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

Xufeng Han[1], **Alexander C. Berg**[1], **Hwamee Oh**[3], **Dimitris Samaras**[1], and **Hoi-Chung Leung**[2]

[1]Department of Computer Science, SUNY at Stony Brook

[2]Department of Psychology, SUNY at Stony Brook

[3]Helen Wills Neuroscience Institute, University of California, Berkeley

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 to 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 to 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.

Keywords

visual working memory; inferior temporal cortex; fusiform gyrus; parahippocampal gyrus; face and scene; fMRI

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

Corresponding author: Hoi-Chung Leung, SUNY Stony Brook, Department of Psychology, Stony Brook, NY 11794-2500, Phone: (631) 632-7820, Fax: (631) 632-7876, hoi-chung.leung@stonybrook.edu.

cortex show functional divisions specializing in categorical representation of objects such as faces, tools, words, etc. (e.g., Epstein and Kanwisher, 1998; Chao et al., 1999). 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 & 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., Kanwisher et al., 1997; Epstein and Kanwisher, 1998). 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 (Oh and Leung, 2010; Lepsien et al., 2005). 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: Oh and Leung, 2010; Gazzaley et al. 2005). 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 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 (Figure 1A) was comprised of 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 Figure 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 (Figure 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 gap 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 degrees, 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. Figure 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 were extracted from the cue phase of the main task. For comparison purposes, we also applied SVM to examine brain activity during the probe phase. See below for details on feature definition and classifier training and testing.

Features—Our features are baseline-corrected voxel responses, specified by time and brain region. For the main task, we divided the cue phase into three time segments: early (7th and 8th scan), middle (9th and 10th scan) and late (11th and 12th scan), and similarly the probe phase into three segments: early (14th and 15th scan), middle (16th and 17th scan) and late (18th and 19th scan). (Scan numbers are counted from the beginning of a trial.) We normalized scans in every segment by subtracting the per voxel average of two baseline scans (1st and 2nd scan) and then averaged the two baseline-subtracted scans. We also constructed features using all segments in the cue or probe phase, in which case we concatenated all 6 baseline-corrected scans in the phase (e.g., Tables 1 and 2). For a particular time choice, voxels were selected from a particular ROI or a combination of ROIs in the baseline-corrected average scan of that time segment. For the localizer task, each face/scene block comprised 8 scans. The first two scans were averaged to form the baseline. The 5th and 6th scans were baseline corrected, and the averaged voxels of the two scans were selected for a region and used as features.

Classifier and cross-validation—The dataset for each subject consisted of 40 trials, 20 face trials and 20 scene trials. We evaluated linear classification accuracy using cross validation, leaving out one trial of each type. For each fold in cross validation, a classifier was trained on data from the remaining 38 trials and evaluated on the left out trial of each type. Each classification task specified the time segment for training and testing as well as the ROI(s) where features were extracted. For instance, one experiment–train with probe data and test with cue data for FG for the middle time segment–used measurements from the voxels in the left and right FG ROI extracted from the 16th and 17th scan, during the middle

segment of the probe phase (and normalized as described above) as training data, and measurements from the same voxels in the 9th and 10th scan, during the middle segment of the cue phase, for testing. Classifiers were trained using the LibLinear SVM package (http://www.csie.ntu.edu.tw/~cjlin/liblinear/) with L2-regularization, L2-loss function, and bias=1. The regularization vs. loss tradeoff parameter C was determined (from the set [0.001, 0.1, 1, 10, 1000]) for each cross-validation fold by using a subset of the training samples (37 trials) of each fold for nested cross validation. The accuracy of a classifier on a dataset is the mean across the accuracy on each fold.

The reported classification accuracy and standard error of the mean for an experiment are computed over the accuracies for the 12 datasets. Statistical significance of classification accuracy above chance (50%) was assessed using a 1-tail one-sample t-test. Bonferroni correction was applied when determining the significance of the p-values for classification tests across the three time segments of the cue/probe phase. Bonferroni correction was also applied when comparing classification performance across regions (e.g., the three face-related ROIs). In addition, we also conducted several permutation tests to evaluate the significance of classification results (following the procedures in Pereira & Botvinick, 2011). In general these tests resulted in confidence estimates similar to those from the t-test. Here we present only results from the t-tests.

