Eye Movements Help Link Different Views in Scene-Selective Cortex

Julie D. Golomb[1], Alice R. Albrecht[2], Soojin Park[1] and Marvin M. Chun[2]

[1]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA and [2]Department of Psychology, Yale University, New Haven, CT 06520, USA

Address correspondence to Julie D. Golomb, McGovern Institute for Brain Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room 46-4141; Cambridge, MA 02139, USA. Email: jgolomb@mit.edu

To explore visual scenes in the everyday world, we constantly move our eyes, yet most neural studies of scene processing are conducted with the eyes held fixated. Such prior work in humans suggests that the parahippocampal place area (PPA) represents scenes in a highly specific manner that can differentiate between different but overlapping views of a panoramic scene. Using functional magnetic resonance imaging (fMRI) adaptation to measure sensitivity to change, we asked how this specificity is affected when active eye movements across a stable scene generate retinotopically different views. The PPA adapted to successive views when subjects made a series of saccades across a stationary spatiotopic scene but not when the eyes remained fixed and a scene translated in the background, suggesting that active vision may provide important cues for the PPA to integrate different views over time as the "same." Adaptation was also robust when retinotopic information was preserved across views when the scene moved in tandem with the eyes. These data suggest that retinotopic physical similarity is fundamental, but the visual system may also utilize oculomotor cues and/or global spatiotopic information to generate more ecologically relevant representations of scenes across different views.

Keywords: parahippocampal place area, retinotopic, scene perception, spatiotopic, viewpoint specificity

Introduction

In order to construct stable images of the world, we must integrate visual information over time. Although researchers debate whether visual information is literally integrated over time in the mind (Irwin 1991; O'Regan 1992; Henderson 1997; Cavanagh et al. 2010), no one doubts the challenge posed by eye movements, namely the rapid changes in visual input received by the retina and transmitted to the brain. For instance, our perception of a large visual scene is comprised of a series of snapshot views; when we move our eyes to actively explore the scene, even small eye movements can make the local foveated elements change completely, yet from these discrete successive views, we perceive a smoothly continuous visual experience.

How the brain recognizes a scene as the "same" has been an oft-studied question. Research into general scene processing using functional magnetic resonance imaging (fMRI) has revealed several scene-selective regions of cortex, most notably the parahippocampal place area (PPA), a region in the medial temporal cortex that robustly responds to scene stimuli (Epstein and Kanwisher 1998). The PPA not only shows an increased response to scenes in general (when compared with other visual stimuli such as faces or objects) but also it can differentiate between individual scenes: The PPA responds very strongly to the presentation of a novel scene but when that same scene is repeated, the response is reduced, a phenomenon known as adaptation or repetition suppression (Schacter and Buckner 1998; Wiggs and Martin 1998). A number of studies have made use of the repetition suppression technique to explore whether a given brain region treats 2 stimuli as the same (Grill-Spector and Malach 2001), measuring PPA adaptation as a proxy for understanding how different aspects of scenes are represented in the brain (Epstein et al. 2003; Yi and Chun 2005; MacEvoy and Epstein 2007; Park et al. 2007; Park and Chun 2009). The PPA represents scenes in a highly specific manner, adapting to identical views of the same scene but not to the same scene from a different viewpoint (Epstein et al. 2003).

Panoramic scenes provide another interesting test of viewpoint specificity. How does the PPA process different overlapping views panning across the same "scene"? Park and Chun (2009) presented subjects with sequences of 3 scenes, either all identical views of the same scene or 3 different but partly overlapping views taken from a single larger panoramic scene. The overlapping views in the panoramic condition were presented consecutively to mimic the everyday experience of moving our eyes across a scene to explore the world, and each view overlapped with the previous view in 66% of its physical details and layout. Despite such strong overlap in physical contents and continuity cues, the PPA did not adapt to these overlapping panoramic views, showing a highly viewpoint-specific representation (Park and Chun 2009). However, a number of studies have shown that visual stability across eye movements depends on efferent feedback or on corollary discharge from the eye movement itself (Stevens et al. 1976; Sommer and Wurtz 2006; Wurtz 2008). For example, during active vision, corollary discharge signaling an impending eye movement can trigger the visual system to suppress the motion transient caused by the eye movement (saccadic suppression: Bridgeman et al. 1975) and to prospectively update the positions of objects in the environment to maintain stability (spatiotopic remapping: Duhamel et al. 1992). When oculomotor cues are disrupted or absent, the percept of visual stability is likewise impaired (Stevens et al. 1976). This leaves open the possibility that the PPA's ability to identify views as similar—originating from the same panoramic scene—might similarly be influenced by oculomotor cues.

To our knowledge, no one has tested how eye movement information is integrated into the PPA representation of a scene, so the current study uses strategically placed fixation locations and eye tracking to explicitly guide eye position across different views. Neural responses in the PPA are compared for overlapping views created by movements of the background scene (with eye position maintained in a consistent location, as in Park and Chun 2009), versus eye movements across a stationary scene. If oculomotor cues are important in establishing a stable representation of a broader scene, then we might expect to see PPA adaptation to overlapping views of

a panoramic image created by eye movements but not by the analogous scene movements. On the other hand, if the PPA is strictly viewpoint specific, there should be no adaptation in either condition.

In addition to exploring how overlapping views of a scene are represented across eye movements, the current study also uses eye movements to investigate which coordinate systems the PPA uses to establish whether successive images are part of the same scene. In the Park and Chun (2009) study, because eye position was never varied, the retinotopic (eye centered) and spatiotopic (world or head centered) images were always bound to each other. Thus, the lack of adaptation in the overlapping view condition could have been due to differences in local retinotopic input with each view or the fact that the global spatiotopic image on the computer screen was changing. To clarify the role of each of these coordinate systems, the current study utilizes a design that manipulates both the physical properties of the scene (identical or partly overlapping "scrolling" views) and eye position (stationary fixating or sequence of saccades).

