants' data were aligned to the left-out participant's space using anatomical alignment, hyperalignment with direct projection (1-step), or hyperalignment via the common model space (2-step). We tested the quality of the maps estimated from other participants' data by calculating the correlation of each with the map estimated from a participant's own data. We also gauged the reliability of the estimates based on participants' own data by calculating Cronbach's alpha

based on variability across runs. To examine local performance and reliability, we did the same analyses with correlations between maps and Cronbach's Alpha in a searchlight analysis with a radius of 15 mm. To test the applicability of our procedure to other category-selective topographies, we extended our analyses to other contrasts (scenes, bodies, and objects).

3. Results

We correlated the whole-cortex contrast map (faces-vs-all) based on a participant's own data with the maps estimated from other participants' data separately for the StudyForrest and Grand Budapest datasets. After 1-step hyperalignment, the mean Pearson correlation values across participants were 0.58 (StudyForrest, N = 15, S.D. = 0.08) and 0.69 (Grand Budapest, N = 21, S.D. = 0.08) (Fig. 2).

Hyperalignment greatly improved the prediction performance compared with anatomical surface alignment. With surface alignment, the average Pearson correlation values across participants were 0.40 (N = 15, S.D. = 0.08) and 0.50 (N = 21, S.D. = 0.06) in the StudyForrest and Grand Budapest datasets, respectively (Fig. 3). The difference between the hyperaligned and the surface-aligned mean correlation values was highly significant (StudyForrest: $t(14) = 17.39$, $p < 0.001$; Grand Budapest: $t(20) = 25.49$, $p < 0.001$).

We also estimated each participant's topography with the 2-step hyperalignment algorithm that projects other participants' localizer data into a common model information space then uses the transpose of the target participant's transformation matrix to project other participants' data from the common model space into his or her cortical anatomy. With this 2-step method, we found similar but slightly weaker results. The mean Pearson correlation values between face-selectivity maps based on a participant's own localizer data and the predicted map after hyperalignment were 0.55 (StudyForrest, N = 15, S.D. = 0.08) and 0.67 (Grand Budapest, N = 21, S.D. = 0.08). These correlations after hyperalignment were significantly better than the correlations after surface alignment (StudyForrest: $t(14) = 15.78$, $p < 0.001$; Grand Budapest: $t(20) = 16.61$, $p < 0.001$), but the estimates with the 1-step procedure were significantly better than estimates with the 2-step procedure (StudyForrest: $t(14) = 11.30$, $p < 0.001$; Grand Budapest: $t(20) = 6.98$, $p < 0.001$).

We compared the correlations between maps estimated from a participant's own data and maps estimated from other participants' data to the reliability of the localizer. We computed the reliability of the contrast maps with Cronbach's Alpha based on variability across the four localizer runs for each set. The mean Cronbach's Alpha between the four localizer runs was 0.60 (N = 15, S.D. = 0.14) in the StudyForrest dataset and was 0.74 (N = 21, S.D. = 0.09) in the Grand Budapest dataset (Fig. 2). These results mean that if we scan each participant for another 4 localizer runs, and compute the correlation between the two maps (4 runs vs. 4 runs), the correlation would be 0.60 and 0.74 on average in the StudyForrest and Grand Budapest datasets, respectively. The dynamic localizer (in Grand Budapest data set) was significantly more reliable than the static localizer (in StudyForrest data set) ($t(34) = 3.76$, $p < 0.001$) despite its

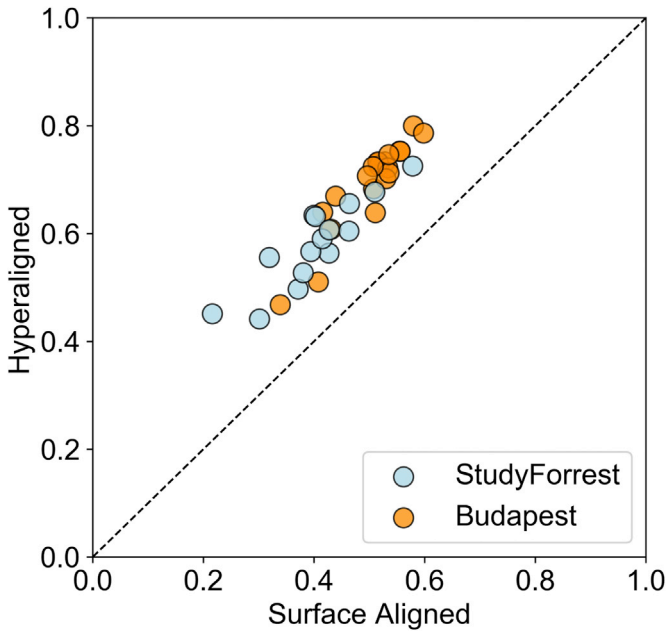

Fig. 3. Correlation values after hyperalignment and surface alignment in individual participants. The Pearson correlation values between face selectivity maps estimated from a participant's own data and data projected into that participant's cortical anatomy using anatomical surface alignment (x-axis) and hyperalignment (y-axis, 1-step algorithm). Individuals in the StudyForrest dataset are indicated with blue dots and those in the Grand Budapest dataset with orange dots. Hyperalignment greatly improved performance for all participants individually.

