Individual Brain Charting third release, probing brain activity during Movie Watching and Retinotopic Mapping

Ana Luísa Pinho, Hugo Richard, Michael Eickenberg, Alexis Amadon, Elvis Dohmatob, Isabelle Denghien, Juan Jesús Torre, Swetha Shankar, Himanshu Aggarwal, Ana Fernanda Ponce, et al.

Individual Brain Charting third release, probing brain activity during Movie Watching and Retinotopic Mapping

Ana Luísa Pinho[1,2,3,*], Hugo Richard[1,4], Michael Eickenberg[1,5], Alexis Amadon[6], Elvis Dohmatob[1,7], Isabelle Denghien[8], Juan Jesús Torre[1], Swetha Shankar[1], Himanshu Aggarwal[1], Ana Fernanda Ponce[1], Alexis Thual[1,8,10], Thomas Chapalain[1], Chantal Ginisty[9], Séverine Becuwe-Desmidt[9], Séverine Roger[9], Yann Lecomte[9], Valérie Berland[9], Laurence Laurier[9], Véronique Joly-Testault[9], Gaëlle Médiouni-Cloarec[9], Christine Doublé[9], Bernadette Martins[9], Gaël Varoquaux[1], Stanislas Dehaene[8,10], Lucie Hertz-Pannier[9,11], and Bertrand Thirion[1]

[1]Université Paris-Saclay, Inria, CEA, Palaiseau, 91120, France
[2]Department of Computer Science, Western University, London, Ontario, Canada
[3]Brain and Mind Institute, Western University, London, Ontario, Canada
[4]Criteo AI Labs, Paris, France
[5]Flatiron Institute, New York, USA
[6]Université Paris-Saclay, CEA, CNRS, BAOBAB, NeuroSpin, 91191, Gif-sur-Yvette, France
[7]Facebook AI Research Paris, France
[8]Cognitive Neuroimaging Unit, INSERM, CEA, Université Paris-Saclay, NeuroSpin center, 91191 Gif/Yvette, France
[9]CEA Saclay/DRF/IFJ/NeuroSpin/UNIACT, France
[10]Collège de France, Paris, France
[11]UMR 1141, NeuroDiderot, Université de Paris, France
[*]corresponding author: Ana Luísa Pinho (agrilopi@uwo.ca)

ABSTRACT

The *Individual Brain Charting* (IBC) is a multi-task functional Magnetic Resonance Imaging dataset acquired at high spatial-resolution, which is intended to facilitate the cognitive mapping of the human brain. It consists in the deep phenotyping of twelve individuals in a fixed environment, covering a broad range of psychological domains that allows, in turn, the investigation of atlasing techniques in functional neuroimaging. Here, we present the inclusion of task data from both naturalistic stimuli and trial-based designs, to uncover core structures of brain activation. We rely on the *Fast Shared Response Model* (FastSRM): an analytical tool that provides a data-driven solution to model naturalistic stimuli, typically containing many features. We show that data from left-out runs can be reconstructed using FastSRM, thus enabling the extraction of functional networks pertaining to vision, audio and language systems. We also present an in-depth study of the topographic organization of the visual system through a retinotopy task. IBC is open access: source plus derivatives imaging data and meta-data are available in public repositories.

Background & Summary

Mapping cognition in the whole human brain calls for the multi-dimensional analysis of the correlates of behavior corresponding to a wide range of psychological domains. Such analysis relies on brain-activation maps obtained from *functional Magnetic Resonance Imaging* (fMRI) that quantify the underlying neural correlates modulated by mental functions across tasks[1–5].

Generalization across task implementations can be achieved by launching data-pooling analyses, such as *meta-analysis*, which involves the aggregation of data derivatives pertaining to different tasks across publications[6–9]. However, while this framework benefits from the combination of multi-task data information, it is still susceptible to inter-subject and inter-site variability as well as loss of information due to sparse peak-coordinate representation, among other limitations[10,11]. An alternative solution is *mega-analysis*, which consists in pooling source data instead[7,12]. Because data can be treated all together and the same processing routines are applied, it mitigates the variability resulting from the aggregation of neuroimaging data from different sources[13].

Yet, mega-analyses do not eliminate inter-subject variability—both within and between tasks—making it difficult to determine whether such variability is elicited by either differences in general cognitive strategies across behavioral conditions or individual functional differences. This problem has been widely recognized in neuroimaging and affects all kinds of group-level analyses. It has been shown to undermine not only the estimation of statistical significance[14] but also the exact demarcation of functional regions according to their contribution in elementary cognitive processes[4]. In the past decade, many studies have thus started to adopt individual analysis, in order to overcome both functional and anatomical inter-subject variability[15–21].

On the other hand, cognitive neuroscience has traditionally relied upon temporal and sparsely controlled designs, using abstract stimuli[22]. Single-task fMRI experiments are usually conceived in this way, wherein experimental designs tightly control the variables and isolate targeted cognitive constructs as means to link them to the function of discrete brain regions. Because this approach has been the mainstay in cognitive neuroscience, many of the publicly available task datasets, as well as the ensuing data-pooling studies are mostly based on this type of stimuli[3,22–26]. By contrast, the study of real-life, dynamic and multimodal sensory stimuli—*aka naturalistic stimuli*—was only recently introduced in the field of cognitive neuroimaging by two seminal papers[27,28], giving rise afterwards to initiatives like the *studyforrest* dataset[29–32]. They are thought to reduce biases inherent to the choice of artificial behavioral settings, and importantly, *a priori* designed contrasts[22]. The main challenges to employ naturalistic-stimuli tasks include: *(1)* experimental implementation, as naturalistic paradigms are lengthy in duration because their high-dimensional feature space requires the collection of a larger amount of data; *(2)* statistical modelling, as the standard *General Linear Model* (GLM) applied to naturalistic stimuli leads to high-dimensional controlled-design models due to the greater amount of features extracted from the paradigms; and *(3)* unsupervised data-driven approaches are preferred because of *(2)*, but high-dimensional imaging data (many voxels) require decomposition methods with scalability.

To obtain a large sample of behavioral features and simultaneously achieve a whole-brain coverage free from inter-subject and inter-site variability, extensive functional mapping at high-spatial resolution of individual brains exposed to a comprehensive collection of task paradigms—comprising trial-based as well as naturalistic designs—is necessary. We thus present herein the third release of the *Individual Brain Charting* (IBC) dataset: a multi-task fMRI-data collection obtained from a permanent cohort of twelve participants acquired with a spatial resolution of 1.5mm. Its task-wise organization combined with a higher spatial resolution make it possible to estimate finer individual representations linked to a broad variety of mental functions[4,5]. It extends previous releases of the IBC dataset[33,34] and is focused foremost on the inclusion of naturalistic tasks[16,35] probing mainly the visual system and to some extent the auditory and language systems. To complement the coverage of the visual system, we have also included the classic retinotopy paradigms dedicated to map the polar angle and eccentricity in the visual cortex[36].

In the present paper, we provide a thorough description of the tasks taking part for this extension and their technical validation. Given the aforementioned challenges posed by the analysis of fMRI data relative to naturalistic paradigms, we give particular emphasis to showcasing the application of the *Fast Shared Response Model* (FastSRM), described in Richard&Thirion (2023)[37].

IBC is an open-access dataset consisting of high-resolution, functional maps of individual brains. It aims at providing quantitative insights about individual differences of elementary processes in cognition, by leveraging a deep behavioral-phenotyping approach. Many sessions pertaining to different tasks are thus undertaken per participant. Unlike longitudinal studies, data collection of each task only takes place once during the lifetime of the project. IBC is thus intended to serve as a source of functional correlates of various cognitive conditions, in order to support research in human neuroscience.

Methods

To avoid ambiguity with MRI-related terms, definitions follow the *Brain-Imaging-Data-Structure* (BIDS) Specification version 1.8.0[38].

Participants

The present release of the IBC dataset consists of brain fMRI data from twelve individuals (two female) acquired between April 2016 and February 2019. The experiments were carried out with the understanding and formal consent of the participants, in accordance with the Helsinki declaration and the French public health regulation.

Detailed description of the age, sex and handedness of the group is provided in Table 1. Age varied between 26 and 40 years old (median = 30 years) upon recruitment and handedness was determined with the *Edinburgh Handedness Inventory*[39]. For more information on the cohort's recruitment, consult Pinho *et al.* (2018)[33] and Pinho *et al.* (2020)[34].

Materials

Stimuli

The stimuli of the tasks (see Experimental Paradigms Section for details about tasks' paradigms) were delivered through custom-made scripts that ensured a fully automated environment and computer-controlled collection of the behavioral data.

Subject ID	Year of recruitment	Age	Sex	Handedness score
sub-01	2015	39.5	M	0.3
sub-04	2015	26.9	M	0.8
sub-05	2015	27.4	M	0.6
sub-06	2015	33.1	M	0.7
sub-07	2015	38.8	M	1
sub-08	2015	36.5	F	1
sub-09	2015	38.5	F	1
sub-11	2016	35.8	M	1
sub-12	2016	40.8	M	1
sub-13	2016	28.2	M	0.6
sub-14	2016	28.3	M	0.7
sub-15	2017	30.3	M	0.9

Table 1. Demographic data of the participants. Age stands for the participants' age upon recruitment.

All protocols were set under Python 2.7. The protocol of the Clips task was adapted from the original study[35] using standard Python libraries; the ones pertaining to the Retinotopy and Raiders tasks were respectively designed with *PsychoPy* v1.90.3[40] and *Expyriment* v0.9.0[41]. Visual stimuli of Clips consisted of the same color natural movies as described in Nishimoto *et al.* (2011)[35], whereas the audiovisual stimuli presented in the Raiders task corresponded to the 2009 DVD edition of *Raiders of the Lost Ark* dubbed to French.

