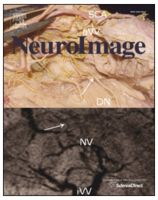
NeuroImage 102 (2014) 484–497

Contents lists available at ScienceDirect

NeuroImage

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



Fine-grained stimulus representations in body selective areas of human occipito-temporal cortex

Natalie Caspari [a], Ivo D. Popivanov [a], Patrick A. De Mazière [b,c], Wim Vanduffel [a,d,e], Rufin Vogels [a], Guy A. Orban [f], Jan Jastorff [a,g,*]

[a] Laboratorium voor Neuro-en Psychofysiologie, KU Leuven, Leuven, Belgium
[b] Department of Healthcare & Technology, KH Leuven, Leuven, Belgium
[c] Department of Computer Sciences, KU Leuven, Leuven, Belgium
[d] Harvard Med. Sch., Boston, MA, USA
[e] MGH Martinos Ctr., Charlestown, MA, USA
[f] Department of Neuroscience, University of Parma, Parma, Italy
[g] Division of Psychiatry, Department of Neuroscience, KU Leuven, Leuven, Belgium

ARTICLE INFO

Article history:
Accepted 30 July 2014
Available online 7 August 2014

Keywords:
Human
Monkey
Extrastriate body area
Fusiform body area
Multi-voxel pattern analysis
Bodies
Functional imaging

ABSTRACT

Neurophysiological and functional imaging studies have investigated the representation of animate and inanimate stimulus classes in monkey inferior temporal (IT) and human occipito-temporal cortex (OTC). These studies proposed a distributed representation of stimulus categories across IT and OTC and at the same time highlighted category specific modules for the processing of bodies, faces and objects. Here, we investigated whether the stimulus representation within the extrastriate (EBA) and the fusiform (FBA) body areas differed from the representation across OTC. To address this question, we performed an event-related fMRI experiment, evaluating the pattern of activation elicited by 200 individual stimuli that had already been extensively tested in our earlier monkey imaging and single cell studies (Popivanov et al., 2012, 2014). The set contained achromatic images of headless monkey and human bodies, two sets of man-made objects, monkey and human faces, four-legged mammals, birds, fruits, and sculptures. The fMRI response patterns within EBA and FBA primarily distinguished bodies from non-body stimuli, with subtle differences between the areas. However, despite responding on average stronger to bodies than to other categories, classification performance for preferred and non-preferred categories was comparable. OTC primarily distinguished animate from inanimate stimuli. However, cluster analysis revealed a much more fine-grained representation with several homogeneous clusters consisting entirely of stimuli of individual categories. Overall, our data suggest that category representation varies with location within OTC. Nevertheless, body modules contain information to discriminate also non-preferred stimuli and show an increasing specificity in a posterior to anterior gradient.

Introduction

In our everyday life, we encounter numerous visual stimuli we can easily identify. Neurons sensitive to object properties relevant for identification and categorization have been described in monkey inferior temporal cortex (IT) and are thought to be present in human occipito-temporal cortex (OTC) (Kourtzi and Connor, 2011). Kiani et al. (2007) recorded neural responses across anterior IT cortex to natural and artificial object images. They found that the categorical structure of the objects was represented by the pattern of activity distributed over the cell population. The major distinction was present between animate and inanimate objects; nevertheless, the category of animate objects was further divided into faces and bodies, which were divided further into several finer grained categories like human faces and monkey faces, or human bodies and four-limbed animal bodies. This organization seems to be similar at least at the core structure in human OTC, where a primary animate–inanimate organization with further sub-division of the animate category into faces and bodies has been reported (Caramazza and Shelton, 1998; Cichy et al., 2014; Connolly et al., 2012; Huth et al., 2012; Kriegeskorte et al., 2008). Whereas several studies have investigated the organization of IT/OTC with respect to different categories, much less is known about their representation within individual face and body selective regions. Yet, comparing the representation within category selective regions with the representation across whole IT/OTC might shed new light on the

---

[*] Corresponding author at: Laboratorium voor Neuro-en Psychofysiologie, KU Leuven Medical School, Herestraat 49, bus 1021, 3000 Leuven, Belgium. Fax: +32 16 345993.
E-mail address: jan.jastorff@med.kuleuven.be (J. Jastorff).

http://dx.doi.org/10.1016/j.neuroimage.2014.07.066
1053-8119/

discussion of modular versus distributed coding of object information in general (Haxby et al., 2001; Reddy and Kanwisher, 2006).

Electrophysiological studies reported a high fraction of face-selective cells and strong face category selectivity within fMRI defined face patches (Issa and DiCarlo, 2012; Tsao et al., 2006). The same conclusion was also drawn from fMRI experiments in monkeys and humans focusing on face selective patches in IT and OTC respectively (Liu et al., 2013; Reddy and Kanwisher, 2006). These results seemed to favor a modular organization with respect to faces. However, little is known about the categorical structure within body selective regions. Two single cell studies of the middle (Popivanov et al., 2014) and presumably anterior (Bell et al., 2011) superior temporal sulcus (STS) body patches in monkeys reported that the majority of neurons responded stronger to body compared to non-body stimuli. However, the selectivity for body stimuli was much lower compared to the one reported for faces in the face patches. Also, both studies reported single cells within the body patches that were highly selective for other categories. Consequently cluster analysis in the mid STS body patch showed an initial clustering of body versus non-body stimuli, where the non-body cluster in turn contained two distinct sub-clusters, perfectly separating faces from inanimate objects (Popivanov et al., 2014). How these results translate to the human is presently unknown, because fine grained stimulus clustering in the extrastriate (EBA (Downing et al., 2001)) and the fusiform (FBA, (Peelen and Downing, 2005; Schwarzlose et al., 2005)) body areas has not been investigated till date.