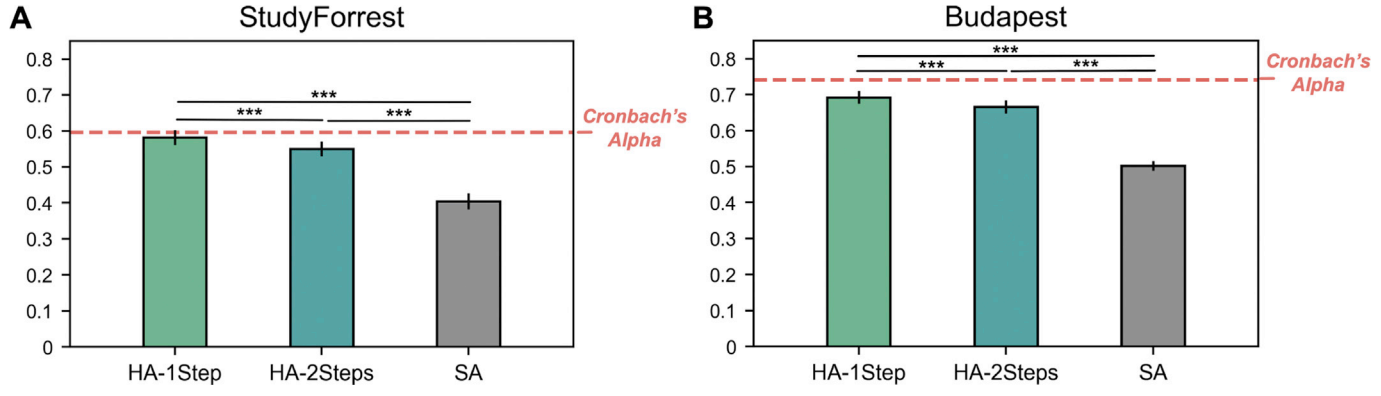

Fig. 2. Cronbach's alpha, mean correlation values after hyperalignment and surface alignment.
The mean Pearson correlation values across participants with 1-step (light green bars), 2-step (dark green bars) hyperalignment methods, and with surface alignment method (gray bar) were plotted for the StudyForrest (Panel **A**) and Grand Budapest dataset (Panel **B**) respectively. Cronbach's Alpha of total four localizer runs was plotted as the red dotted line in both panels. In both datasets, 1-step and 2-step hyperalignment significantly improved the prediction accuracy (1-step: StudyForrest: $t(14) = 17.39$, $p < 0.001$; Grand Budapest: $t(20) = 25.49$, $p < 0.001$; 2-step: StudyForrest: $t(14) = 15.78$, $p < 0.001$; Grand Budapest: $t(20) = 16.61$, $p < 0.001$). The 1-step method had significantly better prediction performance than the 2-step method (StudyForrest: $t(14) = 11.30$, $p < 0.001$; Grand Budapest: $t(20) = 6.98$, $p < 0.001$). The difference between Cronbach's alpha and the mean correlation after 1-step or 2-step hyperalignment was not significant in the StudyForrest dataset (1-step: $t(14) = 0.61$, $p = 0.55$; 2-step: $t(14) = 1.99$, $p = 0.07$). But the differences were significant in the Grand Budapest dataset (1-step: $t(20) = 3.02$, $p = 0.01$; 2-steps: $t(20) = 5.07$, $p < 0.001$). HA: hyperalignment, SA: surface alignment. ***$p < 0.001$.

shorter length (four 234s runs versus four 312s runs, respectively). Cronbach's alpha indicates that the predicted contrast map based on hyperalignment is close to or as good as the real contrast map based on four localizer runs (StudyForrest: $t(14) = 0.61$, $p = 0.55$; Grand Budapest: $t(20) = -3.02$, $p = 0.007$). In the StudyForrest dataset, the predicted contrast map based on hyperalignment was better than the contrast map based on data from three out of four localizer runs in other participants ($t(14) = 2.36$, $p = 0.03$) and in Grand Budapest, the predicted contrast map was comparable to the contrast map based on three localizer runs in other participants ($t(20) = 0.48$, $p = 0.63$).

A scatterplot of the individual correlation values with hyperalignment and with surface alignment (fig 3), shows that predicted maps based on hyperalignment were more accurate than those based on surface alignment in every participant. To further demonstrate how hyperalignment increases individual prediction performance, we further plotted the predicted contrast maps with hyperalignment and those with surface alignment on inflated surface for each participant, together with their measured contrast maps based on localizers (Figs. 4 & 5 for maps of four participants and Supplementary Figs. 1 and 2 for all individual maps). It is striking that hyperalignment was able to recover idiosyncratic topographies (e.g., the location, shape, and size of clusters) in great detail. By contrast, predicted maps based on anatomical alignment were smoother and did not reveal idiosyncratic functional topographies as they are essentially the same for all N-1 groupings of 14 or 20 participants (Fig. 5).