These materials are available in a public GitHub repository dedicated to the behavioral protocols of the tasks featuring the IBC dataset: https://github.com/individual-brain-charting/public_protocols (consult Section Code Availability for further details about the repository).

Eye-Tracker

The video-based, eye-tracker system *EyeLink 1000 Plus* was used for the behavioral, training sessions of the Clips plus Retinotopy tasks (for more information about the training sessions, consult Section Experimental Procedure).

MRI Equipment

The fMRI data were acquired using an MRI scanner Siemens 3T Magnetom Prisma^{fit} along with a Siemens Head/Neck 64-channel coil. Behavioral responses for the Retinotopy tasks were obtained with a MR-compatible, five-button ergonomic pad (Current Designs, Package 932 with Pyka HHSC-1x5-N4) and the MRI-environment audio system for the Raiders task was set with the MR-Confon package.

All sessions were conducted at the NeuroSpin platform of the CEA Research Institute, Saclay, France.

Experimental Procedure

Upon arrival to the research institute, participants were instructed about the execution and timing of the tasks referring to the upcoming session.

Particularly, behavioral training sessions prior to the MRI sessions were conducted for the Clips and Retinotopy tasks. We stress that the center of the visual field must be approximately constant during data acquisition of these tasks, as means to obtain a consistent map of the visual system within and between individuals (consult Section Experimental Paradigms for more details). To this end, participants were prepared during the training sessions to gain perception of their eyes' movements. They were instructed to move them as little as possible while fixating toward a flickering point placed on the center of a screen, which was also displaying video scenes at the same time. These video scenes were excerpts of those presented for the Clips task. An eye-tracker was also coupled to the experimental setup of the training session. Participants were thus provided with a real-time feedback of their eyes' movements in the form of a green moving point displayed on the screen, too. The main goal of this training exercise was to keep the green moving point as close as possible to the flickering one, which was fixed. By this way, participants could then practice how to keep looking continuously toward the center of the screen—i.e. the perimetric origin—for as long as possible.

All MRI sessions were composed of several runs dedicated to one or two tasks as described in Section Experimental Paradigms. The structure of the sessions according to the MRI modality employed at every run is detailed in Table 2. Information on the imaging parameters of the referred modalities can be found in Section Imaging Data.

Session	Modality	Task	Duration of each Run* (min:sec)	Repetitions
Clips 1-3	2D Spin-Echo	-	00:31	PA(×2) + AP(×2)
	BOLD fMRI	Training set	10:50	PA(×2) + AP
	BOLD fMRI	Test set	10:50	PA + AP(×2)
Clips 4	2D Spin-Echo	-	00:31	PA(×2) + AP(×2)
	BOLD fMRI	Training set	10:50	PA(×2) + AP
	BOLD fMRI	Retinotopy wedge clock	05:30	PA(x2) + AP(x2)
	BOLD fMRI	Retinotopy wedge anti clock	05:30	PA(x2) + AP(x2)
	BOLD fMRI	Retinotopy ring expanding	05:30	PA
	BOLD fMRI	Retinotopy ring contracting	05:30	AP
Raiders 1	2D Spin-Echo	-	00:31	PA(×2) + AP(×2)
	BOLD fMRI	Chapter 1	12:28	PA
	BOLD fMRI	Chapter 2	09:53	PA
	BOLD fMRI	Chapter 3	10:28	PA
	BOLD fMRI	Chapter 4	12:37	AP
	BOLD fMRI	Chapter 5	11:33	AP
	BOLD fMRI	Chapter 6	11:32	AP
Raiders 2	2D Spin-Echo	-	00:31	PA(×2) + AP(×2)
	BOLD fMRI	Chapter 7	11:39	PA
	BOLD fMRI	Chapter 8	11:46	PA
	BOLD fMRI	Chapter 9	09:22	AP
	BOLD fMRI	Chapter 10	07:01	AP
	BOLD fMRI	Chapter 1	12:28	AP
	BOLD fMRI	Chapter 2	09:53	AP
	BOLD fMRI	Chapter 3	10:28	AP

Table 2. Plan of the MRI acquisitions for the second extension of the IBC dataset. A BOLD-fMRI run refers to the acquisition of fMRI data on one single task. At least, there were two BOLD runs, corresponding to PA- and AP-phase-encoding directions for each task during a session. The 2D Spin-Echo PA/AP volumes were always acquired before the runs dedicated to the collection of BOLD-fMRI data and repeated afterwards.

* For BOLD fMRI sequences, the durations herein presented account only for the period of the actual acquisition, included in the data volumes. The full duration of each run also included ∼45s of dummy scans, always performed at their beginning.

Experimental Paradigms

Tasks were aggregated in different sessions according to the Table 2. The following sections intend to provide a description of the paradigms employed for each task. Materials used for stimulus presentation (see Section Stimuli) have been made publicly available, together with video annotations of the corresponding protocols, on https://github.com/individual-brain-charting/public_protocols (see Section Code Availability for a more comprehensive description of the repository).

Clips *task*

The *Clips* task stands for a reproduction of the study reported in Nishimoto *et al.* (2011)[35], in which participants were to visualize naturalistic scenes edited as video clips of ten and a half minutes each.

Each run was always dedicated to the data collection of one video clip at a time. As in the original study, runs were grouped in two categories: training and test. Scenes from the training category were shown only once. Contrariwise, scenes from the test category were composed of approximately one-minute-long excerpts extracted from the clips presented during training. Excerpts were concatenated to construct the sequence of every test run; each sequence was predetermined by randomly permuting many excerpts that were repeated ten times each across all runs. The same randomized sequences, employed across test runs, were used to collect data from all participants.

There were twelve and nine runs dedicated to the collection of training data and test data, respectively. Data from nine runs of each category were interspersedly acquired in three full sessions; the three remaining runs devoted to train-data collection were acquired in half of one last session, before the Retinotopy tasks (see Section *Retinotopy* tasks for complete description of these tasks).

To assure the same topographic reference of the visual field for all participants, a colored fixation point was always presented at the center of the images. Such point was changing three times per second to ensure that it was visible regardless the color of the movie. To account for stabilization of the BOLD signal, ten extra seconds of acquisition were added at the beginning and end of every run. The total duration of each run was thus ten minutes and fifty seconds.

Retinotopy *tasks*

The *Retinotopy* tasks refer to the classic retinotopic paradigms—i.e. the *Wedge* and the *Ring* tasks—consisting of four kinds of visual stimuli: *(1-2)* a slowly rotating clockwise or counterclockwise, semicircular checkerboard stimulus, as part of the Wedge task; and *(3-4)* a thick, dilating or contracting ring, as part of the Ring task. The phase of the periodic response at the rotation or dilation/contraction fundamental frequency measured at each voxel relates to the measurement of the perimetric parameters concerning *polar angle* and *eccentricity*, respectively[36].

In the present study, six runs were devoted to this task. Each of them were five-and-a-half minutes long. They were programmed for the same session following the last three "training-data" runs of the Clips task (see Section *Clips* task for complete description of this task.) Four runs were dedicated to the presentation of the rotating checkerboard stimulus (two runs for each direction) and the remaining two were dedicated to the dilating or contracting ring, one at a time.

Similarly to the Clips task, a point was displayed at the center of the visual stimulus in order to keep constant the perimetric origin in all participants. Participants were thus to fixate continuously this point whose color flickered throughout the entire run. To keep the participants engaged in the task, they were instructed that, at the end of every run (i.e. after MRI acquisition was finished), they would be asked which color had most often been presented. They had to select one of the four possible options—i.e. red, green, blue or yellow—by pressing on the corresponding button in the response box.

Raiders *task*

The *Raiders* task was adapted from Haxby *et al.* (2011)[16], in which the full-length action movie *Raiders of the Lost Ark* was presented to the participants. The main goal of the original study was the estimation of the hyperalignment parameters that transform voxel space of functional data into feature space of brain responses, linked to the visual characteristics of the movie displayed.

Similarly, herein, the movie was shown to the IBC participants in contiguous runs determined according to the chapters of the movie defined in the DVD (see Section Stimuli for details about the DVD edition).

This task was completed in two sessions. In order to use the acquired fMRI data in train-test split and cross-validation experiments, we performed three extra-runs at the end of the second session in which the three first chapters of the movie were repeated.

To account for stabilization of the BOLD signal, ten seconds of acquisition were added at the end of the run. The structure of both sessions and duration of their runs are detailed in Table 2.

Data Acquisition

Data across participants were acquired throughout four MRI sessions, whose structure is described in Table 2. Deviations from this structure was only registered for participant 8 (sub-08), whose data pertaining to Run #13 of Raiders task was acquired in

the session dedicated to the *Theory-of-Mind* and *Pain Matrices* task battery from the second release of the IBC dataset (for more details about this release, consult Pinho *et al.* (2020)[34]).

Behavioral Data

Behavioral data from the training sessions for the Clips and Retinotopy tasks were recorded throughout four sessions.

Scores were obtained according to the response accuracy of the participant for a given trial. They indicate how close participant's sight was from the center of the screen during the trial. Each session was composed of four different trials. Therefore, we collected a maximum of sixteen scores per participant across all sessions. They are presented as Supplementary Material. We denote that each participant performed at least one training session. Nevertheless, the completion of the four sessions was not compulsory, because there was a trade-off between extensive training and quality of performance. Since participants could experience some fatigue while fixating their eyes for a continuous period of time, we thus recommended carrying out at least one training session (for more information about how training sessions were conducted, consult Section Experimental Procedure).

