Hyperface: a naturalistic fMRI dataset for investigating human face processing

Matteo Visconti di Oleggio Castello[1,*], Guo Jiahui[2], Ma Feilong[3], Manon de Villemejane[4], James V. Haxby[4], Maria Ida Gobbini[5,6,*]

[1] Department of Neuroscience, University of California, Berkeley, CA 94720
[2] Department of Psychology, School of Behavioral and Brain Sciences, The University of Texas at Dallas, Richardson, TX 75080
[3] Department of Psychology, University of South Carolina, Columbia, SC 29201
[4] Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755
[5] Department of Medical and Surgical Sciences (DIMEC), University of Bologna, Bologna, Italy
[6] ISNB IRCCS, Institute of Neurological Sciences of Bologna, Bologna, Italy
* Correspondence:
Maria Ida Gobbini <mariaida.gobbini@unibo.it>,
Matteo Visconti di Oleggio Castello <mvdoc@berkeley.edu>

Abstract

Faces convey information that guides social behavior, yet neuroimaging studies investigating human face processing typically use static images with small sets of identities under artificial conditions. Controlled designs limit our ability to characterize human face processing under naturalistic conditions or test whether computational models generalize beyond the laboratory. To address this gap, here we release *hyperface*, a naturalistic face viewing fMRI dataset designed to investigate human face processing in response to faces portrayed in videos mimicking more ecologically valid conditions. Twenty-one participants watched 707 unique face video clips that vary systematically in identity, gender, age, ethnicity, expression, and head orientation. Each clip was rated by independent observers, and pairwise similarity judgments were collected through a behavioral arrangement task. Technical validation confirms high data quality with low motion, high tSNR, and high inter-subject correlation in visual and face-processing regions. The *hyperface* dataset is part of a comprehensive experimental framework to investigate human face processing: all 21 participants also watched "The Grand Budapest Hotel," performed a dynamic face localizer task, and 10 participants completed an additional face identity task with personally familiar and visually familiarized faces. These datasets are publicly available and enable within-subject comparisons across paradigms. Together they provide a unique resource for characterizing human face processing under naturalistic conditions and for benchmarking computational models against human brain responses.

Background & Summary

Faces convey rich information that guides our behavior in social contexts. From a single face we can quickly recognize identity, and infer gender, age, emotional state, personality traits, and other socially relevant characteristics of an individual.[1–5] Few other visual objects can convey such rich and socially relevant information. Despite the critical importance of faces in everyday life, neuroimaging studies of face processing typically rely on paradigms that fail to capture the richness and complexity of natural face perception, often using small sets of static images of identities presented under highly controlled conditions. While these controlled designs allow for precise testing of specific hypotheses, they limit our ability to fully characterize face processing under naturalistic conditions, or to test whether computational models of face processing generalize to real-world scenarios.

Naturalistic neuroimaging paradigms address these limitations by enabling investigation of brain function under conditions that better resemble real life.[6–9] However, only a few publicly available datasets so far exist to investigate face processing under naturalistic conditions.[10,11] To address this gap, here we release a naturalistic face viewing fMRI dataset called *hyperface*, in which 21 participants watched 707 unique 4 s face video clips sampled from publicly available YouTube interviews. To investigate face processing under a wide range of conditions, these clips show individuals who vary systematically in identity, gender, age, ethnicity, expression, and head orientation. Each clip was rated by independent observers. Additional pairwise similarity judgments were collected through a behavioral arrangement task in which independent observers were asked to arrange the clips based on perceived similarity of face appearance. In recent work,[12] we used this dataset with hyperalignment and representational similarity analysis to show that DNNs trained for face and object identification tasks successfully capture human cognitive categorical face representation in the behavioral arrangement task, but fail to explain brain representations of identity during dynamic, naturalistic face viewing. This suggests that state-of-the-art visual DNNs are not accurate models of human face processing under naturalistic conditions.[12]

The *hyperface* dataset is part of a comprehensive experimental framework designed to characterize human face processing under naturalistic conditions.[13] This framework includes two additional fMRI datasets already publicly available: a movie-watching dataset in which participants watched Wes Anderson's "The Grand Budapest Hotel,"[11] a socially rich movie with many characters and social relationships; and a controlled fMRI paradigm in which participants viewed faces of personally familiar and visually familiarized individuals with different head orientations.[14] Importantly, all 21 *hyperface* participants also watched "The Grand Budapest Hotel", performed a dynamic face localizer task, and 10 participants completed all experiments.[13] This comprehensive set of experiments enables testing hypotheses about face processing across paradigms, and provides a unique data resource across human neuroscience, computational neuroscience, and artificial intelligence. It allows investigating human face processing with a variety of methods, such as inter-subject correlation (ISC),[15] multivoxel pattern analysis and representational similarity analysis,[16,17] or voxelwise encoding models;[9,18,19] and it enables benchmarking computational models of face processing against human brain responses.[12]