In a searchlight analysis we calculated local correlations between maps estimated from each participant's own data and maps estimated from other participants' data. In searchlights that included strongly face-selective cortical fields (e.g. lateral occipital, ventral temporal, superior temporal sulcus, right lateral prefrontal) mean correlations with estimates based on hyperaligned data exceeded 0.8 (Fig. 6). The lower mean correlations across the whole cortex reflects the contribution of low correlations in cortical areas with low face-selectivity, e.g. sensorimotor and anterior prefrontal cortex.

In supplementary analyses (Supplementary Figs. S3, S4, S5, and S6) we applied the same procedure to estimate category-selective topographies for scenes, bodies, and objects. The results show that these functional topographies also can be estimated from other participants' hyperaligned data with high fidelity.

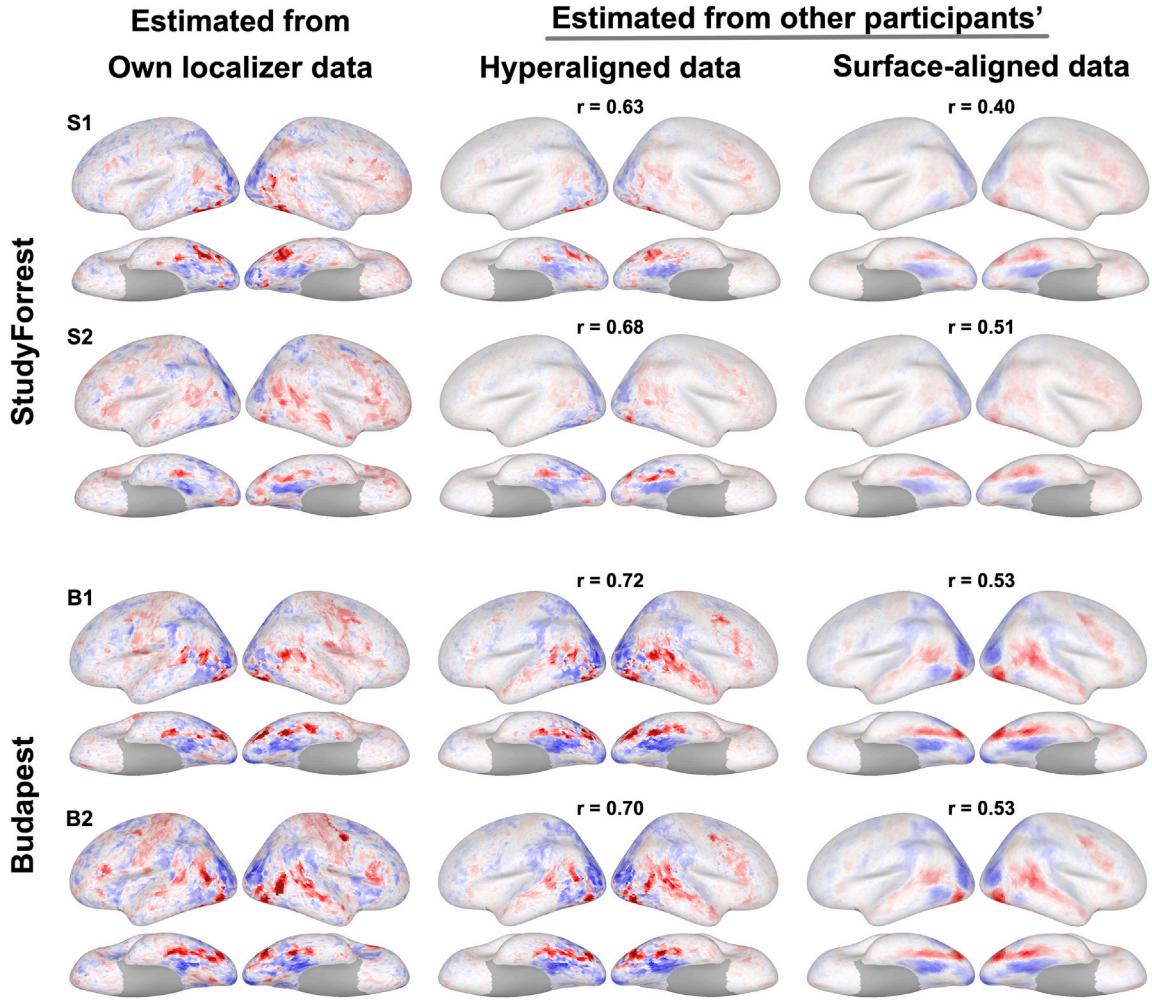

Fig. 4. Contrast maps of sample participants. The whole-brain contrast map for faces-vs-all calculated with four individuals' own localizer runs were plotted on their cortical surfaces (leftmost column) in StudyForrest dataset (upper two rows, subjects S1 & S2) and in Grand Budapest dataset (lower two rows, subjects B1 & B2), respectively. The faces-vs-all contrast identified areas that responded selectively to faces in the lateral temporal cortex, in the ventral temporal cortex, along the STS, and in the lateral prefrontal cortex. The middle column presents the estimated maps from other participants' localizer data projected into target participants' cortical anatomies with hyperalignment. Contrast maps estimated with surface alignment are in the right column. The individual Pearson correlation values between the maps from participants' own data and the predicted maps with hyperalignment were 0.63, 0.68 (StudyForrest), and were 0.72, 0.70 (Grand Budapest). The individual Pearson correlation values between the maps from participants' own data and the predicted maps with surface alignment were 0.40, 0.51 (StudyForrest), and were 0.53, 0.53 (Grand Budapest).