In the Fixation-Identical condition, neither the scene nor the eyes move, and thus both retinotopic and spatiotopic inputs remain the same across the trial. In the Fixation-Scrolling condition, the scene "scrolls" in the background, resulting in both retinotopic and spatiotopic differences across the 3 views. Based on the Park and Chun (2009) study, we should find adaptation in the PPA for the Fixation-Identical but not the Fixation-Scrolling condition. To explore whether this difference is driven by retinotopic or spatiotopic "sameness," the critical test is how the PPA adapts during eye movement trials. In the Saccade-Identical condition, the eyes move across a stationary scene, generating different foveal retinotopic input but the same global spatiotopic input across views. In the Saccade-Scrolling condition, the same eye movements are executed, but now with the scene moving in tandem in the background, such that the eyes are always fixated on the same local point in the scene, creating retinotopically (foveally) identical but spatiotopically different input. If the PPA exhibits strict local viewpoint specificity based solely on retinotopic input, adaptation should be found for the Saccade-Scrolling condition but not the Saccade-Identical condition. If, however, the PPA is more sensitive to the ecologically relevant spatiotopic information, the reverse should be true. Finally, it is possible that both spatiotopic and retinotopic information are represented in the PPA, and thus we would expect to see adaptation to both conditions or to neither if the conjunction is critical. To evaluate the amount of adaptation across conditions, PPA responses for each of the above conditions were compared with baseline no-adaptation conditions (Fixation-Novel or Saccade-Novel), which consisted of sequences of 3 completely different novel scenes presented during both fixation and saccade conditions.

Materials and Methods

Subjects

Twelve subjects (8 female; mean age 22.4, range 19–30) participated in Experiment 1, and 10 subjects (5 female; mean age 24.2, range 21–36) participated in Experiment 2. All subjects were neurologically intact with normal or corrected-to-normal vision. Informed consent was obtained for all subjects, and the study protocols were approved by the Human Investigation Committee of the Yale School of Medicine and the Human Subjects Committee of the Faculty of Arts and Sciences at Yale University.

Experimental Setup

Stimuli were generated using the Psychtoolbox extension (Brainard 1997) for Matlab (The Mathworks, Inc.). During fMRI scanning, stimuli were displayed with an LCD projector onto a screen mounted in the rear of the scanner bore, which subjects viewed from a distance of 79 cm via a mirror attached to the head coil (maximal field of view: 23.5°). Eye position was monitored using a modified ISCAN eye-tracking system (ISCAN, Inc.), in which the camera and infrared source were attached to the head coil above the mirror; pupil and corneal reflection (CR) were recorded at 60 Hz, and gaze angle (pupil–CR) was computed online to confirm accurate fixation.

Experimental Design

Each 6-s trial was composed of 3 scenes presented sequentially (Fig. 1). The scenes were sized 24° × 18° and centered on the screen. There were 3 possible fixation positions, located at horizontal positions of –8°, 0°, and 8° from the center of the scene (–6°, 0°, and 6° for Experiment 2), all centered vertically. The trial began with a fixation dot in either the leftmost or rightmost fixation position. Trial progression was as follows: Fixation 1 (1000 ms), Scene 1 (500 ms), Fixation 1 (500 ms), Fixation 2 (1000 ms), Scene 2 (500 ms), Fixation 2 (500 ms), Fixation 3 (1000 ms), Scene 3 (500 ms), Fixation 3 (500 ms). On half of the trials, the fixation dot remained in the starting location for all 3 events. On the other half the fixation, dot moved to the middle position on Fixation 2 and to the opposite side on Fixation 3; subjects were instructed to move their eyes to the new location as quickly as possible and always stay fixated on the current fixation dot, while passively viewing the scenes. Fixation was monitored to ensure subjects were correctly performing the task.

Within both fixation and saccade trials, there were 3 types of scene conditions. In the "Identical" condition, the same exact scene was presented for each of the 3 events. In the "Scrolling" condition, 3 successive overlapping views taken from a larger panoramic scene were presented. Each view overlapped 66% with the previous view. Finally, in the "Novel" condition, 3 completely different scenes were presented in succession. Each of the 6 conditions was presented for 8 trials per run, with 0, 2, or 4 s between trials. Trials were presented in a pseudorandom order generated separately for each run to minimize serial correlations between conditions, with the additional constraint that each trial's first fixation location had to be in the same position as the last fixation location of the previous trial to reduce unnecessary eye movements.

Stimuli

The image set for Experiment 1 consisted of 320 color photographs of natural indoor and outdoor scenes. Each image was originally sized at 40° × 18°, then cropped into 3 views overlapping by 66% each (Fig. 1*C*). The final cropped stimulus size was 24° × 18°. Stimuli were randomly assigned to one of the 6 conditions, separately for each subject, and stimuli were never repeated across trials.