Additionally, participants were also asked to provide a button-press response at the end of every run concerned with the Retinotopy tasks (for more details about these tasks, consult Section *Retinotopy* tasks). The registry of these behavioral data was held in log files generated by the corresponding stimulus-delivery software.

Imaging Data

FMRI data were collected using a *Gradient-Echo* (GE) pulse, whole-brain *Multi-Band* (MB) accelerated[42,43] *Echo-Planar Imaging* (EPI) T2*-weighted sequence with *Blood-Oxygenation-Level-Dependent* (BOLD) contrasts, using the following parameters: the repetition time (TR) is 2000 ms; the echo time (TE) is 27 ms; the flip angle is 74°; the field-of-view (FOV) is $192 \times 192 \times 140$ mm^3; voxel size is $1.5 \times 1.5 \times 1.5$ mm^3; the slice orientation is axial; slices are acquired in interleaved fashion; in-plane acquisitions were accelerated by a factor (GRAPPA) of 2; and across slices, a multi-band factor of 3 was used. Two different acquisitions for the same task were always performed using two opposite phase-encoding directions: one from *Posterior to Anterior* (PA) and the other from *Anterior to Posterior* (AP). The main purpose was to mitigate geometrical distortions while assuring built-in, within-subject replication of the same tasks.

Spin-Echo (SE) EPI-2D image volumes were acquired in order to correct for spatial distortions, using the following parameters: a TR of 7680 ms; a TE of 46 ms; a FOV of $192 \times 192 \times 140$ mm^3; a voxel size of $1.5 \times 1.5 \times 1.5$ mm^3; axial slice orientation; and acceleration factor (GRAPPA) = 2. Similarly to the GE-EPI sequences, two different acquisitions were also performed using PA and AP phase-encoding directions.

In addition, a *3D magnetization-prepared rapid gradient-echo* (MP-RAGE) T1-weighted anatomical-image volume, covering the whole brain, was acquired with the following parameters: voxel size of $1 \times 1 \times 1$ mm^3; sagittal slice orientation; flip angle of 9°; and FOV of $256 \times 256 \times 160$ mm.

A detailed description of the imaging parameters set for each MRI modality is available in Pinho *et al.* (2018)[33] and in the IBC-dataset documentation: https://individual-brain-charting.github.io/docs/mri_acquisitions.html.

Imaging-Data Analysis

Prior to any neuroimaging-data analysis, the MRI DICOM images were converted to NIfTI format using the *DCM2NII* tool, which is available on https://www.nitrc.org/projects/dcm2nii/. Conversion to NIfTI format also included a full anonymization of the data, i.e. pseudonyms were removed and images were defaced using the Freesurfer-6.0.0 library[44].

Preprocessing

All GE-EPI volumes were collected twice with reversed phase-encoding directions, resulting in pairs of images with distortions going in opposite directions. Susceptibility-induced off-resonance field was estimated from the two SE-EPI volumes, which were collected twice and also using reversed phase-encoding directions. The GE-EPI images were then corrected based on the corresponding deformation model, which was computed using the FSL implementation as described in Andersson *et al.* (2003)[45] and Smith *et al.* (2004)[46].

Source data were then preprocessed using the same pipeline as described in Pinho *et al.* (2018)[33] and Pinho *et al.* (2020)[34], which relies on the *PyPreprocess* library: https://github.com/neurospin/pypreprocess. Pypreprocess stands for a collection of Python scripts—built upon the *Nipype* interface[47]—which are oriented toward a common workflow of fMRI-data preprocessing analysis; it uses precompiled modules of both *SPM12* software package (Wellcome Department of Imaging Neuroscience, London, UK) and *FSL* library (Analysis Group, FMRIB, Oxford, UK) v6.0.0.

GE-EPI volumes of each participant were then aligned among them. To this end, a rigid body transformation was employed, wherein the average volume of all images was used as main reference[48].

The T1-weighted MPRAGE volume and the aligned GE-EPI volumes were then given as input to *FreeSurfer* v6.0.0, in order to extract the surface meshes of the tissue interfaces and a sampling of the functional activation on these meshes, as described in vanEssen *et al.* (2012)[49]. These functional activations were then resampled onto the fsaverage7 template of FreeSurfer[50].

Naturalistic-Data Analysis

Naturalistic stimuli typically imply a profuse amount of data descriptors, which leads to high-dimensional design matrices[51,52]. A data-driven approach was thus employed herein as means to inspect the effects-of-interest elicited by the fMRI data that were acquired for this type of behavioral paradigms.

Because brains exposed to the same stimuli exhibit synchronous activity[28], a shared response can be obtained across different individuals. To this end, we used the FastSRM implementation for fMRI data—as described in Richard&Thirion (2023)[37] —to analyze the preprocessed imaging data of Clips and Raiders and extract a common subjects' response to these tasks together with their individual spatial components.

Formally, consider GE-EPI data preprocessed on the surface (see Section Preprocessing) for $n = 12$ subjects. The brain image at timeframe r, $\mathbf{X}_n^r \in \mathbb{R}^v, r \in 1, \cdots, R$ is modeled for subject n, $n \in 1, \cdots, N$, as a weighted sum of k spatial components stored in $\mathbf{W}_s \in \mathbb{R}^{k \times v}$, wherein v is the number of vertices in the cortical surface mesh. Generalizing for all subjects and time frames, the Deterministic *Shared Response Model* (SRM)[53] is defined by:

$$\mathbf{X}_n = \mathbf{S}\mathbf{W}_n + \mathbf{E}_n \text{ with } \quad \mathbf{W}_n\mathbf{W}_n^\top = \mathbf{I}_k \tag{1}$$

in which $\mathbf{X}_n \in \mathbb{R}^{R \times v}$ is the concatenation of R brain images with v vertices from subject n; $\mathbf{S} \in \mathbb{R}^{R \times k}$ is the concatenation of the weights for every spatial component and it is denoted as the *shared response*; the $\mathbf{W}_n \in \mathbb{R}^{k \times v}$ is the concatenation of k spatial components with v vertices for subject n; $\mathbf{E}_n \in \mathbb{R}^{R \times v}$ is a Gaussian additive noise and $\mathbf{I}_k \in \mathbb{R}^{k \times k}$ is the identity matrix. By analogy with the GLM framework, typically employed in fMRI analysis, the resulting shared response herein obtained is equivalent to the design matrix, while the spatial components are naturally seen as parameter estimates. To estimate $\mathbf{S}$ and $(\mathbf{W})_{n \in \{1,\ldots,N\}}$, one shall consider the following minimization problem:

$$\text{argmin}_{(W_n)_{n=1}^N, \mathbf{S}} \sum_{n=1}^{N} \|\mathbf{X}_n - \mathbf{S}\mathbf{W}_n\|^2 \text{ such that } \forall n \in \{1, \ldots, N\}, W_n^\top W_n = \mathbf{I}_k. \tag{2}$$

This optimization problem is solved by alternate minimization. Solving with respect to $\mathbf{W}_n$ while the other quantities are fixed, it yields the following updates:

$$\mathbf{W}_n = \mathbf{U}_n\mathbf{V}_n \text{ where } \mathbf{U}_n, \mathbf{D}_n, \mathbf{V}_n = \text{SVD}(\mathbf{S}^T\mathbf{X}_n) \tag{3}$$

where SVD stands for *Singular Value Decomposition*.

Solving with respect to $\mathbf{S}$, while other quantities are fixed, it yields the following update:

$$\mathbf{S} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{X}_n\mathbf{W}_n^T. \tag{4}$$

Updates in equations (3) and (4) are performed alternatively until the gradient of the objective in equation (2) is small enough.

The FastSRM is, in turn, a memory-efficient computational algorithm of the SRM, especially suitable for datasets with large v. Because the IBC dataset was acquired at high-spatial resolution (see Sections Background & Summary and Imaging Data for more details), v is on the order of 10^5. The method works by applying a *Principal Component Analysis* (PCA) to each subject with a number of components set to the number of timeframes R in the dataset (in the IBC dataset, R is on the order of 10^3). Since R is much smaller than v, the reduced data $\widehat{\mathbf{X}}_n \in \mathbb{R}^{R \times R}$ are much smaller in size. We then estimate $\mathbf{W}_n$ and S replacing $\mathbf{X}_n$ by $\widehat{\mathbf{X}}_n$ in equations (3) and (4). As proved by Richard&Thirion (2023)[37], the estimate $\widehat{\mathbf{S}}$ of $\mathbf{S}$ we obtain using the reduced data is the same as if the full data were used.

For the present study, the number of iterations used to fit the data in the FastSRM model was 10^4. After, we recover the spatial components $\mathbf{W}_n$ for every subject n through orthonormal regression—as defined by (3)—using $\widehat{\mathbf{S}}$ and the original data $\mathbf{X}_n$.

Within this framework, we can extract the subject-specific effects-of-interest elicited by the tasks through the individual spatial components $\mathbf{W}_n$. To be able to statistically validate these findings, we set an analysis based on *Cross Validation* (CV), in which we assessed whether reconstructed individual data of one run predicted the original individual data for that run. As referred in Table 2, data per participant and for one task were collected in different runs during one or more sessions and, consequently, data from one run pertain only to one task. Moreover, Raiders and Clips data were analyzed separately. By this way, this prediction framework is task specific.