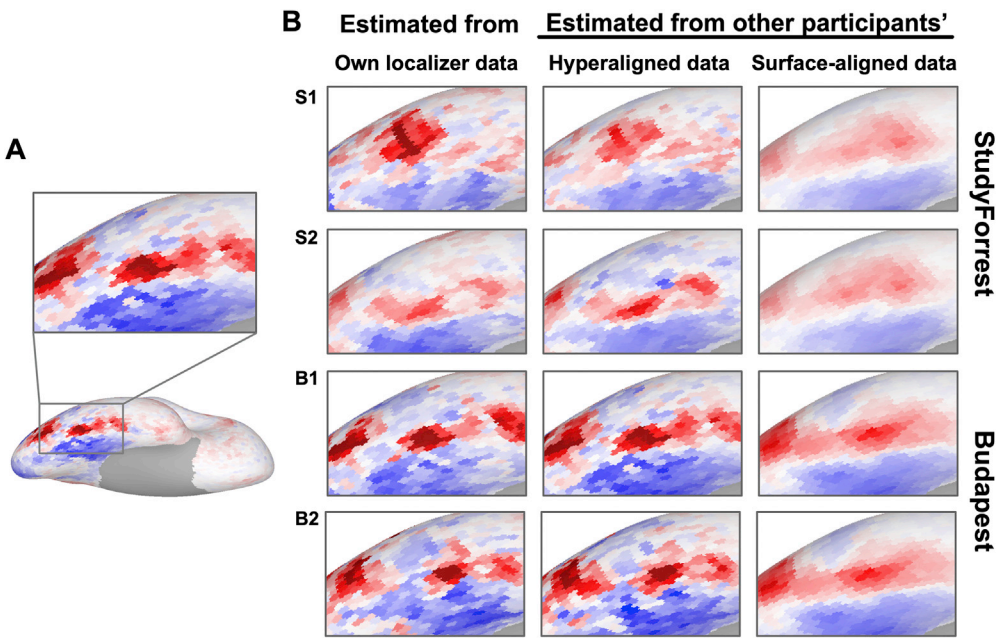

Fig. 5. Enlarged sample contrast maps of ventral temporal cortex. A. Part of the right ventral temporal cortex of one individual's face-vs-all topography estimated from other participants' hyperaligned data is enlarged. This panel shows how the enlarged images were created in panel B. **B.** Enlargement of the same individuals' face-vs-all topographies as in Fig. 4 calculated from their own localizer data, from other participants' 1-step hyperaligned and surface-aligned data in StudyForrest dataset (subjects S1 & S2) and in Grand Budapest dataset (subjects B1 & B2). Compared with surface alignment, hyperalignment was able to recover idiosyncratic topographies (e.g., the location, shape, and size of clusters) in great detail.

4. Discussion

In this study, using two datasets with different movies (Forrest Gump and Grand Budapest Hotel) and different localizers (one static and one dynamic localizer), we showed that a new subject's idiosyncratic functional topography can be estimated with high fidelity based only on that subject's naturalistic movie data using hyperalignment and other subjects' movie and localizer data. We also present a new optimized 1-step hyperalignment procedure for this application that outperforms our previous algorithm. Instead of projecting information encoded in idiosyncratic representational spaces into a common model space first (2-step procedure), the 1-step method directly projects each of the other subjects' data into the target subject's cortical space without the intermediate step of a group common model space. The optimized method generated better estimates of individually variable topographic maps though both procedures produced excellent performance. Importantly, our findings lay the foundation for developing a computational tool with a database from a normative group that could allow researchers to estimate new subjects' functional topographies with high-fidelity by collecting only a movie-viewing data set and then deriving the individualized topographies with the normative database.