The image set for Experiment 2 consisted of 3D virtual environment scenes created with Complete Home Designer 5 (Data Becker, 2001). The overlapping views used in Experiment 1 were taken from single panoramic 2D scenes and thus reflected viewpoint translation but not rotation and motion distortion effects experienced in the real world. The 3D virtual environment scenes allowed viewpoint changes including distortions of motion parallax effects as the retinal positions of near and far surfaces move relative to each other, creating more ecologically valid viewpoint changes. Eighty unique virtual indoor and outdoor scenes were created. For 32 of these scenes, 3 different snapshots were created simulating overlapping views for the scrolling conditions (Fig. 1*D*). Because these virtual overlapping views were not simple image translations, the exact percent overlap cannot be precisely determined, but each successive view overlapped by roughly 75% with the previous view. The 3D overlapping view stimuli were randomly distributed among the Fixation-Scrolling and Saccade-Scrolling conditions for Experiment 2. For the remaining forty-eight 3D scenes, only one viewpoint was simulated. These single view stimuli were randomly distributed among the 4 remaining conditions. Stimulus size for all images was 24° × 18°, and stimuli were never repeated across trials.
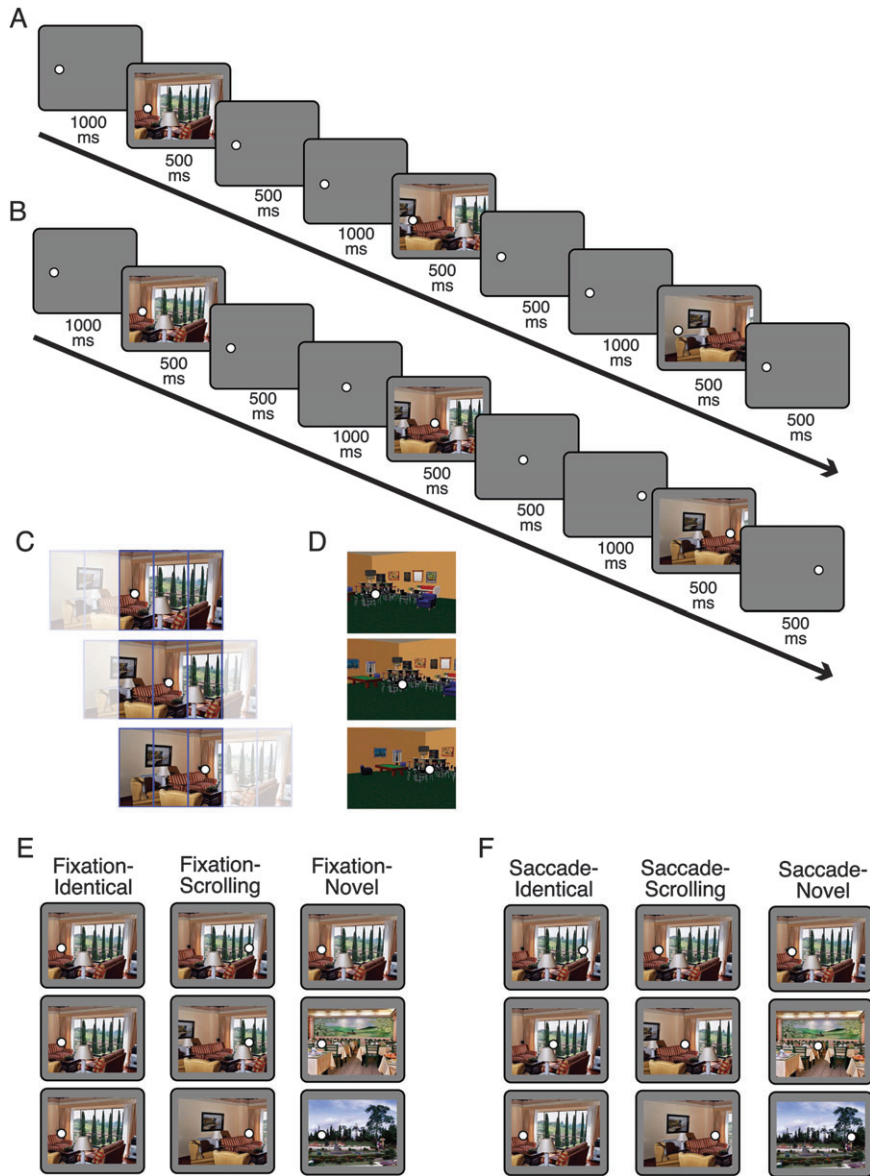
Figure 1. Task design and stimuli. (*A*, *B*) Trial timing. Each trial consisted of the following sequence of events: Fixation 1, Scene 1, Fixation 1, Fixation 2, Scene 2, Fixation 2, Fixation 3, Scene 3, Fixation 3. (*A*) On Fixation trials, the fixation dot remained in the same location (left or right) for the duration of the trial. (*B*) On Saccade trials, the fixation dot jumped from left to middle to right or right to middle to left; saccade direction was equally distributed across all 3 conditions. (*C*) Illustration of stimulus construction for Experiment 1. Panoramic scenes were divided into 5 panels, only 3 of which were shown in a given view. The resulting views overlapped by 66% in each step. Fixation locations were chosen so that local retinotopic stimulation would be identical across the 3 Saccade-Scrolling views (shown). (*D*) Sample 3D virtual environment stimuli from Experiment 2. (*E*, *F*) Example images for a single trial of each of the 3 Fixation conditions (*E*) and Saccade conditions (*F*). Note that the same scene is reused here for illustration purposes, but individual scenes were never repeated across trials in the main task.

Localizer Task

Scene-selective cortical regions were identified using an independent localizer run. Subjects viewed alternating blocks of scenes and faces, for a total of 12 blocks. Within each block, 16 images (24° × 18°) were presented for 500 ms each, interspersed with 500 ms fixations. Subjects performed a one-back repetition task, pressing a button whenever a repeat was detected. Localizer faces were selected from a pool of 32 color photographs of male and female faces. Localizer scenes were taken from a separate pool of 32 scene images not used in the main task.

fMRI Data Acquisition

MRI scanning was carried out with a Siemens Trio 3 T scanner using an 8-channel receiver array head coil. Functional data were acquired with a T_2^*-weighted gradient-echo sequence (time repetition = 2000 ms,

time echo = 25 ms, flip angle = 90°, matrix = 64 × 64). Thirty-four axial slices (3.5 mm thick, 0 mm gap) were taken oriented parallel to the anterior commissure–posterior commissure line. In Experiment 1, 4 functional runs of the main task and one functional localizer run were collected; in Experiment 2, one run of the main task and one functional localizer run were collected.

fMRI Preprocessing and Analysis

Preprocessing of the data was done using Brain Voyager QX (Brain Innovation). The first 6 volumes of each functional run were discarded, and the remaining data were corrected for slice acquisition time and head motion, spatially smoothed with a 4 mm full-width at half-maximum kernel, temporally high-pass filtered with a 128 s period cutoff, normalized into Talairach space (Talairach and Tournoux 1988), and interpolated into 3 mm isotropic voxels.

Multiple regression analyses were performed separately for each subject to obtain subject-specific regions of interest (ROIs) from the localizer run. A whole-brain contrast of the localizer run was performed for scenes greater than faces, and the single peak voxel displaying the maximal contrast was selected within each of the following 6 regions: left and right parahippocampal gyrus/collateral sulcus (PPA: mean Talairach coordinates −26, −42, −6 for left and 25, −40, −7 for right), left and right transverse occipital sulcus (TOS: mean Talairach coordinates −31, −78, 14 for left and 30, −76, 17 for right), and left and right retrosplenial cortex (RSC: mean Talairach coordinates −16, −56, 15 for left and 15, −54, 15 for right). Spherical ROIs (4 mm radius) were then created around the peak voxels and applied to the main task.

For the main task, a random-effects general linear model (GLM), using a canonical hemodynamic response function (HRF), was used to extract beta weights for each ROI for each of the 6 conditions, plus a null fixation condition. A separate GLM using 10 finite impulse response (FIR) functions was also conducted within each ROI to generate timecourses of activity for each condition. Responses were collapsed across hemispheres, and adaptation indices were calculated using the Fixation-Novel and Saccade-Novel conditions as baselines for the other Fixation and Saccade conditions, respectively (e.g., Adaptation Index for Saccade-Scrolling = [Saccade-Scrolling − Saccade-Novel]/[Saccade-Scrolling + Saccade-Novel]). Repeated measures analyses of variance (ANOVAs) and paired t-tests were performed to compare adaptation indices across conditions.