RESULTS

Classification of activation patterns during probe recognition

To compare with previous findings, we first tested how accurately we could classify the probe category (face vs. scene) using activation patterns of the FG and ventral PHG from both hemispheres during the probe stage of the trials. Table 1 shows the average within-subject classification accuracies for classifiers trained and tested using fMRI data from the three time segments (early, middle and late) of the probe phase. Classification performance was well above chance (50%) for discriminating activation patterns during the middle time segment of the probe phase (>70% mean accuracy with p's<0.001; see Fig. 2B, yellow bars). These results were comparable to previous MVPA results for picture recognition (e.g., Morgan et al., 2011; Walther et al., 2009). It should also be noted that classification performance varied across the three time segments of the probe phase, with lower classification accuracy for the early and late time segments (though all above chance, *p's*<0.025 with Bonferroni correction, except for PHG during the late time segment). Similar results on classification of probe category were obtained by training classifiers with data from the localizer task (see Fig. 2B, blue bars). It should be mentioned that the differences in mean classification accuracies produced by training on localizer data versus probe data were not significant (p's>0.15), while some differences were expected as the localizer task was a completely separate task with a different design and smaller total number of scans.

Classification of activation patterns during selective maintenance

A primary goal of this study was to examine whether the FG and PHG carry information about which visual category is selectively maintained in working memory. Using the leave-one-out cross-validation method, classifiers were trained and tested using data from the three time segments of the cue phase. Classification performance was significantly above chance for distinguishing activation patterns during the middle time segment of the cue phase (>62% mean accuracy with p's<0.006 with Bonferroni correction), but was around chance for the early and late time segments (see Table 2A).

To further test whether the neural activity patterns were representing the cued visual category and to rule out the potential confound of word classification in the above test

results, additional classifiers were trained with data from the localizer task and the probe phase where no text cue was presented; these classifiers were tested on data from the cue phase. Classifiers trained with the probe data showed well above chance performance on differentiating the activation patterns in both FG and PHG by visual category during the middle time segments of the cue phase (>62% mean accuracy with $p<0.001$ and $p<0.0001$, respectively; see Table 2B). Classifiers trained with the localizer data showed slightly lower but still above chance performance (>58% mean accuracy with p's<0.01). See Figure 2A for a summary of the classification results using different training and testing combinations.

We further tested whether selective maintenance might be restricted to the left or right hemisphere of the FG and PHG (see Figure 3A). Since the PHG activation cluster was big, we also examined other subdivisions of the PHG (a more dorsal and a more posterior part) in comparison to the ventral PHG data described above (Arcaro et al., 2009; Sewards, 2010). For classifiers trained with probe data and tested on cue data (middle time segment), classification performance was significantly above chance for both the right and the left FG (both ~61%, p's<0.001). Classification performance for the dorsal-lateral and posterior-lateral portions of the PHG was comparable to the ventral-medial portion (reported above) (all p's<0.007, except for the right posterior part, p>0.17). Not surprisingly, classification of the probe data by visual category was well above chance for all subdivisions of FG and PHG (all p's <0.001; see Fig. 3B). Taken together, there were no clear hemispheric differences in classification results for either FG or PHG.

In sum, while previous univariate analysis did not yield significant differential activity in the FG for selective maintenance of faces, the MVPA analysis was able to use the spatial activation patterns from the middle time segment of the post-cue period in most individuals to differentiate the task relevant visual category.

Classification of activation patterns in other visual association regions

Besides the FG and PHG, we examined several other temporal and occipital regions that have been previously implicated in face and scene processing (see Methods). In terms of the average activity amplitude, only scene-related regions but not face-related regions showed differential effects by the cued visual category during selective maintenance (see Fig. 1B). Using the same approach described above, we conducted visual-category classification tests to examine the spatial response patterns of these visual cortical regions during the middle time segment of the probe and cue phases. Classifiers were trained with data from the probe phase. Figure 4 shows the classification results. Besides plotting the overall classification accuracy for each region as in the other figures, we plotted the classification accuracies by trial type (face/scene cue) in this figure.