In our earlier work (Guntupalli et al., 2016; Haxby et al., 2011) we have shown that individual-specific category-selectivity topographies and retinotopy can be estimated using hyperalignment. Our previous work on category-selective topographies mainly focused on the analysis of the posterior ventral temporal cortex. The current results replicated and expanded the previous findings to the whole cortex with strong correlations between face-selectivity contrast maps estimated from a participant's own localizer data and with contrast maps estimated from other participants' data in lateral occipital, superior temporal sulcal, and right lateral prefrontal cortices. In addition, we used Cronbach's alpha to estimate the internal reliability of localizer runs, which is more accurate than previous split-half correlations. In the previous analysis, hyperalignment-derived individual transformation matrices projected localizer data into a common model representational space. When estimating the target participant's individualized functional map, other participants' localizer data were projected into the target subject's cortical anatomy from the common space with the transpose of that participant's transformation matrix (two-step procedure). By contrast, the one-step procedure allows other participants' data to be projected directly into the target participant's cortical anatomy, avoiding accumulation of errors over two transformation steps. While the 1-step method boosts accuracy of estimates of selective topographies, as shown here, the 2-step method provides a group common model space that is more appropriate and efficient for other analyses, such as studies with between-subject multivariate pattern classification (bsMVPC, Guntupalli et al., 2016; Haxby et al., 2011), intersubject and between area correlation of representational geometries (Guntupalli et al., 2016, 2018; Nastase et al., 2017; Visconti di Oleggio Castello et al., 2017), modeling encoding in a common neural representational space (Van Uden et al., 2018), and investigating individual differences in fine-scale functional architecture (Feilong et al., 2018). One thing to keep in mind is that the one-step procedure can be computationally expensive in studies with large number of participants. The number of pairwise transformation matrices needed in the one-step method increases with the square of the sample size whereas in the two-step method the number of transformation matrices increases linearly with sample size.

We found more reliable estimates using the dynamic localizer than the static localizer, consistent with previous reports on the increased power of dynamic localizers (Fox et al., 2009; Pitcher et al., 2011). The maps show that the dynamic localizer better maps face-selectivity in lateral temporal and prefrontal cortices, and that this is captured well by the movie-viewing derived transformations. We would predict that the StudyForrest data also could reveal this more extensive face-selectivity map with high fidelity if dynamic localizer data were available in those subjects. The variable results with different localizers indicate that no localizer should be considered the "ground truth". Furthermore, the responsivity of the network defined by "face selectivity" also encompasses modulation of responses to nonface information related to the animacy continuum (Connolly et al., 2012, 2016; Çukur et al., 2013; Sha et al., 2014; Thorat et al., 2019), biological motion (Beauchamp et al., 2003; Bonda et al., 1996; Çukur et al., 2013; Gobbini et al., 2007; Grossman and Blake, 2002), social interaction (Castelli et al., 2002; Çukur et al., 2013; Gobbini et al., 2007; Schultz et al., 2003), and agentic behavior (Gobbini et al., 2011; Nastase et al., 2017), suggesting that the ground truth for the functional selectivity of the system identified by face-selectivity may extend well beyond face-selectivity *per se*. Thus, any functional localizer can provide only an imprecise estimate of a topography whose function is only partially defined.

We showed further that transformation matrices derived from StudyForrest and Grand Budapest also can be applied to estimate individual participants' scene-, body-, and object-selective topographies with

