



# Voxelwise encoding models of body stimuli reveal a representational gradient from low-level visual features to postural features in occipitotemporal cortex

Giuseppe Marrazzo<sup>a</sup>, Federico De Martino<sup>a,b</sup>, Agustin Lage-Castellanos<sup>a,c</sup>, Maarten J. Vaessen<sup>a</sup>, Beatrice de Gelder<sup>a,\*</sup>

<sup>a</sup> Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Limburg 6200 MD, Maastricht, the Netherlands

<sup>b</sup> Center for Magnetic Resonance Research, Department of Radiology, United States and Department of NeuroInformatics, University of Minnesota, Minneapolis MN 55455, USA

<sup>c</sup> Cuban Center for Neuroscience, Street 190 e/25 and 27 Cubanacán Playa Havana, CP 11600, Cuba

## ARTICLE INFO

### Keywords:

Body representation  
Encoding models  
Occipitotemporal cortex  
Banded ridge regression  
Extrastriate body area

## ABSTRACT

Previous research on body representation in the brain has focused on category-specific representation, using fMRI to investigate the response pattern to body stimuli in occipitotemporal cortex. But the central question of the specific computations involved in body selective regions has not been addressed so far. This study used ultra-high field fMRI and banded ridge regression to investigate the computational mechanisms of coding body images, by comparing the performance of three encoding models in predicting brain activity in occipitotemporal cortex and specifically in the extrastriate body area (EBA). Our results indicate that bodies are encoded in occipitotemporal cortex and in the EBA according to a combination of low-level visual features and postural features.

## 1. Introduction

Faces and bodies are amongst the most frequently encountered visual objects and provide essential information about the behaviour of conspecifics. In contrast to face perception, body perception is still poorly understood. Mainstream research on body representation in humans has focussed on category specific body representation in the brain, investigated with fMRI to identify conceptual category defined functional selectivity. Initially a body category selective area was reported in the middle occipital/temporal gyrus, the extrastriate body area (EBA) (Downing et al., 2001). Later a second body selective area was described in the fusiform cortex and labelled fusiform body area (FBA) (Peelen and Downing, 2005). Studies on body representation in nonhuman primates using fMRI as well as invasive electrophysiology resulted in a similar situation of multiple body sensitive patches in temporal cortex (Vogels, 2022). Given multiple category selective areas in human as well as in nonhuman primates, the central issue is to understand how body images are coded in the different body selective areas and how to account for the observed body selectivity.

Previous studies tried to understand the functional role of EBA vs FBA using the contrast between representation of body parts vs. configural processing of the whole body image; for reviews see (de Gelder and Poyo Solanas, 2021; de Gelder et al., 2010; Downing and Peelen, 2011,

2016; Peelen and Downing, 2007). This issue was addressed in several studies but there is currently no consensus in the literature on assigning one or the other to either EBA or FBA. One study using body parts comparing perception of actor identity and action showed how identity is represented in EBA and action in ventral premotor cortex (Urgesi et al., 2007). Other studies showed that EBA encodes image details pertaining to the shape, posture and position of the body (Downing and Peelen, 2011). Another study found that EBA was selective for body parts and the more anterior FBA for whole bodies and their overall configuration (Taylor and Downing, 2011; Taylor et al., 2007). A few studies also showed that EBA activity is influenced by the specific task (Bracci et al., 2017; Bugatus et al., 2017; Marrazzo et al., 2021) as well as by semantic attributes like gender and emotional expression (Chan et al., 2004; Ewbank et al., 2011; Foster et al., 2022, 2019; Kanwisher, 2017; Peelen et al., 2007; Taylor et al., 2010; Vangeneugden et al., 2014).

More recent studies have moved away from the part vs. whole distinction as a means of clarifying functional roles between two different body selective areas and explored how using computational models can be used to bridge the gap between low-level physical features of the stimulus taking place in early visual cortex and the generation of semantic concepts. In view of its location in temporal cortex the kind of coding to expect in EBA may be related to computing some subsymbolic body features rather than semantics defined over whole bodies and body parts

\* Corresponding author.

E-mail address: [b.degelder@maastrichtuniversity.nl](mailto:b.degelder@maastrichtuniversity.nl) (B. de Gelder).

<https://doi.org/10.1016/j.neuroimage.2023.120240>.

Received 14 March 2023; Received in revised form 16 June 2023; Accepted 19 June 2023

Available online 20 June 2023.

1053-8119/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

(Bracci et al., 2019). Recent computational approaches move the classical debate on parts vs. whole category representation forward to more analytical investigations on how higher order category representation emerges in the brain from posterior to more anterior visual areas via intermediate midlevel representations (de Gelder and Poyo Solanas, 2021; Ratan Murty, Bashivan, Abate, DiCarlo, and Kanwisher, 2021). Candidate features are overall shape representation (whole-body shape) and related to that, viewpoint tolerance, an important dimension in the posterior to anterior gradient of object recognition. Studies in non-human primates that use single cell recordings indicate that moving from posterior to anterior temporal cortex, body patch neurons increase their selectivity for body identity and posture, while there is a decrease in viewpoint selectivity. Specifically, recordings in body selective patches, middle superior temporal body (MSB) and anterior superior temporal body (ASB) showed strong viewpoint selectivity for the former and conversely, high tolerance for the latter (Kumar et al., 2019). Furthermore, Caspari and colleagues using the same set of category stimuli as Kumar and colleagues, showed similar decoding pattern between monkeys and humans in body selective regions, suggesting an homology between the human EBA and monkey MSB as well as the human FBA and monkey ASB (Caspari et al., 2014; Kumar et al., 2019).

This suggests a general principle of object coding in the inferior temporal cortex (IT): a greater tolerance to image transformations that preserve identity (Bao et al., 2020) and, in the case of bodies, posture, for more anterior patches. The monkey data fits human fMRI work that found viewpoint-invariant decoding of body identity in FBA but not EBA (Foster et al., 2021), but as noted above, results of between-area differences in fMRI multi voxel pattern analysis (MVPA) are difficult to interpret (Dubois et al., 2015).

An important question is whether a similar posterior to anterior organisation can be found for EBA by using ultra-high field fMRI in combination with computational hypotheses. One popular approach to test and compare different computational hypotheses of brain function is to use (linearized) encoding (Kay et al., 2008; Naselaris et al., 2011) approaches. In these approaches brain activity (e.g. the blood oxygen level dependant (BOLD) signals in a voxel or a brain area in fMRI) is predicted from the features of (different) computational models, and their accuracy can be compared to adjudicate between competing models or partitioned with the respect to the variance explained by each of the models (Dumoulin and Wandell, 2008; Dupré la Tour, Eickenberg, and Gallant, 2022; Moerel et al., 2012; Nunez-Elizalde et al., 2019; Santoro et al., 2014; Thirion et al., 2006; Wandell et al., 2007).

Candidate encoding models for predicting activity in visual cortex are Gabor filters. These filters have been shown to be powerful tools to simulate receptive fields of neurons in the visual system and are used to decompose visual images into different spatial frequency and orientation components (Adelson and Bergen, 1985; Carandini et al., 2005; Jones and Palmer, 1987; Kay et al., 2008; Naselaris et al., 2011).

Literature on body perception has not yet proposed a clear computational hypothesis that could explain activity in occipitotemporal cortex in response to visual body stimuli. Previous studies have shown the sensitivity of EBA and FBA to bodies and body parts presented in different visual formats like stick figures, line drawings and silhouettes (Amoruso et al., 2011; Downing et al., 2001; Peelen and Downing, 2005; Schwarzlose et al., 2005; Spiridon et al., 2006; Weiner and Grill-Spector, 2010). Therefore, a candidate for predicting visual responses to body stimuli could be a model that encodes postural features like for instance the position of the joints. This representation could be viewpoint dependant like the orthogonal projection of a three-dimensional stimulus on one of the planes of the reference frame or could be viewpoint independent. If so, a model which encodes a three-dimensional representation of the joints would be more appropriate.

We used ultra-high field fMRI and linearized encoding to evaluate to what extent the response in occipitotemporal cortex and specifically the extrastriate body areas can be explained on the one hand by low-level visual features (Gabor) (Nishimoto et al., 2011) and on the other

by the features extracted by two computational models that represent the postural features of the body (kp2d, kp3d) (Loper et al., 2015; Pavlakos et al., 2019).

## 2. Material and methods

### 2.1. Participants

20 right-handed subjects (8 males, mean age =  $24.4 \pm 3.4$  years) participated in this study. They all had normal (or corrected to normal) vision and were recruited from Maastricht University student cohorts. All subjects were naïve to the task and the stimuli and received monetary compensation for their participation (7.5 € VVV vouchers/per hour or a bank transfer for the same amount; 4 h in total, 30 €). Scanning sessions took place at the neuroimaging facility Scannexus at Maastricht University. All experimental procedure conformed to the Declaration of Helsinki and the study was approved by the Ethics Committee of Maastricht University.

