Coarse-to-fine Categorization of Visual Scenes in Scene-selective Cortex

Benoit Musel[1,2], Louise Kauffmann[1], Stephen Ramanoël[1], Coralie Giavarini[1], Nathalie Guyader[1], Alan Chauvin[1], and Carole Peyrin[3]

Abstract

■ Neurophysiological, behavioral, and computational data indicate that visual analysis may start with the parallel extraction of different elementary attributes at different spatial frequencies and follows a predominantly coarse-to-fine (CtF) processing sequence (low spatial frequencies [LSF] are extracted first, followed by high spatial frequencies [HSF]). Evidence for CtF processing within scene-selective cortical regions is, however, still lacking. In the present fMRI study, we tested whether such processing occurs in three scene-selective cortical regions: the parahippocampal place area (PPA), the retrosplenial cortex, and the occipital place area. Fourteen participants were subjected to functional scans during which they performed a categorization task of indoor versus outdoor scenes using dynamic scene stimuli. Dynamic scenes were composed of six filtered images of the same scene, from LSF to HSF or from HSF to LSF, allowing us to mimic a CtF or the reverse fine-to-coarse (FtC) sequence. Results showed that only the PPA was more activated for CtF than FtC sequences. Equivalent activations were observed for both sequences in the retrosplenial cortex and occipital place area. This study suggests for the first time that CtF sequence processing constitutes the predominant strategy for scene categorization in the PPA. ■

INTRODUCTION

One of the main functions of the visual system is to categorize the environment. A considerable number of studies on the visual system in humans and animals suggest that spatial frequencies are crucial in the visual categorization of scenes. In terms of signal representation, the image of a scene can be expressed in the Fourier domain as both amplitude and phase spectra (Hughes, Nozawa, & Kitterle, 1996; Tolhurst, Tadmor, & Chao, 1992; Field, 1987, Ginsburg, 1986). The amplitude spectrum decomposes the scene in terms of spatial frequencies and orientations, and the phase spectrum describes the relationship between spatial frequencies. In primates, the primary visual cortex is mainly dominated by simple and complex cells that respond preferentially to orientations and spatial frequencies (Shams & von der Malsburg, 2002; De Valois, Albrecht, & Thorell, 1982; De Valois, Yund, & Hepler, 1982; Poggio, 1972). In humans, simulations and psychophysical experiments have shown that information from low/medium spatial frequencies is sufficient to allow rapid scene categorization (Guyader, Chauvin, Peyrin, Herault, & Marendaz, 2004; Torralba & Oliva, 2003; Schyns & Oliva, 1994). For example, Schyns and Oliva (1994) used hybrid stimuli made of two superimposed images of natural scenes, taken from different semantic categories and containing different spatial frequencies (e.g., a highway scene in low spatial frequencies superimposed on a city scene in high spatial frequencies). The perception of these hybrid scenes was dominated by low spatial frequency information when presentation time was very brief (30 msec), but by high spatial frequency information when presentation time was longer (150 msec), suggesting precedence of low spatial frequencies (LSF) over high spatial frequencies (HSF) in the visual processing time course. These data support the influential neurobiological models of visual recognition (Peyrin et al., 2010; Bar, 2003; Bullier, 2001). According to these models, visual analysis starts with the parallel extraction of different elementary visual attributes at different spatial frequencies in a predominantly coarse-to-fine (CtF) processing sequence. The LSF in a scene, conveyed by fast magnocellular visual channels, might therefore activate visual pathways and subsequently access the occipital cortex and high-order areas in the dorsal visual cortex (extending to the parietal and frontal cortices) and ventral visual cortex (to the inferotemporal cortex) more rapidly than HSF. The rapid analysis of LSF allows an initial perceptual parsing of visual inputs before their complete propagation along the ventral visual stream, which ultimately mediates object recognition. This initial

[1]Université Grenoble Alpes, [2]University of Geneva, [3]CNRS, LPNC UMR 5105, Grenoble, France

Journal of Cognitive Neuroscience 26:10, pp. 2287–2297
doi:10.1162/jocn_a_00643

low-pass visual analysis might serve to refine the subsequent processing of HSF, conveyed more slowly by parvocellular visual channels to the ventral visual cortex.

The ventral visual cortex contains a mosaic of different areas that respond selectively to different categories of visual stimuli (Spiridon & Kanwisher, 2002; Haxby et al., 2001; Lerner, Hendler, Ben-Bashat, Harel, & Malach, 2001). For example, faces selectively activate a lateral region of the fusiform gyrus, called the fusiform face area (FFA; Kanwisher, McDermott, & Chun, 1997), whereas man-made objects primarily activate the lateral occipital complex (LOC; Grill-Spector, Kourtzi, & Kanwisher, 2001; Malach et al., 1995). A number of fMRI studies have specifically investigated the cerebral structures involved in complex scene processing compared with other visual stimuli (e.g., faces and objects). Most studies agree that a prominent region in the inferotemporal cortex, known as the parahippocampal place area (PPA), and the retrosplenial cortex (RSC) are regions of the human cortex primarily involved in processing "spatial layout" during the perception of scenes (Epstein & Ward, 2010; Epstein, 2005, 2008; Epstein & Higgins, 2007; Epstein, Graham, & Downing, 2003; Epstein, Harris, Stanley, & Kanwisher, 1999; Epstein & Kanwisher, 1998), navigationally relevant spatial information within a familiar real world environment (Vass & Epstein, 2013; Epstein, Higgins, Jablonski, & Feiler, 2007) or contextual associations (Bar, Aminoff, & Ishai, 2008; Bar, Aminoff, & Schacter, 2008; Aminoff, Gronau, & Bar, 2007; Bar, 2004, 2007; Bar & Aminoff, 2003). A region around the transverse occipital sulcus was also more recently found to be involved in both the "spatial layout" processing and the semantic categorization process during the perception of scenes (Dilks, Julian, Paunov, & Kanwisher, 2013). As a result, this region is now called the occipital place area (OPA).

However, the specific functions supported by scene-selective regions during the categorization of scenes remain unclear. We do not as yet know if scene-selective regions use CtF categorization because, to our knowledge, no imaging studies have explored the effects of different spatial frequency orders during the explicit categorization of complex visual scenes within these regions. In the present fMRI study, we measured the activation of scene-preferring regions during the categorization of dynamic natural scene stimuli (see Musel, Chauvin, Guyader, Chokron, & Peyrin, 2012), in which they resorted to either a CtF sequence or a reverse fine-to-coarse (FtC) sequence. We first identified scene-selective regions in each individual using a localizer adapted from previous studies (Walther, Caddigan, Fei-Fei, & Beck, 2009; Bar, Aminoff, & Ishai, 2008; Epstein et al., 2003; Epstein & Kanwisher, 1998) in which participants viewed grayscale photographs of scenes, faces, and common objects. The contrast between scenes and other categories was intended to enable the localization of the regions involved in the perception of scenes. Once localized, we compared activation elicited by CtF and FtC dynamic scenes within the areas defined as the PPA, RSC, and OPA.

