NeuroImage 73 (2013) 8–15

Contents lists available at SciVerse ScienceDirect

NeuroImage

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

Multi-voxel pattern analysis of selective representation of visual working memory in ventral temporal and occipital regions

Xufeng Han [a], Alexander C. Berg [a], Hwamee Oh [c], Dimitris Samaras [a], Hoi-Chung Leung [b,*]

[a] Department of Computer Science, State University of New York at Stony Brook, USA
[b] Department of Psychology, State University of New Yrok at Stony Brook, USA
[c] Helen Wills Neuroscience Institute, University of California, Berkeley, USA

ARTICLE INFO

Article history:
Accepted 26 January 2013
Available online 4 February 2013

Keywords:
Visual working memory
Inferior temporal cortex
Fusiform gyrus
Parahippocampal gyrus
Face and scene
fMRI

ABSTRACT

While previous results from univariate analysis showed that the activity level of the parahippocampal gyrus (PHG) but not the fusiform gyrus (FG) reflects selective maintenance of the cued picture category, present results from multi-voxel pattern analysis (MVPA) showed that the spatial response patterns of both regions can be used to differentiate the selected picture category in working memory. The ventral temporal and occipital areas including the PHG and FG have been shown to be specialized in perceiving and processing different kinds of visual information, though their role in the representation of visual working memory remains unclear. To test whether the PHG and FG show spatial response patterns that reflect selective maintenance of task-relevant visual working memory in comparison with other posterior association regions, we reanalyzed data from a previous fMRI study of visual working memory with a cue inserted during the delay period of a delayed recognition task. Classification of FG and PHG activation patterns for the selected category (face or scene) during the cue phase was well above chance using classifiers trained with fMRI data from the cue or probe phase. Classification of activity in other temporal and occipital regions for the cued picture category during the cue phase was relatively less consistent even though classification of their activity during the probe recognition was comparable with the FG and PHG. In sum, these findings suggest that the FG and PHG carry information relevant to the cued visual category, and their spatial activation patterns during selective maintenance seem to match those during visual recognition.

Introduction

Studies of human and nonhuman primates have consistently shown that the ventral temporal and occipital regions are involved in the perception and recognition of visual stimuli (see review by Ungerleider and Haxby, 1994). These visual association regions in the posterior cortex show functional divisions specializing in categorical representation of objects such as faces, tools, words, etc. (e.g., Chao et al., 1999; Epstein and Kanwisher, 1998). It has been proposed that these regions are also involved in supporting visual working memory – the short-term representation of visual stimuli that are no longer physically available (Postle, 2006; Ranganath and D'Esposito, 2005). Neuroimaging findings, however, have been inconsistent thus far. Some showed that the inferior temporal region (e.g., the lateral fusiform gyrus) was active in tasks requiring holding faces (e.g., Druzgal and D'Esposito, 2003; Postle et al., 2003; Ranganath et al., 2004) and in tasks requiring refreshing recently seen faces (e.g., Johnson et al., 2007). Others, however, showed that the activity in the inferior temporal region was not long lasting (Jha and McCarthy, 2000) and subject to interference (Miller et al., 1993; Sreenivasan et al., 2007; but see Yoon et al., 2006 for different results).

Some investigators further examined the selectivity of the posterior visual association regions in representing specific visual working memory. Face and/or scene images were used as task stimuli in neuroimaging studies since the fusiform (FG) and parahippocampal gyri (PHG) are known to be more specialized in processing faces and scenes, respectively (e.g., Epstein and Kanwisher, 1998; Kanwisher et al., 1997). Participants were cued to remember a particular category of visual stimuli (e.g., Remember face but ignore scene, and vice versa), with the cue presented either prior to stimulus presentation for selective encoding (Gazzaley et al., 2005; Nobre et al., 2004) or after, for selective maintenance (Lepsien et al., 2005; Oh and Leung, 2010). Across studies, the PHG consistently showed elevated activity during selective encoding and selective maintenance of scene images. The FG, however, did not always show differential activity for selective processing of faces (compare: Gazzaley et al., 2005; Oh and Leung, 2010). A recent fMRI study reported that neither PHG nor FG was modulated by the number of face/scene images to be selectively maintained in working memory (Lepsien et al., 2011). Thus, it is unclear to what

* Corresponding author at: SUNY Stony Brook, Department of Psychology, Stony Brook, NY 11794–2500, USA. Fax: +1 631 632 7876.
E-mail address: hoi-chung.leung@stonybrook.edu (H.-C. Leung).

1053-8119/$ – see front matter
http://dx.doi.org/10.1016/j.neuroimage.2013.01.055

extent the different posterior association regions are involved in representing task-relevant visual working memory.

Most previous studies reviewed above applied univariate analysis to determine whether or not a brain region is activated while particular visual information is assumed to be held in working memory. Using multiple voxel pattern analysis (MVPA), recent studies successfully showed differential spatial patterns of activation in both striate and extrastriate areas for holding visual features (e.g., orientations, Harrison and Tong, 2009) and visual categories (e.g., faces, scenes and objects; Lewis-Peacock and Postle, 2008; Lewis-Peacock et al., 2012 [Experiment 1]). Through reanalyzing data from the second experiment of the Lewis-Peacock et al. (2012) study, Lewis-Peacock and Postle (2012) showed that their results on classification of task-relevant category (out of three potential categories: pseudowords, words, and line orientations) during the delay period were not affected even after excluding the suprathreshold voxels identified by the general linear model (GLM) as category-specific. Here, we further examined the activation patterns of the FG and PHG as well as other specific temporal/occipital regions in response to cued selective maintenance of task-relevant visual working memory in the presence of no-longer-relevant working memory.