Eye Tracking

Eye position was continuously tracked and recorded for each trial. The eye tracker was calibrated at the beginning of the experiment and recalibrated as necessary between runs. For one subject, the eye tracker malfunctioned and fixation was monitored manually over the video feed. Because accurate fixation and eye movements were a critical part of this experiment, in addition to monitoring eye position in the scanner, all subjects were first trained on the task with an eye tracker outside the scanner. Subjects were also given feedback on their eye-tracking performance after each run during both practice and main tasks. Figure 2 shows a series of sample eye traces from a single subject for 4 consecutive trials in the main fMRI task (Experiment 1), illustrating successful eye position behavior on both fixation and saccade trials. On saccade trials, average saccadic latency across all subjects was 317.2 ms (standard deviation [SD] 54.6 ms) for the first (unpredictable) saccade and 209.5 ms (SD 46.8 ms) for the second (predictable) saccade. There were no significant differences in latency for saccade between Saccade-Identical, Saccade-Scrolling, and Saccade-Novel conditions (all t's < 1).

Results

PPA Timecourses

Because every trial consisted of a series of 3 images presented in succession, all blood oxygen level–dependent (BOLD) responses reflect the hemodynamic response to the set of 3 images. If the PPA adapts to repeated views of the same image, then we should expect a reduced BOLD response for 3 identical images compared with 3 different images. This pattern is obvious in the left panel of Figure 3, which illustrates the timecourse of the PPA BOLD response for the 3 types of Fixation conditions. In the absence of eye movements, clear adaptation can be seen when identical views of the same scene are repeated (Fixation-Identical) compared with a series of different scenes (Fixation-Novel). Consistent with previous work (Park and Chun 2009), when different but overlapping views of the same panoramic scene were presented (Fixation-Scrolling), little if any PPA adaptation was present compared with the novel images.

In the Saccade conditions, a strikingly different pattern emerged. The PPA response appeared reduced for both types of repeated images compared with the novel scenes. Adaptation
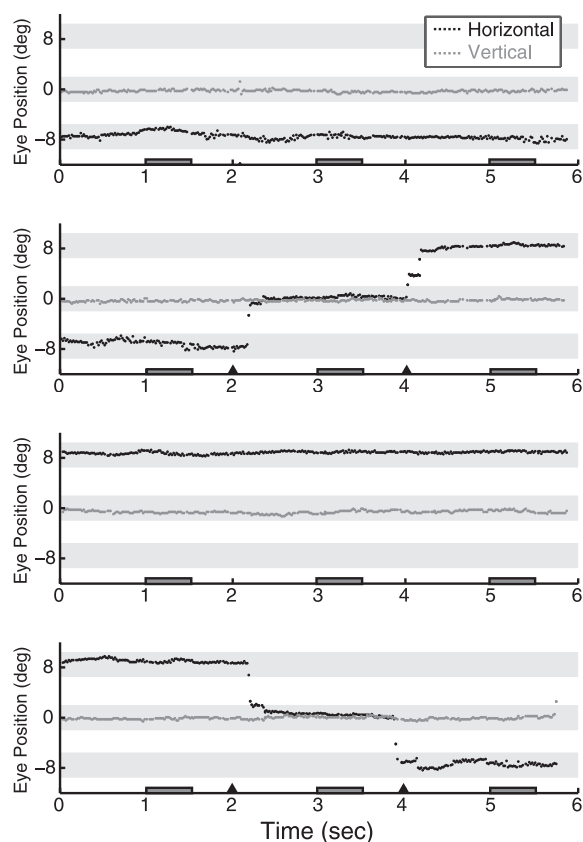

Figure 2. Sample eye traces. Eye traces are shown for a single representative subject on 4 consecutive trials of Experiment 1. Horizontal and vertical eye position are plotted in degrees with zero indicating the center of the screen. Vertical eye position always remained at center. On half of the trials horizontal eye position also remained stable, at either the left fixation position (panel 1) or right fixation position (panel 3). On the other half of the trials, a series of horizontal eye movements was made from left to middle to right positions (panel 2) or right to middle to left (panel 4). Eye position over the full 6 s trial is shown. Dark gray boxes along the x-axis indicate the 3 stimulus periods on each trial, and triangles indicate fixation location changes on saccade trials. Light gray shading illustrates the fixation locations ±2 degrees: that is, the fixation window that subjects needed to remain within for successful fixation.

was found when eye movements were made across scrolling images (where retinotopic input was repeated) and when eye movements were made across identical images (where spatio-topic input was repeated), although the magnitude of adaptation appeared slightly greater for the retinotopic repeats.

PPA Adaptation Index

To quantify these effects, we modeled the BOLD response using a canonical HRF and calculated adaptation indices for the Identical and Scrolling conditions compared with Novel; this was calculated separately for Fixation and Saccade trials. PPA adaptation indices for Experiment 1 are shown in Figure 4A. Significant adaptation was found for the Fixation-Identical (t_{11} = 3.80, P = 0.003), Saccade-Identical (t_{11} = 3.61, P = 0.004), and Saccade-Scrolling (t_{11} = 3.61, P = 0.004) conditions (Bonferroni-corrected alpha: 0.05/4 = 0.015) but not for Fixation-Scrolling (t < 1). To compare the magnitude of adaptation across these 4 conditions, we conducted a 2 (Fixation/Saccade) × 2 (Identical/Overlap) ANOVA. The main effects of Fixation/Saccade (F < 1) and Identical/Scrolling ($F_{1,11}$ = 2.36, P = 0.15) were not significant, but the interaction was
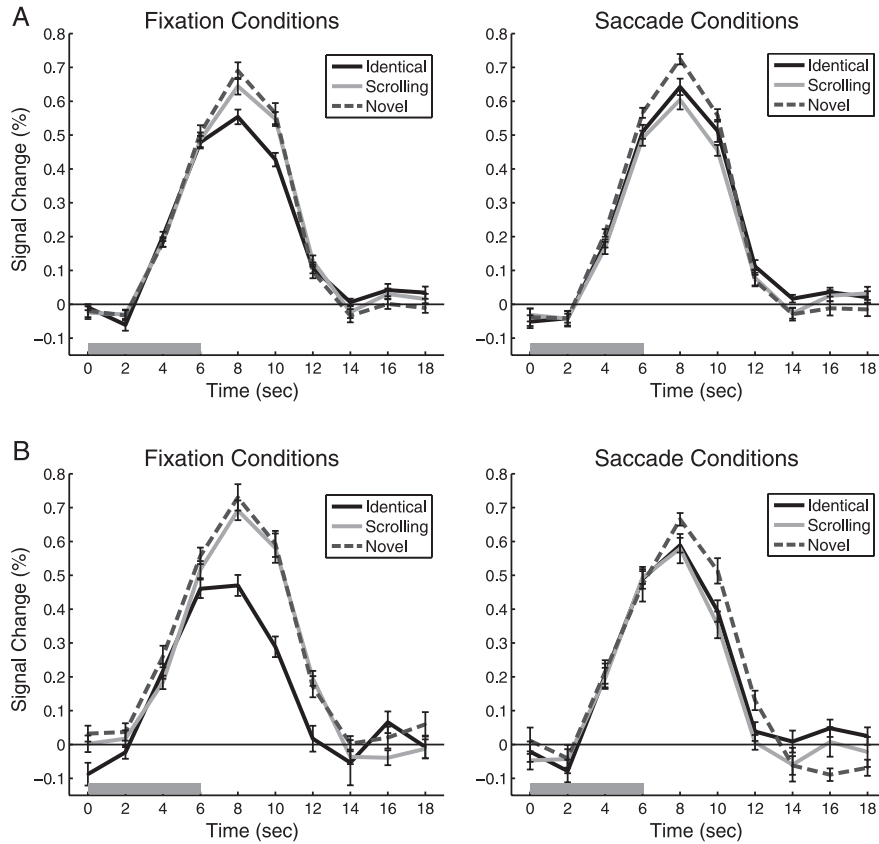
Figure 3. PPA Timecourses. FIR timecourses are shown for each of the 3 Fixation conditions and Saccade conditions. (*A*) Experiment 1, $N = 12$. (*B*) Experiment 2, $N = 10$. The 6-s trial stimulation period is indicated by the gray bar. Error bars are standard error of the mean after normalization to remove between-subject variability at each timepoint (Loftus and Masson 1994).