METHODS

Participants

Sixteen right-handed participants (eight men; 23 ± 2 years) with normal or corrected-to-normal vision and no history of neurological disorders were included in this experiment. All participants gave their informed written consent before participating in the study, which was approved by the local ethics committee. All participants were submitted to two experiments: the dynamic scene experiment and the localizer experiment.

Stimuli and Procedure in the Dynamic Scene Experiment

Stimuli consisted of 40 black and white photographs (256-level grayscale, 1024 × 768 pixels) of man-made scenes classified into two distinct categories (20 indoor scenes and 20 outdoor scenes) with a visual angle of 24 × 18 degrees. Exemplars from both categories were chosen to ensure similarity of amplitude spectrum and to prevent categorization from being based on this type of visual cue (Guyader et al., 2004) and to avoid contrast energy differences between categories that could interfere with the sequence of spatial frequency processing. In both categories, images had the same distribution of energy in spatial frequencies and dominant orientations (as shown by the mean amplitude spectrum of nonfiltered natural scenes in each category; Figure 1A). To ensure that the chosen scenes have similar amplitude spectra, we first calculated the mean amplitude spectrum for the 20 indoor scenes (mean AS indoor) and the 20 outdoor scenes (mean AS outdoor). Then, for each scene, we calculated two 2-D correlation coefficients, one between the scene's amplitude spectrum and the mean AS indoor and the other one between the scene's amplitude spectrum and the mean AS outdoor. The mean AS of the category corresponding to the scene of interest was calculated by excluding the scene's amplitude spectrum (i.e., for an indoor scene, the mean AS indoor was calculated based on the 19 remaining indoor scenes, whereas the mean AS outdoor was calculated based on the 20 outdoor scenes). The 2-D correlation coefficient was calculated using the Matlab function "corr2d." A 2 × 2 variance analysis (ANOVA) with the category of the Scene (indoor and outdoor) and the category of the Mean AS (indoor and outdoor) as within-subject factors were conducted on the 2-D correlation coefficients. Results show that the 2-D correlation coefficients calculated between indoor scenes and the mean AS indoor did not significantly differ from those calculated between indoor scenes and the mean AS outdoor (0.76 ± 0.05 and 0.76 ± 0.06, respectively; $F(1, 38) < 1$). Similarly, the 2-D correlation

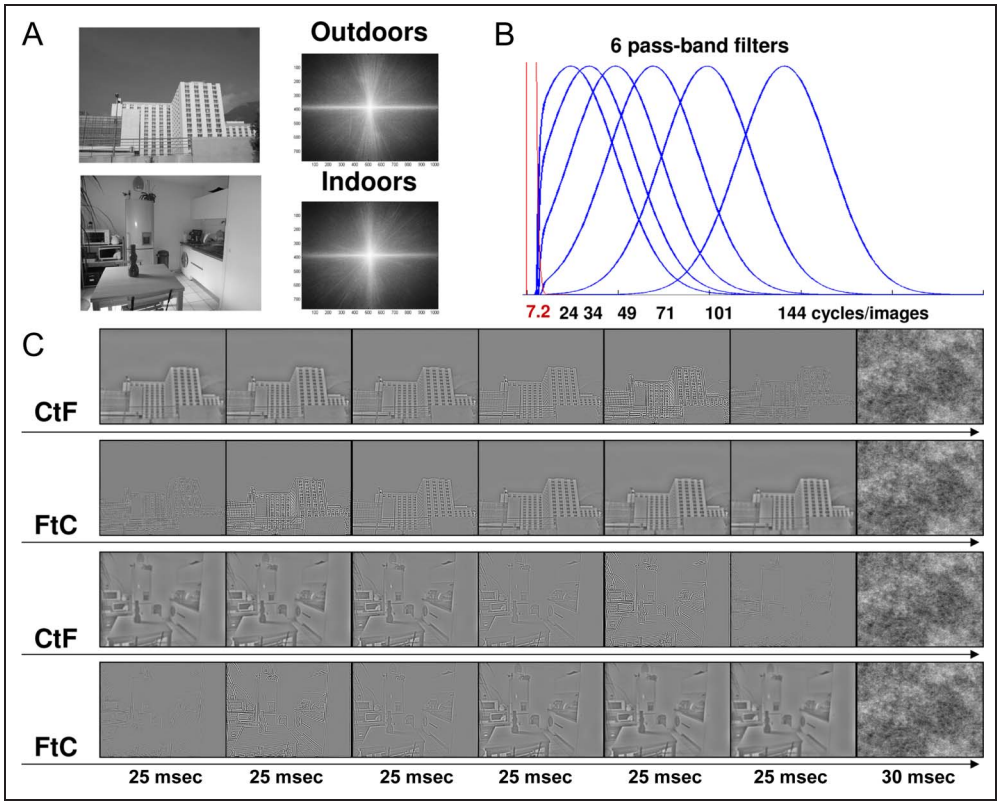

Figure 1. (A) Example of scenes belonging to different categories (outdoors and indoors). Mean amplitude spectra of each category. On each amplitude spectrum, low spatial frequencies are close to the center, and high spatial frequencies are on the periphery. Vertical orientations are represented on the *x* axis, and horizontal orientations are on the *y* axis. (B) Example of six spatial frequency filtered images of scenes depicting (C) the CtF and FtC sequences.

coefficients calculated between outdoor scenes and the mean AS outdoor did not significantly differ from those calculated between outdoor scenes and the mean AS indoor (0.78 ± 0.05 and 0.78 ± 0.04, respectively; $F(1, 38) < 1$).

Furthermore, outdoor and indoor categories were equivalent in terms of visual cluttering (subband entropy measures; see Rosenholtz, Li, & Nakano, 2007). The mean subband entropy was equivalent for outdoors and indoors (2.95 ± 0.16 and 2.95 ± 0.14, respectively; $F(1, 38) < 1$). Stimuli were created using the image processing toolbox on MATLAB (Mathworks, Inc., Sherborn, MA). On the basis of our previous studies (Musel et al., 2012), each scene was filtered by six band-pass filters with different central spatial frequencies: 24, 34, 49, 71, 101, 144 cycles/image and a standard deviation of 25.6 cycles/image (or 1, 1.42, 2.04, 2.96, 4.21, 6 cycles/degree and a standard deviation of 1.07 cycles/degree). The central frequencies of filters followed a logarithmic scale. This enabled us to obtain a better sample of the amplitude spectrum of natural scene, in which the energy decreases as frequency increases ("1/f" shape; Field, 1987) and more filters centered on low spatial frequencies (for a similar approach, see Willenbockel et al., 2010). For the first filter, we removed information contained in the frequencies below 7.2 cycles/image (or 0.3 cycles/degree) and retained higher LSF to apply a band-pass filter even for the lower spatial frequency image (see Figure 1). The cut-off frequencies at 67% of the height of each Gaussian were, therefore, [7 47] [11 58] [26 72] [47 93] [78 124] [121 167] cycles/image (i.e., [0.29 1.96] [0.46 2.42] [1.08 3] [1.96 3.88] [3.25 5.17] [5.04 6.97] cycles/degree). Images were then normalized to obtain a mean luminance of 128 with a standard deviation of 25.5 (i.e., root mean square [RMS]) on a 256 gray-level scale. This resulted in six versions of each scene, all accurately categorized,[1] and from these we created a movie. For each scene, we created two movies: one following a CtF sequence and one following a FtC sequence. Each movie lasted 150 msec and was composed of the same scene filtered in the six different frequency bands (presented for 25 msec). Stimuli were displayed using E-prime software (E-prime Psychology Software Tools, Inc., Pittsburgh, PA) and back-projected on a translucent screen positioned at the rear of the magnet. Participants viewed this screen at a distance of about 222 cm via a mirror fixed on the head coil. We used a backward mask, built with 1/f white noise, to prevent retinal persistence of the scene.