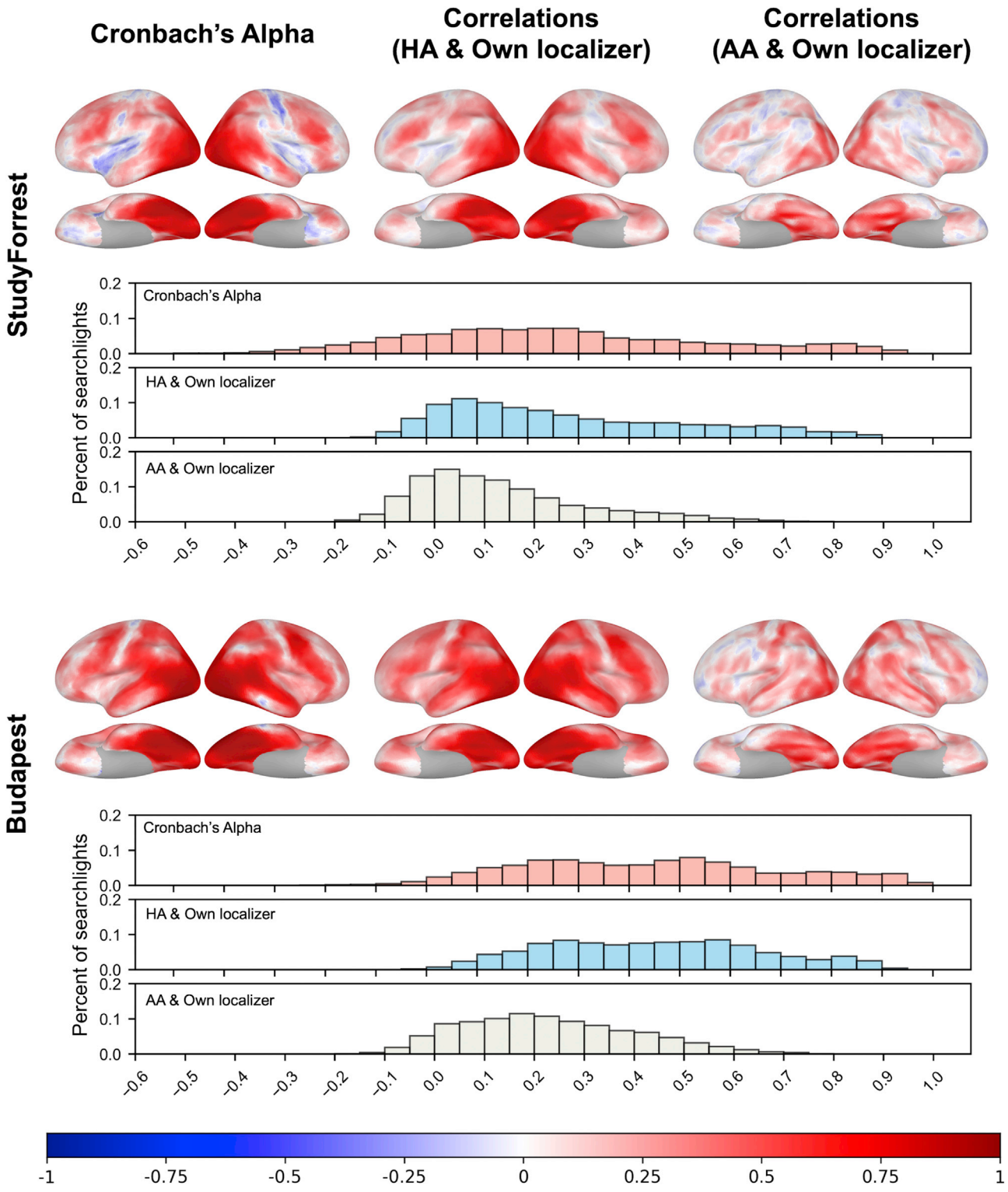

Fig. 6. Searchlight analysis of Cronbach's alpha, correlations between maps from participant's own localizer data and data estimated from others' hyperaligned or surface-aligned data. The upper two rows are results based on the StudyForrest data set, and the bottom two rows are based on the Grand Budapest data set. Within each dataset, Cronbach's alpha across four localizer runs, mean correlations between maps from participant's own localizer data and data estimated from others' 1-step hyperaligned or surface-aligned data were calculated in each searchlight (15 mm radius) and plotted on the cortical surface. Histogram plots with bin size of 0.05 are displayed for Conbach's alpha and those two maps accordingly.

high-fidelity. These results show that data collected during movie viewing can be used to estimate multiple category-selective topographies. Elsewhere, we also have shown that we can map retinotopy in early visual areas from other participants' retinotopy localizer data that are projected into a new participant's occipital cortices with hyperalignment transformation matrices based on movie viewing data (Guntupalli et al., 2016). There are numerous other functional localizers in other perceptual and cognitive domains, such as simple visual motion, biological motion, tonotopy, voice perception, music perception, language, calculations, working memory, and theory of mind. Because

naturalistic movies like Forrest Gump and Grand Budapest Hotel include people, human actions, conversations, social interactions, background music etc., we predict that hyperalignment transformation matrices based on these movies also will work for localizers of functional topographies for audition, language, and social cognition. Some high-level cognitive processes, such as calculation, working memory, and logical reasoning, may be less well sampled by movie viewing, and further work is necessary to test whether hyperalignment based on movie-viewing data can be used to estimate topographies for these other domains of high-level cognition. Researchers may need to select or develop movies that involve more calculations or logical inference-making to afford accurate estimation of topographies in these domains.

Functional topographies are individually variable, making it necessary to estimate a topography for each individual. Functional localizers are inefficient because they estimate one or, at most, a few topographies at a time. Hyperalignment, by contrast, models topographies as overlapping topographic basis functions (Guntupalli et al., 2016, 2018; Haxby et al., 2011), giving it the capacity to model multiple topographies as different mixtures of these basis functions. Basing hyperalignment on patterns of brain activity elicited by viewing and listening to a naturalistic, dynamic stimulus lends even more power by engaging multiple cognitive systems in parallel, better approaching natural cognition. The result is that the approach we present here has the potential to estimate an unlimited variety of functional topographies at the individual level based on the responses to a single naturalistic, dynamic stimulus. A tool to take advantage of this capacity would require a normative database of participants who were scanned during movie viewing and during functional localizers for different perceptual and cognitive functions, as well as a software tool for calculating transformation matrices and projecting the data from the normative sample into new brains' cortical anatomies. This software tool could be used with fMRI data at different resolutions mapped to the cortical surface or in its original volumetric space. Hyperalignment can rotate a cortical information space based on cortical vertices from a normative database into a new brain's cortical information space that is similarly mapped to cortical vertices or that is in voxels at low or high resolution. Given such a database with a shared software tool, investigators would need to scan their participants only during movie viewing and a wide range of idiosyncratic functional topographies could then be estimated individually based on localizer data projected from the brains in the normative sample into the new participants' cortical anatomies.