Methods

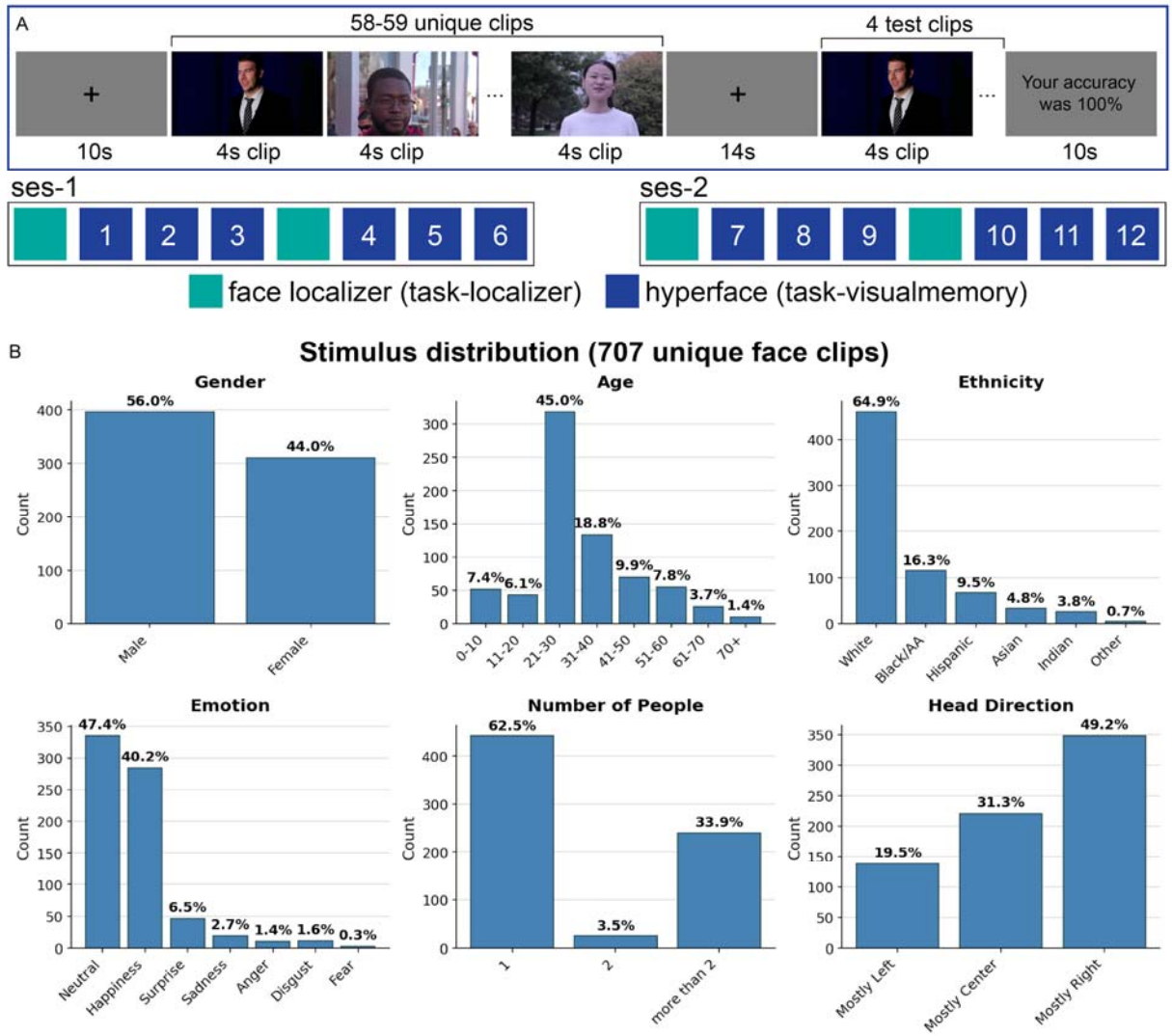


Figure 1. Experimental design and stimulus distribution of face stimuli. A. In each *hyperface* functional run participants watched 58-59 short face clips (4 s each, no sound), then performed an attention check by indicating whether four face clips had been presented in the run. Data were collected across two scanning sessions. Each session included six *hyperface* runs (*task-visual memory*) and two dynamic face localizer runs (*task-localizer*). **B.** 707 unique face clips were selected from publicly available YouTube interviews and rated on Amazon Mechanical Turk for Gender, Age, Ethnicity, Emotion, Number of People, and Head Direction. Each clip was rated by at least eight Mechanical Turk workers, and the final rating was assigned as the modal response (see Methods).

Participants

Twenty-one participants (11 females, mean age 27.3 ± 2.3 years, range 22–31) took part in the experiment. All had normal or corrected-to-normal vision and used a custom-fitted CaseForge headcase (https://github.com/gallantlab/headcase-pipeline) to minimize head motion. The study was approved by the Dartmouth Committee for the Protection of Human Subjects.

The same participants also watched "The Grand Budapest Hotel" in separate scanning sessions (see ref.[11] for more information). Moreover, 10 of the 21 participants completed two additional scanning sessions involving a face identity task with personally familiar and visually familiarized faces (see ref.[14] for more information).

Stimuli

We created 707 unique face video clips (4 s each, no audio) by sampling publicly available YouTube interviews. The clips showed different individuals varying in age, ethnicity, gender, expression, and head orientation. Clips were cropped as necessary to remove distracting text, and audio was removed.

Behavioral rating task. A total of 121 Amazon Mechanical Turk workers participated in a behavioral rating task. In each trial, workers watched a video clip and rated the stimulus on five features: perceived gender (male/female), age (0–10, 11–20, 21–30, 31–40, 41–50, 51–60, 61–70, 70+), ethnicity (White, Black or African American, Asian, Indian, Hispanic or Latino, Other), expression (neutral, happiness, surprise, anger, disgust, sadness, fear), and head orientation (mostly left, mostly center, mostly right). All 707 clips were divided into 15 independent sessions (approximately 47 clips each). Each worker completed one session. At least eight workers completed each session, and the final rating for each clip was assigned as the modal response across workers.

Behavioral arrangement task. An independent group of 39 Amazon Mechanical Turk workers performed a multi-arrangement task to obtain pairwise similarity judgments for all face clips.[12,20] Within each trial, all stimuli from one scanning run (58–59 clips) were displayed as thumbnails around the circumference of a white circle on a gray background. Workers arranged the thumbnails within the circle based on perceived similarity of face appearance. Hovering over a thumbnail triggered a larger, dynamic display of the video clip; right-clicking replayed the video. Each worker completed three randomly selected trials from the 12 total runs. At least 10 workers arranged the stimuli of each scanning run.

Experimental Procedure

Participants completed two scanning sessions. Each session comprised one gradient echo recalled (GRE) fieldmap scan, eight functional runs, and one diffusion-weighted imaging (DWI) scan preceded by two spin-echo sequences with opposite phase-encoding directions for distortion correction. The first session also included one T2-weighted anatomical scan. A T1-weighted anatomical scan was collected in a separate session in which participants watched “The Grand Budapest Hotel”; this T1-weighted scan is also included in the hyperface data release. The order of the functional runs was fixed across sessions: one localizer run, three hyperface runs, one localizer run, and three hyperface runs (see Figure 1). All stimuli were back-projected onto a screen and subtended approximately 16.27° × 9.17° (W × H) of visual angle. During each hyperface run, participants viewed a continuous stream of at least 58 face clips (4 s each). Participants were instructed to attend to the identity of the individuals portrayed. The clips were followed by a 14-s blank period and an attention check in which four face clips were presented sequentially (see Figure 1). Participants indicated whether each individual had appeared in the preceding run. Two of the four clips were matches and two were non-matches.