Among face-related regions, classification of FG activity for selective maintenance of face/scene visual category was on average more accurate than OFA and STS (e.g., 63%, 56%, and 55%, respectively, as shown by the blue bars in Fig. 4A). The mean classification accuracy was significantly above chance for FG (p<0.003) but not for OFA and STS (p's>0.08), after correction for multiple comparison. The overall classification accuracy for probe category, however, was comparable between FG and OFA (both about 70%; Fig. 4D) and lower for STS (61%), though all well above chance (p's<0.002, corrected for multiple comparison). The overall classification results for the scene-related regions were more comparable, with classification of post-cue activation pattern slightly more accurate for ventral PHG than TOS and RSC (64%, 60%, 60%, respectively; blue bars in Fig. 4A; all above chance, p's<0.02 after correction for multiple comparison). It is worth mentioning that relative to FG or PHG alone, combining data from the three face-related regions or the three scene-related regions did not change classification accuracy by much (<0.5% difference in most cases). Intriguingly, for face-related regions such as the FG and OFA,

better classification accuracy was achieved for their post-cue response patterns on face trials (65% or higher) while classification was around chance level for their response patterns on scene trials (~50%); and such differences in classification were significant (pair t-tests: p's<0.04). In other words, most face trials were correctly classified as face trials while less than half of the scene trials were correctly classified as scene trials in these face-related ROIs. The opposite was not observed with the scene-related regions (p's>0.1), suggesting different ways of representing visual category information in working memory in the PHG versus FG.

Lastly, we conducted additional classification tests using activation patterns from brain areas that showed similar level of activity on face and scene trials. The mask included primarily frontal and parietal regions, defined using the conjunction map from our previous analysis (Oh and Leung, 2010). The classification accuracy for activation patterns of these regions during the cue phase, albeit significantly above chance (p's < 0.05), was all lower than 60% (between 55 to 60%, with classifiers trained by either cue or probe phase data; data not shown).

DISCUSSION

A primary feature of visual working memory is the flexible representation of visual information most relevant to the current task demand. We utilized MVPA to examine whether the spatial response patterns of the individually defined ventral temporal and occipital regions reflect the selected visual category in working memory. Although previously we only found differential level of activity in the PHG for selective representation of scene images and no significant effects for the FG, present results from MVPA revealed differentiable spatial response patterns in both regions in correspondence to the cued visual category during post-cue maintenance and probe recognition. Classification of activation patterns for selective maintenance of face/scene category was well above chance even when classifiers were trained with data samples from the probe phase and the separate localizer task. For some other temporal and occipital regions that also showed differential response patterns to face versus scene probes, the classification of these regions' activity patterns during selective maintenance was less consistent (i.e., not always significantly above chance). These findings suggest that the ventral occipital and temporal regions, especially the FG and PHG carry visual category representations relevant to the task, and their activity patterns during working memory maintenance may resemble those during probe recognition.

Findings from this study support the previous assertion that the ventral temporal and occipital areas contribute to keeping visual information in working memory. In particular, the present results from MVPA address the inconsistent results from previous univariate analysis on the involvement of FG in maintaining face working memory (Druzgal and D'Esposito, 2003; Jha and McCarthy, 2000; Oh and Leung, 2010). By examining the spatial patterns of brain activation, we found that not only the PHG but also the FG exhibits differential activity patterns in correspondence to the cued visual category during the post-cue delay period. This was not likely caused by pure verbalization of the cue stimulus as classification was well above chance even when classifiers were trained with fMRI data from the probe phase and the completely separate localizer task. In fact, classification results on post-cue scans were worse when training with post-cue scans than when training with probe scans. This could be due to neural activity during the post-cue period being typically lower in amplitude and more variable across trials, whereas neural activity in response to the probe is usually stronger and more consistent across trials. The reliability of a classifier is dependent on the signal-to-noise ratio of the training data.

A recent study manipulated three categories of visual stimuli in a working memory task (line segments, pseudowords and words), and found that brain activation patterns reflect the category currently relevant to the task demand but not the irrelevant category (Experiment 2 data from Lewis-Peacock et al., 2012; Lewis-Peacock & Postle, 2012). Our investigation extended such findings to specific visual association regions known to be involved in categorical processing. While our analysis cannot completely rule out whether or not the no-longer-relevant visual category was still represented in the visual association regions, examining the classification results for the two types of cue trials separately allowed us to gain some insight on how a region behaves on trials where the cued category matched its presumed categorical preference compared to trials of the non-preferred category (see Fig. 4). On the one hand, the classification accuracy of activation patterns of the FG reached 70% for classifying the face trials as face trials but was at chance for classifying the scene trials as scene trials. Differences in classification of FG activation patterns on scene trials across subjects ranged from 20–85% accuracy, which suggest that some subjects might still maintain a face representation in the FG on scene trials while others might simply show some sort of "non-face" response patterns (or degradation in face representation) on scene trials. On the other hand, classification of activation patterns of the PHG on face trials and scene trials was similarly above chance. These findings together with our previous finding of significantly reduced average activity in the PHG on face trials relative to scene trials suggest that the PHG's activity pattern might have changed on face trials (e.g., degraded) leading to the correct classification of its activity on face trials as "non-scene" trials. Taken together, our results support the notion that the FG and PHG are specialized in representing the task-relevant visual category in working memory (Ranganath et al., 2004). Importantly, our results further suggest that the FG and PHG probably use fundamentally different mechanisms in representing visual information and show different levels of degradation in representing its preferred visual category depending on the task goal. Perhaps reduction of face representation primarily occurs on situations demanding greater attention to a different category as in some other studies (Lewis-Peacock et al., 2012; Seidl et al., 2012). However, these previous studies did not examine the individual face- and scene-related regions so it is unclear whether or not there are differences in visual representation across different visual association regions. It is also possible that the SVM could be further optimized to distinguish multiple representations held in working memory. Further studies with different task designs are required to resolve this issue.