A block design paradigm was used with CtF and FtC movies. The dynamic scene experiment consisted of four functional runs. Each functional scan lasted 5 min and was composed of eight 25-sec task blocks (four CtF blocks and four FtC blocks) including 10 dynamic scenes (five indoors and five outdoors), interspersed with four 25-sec blocks with a fixation dot in the center of the screen (Fixation condition) displayed against a gray background. Each scene was presented in the CtF and FtC movie conditions within a run, but a given scene appear only once in each sequence condition within a run. It should be noted that a block design paradigm did not allow us to

analyze individual response to trials as an event-related paradigm allowed, and it did not allow us to investigate the neural correlates of priming effect in our study. However, a block design paradigm increases the statistical power (Friston, Zarahn, Josephs, Henson, & Dale, 1999) and the relatively large BOLD signal change relative to baseline (Glover, 1999). The robustness of activations is particularly important for this study in which we used two experimental conditions involving the exact same images for the same total exposure duration but differing only by their relative temporal order. The choice of a block design paradigm was guided by an unpublished previous work in which the use of an event-related paradigm with CtF and FtC movies, as well as null events, did not reveal significant activation.

Each functional run consisted of 80 experimental trials. Each stimulus was displayed for 150 msec, followed by a mask for 30 msec and a fixation dot in the center of the screen. The average interval between the onsets of two successive stimuli was 2.5 sec. Participants had to give a categorical answer on movies ("indoors" or "outdoors") by pressing the corresponding key with the forefinger and the middle finger of their dominant hand. They were instructed to fixate the center of the screen (fixation dot) during the whole run and to respond as quickly and as accurately as possible by pressing one of two response buttons. Half the participants had to answer "indoor" with the forefinger and "outdoor" with the middle finger, whereas the other half had to answer "indoor" with the middle finger and "outdoor" with the forefinger. Response accuracy (ACC) and RTs (in milliseconds) were recorded.

Stimuli and Procedure in the Localizer Experiment

Following the main dynamic scene experiment, we performed a separate functional localizer experiment to localize the functional ROIs specifically involved in the processing of natural scenes. The localizer experiment was adapted from previous studies (Walther et al., 2009; Bar, Aminoff, & Ishai, 2008; Epstein & Kanwisher, 1998). Participants viewed scrambled and intact versions of grayscale photographs of scenes, faces, and common objects in a block design paradigm. Scene pictures used in the localizer experiment were not shown during the dynamic scene experiment. Stimuli were black and white photographs (256 grayscale), all sized 700 × 700 pixels (with a visual angle of 16.4 × 16.4 degrees), and RMS contrast of images was normalized. Scrambled pictures of these stimuli were created by dividing and randomizing intact picture scenes into 100 × 100 pixel squares. Participants viewed intact and scrambled versions of photographs of scenes, faces, and common objects in separate blocks of a block design paradigm. The localizer experiment consisted of two functional runs. Each functional scan lasted 3 min 30 sec and was composed of ten 15-sec task blocks (one block for each stimulus type) including 15 different photographs of the same type, interspersed with four 15-sec blocks with a fixation dot in the center of the screen displayed against a gray background. Participants performed a "1-back" repetition detection task. They were instructed to press a button whenever they saw two identical stimuli repeated. This task ensured that participants paid at least as much attention to unmotivating stimuli (e.g., scrambled images) as to more interesting stimuli (scenes, faces, and objects). Only two repetitions were presented per block. Each stimulus was presented for 300 msec, with a 700-msec ISI with a fixation dot in the center of the screen. The PPA, RSC, and OPA were identified in both hemispheres mainly by a [Scenes > Faces + Objects] contrast.

fMRI Acquisition

Experiments were performed on a whole-body 3T Philips scanner (Philips Medical Systems, Eugene, OR) at the University Hospital Center of Grenoble (France). For all functional dynamic scene and localizer scans, the manufacturer-provided gradient-echo/T2* weighted EPI method was used. Thirty-nine adjacent axial slices parallel to the bicommissural plane were acquired in interleaved mode. Slice thickness was 3.5 mm. The in-plane voxel size was 3 × 3 mm (216 × 216 mm field of view acquired with a 72 × 72 pixel data matrix; reconstructed with zero filling to 128 × 128 pixels). The main sequence parameters were repetition time = 2.5 sec, echo time = 30 msec, flip angle = 77°. To correct images for geometric distortions induced by local B0 inhomogeneity, a B0 field map was obtained from two gradient-echo data sets acquired with a standard 3-D FLASH sequence (Δecho time = 9.1 msec). The field map was subsequently used during data processing. Finally, a T1-weighted high-resolution 3-D anatomical volume was acquired by using a 3-D Modified Driven Equilibrium Fourier Transform sequence (field of view = 256 × 224 × 176 mm; resolution = 1.333 × 1.750 × 1.375 mm; acquisition matrix = 192 × 128 × 128 pixels; reconstruction matrix = 256 × 128 × 128 pixels).

Data Analysis

Data analysis was performed using the general linear model (Friston et al., 1995) for block designs in SPM8 (Wellcome Department of Imaging Neuroscience, London, UK, www.fil.ion.ucl.ac.uk/spm) implemented in MATLAB 7 (Mathworks, Inc., Sherborn, MA). Individual scans were realigned, time-corrected, normalized to the MNI space and spatially smoothed by an 8-mm FWHM Gaussian kernel. Time series for each voxel were high-pass filtered (1/128 Hz cutoff) to remove low-frequency noise and signal drift.

Individual ROIs were isolated based on the two localizer scans. The fMRI signal in the localizer runs was analyzed using single-participant general linear model. For each participant, five conditions of interest (scenes, faces, objects, scrambled scenes, and fixation) were modeled as five regressors convolved with a canonical

hemodynamic response function. Movement parameters derived from realignment corrections (three translations and three rotations) were also entered in the design matrix as additional factors of no interest. The areas responding to scenes were defined independently for each participant using two contrasts: [Scenes > Faces + Objects] and [Scenes > Scrambled scenes]. The contrast eliciting greater activity in ROIs for all participants was the one used to continue the analysis of data. Significant voxel clusters on individual *t* maps were identified using a false discovery correction at qFDR < 0.05 to control the overall false positive rate (Benjamini & Hochberg, 1995). Scene-selective voxel clusters were located in the PPA, RSC, and OPA. To facilitate comparisons with other studies, a transformation of MNI into Talairach and Tournoux (1988) coordinates was performed using the MNI2TAL function (created by Matthew Brett, available at www.mrc-cbu.cam.ac.uk/Imaging). These clusters were selected as ROIs for the analysis of data in the dynamic scene experiment in which categorization of CtF and FtC sequences was examined.