### 2.2. Stimuli

#### 2.2.1. Main experiment stimuli

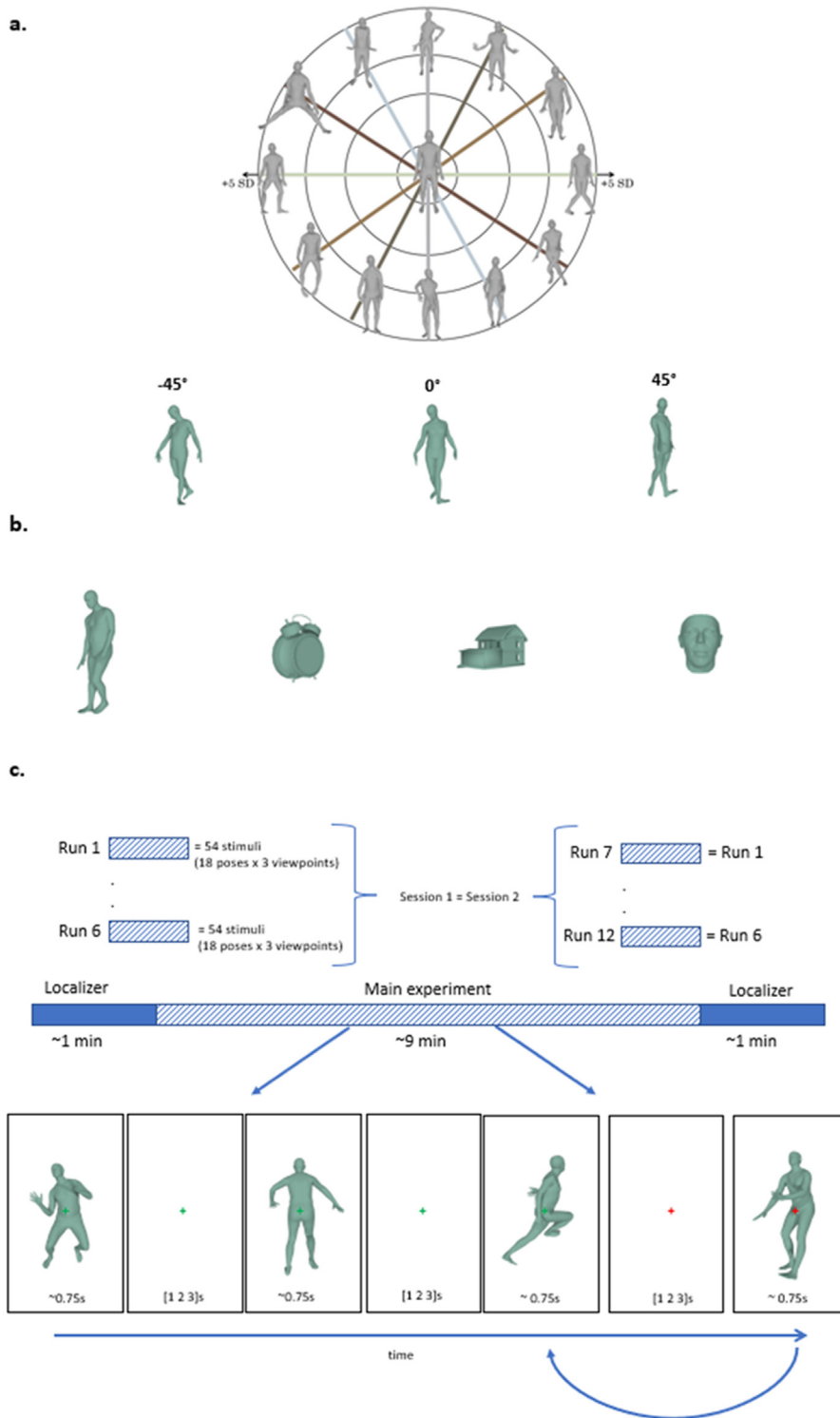
The stimulus set consisted of 108 pictures of 3D rendered body meshes shown in different orientations:  $0^\circ$  (frontal),  $-45^\circ$  (left rotated) and  $45^\circ$  (right rotated) for a total of 324 unique images. Examples of the stimuli in the different orientation are shown in Fig. 1a. 3D rendered body meshes were generated via VPoser, a variational autoencoder (VAE) trained to learn a 32-dimensional (normal distribution) latent representation of Skinned Multi-Person Linear Model (SMPL) parameters (Loper et al., 2015; Pavlakos et al., 2019). The stimuli used in the study were generated via randomly sampling the latent space and generating via the decoder part of the VPoser, the associated body image. To also sample images sufficiently distant from the mean image (and thus maintain a sufficiently large variability of poses in the stimulus set), we sampled the latent space within three distinct shells defined by the standard deviations from the mean pose (Fig. 1a). Ultimately, the body images were generated by transferring the decoded SMPL parameters to a posed mesh. The resulting body poses had mean widths and heights of  $2.43^\circ \times 5.22^\circ$  of visual angle and were colour rendered (mean RGB: 120,157,144).

#### 2.2.2. Localizer stimuli

Stimuli for the localizer experiment consisted of 3D rendered images depicting four object categories: faces, bodies, tools, and houses (Fig. 1b). The stimuli were colour rendered using the same colour for the main experiment stimuli (mean RGB: 120,157,144). None of the stimuli from the localizer were used in the main experiment.

### 2.3. Behavioural validation

Stimuli used in the main experiment were generated via the VAE. Random sampling from the latent space allowed us to produce a varied set of body poses but did not allow us to control the stimuli for the possible presence of semantic body attributes like action or emotion. Therefore, we asked 113 participants (25 excluded for missing data, 88 in total: 29 males, mean age =  $23 \pm 4$  years, 72 right handed) to rate the stimuli using a questionnaire consisting of both categorical and likert-scale questions. Participants were presented with 1/3 (108) of the total stimuli (324) for 750 ms each. For each participant, the stimuli were pseudo-randomized (108 stimuli randomly selected for each participant, but evenly distributed so that each stimulus got approximately the same number of answers). After each presentation, participants were asked to answer 6 questions about the emotional expression, action content; salience of specific body parts; implied body movement and realism of the posture (see supplementary material).



**Fig. 1.** Stimuli and experimental procedure. (a) (top) Stimuli were generated by randomly sampling the latent space of the VAE (Loper et al., 2015; Pavlakos et al., 2019). The 32-dimensional latent space was sampled in three shells, defined by the value of the standard deviation from the mean pose, to ensure variability amongst the generated poses. (a) (bottom) 108 unique poses were generated from three different viewpoints: 0° (frontal), -45° (left rotated) and 45° (right rotated) for a total of 324 unique stimuli. (b) Body sensitive areas were identified by mean of a localizer using stimuli selected from four different object categories: bodies, tools, houses, and faces. These stimuli underwent the same rendering process as the stimuli of the main experiment. (c) During the main experiment participants performed a one-back task. They fixated on the green cross and were presented with pictures of body poses each for approximately 750 ms followed by a blank screen which appeared for 1, 2 or 3 s. When the fixation cross turned red, they had to report by button press whether the current stimulus matched the previously presented one with the same viewpoint.

#### 2.4. MRI acquisition and experimental procedure

Participants viewed the stimuli while lying supine in the scanner. Stimuli were presented on a screen positioned behind participant's head at the end of the scanner bore (distance screen/eye = 99 cm) which the participants could see via a mirror attached to the head coil. The screen had a resolution of 1920 × 1200 pixels, and its angular size was 16° (horizontal) × 10° (vertical). The experiment was coded in Matlab (v2018b

The MathWorks Inc., Natick, MA, USA) using the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

Each participant underwent two MRI sessions, we collected a total of twelve functional runs (six runs per session) and one set of anatomical images. Images were acquired in a 7T MR scanner (Siemens Magnetom) using a 32-channel (NOVA) head coil. Anatomical images were collected via a T1-weighted MP2RAGE: 0.7 mm isotropic, repetition time (TR) = 5000 ms, echo time (TE) = 2.47 ms, matrix size = 320 × 320, num-

ber of slices = 240. The functional dataset covered the entire brain and was acquired via T2\*-weighted Multi-Band accelerated 2D-EPI BOLD sequence, multiband acceleration factor = 3, voxel size = 1.6 mm isotropic, TR = 1000 ms, TE = 20 ms, number of slices = 68 without gaps; matrix size = 128 × 128.

Each run consisted of three main sections: 1) two short localizers parts (approximately one minute at the beginning and at the end of the run), during which images were presented in blocks of categories (faces, bodies, tools, and houses), and 2) a main experimental part where stimuli (body images different from the ones used in the localizer) were presented following a fast event-related design. Participants were asked to fixate at all times on the green cross at the centre of the screen.

Each localizer block contained six images which were presented for 750 ms and followed by 250 ms blank screen. Each block lasted six seconds followed by a fixation period of eight seconds and each category block was presented once at the beginning and once at the end of each run (24 blocks per condition across the 12 runs). During the localizer participants did not perform any task.

During the main experiment, stimuli were presented for 750 ms with an inter stimulus interval that was pseudo-randomised to be 1, 2 or 3 TRs. To keep attention on the stimuli, participants performed a one-back task on stimulus identity. Following a visual cue (colour change of the fixation cross), they were asked to report via a button press whether the current stimulus was the same as the previous one (same posture and viewpoint) (Fig. 1c). Within each run, the experimental section consisted of the presentation of 54 stimuli (18 unique poses × 3 viewpoints) repeated 3 times each. Six target trials were added for a total of 168 trials. Across the two sessions each of the 108 unique poses were repeated 18 times (3 repetitions × 3 viewpoints × 2 sessions) across the 12 runs, whereas the 324 unique stimuli were repeated 6 times (3 repetitions × 2 sessions).