The present findings also suggest that the activation patterns of the posterior association regions during selective maintenance in the absence of the physical stimuli are similar to the patterns during probe recognition in the presence of the physical stimuli. Our findings are in line with an fMRI study by Lewis-Peacock and Postle (2008). They trained their participants to remember face-scene, face-object, object-scene pairs, and examined brain activation patterns in correspondence to the retrospective and prospective representation of recently presented visual images and the associated images, respectively. Their findings suggested that visual working memory is manifested as a reactivation of visual representations from long term memory. The posterior association regions indeed showed differential activation patterns during memory search of famous people, famous places and common objects from past experiences (Polyn et al., 2005). In our task, however, the face and scene target images were trial unique and novel or unfamiliar to the subjects, so the working memory representation during the post-cue delay period likely resembles the perceptual representation during the probe recognition stage, at least in terms of visual category. Thus, our findings in general support the hypothesis that the posterior association regions carry similar representations for visual perception, short-term, and long-term memory (Courtney & Ungerleider, 1997; Postle, 2006). The working memory representation of visual stimuli nonetheless may not be identical to the perceptual representation of visual stimuli, as the classification of activation patterns during probe recognition was 10% higher than the

classification of activation patterns during post-cue maintenance. It could be that the presence of a physical stimulus during the probe phase made a difference in category-specific pattern activity (but see above for alternative explanations). Future studies should investigate the cause of this degradation in classification performance for working memory maintenance as opposed to object recognition.

SUMMARY

We applied MVPA to show that the spatial response patterns of specific posterior visual association regions including the fusiform gyrus and the parahippocampal gyrus carry information reflecting the task-relevant visual category in working memory. These findings of selective visual representation revealed by the multi-voxel activation patterns in the fusiform gyrus complimented our previous analysis which showed no differences in this region's average activity across the task conditions (Oh and Leung, 2010). In addition, our findings suggest that the activation patterns in these higher-order visual cortical regions associated with visual working memory representations are similar to the patterns associated with visual perception and recognition of different visual categories. It is possible that the visual representation maintained in working memory is a reinstatement of recent perceptual experience (see Postle, 2006; Magnussen, 2000).

Acknowledgments

References

Arcaro MJ, McMains SA, Singer BD, Kastner S. Retinotopic organization of human ventral visual cortex. Journal of Neuroscience. 2009; 29:10638–10652. [PubMed: 19710316]

Chao LL, Haxby JV, Martin A. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. Nature Neuroscience. 1999; 2:913–919.

Courtney SM, Ungerleider LG. What fMRI has taught us about human vision. Current Opinion in Neurobiology. 1997; 7:554–561. [PubMed: 9287197]

Druzgal TJ, D'Esposito M. Dissecting contributions of prefrontal cortex and fusiform face area to face working memory. Journal of Cognitive Neuroscience. 2003; 15:771–784. [PubMed: 14511531]

Epstein R, Kanwisher N. A cortical representation of the local visual environment. Nature. 1998; 392:598–601. [PubMed: 9560155]

Ester EF, Serences JT, Awh E. Spatially global representations in human primary visual cortex during working memory maintenance. Journal of Neuroscience. 2009; 29:15258–15265. [PubMed: 19955378]

Fox CJ, Iaria G, Barton JJS. Defining the face processing network: optimization of the functional localizer in fMRI. Human Brain Mapping. 2009; 30:1637–1651. [PubMed: 18661501]

Friston K, Ashburner J, Frith C, Poline J, Heather J, Frackowiak R. Spatial registration and normalization of images. Human Brain Mapping. 1995; 3:165–189.