For the dynamic scene experiment, three conditions of interest (CtF, FtC, and Fixation) were modeled as three regressors convolved with a canonical hemodynamic response function. Note that the CtF and FtC experimental conditions involved the exact same images for the same total exposure duration (150 msec) but differed by their relative temporal order, which was too close to produce distinct hemodynamic responses to each of the six images during fMRI. RTs for each trial and movement parameters derived from realignment corrections (three translations and three rotations) were also entered in the design matrix as additional factors of no interest to account for RT-related variance and head motion, respectively. Parameter estimates (percent signal change relative to the global mean intensity of signal) of block responses were then extracted from the scene-selective ROIs for each participant. The average parameter of activity was calculated for CtF, FtC, and Fixation conditions. These values were submitted to a repeated-measure ANOVA with Conditions (CtF, FtC, and Fixation), Regions (PPA, RSC, and OPA), and Hemispheres (Left and Right) as within-subject factors.

RESULTS

Behavioral Results

Two 2 × 2 variance analyses (ANOVA) with Sequences (CtF and FtC) and Categories (outdoor and indoor) as within-subject factors were conducted on mean error rates (mER) and mean correct RTs (mRT). The ANOVA conducted on mER showed no effect of Sequences (CtF: 3.98 ± 4.68%; FtC: 5.43 ± 8.41%; $F(1, 16) = 1.48$, $p = .24$), but a main effect of Categories ($F(1, 15) = 9.91$, $p < .05$). Participants made more errors when categorizing indoor (5.47 ± 7.25%) than outdoor scenes (3.95 ± 6.32%). No interaction was observed between Sequences and Categories ($F(1, 15) < 1$). The ANOVA conducted on mRT showed that participants categorized CtF sequences more quickly than FtC sequences (CtF: 640 ± 128 msec; FtC: 657 ± 130 msec; $F(1, 15) = 7.50$, $p < .05$). There was no effect of Categories (outdoor: 650 ± 125 msec; indoor: 647 ± 132 msec; $F(1, 15) < 1$), and no interaction was observed between Sequences and Categories ($F(1, 15) < 1$; Figure 2).

fMRI Results

PPA, RSC, and OPA ROIs were defined in each individual based on the independent localizer experiment. This served as the structural constraint for the analysis of data

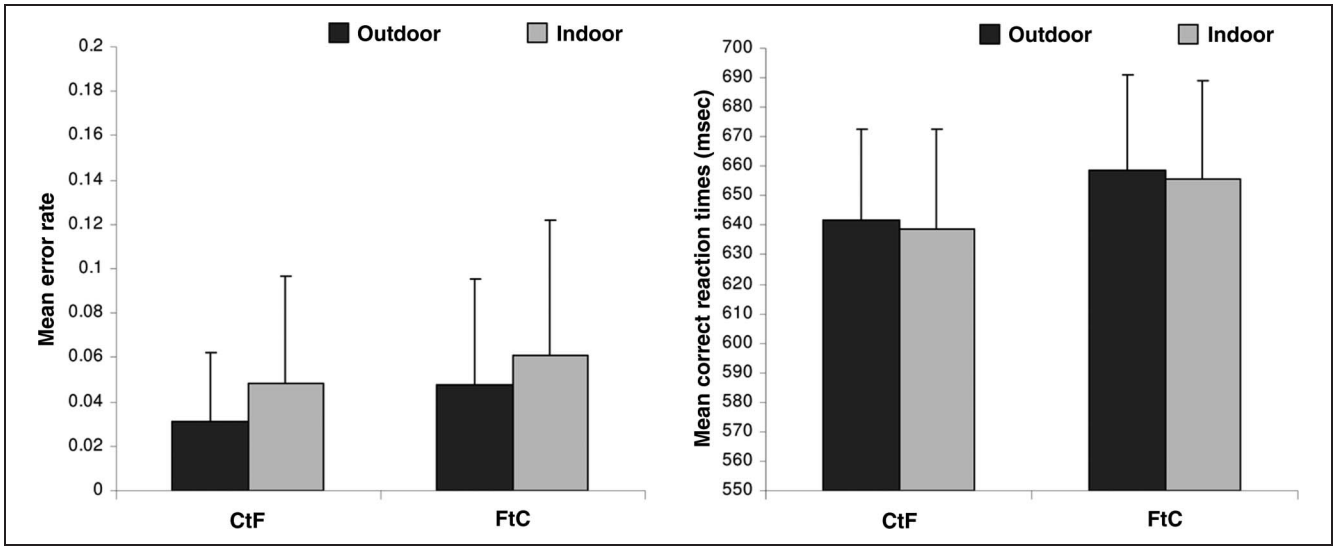

Figure 2. Mean error rates and mean correct RTs in milliseconds according to CtF and FtC sequences and scene categories (Outdoor and Indoor). Error bars correspond to standard errors.