Extending this approach to other populations, such as children, or other cultural groups will present further challenges for selecting appropriate movies and developing databases that allow adjustment for factors such as age.

The procedure that we present here produces continuous maps of selectivity that are estimated from other participants' data. These maps can be further processed to identify category-selective regions, such as the OFA or FFA, using the same methods that are used to threshold and cluster a participant's own localizer data (see Supplementary Fig. S7). Because methods vary for identifying category-selective regions, we refrain from endorsing one method here and leave that step to the investigators' discretion.

Author contribution

MIG and JVH conceived and designed the experiment and supervised the project; MIG, JVH and GJ developed the methods of analysis and models; GJ performed the analyses; MVdOC collected the data; software was designed by JVH, JSG, GJ and MF; GJ wrote the first draft of the manuscript and MIG, JVH, MVdOC, JSG, MF and VC provided critical comments and suggestions for revision.

Funding

This work was supported by NSF grants 1607845 (JVH) and 1835200 (MIG).

Acknowledgements

We thank Yaroslav O. Halchenko, Samuel A. Nastase, Andrew C. Connolly for their helpful discussions and comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.116458.

References

Beauchamp, M.S., Lee, K.E., Haxby, J.V., Martin, A., 2003. fMRI responses to video and point-light displays of moving humans and manipulable objects. J. Cogn. Neurosci. 15, 991–1001.
Behzadi, Y., Restom, K., Liau, J., Liu, T.T., 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage 37, 90–101.
Bonda, E., Petrides, M., Ostry, D., Evans, A., 1996. Specific involvement of human parietal systems and the amygdala in the perception of biological motion. J. Neurosci. 16, 3737–3744.
Castelli, F., Frith, C., Happé, F., Frith, U., 2002. Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. Brain 125, 1839–1849.
Connolly, A.C., Guntupalli, J.S., Gors, J., Hanke, M., Halchenko, Y.O., Wu, Y.-C., Abdi, H., Haxby, J.V., 2012. The representation of biological classes in the human brain. J. Neurosci. 32, 2608–2618.
Connolly, A.C., Sha, L., Guntupalli, J.S., Oosterhof, N., Halchenko, Y.O., Nastase, S.A., Visconti di Oleggio Castello, M., Abdi, H., Jobst, B.C., Gobbini, M.I., et al., 2016. How the human brain represents perceived dangerousness or "predacity" of animals. J. Neurosci. 36, 5373–5384.
Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.
Çukur, T., Huth, A.G., Nishimoto, S., Gallant, J.L., 2013. Functional subdomains within human FFA. J. Neurosci. 33, 16748–16766.
Esteban, O., Markiewicz, C.J., Blair, R.W., Moodie, C.A., Isik, A.I., Erramuzpe, A., Kent, J.D., Goncalves, M., DuPre, E., Snyder, M., et al., 2019. fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat. Methods 16, 111.
Feilong, M., Nastase, S.A., Guntupalli, J.S., Haxby, J.V., 2018. Reliable individual differences in fine-grained cortical functional architecture. Neuroimage 183, 375–386.
Fischl, B., 2012. FreeSurfer. Neuroimage 62, 774–781.
Fischl, B., Sereno, M.I., Dale, A.M., 1999. Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. Neuroimage 9, 195–207.
Fox, C.J., Iaria, G., Barton, J.J.S., 2009. Defining the face processing network: optimization of the functional localizer in fMRI. Hum. Brain Mapp. 30, 1637–1651.
Gobbini, M.I., Koralek, A.C., Bryan, R.E., Montgomery, K.J., Haxby, J.V., 2007. Two takes on the social brain: a comparison of theory of mind tasks. J. Cogn. Neurosci. 19, 1803–1814.
Gobbini, M.I., Gentili, C., Ricciardi, E., Bellucci, C., Salvini, P., Laschi, C., Guazzelli, M., Pietrini, P., 2011. Distinct neural systems involved in agency and animacy detection. J. Cogn. Neurosci. 23, 1911–1920.
Greve, D.N., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. Neuroimage 48, 63–72.
Grossman, E.D., Blake, R., 2002. Brain areas active during visual perception of biological motion. Neuron 35, 1167–1175.
Guntupalli, J.S., Hanke, M., Halchenko, Y.O., Connolly, A.C., Ramadge, P.J., Haxby, J.V., 2016. A model of representational spaces in human cortex. Cerebr. Cortex 26, 2919–2934.
Guntupalli, J.S., Wheeler, K.G., Gobbini, M.I., 2017. Disentangling the representation of identity from head view along the human face processing pathway. Cerebr. Cortex 27, 46–53.
Guntupalli, J.S., Feilong, M., Haxby, J.V., 2018. A computational model of shared fine-scale structure in the human connectome. PLoS Comput. Biol. 14, e1006120.
Hanke, M., Halchenko, Y.O., Sederberg, P.B., Hanson, S.J., Haxby, J.V., Pollmann, S., 2009. PyMVPA: a Python toolbox for multivariate pattern analysis of fMRI data. Neuroinformatics 7, 37–53.
Hanke, M., Baumgartner, F.J., Ibe, P., Kaule, F.R., Pollmann, S., Speck, O., Zinke, W., Stadler, J., 2014. A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie. Sci. Data 1, 140003.
Hanke, M., Adelhöfer, N., Kottke, D., Iacovella, V., Sengupta, A., Kaule, F.R., Nigbur, R., Waite, A.Q., Baumgartner, F., Stadler, J., 2016. A studyforrest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation. Sci. Data 3, 160092.
Haxby, J.V., Gobbini, M.I., 2011. Distributed neural systems for face perception. In: Oxford Handbook of Face Perception. Oxford University Press, Oxford, New York, pp. 93–110.
Haxby, J.V., Horwitz, B., Ungerleider, L.G., Maisog, J.M., Pietrini, P., Grady, C.L., 1994. The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. J. Neurosci. 14, 6336–6353.