Preprocessing was performed using BrainVoyager software (v22.2, Brain Innovation B.V., Maastricht, the Netherlands) and FSL (Jenkinson et al., 2012; Smith et al., 2004; Woolrich et al., 2009). The following steps were performed in BrainVoyager unless indicated otherwise. EPI Distortion was corrected using the Correction based on Opposite Phase Encoding (COPE) plugin in BrainVoyager, where the amount of distortion is estimated based on volumes acquired with opposite phase-encoding (PE) with respect to the PE direction of the main experiment volumes (Fritz et al., 2014), after which subsequent corrections is applied to the functional volumes. Other preprocessing steps included: scan slice time correction using cubic spline, 3D motion correction using trilinear/sinc interpolation and high-pass filtering (GLM Fourier) cut off 3 cycles per run. During the 3D motion correction all the runs were aligned to the first volume of the first run. Anatomical images were resampled at 0.5 mm isotropic resolution using sinc interpolation and then normalized to Talairach space (Talairach and Tournoux, 1988). To ensure a correct functional-anatomical and functional-functional alignment, the first volume of the first run was coregistered to the anatomical data in native space using boundary based registration (Greve and Fischl, 2009). Volume Time Courses (VTCs) were created for each run in the normalized space (sinc interpolation) and exported in nifti format for further processing in FSL. To further reduce non-linear intersession distortions, functional images were additionally corrected using the fnirt command in FSL (Andersson et al., 2007) using as template the first volume of the first run in normalized space. Prior to the encoding analysis (and following an initial general linear model [GLM] analysis aimed at identifying regions of interest based on the response to the localizer blocks), we performed an additional denoising step of the functional time series by regressing out the stimulus onset (convolved with a canonical hemodynamic response function [HRF]) and the motion parameters.

Segmentation of white matter (WM) and grey matter (GM) boundary was performed in BrainVoyager using the deep learning-based segmentation algorithm and in house Matlab scripts. The resulting boundaries were then inflated to a reference sphere and aligned using cortex based

alignment (CBA) (Goebel et al., 2006). The aligned meshes were averaged to create a group WM-GM mesh for each hemisphere.

## 2.5. Voxels selection for encoding analysis

The functional time series of each participant were analysed using a fixed-effect GLM with 5 predictors (4 for the localizer blocks and 1 modelling the responses to all the stimuli in the main experiment). Motion parameters were included in the design matrix as nuisance regressors. The estimated regressor coefficients representing the response to the localizer blocks were used for voxel selection. A voxel was selected for the encoding analysis if significantly active ( $q(\text{FDR}) < 0.05$ ) within the main effect of the localizer (Body, Houses, Tools, Faces – Fig. 2). Note that this selection is unbiased to the response to the main stimuli presented in the experimental section of each run.

To assess the spatial consistency of activation to the localizer across subjects, we created a probabilistic functional map depicting, at each spatial location, the percentage of subjects for which that location was significantly ( $q(\text{FDR}) < 0.05$ ) modulated by the localizer blocks (Fig. 3a).

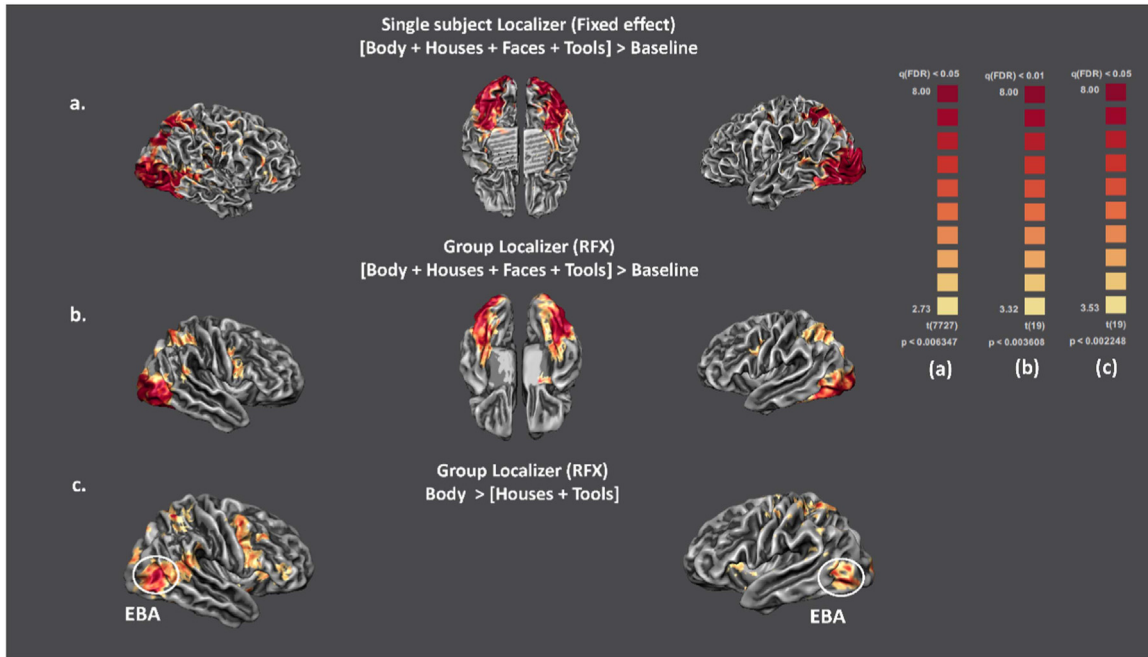
## 2.6. Functional ROI definition

We defined body selective regions at the group level using a random-effect GLM (RFX-GLM), in which EBA was defined using the localizer contrast Body > Objects([Houses + Tools]) (Ross et al., 2020) with a statistical threshold of  $q(\text{FDR}) < 0.05$ . Functional images from every participant were spatially smoothed using a gaussian filter (FWHM = 4 mm) and then entered the RFX GLM in which we defined 5 predictors of interest (4 for the localizer 1 for modelling the responses to the main experiment). For each participant, we regressed out signals coming from head motion by including motion parameters in the design matrix. Responses from each subject were selected via intersection with the group ROI definition of EBA and the single subject localizer's main effect map (see 2.5). Fig. 3b projects the group definition of EBA onto the probabilistic functional map of the localizer's main effect.

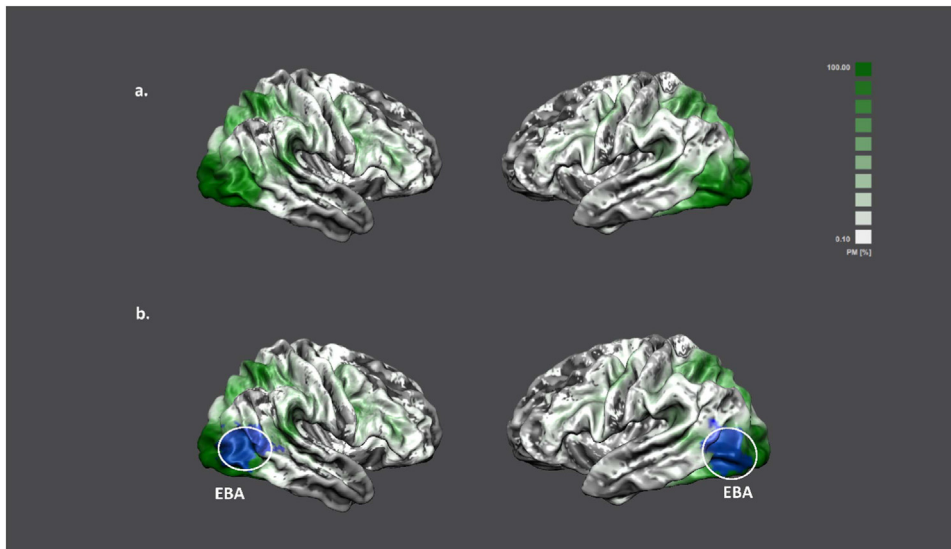
The group level body sensitive ROIs were intersected with the single subject activation maps to obtain individual ROIs. Note again that while this procedure makes use of the same data (localizer) twice, its purpose was to define single subject regions to be subsequently used for encoding analysis which was performed on an independent portion of the data set. Fig. 3b reports the overlap between EBA defined at the group level and the probabilistic activation maps of the localizer's main effect.

## 2.7. Encoding models

In order to understand what determines the response to body images we tested several hypotheses, represented by different computational models, using fMRI encoding (Allen et al., 2018; Kay et al., 2008; Naselaris et al., 2011; Santoro et al., 2014). We compared the performance (accuracy in predicting left out data) of three encoding models. The first represented body stimuli using the position of joints in two dimensions (kp2d) using 54 keypoints (joints, hand and facial features like eyeballs, neck and jaw) plus one keypoint for global rotation extracted during the stimulus creation using VPoser (Pavlakos et al., 2019). This encoding model extracts for each pose the orthogonal projection of the pose's spatial coordinates on the camera plane which ultimately constitutes the image coordinates (i.e. x,y) of the keypoints. Therefore, this model has 110 features (55 kp \* 2 dimensions). The second model represented body stimuli using the three-dimensional position of the keypoints (kp3d) extracted from VPoser. This representation differs from the kp2d one by adding the third dimension (no projection on the camera plane), resulting in an encoding model with 165 features (55 kp \* 3 dimensions). It is important to note that the main difference between the kp2d and kp3d representations is that the latter is viewpoint invariant as the position of the joints is independent from the angle under which the object is observed.