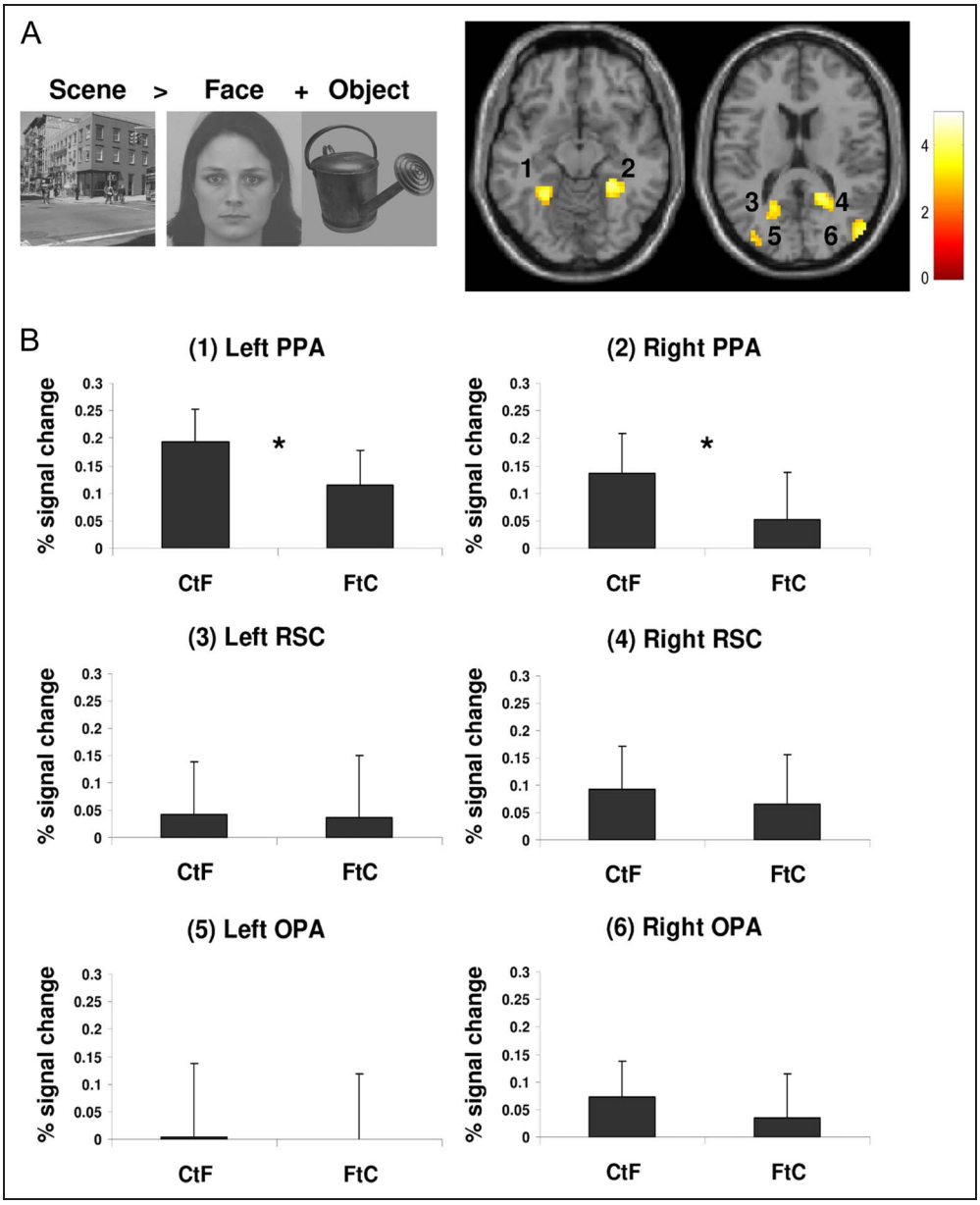

Figure 3. (A) Cerebral regions activated during the perception of scenes compared with faces and objects ([Scenes > Faces + Objects] contrast): (1) left PPA, (2) right PPA, (3) left restrosplenial cortex/RSC, (4) right RSC, (5) left OPA, (6) right OPA. ROIs are illustrated on a representative participant. (B) The ROIs were defined independently for each participant by contrasting intact scenes to meaningful intact other stimuli: [Scenes > Faces + Objects]. Signal changes relative to the global mean intensity of signal were then extracted from the scene-selective ROIs for each participant and each sequence (CtF and FtC). Graphics represent for each ROI the mean percentage of signal change for 16 participants. Error bars indicate 95% confidence intervals. * indicate significant differences.

in the dynamic scene experiment in which categorization of CtF and FtC sequence was examined. Using the contrast in which scenes were compared with faces and common objects, greater activity elicited by scenes was observed in different scene-selective regions. Scenes elicited stronger activation than faces and objects ([Scenes > Faces + Objects] contrast; Figure 3A) within the bilateral parahippocampal gyrus (including the PPA), cingulate gyrus (including the RSC), and the occipital gyrus (in the OPA) for all participants. Peak coordinates of ROIs were consistent with previous studies (Talairach coordinates are reported in Table 1). As the clusters were generally large involving several ROIs, small sphere ROIs (3-mm radius) were created at the peak of activation clusters. To ensure that the sphere only contains voxels that were truly activated, these spheres were masked

Table 1. Mean Coordinates of Scene-selective Regions Identified by the [Scenes > Faces + Objects] Contrast in the Localizer Experiment

	Mean x		*Mean y*		*Mean z*	
Right PPA	26	*±2.9*	−38	±3	−10	*±5.5*
Left PPA	−26	*±4.4*	−43	±4.7	−8.2	*±2.9*
Right RSC	21	*±4.2*	−53	±7.7	15	*±11*
Left RSC	−19	*±4.2*	−57	±7.4	14	*±11*
Right OPA	−29	*±9*	−78	±6.5	28	*±18*
Left OPA	−22	*±8*	−70	±4	27	*±15*

Coordinates (*x*, *y*, *z*) are indicated in the Talairach space.
Standard errors are shown in *italics*.

with the thresholded activation map (Poldrack, 2007). It should be noted that the contrast in which scenes were compared with scrambled images did not allow to identify all the ROIs for all participants. Parameter estimates (percent signal change relative to the global mean intensity of signal) of block responses were then extracted from these six sphere ROIs for each participant. The average parameter of activity was calculated for CtF, FtC, and Fixation condition. These values were submitted to a repeated-measures ANOVA with Conditions (CtF, FtC, and Fixation), Regions (PPA, RSC, and OPA), and Hemispheres (Left and Right) as within-subject factors.

The ANOVA revealed a significant interaction between the ROIs and the Sequences ($F(2, 30) = 3.35, p < .05$). Planned comparisons showed that the CtF periods elicited greater activation than FtC periods only within the PPA ($F(1, 15) = 12.52, p < .05$; RSC: $F(1, 15) < 1$; OPA: $F(1, 15) < 1$; Figure 3B). Hemispheres did not interact with Sequences in either region (PPA: $F(1, 15) < 1$; RSC: $F(1, 15) = 2.29, p = .15$; OPA: $F(1, 15) < 1$). The right and left PPA were more activated for the CtF than FtC (right PPA: $F(1, 15) = 13.27, p < .05$; left PPA: $F(1, 15) = 9.93, p < .05$). This was not the case for the right and left RSC (right RSC: $F(1, 15) = 1.50, p = .24$; left RSC: $F(1, 15) < 1$) or the right and left OPA (right OPA: $F(1, 15) = 1.48, p = .24$; left OPA: $F(1, 15) < 1$). Furthermore, CtF and FtC sequences showed greater activation than Fixation periods within the right PPA (CtF vs. Fixation: $F(1, 15) = 42.50, p < .05$; FtC vs. Fixation: $F(1, 15) = 23.40, p < .05$) and left PPA (CtF vs. Fixation: $F(1, 15) = 32.60, p < .05$; FtC vs. Fixation: $F(1, 15) = 31.63, p < .05$). This was not the case for the other scene-selective regions (all $F(1, 15) < 1$).

