The Neural Substrates and Timing of Top–Down Processes during Coarse-to-Fine Categorization of Visual Scenes: A Combined fMRI and ERP Study

Carole Peyrin[1,2,3], Christoph M. Michel[3], Sophie Schwartz[3], Gregor Thut[3], Mohamed Seghier[3], Theodor Landis[4], Christian Marendaz[1], and Patrik Vuilleumier[3]

Abstract

■ Spatial frequencies in an image influence visual analysis across a distributed, hierarchically organized brain network. Low spatial frequency (LSF) information may rapidly reach high-order areas to allow an initial coarse parsing of the visual scene, which could then be "retroinjected" through feedback into lower level visual areas to guide finer analysis on the basis of high spatial frequency (HSF). To test this "coarse-to-fine" processing scheme and to identify its neural substrates in the human brain, we presented se-

quences of two spatial-frequency-filtered scenes in rapid succession (LSF followed by HSF or vice versa) during fMRI and ERPs in the same participants. We show that for low-to-high sequences (but not for high-to-low sequences), LSF produces a first increase of activity in prefrontal and temporo-parietal areas, followed by enhanced responses to HSF in primary visual cortex. This pattern is consistent with retroactive influences on low-level areas that process HSF after initial activation of higher order areas by LSF. ■

INTRODUCTION

Natural scenes such as landscapes or indoor environments usually contain many visual objects disposed in complex three-dimensional layouts. It remains unresolved how the visual system can rapidly organize information distributed across the visual field to allow more detailed inspection and accurate identification of selected objects in the scene. Influential theories of visual recognition have speculated that spatial frequency (SF) content may impose a specific temporal hierarchy in the processing of visual inputs (Hegde, 2008; Bar, 2003; Bullier, 2001). According to these models, visual analysis may start with a parallel extraction of different elementary attributes at different SF, but with a predominant "coarse-to-fine" (low-to-high SF) sequence that privileges low spatial frequencies (LSFs) at initial stages of visual processing and high spatial frequencies (HSFs) at later stages. The LSF in a scene, conveyed by fast magnocellular visual channels, might thus activate visual pathways and then reach high-order areas in the dorsal stream (parietal and frontal) more rapidly than HSF, allowing an initial perceptual parsing of the visual inputs prior to their complete propagation along the ventral stream (inferotemporal) that ultimately mediates object recognition (Bullier, 2001). This initial low-pass vi-

sual analysis might serve to refine the subsequent processing of HSF conveyed more slowly by parvocellular visual channels to the ventral stream.

In the present study, we combined fMRI and ERPs to assess the neural substrates underlying such "coarse-to-fine" processing sequence during perception and categorization of complex visual scenes. To date, empirical evidence in support of "coarse-to-fine" processing in human vision mostly comes from psychophysical studies. Early work using gratings of different SF as stimuli (Breitmeyer, 1975) showed that LSF channels have shorter latencies and shorter integration time, relative to HSF, suggesting that LSFs are transmitted faster than HSF through the visual system. Other studies using hierarchical stimuli such as global forms composed of several local elements (Navon, 1977) demonstrated faster identification of global than local shapes, a finding that could be also attributed to a general principle of "coarse-to-fine" SF analysis on the basis of the assumption that global information is conveyed by LSF but local information by HSF (Lamb & Yund, 1993; Badcock, Whitworth, Badcock, & Lovegrove, 1990; Schulman, Sulivan, Gisch, & Sadoka, 1986). Importantly, "coarse-to-fine" analysis was found with more ecological visual stimuli, such as natural scene images (Schyns & Oliva, 1994, 1997; Parker, Lishman, & Hughes, 1992). For example, using "hybrid" stimuli made of two superimposed images belonging to different semantic categories and containing different SF (e.g., a highway scene in LSF superimposed on a city scene in HSF), Schyns and Oliva (1994) have shown that perception is

[1]Université Pierre Mendès-France, Grenoble, France, [2]Centre National de la Recherche Scientifique (CNRS), Grenoble, France, [3]University Medical Center, Geneva, Switzerland, [4]University Hospital of Geneva, Switzerland

dominated by LSF information when presentation time is brief (30 msec) but by HSF information when presentation time is longer (150 msec). Furthermore, when two successive hybrids are presented with a low-to-high (LtH) sequence for one scene and an inverse high-to-low (HtL) sequence for the other, perception is dominated by the scene shown in the LtH sequence.