significant ($F_{1,11} = 10.65$, $P = 0.008$). This interaction was not significantly modified by hemisphere when left and right PPAs were analyzed separately ($F < 1$), and the same pattern was found when the peaks of the FIR were used instead of overall beta weights (interaction term for peaks: $F_{1,11} = 13.48$, $P = 0.004$). Post hoc paired *t*-tests revealed that adaptation was significantly greater for Fixation-Identical compared with Fixation-Scrolling ($t_{11} = 3.76$, $P = 0.003$); the reversal trended toward significance in the Saccade conditions, with adaptation greater for Saccade-Scrolling than Saccade-Identical ($t_{11} = -1.89$, $P = 0.086$).

Experiment 2: Virtual 3D Scenes

To test the robustness of these effects across different types of viewpoint changes, we created a set of virtual stimuli simulating natural 3D motion and rotation. PPA adaptation indices are shown in Figure 4*B*. As in Experiment 1, significant adaptation was found for the Fixation-Identical ($t_9 = 3.26$, $P = 0.010$), Saccade-Identical ($t_9 = 2.4$, $P = 0.040$), and Saccade-Scrolling ($t_9 = 2.44$, $P = 0.037$) conditions but not for Fixation-Scrolling ($t < 1$). The 2×2 ANOVA interaction was significant ($F_{1,9} = 24.76$, $P = 0.001$), while neither main effect of Fixation/Saccade ($F < 1$) or Identical/Scrolling ($F_{1,9} = 3.40$, $P = 0.098$) reached significance. Post hoc paired *t*-tests revealed that adaptation was significantly greater for Fixation-Identical compared with Fixation-Scrolling ($t_9 = 4.34$, $P = 0.002$), with no significant difference between Saccade-Scrolling and Saccade-Identical ($t < 1$). The magnitude of Fixation-Identical adaptation was considerably higher than in Experiment
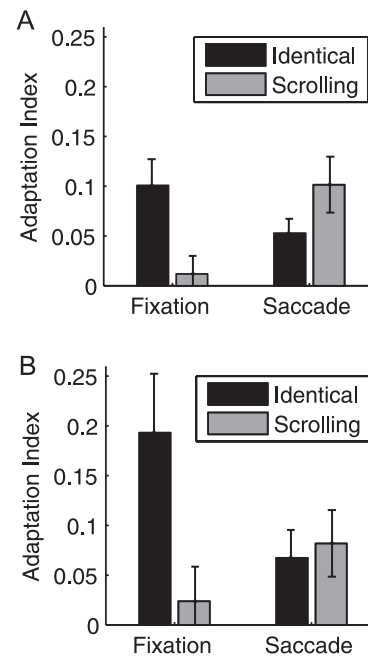

Figure 4. PPA adaptation indices. Adaptation indices were calculated using the Fixation-Novel and Saccade-Novel conditions as baselines (see text). (*A*) Experiment 1, $N = 12$. (*B*) Experiment 2, $N = 10$. Error bars are standard error of the mean.

1; however, this is likely due to differences in stimulus properties between the 2 experiments, and the possibility that the 3D virtual stimuli might have been more attentionally engaging, a factor known to modulate adaptation effects (Yi and Chun 2005). Regardless, the overall pattern of adaptation is remarkably similar across the 2 experiments.

Other Scene-Selective Areas

In addition to the PPA, in Experiment 1, we localized 2 additional scene-selective regions: TOS (12 subjects) and RSC (10 subjects). Adaptation indices for these 2 regions are shown in Figure 5. In the TOS, significant adaptation was found for the Fixation-Identical ($t_{11} = 2.36$, $P = 0.038$) and Saccade-Scrolling ($t_{11} = 2.55$, $P = 0.027$) conditions but not for Fixation-Scrolling or Saccade-Identical (both t's < 1). The 2 × 2 ANOVA interaction was significant ($F_{1,11} = 7.01$, $P = 0.023$), while neither main effect of Fixation/Saccade or Identical/Scrolling reached significance (both F's < 1). In the RSC, the data were much noisier, and adaptation for none of the 4 conditions reached significance (Fixation-Identical: $t_9 = 1.84$, $P = 0.099$; Fixation-Scrolling: $t < 1$; Saccade-Identical: $t < 1$; Saccade-Scrolling: $t_9 = 1.58$, $P = 0.15$). The overall pattern was similar to the TOS, with a significant ANOVA interaction ($F_{1,9} = 10.86$, $P = 0.009$) and neither main effect significant (both F's < 1). Because the effects in TOS and especially RSC were generally noisier than in the PPA, given the reduced experimental power and subject numbers in Experiment 2, we did not extend the analyses in Experiment 2 beyond the PPA.

We also investigated a visually active nonscene-selective region in early visual cortex (selected bilaterally from the

Figure 5. TOS and RSC adaptation indices. Adaptation indices were calculated using the Fixation-Novel and Saccade-Novel conditions as baselines (see text). (*A*) TOS, Experiment 1, *N* = 12. (*B*) RSC, Experiment 1, *N* = 10. Error bars are standard error of the mean.

localizer contrast of all > fixation). Adaptation was not found for either saccade condition (both t's < 1). Although one might expect retinotopic (Saccade-Scrolling) adaptation in low-level visual areas, several factors could be contributing to its absence, including the likelihood that the eye movements themselves activated early visual cortex, which could cause a release from adaptation or an increase in activation that would wash out adaptation effects, as well as the fact that our early visual ROI was not restricted to foveal eccentricities and thus could be responding to other visual information (such as the outline of the scene or projection screen) that would have varied retinotopically with each eye movement regardless of condition. In contrast, because the PPA is scene selective, it is not susceptible to this potential concern.