Concretely, it consists in a double K-fold cross validation based on *co-smoothing*[54], wherein $K = 3$ for $N = 12$ subjects and $K = 2$ for R runs. The total number of runs per task is different. For Raiders, $R = 13$ and, therefore, the size of the two folds are six and seven. For Clips, data were split in terms of number of runs for training and test, which were predefined according to

the acquisition design of the task (see *Clips* task for more details); the size of the training and test sets are thus twelve and nine, respectively. This CV procedure served therefore as a scoring algorithm. It computes the correlation coefficient between predicted brain-activity maps of test runs and subjects —which were obtained from the shared response and $k = 20$ spatial components learnt on the training runs— and the corresponding original data. The resulting vertex-wise correlation coefficients represent a similarity measure of the shared content between subjects. This procedure is depicted in further detail on Figure 1.

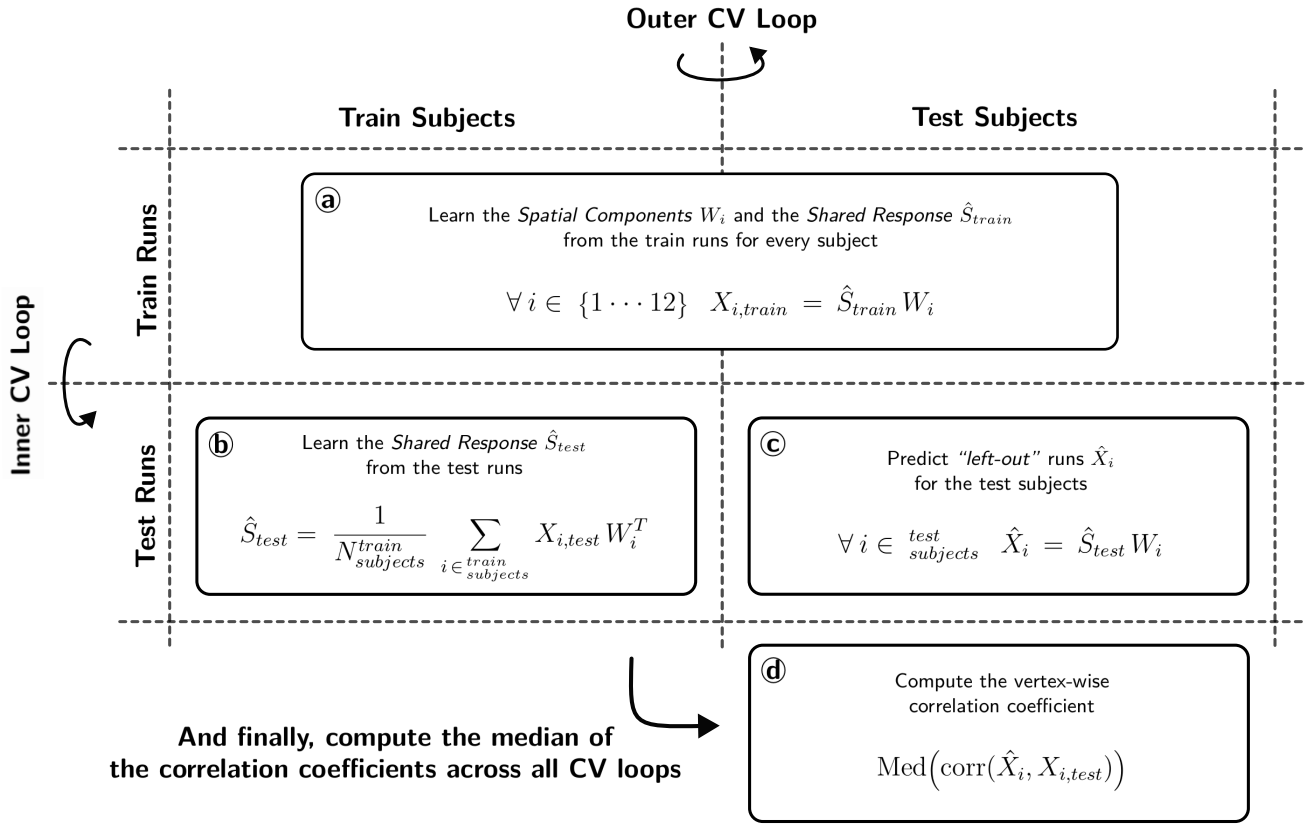

Figure 1. Description of the co-smoothing procedure to compute the jointly activated brain areas using FastSRM. The algorithm runs two nested CV loops. One first outer loop executes a random split of the cohort of twelve participants into a train set and a test set, respectively composed of eight and four subjects. One inner loop executes a random split of the group of runs into a train set and a test set: respectively twelve and nine for *Clips* and, interchangeably seven or six for *Raiders*. For every turn of the nested CV loops: (a) the $k = 20$ spatial components specific to each subject on the train runs are computed through alternate minimization (see equations 3 and 4) together with their shared response, which is then used to compute the individual components of test subjects on the same runs; then, (b) assuming that the same features of the train runs will be found on the test runs, we fit the individual responses of the train subjects on the test runs in order to compute their shared response; (c) test runs are then predicted through their shared response computed in (b); and, (d) the vertex-wise correlation between the predicted runs and the corresponding original data is computed. For every subject, the vertex-wise median of the correlations was estimated across runs within each split-half set. The vertex-wise median of the correlations across all subjects was then estimated from the individual median correlations. This final coefficient represents the similarity of activated regions across subjects for each task.

To compute the group-level significance of these individual estimates, we performed a mass-univariate group analysis with permuted *Ordinary Least Squares* (OLS) using a sign flipping permutation scheme.

We also assessed what regions are significantly different in the performance of Raiders *versus* Clips by computing a two-sided Paired $t-$test in every vertex, between the individual correlation coefficients of the two tasks. In order to obtain a precise labeling of the functional regions covered by these results, we computed the proportion of significant vertices present in the areas and regions derived from the cortical parcellation of Glasser *et al.* (2016)[55]. This estimation was performed using the projection of the HCP-MMP1.0 parcellation onto the fsaverage7 template, which is available on Mills (2016)[56].

The FastSRM encoding analysis was implemented using the *IdentifiableFastSRM* module of the FastSRM package that can

be found on: https://github.com/hugorichard/FastSRM. Statistical analyses and plotting were performed using *SciPy* v1.6.1 (https://www.scipy.org/) and *Nilearn* v0.9.2[57] (https://nilearn.github.io/).

Retinotopic-Data Analysis

The retinotopic mapping data (rotating wedge and expanding/shrinking rings) were analyzed using the standard frequency-domain analysis described in Sereno *et al.* (1995)[36] and Warnking *et al.* (2002)[58]: sine/cosine regressors were specified at the frequency of the periodic stimulus position change ($\frac{1}{32}Hz$). The magnitude of the reponse of the voxels was estimated across six sessions, and tested for significance using an $F-$test, thresholded at $p < 0.001$ and uncorrected for multiple comparisons. The phase of the reponse in each voxel with a significant effect was then determined for each session, by comparing the relative magnitude of the sine and cosine regressors. The phase information was combined for wedge- and ring-stimuli separately, in order to cancel the hemodynamic-induced phase delay. This is possible because the stimuli were presented in opposite motions (clockwise *versus* anti-clockwise for the wedge and expanding *versus* contracting for the ring). In addition, results were averaged across replications (wedge experiments). The phase estimate therefore defines in polar coordinates (eccentricity and polar angle) the visual field position that elicits a maximal amount of activation in each voxel.

All these steps were computed on the data sampled on the *fsaverage7* template using the Freesurfer software[50]. For visualization, the surface-based polan angle and eccentricity maps were displayed using the flat representation available through the Pycortex tool[59].

Data Records

The online access of source data is assured by the *Human Brain Project* (HBP) EBRAINS platform[60] as well as the OpenNeuro public repository[61] under the accession number *ds*002685[62]. The NIfTI files as well as paradigm descriptors and imaging parameters are organized per run for subject and each session, according to BIDS Specification. For more details, consult 'Data Records' sections of Pinho *et al.* (2018)[33] and Pinho *et al.* (2020)[34]: the data descriptors of the IBC first and second releases, respectively.

The individual and unthresholded z-maps, obtained from the contrast maps of the experimental conditions concerned with the *Retinotopy* (see Section Retinotopy Study) tasks, can be found in the *NeuroVault* repository[63], under the collection with the id=6618[64].

Technical Validation

Behavioral results were obtained through an elementary assessment of curated behavioral data. They are reported in Section Behavioral Results.

All imaging results were obtained following the methodological procedures, as presented in Sections Naturalistic-Data Analysis and Retinotopic-Data Analysis, applied to task-fMRI data previously preprocessed on the surface, as described in Section Preprocessing. They are respectively reported in Sections FastSRM-Encoding Study and Retinotopy Study.

Behavioral Results

In the Retinotopy tasks, participants were to give a button-press response at the end of every run corresponding to the color of the flickering fixation point they saw most times. There was only one correct answer out of the four possible options provided for each task run, which was fixed across participants. Table 1 displays the individual response accuracy achieved for all task runs. These scores are presented as percentages of the individual correct responses with respect to the total number of (correct) responses; because there were six runs dedicated to the Retinotopy tasks (see Table 2), six responses were obtained. The average $\pm$ standard deviation of the response accuracy across participants (excluding participant 5) are $56 \pm 20\%$, i.e. higher than chance level (25%). These results show that overall participants' fixation was good enough to perform a discrimination task during the course of the run, thus suggesting that fixation was held properly.

FastSRM-Encoding Study

Figure 2 – a shows the t-tests' results of the measures of performance across subjects of the FastSRM algorithm in terms of pearson-correlation coefficients (see Section Naturalistic-Data Analysis and Figure 1 for details), which were obtained between predicted and original data. These results yield functional regions that are synchronously activated across subjects for each task, because predicted data is herein estimated from a shared response to the same stimuli—learnt from the model—which is assumed to be observed in all individuals.