To address this issue, we collected individual human fMRI responses to 200 stimuli showing monkey and human bodies and faces, four-legged mammals, birds, fruits, sculptures and man-made objects. The stimulus set was identical to the one used in our previous monkey imaging and single cell studies (Popivanov et al., 2012, 2014). We compared the representational similarity among these stimuli at the fMRI voxel level within body selective regions and also across whole human OTC to investigate the following questions: 1) What is the fine grained categorical structure representing a large number of animate and inanimate objects in EBA and FBA, 2) is there any difference in the representation between body selective regions and whole OTC and finally, 3) can we use the representational structure to address potential homologies between human and monkey body-selective regions?

Materials and methods

Participants

Eight volunteers (4 females, mean age 25 years, range 23–32 years) participated in the experiment. All participants were right-handed, had normal or corrected-to-normal visual acuity and no history of mental illness or neurological diseases. The study was approved by the Ethical Committee of KU Leuven Medical School and all volunteers gave their written informed consent in accordance with the Helsinki Declaration prior to the experiment. One subject (male) was excluded from the analysis, because of fatigue and decreased fixation performance during the experiment. The present data are based on the remaining seven subjects.

Stimuli

The stimuli were identical to those used in the monkey fMRI study of Popivanov et al. (2012). Ten classes of achromatic images – human and monkey bodies (excluding the head), human and monkey faces, four-legged mammals, birds, two classes of man-made objects (matched in terms of low level properties either to the human or to the monkey bodies), fruits/vegetables and body-like sculptures (by the British artist H. Moore) – served as stimuli. Each class consisted of 20 images (Fig. 1a).

The images of human bodies were from Downing et al. (2001), yet two stimuli were exchanged with other bodies to better match the low level properties, as described below. The human face stimuli were obtained from the Tarrlab stimulus repository (courtesy of M. J. Tarr — http://www.tarrlab.org/) and the NBU Faces Database (http://nbufaces.yobul.com/ENAboutDatabase.aspx). These photographs depicted different individuals and were taken from different viewpoints. The images of monkey bodies and faces were cut from photographs of 4 male monkeys from our colony. They depicted headless bodies in different poses and faces that varied in both orientation and viewpoint (varying between frontal and profile views). All other stimuli were generated from images downloaded from the public domain.

Since monkey IT neurons are known to be sensitive to low-level image characteristics like aspect ratio (Kayaert et al., 2005) and area (Baldassi et al., 2013), we carefully controlled several of the low-level properties across the different classes of stimuli. The mean aspect ratio was computed as the ratio between the first and the second Eigen vectors of the object's image (Fig. 1b). As one can expect, there was a systematic difference in mean aspect ratio when the human bodies, depicted predominantly standing upright, were compared to the monkey bodies, which were depicted on four legs or seated. For this reason, two classes of manmade objects were used — one matching the aspect ratio of the human bodies and the other matching the aspect ratio of the monkey bodies.

The mean luminance values (Fig. 1c) were equated across classes. The mean contrast was defined as the root-mean-square contrast (Peli, 1990) normalized by the mean luminance of the object's image (Fig. 1d). As the mean contrast of the sculptures was on average highly similar to the mean contrast of the human bodies (Fig. 1d), variation in contrast cannot account for differences in body and non-body classes, in general. The images were resized so that the average area per class was matched across classes (Fig. 1e) while still allowing some variation in area (range 3.7° (square root of the area) to 6.7°) within each class. The area of the human bodies and the human control objects was smaller than the other classes, as their vertical extent would otherwise exceed 15° because of their distinct aspect ratio. This variation in size avoided clustering based on low-level image features. The mean vertical and horizontal extent of all images was 8.3° and 6.7° of visual angle, respectively. All image transformations (e.g. isolating the image from background, resizing, rotating, and adjusting the grayscale values) were performed using Adobe Photoshop CS3 and the Image Processing Toolbox of Matlab (The Mathworks, Inc., Natick, MA). The images were embedded into pink noise backgrounds, filling the entire display and having the same mean luminance as the images. Pink noise has a spatial frequency power spectrum similar to that of natural images. Each image was presented on top of 5 different backgrounds that were changed randomly during the experiment. A red fixation dot was superimposed on every stimulus as well as during the inter stimulus interval (ISI).

Procedure

Each subject participated in 5 scanning sessions, on five separate days. These included four sessions for the main event-related experiment and one session for the localizer scans. A single scanning session lasted about 90 min. Before scanning, participants were familiarized with the different conditions outside the scanner and were instructed to maintain fixation on the red fixation dot throughout the experiment.

Event-related experiment

Each image was presented for 480 ms with an ISI jittering between 2300 ms and 3500 ms. A single time series (run) lasted 528 s and 110 images were presented, 10 for each category, plus 10 images of the baseline fixation condition, which consisted of a red fixation dot superimposed on a pink noise background. The background pattern was changed with the presentation of each stimulus. As the stimulus