DISCUSSION

The present fMRI study investigated for the first time the temporal order of spatial frequency processing within scene-selective regions. For this purpose, we created two types of movies composed of the same scene filtered in six different spatial frequency bands but displayed in different orders, one following a CtF sequence and one following a FtC sequence. These sequences therefore imposed the order of spatial frequency processing. Although this procedure was obviously not physiological, it allowed us to experimentally mimic the sequential processing of spatial frequencies postulated by influential model of visual categorization (Peyrin et al., 2010; Hegde, 2008; Bar, 2003; Schyns & Oliva, 1994). Crucially, all visual information content was the same in both sequences, and only the order of spatial frequency images changed. None of the participants realized that we manipulated the spatial frequency order. However, when we used these stimuli in a previous behavioral study (Musel et al., 2012), we showed that CtF sequences were categorized more rapidly than FtC sequences in young adults. This provided new arguments in favor of a predominantly CtF categorization of natural scenes and a new experimental tool, which imposes a CtF processing and allows investigations of the neural substrates of CtF processing. In this study, we replicated the behavioral results of Musel et al. (2012). Participants categorized CtF more rapidly than FtC sequences. Furthermore, in terms of scene-selective regions, only the PPA showed stronger activation during the CtF than FtC categorization of filtered scenes. Sequences of filtered scenes (with blank screens occurring between scenes) were also used by Schettino, Loeys, Delplanque, and Pourtois (2011) in an evoked potential study to investigate the neural correlates of the accumulation of visual information during object recognition and their time course. For this purpose, the authors used sequences in which the first scene was always in LSF and the scene was gradually revealed in six successive images by progressively adding HSF information. In line with our results, these authors observed that activation in the parahippocampal cortex decreases when the spatial frequency content of scenes increases, suggesting that this region is sensitive to the primary processing of LSF information, even if this study did not investigate explicit CtF processing. CtF processing of faces in high-level visual cortex was recently the central focus of Goffaux et al. (2011), who showed an intriguing effect of spatial frequencies in a face-selective region, the FFA. By manipulating exposure duration and the spatial frequency content of faces, the authors showed that the FFA responded more strongly to LSF for short exposure durations of faces, whereas it responded more strongly to HSF for longer exposure durations. These results suggest that CtF processing is the predominant strategy in the most prominent regions of the ventral visual stream (inferotemporal cortex). Interestingly, the authors used scrambled faces (phase of the face images was scrambled in the Fourier domain via random permutation) as control stimuli, from which no face representation can be extracted. Therefore, the activation elicited by contrasting intact and scrambled faces are related to high-level representations of faces and not to low-level aspects of spatial frequency processing. Note, however, that participants had to perform an intact versus scrambled face categorization task. This task possibly engaged other cognitive demands and neural processes that the ones usually involved in face recognition. In our present dynamic scene experiment, we did not use scrambled stimuli because we were rather concerned by the effect of temporal order of different spatial frequency bands during an explicit categorization of dynamic scene and also because the CtF and FtC movies involved the exact same images and thus the same low-level visual information.

This study does not directly address the role played by the PPA in scene perception, but it suggests that the PPA would be selectively tuned to the processing of LSF before HSF information. Early studies (Epstein, 2008; Downing, Chan, Peelen, Dodds, & Kanwisher, 2006; Epstein et al., 1999; Epstein & Kanwisher, 1998) assumed

that the PPA responds more strongly to images of real-world scenes (such as cityscapes and landscapes) than to scrambled images and other meaningful visual stimuli (such as faces and objects). We obtained similar results in our localizer experiment. According to Epstein and colleagues (Epstein & Ward, 2010; Epstein, 2005, 2008; Epstein & Kanwisher, 1998), the PPA encodes the geometric structure (i.e., spatial layout) of scenes. The PPA is sensitive to real scenes without discrete objects (an empty room) and to complex meaningful scenes containing multiple objects (the same room furnished). However, it responds weakly to objects lacking a 3-D spatial context (objects in this room on a blank background). The perception of 3-D spatial information usually requires global perception of the scene and might therefore be preferentially performed on a first LSF-based analysis (Farell, Li, & McKee, 2004). Bar and colleagues (Bar, Aminoff, & Ishai, 2008; Bar, Aminoff, & Schacter, 2008; Aminoff et al., 2007; Bar, 2004; Bar & Aminoff, 2003) suggest that the PPA would be more particularly sensitive to visual contextual associations in scenes. These authors have shown in several studies that the PPA responds significantly more strongly to images with highly associative contextual objects both in spatial (e.g., a traffic light is strongly associated with a street context) and in nonspatial domains (e.g., a crown is strongly associated with royalty, but not with a specific place) compared with images of equal visual qualities but containing objects with weak contextual associations (e.g., a cell phone, which is not strongly associated with a single context). The PPA should therefore be viewed not as being exclusively dedicated to the analysis of place, but rather as more generally mediating spatial contextual associations (Aminoff et al., 2007). Computational data again suggest that the early processing of LSF information is sufficient to extract the context of a scene (Torralba & Oliva, 2003; Oliva & Torralba, 2001). Therefore, this study would provide support for both the spatial layout and contextual hypothesis of PPA function.

PPA sensitivity to LSF (when displayed before HSF) could be linked to previous fMRI results showing greater activation of the parahippocampal gyrus for LSF than HSF scene categorization (Peyrin, Baciu, Segebarth, & Marendaz, 2004). This result differed from recent fMRI studies conducted by Rajimehr, Devaney, Bilenko, Young, and Tootell (2011), who showed that the PPA responds preferentially to spatial discontinuities in geometrical shapes (such as boundary edges and corners) and HSF in natural images. The discrepancy between our results and those of Rajimehr et al. (2011) could be because of several experimental parameters. First, Goffaux et al. (2011) clearly demonstrated that the presentation time of stimuli could drastically modify the level of activity of category-specific regions. Face-specific regions were more strongly activated by LSF faces in the early stages of visual processing (up to 75 msec of face exposure), and this activation decreased as a function of exposure duration (mostly to 150 msec). In contrast, activation in response to HSF faces increased over time within these regions. In Rajimehr et al. (2011), the presentation time of stimuli (checkerboard, face and scene) was always beyond 500 msec. Such a long presentation time may favor the analysis of HSF information and the stronger activity of the PPA for HSF stimuli. Second, like psychophysical studies, which revealed a certain degree of flexibility in the extraction and analysis of spatial frequencies depending on task demands (Rotshtein, Schofield, Funes, & Humphreys, 2010; Oliva & Schyns, 1997; Schyns & Oliva, 1997), we believe that the sensitivity of the PPA to a particular bandwidth may also depend on the demands of the visual task. Thus, the PPA would be more activated for LSF when visual tasks require a global visual processing and for HSF when visual tasks require finer and more detailed visual processing. This assumption could explain why Zeidman, Mullally, Schwarzkopf, and Maguire (2012) observed stronger activation of the PPA in response to HSF than LSF 3-D spaces. In this study, the spaces were depicted by positioning small white dots on a black background following an exponential distribution and then either displayed in HSF (intact dot pattern) or filtered in LSF. Participants had to detect whether a small proportion of dots disappeared irrespective of spatial frequency content. This type of detection task involves a fine analysis of visual information, which may favor the analysis of HSF spaces by the PPA. In the same way, LSF information could be sufficient to allow rapid discrimination and categorization of indoor and outdoor scenes by the PPA. It should however be noted that, in this study, we used dynamic scenes containing both LSF and HSF information displayed in different relative orders. It is therefore possible that each image that made up the movies may have induced differential activation within the PPA. Future studies manipulating explicitly spatial frequencies in different stimuli are needed to fully investigate the spatial frequency selectivity of the PPA during scene categorization.