**Fig. 2.** Univariate analysis. (a) Brain maps showing the responses for the main effect of the localizer in a single subject computed with a fixed-effect GLM. This map was created in volume space ( $q(\text{FDR}) < 0.05$ ) and overlaid on the subject mesh for visualization purposes. (b) Brain activation for the main effect of the localizer obtained when including all the subjects in a RFX GLM. The activation map is corrected for multiple comparison at  $q(\text{FDR}) < 0.01$  and is cluster thresholded (cluster size = 25). (c) Body selective regions obtained by contrasting the localizer conditions Body > Objects (Houses + Tools). As in (a), the statistical thresholding of the map was performed in volume space ( $q(\text{FDR}) < 0.05$ ) and then overlaid on the group average mesh for visualization purposes. We used this contrast to obtain a group definition of EBA which was intersected with single subjects' activation maps for the subsequent ROI analysis.



**Fig. 3.** Probabilistic map of the main effect of the localizer. (a) This brain map shows the extent of the overlap between participants within the main effect of the localizer computed for each participant across all runs via a fixed-effect GLM. This overlap is expressed via a probability map where at each spatial location the percentage of the relative number of subjects leading to significant activity is reported (low probability  $\rightarrow$  high probability: white  $\rightarrow$  green). (b) In the second row, we overlay the binarized (suprathreshold voxels  $q(\text{FDR}) < 0.05 = 1$ ) group definition of EBA (in blue) (see Fig. 2c) on the probabilistic map. This shows that most (90–100%) of the participants shared significant responses ( $q(\text{FDR}) < 0.05$ ) within the region of interest.

The last encoding model we tested is a Gabor filtering of the images (Adelson and Bergen, 1985; Nishimoto and Gallant, 2011; Nishimoto et al., 2011; Watson and Ahumada, 1985). In this procedure, each stimulus was transformed into the Commission internationale de l'éclairage (CIE)  $L^*A^*B^*$  colour space and the luminance signals then passed through a bank of 1425 spatial Gabor filters differing in position, orientation, and spatial frequency (DeAngelis et al., 1993; Jones and Palmer, 1987; Nishimoto et al., 2011). Ultimately, the filters output underwent a logarithmic non-linear compression in which large values were scaled down more than small values. For details on this procedure we refer to the original publication (Nishimoto et al., 2011).

### 2.8. Banded ridge regression and model estimates'

Generally, in the linearized encoding framework (as applied in fMRI) the information explained in brain activity is obtained via L2-regularized (ridge) regression (Hoerl and Kennard, 1970). Ridge regression is a powerful tool which allows to improve performance of encoding models whose features are nearly collinear, and it minimizes overfitting. When dealing with more than one encoding model, ridge regression can either estimate parameters of a joint feature space (combining all feature spaces in one encoding model) or obtain model estimates from each encoding model separately. Fitting a joint model with ridge regression

allows considering the complementarity of different feature spaces but subjects all models (feature sets) to a unique regularization. As the optimal regularization required when fitting each individual feature space may differ (since it depends, amongst others, on factors such as number of features and features covariances) (Dupré la Tour et al., 2022), fitting a joint model with one regularization parameter may be suboptimal and can be extended to banded ridge regression. In banded ridge regression, separate regularization per parameters for each feature space are optimized, which ultimately improves model performance by reducing spurious correlations and ignoring non-predictive feature spaces (Dupré la Tour et al., 2022; Nunez-Elizalde et al., 2019). In the present work we used banded ridge regression to fit the three encoding models and performed a decomposition of the variance explained by each of the models following established procedures (Dupré la Tour et al., 2022). All analyses were performed using a publicly available repository in Python (Himalaya, <https://github.com/gallantlab/himalaya>).

Model training and testing were performed in cross-validation (3-folds: training on 8 runs [216 stimuli] and testing on 4 runs [108 stimuli]). For each fold, the training data were additionally split in training set and validation set using split-half crossvalidation. Within the (split-half) training set a combination of random search and gradient descent (Dupré la Tour et al., 2022) was used to choose the model (regularization strength and model parameters) that maximized the prediction accuracy on the validation set. Ultimately, the best model over the two (split-half) folds was selected to be tested on the yield out test data (4 runs). The fMRI predicted time courses were estimated as follows. Within each fold, the models' representations of the training stimuli were normalized (each feature was standardized to zero mean and unit variance withing the training set). The feature matrices representing the stimuli were then combined with the information of the stimuli onset during the experimental runs. This resulted in an experimental design matrix (nrTRs x NrFeatures) in which each stimulus was described by its representation by each of the models. To account for the hemodynamic response, we delayed each feature of the experimental design matrix (15 delays spanning 15 s). The same procedure was applied to the test data, with the

only difference that when standardizing the model matrices, the mean and standard deviation obtained from the training data were used. We used banded ridge regression to determine the relationship between the fMRI response at each voxel, which significantly responded to the localizer stimuli ( $p(\text{FDR}) < 0.05$ ), and the features of the encoding models (stimulus representations).

For each cross-validation, we assessed the accuracy of the model in predicting fMRI time series by computing the correlation between the predicted fMRI response to novel stimuli (4 runs, 108 stimuli) and the actual responses. The accuracy obtained across the three folds were then averaged. To obtain the contribution of each of the models to the overall accuracy we computed the partial correlation between the measured time series and the prediction obtained when considering each of the models individually (Dupré la Tour et al., 2022).

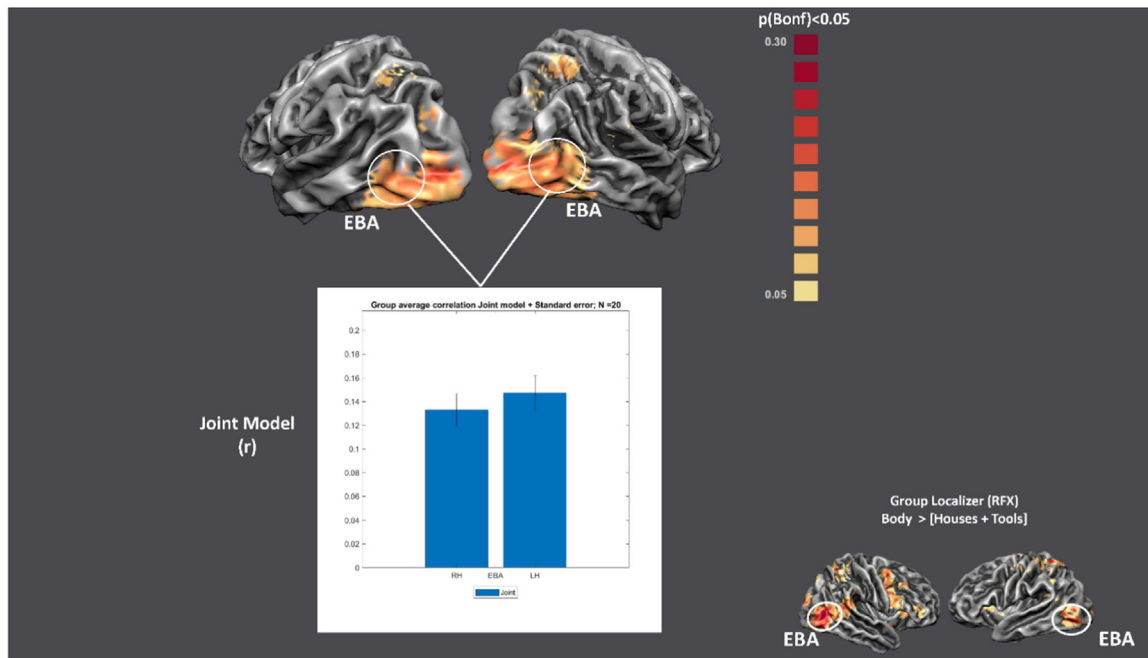
## 2.9. Group maps and statistical inference

To evaluate the statistical significance of the model fittings, accuracy maps of each subject were projected on the cortex based aligned group WM-GM mesh. We computed the probability of the mean accuracy (across subjects) to be higher than chance by sign flipping (10,000 times) the correlations. This procedure allowed estimating a non-parametric null distribution for each vertex, which was used to obtain a significance value for the mean accuracy. We accounted for the multiple comparisons by correcting the p-values using Bonferroni correction (i.e. dividing by the number of tests, equal to the number of vertices in the analysis).

## 3. Results

### 3.1. Behavioural analysis

The analysis of the responses to the questionnaires revealed that no action was recognised for 92% of the stimuli (298 out of 324). Likewise, no emotion was recognised for 97% of the stimuli (314 out of 324).



**Fig. 4.** Joint model performance. Group Prediction accuracy for the joint model (kp2d, kp3d, Gabor). Statistical significance was assessed via permutation test (subject wise sign-flipping, 10,000 times), and correction for multiple comparison was performed using Bonferroni correction ( $p < 0.05$ ). The bar plot depicts the group (mean + standard error) correlation coefficient between the joint model predictions and brain response to novel stimuli (test stimuli) across participants in bilateral EBA. We did not find any significant difference across hemispheres (two-sample  $t$ -test,  $p = 0.481$ ). For reference, the bottom right corner shows the functional definition of EBA already presented in Fig. 2c.