Higher significance was obtained throughout occipital areas for both Raiders (Figure 2 – a-left) and Clips (Figure 2 – a-right) tasks. Besides, same results were also attained in the *Inferior Parietal Lobe* and posterior areas of the *Superior Temporal Sulcus* only for the Raiders task. Overall, these results suggest a prominent recruitment of the visual system during performance

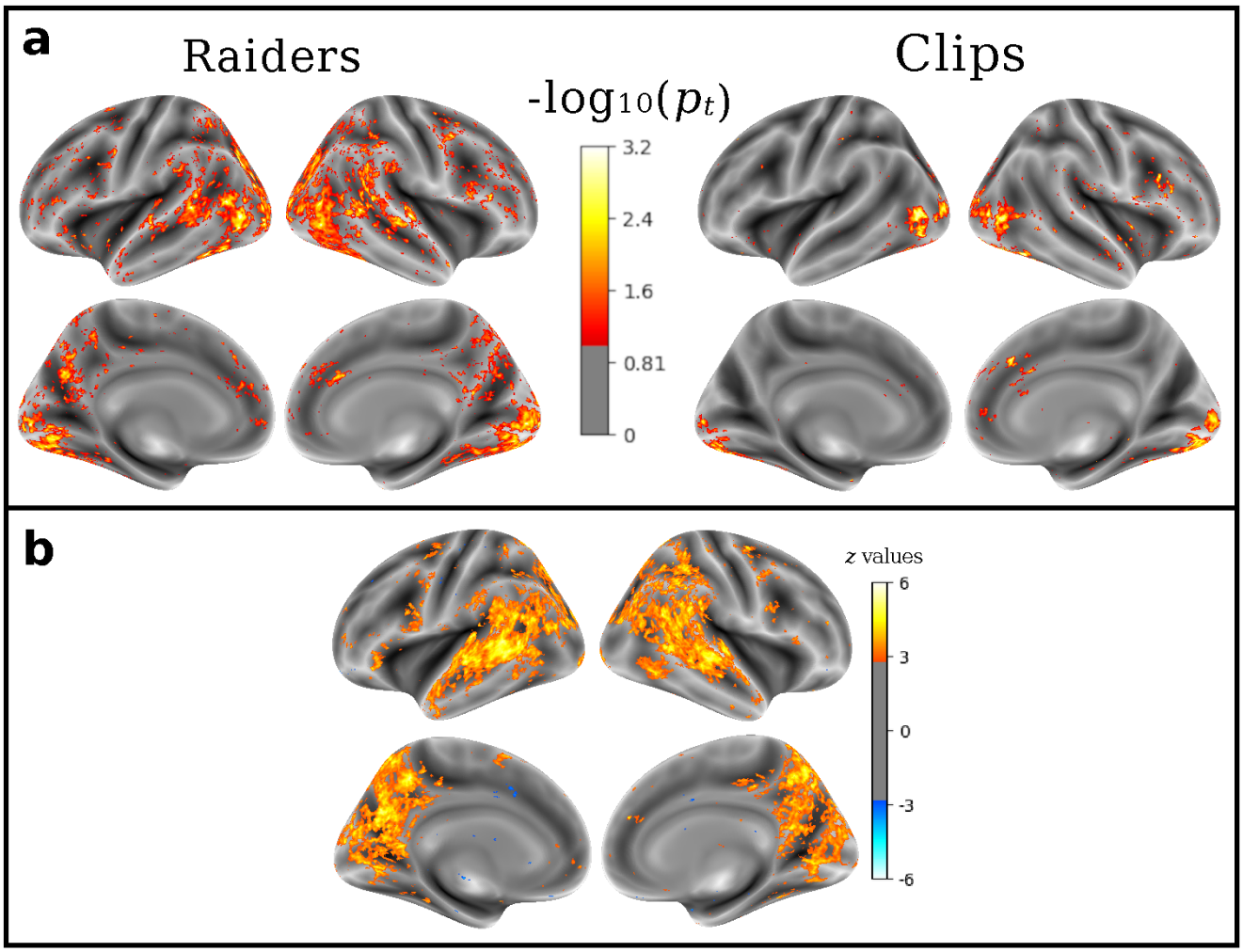

Figure 2. Group-level brain activation per task and between tasks. (a) Negative $\log_{10}$ p-values associated with the significance one-sample group-level test accuracy prediction in the reconstruction task. Pearson-correlation coefficient obtained for every vertex from a double $K-$fold cross-validation experiment across subjects (K=3) and runs (K=2) of *(left)* Raiders task and *(right)* Clips task. Data of test subjects performing test runs were reconstructed from the projection of the shared response of train subjects while performing test runs onto the individual components of test subjects while performing train runs. Predictions between original and reconstructed data were performed for every subject and run. To obtain the group-level estimation of the coefficient and associated $p-$value at every vertex, the vertex-wise median of the coefficients and combined $p-$values were first taken within split-halfs and, second, between split-halfs for every subject. To assess the group-level significance of these estimates at every vertex, we computed a mass-univariate analysis with permuted OLS and took the maximum *t-score* across permutations. Associated family-wise corrected p-values of permutations across vertices were estimated and corresponding negative $\log_{10}$ p-values thresholded to 0.1 are displayed on the panel. **(b)** Group-level $z-$maps displaying brain activation significantly different between Raiders and Clips tasks. Statistical results were determined through a vertex-wise paired $t-$test between the individual pearson-correlation coefficients of the two tasks and standardized afterwards. Statistical significance was assessed using an FDR-corrected threshold $q = 0.05$. Clusters depicted by the orange/yellow scale represent brain activation significantly higher for Raiders than Clips; conversely, clusters depicted by the dark/light blue represent brain activation significantly higher for Clips than Raiders. One can clearly observe that the orange-yellow clusters surpass in number and size the blue clusters. These results thus highlight the predominant recruitment of extra brain regions during performance of Raiders, which relate to the additional cognitive domains pertaining to auditory and language comprehension that are involved in this task.

Subject ID	Response Accuracy (%)
sub-01	83
sub-04	67
sub-05	-
sub-06	50
sub-07	33
sub-08	67
sub-09	67
sub-11	67
sub-12	33
sub-13	33
sub-14	33
sub-15	83
Chance level: 25 %	

Supplementary Table 1. Response accuracy (%) of behavioral performance for the Retinotopy tasks. Scores were estimated based on the correct answers of each participant across the six Retinotopy runs, i.e. comprising both the Wedge and the Ring tasks. The absence of response accuracy for subject 05 relate to loss of behavioral data due to a transient malfunctioning of the equipment; we stress that this issue refers to a misregistration of the data in the log files generated by the stimulus-delivery software and, thus, agnostic to subjects' performance.

of both Raiders and Clips tasks and an additional recruitment of the auditory and language-comprehension systems during performance of the Raiders task. They also reflect the main behavioral differences underlying these two tasks, i.e. they highlight the fact that while both tasks refer to naturalistic visual stimuli, only Raiders refers to naturalistic audio stimuli.

To obtain a clear distinction of the regions exhibiting greater contributions for Raiders than Clips and vice-versa, we further inspected the $z-$maps from Figure 2 – b depicting clusters whose difference of activations between the two tasks is significant. Results massively display regions in which the magnitude of their activation is significantly higher for Raiders than Clips, whereas the amount of regions showing results in the opposite direction is residual. The identification of the functional territories covered by the clusters was determined in agreement with the parcellation of Glasser *et al.* (2016)[55], which comprises 180 neocortical areas that are subsequently grouped in 22 main regions. Supplementary Table 2 presents, by descending order, the list of areas that contain a proportion of significant vertices larger than 5% in both hemispheres. The correspondence between area and main region was established according to the primary section in which the given area is described by the cortical parcellation (for further details, consult Table 1 of the *Neuroanatomical Supplementary Results of* Glasser *et al.* (2016)[55]). We identified 94 areas belonging to 21 main regions displaying higher activation for Raiders than Clips; no areas displaying higher activation for Clips than Raiders were observed above the same threshold. These results confirm the recruitment of additional brain networks necessary in the performance of cognitive tasks—namely auditory, speech and narrative comprehension—that are present in Raiders but not in Clips (see Section Experimental Paradigms for more details). Contributions to these networks come largely from the *Auditory Association Cortex*, *Early Auditory Cortex*, *Temporo-Parieto-Occipital Junction*, *Posterior Cingulate Cortex*, *Parietal Cortex* and *Lateral Temporal Cortex* (see Supplementary Table 2).