We applied MVPA to previously published data (Oh and Leung, 2010) and conducted within-subject analysis to examine the activation patterns in the FG, PHG and other ventral temporal and occipital regions during selective maintenance of face/scene images. The task (Fig. 1A) comprised three phases: initial encoding (remembering two pictures, a face and a scene), selective maintenance (maintaining one of the two pictures according to a text cue), and recognition (judging whether the probe image is an exact match of the cued picture). We first trained and tested classifiers using activation patterns from the cue phase and examined classification performance across time during selective maintenance. In addition, we trained classifiers using activation patterns from the probe phase and from a separate localizer task, and tested these different classifiers on the cue-phase data to confirm that classification results for selective maintenance of faces/scenes are not due to the word cue itself. We were particularly interested in the FG and other ventral temporal and occipital regions involved in face processing since many of these regions did not show differential activity during selective maintenance in previous univariate analysis (see Fig. 1B).

Methods

We used the 12 datasets from a study published by Oh and Leung (2010). A detailed description of the experimental procedure and image preprocessing can be found in that paper. Here, we provide a brief summary on the task design and image acquisition and processing procedures.

Working memory and localizer tasks

The fMRI data were collected while participants performed a visual working memory task and a localizer task. For the main visual working memory task (Fig. 1A), we used a variant of the delayed recognition paradigm with a cue inserted during the delay period to study selective maintenance of faces or scenes. At the beginning of each trial, a fixation point (a small green square) was presented for 3 s and, as a warning, it turned into red color briefly before stimulus presentation. Two pictures (a face and a scene) were presented sequentially in counterbalanced order, each for 800 ms, with a 200-ms gap in between. A mask (black-and-white checkerboard) was displayed for 200 ms after the offset of the second stimulus. After a delay of 2.5 s, a cue word (e.g., "face", "scene") was presented in the center of the screen for 1 s. This cue indicated the picture category relevant for the recognition test 9.5-s later. All cues were 100% valid. For trials with the face cue, the participants would only need to continually hold the memorized face picture as the probe would be either the cued face or a new face. It was the opposite for trials with the scene cue. The participants made button presses to indicate whether or not the probes matched the to-be-remembered picture. The inter-trial interval (ITI) varied between 8 and 14 s with a mean of 11 s. There were 20 trials with the face cue and 20 trials with the scene cue.

The localizer task was in a 1-back working memory format. There were 8 task blocks (4 face blocks and 4 scene blocks) separated by resting fixation blocks. Each block was 16 s long. Within each task block, eight pictures were sequentially presented, each for 800 ms, with a gap of 1.2 s between the stimuli. The participants made a same/different response to each picture indicating whether or not it matched the preceding one.

Image data acquisition, preprocessing and defining ROIs

Anatomical and functional MR images were acquired with a 3 T Philips Achieva system using the standard quadrature head coil (8 channels). The acquisition parameters for the echo-planar (EPI) images were as follows for the main working memory task: 24 axial-oblique 5-mm slices/volume, 245 volumes/run, TR = 1.5 s, TE = 30 ms, flip angle = 80°, FOV = 220 × 220 mm, matrix = 64 × 64 and ascending acquisition from the bottom slice. Similar parameters were applied for the localizer task, except that a 2-s TR was used instead. All preprocessing steps were conducted using SPM2 (Wellcome Department of Cognitive Neurology, London, UK.) as reported in Oh and Leung (2010). Functional images were corrected for differences in slice timing and head motion. Images were normalized to the MNI gray matter template (Friston et al., 1995). We used smoothed images (8 mm Gaussian kernel), since we found little or no differences in our classification results using either nonsmoothed or smoothed images in a preliminary test.

Image data from the localizer task were used to define the visual association regions for each subject. Fig. 1C illustrates the locations of the regions of interest (ROIs). The primary regions of interest included three inferior temporal and occipital areas that showed greater activation in the face > scene contrast (FG, OFA [occipitotemporal face area], STS [posterior superior temporal sulcus]), and three areas that showed greater activation in the scene > face contrast (PHG, TOS [transverse occipital sulcus], and RSC [retrosplenial cortex]). These areas were defined in each hemisphere following the literature (Fox et al., 2009; Nasr et al., 2011) and guided by anatomy and group-level contrasts. Contrasts were thresholded at $t > 3$. ROIs were defined as spheres (radius = 3 voxels or 10.5 mm; approximately 123 voxels in volume) centered on the coordinates of the peak of the suprathreshold clusters in each individual. For a few subjects, we either used a lower threshold or used the contrast with fixation baseline to identify the coordinates; this was the case for TOS (1 right, 1 left), RSC (3 right, 3 left), OFA (3 left, 3 right), FG (1 left, 1 right), STS (3 right, 2 left). For the two subjects where we could not identify activations in the RSC even at a lower threshold ($t > 1$), we used the mean coordinates from the other subjects in the group.

Pattern classification analysis

We applied linear support vector machines (SVMs) to examine the spatial response patterns in specific brain regions during the cue phase of the main task for predicting the face/scene cues. Regions were selected from the 6 ROIs (FG, OFA, STS, PHG, TOS and RSC) and some of their combinations, e.g., all face-related ROIs or all scene-related ROIs. All classification experiments used *binary classification* designed to distinguish between trials where subjects were cued to remember faces (*face trials*) and trials where subjects were cued to remember scenes (*scene trials*). The training features were voxel responses within an ROI extracted either from the cue or probe phase of the main task or from the separate localizer task. The test features