Clip presentation order in the hyperface runs was counterbalanced across participants. The 707 clips were divided into 12 blocks of 58–59 clips each, one block per functional run. To ensure each block contained a balanced distribution of face attributes (gender, age, ethnicity, emotion, number of people, head direction), 10,000 random permutations of the stimulus order were generated, and we selected the permutation that minimized the summed squared deviation from a uniform distribution of attributes across blocks. Each block was also assigned four catch-trial clips: two matches (randomly selected from within the block) and two non-matches (randomly selected from a different block). The 12 blocks were then assigned to two sessions of six blocks each. Session order was counterbalanced across participants, and block order within each session was randomized for each participant.

During each localizer run, participants viewed 3-s clips of faces (F), bodies (B), scenes (S), objects (O), and scrambled objects (Sc).[21] Clips were presented continuously without intervening blank periods. Participants pressed a button whenever a clip repeated. Block order followed a fixed structure: an 18-s fixation block, five category blocks (18 s each) in random order, an 18-s fixation block, and five category blocks in reversed order (e.g., B-S-F-O-Sc followed by Sc-O-F-S-B, see ref.[21] for more details). The order of the category blocks was randomized across runs.

Imaging Parameters

All data were acquired using a 3T Siemens Magnetom Prisma MRI scanner (Siemens, Erlangen, Germany) with a 32-channel phased-array head coil at the Dartmouth Brain Imaging Center. Functional, blood oxygenation level-dependent (BOLD) images were acquired in an interleaved fashion using gradient-echo echo-planar imaging with pre-scan normalization, fat suppression, a 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 1): TR / TE = 1000 / 33 ms, flip angle = 59°, resolution = 2.5 mm^3 isotropic voxels, matrix size = 96 x 96, FoV = 240 x 240 mm, 48 axial slices with full cortical coverage and no gap, anterior-posterior phase encoding. At the beginning of each run, three dummy scans were acquired to allow for signal stabilization. At the beginning of each imaging session, a single dual-echo gradient echo (GRE) fieldmap was acquired for spatial distortion correction: TR = 675 ms, TE1/TE2 = 4.92/7.38 ms, flip angle = 60°, resolution = 2.5 mm^3 isotropic voxels, matrix size = 96 x 96, 48 axial slices, anterior–posterior phase encoding.
A T2-weighted structural scan was acquired with an in-plane acceleration factor of 4 using GRAPPA:
TR/TE = 3200/563 ms, flip angle = 120°, resolution = 0.9375 x 0.9375 x 0.9 mm voxels, matrix size = 256 x 256, FoV = 240 x 240 x 172.8 mm, 192 sagittal slices, ascending acquisition, anterior–posterior phase encoding, no fat suppression, 3 min 21 s total acquisition time.

A T1-weighted scan acquired during the Grand Budapest Hotel session[11] is included in this dataset. This scan had the following parameters: 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 x 0.9375 x 0.9 mm voxels, matrix size = 256 x 256, FoV = 240 x 240 x 172.8 mm, 192 sagittal slices, ascending acquisition, anterior–posterior phase encoding, no fat suppression, 5 min 21 s total acquisition time.
Diffusion-weighted images were acquired using a single-shot spin-echo echo-planar imaging (SE-EPI) sequence with pre-scan normalization, fat suppression, and a multiband (i.e., simultaneous multi-slice; SMS) acceleration factor of 4: TR/TE = 2600/103 ms, flip angle = 90°, resolution = 2.0 mm^3 isotropic voxels, matrix size = 120 x 120, FoV = 240 x 240 mm, 64 axial slices with no gap, posterior–anterior phase encoding. Diffusion weighting was applied along 96 non-collinear directions (HARDI scheme) with b = 1000 s/mm^2, interleaved with 11 non-diffusion-weighted volumes (b = 0 s/mm^2) distributed throughout the acquisition for a total of 107 volumes and approximately 4 min 38 s total acquisition time. For susceptibility-induced distortion correction of the diffusion-weighted images, a pair of spin-echo EPI scans was acquired with opposite phase encoding directions (anterior–posterior and posterior–anterior) using the same geometric parameters as the DWI acquisition: TR/TE = 9700/96 ms, flip angle = 90°, resolution = 2.0 mm^3 isotropic voxels, matrix size = 120 x 120, FoV = 240 x 240 mm, 64 axial slices with no gap, fat suppression. Each phase encoding direction included 6 non-diffusion-weighted volumes (b = 0 s/mm^2), with an effective echo spacing of 0.64 ms and total readout time of 76.2 ms.

Dataset Preparation

DICOM data were converted to BIDS format (version 1.10.1) using HeuDiConv[22,23] with the ReproIn heuristic.[24] T1- and T2- weighted anatomical volumes were defaced using Pydeface.[25]

Preprocessing

The text in this subsection is copied from the output of fMRIPrep as recommended by its developers (see https://www.nipreps.org/intro/transparency/#citation-boilerplates). Minor edits were performed for consistency and formatting.
Results included in this manuscript come from preprocessing performed using fMRIPrep 25.1.4[26,27] (RRID:SCR_016216), which is based on Nipype 1.10.0[28,29] (RRID:SCR_002502).

Preprocessing of B0 inhomogeneity mappings. A total of 2 fieldmaps were found available within the input BIDS structure for this particular subject. A B0 nonuniformity map (or fieldmap) was estimated from the phase-drift maps

measure with two consecutive GRE (gradient-recalled echo) acquisitions. The corresponding phase-maps were phase-unwrapped with prelude (FSL).

Anatomical data preprocessing. A total of 1 T1-weighted (T1w) images were found within the input BIDS dataset. The T1w image was corrected for intensity non-uniformity (INU) with *N4BiasFieldCorrection*,[30] distributed with ANTs 2.6.2[31] (RRID:SCR_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the *antsBrainExtraction.sh* workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using *fast* (FSL, RRID:SCR_002823).[32] Brain surfaces were reconstructed using *recon-all* (FreeSurfer 7.3.2, RRID:SCR_001847),[33] and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR_002438).[34] A T2-weighted image was used to improve pial surface refinement. Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with *antsRegistration* (ANTs 2.6.2), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization and accessed with TemplateFlow[35] (24.2.2): ICBM 152 Nonlinear Asymmetrical template version 2009c[36] (RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym).