Region Name	**Area Name**	**Proportion (%) of significant vertices in the left hemisphere**	**Proportion (%) of significant vertices in the right hemisphere**
Auditory Association Cortex	Area STSd posterior	99	93
	Auditory 5 Complex	90	88
	Area STSv posterior	78	61
	Auditory 4 Complex	76	48
	Area STSd anterior	64	25
	Area STSv anterior	58	9
	Area STGa	31	24
Temporo-Parieto-Occipital Junction	Area TemporoParietoOccipital Junction 1	90	95
	Superior Temporal Visual Area	89	84
	PeriSylvian Language Area	77	70
	Area TemporoParietoOccipital Junction 3	70	67
	Area TemporoParietoOccipital Junction 2	74	54

Posterior Cingulate Cortex	Area 7m	92	90
	Parieto-Occipital Sulcus Area 1	89	65
	PreCuneus Visual Area	82	70
	Area 31pd	70	68
	Area ventral 23 a+b	64	70
	Parieto-Occipital Sulcus Area 2	66	62
	Area 31p ventral	73	34
	Dorsal Transitional Visual Area	65	41
	Area dorsal 23 a+b	28	26
	Area 31a	29	26
	ProStriate Area	34	12
	RetroSplenial Complex	9	12
Superior Parietal Cortex	Medial Area 7P	88	92
	Lateral Area 7P	77	90
	Medial Area 7A	80	68
	Medial IntraParietal Area	72	53
	Anterior IntraParietal Area	4	78
	Area Lateral IntraParietal dorsal	16	56
	Area 7PC	20	43
	Lateral Area 7A	27	27
	Ventral IntraParietal Complex	22	23
	Area Lateral IntraParietal ventral	8	13
Inferior Parietal Cortex	Area PGi	74	78
	Area PGp	56	66
	Area IntraParietal 0	72	49
	Area PGs	45	53
	Area IntraParietal 1	37	49
	Area PFm Complex	20	62
	Area IntraParietal 2	24	57
	Area PF Complex	32	43
	Area PF opercular	31	4
	Area PFt	2	26
Early Auditory Cortex	Lateral Belt Complex	78	48
	ParaBelt Complex	61	43
	Primary Auditory Cortex	10	27
	Area PFcm	20	16
	RetroInsular Cortex	10	4
Dorsal Stream Visual Cortex	Sixth Visual Area	62	60
	Area V3A	56	37
	IntraParietal Sulcus Area 1	64	21
	Area V3B	49	19
	Area V6A	13	28
	Seventh Visual Area	24	0
Lateral Temporal Cortex	Area PHT	65	54
	Area TE1 anterior	16	2
	Area TE1 posterior	13	6
	Area TG dorsal	11	4
	Area TE1 Middle	13	0
	Area TG Ventral	9	0
MT+Complex and Neighboring Visual Areas	Area FST	47	56
	Area V3CD	43	29
	Area PH	12	21
	Medial Superior Temporal Area	3	22
	Area Lateral Occipital 3	20	3
Primary Visual Cortex (V1)	Primary Visual Cortex	45	38
Early Visual Cortex	Second Visual Area	31	27
	Third Visual Area	18	23
	Fourth Visual Area	7	5
Premotor Cortex	Rostral Area 6	42	14
	Area 55b	20	32
	Area 6 anterior	18	23
	Premotor Eye Field	30	4
	Frontal Eye Fields	11	13
	Dorsal area 6	3	12
	Ventral Area 6	5	4
Paracentral Lobular and Mid Cingulate Cortex	Area 23c	16	26
	Supplementary and Cingulate Eye Field	10	0
Inferior Frontal Cortex	Area 45	25	6
	Area IFJp	17	12

	Area 44	16	13
	Area 47l (47 lateral)	21	0
Ventral Stream Visual Cortex	VentroMedial Visual Area 3	8	22
	VentroMedial Visual Area 2	11	4
Medial Temporal Cortex	ParaHippocampal Area 3	9	5
	ParaHippocampal Area 2	8	2
Dorsolateral Prefrontal Cortex	Inferior 6-8 Transitional Area	3	14
	Superior Frontal Language Area	10	3
	Area anterior 9-46v	11	0
Somatosensory and Motor Cortex	Area 2	1	10
Posterior Opercular Cortex	Area OP4/PV	12	0
Insular and Frontal Opercular Cortex	Area Frontal Opercular 5	12	0
Anterior Cingulate and Medial Prefrontal Cortex	Area p32	7	2

Supplementary Table 2. Proportions of vertices within brain areas, at each hemisphere, whose activation significantly differs between Raiders and Clips tasks. Only areas with a proportion of significant vertices $\geq 5\%$, in both hemispheres taken together, are listed in the table. Names of areas and regions as well as their correspondence were obtained from the cortical parcellation of Glasser *et al.* (2016)[55]. We note that only areas displaying higher magnitude of activation for Raiders than Clips are reported in the table, since no areas with $\geq 5\%$ of significant vertices (in both hemispheres) were found for the opposite direction.

Retinotopy Study

Figure 3 shows the retinotopic organization of the visual field in the human brain elicited by the Retinotopy tasks (see Sections Materials, Experimental Procedure, *Retinotopy* tasks for further details about the implementation of these classic retinotopy paradigms). The topographic projection to the V1-4 brain areas of the top-down and left-right reversed representation in the retina of the visual stimuli is mapped for every participant. As in Sereno *et al.* (1995)[36], the polar angle of this projection is obtained from conditions pertaining to the Wedge task *(left column)*, whereas its eccentricity is obtained from those pertaining to the Ring task *(right column)*. Overall, one can clearly notice a consistent spatial encoding of the visual field through these polar coordinates across all individuals.

Usage Notes

Our results show that functional-imaging data featuring the hthird release of the IBC dataset reflect response to behavior during performance of the corresponding tasks. We also show the feasibility of extracting the same type of data derivatives —i.e. contrast maps— from tasks pertaining to different types of experimental designs. Concretely, we demonstrate that cognitive networks of functional data collected from naturalistic paradigms can be extracted using FastSRM —an unsupervised data-driven method— without explicitly model features of the stimuli. This is particularly useful toward computational efficiency in high-dimensional regimes thanks to various shortcuts described in[37]. In addition, we demonstrate that results obtained for every task—which are adapted from previous studies— are in agreement with the ones originally reported.

The collection of new data continues till Fall 2023 and, thus, a final release is expected in 2024. Tasks featuring these releases will comprise not only other sensory modalities in greater depth but also high-order cognitive modules that will complement those from past releases. For instance, we plan to attain a better coverage of the auditory system with the inclusion of tasks on tonotopy, auditory language comprehension and listening of naturalistic sounds. Other tasks on biological motion, motor inhibition, finger tapping, visual perception (e.g. color, scenes and faces), stimulus salience, working memory, emotional memory, spatial navigation, risk-based decision making, reward processing, language and arithmetic processing will also integrate these future releases.

Although the IBC dataset is dedicated foremost to task-fMRI data, future releases will be also dedicated to resting-state fMRI data as well as to other MRI modalities, concretely high-resolution T1- and T2-weighted, diffusion-weighted and myelin water fraction.

The official website of the IBC dataset (https://individual-brain-charting.github.io/docs/) can be consulted anytime for a continuous update about its releases.

Code Availability

Metadata concerning the stimuli presented during the BOLD fMRI runs are publicly available at https://github.com/individual-brain-charting/public_protocols. They include: *(1)* the task-stimuli protocols; *(2)* demo presentations of the tasks as video annotations; *(3)* instructions to the participants; and *(4)* scripts to extract paradigm descriptors from log files for the GLM estimation. Regarding the task-stimuli protocols, Clips consist in a reproduction of the protocol featuring the original study,