Participants reported that they focused on the overall body pose in 65% of the cases (211 out of 324), on the hands in 20% (64 out of 324) of the cases and on the arms for 11% (38 out of 324). The full report on the behavioural results is found in the supplementary material.

### 3.2. Univariate analysis and voxels selection for encoding

In each subject, voxels that significantly ( $q(\text{FDR}) < 0.05$ ) responded to the localizer conditions (main effect) were selected for the encoding analysis (Fig. 2a). At the group level, we observed significant ( $q(\text{FDR}) < 0.01$ ) activation in occipitotemporal cortex as well as parietal cortex, in the occipital gyrus (superior/middle/inferior) (SOG/MOG/IOG), fusiform gyrus (FG), lingual gyrus (LG), middle temporal gyrus (MTG), superior parietal lobule (SPL), intraparietal sulcus (IPS), inferior temporal sulcus (ITS), lateral occipital sulcus (LOS), superior temporal sulcus (STS) (Fig. 2b). Subtracting the response to object stimuli from the response to body stimuli allowed us to define EBA. This cluster spanned the MOG, MTG as well as the ITS (Fig. 2c). The voxels selection for the encoding analysis was performed at the individual level and based on the main effect. A probabilistic map (computed by counting the number of subjects for which a given voxel was included in the analysis) showed a consistent overlap with the functionally defined EBA (Fig. 3).

### 3.3. Encoding results

The voxels selected using the response to the localizer were submitted to the encoding analysis. The response to the body stimuli presented in the main experiment (data independent from the localizer) were modelled using banded ridge regression. The group performance of the joint (three) encoding model is shown in Fig. 4. The accuracy of the joint (kp2d, kp3d, Gabor) encoding model at the group level (after statistical testing and correction for multiple comparisons) is shown in Fig. 4. We found that when combining information from the three models we could significantly predict responses to novel stimuli

(Fig. 4) throughout the ventral visual cortex (SOG, MOG, IOG, ITG, MTG, FG, LOS), and in parietal cortex (SPL). We assessed spatial differences in how models contributed to the fMRI response by colour coding the relative contribution of each of the models to the overall prediction accuracy (Fig. 5). The response to bodies in early visual cortical areas was in average better explained by the Gabor model (blue-purple-dark magenta).

Moving to higher visual cortical areas corresponded to a shift in the relative contribution towards a combination of kp2d and Gabor (magenta), while in EBA the model that contributed most to the prediction accuracy was kp2d (magenta - light magenta - pink). When considering EBA (Fig. 5), the joint model significantly predicted brain responses to test stimuli (Fig. 4), and the kp2d model accounts for approximately 50–60% of the variance of this prediction (Fig. 5b). It is worth noting that when considering the spatial distribution of relative model contributions to the prediction accuracy (Fig. 5), the posterior part of EBA, specifically the posterior part of lateral occipital sulcus (LOS) was best explained by the Gabor model (dark magenta area), while the anterior part of LOS showed lighter shades of magenta indicating that the leading representation is kp2d.

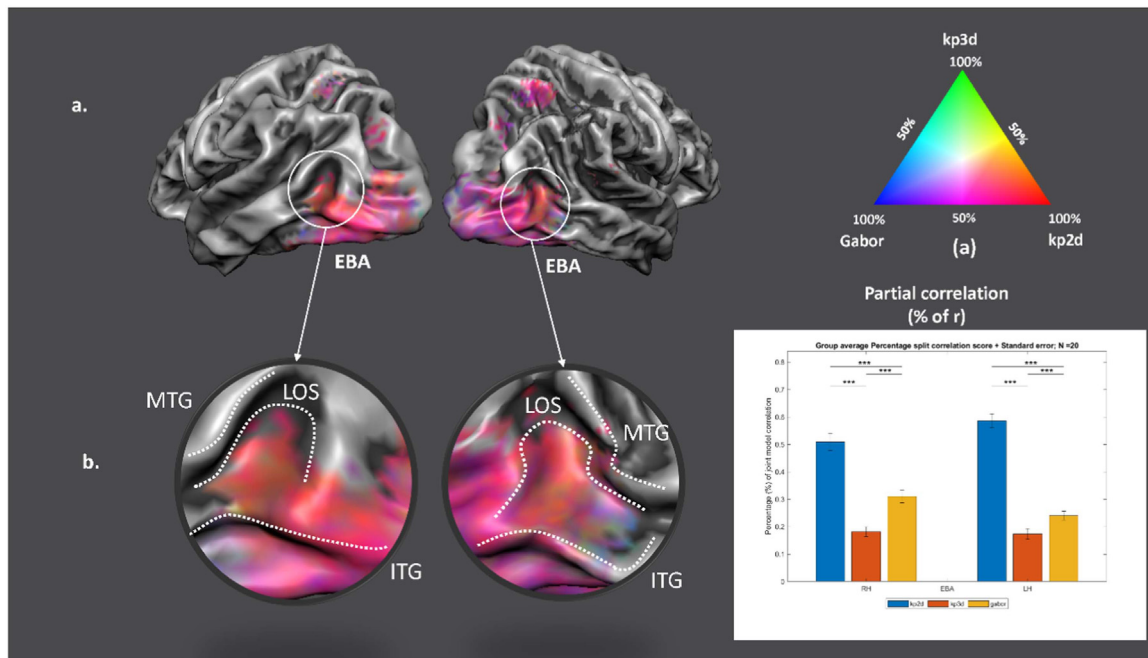
## 4. Discussion

In this study, we used ultra-high field fMRI to determine the main (stimulus) features that drive brain responses to still body stimuli in occipitotemporal cortex, with a particular focus to the responses in the extrastriate body area (EBA). We compared the performance of three encoding models using banded ridge regression. At the group level, we observed that a combination of the three models (kp2d, kp3d, Gabor) could significantly (permutation testing and corrected for multiple comparison) predict fMRI BOLD responses in occipitotemporal cortex and

in parietal cortex (SPL). The partial correlation analysis revealed significant differences (at the group level) between the separate models in EBA, with approximately 50% of the joint model performance (correlation) being explained by kp2d, 30% by Gabor and 20% by kp3d (see Fig. 5). These results lead us to conclude that a combination of low-level visual features and postural features may play a role in encoding responses to body stimuli in EBA. The total variance of the joint model is arguably low (albeit significantly different from zero, see Fig. 4) compared to previous studies investigating the encoding of (visual) stimuli features in fMRI responses (Kay et al., 2008; Naselaris et al., 2011; Naselaris et al., 2009; Nishimoto and Gallant, 2011; Nishimoto et al., 2011; Nunez-Elizalde et al., 2019). There are several issues why this may be the case. Firstly, the fMRI data we collected may be noisy (noisier than other dataset used to evaluate the encoding of models in other cortical areas). To evaluate the data quality, and most importantly the best attainable performance by any model given the data quality, it is common practice to compute noise ceilings. We performed fMRI encoding at the level of whole time series, thus the most appropriate metric for the computation of the noise ceiling would be the correlation of two identical repetitions of the test data (Lage-Castellanos et al., 2019). Unfortunately, the design we used prevents us from performing such analysis as the order (and jittering) of the stimuli is fully randomized across runs. Secondly, previous fMRI studies that considered the variance explained by a computational model in visual areas, have done so on responses elicited by a varied stimulus set including (scenes) with e.g. different categories. This stimulus choice increases the stimulus related variance of the response as (for example) it exploits the regional differences in category preference. That is, using such a stimulus set EBA would activate substantially to stimuli in which bodies are present but very little when bodies are not present in the stimulus. We on the other hand focused only on body stimuli, consequently reducing the stimulus related variance. Even with similar data quality, such a choice is expected to reduce the noise ceiling as the ratio between noise and stimulus related variance is in this case increased in favour of the noise. Thirdly, the joint model we consider may not be able to fully capture and perfectly explain the variability in EBA.

EBA was originally defined as a category selective area associated with body representation but the computations underlying this selective response are not yet well understood. Previous proposals stressed the role of EBA for individual body parts but not whole body images (Downing et al., 2001; Downing and Peelen, 2011; Peelen and Downing, 2007). These results are difficult to combine with evidence that EBA is selective for human bodies when only represented as stick figures, line drawings or silhouettes (Atkinson et al., 2012). Our findings are consistent with the latter hypothesis as the kp2d/3d model explained (significantly) more variance than the competing model in EBA.