Gazzaley A, Cooney JW, McEvoy K, Knight RT, D'Esposito M. Top-down enhancement and suppression of the magnitude and speed of neural activity. Journal of Cognitive Neuroscience. 2005; 17:507–517. [PubMed: 15814009]

Harrison SA, Tong F. Decoding reveals the contents of visual working memory in early visual areas. Nature. 2009; 458:632–635. [PubMed: 19225460]

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science. 2001; 293:2425–2430. [PubMed: 11577229]

Haynes J, Rees G. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nature Neuroscience. 2005; 8:686–691.

Jha AP, McCarthy G. The influence of memory load upon delay-interval activity in a working-memory task: an event-related functional MRI study. Journal of Cognitive Neuroscience. 2000; 12(Suppl 2):90–105. [PubMed: 11506650]

Johnson MR, Mitchell KJ, Raye CL, D'Esposito M, Johnson MK. A brief thought can modulate activity in extrastriate visual areas: Top-down effects of refreshing just-seen visual stimuli. NeuroImage. 2007; 37:290–299. [PubMed: 17574442]

Kamitani Y, Tong F. Decoding the visual and subjective contents of the human brain. Nature Neuroscience. 2005; 8 (5):679–685.

Kanwisher N, McDermott J, Chun M. The fusiform face area: A module in human extrastriate cortex specialized for face perception. Journal of Neuroscience. 1997; 17:4302–4311. [PubMed: 9151747]

Lepsien J, Griffin IC, Devlin JT, Nobre AC. Directing spatial attention in mental representations: Interactions between attentional orienting and working-memory load. NeuroImage. 2005; 26:733–743. [PubMed: 15955482]

Lepsien J, Thornton I, Nobre AC. Modulation of working-memory maintenance by directed attention. Neuropsychologia. 2011; 49:1569–1577. [PubMed: 21420421]

Lewis-Peacock JA, Drysdale AT, Oberauer K, Postle BR. Neural Evidence for a Distinction between Short-term Memory and the Focus of Attention. Journal of Cognitive Neuroscience. 2012; 24:61–79.

Lewis-Peacock JA, Postle BR. Temporary activation of long-term memory supports working memory. Journal of Neuroscience. 2008; 28:8765–8771. [PubMed: 18753378]

Lewis-Peacock JA, Postle BR. Decoding the internal focus of attention. Neuropsychologia. 2012; 50(4):470–8. [PubMed: 22108440]

Magnussen S. Low-level memory processes in vision. Trends in Neurosciences. 2000; 23:247–251. [PubMed: 10838593]

Miller EK, Li L, Desimone R. Activity of neurons in anterior inferior temporal cortex during a short-term memory task. Journal of Neuroscience. 1993; 13:1460–1478. [PubMed: 8463829]

Morgan LK, Macevoy SP, Aguirre GK, Epstein RA. Distances between real-world locations are represented in the human hippocampus. Journal of Neuroscience. 2011; 31:1238–1245. [PubMed: 21273408]

Nasr S, Liu N, Devaney KJ, Yue X, Rajimehr R, Ungerleider LG, Tootell RBH. Scene-Selective Cortical Regions in Human and Nonhuman Primates. Journal of Neuroscience. 2011; 31:13771–13785. [PubMed: 21957240]

Nobre AC, Coull JT, Maquet P, Frith CD, Vandenberghe R, Mesulam MM. Orienting Attention to Locations in Perceptual Versus Mental Representations. Journal of Cognitive Neuroscience. 2004; 16:363–373. [PubMed: 15072672]

Oh H, Leung HC. Specific and nonspecific neural activity during selective processing of visual representations in working memory. Journal of Cognitive Neuroscience. 2010; 22:292–306. [PubMed: 19400681]

Pereira F, Botvinick M. Information mapping with pattern classifiers: A comparative study. NeuroImage. 2011; 56(2):476–496. [PubMed: 20488249]

Polyn SM, Natu VS, Cohen JD, Norman KA. Category-specific cortical activity precedes retrieval during memory search. Science. 2005; 310:1963–1966. [PubMed: 16373577]

Postle BR. Working memory as an emergent property of the mind and brain. Neuroscience. 2006; 139:23–38. [PubMed: 16324795]

Postle BR, Druzgal TJ, D'Esposito M. Seeking the neural substrates of visual working memory storage. Cortex. 2003; 39:927–946. [PubMed: 14584560]

Ranganath C, Cohen MX, Dam C, D'Esposito M. Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory retrieval. Journal of Neuroscience. 2004; 24:3917–3925. [PubMed: 15102907]