Discussion

The PPA is well known for its selectivity for scenes compared with other types of stimuli (Epstein and Kanwisher 1998), as well as its specificity in representing individual viewpoint-dependent scenes (Epstein et al. 2003). With the eyes fixated, the PPA response was significantly attenuated for successive presentations of the same exact scene but not for successive presentations of different but overlapping views of that scene, replicating earlier findings (Park and Chun 2009). In these Fixation conditions, because the eyes never changed position, spatiotopic and retinotopic contributions to adaptation could not be separated—spatiotopic and retinotopic inputs were either both identical (Fixation-Identical) or both changing (Fixation-Scrolling). In the critical Saccade conditions, we were able to separately explore the contributions of each. Interestingly, significant adaptation was found for both spatiotopic repeats (where the eyes moved to explore a stationary scene) and retinotopic repeats (where the eyes and scene moved together). These effects cannot be attributable to the eye movements themselves; adaptation was always referenced to a baseline condition of novel images over which identical eye movements were executed. This pattern was replicated with a different set of subjects in Experiment 2 using virtual 3D scenes, which provided a more ecologically valid viewpoint change.

It is striking that the PPA did not adapt to successive views when the eyes remained stationary and the scene moved in the background (Fixation-Scrolling) but it did adapt in the condition where the eyes moved to analogous points across the stationary scene (Saccade-Identical). This indicates that retinotopic input is not the sole factor in PPA adaptation. It also suggests that our natural use of eye movements to explore a visual scene cannot be emulated simply by moving the scene in the background to match the retinotopic changes. Of course, the timing and nature of the eye movements in our task were still not as natural as one might experience in the real world; the longer delay between eye movements and stereotyped movements themselves were necessary to carefully control visual input and ensure that subjects were successfully fixating on each location. An important avenue for future research will be determining how these effects vary under true free-viewing conditions.

The difference between the Fixation-Scrolling and Saccade-Identical results—which were matched for retinotopic input—could be attributable to one of 2 things: Either (1) the act of executing an eye movement provides important cues for visual stability, such as corollary discharge or efferent oculomotor feedback during remapping (Duhamel et al. 1992; Sommer
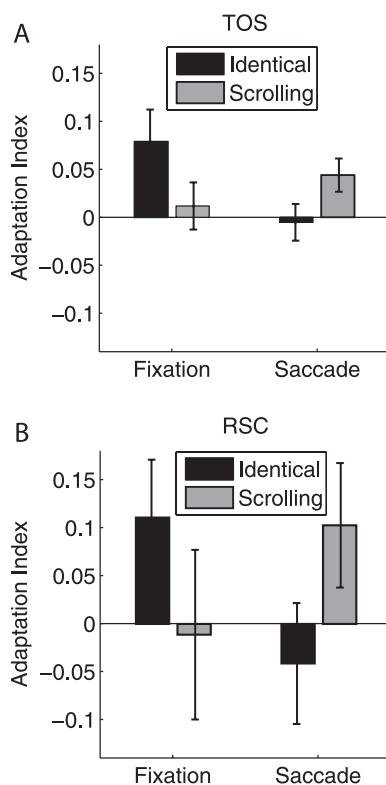
and Wurtz 2006) or modulatory input from eye position gain fields (Andersen et al. 1997) and/or (2) the PPA utilizes global spatiotopic information—regardless of eye position—to compare whether a subsequently viewed scene is the same as a previous one. The existence of explicit spatiotopic maps has remained under debate (for review, see Cavanagh et al. 2010), although spatiotopic fMRI adaptation effects have been previously reported in the lateral occipital complex (LOC: McKyton and Zohary 2007). Our results cannot differentiate between these 2 possibilities, but it will be an important goal for future research. Regardless, these findings suggest that models of scene recognition need to take into account not only the images being viewed but also how those images are acquired by the viewer and the role of eye movements in this process.

In addition to the spatiotopic repetition effects, the PPA exhibited strong adaptation to the retinotopic repeats. Retinotopic adaptation may not be particularly surprising given the sensitivity of the PPA to the physical similarity between images (Grill-Spector and Malach 2001; Xu et al. 2007), but previous studies have not explored whether this physical similarity should be defined in retinotopic or spatiotopic coordinates. Furthermore, if making a sequence of eye movements across a scene provides a more natural means of exploring the scene, then the retinotopic Saccade-Scrolling condition, where the scene shifts with the eyes, is arguably the most unnatural. In this condition, oculomotor feedback should suggest that the visual representation was updated since the eyes landed on a new location in the scene, yet the actual visual input had not changed accordingly. Although efferent feedback may be imprecise or incomplete (Bridgeman 2007), visual information present before and after a saccade can also be used to aid stable representations across eye movements (McConkie and Currie 1996), so when this information does not update as expected, that would seem to jeopardize stability. However, despite the artificial nature of the Saccade-Scrolling condition, the PPA treated the successive views as if they were completely identical; adaptation was just as strong as adaptation to repeated identical images at fixation. This indicates that local retinotopic information, which was preserved in this Saccade-Scrolling condition, plays a large role in the representation of scenes in the PPA, an idea consistent with recent results suggesting that the PPA is considerably more sensitive to retinal location than initially thought (Schwarzlose et al. 2008; Arcaro et al. 2009). It also indicates that the design succeeded at equating retinotopic input across eye movements, despite the likely existence of microsaccades and other occasional small drifts in eye position. (Note that in Experiment 2, Saccade-Scrolling adaptation was relatively weaker, although in this experiment the 3D viewpoint changes were not simple translations of the image, and thus retinotopic input was not completely identical across views, decreasing the physical similarity between images in this condition.) Given that real-world scenes do not typically move when the eyes move, perhaps these retinotopic cues have evolved as a reliable indicator of physical similarity such that the PPA relies more on retinotopic physical similarity than potentially competing top-down cues.

These data could have important implications for comparisons of retinotopic and spatiotopic representations in general. A number of recent studies have used eye movements to dissociate these eye-centered and world-centered frames of reference. Spatiotopic representations have been reported for certain behavioral phenomena (Hayhoe et al. 1991; Shimojo et al. 1996; Melcher and Morrone 2003; Burr et al. 2007; Ong et al.