The Gabor model proposed by Nishimoto et al. (2011) was specifically constructed to encode low-level visual features such as spatial frequency, location, size and object orientation. Gabor based models have been shown to be powerful tool for inferring (encoding/decoding) (Dumoulin and Wandell, 2008; Haynes and Rees, 2006; Kamitani and Tong, 2005; Kay et al., 2008; Miyawaki et al., 2008; Naselaris et al., 2009; Nishimoto et al., 2011; Nunez-Elizalde et al., 2019; Thirion et al., 2006; Wandell et al., 2007) brain activity inside and outside early visual cortex. Our findings suggest that the variance explained by the Gabor model shows a decreasing gradient from early to higher-level visual cortex. This suggests that within early sensory regions (superior occipital gyrus, blue patches in Fig. 4b) Gabor features are critical for predicting BOLD responses to body stimuli. Conversely, the variance explained by kp2d shows the opposite gradient, and it is highest in EBA. This suggests that postural features are critical in driving the response to body pictures in EBA. Interestingly, the transition between low-level features driving the response in early areas and mid-level (postural) features driving the response in high-level visual cortex (EBA) at the group level is smooth and suggests a dynamic, stimulus dependent, representation of bodies (de Gelder and Poyo Solanas, 2021). Likewise, simi-



**Fig. 5.** Comparison between encoding models. (a) RGB map in which each vertex is colour coded according to the relative contribution of each model to the accuracy of the joint model (red = 100% kp2d; blue = 100% Gabor; green = 100% kp3d). (b) In EBA, the information contained in the joint model predictions which significantly correlates with BOLD activity is split across models with kp2d accounting for 50–60% of the variance, Gabor approximately 25–30% of the variance and kp3d the remaining 15–20%. We tested for statistical difference across models' pair (solid lines at the top), using a two-sample *t*-test (\*\*\*)  $p < 0.0001$  (see bar plot). Additionally, the variance explained follows a gradient from the posterior part (posterior ITG/LOS) to the anterior (anterior LOS) of EBA, with darker shades of magenta in the posterior part indicating higher representation of low-level body features (Gabor), and lighter shades of magenta in the anterior part indicating higher representation of mid-level features (kp2d-kp3d). ITG = inferior temporal gyrus; MTG = middle temporal gyrus; LOS = lateral occipital sulcus.

lar patterns can be seen at the single subject level (see Supplementary material).

Another important point is the performance difference between kp2d and kp3d. These models represent body poses as the spatial location of specific keypoints (joints, hand, head etc.). In the case of kp3d, the keypoints represent the 3D coordinates used by VPoser to pose the mesh (Loper et al., 2015; Pavlakos et al., 2019) and construct the actual stimulus. Similarly, kp2d represents the orthogonal projection of the 3D coordinates on the camera plane. Therefore, the only difference between the models is that kp3d is isotropic (invariant across viewpoints), whereas the features of kp2d change across different view of the same pose. Our findings show that between kp2d/3d, banded ridge almost always selects the former as predictive and consider the latter as redundant. This is reflected in the percentage maps where on average kp2d outperforms kp3d. One possible explanation for this result is that the information contained in the 3rd dimension of kp3d was not needed to explain the variance in the data and, as a result, the selected feature space was kp2d for most of the voxels, suggesting that the viewpoint information is encoded in EBA. Previous research has shown that EBA is sensitive to body orientation (Chan et al., 2004; Ewbank et al., 2011; Foster et al., 2022; Taylor et al., 2010; Vangeneugden et al., 2014), although we did not find significant differences in brain activity when looking at differences between viewpoints (RFXGLM with three viewpoints as predictors of interest). This result is in line with what has been shown in single cell recordings on primates, where the MSB (analogous of the EBA in humans) showed strong viewpoint selectivity (Kumar et al., 2019).

It is worth mentioning that our stimuli were specifically controlled for not containing high-level stimulus attributes (i.e. emotion and action information) and validated using behavioural ratings (see 2.3, 3.1 and supplementary information) because our focus was on the representation of bodies in the occipitotemporal cortex (i.e. to focus primarily on bottom-up processing). This is also why our modelling efforts concentrate on models that are image computable (i.e.

they extract information from the stimulus directly). Previous literature has proposed that EBA/FBA might interact with the fronto-insular-temporal social context network (Amoruso et al., 2011; Downing and Peelen, 2011). Similarly, other studies have shown that activity in EBA is modulated by emotional body expressions (Costantini et al., 2005; de Gelder, de Borst, and Watson, 2015; Marsh et al., 2010; Moro et al., 2008; Pichon et al., 2012; Saxe et al., 2006; Tamietto et al., 2015; Van den Stock, Hortensius, Sinke, Goebel, and de Gelder, 2015). Moreover, a recent study has shown that unique information from the posture feature limb contractions is involved in fearful body expression perception (Poyo Solanas, Vaessen, and de Gelder, 2020). This literature indicates that social context can modulate activity in EBA/FBA if specific information regarding body or body parts is crucial to infer higher level attributes. Therefore, our goal was to control as much as possible for top-down modulation triggered when the stimuli have socially relevant attributes like emotions or actions.

Our results corroborate the notion that the functional EBA definition spans several anatomical regions with potentially different roles. Specifically, the EBA may be subdivided in three anatomical regions (Weiner and Grill-Spector, 2011) located respectively in the inferior temporal gyrus (ITG), middle temporal gyrus (MTG) and lateral occipital sulcus (LOS). When looking closely at the partial correlation patterns in EBA around the anatomical landmarks ITG, MTG, LOS (Fig. 5b and the barplot Fig. 5b) we see that not all the variance can be explained by combining the kp2d model with the Gabor model. This is graphically represented in Fig. 5b, where we find a green (or green derived) colour in the anterior part of EBA (anterior LOS/ITG), indicating that the variance explained by kp3d model is on average located in the more anterior portion of EBA. Specifically, bodies in the anterior portion of EBA are represented as a combination of kp2d/Gabor features with the kp3d model (yellow/light-blue patches in Fig. 5b). This finding might indicate that, as shown for early sensory regions, body representation in EBA is differentially encoded, going from a low-level representation (Gabor



like/blue patches) in pITG/pLOS, to a mid-level (viewpoint dependant) postural representation (kp2d, light-magenta, orange, pink patches) in the (middle) LOS to a high-level (viewpoint independent) postural representation (kp3d) in aITG/aLOS (green, light-blue, yellow patches).

Concerning the other major body selective region FBA, we observed that for the voxels significantly responding to localizer stimuli, the group definition of this region was not consistent across participants. Moreover, amongst the voxels functionally identified as part of the FBA, only few survived the statistical correction for multiple comparison of the encoding analysis. For completeness, we include the results of the encoding compared to EBA in the supplementary information. Briefly, the joint model performs significantly worse in FBA than in EBA, this could be due to low signal to noise ratio in the area. Nonetheless, the barplot depicting the percentage of the correlation explained by each model reveals a similar behaviour to what has been presented for EBA. The main difference is that kp3d model has an increase (from 20 to 25%) in percentage of correlation explained in FBA, at the expense of the correlation explained by kp2d. This is consistent with the fact that FBA has higher viewpoint tolerance than EBA as is expected if FBA is more involved in higher cognitive processing of body information like personal identity (Ewbank et al., 2011; Foster et al., 2021, 2019).

## 5. Conclusions

Taken together, our results suggest that the EBA encodes features pertaining specifically to posture. This representation appears to be viewpoint dependant posteriorly (pITG/pLOS) whereas greater viewpoint tolerance arises anteriorly (aITG/aLOS). On this account, the body selectivity observed in many studies in EBA is rooted in body specific feature representation that is not yet dependant of high order body categorisation processes. Future research must investigate whether these body selective features are rooted in uniquely defined biomechanical constraints, in human skeleton keypoint priors or also in sensorimotor processes.

## Data availability statement

Data and code are available for download at <https://doi.org/10.34894/TZA9S8>.

## Declaration of Competing Interest

None

## Credit authorship contribution statement

**Giuseppe Marrazzo:** Conceptualization, Investigation, Software, Formal analysis, Validation, Visualization, Writing – original draft, Writing – review & editing. **Federico De Martino:** Conceptualization, Investigation, Supervision, Validation, Writing – review & editing. **Agustin Lage-Castellanos:** Conceptualization, Supervision, Validation, Writing – review & editing. **Maarten J. Vaessen:** Conceptualization, Investigation, Software. **Beatrice de Gelder:** Conceptualization, Project administration, Resources, Supervision, Funding acquisition, Writing – original draft, Writing – review & editing.

## Acknowledgments

This work was supported by This work was supported by the European Research Council (ERC) FP7-IDEAS-ERC (Grant agreement number 295673; Emobodies), by the ERC Synergy grant (Grant agreement 856495; Relevance), by the Future and Emerging Technologies (FET) Proactive Program H2020-EU.1.2.2 (Grant agreement 824160; EnTime-Ment), by the Industrial Leadership Program H2020-EU.1.2.2 (Grant agreement 825079; MindSpaces), by the Horizon 2020 Programme H2020-FETPROACT430 2020–2 (grant 101017884 GuestXR).

The authors are grateful to Rufin Vogels and Giancarlo Valente for fruitful discussion and comments on the original draft and to anonymous reviewer for insightful comments and useful criticism. The authors declare no competing financial interests.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.neuroimage.2023.120240](https://doi.org/10.1016/j.neuroimage.2023.120240).