Ranganath C, D'Esposito M. Directing the minds eye: prefrontal, inferior and medial temporal mechanisms for visual working memory. Current Opinion in Neurobiology. 2005; 15:175–182. [PubMed: 15831399]

Sapountzis P, Schluppeck D, Bowtell R, Peirce JW. A comparison of fMRI adaptation and multivariate pattern classi_cation analysis in visual cortex. NeuroImage. 2010; 49:1632–1640. [PubMed: 19815081]

Seidl KN, Peelen MV, Kastner S. Neural evidence for distracter suppression during visual search in real-world scenes. J Neurosci. 2012; 32(34):11812–9. [PubMed: 22915122]

Sewards TV. Neural structures and mechanisms involved in scene recognition: a review and interpretation. Neuropsychologia. 2011; 49:277–298. [PubMed: 21095199]

Sreenivasan KK, Katz J, Jha AP. Temporal characteristics of top-down modulations during working memory maintenance: an event-related potential study of the N170 component. Journal of Cognitive Neuroscience. 2007; 19:1836–1844. [PubMed: 17958486]

Ungerleider LG, Haxby JV. "What" and "where" in the human brain. Current Opinion in Neurobiology. 1994; 4:157–165. [PubMed: 8038571]

Walther DB, Caddigan E, Fei-Fei L, Beck DM. Natural scene categories revealed in distributed patterns of activity in the human brain. Journal of Neuroscience. 2009; 29:10573–10581. [PubMed: 19710310]

Yoon JH, Curtis CE, D'Esposito M. Differential effects of distraction during working memory on delay-period activity in the prefrontal cortex and the visual association cortex. NeuroImage. 2006; 29:1117–1126. [PubMed: 16226895]

Highlights

We studied spatial response patterns of visual cortical areas during working memory.

Multivoxel pattern analysis reveals selective maintenance of face or scene pictures.

Activity of fusiform and parahippocampal areas reflects the task relevant category.