2009; Pertzov et al. 2010) and neural populations (Duhamel et al. 1997; Snyder et al. 1998; d'Avossa et al. 2007; McKyton and Zohary 2007), and it has been suggested that spatiotopic representations become more common with more complex stimuli and later visual brain regions (Andersen et al. 1997; Wurtz 2008; Melcher and Colby 2008). However, an increasingly prevalent theme is that natively retinotopic representations might not automatically accommodate spatiotopy (Gardner et al. 2008; Golomb et al. 2008; Afraz and Cavanagh 2009; Knapen et al. 2009; Cavanagh et al. 2010; Golomb et al. 2010). The current results are consistent with both ideas; larger receptive fields and more position-invariant representations (MacEvoy and Epstein 2007) may enable spatiotopic representations in the PPA, but the retinotopic reference frame still dominates. When compelling cues are present, such as active oculomotor signals, this native retinotopic information might be dynamically transformed into more ecologically relevant spatiotopic representations.

It is notable that we find spatiotopic effects in the PPA, a ventral stream area, when spatiotopic representations and eye movement effects are more typically associated with the dorsal stream (Galletti et al. 1993; Duhamel et al. 1997; Snyder et al. 1998; Mullette-Gillman et al. 2009, although there has been a previous report of spatiotopic adaptation in ventral area LOC: McKyton and Zohary 2007). However, as noted above, it is unclear whether our adaptation effects reflect a true coherent spatiotopic representation or whether the retinotopic representations are dynamically updated by eye movement cues to create the appearance of spatiotopy. Indeed, a recent report suggests that both dorsal and ventral regions contain information about retinotopic position and eye position, without containing any explicit information about spatiotopic position (Golomb and Kanwisher, unpublished data). An additional possibility is that with its larger receptive fields, the PPA might be sensitive to repeated information anywhere within the larger area (MacEvoy and Epstein 2007), even if the exact retinotopic position changes, something analogous to a level between retinotopic and true spatiotopic representations. Regardless of the exact mechanism underlying the Saccade-Identical adaptation, it is clear that scene representations in the PPA reflect more than pure retinotopic position and are less viewpoint specific than previously thought.

Interestingly, in the TOS and RSC scene-selective regions, we found only retinotopic adaptation. The TOS bias toward retinotopic adaptation is consistent with its location near earlier visual areas with known retinotopic organization (e.g., area V3A: Tootell et al. 1997). The RSC result is somewhat surprising, however, because RSC is thought to serve an integrating function across different views of scenes (Epstein 2008; Park and Chun 2009; Vann et al. 2009) and reference frames (Committeri et al. 2004; Byrne et al. 2007; Epstein 2008; Vann et al. 2009) to accommodate broader maps of the world. Surprisingly, not only was spatiotopic adaptation absent but RSC did not adapt to overlapping scenes at fixation, despite this condition producing significant RSC adaptation in previous studies (Park and Chun 2009). However, in the current study, the RSC responses were noisy across all conditions, and the overall magnitude of RSC response was considerably lower than that of the other scene-selective regions (peak activation of 0.25% signal change for RSC compared with 0.6-0.7% signal change for PPA and TOS). It thus appears that our task was simply not effective at generating RSC responses in general. Although the scene stimuli were highly similar to those used in the Park and Chun (2009) study, subjects

were not performing the same task on these stimuli. In the previous study, subjects were instructed to attend to the scenes and attempt to memorize their layout and details for a subsequent memory test, while in the current study, the primary task was to maintain fixation at the indicated fixation location while passively viewing the scenes. Prior work has demonstrated that RSC activity is task dependent, exhibiting the strongest responses during location tasks (Epstein et al. 2007). Perhaps during passive viewing, without any explicit task-emphasis on the scenes themselves, the RSC simply was not reliably activated by our task, which could explain the lack of adaptation effects.

The PPA is less sensitive to task manipulations (Epstein et al. 2007), yet more sensitive to stimulus changes (Epstein 2008). The ability to link scenes across changes in viewpoint, particularly across eye movements, is clearly critical for successful behavior. However, the ability to differentiate between scenes is also important. The current study allows us to explore whether retinotopic and spatiotopic cues are necessary and/or sufficient for these abilities. We report that the PPA is sensitive to both types of repetition but more sensitive to local retinotopic information. In other words, the PPA exhibits the same magnitude of adaptation whenever the retinotopic image is preserved, regardless of whether total spatiotopic overlap of the scene is 100% across the 3 views (Fixation-Identical condition) or 33% (Saccade-Scrolling condition). This suggests that while neither retinotopic or spatiotopic information alone may be necessary to construct a stable percept in the PPA, retinotopic information may be sufficient. However, the fact that significant adaptation was found when retinotopic information was disrupted as the eyes moved across a stationary spatiotopic scene reveals that representations of viewpoint invariance in the PPA may be constructed from both retinotopic and spatiotopic cues. Perhaps the strongest cue for the PPA is not what makes a scene the same but what signals a change in scene: if neither retinotopic nor spatiotopic information is repeated across views, it is not productive to integrate across them.

Funding

Notes

References

Afraz A, Cavanagh P. 2009. The gender-specific face aftereffect is based in retinotopic not spatiotopic coordinates across several natural image transformations. J Vis. 9(10):10.1–10.17.

Andersen RA, Snyder LH, Bradley DC, Xing J. 1997. Multimodal representation of space in the posterior parietal cortex and its use in planning movements. Annu Rev Neurosci. 20:303–330.

Arcaro MJ, McMains SA, Singer BD, Kastner S. 2009. Retinotopic organization of human ventral visual cortex. J Neurosci. 29:10638–10652.

Brainard DH. 1997. The psychophysics toolbox. Spat Vis. 10:433–436.

Bridgeman B. 2007. Efference copy and its limitations. Comput Biol Med. 37:924–929.

Bridgeman B, Hendry D, Stark L. 1975. Failure to detect displacement of the visual world during saccadic eye movements. Vision Res. 15:719–722.

Burr D, Tozzi A, Morrone MC. 2007. Neural mechanisms for timing visual events are spatially selective in real-world coordinates. Nat Neurosci. 10:423–425.

Byrne P, Becker S, Burgess N. 2007. Remembering the past and imagining the future: a neural model of spatial memory and imagery. Psychol Rev. 114:340–375.

Cavanagh P, Hunt AR, Afraz A, Rolfs M. 2010. Visual stability based on remapping of attention pointers. Trends Cogn Sci. 14:147–153.