Functional data preprocessing. For each of the 16 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume was generated, using a custom methodology of fMRIPrep, for use in head motion correction. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using *mcflirt* (FSL).[37] The estimated fieldmap was then aligned with rigid-registration to the target EPI (echo-planar imaging) reference run. The field coefficients were mapped on to the reference EPI using the transform. The BOLD reference was then co-registered to the T1w reference using *bbregister* (FreeSurfer) which implements boundary-based registration.[38] Co-registration was configured with six degrees of freedom. The aligned T2w image was used for initial co-registration. Several confounding time-series were calculated based on the preprocessed BOLD: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power[39] (absolute sum of relative motions) and Jenkinson[37] (relative root mean square displacement between affines). FD and DVARS are calculated for each functional run, both using their implementations in Nipype (following the definitions by Power et al.[39]). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor,[40]). Principal components are estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, a mask of pixels that likely contain a volume fraction of GM is subtracted from the aCompCor masks. This mask is obtained by dilating a GM mask extracted from the FreeSurfer's aseg segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each [41]. Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. Additional nuisance timeseries are calculated by means of principal components analysis of the signal found within a thin

band (crown) of voxels around the edge of the brain, as proposed by ref.[42]. The BOLD time-series were resampled onto the following surfaces (FreeSurfer reconstruction nomenclature): *fsaverage6*. All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using *nitransforms*, configured with cubic B-spline interpolation. Non-gridded (surface) resamplings were performed using *mri_vol2surf* (FreeSurfer).
Many internal operations of fMRIPrep use Nilearn 0.11.1[43] (RRID:SCR_001362), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in fMRIPrep's documentation.

Data Validation

The functional data preprocessed by fMRIPrep were denoised using custom Python scripts. The following nuisance parameters were regressed out from the functional time series using ordinary least-squares regression: six motion parameters and their derivatives, global signal, framewise displacement, the first six noise components estimated by aCompCor, and polynomial trends up to second order. All metrics of interest were computed on data denoised as described, either in volume space or in surface space. No additional spatial smoothing or temporal filtering was performed.

Temporal Signal-to-Noise Ratio (tSNR)

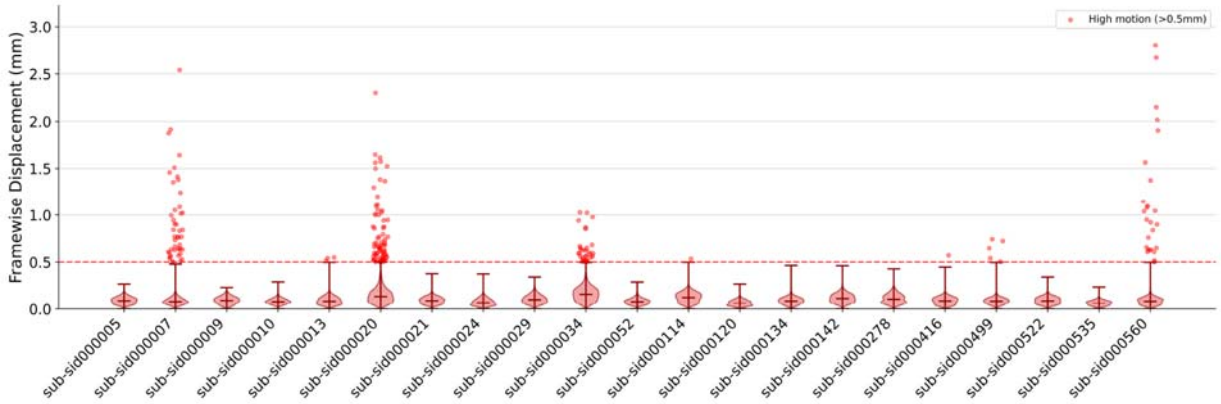
We computed tSNR for each preprocessed functional run in native subject space (*space-T1w*) to avoid spatial smoothing from template normalization. For each voxel, tSNR was calculated as the temporal mean divided by the temporal standard deviation. To visualize tSNR variation across brain areas at the group level, we repeated the analysis on data resampled to the fsaverage6 surface.

Inter-Subject Correlation (ISC)

ISC was computed to estimate the consistency of brain responses to the face clips. BOLD time series were projected to the template surface fsaverage6. Functional runs were reordered across participants to obtain the same stimulus order. Data corresponding to the initial buffer phase and memory test trials were discarded. To compute ISC, for each cortical vertex we correlated each participant's time series with the mean time series of the remaining 20 participants. This procedure was repeated for all participants, and a median ISC map was computed at the group level.

Data Records

To facilitate data sharing and the use of tools such as fMRIPrep and MRIQC,[26,44] the unpreprocessed data were standardized following the Brain Imaging Data Structure[45] (version 1.10.1). The hyperface raw dataset is available on OpenNeuro (https://openneuro.org/datasets/ds007329). This dataset includes the unpreprocessed data, the 707 face clips used in the experiment, as well as QA results and extensive derivative data used in ref. 12 (see Figure 2). Additional fMRIPrep derivative data is available athttps://openneuro.org/datasets/ds007384, and FreeSurfer derivative data is available at https://openneuro.org/datasets/ds007378. Presentation and QA scripts are available at https://github.com/mvdoc/hyperface-data-paper.