Activation patterns during selective maintenance and object recognition are similar.
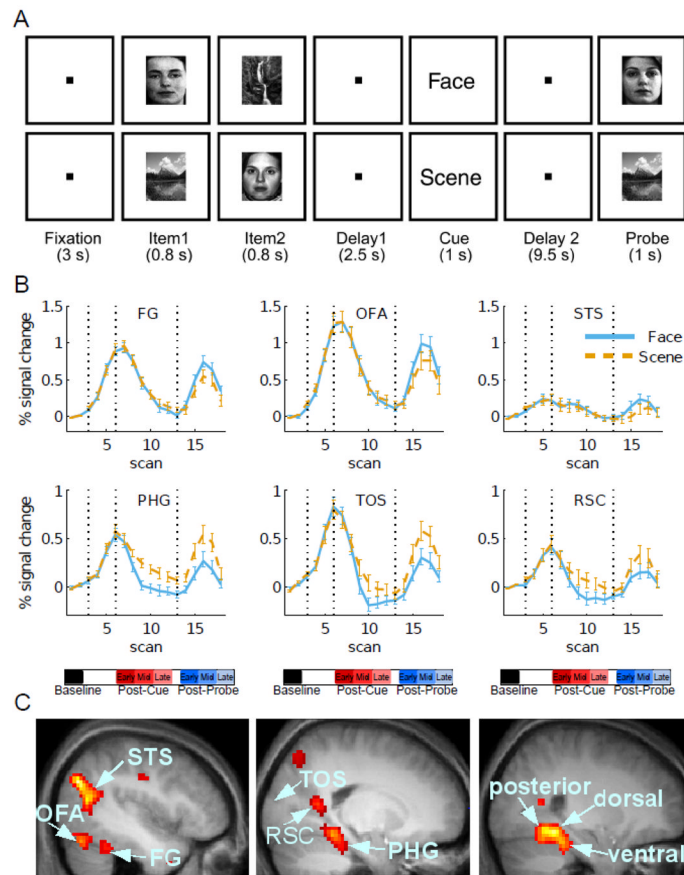
Figure 1.
(A) Visual working memory task with a selection cue. A face and a scene picture were
presented (each for 800 ms counterbalanced in order) at the beginning of each trial for
encoding. A word cue ("Face"/ "Scene") was then presented for 1000 ms to indicate the
relevant picture category for selective maintenance. After a 9.5-s delay, a probe stimulus
was presented for 1000 ms for the participants to make a match/non-match recognition
judgment. (B) Average percent signal change in fMRI signal across time in regions
associated with face and scene processing during the working memory task. The vertical
dotted lines mark the onset of stimulus presentation, cue, and probe. The scales in the
bottom show the division of the cue and probe phases into early, middle and late time
segments, which were used in training and testing SVM classifiers. (C) Functionally defined
regions of interest. The rightmost image shows the three subdivisions of the PHG for finer
scale MVPA. In the middle image, one of the arrows shows the approximate location of
TOS while activation was not found in the selected slice but in nearby slices. Abbreviations:
FG, fusiform gyrus; OFA, occipitotemporal cortex or occipital face area; STS, superior
temporal sulcus; PHG, parahippocampal gyrus; TOS, transverse occipital sulcus; RSC,
retrosplenial cortex. Legend: blue line–trials with the face cue, orange dotted line–trials with
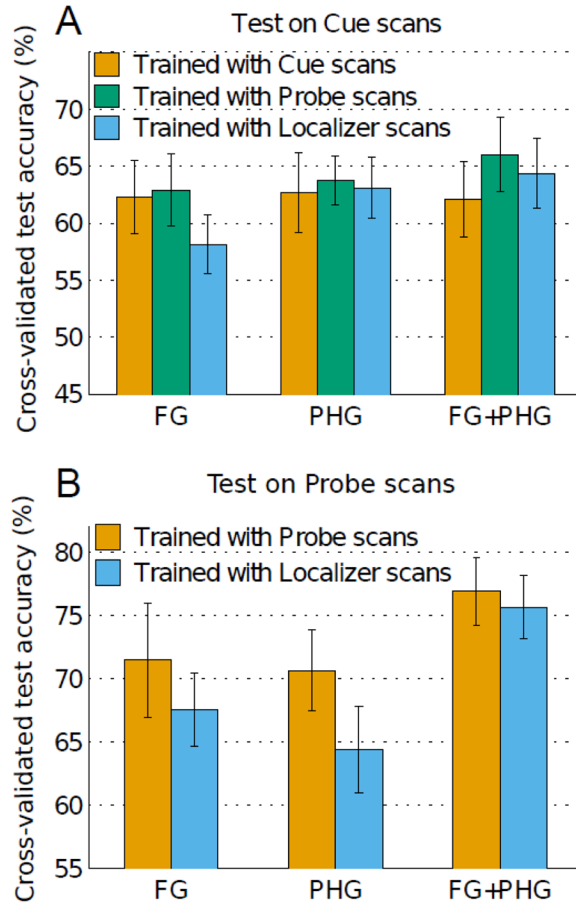the scene cue.