Concerning the other scene-selective regions, although both RSC and OPA were more activated by the passive viewing of scenes compared with faces and objects in the Localizer experiment, they were not more activated by the categorization of CtF and FtC sequences compared with fixation periods in the dynamic scene experiment. These results suggest that these two regions were not involved in the processing of spatial frequencies or the categorization process. However, according to evidence previously provided by Dilks et al. (2013), the OPA was shown to be explicitly involved in the categorization of scenes such as beaches, forests, cities, or kitchens (compared with the categorization of objects as a camera, chair, car, or shoes). The authors even suggest that the OPA was preferentially involved in the processing of low-level visual features in scenes such as spatial frequencies or spatial envelope properties (i.e., the representation of the shape of a scene based on its degree of naturalness, openness, roughness, expansion, and ruggedness of the scene; Oliva & Torralba, 2001). Our study did not reveal significant activation of the OPA for either CtF or FtC sequences relative to fixation periods, excluding the potential specialization

of the OPA for spatial frequency processing during the categorization. It should be noted that Dilks et al. (2013) used a task involving the perception of scenes and objects whose spatial envelope properties and spatial representations differed greatly. This suggests that the OPA might be more involved in the spatial discrimination between scenes and other categories through the analysis of the spatial envelope than in the analysis of spatial frequency content used in the categorization process. In keeping with this assumption, the two categories used in this study (outdoor and indoor categories) were composed exclusively of man-made environments. Outdoor scenes consisted mainly of front views of buildings or close-up views of cities. This drastically reduced the degree of expansion that usually characterizes outdoors and allowed them to be distinguished from indoor scenes. Furthermore, both categories had similar amplitude spectra and equivalent levels of visual cluttering to prevent task performance being based on these low-level visual features. The two categories therefore contained large objects with flat surfaces (either a kitchen cupboard door for the indoor category or shutters on a building for the outdoor category), although outdoor scenes are usually more highly textured and made up of small elements. Given the visual properties of our scenes and based on previous computational works (Oliva & Torralba, 2001), we assumed similarity of spatial envelope in both categories. Finally, mean luminance and contrast were normalized for the six images that composed the movie. As a result, the absence of activation in the OPA during our categorization task may have been because of an irrelevant analysis of the amplitude spectrum (i.e., the distribution of energy in spatial frequencies and dominant orientations), the spatial envelope properties or the contrast to perform the visual task. Alternatively, the OPA could be selective to a particular spatial frequency range within the CtF and FtC sequences, but the selective activation would be disturbed by the other spatial frequency ranges.

The RSC, region of the posterior cingulate cortex, is mainly involved in spatial memory and navigation (Kravitz, Saleem, Baker, & Mishkin, 2011; Vann, Aggleton, & Maguire, 2009; Aggleton & Vann, 2004; Harker & Whishaw, 2004). These interpretations are mainly based on the extensive connections linking the RSC and the visual and hippocampal regions and the presence of cells responding to specific directions of the head (Vann et al., 2009; Cho & Sharp, 2001). However, in a number of recent studies, involvement of this region was also evidenced during the perception of natural scenes (Park, Brady, Greene, & Oliva, 2011; Park & Chun, 2009; Walther et al., 2009; Bar, Aminoff, & Ishai, 2008; Sung, Kamba, & Ogawa, 2008; Aminoff et al., 2007; Epstein & Higgins, 2007) in the coding of "spatial layout" and location (Vass & Epstein, 2013; Epstein & Higgins, 2007) or contextual associations (Bar, Aminoff, & Ishai, 2008; Aminoff et al., 2007). More particularly, the RSC may support an expansive spatial representation of the scenes, which allows the current view of the scene to be replaced into a larger spatial environment extending beyond the borders of the image (Kravitz et al., 2011; Park et al., 2011; Epstein & Higgins, 2007). Using three pictures from panoramic scenes representing different views of the same scene in an fMRI paradigm of adaptation, Park et al. (2011) found a significant attenuation of activation in the RSC for repeated panoramic scenes suggesting that different viewpoints of the same scene are integrated into a single visual context. It has also been demonstrated that the RSC responds more strongly to the identification of specific locations around a familiar environment than to general judgments about the scene category (Epstein & Higgins, 2007). Consequently, it could explain why the RSC was not activated during our categorization task.

To sum up, we have demonstrated here that the CtF strategy is a plausible modus operandi in the PPA. This result provides critical support for influential models of visual perception based mainly on a spatial frequency analysis, which follows a CtF strategy (Peyrin et al., 2010; Hegde, 2008; Bar, 2003; Schyns & Oliva, 1994). It also provides additional data for the development of a neurally grounded model of scene perception.

Acknowledgments

This work was supported by the RECOR ANR grant (ANR-12-JHS2-0002-01 RECOR). Benoit Musel and Louise Kauffmann were supported by Région Rhône-Alpes (ARC2 and Cible grants). The authors warmly thank the "Délégation à la Recherche Clinique et à l'Innovation" of the University Hospital of Grenoble for sponsoring. We thank Catherine Dal Molin for the English revision of the manuscript.

Reprint requests should be sent to Carole Peyrin, CNRS UMR 5105, Université Pierre Mendès France, BP 47, 38040 Grenoble Cedex 09, France, or via e-mail: carole.peyrin@upmf-grenoble.fr.

Note

1. A control behavioral study was conducted on 17 participants to verify that the six spatial frequency band images were correctly categorized. Whatever the spatial frequency band, the mean error rate was low (mER for the pass-band centered on 24 cycles/image [1 cycle/degree]: 6.03 ± 1.25%; on 34 cycles/image [1.42 cycles/degree]: 4.59 ± 0.95%; on 49 cycles/image [2.04 cycles/degree]: 5.59 ± 1.02%; on 71 cycles/image [2.96 cycles/degree]: 5.15 ± 1.36%; on 101 cycles/image [4.21 cycles/degree]: 4.71 ± 1.02%; on 144 cycles/image [6 cycles/degree]: 5.29 ± 1.21%) and there was no effect of the spatial frequency band, $F(5, 80) < 1$.

REFERENCES

Aggleton, J. P., & Vann, S. D. (2004). Testing the importance of the retrosplenial navigation system: Lesion size but not strain matters: A reply to Harker and Whishaw. *Neuroscience and Biobehavioral Reviews, 28,* 525–531.

Aminoff, E., Gronau, N., & Bar, M. (2007). The parahippocampal cortex mediates spatial and nonspatial associations. *Cerebral Cortex, 17,* 1493–1503.

