NeuroImage xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

NeuroImage

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

The encoding of category-specific versus nonspecific information in human inferior temporal cortex

Bingbing Guo, Ming Meng *

Dartmouth College, Department of Psychological and Brain Sciences, 6207 Moore Hall, Hanover, NH 03755, USA

ARTICLE INFO

Article history:
Received 29 January 2015
Accepted 3 April 2015
Available online xxxx

Keywords:
Face processing
Object recognition
Visual contrast
Inferior temporal cortex
Multivariate pattern analysis
fMRI

ABSTRACT

Several brain areas in the inferior temporal (IT) cortex, such as the fusiform face area (FFA) and parahippocampal place area (PPA), are hypothesized to be selectively responsive to a particular category of visual objects. However, how category-specific and nonspecific information may be encoded at this level of visual processing is still unclear. Using fMRI, we compared averaged BOLD activity as well as multi-voxel activation patterns in the FFA and PPA corresponding to high-contrast and low-contrast face and house images. The averaged BOLD activity in the FFA and PPA was modulated by the image contrast regardless of the stimulus category. Interestingly, unlike the univariate averaged BOLD activity, multi-voxel activation patterns in the FFA and PPA were barely affected by variations in stimulus contrast. In both the FFA and PPA, decoding the categorical information about whether participants saw faces or houses was independent of stimulus contrast. Moreover, the multivariate pattern analysis (MVPA) results were highly stable when either the voxels that were more sensitive to stimulus contrast or the voxels that were less sensitive were used. Taken together, these findings demonstrate that both category-specific (face versus house) information and nonspecific (image contrast) information are available to be decoded orthogonally in the same brain areas (FFA and PPA), suggesting that complementary neural mechanisms for processing visual features and categorical information may occur in the same brain areas but respectively be revealed by averaged activity and multi-voxel activation patterns. Whereas stimulus strength, such as contrast, modulates overall activity amplitudes in these brain areas, activity patterns across populations of neurons appear to underlie the representation of object category.

Introduction

The inferior temporal (IT) cortex is known to play an important role in visual categorization and recognition (Miller et al., 1991; Hung et al., 2005). The early visual areas in the occipital lobe are thought to process low-level features (Hubel and Wiesel, 1962, 1965), while the synthesis of these features to form high-level object representations is presumed to occur in the IT cortex (Felleman and Van Essen, 1991; DiCarlo and Maunsell, 2000; Yamins et al., 2014). Consistent with this notion, previous neurophysiological and neuroimaging studies have shown object-selective properties of IT neurons in nonhuman primates (Gross et al., 1972; Desimone et al., 1984; Tsao et al., 2003, 2006; Pinsk et al., 2005; Freiwald et al., 2009; Freiwald and Tsao, 2010). Through fMRI measurements of BOLD activity, a few regions in the human IT cortex have also been identified to be selectively responsive to some particular categories of objects (Kanwisher, 2010). Most notably, an area in the lateral fusiform gyrus has been hypothesized to be selectively responsive to images of faces (the fusiform face area, FFA) (Kanwisher et al., 1997), whereas the parahippocampal place area (PPA) (Epstein and Kanwisher, 1998) has been hypothesized to respond to images of places more strongly than to images of faces.

How would perceptual processing lead to categorical representations in the IT cortex? Note that visual features such as contrast are not specific to any object categories and can vary continuously in a physically measurable dimension. There may be a detection threshold, but if, for example, neurons in the FFA were truly the "face cells", one would assume that the response function of these neurons should be invariant to low-level visual features such as contrast, as long as the contrast is above the detection threshold and the face stimuli can be recognized. If, however, the non-specific information modulates responses in the FFA, some other representation mechanism should be able to encode the category-specific information that is invariant to low-level features at a later stage. Indeed, a recent fMRI study found that lower-level visual features significantly modulate the averaged BOLD activation in the FFA (Yue et al., 2011). This result suggests that averaged BOLD activity in the FFA potentially reflects stimulus properties that are non-specific to faces, and may not be directly used to encode category-specific information that should be invariant to low-level features. It remains unknown whether the contrast of non-face images may also modulate FFA activity, or whether other brain areas in the IT cortex (e.g., PPA) may be modulated by stimulus contrast.

* Corresponding author. Fax: +1 603 646 1419.
E-mail address: ming.meng@dartmouth.edu (M. Meng).

http://dx.doi.org/10.1016/j.neuroimage.2015.04.006
1053-8119/

An alternative to the hypothesis that object categories are encoded by category-selective neurons is the possibility that the encoding of object categories is accomplished through the activity of a distributed population of neurons (Gochin et al., 1994; Haxby et al., 2001). If a distributed ensemble of neurons in the IT cortex encodes faces (Haxby et al., 2000; Nestor et al., 2011), activation patterns would reflect the categorical representation of faces better, instead of univariate averaged BOLD activity. On the one hand, signal strength of the input (e.g. high contrast) may drive neurons to fire more vigorously, leading to overall increased BOLD activity (Heeger et al., 2000). On the other hand, activation patterns do not have to change, providing the possibility for the encoding of object categories. Hypothetically, the transition of local averaged neuronal activity to an activation pattern of distributed populations of neurons would fit naturally with the non-linear transformation of encoding from continuous visual features to discrete categorical information. We test this hypothesis in the present study by comparing univariate averaged BOLD activity with the results from multivariate pattern analysis (MVPA) of fMRI activity. MVPA uses machine-learning techniques (e.g., PyMVPA, see Hanke et al., 2009) to discriminate multi-voxel patterns of brain activity corresponding to different experimental conditions (e.g., seeing faces versus seeing houses). Whereas univariate averaged BOLD activity may correlate with the local averaged neuronal activity (Bandettini and Ungerleider, 2001; Heeger and Ress, 2002; Logothetis et al., 2001), activation patterns of a distributed population of neurons can be revealed by the MVPA (e.g., Haxby et al., 2001). If categorical representation occurs through encoding by populations of neurons in the IT cortex, we expect that the decoding through MVPA should be independent of stimulus contrast, despite the fact that averaged BOLD activity is being significantly modulated by stimulus contrast (e.g., see Yue et al., 2011). If, however, the encoding in the IT cortex is not independent of stimulus contrast, MVPA results would also be modulated by stimulus contrast.