Figure 2.
Mean accuracy in classification of face/scene during selective maintenance and probe recognition using voxels in the FG and ventral PHG. (A) SVM classifiers were trained with scans from the middle time segment of the cue phase (orange), probe phase (green) and localizer blocks (blue). Classifiers were tested on scans from the middle time segment of the cue phase. (B) Classifiers were trained with scans from the middle time segment of the probe phase (orange) and localizer blocks (blue), and tested on scans from the middle time segment of the probe phase.
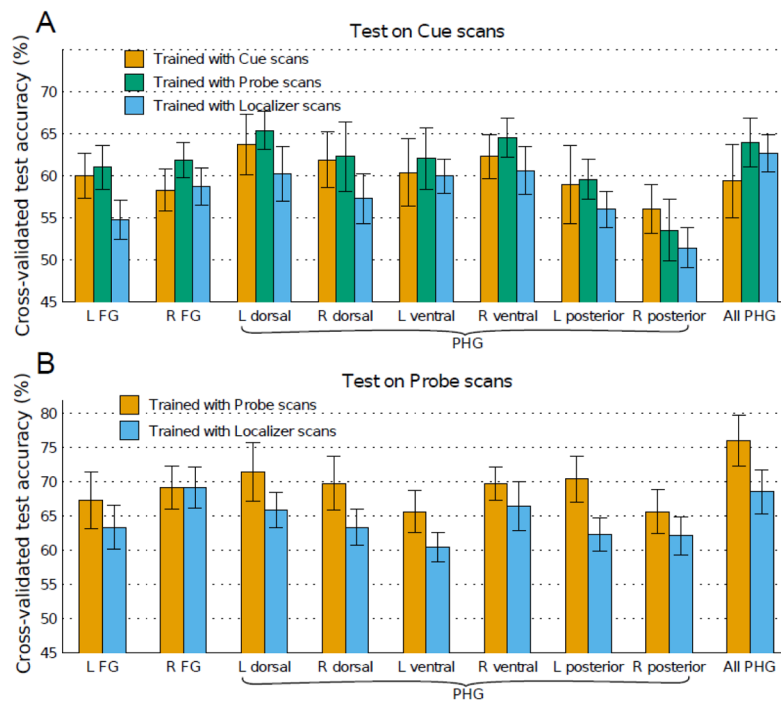
Figure 3.
Mean accuracy in classification of face/scene during selective maintenance (A) and probe recognition (B) using voxels in the different parts (dorsal, ventral and posterior) of the PHG in the left and right hemisphere. The left and right FG are shown for comparison. Abbreviations: L- left, R- right.
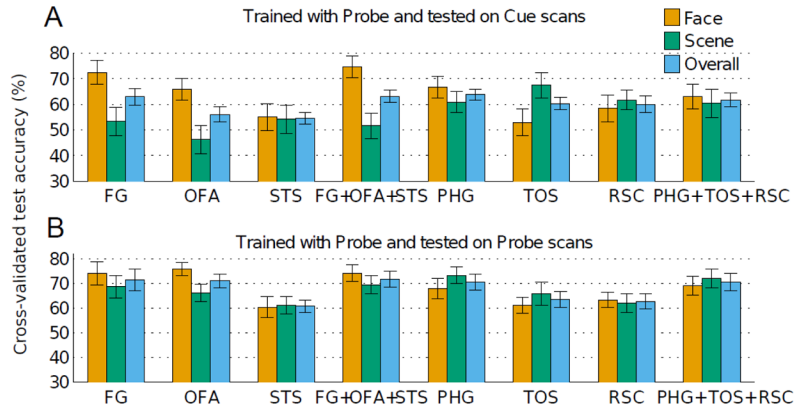
Figure 4.
Mean accuracy in classification of face/scene during selective maintenance (A) and probe recognition (B) using voxels in ventral temporal and occipital regions associated with face and scene processing. Three bars are shown for each region, indicating classification performance for the face trials (orange; i.e., number of correctly classified face trials out of total number of face trials), the scene trials (green; i.e., number of correctly classified scene trials out of total number of scene trials) and the overall accuracy (average of the two, blue).

Table 1

Average accuracy in classification of FG and PHG activation patterns across time (early, middle and late, and their combination [All]) for face and scene probe recognition. Classifiers were trained and tested with data from the probe phase, with features combined from both hemispheres for each brain region. Note training and testing were done only on the corresponding time segments and never across time segments. Standard errors of mean are shown in parentheses. All values are significantly above chance (50%) except for the late time segment of vPHG. FG, fusiform gyrus; vPHG, ventral parahippocampal gyrus.

	All	**Early**	**Middle**	**Late**
FG	71.88% (4.21%)	64.17% (3.32%)	71.46% (4.51%)	58.75% (2.37%)
vPHG	65.21% (3.26%)	59.38% (3.29%)	70.63% (3.23%)	50.24% (3.56%)
FG+vPHG	78.12% (2.52%)	70.00% (2.04%)	76.88% (2.70%)	56.67% (2.23%)

Table 2

Average accuracy in classification of FG and PHG activation patterns across time for selective maintenance of face and scene working memory. Features were combined from both hemispheres for each brain region. Note training and testing were done only on the corresponding time segments and never across time segments. Standard errors of mean are shown in parentheses. Values significantly above chance (50%) are shown in bold.

(A) Classifiers were trained and tested with data from the cue phase.

	All	Early	Middle	Late
FG	55.21% (3.58%)	49.17% (2.91%)	**62.29% (3.25%)**	52.50% (3.40%)
vPHG	**60.21% (3.61%)**	52.29% (2.81%)	**62.71% (3.50%)**	53.96% (4.43%)
FG+vPHG	**60.83% (3.61%)**	50.62% (2.84%)	**62.08% (3.34%)**	56.46% (3.70%)

(B) Classifiers were trained with data from the probe phase and tested on data from the cue phase.

	All	Early	Middle	Late
FG	**59.58% (2.34%)**	49.58% (2.42%)	**62.92% (3.18%)**	**57.71% (1.86%)**
vPHG	**57.29% (2.27%)**	53.54% (2.86%)	**63.75% (2.14%)**	53.54% (2.47%)
FG+vPHG	**61.46% (2.27%)**	53.12% (2.65%)	**66.04% (3.22%)**	54.17% (2.35%)