Committeri G, Galati G, Paradis AL, Pizzamiglio L, Berthoz A, LeBihan D. 2004. Reference frames for spatial cognition: different brain areas are involved in viewer-, object-, and landmark-centered judgments about object location. J Cogn Neurosci. 16:1517–1535.

d'Avossa G, Tosetti M, Crespi S, Biagi L, Burr DC, Morrone MC. 2007. Spatiotopic selectivity of BOLD responses to visual motion in human area MT. Nat Neurosci. 10:249–255.

Duhamel JR, Bremmer F, BenHamed S, Graf W. 1997. Spatial invariance of visual receptive fields in parietal cortex neurons. Nature. 389:845–848.

Duhamel JR, Colby CL, Goldberg ME. 1992. The updating of the representation of visual space in parietal cortex by intended eye movements. Science. 255:90–92.

Epstein R, Graham KS, Downing PE. 2003. Viewpoint-specific scene representations in human parahippocampal cortex. Neuron. 37:865–876.

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

Epstein RA. 2008. Parahippocampal and retrosplenial contributions to human spatial navigation. Trends Cogn Sci. 12:388–396.

Epstein RA, Parker WE, Feiler AM. 2007. Where am I now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. J Neurosci. 27:6141–6149.

Galletti C, Battaglini PP, Fattori P. 1993. Parietal neurons encoding spatial locations in craniotopic coordinates. Exp Brain Res. 96:221–229.

Gardner JL, Merriam EP, Movshon JA, Heeger DJ. 2008. Maps of visual space in human occipital cortex are retinotopic, not spatiotopic. J Neurosci. 28:3988–3999.

Golomb JD, Chun MM, Mazer JA. 2008. The native coordinate system of spatial attention is retinotopic. J Neurosci. 28:10654–10662.

Golomb JD, Nguyen-Phuc AY, Mazer JA, McCarthy G, Chun MM. 2010. Attentional facilitation throughout human visual cortex lingers in retinotopic coordinates after eye movements. J Neurosci. 30:10493–10506.

Grill-Spector K, Malach R. 2001. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychol (Amst). 107:293–321.

Hayhoe M, Lachter J, Feldman J. 1991. Integration of form across saccadic eye movements. Perception. 20:393–402.

Henderson JM. 1997. Transsaccadic memory and integration during real-world object perception. Psychol Sci. 8:51–55.

Irwin DE. 1991. Information integration across saccadic eye movements. Cognit Psychol. 23:420–456.

Knapen T, Rolfs M, Cavanagh P. 2009. The reference frame of the motion aftereffect is retinotopic. J Vis. 9(5):16.1–16.7.

Loftus GR, Masson MEJ. 1994. Using confidence intervals in within-subject designs. Psychon Bull Rev. 1:476–490.

MacEvoy SP, Epstein RA. 2007. Position selectivity in scene- and object-responsive occipitotemporal regions. J Neurophysiol. 98:2089–2098.

McConkie GW, Currie CB. 1996. Visual stability across saccades while viewing complex pictures. J Exp Psychol Hum Percept Perform. 22:563–581.

McKyton A, Zohary E. 2007. Beyond retinotopic mapping: the spatial representation of objects in the human lateral occipital complex. Cereb Cortex. 17:1164–1172.

Melcher D, Colby CL. 2008. Trans-saccadic perception. Trends Cogn Sci. 12:466–473.

Melcher D, Morrone MC. 2003. Spatiotopic temporal integration of visual motion across saccadic eye movements. Nat Neurosci. 6:877–881.

Mullette-Gillman OA, Cohen YE, Groh JM. 2009. Motor-related signals in the intraparietal cortex encode locations in a hybrid, rather than eye-centered reference frame. Cereb Cortex. 19:1761–1775.

O'Regan JK. 1992. Solving the "real" mysteries of visual perception: the world as an outside memory. Can J Psychol. 46:461–488.

Ong WS, Hooshvar N, Zhang M, Bisley JW. 2009. Psychophysical evidence for spatiotopic processing in area MT in a short-term memory for motion task. J Neurophysiol. 102:2435-2440.

Park S, Chun MM. 2009. Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. Neuroimage. 47:1747-1756.

Park S, Kim MS, Chun MM. 2007. Concurrent working memory load can facilitate selective attention: evidence for specialized load. J Exp Psychol Hum Percept Perform. 33:1062-1075.

Pertzov Y, Zohary E, Avidan G. 2010. Rapid formation of spatiotopic representations as revealed by inhibition of return. J Neurosci. 30:8882-8887.

Schacter DL, Buckner RL. 1998. Priming and the brain. Neuron. 20:185-195.

Schwarzlose RF, Swisher JD, Dang S, Kanwisher N. 2008. The distribution of category and location information across object-selective regions in human visual cortex. Proc Natl Acad Sci U S A. 105:4447-4452.

Shimojo S, Tanaka Y, Watanabe K. 1996. Stimulus-driven facilitation and inhibition of visual information processing in environmental and retinotopic representations of space. Cogn Brain Res. 5:11-21.

Snyder LH, Grieve KL, Brotchie P, Andersen RA. 1998. Separate body- and world-referenced representations of visual space in parietal cortex. Nature. 394:887-891.

Sommer MA, Wurtz RH. 2006. Influence of the thalamus on spatial visual processing in frontal cortex. Nature. 444:374-377.

Stevens JK, Emerson RC, Gerstein GL, Kallos T, Neufeld GR, Nichols CW, Rosenquist AC. 1976. Paralysis of the awake human: visual perceptions. Vision Res. 16:93-98.

Talairach J, Tournoux P. 1988. Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging. New York: Thieme.

Tootell RB, Mendola JD, Hadjikhani NK, Ledden PJ, Liu AK, Reppas JB, Sereno MI, Dale AM. 1997. Functional analysis of V3A and related areas in human visual cortex. J Neurosci. 17:7060-7078.

Vann SD, Aggleton JP, Maguire EA. 2009. What does the retrosplenial cortex do? Nat Rev Neurosci. 10:792-802.

Wiggs CL, Martin A. 1998. Properties and mechanisms of perceptual priming. Curr Opin Neurobiol. 8:227-233.

Wurtz RH. 2008. Neuronal mechanisms of visual stability. Vision Res. 48:2070-2089.

Xu Y, Turk-Browne NB, Chun MM. 2007. Dissociating task performance from fMRI repetition attenuation in ventral visual cortex. J Neurosci. 27:5981-5985.

Yi DJ, Chun MM. 2005. Attentional modulation of learning-related repetition attenuation effects in human parahippocampal cortex. J Neurosci. 25:3593-3600.