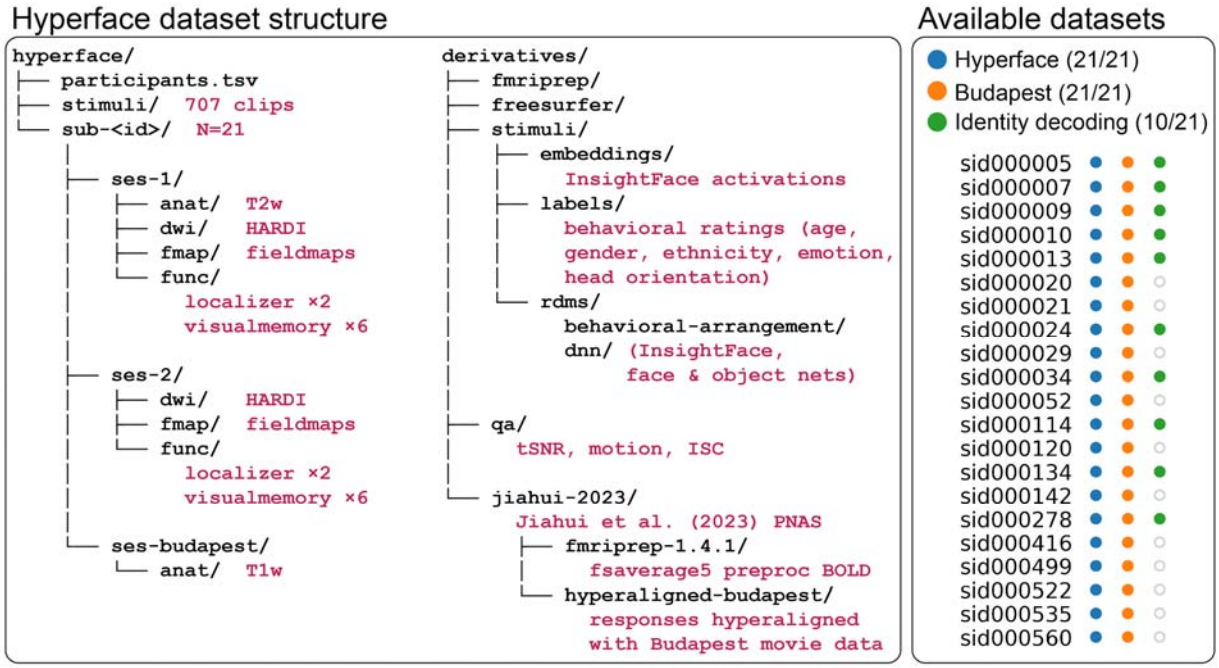


Figure 2. Dataset structure and available related datasets. (Left) The hyperface dataset follows the Brain Imaging Data Structure (BIDS) format.[45] Raw data include the 707 face clips used in the experiment (*stimuli*) and two hyperface sessions (*ses-1*, *ses-2*), each comprising one GRE fieldmap scan, eight functional runs, and one diffusion-weighted imaging (DWI) scan; *ses-1* also includes a T2-weighted anatomical scan. A T1-weighted anatomical scan from a separate session (*ses-budapest*) is included from the publicly available Grand Budapest Hotel dataset. The derivatives folder contains two sets of preprocessed data: fMRIPrep 25.1.4 and FreeSurfer 7.3.2 outputs used for technical validation in this manuscript, and *jiahui-2023*, which includes preprocessed and hyperaligned data used in ref. 12. Quality assurance (*qa*) files used for technical validation are also provided. Stimulus derivatives include InsightFace embeddings for all face clips, behavioral ratings from Mechanical Turk, and representational dissimilarity matrices (RDMs) derived from a behavioral arrangement task and from face- and object-specific deep neural networks (see ref.[12] for details). **(Right)** Related publicly available datasets for the 21 hyperface participants. All participants also watched the second part of The Grand Budapest Hotel in the scanner.[11] Ten participants completed two sessions of a controlled task used for face identity decoding with personally familiar and visually familiar faces.[14] Together, these three fMRI datasets provide a unique resource for characterizing human face processing under naturalistic conditions and for benchmarking computer vision algorithms.

Figure 3. Framewise displacement for each participant across hyperface runs. Each violin plot shows the distribution of framewise displacement (FD) values across all functional volumes for one participant. Red dots indicate individual volumes with FD > 0.5 mm (dashed line). Subject motion was low across the dataset: median FD was 0.09 mm (range: 0.06–0.16 mm), well below the 0.5 mm threshold. Thirteen of 21 participants had fewer than 5% of volumes marked as motion outliers by fMRIPrep (FD > 0.5 mm or standardized DVARS > 1.5; see Methods).

Technical Validation

We validated this dataset following our previously published approach to ensure data quality across separate domains.[11] First, we analyzed participant motion to quantify behavioral sources of noise. Second, we estimated temporal signal-to-noise ratio (tSNR) for each voxel to verify comparable SNR levels across participants and to identify areas with low SNR. Third, we computed inter-subject correlation (ISC) as a check that the stimulus evoked consistent brain responses across participants.

Motion. Participant motion was low throughout the dataset. The median framewise displacement across participants was 0.09 mm (range: 0.06–0.16 mm; Figure 3). The median percentage of volumes marked as motion outliers was 3.76% (range: 0.41–21.18%), with 13 of 21 participants having fewer than 5% outlier volumes. fMRIPrep defines an outlier as a volume with framewise displacement greater than 0.5 mm or standardized DVARS greater than 1.5 (see Methods).

Temporal SNR. Temporal SNR (tSNR) was estimated for all participants in native anatomical space to minimize interpolation that would artificially inflate tSNR, and in *fsaverage6* template space for qualitative assessment across cortical areas. The mean whole-brain tSNR was 83.0 ± 4.0, comparable to previous datasets at 3T.[11,46,47] As expected for a multiband BOLD EPI sequence, tSNR was highest in occipital and dorsal areas near the receiver coils and lowest in anterior temporal and orbitofrontal cortex near air-tissue boundaries (Figure 4).

Inter-Subject Correlation. ISC identifies regions where brain activity in response to the face clips was consistent across participants. Visual areas showed the highest ISC values, as expected from a dynamic visual stimulus (Figure 5). Positive ISC values were also found in face-processing regions such as the occipital face area (OFA), fusiform face area (FFA), and superior temporal sulcus (STS), as well as dorsal precuneus[48] and right inferior frontal gyrus.[14,49] Note that because this analysis was based on anatomical alignment, these ISC estimates represent a lower bound of what can be achieved with this dataset. Previous work has shown that hyperalignment[50–53] provides much higher ISC values compared to anatomical alignment.[12,50]

Taken together, these analyses validate the high quality of this dataset. Low motion reflects good participant compliance during scanning. Comparable tSNR levels indicate consistent data quality across participants. ISC reveals shared brain responses in visual and face-processing areas, confirming that the naturalistic face stimuli elicited reliable brain activity across participants.