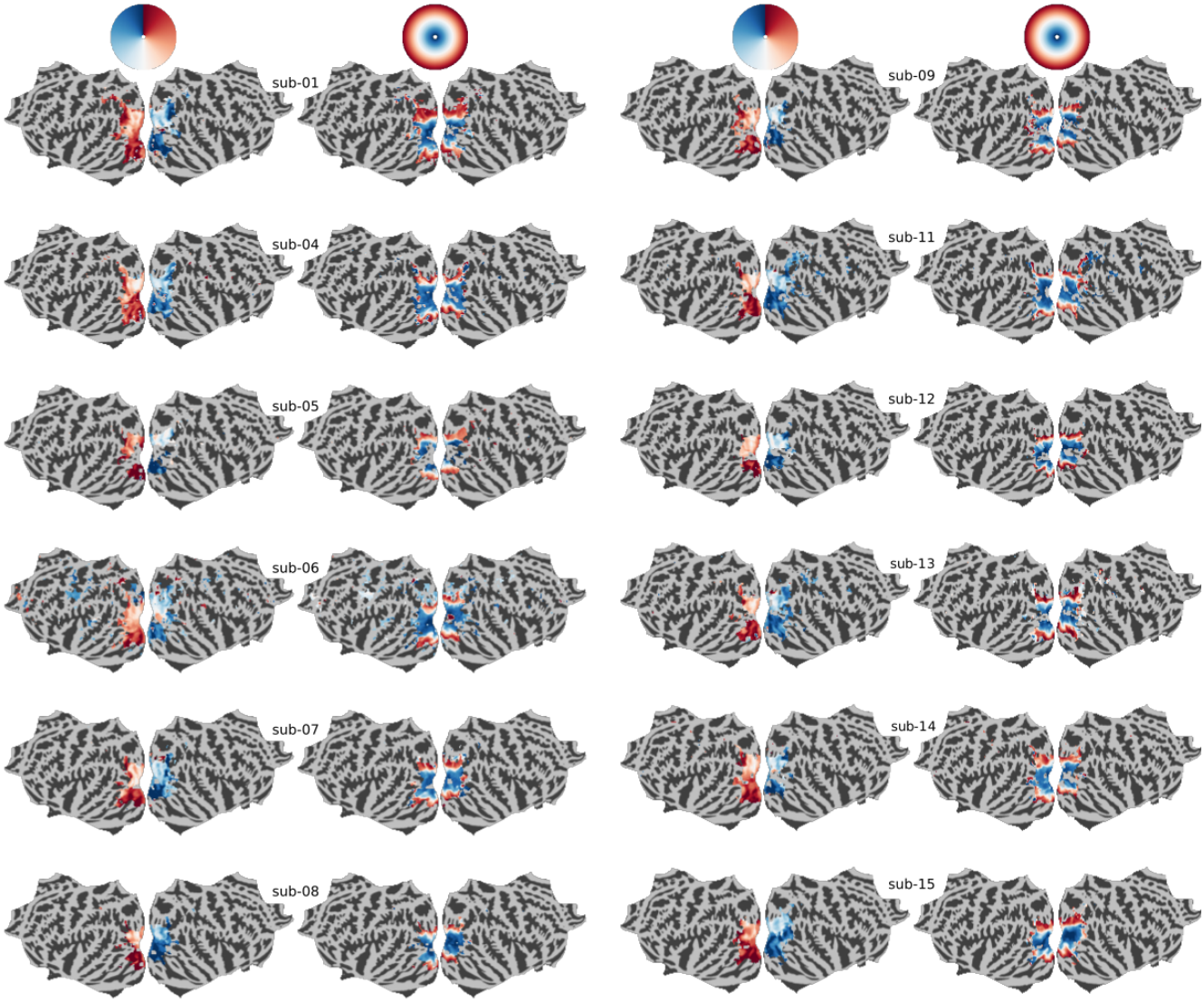

Figure 3. Individual, flat and binary maps of retinotopy in the visual field. (top) The visual field is encoded through polar coordinates: polar angle(left) and eccentricity (right). These polar coordinates are mapped on a flattened representation of the cortical surfaces extracted from the twelve IBC subjects: sub-01, sub-04, sub-05, sub-06, sub-07 and sub-08 on the left side; sub-09, sub-11, sub-12, sub-13, sub-14 and sub-15 on the other right side. One shall note the striking similarity of these maps across individuals. Individual binary maps for fixed effects are displayed for every participant, using an FDR-corrected threshold $q = 0.05$.

only with minor adjustments and most of them concerned with experimental settings; Retinotopy and Raiders were re-written from scratch in Python with no change of the design referring to the original paradigms.

The scripts used for data analysis are publicly available under the Simplified BSD license: https://github.com/individual-brain-charting/public_analysis_code.

Full description of the experimental designs, acquisition parameters and analysis pipeline for all tasks featuring this release as well as conditions and contrasts for the retinotopy tasks can be found on the online documentation of the IBC dataset: https://individual-brain-charting.github.io/docs/tasks.html.

References

1. Genon, S., Reid, A., Langner, R., Amunts, K. & Eickhoff, S. B. How to Characterize the Function of a Brain Region. *Trends Cogn Sci* **22**, 350–364. URL https://doi.org/10.1016/j.tics.2018.01.010 (2018).
2. Varoquaux, G. *et al.* Atlases of cognition with large-scale brain mapping. *PLoS Comput Biol* **14**. URL https://doi.org/10.1371/journal.pcbi.1006565 (2018).
3. King, M., Hernandez-Castillo, C. R., Poldrack, R. A., Ivry, R. B. & Diedrichsen, J. Functional boundaries in the human cerebellum revealed by a multi-domain task battery. *Nat Neurosci* **22**, 1371–1378. URL https://doi.org/10.1038/s41593-019-0436-x (2019).
4. Pinho, A. L. *et al.* Subject-specific segregation of functional territories based on deep phenotyping. *Hum Brain Mapp* **42**, 841–870. URL https://doi.org/10.1002/hbm.25189 (2021).
5. Thirion, B., Thual, A. & Pinho, A. L. From deep brain phenotyping to functional atlasing. *Curr Opin Behav Sci* **40**, 201–212. URL https://doi.org/10.1016/j.cobeha.2021.05.004 (2021). Deep Imaging - Personalized Neuroscience.
6. Wager, T. D., Lindquist, M. & Kaplan, L. Meta-analysis of functional neuroimaging data: current and future directions. *Soc Cogn Affect Neurosci* **2**, 150–158. URL https://doi.org/10.1093/scan/nsm015 (2007).
7. Costafreda, S. Meta-Analysis, Mega-Analysis, and Task Analysis in fMRI Research. *Philosophy, Psychiatry, & Psychology* **18**, 275–277. URL https://doi.org/10.1353/ppp.2011.0049 (2011).
8. Müller, V. I. *et al.* Ten simple rules for neuroimaging meta-analysis. *Neurosci Biobehav Rev* **84**, 151 – 161. URL https://doi.org/10.1016/j.neubiorev.2017.11.012 (2018).
9. Gurevitch, J., Koricheva, J., Nakagawa, S. & Stewart, G. Meta-analysis and the science of research synthesis. *Nature* **555**, 175–182. URL https://doi.org/10.1038/nature25753 (2018).
10. Salimi-Khorshidi, G., Smith, S. M., Keltner, J. R., Wager, T. D. & Nichols, T. E. Meta-analysis of neuroimaging data: A comparison of image-based and coordinate-based pooling of studies. *Neuroimage* **45**, 810–823. URL https://doi.org/10.1016/j.neuroimage.2008.12.039 (2009).
11. Samartsidis, P., Montagna, S., Johnson, T. D. & Nichols, T. E. The Coordinate-Based Meta-Analysis of Neuroimaging Data. *Stat Sci* **32**, 580 – 599. URL https://doi.org/10.1214/17-STS624 (2017).
12. Costafreda, S. Pooling fMRI data: meta-analysis, mega-analysis and multi-center studies. *Front Neuroinform* **3**, 33. URL https://doi.org/10.3389/neuro.11.033.2009 (2009).
13. Carp, J. On the Plurality of (Methodological) Worlds: Estimating the Analytic Flexibility of fMRI Experiments. *Front Neurosci* **6**, 149. URL https://doi.org/10.3389/fnins.2012.00149 (2012).
14. Thirion, B. *et al.* Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *Neuroimage* **35**, 105 – 120. URL https://doi.org/10.1016/j.neuroimage.2006.11.054 (2007).
15. Fedorenko, E., Behr, M. K. & Kanwishera, N. Functional specificity for high-level linguistic processing in the human brain. *Proc Natl Acad Sci U S A* **108**, 16428–33. URL https://doi.org/10.1073/pnas.1112937108 (2011).
16. Haxby, J. V. *et al.* A Common, High-Dimensional Model of the Representational Space in Human Ventral Temporal Cortex. *Neuron* **72**, 404 – 416. URL https://doi.org/10.1016/j.neuron.2011.08.026 (2011).
17. Nieto-Castañón, A. & Fedorenko, E. Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *Neuroimage* **63**, 1646 – 1669. URL https://doi.org/10.1016/j.neuroimage.2012.06.065 (2012).
18. Laumann, T. O. *et al.* Functional System and Areal Organization of a Highly Sampled Individual Human Brain. *Neuron* **87**, 657 – 670. URL https://doi.org/10.1016/j.neuron.2015.06.037 (2015).

19. Braga, R. M. & Buckner, R. L. Parallel Interdigitated Distributed Networks within the Individual Estimated by Intrinsic Functional Connectivity. *Neuron* **95**, 457–471.e5. URL https://doi.org/10.1016/j.neuron.2017.06.038 (2017).

20. Gordon, E. M. *et al.* Precision Functional Mapping of Individual Human Brains. *Neuron* **95**, 791 – 807.e7. URL https://doi.org/10.1016/j.neuron.2017.07.011 (2017).

21. Chang, N. *et al.* BOLD5000, a public fMRI dataset while viewing 5000 visual images. *Sci Data* **6**, 49. URL https://doi.org/10.1038/s41597-019-0052-3 (2019).

22. Sonkusare, S., Breakspear, M. & Guo, C. Naturalistic stimuli in neuroscience: critically acclaimed. *Trends Cogn Sci* **23**, 699–714. URL https://doi.org/10.1016/j.tics.2019.05.004 (2019).

23. Pinel, P. *et al.* Fast reproducible identification and large-scale databasing of individual functional cognitive networks. *BMC Neurosci* **8**, 91. URL https://doi.org/10.1186/1471-2202-8-91 (2007).

24. Barch, D. M. *et al.* Function in the human connectome: Task-fMRI and individual differences in behavior. *Neuroimage* **80**, 169–89. URL https://doi.org/10.1016/j.neuroimage.2013.05.033 (2013).

25. Orfanos, D. P. *et al.* The Brainomics/Localizer database. *Neuroimage* **144, Part B**, 309 – 314. URL https://doi.org/10.1016/j.neuroimage.2015.09.052 (2017). Data Sharing Part {II}.

26. Pinel, P. *et al.* The functional database of the ARCHI project: Potential and perspectives. *Neuroimage* **197**, 527 – 543. URL https://doi.org/10.1016/j.neuroimage.2019.04.056 (2019).

27. Bartels, A. & Zeki, S. Functional brain mapping during free viewing of natural scenes. *Human Brain Mapping* **21**, 75–85. URL https://doi.org/10.1002/hbm.10153 (2004).

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

29. Hanke, M. *et al.* A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie. *Sci Data* **1**. URL https://doi.org/10.1038/sdata.2014.3 (2014).

30. Hanke, M. *et al.* High-resolution 7-Tesla fMRI data on the perception of musical genres–an extension to the studyforrest dataset. *F1000Res* **4**, 174. URL https://doi.org/10.12688/f1000research.6679.1 (2015).

31. Hanke, M. *et al.* A studyforrest extension, simultaneous fMRI and eye gaze recordings during prolonged natural stimulation. *Sci Data* **3**. URL https://doi.org/10.1038/sdata.2016.92 (2016).

32. Sengupta, A. *et al.* A studyforrest extension, retinotopic mapping and localization of higher visual areas. *Sci Data* **3**. URL https://doi.org/10.1038/sdata.2016.93 (2016).

33. Pinho, A. L. *et al.* Individual Brain Charting, a high-resolution fMRI dataset for cognitive mapping. *Sci Data* **5**, 180105. URL https://doi.org/10.1038/sdata.2018.105 (2018).