## References

- Adelson, E.H., Bergen, J.R., 1985. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Amer. A* 2 (2), 284–299. doi:10.1364/JOSAA.2.000284.
- Allen, E.J., Moerel, M., Lage-Castellanos, A., De Martino, F., Formisano, E., Oxenham, A.J., 2018. Encoding of natural timbre dimensions in human auditory cortex. *Neuroimage* 166, 60–70. doi:10.1016/j.neuroimage.2017.10.050.
- Amoruso, L., Couto, J., Ibanez, A., 2011. Beyond extrastriate body area (EBA) and fusiform body area (FBA): context Integration in the meaning of actions. *Front. Hum. Neurosci.* 5. doi:10.3389/fnhum.2011.00124.
- Atkinson, A.P., Vuong, Q.C., Smithson, H.E., 2012. Modulation of the face- and body-selective visual regions by the motion and emotion of point-light face and body stimuli. *Neuroimage* 59 (2), 1700–1712. doi:10.1016/j.neuroimage.2011.08.073.
- Bao, P., She, L., McGill, M., Tsao, D.Y., 2020. A map of object space in primate inferotemporal cortex. *Nature* 583 (7814), 103–108. doi:10.1038/s41586-020-2350-5.
- Bracci, S., Daniels, N., Op De Beeck, H., 2017. Task context overrules object- And category-related representational content in the human parietal cortex. *Cereb. Cortex* 27 (1), 310–321. doi:10.1093/cercor/bhw419.
- Bracci, S., Ritchie, J.B., Kalfas, I., Op De Beeck, H., 2019. The ventral visual pathway represents animal appearance over animacy, unlike human behavior and deep neural networks. *J. Neurosci.* 39 (33), 6513–6525. doi:10.1523/jneurosci.1714-18.2019.
- Brainard, D.H., 1997. The psychophysics toolbox. *Spat. Vis.* 10 (4), 433–436. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/9176952>.
- Bugatus, L., Weiner, K.S., Grill-Spector, K., 2017. Task alters category representations in prefrontal but not high-level visual cortex. *Neuroimage* 155, 437–449. doi:10.1016/j.neuroimage.2017.03.062.
- Carandini, M., Demb, J.B., Mante, V., Tolhurst, D.J., Dan, Y., Olshausen, B.A., Rust, N.C., 2005. Do we know what the early visual system does? *J. Neurosci.* 25 (46), 10577–10597. doi:10.1523/jneurosci.3726-05.2005.
- Caspari, N., Popivanov, I.D., De Mazière, P.A., Vanduffel, W., Vogels, R., Orban, G.A., Jastorff, J., 2014. Fine-grained stimulus representations in body selective areas of human occipito-temporal cortex. *Neuroimage* 102, 484–497. doi:10.1016/j.neuroimage.2014.07.066.
- Chan, A.W.Y., Peelen, M.V., Downing, P.E., 2004. The effect of viewpoint on body representation in the extrastriate body area. *Neuroreport* 15 (15). Retrieved from [https://journals.lww.com/neuroreport/Fulltext/2004/10250/The\\_effect\\_of\\_viewpoint\\_on\\_body\\_representation\\_in\\_21.aspx](https://journals.lww.com/neuroreport/Fulltext/2004/10250/The_effect_of_viewpoint_on_body_representation_in_21.aspx).
- Costantini, M., Galati, G., Ferretti, A., Caulo, M., Tartaro, A., Romani, G.L., Aglioti, S.M., 2005. Neural systems underlying observation of humanly impossible movements: an fMRI study. *Cereb. Cortex* 15 (11), 1761–1767. doi:10.1093/cercor/bhi053.
- de Gelder, B., de Borst, A.W., Watson, R., 2015. The perception of emotion in body expressions. *WIREs Cogn. Sci.* 6 (2), 149–158. doi:10.1002/wcs.1335.
- de Gelder, B., Poyo Solanas, M., 2021. A computational neuroethology perspective on body and expression perception. *Trends Cogn. Sci. Regul. Ed.* 25 (9), 744–756. doi:10.1016/j.tics.2021.05.010.
- de Gelder, B., Van den Stock, J., Meeren, H.K.M., Sinke, C.B.A., Kret, M.E., Tamietto, M., 2010. Standing up for the body. Recent progress in uncovering the networks involved in the perception of bodies and bodily expressions. *Neurosci. Biobehav. Rev.* 34 (4), 513–527. doi:10.1016/j.neubiorev.2009.10.008.
- DeAngelis, G.C., Ohzawa, I., Freeman, R.D., 1993. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *J. Neurophysiol.* 69 (4), 1091–1117. doi:10.1152/jn.1993.69.4.1091.
- Downing, P.E., Jiang, Y., Shuman, M., Kanwisher, N., 2001. A cortical area selective for visual processing of the human body. *Science* 293 (5539), 2470–2473. doi:10.1126/science.1063414.
- Downing, P.E., Peelen, M.V., 2011. The role of occipitotemporal body-selective regions in person perception. *Cogn. Neurosci.* 2 (3–4), 186–203. doi:10.1080/17588928.2011.582945.
- Downing, P.E., Peelen, M.V., 2016. Body selectivity in occipitotemporal cortex: causal evidence. *Neuropsychologia* 83, 138–148. doi:10.1016/j.neuropsychologia.2015.05.033.
- Dubois, J., de Berker, A.O., Tsao, D.Y., 2015. Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *J. Neurosci.* 35 (6), 2791–2802. doi:10.1523/jneurosci.4037-14.2015.
- Dumoulin, S.O., Wandell, B.A., 2008. Population receptive field estimates in human visual cortex. *Neuroimage* 39 (2), 647–660. doi:10.1016/j.neuroimage.2007.09.034.
- Dupré la Tour, T., Eickenberg, M., Gallant, J.L., 2022. Feature-space selection with banded ridge regression. *Neuroimage* 264, 119728. doi:10.1016/j.neuroimage.2022.119728.
- Ewbank, M.P., Lawson, R.P., Henson, R.N., Rowe, J.B., Passamonti, L., Calder, A.J., 2011. Changes in “top-down” connectivity underlie repetition suppression in the ventral visual pathway. *J. Neurosci.* 31 (15), 5635–5642. doi:10.1523/jneurosci.5013-10.2011.
- Foster, C., Zhao, M., Bolkart, T., Black, M.J., Bartels, A., Bühlhoff, I., 2021. Separated and overlapping neural coding of face and body identity. *Hum. Brain Mapp.* 42 (13), 4242–4260. doi:10.1002/hbm.25544.