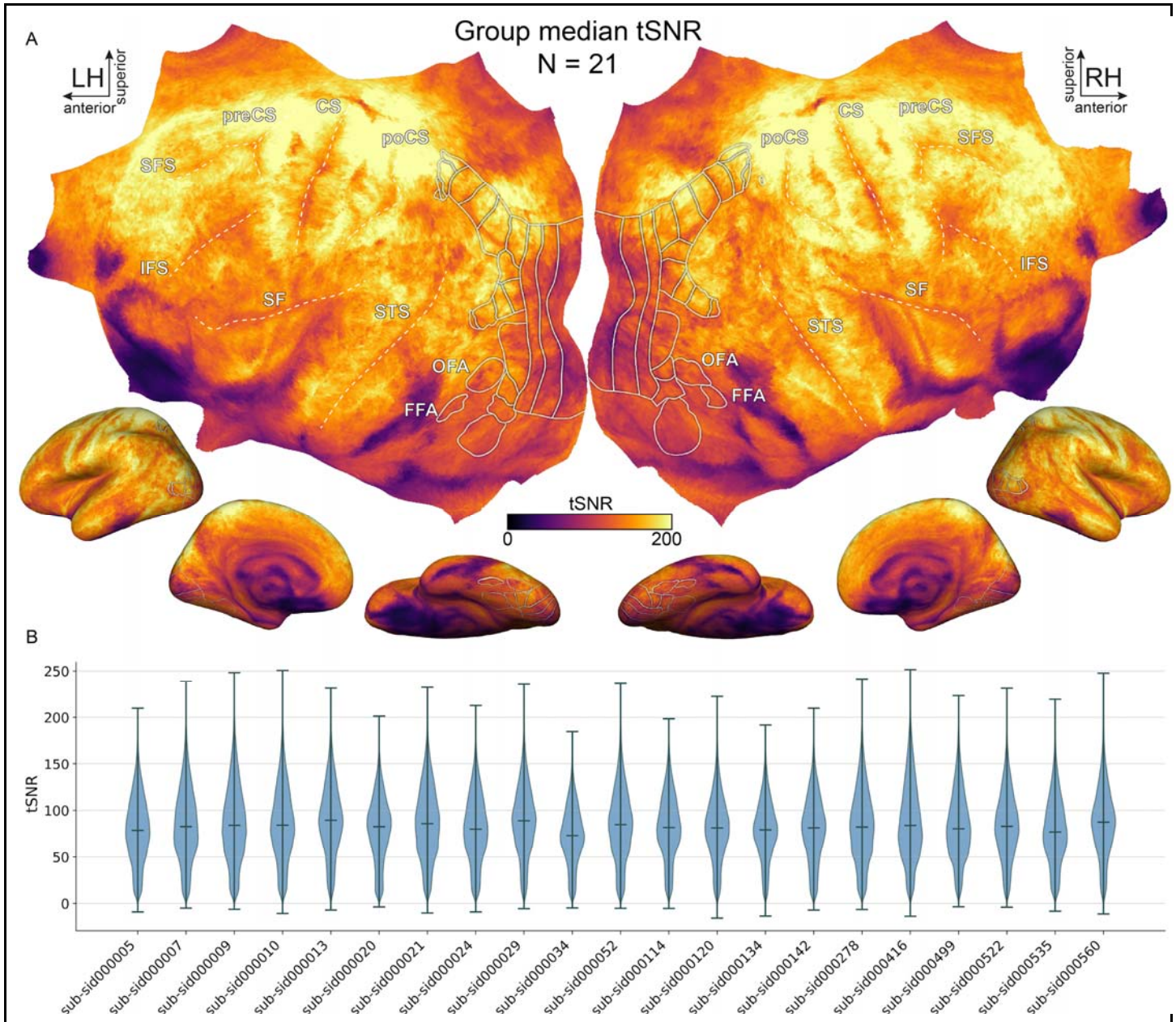


Figure 4. Temporal signal-to-noise ratio (tSNR) across participants. A. Group median tSNR (N=21) computed on data projected to *fsaverage6* surface space, displayed on a flatmap (top) and inflated surfaces (bottom). White outlines indicate anatomical landmarks and visual ROIs. tSNR is high across most of the cortex, with expected signal dropout near air-tissue boundaries in anterior temporal and orbitofrontal cortex. **B.** Violin plots show the distribution of voxelwise tSNR values for each participant, computed in native space. Group-averaged tSNR was 83.0 ± 4.0, consistent with other publicly available 3T datasets with similar voxel size. Abbreviations: CS, central sulcus; FFA, fusiform face area; IFS, inferior frontal sulcus; OFA, occipital face area; poCS, postcentral sulcus; preCS, precentral sulcus; SF, sylvian fissure; SFS, superior frontal sulcus; STS, superior temporal sulcus.

Code Availability

All the code to replicate the QA analyses is available at https://github.com/mvdoc/hyperface-data-paper. The code uses several Python packages such as numpy,[54] scipy,[55] pandas, matplotlib,[56] seaborn, nibabel,[57] nilearn,[43] scikit-learn,[58] and pycortex.[59] Additional code associated with our previous publication using this dataset[12] is available at https://github.com/GUO-Jiahui/face_DCNN.