34. Pinho, A. L. *et al.* Individual Brain Charting dataset extension, second release of high-resolution fMRI data for cognitive mapping. *Sci Data* **7**. URL https://doi.org/10.1038/s41597-020-00670-4 (2020).

35. Nishimoto, S. *et al.* Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies. *Curr Biol* **21**, 1641–6. URL https://doi.org/10.1016/j.cub.2011.08.031 (2011).

36. Sereno, M. *et al.* Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* **268**, 889–893. URL https://doi.org/10.1126/science.7754376 (1995).

37. Richard, H. & Thirion, B. Fastsrm: A fast, memory efficient and identifiable implementation of the shared response model. *Aperture Neuro* **3**, 1–13. URL https://doi.org/10.52294/001c.87954 (2023).

38. Gorgolewski, K. *et al.* The brain imaging data structure: a standard for organizing and describing outputs of neuroimaging experiments. *Sci Data* **3**, 160044. URL https://doi.org/10.1038/sdata.2016.44 (2016).

39. Oldfield, R. C. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* **9**, 97–113. URL https://doi.org/10.1016/0028-3932(71)90067-4 (1971).

40. Peirce, J. *et al.* PsychoPy2: Experiments in behavior made easy. *Behav Res Methods* **51**, 195–203. URL https://doi.org/10.3758/s13428-018-01193-y (2019).

41. Krause, F. & Lindemann, O. Expyriment: A Python library for cognitive and neuroscientific experiments. *Behav Res Methods* **46**, 416–428. URL https://doi.org/10.3758/s13428-013-0390-6 (2014).

42. Moeller, S. *et al.* Multiband multislice GE-EPI at 7 Tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magn Reson Med* **63**, 1144–53. URL https://doi.org/10.1002/mrm.22361 (2010).

43. Feinberg, D. A. *et al.* Multiplexed Echo Planar Imaging for Sub-Second Whole Brain fMRI and Fast Diffusion Imaging. *PLoS One* **5**, 1–11. URL https://doi.org/10.1371/journal.pone.0015710 (2010).

44. Bischoff-Grethe, A. *et al.* A technique for the deidentification of structural brain MR images. *Hum Brain Mapp* **28**, 892–903. URL https://doi.org/10.1002/hbm.20312 (2007).

45. Andersson, J. L., Skare, S. & Ashburner, J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage* **20**, 870 – 888. URL https://doi.org/10.1016/S1053-8119(03)00336-7 (2003).

46. Smith, S. *et al.* Advances in functional and structural {MR} image analysis and implementation as {FSL}. *Neuroimage* **23, Supplement 1**, S208 – S219. URL https://doi.org/10.1016/j.neuroimage.2004.07.051 (2004).

47. Gorgolewski, K. *et al.* Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Front Neuroinform* **5**, 13. URL https://doi.org/10.3389/fninf.2011.00013 (2011).

48. Friston, K., Frith, C., Frackowiak, R. & Turner, R. Characterizing Dynamic Brain Responses with fMRI: a Multivariate Approach. *Neuroimage* **2**, 166–172. URL https://doi.org/10.1006/nimg.1995.1019 (1995).

49. van Essen, D. C., Glasser, M. F., Dierker, D. L., Harwell, J. & Coalson, T. Parcellations and Hemispheric Asymmetries of Human Cerebral Cortex Analyzed on Surface-Based Atlases. *Cereb Cortex* **22**, 2241. URL https://doi.org/10.1093/cercor/bhr291 (2012).

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

51. Willems, R. M., Nastase, S. A. & Milivojevic, B. Narratives for neuroscience. *Trends Cogn Sci* **43**, 271–273. URL https://doi.org/10.1016/j.tins.2020.03.003 (2020).

52. Wheatley, T., Boncz, A., Toni, I. & Stolk, A. Beyond the isolated brain: The promise and challenge of interacting minds. *Neuron* **103**, 186–188. URL https://doi.org/10.1016/j.neuron.2019.05.009 (2019).

53. Chen, P.-H. C. *et al.* A reduced-dimension fmri shared response model. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M. & Garnett, R. (eds.) *Advances in Neural Information Processing Systems* (Curran Associates, Inc.), **28**. URL https://proceedings.neurips.cc/paper/2015/file/b3967a0e938dc2a6340e258630febd5a-Paper.pdf (2015).

54. Wu, A., Pashkovski, S., Datta, S. R. & Pillow, J. W. Learning a latent manifold of odor representations from neural responses in piriform cortex. In Bengio, S. *et al.* (eds.) *Advances in Neural Information Processing Systems* (Curran Associates, Inc.), **31**. URL https://proceedings.neurips.cc/paper/2018/file/17b3c7061788dbe82de5abe9f6fe22b3-Paper.pdf (2018).

55. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178. URL https://doi.org/10.1038/nature18933 (2016).

56. Mills, K. HCP-MMP1.0 projected on fsaverage. *Figshare. Dataset version 2*. URL https://doi.org/10.6084/m9.figshare.3498446.v2 (2016).

57. Abraham, A. *et al.* Machine learning for neuroimaging with scikit-learn. *Front Neuroinform* **8**, 14. URL https://doi.org/10.3389/fninf.2014.00014 (2014).

58. Warnking, J. *et al.* fMRI Retinotopic Mapping—Step by Step. *Neuroimage* **17**, 1665–1683. URL https://doi.org/10.1006/nimg.2002.1304 (2002).

59. Gao, J. S., Huth, A. G., Lescroart, M. D. & Gallant, J. L. Pycortex: an interactive surface visualizer for fMRI. *Front Neuroinform* **9**, 23. URL https://doi.org/10.3389/fninf.2015.00023 (2015).

60. Pinho, A. L. *et al.* Individual Brain Charting (IBC). *EBRAINS. Dataset*. v3.0. URL https://doi.org/10.25493/SM37-TS4 (2021).

61. Poldrack, R. *et al.* Toward open sharing of task-based fMRI data: the openfMRI project. *Front Neuroinform* **7**, 12. URL https://doi.org/10.3389/fninf.2013.00012 (2013).

62. Pinho, A. L. *et al.* IBC. *OpenNeuro. Dataset*. ds002685 v1.3.1. URL https://doi.org/10.18112/openneuro.ds002685.v1.3.1 (2021).

63. Gorgolewski, K. *et al.* NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Front Neuroinform* **9**, 8. URL https://doi.org/10.3389/fninf.2015.00008 (2015).

64. Pinho, A. L. *et al.* *NeuroVault*. id collection=6618. URL https://identifiers.org/neurovault.collection:6618 (2020).

Acknowledgements

We are very thankful to all volunteers who have accepted to be part of this ten-year long study, with many MRI repeated scans over this period of time.

We thank to the *Gallant Lab* for having provided the behavioral protocol and stimuli of naturalistic scenes used to develop the Clips tasks. We also thank to the Center for Magnetic Resonance Research, University of Minnesota for having kindly provided the Multi-Band Accelerated EPI Pulse Sequence and Reconstruction Algorithms. This project has received funding from the European Union's Horizon 2020 Framework Program for Research and Innovation under Grant Agreement No 720270 (Human Brain Project SGA1), 785907 (Human Brain Project SGA2) and 945539 (Human Brain Project SGA3). Ana Luísa Pinho is the recipient of a BrainsCAN Postdoctoral Fellowship at Western University, funded by the Canada First Research Excellence Fund (CFREF).

Author Contributions Statement

A.L.P. set the task protocols, performed the MRI acquisitions, performed the analysis of the naturalistic-task fMRI data, post-processed the behavioral data, produced video annotations, contributed to the IBC public analysis pipeline, contributed to Nilearn, contributed to Pypreprocess, wrote the IBC documentation and wrote the paper. H.R. developed FastSRM. M.E. designed the Retinotopy tasks, help setting the Clips task and contributed to Nilearn. A.A. set the MRI sequences. E.D. developed Pypreprocess and Nilearn. I.D. assisted in the implementation of the task protocols, performed some MRI Sessions and contributed to Pypreprocess. J.J.T. wrote the IBC documentation, performed some MRI Sessions, contributed to the IBC public analysis pipeline and contributed to Nilearn. S.S. wrote the IBC documentation, contributed to Pypreprocess, contributed to the IBC public analysis pipeline and assisted on preliminary analysis of naturalistic-task fMRI data. H.A. wrote the documentation, contributed to Pypreprocess and contributed to the IBC public analysis pipeline. A.F.P. wrote the documentation and contributed to the IBC public analysis pipeline. A.T. contributed to the IBC public analysis pipeline and contributed to Nilearn. T.C. conceptualized Figure 1. C.G., S.B-D., S.R.,Y.L. and V.B. performed the MRI acquisitions plus visual inspection of the neuroimaging data for quality-checking. L.L., V.J-T. and G.M-C. recruited the participants and managed routines related to appointment scheduling and ongoing medical assessment. C.D. recruited the participants and managed the scientific communication of the project with them. B.M. managed regulatory issues. G.V. developed Nilearn, contributed to Pypreprocess and advised on the study design plus analysis pipeline. S.D. conceived the general design of the study. L.H-P. conceived the general design of the study and wrote the ethical protocols. B.T. conceived the general design of the study, managed the project, wrote the ethical protocols, performed some MRI acquisitions, pre-processed fMRI data, performed the analysis of the retinotopy fMRI data, developed FastSRM, developed Nilearn, developed Pypreprocess, developed the IBC public analysis pipeline, uploaded IBC collections on EBRAINS + OpenNeuro + NeuroVault, wrote the IBC documentation and wrote the paper.

Competing Interests

The authors declare no competing interests.