- Foster, C., Zhao, M., Bolkart, T., Black, M.J., Bartels, A., Bühlhoff, I., 2022. The neural coding of face and body orientation in occipitotemporal cortex. *Neuroimage* 246, 118783. doi:10.1016/j.neuroimage.2021.118783.
- Foster, C., Zhao, M., Romero, J., Black, M.J., Mohler, B.J., Bartels, A., Bühlhoff, I., 2019. Decoding subcategories of human bodies from both body- and face-responsive cortical regions. *Neuroimage* 202, 116085. doi:10.1016/j.neuroimage.2019.116085.
- Fritz, L., Mulders, J., Bremner, H., Peters, J., Bastiani, M., Roebroeck, A., Goebel, R., 2014. Comparison of EPI distortion correction methods at 3T and 7T. In: Proceedings of the Paper presented at the Annual Meeting of the Organization for Human Brain Mapping, Hamburg, Germany, June 2014.
- Goebel, R., Esposito, F., Formisano, E., 2006. Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum. Brain Mapp.* 27 (5), 392–401.
- Greve, D.N., Fischl, B., 2009. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* 48 (1), 63–72.
- Haynes, J.D., Rees, G., 2006. Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* 7 (7), 523–534. doi:10.1038/nrn1931.
- Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12 (1), 55–67. doi:10.2307/1267351.
- Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. FSL. *Neuroimage* 62 (2), 782–790. doi:10.1016/j.neuroimage.2011.09.015.
- Jones, J.P., Palmer, L.A., 1987. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiol.* 58 (6), 1233–1258. doi:10.1152/jn.1987.58.6.1233.
- Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* 8 (5), 679–685. doi:10.1038/nn1444.
- Kanwisher, N., 2017. The quest for the FFA and where it led. *J. Neurosci.* 37 (5), 1056–1061. doi:10.1523/JNEUROSCI.1706-16.2016.
- Kay, K.N., Naselaris, T., Prenger, R.J., Gallant, J.L., 2008. Identifying natural images from human brain activity. *Nature* 452 (7185), 352–355. doi:10.1038/nature06713.
- Andersson J.L., Jenkinson M., & Smith S. (2007). Non-linear registration, aka Spatial normalisation FMRIB technical report TR07JA2. FMRIB Analysis Group of the University of Oxford, 2(1), e21. [www.fmrib.ox.ac.uk/analysis/techrep/tr07ja2/tr07ja2.pdf](http://www.fmrib.ox.ac.uk/analysis/techrep/tr07ja2/tr07ja2.pdf)
- Kleiner, M., Brainard, D.H., Pelli, D., 2007. What's new in. *Psychtoolbox-3? Perception*, 36. *ECVP Abstract Supplement*.
- Kumar, S., Popivanov, I.D., Vogels, R., 2019. Transformation of visual representations across ventral stream body-selective patches. *Cereb. Cortex* 29 (1), 215–229. doi:10.1093/cercor/bhx320.
- Lage-Castellanos, A., Valente, G., Formisano, E., De Martino, F., 2019. Methods for computing the maximum performance of computational models of fMRI responses. *PLoS Comput. Biol.* 15 (3), e1006397. doi:10.1371/journal.pcbi.1006397.
- Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J., 2015. SMPL: a skinned multi-person linear model. *ACM Trans. Graph.* 34 (6). doi:10.1145/2816795.2818013, Article 248.
- Marrazzo, G., Vaessen, M.J., de Gelder, B., 2021. Decoding the difference between explicit and implicit body expression representation in high level visual, prefrontal and inferior parietal cortex. *Neuroimage* 243, 118545. doi:10.1016/j.neuroimage.2021.118545.
- Marsh, A.A., Kozak, M.N., Wegner, D.M., Reid, M.E., Yu, H.H., Blair, R.J.R., 2010. The neural substrates of action identification. *Soc. Cogn. Affect. Neurosci.* 5 (4), 392–403. doi:10.1093/scan/nsq004.
- Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M.A., Morito, Y., Tanabe, H.C., Kamitani, Y., 2008. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron* 60 (5), 915–929. doi:10.1016/j.neuron.2008.11.004.
- Moerel, M., De Martino, F., Formisano, E., 2012. Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. *J. Neurosci.* 32 (41), 14205–14216. doi:10.1523/JNEUROSCI.1388-12.2012.
- Moro, V., Urgesi, C., Pernigo, S., Lanteri, P., Pazzaglia, M., Aglioti, S.M., 2008. The neural basis of body form and body action agnosia. *Neuron* 60 (2), 235–246. doi:10.1016/j.neuron.2008.09.022.
- Naselaris, T., Kay, K.N., Nishimoto, S., Gallant, J.L., 2011. Encoding and decoding in fMRI. *Neuroimage* 56 (2), 400–410. doi:10.1016/j.neuroimage.2010.07.073.
- Naselaris, T., Prenger, R.J., Kay, K.N., Oliver, M., Gallant, J.L., 2009. Bayesian reconstruction of natural images from human brain activity. *Neuron* 63 (6), 902–915. doi:10.1016/j.neuron.2009.09.006.
- Nishimoto, S., Gallant, J.L., 2011. A Three-dimensional spatiotemporal receptive field model explains responses of area mt neurons to naturalistic movies. *J. Neurosci.* 31 (41), 14551–14564. doi:10.1523/jneurosci.6801-10.2011.
- Nishimoto, S., Vu, A.T., Naselaris, T., Benjamini, Y., Yu, B., Gallant, J.L., 2011. Reconstructing visual experiences from brain activity evoked by natural movies. *Curr. Biol.* 21 (19), 1641–1646. doi:10.1016/j.cub.2011.08.031.
- Nunez-Elizalde, A.O., Huth, A.G., Gallant, J.L., 2019. Voxelwise encoding models with non-spherical multivariate normal priors. *Neuroimage* 197, 482–492. doi:10.1016/j.neuroimage.2019.04.012.
- Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A.A.A., Tzionas, D., Black, M.J., 2019. Expressive body capture: 3d hands, face, and body from a single image. In: Proceedings of the Paper presented at theIEEE/CVF Conference on Computer Vision and Pattern Recognition.
- Peelen, M.V., Atkinson, A.P., Andersson, F., Vuilleumier, P., 2007. Emotional modulation of body-selective visual areas. *Soc. Cogn. Affect. Neurosci.* 2 (4), 274–283. doi:10.1093/scan/nsm023.
- Peelen, M.V., Downing, P.E., 2005. Selectivity for the human body in the fusiform gyrus. *J. Neurophysiol.* 93 (1), 603–608. doi:10.1152/jn.00513.2004.
- Peelen, M.V., Downing, P.E., 2007. The neural basis of visual body perception. *Nat. Rev. Neurosci.* 8 (8), 636–648. doi:10.1038/nrn2195.
- Pelli, D.G., 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442.
- Pichon, S., de Gelder, B., Grèzes, J., 2012. Threat prompts defensive brain responses independently of attentional control. *Cereb. Cortex* 22 (2), 274–285. doi:10.1093/cercor/bhr060.
- Poyo Solanas, M., Vaessen, M.J., de Gelder, B., 2020. Computation-based feature representation of body expressions in the human brain. *Cereb. Cortex* 30 (12), 6376–6390. doi:10.1093/cercor/bhaa196.
- Ratan Murty, N.A., Bashivan, P., Abate, A., DiCarlo, J.J., Kanwisher, N., 2021. Computational models of category-selective brain regions enable high-throughput tests of selectivity. *Nat. Commun.* 12 (1), 5540. doi:10.1038/s41467-021-25409-6.
- Ross, P., de Gelder, B., Crabbe, F., Grosbras, M.H., 2020. A dynamic body-selective area localizer for use in fMRI. *MethodsX* 7, 100801. doi:10.1016/j.mex.2020.100801.
- Santoro, R., Moerel, M., De Martino, F., Goebel, R., Ugurbil, K., Yacoub, E., Formisano, E., 2014. Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. *PLoS Comput. Biol.* 10 (1), e1003412. doi:10.1371/journal.pcbi.1003412.
- Saxe, R., Jamal, N., Powell, L., 2006. My body or yours? the effect of visual perspective on cortical body representations. *Cereb. Cortex* 16 (2), 178–182. doi:10.1093/cercor/bhi095.
- Schwarzlose, R.F., Baker, C.I., Kanwisher, N., 2005. Separate face and body selectivity in the fusiform gyrus. *J. Neurosci.* 25 (47), 11055–11059. doi:10.1523/jneurosci.2621-05.2005.
- Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23 (Suppl 1), S208–S219. doi:10.1016/j.neuroimage.2004.07.051.
- Spiridon, M., Fischl, B., Kanwisher, N., 2006. Location and spatial profile of category-specific regions in human extrastriate cortex. *Hum. Brain Mapp.* 27 (1), 77–89. doi:10.1002/hbm.20169.
- Talairach, J., Tournoux, P., 1988. *Co-planar Stereotaxic Atlas of the Human brain: 3-Dimensional Proportional system: An approach to Cerebral Imaging*. Thieme Medical Publishers, Inc., New York.
- Tamietto, M., Cauda, F., Celeghini, A., Diano, M., Costa, T., Cossa, F.M., de Gelder, B., 2015. Once you feel it, you see it: insula and sensory-motor contribution to visual awareness for fearful bodies in parietal neglect. *Cortex* 62, 56–72. doi:10.1016/j.cortex.2014.10.009.
- Taylor, J.C., Downing, P.E., 2011. Division of labor between lateral and ventral extrastriate representations of faces, bodies, and objects. *J. Cogn. Neurosci.* 23 (12), 4122–4137. doi:10.1162/jocn\_a.00091.
- Taylor, J.C., Wiggett, A.J., Downing, P.E., 2007. Functional MRI analysis of body and body part representations in the extrastriate and fusiform body areas. *J. Neurophysiol.* 98 (3), 1626–1633. doi:10.1152/jn.00012.2007.
- Taylor, J.C., Wiggett, A.J., Downing, P.E., 2010. fMRI-adaptation studies of viewpoint tuning in the extrastriate and fusiform body areas. *J. Neurophysiol.* 103 (3), 1467–1477. doi:10.1152/jn.00637.2009.
- Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J.B., LeBihan, D., Dehaene, S., 2006. Inverse retinotopy: inferring the visual content of images from brain activation patterns. *Neuroimage* 33 (4), 1104–1116. doi:10.1016/j.neuroimage.2006.06.062.
- Urgesi, C., Candidi, M., Ionta, S., Aglioti, S.M., 2007. Representation of body identity and body actions in extrastriate body area and ventral premotor cortex. *Nat. Neurosci.* 10 (1), 30–31. doi:10.1038/nn1815.
- Van den Stock, J., Hortensius, R., Sinke, C.B.A., Goebel, R., de Gelder, B., 2015. Personality traits predict brain activation and connectivity when witnessing a violent conflict. *Sci. Rep.* 5 (1), 13779. doi:10.1038/srep13779.
- Vangeneugden, J., Peelen, M.V., Tadin, D., Battelli, L., 2014. Distinct neural mechanisms for body form and body motion discriminations. *J. Neurosci.* 34 (2), 574–585. doi:10.1523/jneurosci.4032-13.2014.
- Vogels, R., 2022. More than the face: representations of bodies in the inferior temporal cortex. *Annu. Rev. Vis. Sci.* 8 (1), 383–405. doi:10.1146/annurev-vision-100720-113429.
- Wandell, B.A., Dumoulin, S.O., Brewer, A.A., 2007. Visual field maps in human cortex. *Neuron* 56 (2), 366–383. doi:10.1016/j.neuron.2007.10.012.
- Watson, A.B., Ahumada, A.J., 1985. Model of human visual-motion sensing. *J. Opt. Soc. Amer. A* 2 (2), 322–342. doi:10.1364/JOSAA.2.000322.
- Weiner, K.S., Grill-Spector, K., 2010. Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *Neuroimage* 52 (4), 1559–1573. doi:10.1016/j.neuroimage.2010.04.262.
- Weiner, K.S., Grill-Spector, K., 2011. Not one extrastriate body area: using anatomical landmarks, hMT+, and visual field maps to parcellate limb-selective activations in human lateral occipitotemporal cortex. *Neuroimage* 56 (4), 2183–2199. doi:10.1016/j.neuroimage.2011.03.041.
- Woolrich, M.W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Smith, S.M., 2009. Bayesian analysis of neuroimaging data in FSL. *Neuroimage* 45 (1 Suppl), S173–S186. doi:10.1016/j.neuroimage.2008.10.055.