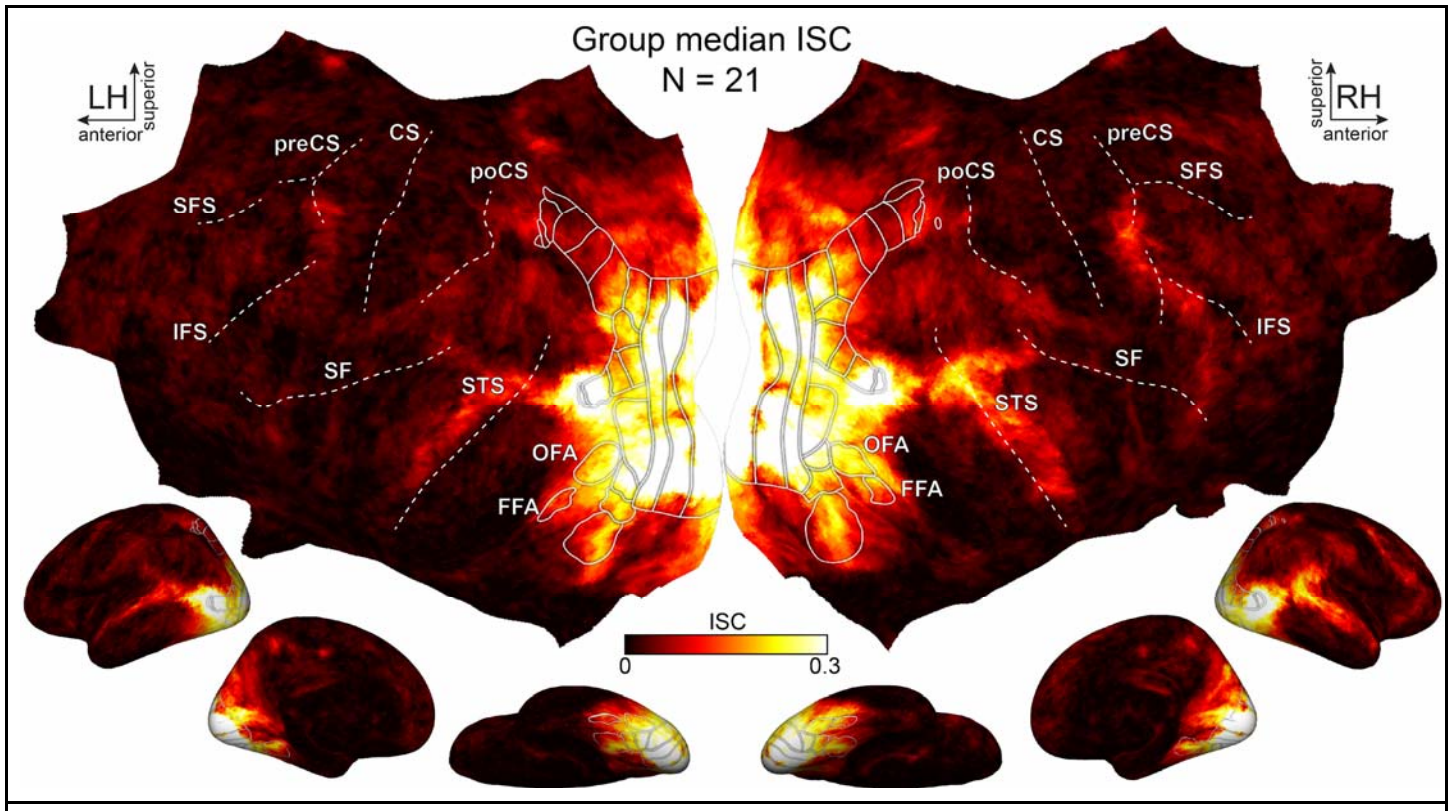


Figure 5. Group median inter-subject correlation (ISC). Flatmap and inflated surfaces show group median ISC (N=21) computed in *fsaverage6* space. White outlines indicate anatomical landmarks and visual ROIs. As expected from a dynamic visual stimulus, high ISC is found in primary visual and motion-sensitive areas. High ISC is also found in face-processing regions including the occipital face area (OFA), fusiform face area (FFA), and superior temporal sulcus (STS), as well as dorsal precuneus[48] and right inferior frontal gyrus.[14,48,49] Abbreviations: CS, central sulcus; FFA, fusiform face area; IFS, inferior frontal sulcus; OFA, occipital face area; poCS, postcentral sulcus; preCS, precentral sulcus; SF, sylvian fissure; SFS, superior frontal sulcus; STS, superior temporal sulcus.

Acknowledgments

We would like to thank Yaroslav Halchenko, Sam Nastase, Vassiki Chauhan, and the members of the Gobbini and Haxby labs for helpful discussions during the development of this project.

Author Contributions

M.V.d.O.C. designed the experiment, wrote the presentation code, collected and analyzed the data, curated the data, and wrote the manuscript. G.J. and F.M. collected additional behavioral data, generated stimulus features, analyzed the data, and provided feedback on the manuscript. M.d.V. collected, edited, and curated the stimuli. M.I.G. and J.V.H. designed the experiment, obtained funding, and provided feedback on the manuscript. All authors read and approved the manuscript.

Funding

This work was supported by NSF award #1835200 to M. Ida Gobbini and James V. Haxby.

References

1. Jack, R. E. & Schyns, P. G. The human face as a dynamic tool for social communication. *Curr. Biol.* **25**, R621–R634 (2015).

2. Haxby, J. V., Hoffman, E. A. & Gobbini, M. I. Human neural systems for face recognition and social communication. *Biol. Psychiatry* **51**, 59–67 (2002).

3. Haxby, J. V., Hoffman, E. A. & Gobbini, M. I. The distributed human neural system for face perception. *Trends Cogn. Sci.* **4**, 223–233 (2000).

4. Todorov, A., Mandisodza, A. N., Goren, A. & Hall, C. C. Inferences of competence from faces predict election outcomes. *Science* **308**, 1623–1626 (2005).

5. Todorov, A., Gobbini, M. I., Evans, K. K. & Haxby, J. V. Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia* **45**, 163–173 (2007).

6. Nastase, S. A., Goldstein, A. & Hasson, U. Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *Neuroimage* **222**, 117254 (2020).

7. Matusz, P. J., Dikker, S., Huth, A. G. & Perrodin, C. Are we ready for real-world neuroscience? *J. Cogn. Neurosci.* **31**, 327–338 (2019).

8. Wu, M. C.-K., David, S. V. & Gallant, J. L. Complete functional characterization of sensory neurons by system identification. *Annu. Rev. Neurosci.* **29**, 477–505 (2006).

9. Visconti di Oleggio Castello, M., Deniz, F., Dupré la Tour, T. & Gallant, J. L. Encoding models in functional magnetic resonance imaging: the Voxelwise Encoding Model framework. *PsyArXiv* (2025) doi:10.31234/osf.io/nt2jq_v2.

10. Chen, P. *et al.* An fMRI dataset in response to large-scale short natural dynamic facial expression videos. *Sci Data* **11**, 1247 (2024).

11. Visconti di Oleggio Castello, M., Chauhan, V., Jiahui, G. & Gobbini, M. I. An fMRI dataset in response to 'The Grand Budapest Hotel', a socially-rich, naturalistic movie. *Sci Data* **7**, 383 (2020).

12. Jiahui, G. *et al.* Modeling naturalistic face processing in humans with deep convolutional neural networks. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2304085120 (2023).

13. Visconti di Oleggio Castello, M. Characterizing feature representations in the human face-processing network with multivariate analyses and encoding models. PhD thesis, Dartmouth College (2018).

14. Visconti di Oleggio Castello, M., Haxby, J. V. & Gobbini, M. I. Shared neural codes for visual and semantic information about familiar faces in a common representational space. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2110474118 (2021).