Haxby, J.V., Hoffman, E.A., Gobbini, M.I., 2000. The distributed human neural system for face perception. Trends Cogn. Sci. 4, 223–233.
Haxby, J.V., Guntupalli, J.S., Connolly, A.C., Halchenko, Y.O., Conroy, B.R., Gobbini, M.I., Hanke, M., Ramadge, P.J., 2011. A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron 72, 404–416.
Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825–841.
Nastase, S.A., Connolly, A.C., Oosterhof, N.N., Halchenko, Y.O., Guntupalli, J.S., Visconti di Oleggio Castello, M., Gors, J., Gobbini, M.I., Haxby, J.V., 2017. Attention selectively reshapes the geometry of distributed semantic representation. Cerebr. Cortex 27, 4277–4291.
Pitcher, D., Dilks, D.D., Saxe, R.R., Triantafyllou, C., Kanwisher, N., 2011. Differential selectivity for dynamic versus static information in face-selective cortical regions. Neuroimage 56, 2356–2363.
Power, J.D., Mitra, A., Laumann, T.O., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2014. Methods to detect, characterize, and remove motion artifact in resting state fMRI. Neuroimage 84, 320–341.
Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., Saxe, R., 2018. Development of the social brain from age three to twelve years. Nat. Commun. 9, 1–12.
Saxe, R., Brett, M., Kanwisher, N., 2006. Divide and conquer: a defense of functional localizers. Neuroimage 30, 1088–1096.
Schultz, R.T., Grelotti, D.J., Klin, A., Kleinman, J., Van der Gaag, C., Marois, R., Skudlarski, P., 2003. The role of the fusiform face area in social cognition: implications for the pathobiology of autism. Philos. Trans. R. Soc. Lond. B Biol. Sci. 358, 415–427.
Sengupta, A., Kaule, F.R., Guntupalli, J.S., Hoffmann, M.B., Häusler, C., Stadler, J., Hanke, M., 2016. A studyforrest extension, retinotopic mapping and localization of higher visual areas. Sci. Data 3, 160093.
Sha, L., Haxby, J.V., Abdi, H., Guntupalli, J.S., Oosterhof, N.N., Halchenko, Y.O., Connolly, A.C., 2014. The animacy continuum in the human ventral vision pathway. J. Cogn. Neurosci. 27, 665–678.
Thorat, S., Proklova, D., Peelen, M.V., 2019. The Nature of the Animacy Organization in Human Ventral Temporal Cortex. ArXiv:1904.02866 [q-Bio].
Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging 29, 1310–1320.
Van Uden, C.E., Nastase, S.A., Connolly, A.C., Feilong, M., Hansen, I., Gobbini, M.I., Haxby, J.V., 2018. Modeling semantic encoding in a common neural representational space. Front. Neurosci. 12.
Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L.C., Castellanos, F.X., 2015. Inscapes: a movie paradigm to improve compliance in functional magnetic resonance imaging. Neuroimage 122, 222–232.
Visconti di Oleggio Castello, M., Halchenko, Y.O., Guntupalli, J.S., Gors, J.D., Gobbini, M.I., 2017. The neural representation of personally familiar and unfamiliar faces in the distributed system for face perception. Sci. Rep. 7, 12237.
Zhen, Z., Kong, X.-Z., Huang, L., Yang, Z., Wang, X., Hao, X., Huang, T., Song, Y., Liu, J., 2017. Quantifying the variability of scene-selective regions: Interindividual, interhemispheric, and sex differences. Human Brain Mapping 38, 2260–2275.
Zhen, Z., Yang, Z., Huang, L., Kong, X., Huang, Y., Song, Y., Liu, J., 2015. Quantifying interindividual variability and asymmetry of face-selective regions: A probabilistic functional atlas. Neuroimage 113, 13–15.