Materials and methods

Seventeen healthy adults, (mean age = 26 years, seven females) with normal or corrected-to-normal visual acuity, participated in this study. The data from five participants were excluded from further analysis because of excessive head motion (>3 mm). All participants gave written informed consent. The study was approved by the Committee for the Protection of Human Subjects at Dartmouth College.

MRI acquisitions

Scanning was performed on a 3.0 T Philips Achieva Intera scanner with a 32-channel head coil at the Dartmouth Brain Imaging Center. The BOLD signals were collected using an echo-planar imaging (EPI) sequence (TR = 2000 ms, TE = 35 ms, flip angle = 90°, FOV = 240 mm, voxel size = 3 × 3 × 3 mm, 35 slices). For each participant, a high-resolution magnetization-prepared rapid-acquisition gradient echo (MPRAGE) anatomical scan was acquired at the end of the scan session (TR = 8.2 ms, TE = 3.8 ms, flip angle = 8°, FOV = 240 mm, voxel size = 1 × 1 × 1 mm, 222 slices). During the EPI scans, visual stimuli were presented to a screen located at the back of the scanner via a LCD projector (Panasonic PT-D4000U, 1024 × 768 pixel resolution) using MATLAB 2011b with Psychtoolbox (Brainard, 1997; Pelli, 1997). Participants viewed the stimuli using a mirror placed within the head coil. Stimuli subtended a visual angle of 8.7°.

ROI localizer runs

To localize the FFA and PPA as the regions of interest (ROIs), visual stimuli were chosen from an independent set of gray-scale images of faces and houses. The localizer scans consisted of an alternating block design, with 5 stimulation blocks presenting face images and 5 stimulation blocks presenting house images interleaved with 16-s periods of a blank screen with a fixation cross in the center. Each stimulation block was also 16-s long. In total, each localizer scan was 336-s long, consisting of 11 periods of fixation and 5 blocks for each of the two stimulation categories. In each stimulation block, 16 faces (or houses) were presented (500-ms per image, with a 500-ms interstimulus interval). Each participant completed two localizer scans. During these scans, participants performed a one-back task in which they were asked to make a key-press whenever an image was repeated consecutively.

Experimental runs

The stimuli set included gray-scale images of eight conditions: low-contrast faces, high-contrast faces, low-contrast houses, high-contrast houses and four conditions for a separate study that is not directly relevant. The high contrast level was defined with root mean square (RMS) = 0.25 in normalized unit, whereas the low contrast level RMS = 0.025 in normalized unit. Corresponding contrasts of faces and houses were made equal by using the SHINE toolbox (Willenbockel et al., 2010). The experimental scans consisted of an alternating block design, with 16-s blocks of stimulation interleaved with 16-s of fixation periods. In total, each experimental scan was 272-s long, consisting of 9 fixation periods and 1 stimulation block for each of the eight conditions. In each stimulation block, 8 images from a condition were presented (1700-ms per image, with a 300-ms interstimulus interval). Each participant completed 9–10 experimental scans. During these scans, participants performed a color detection task in which they were asked to press a button whenever the entire image was presented in red for 200-ms at random times across every stimulation block. The purpose of this task was to ensure that participants had been attentive to the stimuli.

Data analysis

AFNI (http://afni.nimh.nih.gov/afni) was used for preprocessing the MRI data. EPIs were motion corrected to the image acquired closest to the anatomical images, spatially smoothed with a 4-mm full width at half maximum (FWHM) filter, and temporally filtered to remove baseline drifts. The anatomical images were aligned to the functional images to avoid the additional resampling and interpolation of functional images.

Data from the localizer scans were further submitted to a General Linear Model (GLM) analysis, which allowed the calculation of beta coefficient values associated with block conditions. ROIs were individually defined for each participant based on activation maps from the GLM analysis. Each ROI was defined as a continuous cluster of activated voxels corresponding to the following statistical contrasts: the FFA was defined in the right middle fusiform gyrus as responding more strongly to faces than to houses ($p < 10^{-4}$, uncorrected), and the PPA was defined in the right parahippocampal gyrus as responding more strongly to houses than to faces ($p < 10^{-4}$, uncorrected). To control for any potential confounding effects of ROI size, the FFA and PPA were localized with roughly the same number of voxels (~40 voxels). We focus on results of ROIs in the right hemisphere here. Results in the left hemisphere are similar, except that by using the same criteria, we were only able to successfully identify the left FFA with ~40 voxels in eight participants. This is consistent with previous reports about lateralization of the FFAs in humans (e.g., Kanwisher et al., 1997; Behrmann and Plaut, 2013).

For comparison, we also used an ROI in the early visual cortex. The Brodmann area 17 (BA17) was localized using an anatomical mask as well as BOLD contrasts during ROI localizer scans. The anatomical mask of BA17 (TT_N27 template) was individually aligned to the anatomical images of each participant. Using the BOLD contrasts acquired through analyzing the ROI localizer scans, activated voxels were localized in the calcarine sulcus that responded more strongly during stimulation blocks than during fixation periods ($p < 10^{-4}$, uncorrected). The

Please cite this article as: Guo, B., Meng, M., The encoding of category-specific versus nonspecific information in human inferior temporal cortex, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.04.006