15. Nastase, S. A., Gazzola, V., Hasson, U. & Keysers, C. Measuring shared responses across subjects using intersubject correlation. *Soc. Cogn. Affect. Neurosci.* **14**, 667–685 (2019).

16. Haxby, J. V., Connolly, A. C. & Guntupalli, J. S. Decoding neural representational spaces using multivariate pattern analysis. *Annu. Rev. Neurosci.* **37**, 435–456 (2014).

17. Kriegeskorte, N., Mur, M. & Bandettini, P. Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2**, 4 (2008).

18. Dupré la Tour, T., Visconti di Oleggio Castello, M. & Gallant, J. L. The Voxelwise Encoding Model framework: a tutorial introduction to fitting encoding models to fMRI data. *Imaging Neuroscience* **3**, imag_a_00575 (2025).

19. Naselaris, T., Kay, K. N., Nishimoto, S. & Gallant, J. L. Encoding and decoding in fMRI. *Neuroimage* **56**, 400–410 (2011).

20. Kriegeskorte, N. & Mur, M. Inverse MDS: inferring dissimilarity structure from multiple item arrangements. *Front Psychol* **3**, 245 (2012).

21. Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C. & Kanwisher, N. Differential selectivity for dynamic versus static information in face-selective cortical regions. *Neuroimage* **56**, 2356–2363 (2011).

22. Halchenko, Y. O. *et al.* HeuDiConv - flexible DICOM conversion into structured directory layouts. *J Open Source Softw* **9**, 5839 (2024).

23. Halchenko, Y. O. *et al. HeuDiConv — flexible DICOM conversion into structured directory layouts*. (Zenodo, 2025). doi:10.5281/ZENODO.17195198.

24. Visconti di Oleggio Castello, M. *et al. ReproNim/reproin*. (Zenodo, 2023). doi:10.5281/zenodo.1207117.

25. Gulban, O. F. *et al. Poldracklab/pydeface: PyDeface v2.0.2*. (Zenodo, 2022). doi:10.5281/ZENODO.6856482.

26. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116 (2019).

27. Markiewicz, C. J., Esteban, O., Goncalves, M., Poldrack, R. A. & Gorgolewski, K. J. *fMRIPrep: a robust preprocessing pipeline for functional MRI*. (Zenodo, 2026). doi:10.5281/ZENODO.18248358.

28. Gorgolewski, K. *et al.* Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python.*Front. Neuroinform.* **5**, 13 (2011).

29. Esteban, O. *et al. Nipy/nipype*. (Zenodo, 2020). doi:10.5281/ZENODO.596855.

30. Tustison, N. J. *et al.* N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**, 1310–1320 (2010).

31. Avants, B. B., Epstein, C. L., Grossman, M. & Gee, J. C. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* **12**, 26–41 (2008).

32. Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* **20**, 45–57 (2001).

33. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage* **9**, 179–194 (1999).

34. Klein, A. *et al.* Mindboggling morphometry of human brains. *PLoS Comput. Biol.* **13**, e1005350 (2017).

35. Ciric, R. *et al.* TemplateFlow: FAIR-sharing of multi-scale, multi-species brain models. *Nat Methods* **19**, 1568–1571 (2022).

36. Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R. & Collins, D. L. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage* **47**, S102 (2009).

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

38. Greve, D. N. & Fischl, B. Accurate and robust brain image alignment using boundary-based registration.

Neuroimage **48**, 63–72 (2009).

39. Power, J. D. *et al.* Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* **84**, 320–341 (2014).

40. Behzadi, Y., Restom, K., Liau, J. & Liu, T. T. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage* **37**, 90–101 (2007).

41. Satterthwaite, T. D. *et al.* An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage* **64**, 240–256 (2013).

42. Patriat, R., Reynolds, R. C. & Birn, R. M. An improved model of motion-related signal changes in fMRI. *Neuroimage* **144**, 74–82 (2017).

43. Abraham, A. *et al.* Machine learning for neuroimaging with scikit-learn. *Front. Neuroinform.* **8**, 14 (2014).

44. Esteban, O. *et al.* MRIQC: advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS One* **12**, e0184661 (2017).

45. Gorgolewski, K. J. *et al.* The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci Data* **3**, 160044 (2016).

46. Nastase, S. A. *et al.* The 'Narratives' fMRI dataset for evaluating models of naturalistic language comprehension. *Sci Data* **8**, 250 (2021).

47. Aliko, S., Huang, J., Gheorghiu, F., Meliss, S. & Skipper, J. I. A naturalistic neuroimaging database for understanding the brain using ecological stimuli. *Sci Data* **7**, 347 (2020).

48. Visconti di Oleggio Castello, M., Halchenko, Y. O., Swaroop Guntupalli, J., Gors, J. D. & Gobbini, M. I. The neural representation of personally familiar and unfamiliar faces in the distributed system for face perception. *Sci. Rep.* **7**, 12237 (2017).

49. Guntupalli, J. S., Wheeler, K. G. & Gobbini, M. I. Disentangling the representation of identity from head view along the human face processing pathway. *Cereb. Cortex* **27**, 46–53 (2017).

50. Haxby, J. V., Guntupalli, J. S., Nastase, S. A. & Feilong, M. Hyperalignment: modeling shared information encoded in idiosyncratic cortical topographies. *Elife* **9**, e56601 (2020).

51. Haxby, J. V. *et al.* A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron* **72**, 404–416 (2011).

52. Guntupalli, J. S. *et al.* A model of representational spaces in human cortex. *Cereb. Cortex* **26**, 2919–2934 (2016).

53. Feilong, M. *et al.* The individualized neural tuning model: precise and generalizable cartography of functional architecture in individual brains. *Imaging Neuroscience* **1**, 1–34 (2023).

54. van der Walt, S., Colbert, S. C. & Varoquaux, G. The NumPy array: a structure for efficient numerical computation. *Comput. Sci. Eng.* **13**, 22–30 (2011).

55. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

56. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

57. Brett, M. *et al. Nipy/nibabel*. (Zenodo, 2024). doi:10.5281/zenodo.591597.

58. Pedregosa, F. *et al.* Scikit-learn: machine learning in Python.*J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

59. Gao, J. S., Huth, A. G., Lescroart, M. D. & Gallant, J. L. Pycortex: an interactive surface visualizer for fMRI. *Front. Neuroinform.* **9**, 23 (2015).