Bar, M. (2003). A cortical mechanism for triggering top–down facilitation in visual object recognition. *Journal of Cognitive Neuroscience, 15,* 600–609.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5,* 617–629.
Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences, 11,* 280–289.
Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron, 38,* 347–358.
Bar, M., Aminoff, E., & Ishai, A. (2008). Famous faces activate contextual associations in the parahippocampal cortex. *Cerebral Cortex, 18,* 1233–1238.
Bar, M., Aminoff, E., & Schacter, D. L. (2008). Scenes unseen: The parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. *Journal of Neuroscience, 28,* 8539–8544.
Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B, 57,* 289–300.
Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews, 36,* 96–107.
Cho, J., & Sharp, P. E. (2001). Head direction, place, and movement correlates for cells in the rat retrosplenial cortex. *Behavioral Neuroscience, 115,* 3–25.
De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research, 22,* 545–559.
De Valois, R. L., Yund, E. W., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research, 22,* 531–544.
Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *Journal of Neuroscience, 33,* 1331–1336a.
Downing, P. E., Chan, A. W., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain specificity in visual cortex. *Cerebral Cortex, 16,* 1453–1461.
Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: Recognition, navigation, or encoding? *Neuron, 23,* 115–125.
Epstein, R. A. (2005). The cortical basis of visual scene processing. *Visual Cognition, 12,* 954–978.
Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences, 12,* 388–396.
Epstein, R. A., Graham, K. S., & Downing, P. E. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron, 37,* 865–876.
Epstein, R. A., & Higgins, J. S. (2007). Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. *Cerebral Cortex, 17,* 1680–1693.
Epstein, R. A., Higgins, J. S., Jablonski, K., & Feiler, A. M. (2007). Visual scene processing in familiar and unfamiliar environments. *Journal of Neurophysiology, 97,* 3670–3683.
Epstein, R. A., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392,* 598–601.
Epstein, R. A., & Ward, E. J. (2010). How reliable are visual context effects in the parahippocampal place area? *Cerebral Cortex, 20,* 294–303.
Farell, B., Li, S., & McKee, S. P. (2004). Coarse scales, fine scales, and their interactions in stereo vision. *Journal of Vision, 4,* 488–499.
Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A, 4,* 2379–2394.
Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R. S., et al. (1995). Analysis of fMRI time-series revisited. *Neuroimage, 2,* 45–53.
Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N., & Dale, A. M. (1999). Stochastic designs in event-related fMRI. *Neuroimage, 10,* 607–619.
Ginsburg, A. P. (1986). Spatial filtering and visual form perception. In K. Boff, L. Kauman, & J. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2, Chap. 34, pp. 1–41). New York: Wiley.
Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage, 9,* 416–429.
Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2011). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cerebral Cortex, 21,* 467–476.
Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research, 41,* 1409–1422.
Guyader, N., Chauvin, A., Peyrin, C., Herault, J., & Marendaz, C. (2004). Image phase or amplitude? Rapid scene categorization is an amplitude-based process. *Comptes Rendus Biologies, 327,* 313–318.
Harker, K. T., & Whishaw, I. Q. (2004). A reaffirmation of the retrosplenial contribution to rodent navigation: Reviewing the influences of lesion, strain, and task. *Neuroscience and Biobehavioral Reviews, 28,* 485–496.
Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293,* 2425–2430.
Hegde, J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Progress in Neurobiology, 84,* 405–439.
Hughes, H. C., Nozawa, G., & Kitterle, F. L. (1996). Global precedence, spatial frequency channels, and the statistic of the natural image. *Journal of Cognitive Neuroscience, 8,* 197–230.
Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302–4311.
Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience, 12,* 217–230.
Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., & Malach, R. (2001). A hierarchical axis of object processing stages in the human visual cortex. *Cerebral Cortex, 11,* 287–297.
Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences, U.S.A., 92,* 8135–8139.
Musel, B., Chauvin, A., Guyader, N., Chokron, S., & Peyrin, C. (2012). Is coarse-to-fine strategy sensitive to normal aging? *PLoS One, 7,* e38493.
Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology, 34,* 72–107.
Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal in Computer Vision, 42,* 145–175.
Park, S., Brady, T. F., Greene, M. R., & Oliva, A. (2011). Disentangling scene content from spatial boundary: Complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *Journal of Neuroscience, 31,* 1333–1340.
Park, S., & Chun, M. M. (2009). Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *Neuroimage, 47,* 1747–1756.

Peyrin, C., Baciu, M., Segebarth, C., & Marendaz, C. (2004). Cerebral regions and hemispheric specialization for processing spatial frequencies during natural scene recognition. An event-related fMRI study. *Neuroimage, 23,* 698–707.

Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., et al. (2010). The neural substrates and timing of top–down processes during coarse-to-fine categorization of visual scenes: A combined fMRI and ERP study. *Journal of Cognitive Neuroscience, 22,* 2768–2780.

Poggio, G. F. (1972). Spatial properties of neurons in striate cortex of unanesthetized macaque monkey. *Investigative Ophthalmology, 11,* 368–377.

Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Scan, 2,* 67–70.

Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C., & Tootell, R. B. (2011). The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. *PLoS Biology, 9,* e1000608.

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision, 7,* 17.11–17.22.

Rotshtein, P., Schofield, A., Funes, M. J., & Humphreys, G. W. (2010). Effects of spatial frequency bands on perceptual decision: It is not the stimuli but the comparison. *Journal of Vision, 10,* 25.

Schettino, A., Loeys, T., Delplanque, S., & Pourtois, G. (2011). Brain dynamics of upstream perceptual processes leading to visual object recognition: A high density ERP topographic mapping study. *Neuroimage, 55,* 1227–1241.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependant scene recognition. *American Psychological Society, 5,* 195–200.

Schyns, P. G., & Oliva, A. (1997). Flexible, diagnosticity-driven, rather than fixed, perceptually determined scale selection in scene and face recognition. *Perception, 26,* 1027–1038.

Shams, L., & von der Malsburg, C. (2002). The role of complex cells in object recognition. *Vision Research, 42,* 2547–2554.

Spiridon, M., & Kanwisher, N. (2002). How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron, 35,* 1157–1165.

Sung, Y. W., Kamba, M., & Ogawa, S. (2008). Building-specific categorical processing in the retrosplenial cortex. *Brain Research, 1234,* 87–93.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotactic atlas of the human brain*. Thieme Medical, New York.

Tolhurst, D. J., Tadmor, Y., & Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic and Physiological Optics, 12,* 229–232.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network, 14,* 391–412.

Vann, S. D., Aggleton, J. P., & Maguire, E. A. (2009). What does the retrosplenial cortex do? *Nature Reviews Neuroscience, 10,* 792–802.

Vass, L. K., & Epstein, R. A. (2013). Abstract representations of location and facing direction in the human brain. *Journal of Neuroscience, 33,* 6133–6142.

Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural scene categories revealed in distributed patterns of activity in the human brain. *Journal of Neuroscience, 29,* 10573–10581.

Willenbockel, V., Fiset, D., Chauvin, A., Blais, C., Arguin, M., Tanaka, J., et al. (2010). Does face inversion change spatial frequency tuning? *Journal of Experimental Psychology: Human Perception and Performance, 36,* 122–135.

Zeidman, P., Mullally, S. L., Schwarzkopf, D. S., & Maguire, E. A. (2012). Exploring the parahippocampal cortex response to high and low spatial frequency spaces. *NeuroReport, 23,* 503–507.