How and where in the brain low and high SF information is differentially analyzed and eventually merged during visual processing remain unsettled questions. Traditional models have generally assumed that different visual cues are combined at successive stages along the cortical hierarchy (Riesenhuber & Poggio, 1999; Biederman, 1995), suggesting that low and high SF might converge only in higher level visual areas within the inferior temporal cortex (such as the fusiform or parahippocampal cortex (Rotshtein, Vuilleumier, Winston, Driver, & Dolan, 2007; Bar et al., 2001). On the other hand, on the basis of neurophysiological recordings in nonhuman primates (Hupe et al., 2001), Bullier (2001) proposed that rapid LSF analysis, predominantly carried out in the dorsal visual stream, might be "retro-injected" through feedback signals into low-level areas (e.g., primary visual cortex, V1) where this could act to influence subsequent HSF analysis and guide further processing through the ventral visual stream. V1 might therefore serve as an "active blackboard" integrating computations accomplished by higher order cortical areas. However, to date, the neural architecture and the temporal dynamics of such top–down mechanisms have never been systematically investigated with a direct test of the preferential LtH processing sequence during visual scene recognition in humans. Another dynamic model of vision (Bar, 2003) also suggested that LSF may rapidly reach prefrontal areas, allowing a first "interpretation" of visual inputs that is then fed back to ongoing bottom–up analysis

in temporal cortex and was recently tested by a combined (MEG) magneto-encephalographic and fMRI study (Bar et al., 2006). These authors demonstrated earlier activations in prefrontal than temporal cortex during recognition of single objects, and these differential responses were driven by LSF in the image, consistent with a top–down mechanism initiated in the pFC. However, this pioneer study did not directly address the crucial issue of delayed "retro-injection" and SF integration in V1 (Bullier, 2001). Here, we specifically investigated the neural architecture and sequence of SF integration by directly manipulating the temporal order of SF inputs.

In our first experiment, we used fMRI to identify brain regions preferentially activated by a "coarse-to-fine" analysis of visual scenes and could delineate the neural effects of any "retro-injection" of LSF (Bullier, 2001) by comparing cortical activity during conditions imposing a "coarse-to-fine" relative to a reverse "fine-to-coarse" sequence with the same visual inputs. To constrain SF processing according to these different sequences, we presented brief displays of two successive images, each with opposite SF contents (either LSF or HSF; see Figure 1), therefore allowing us to experimentally "decompose" the visual inputs in either an LtH or an HtL sequence of SF. These sequences could thus experimentally "mimic" and impose a "coarse-to-fine" versus "fine-to-coarse" perceptual analysis using a controlled breakdown of SF information. Each scene in a sequence could belong to one of three categories (city, beach, or indoor). Half of the sequences displayed two scenes from the same category, whereas the other half displayed two scenes from different categories. The participants had to judge whether the two successive scenes belonged to the same category.

In a second experiment, we examined the temporal dynamics of cortical activations elicited by LtH relative to HtL

Figure 1. Examples of stimuli used in the scene-matching task. Each image was a natural scene (city, beach, or indoor) filtered in either LSF (<4 cycles/degree) or HSF (>6 cycles/degree). Each experimental trial consisted of a brief sequence during which two SF-filtered images were displayed for 100 msec in rapid succession with either an LSF image followed by an HSF image (LtH sequence) or an HSF image followed by an LSF image (HtL sequence). Half of the sequences displayed two scenes form the same category, whereas the other half displayed two scenes

from different categories. Participants had to judge whether the two scenes were or were not from the same category. During fMRI, the first scene was displayed in the CVF, LVF, or RVF. Because fMRI results showed no effect of the first image position, the first image was always presented in the central field during ERP recordings.
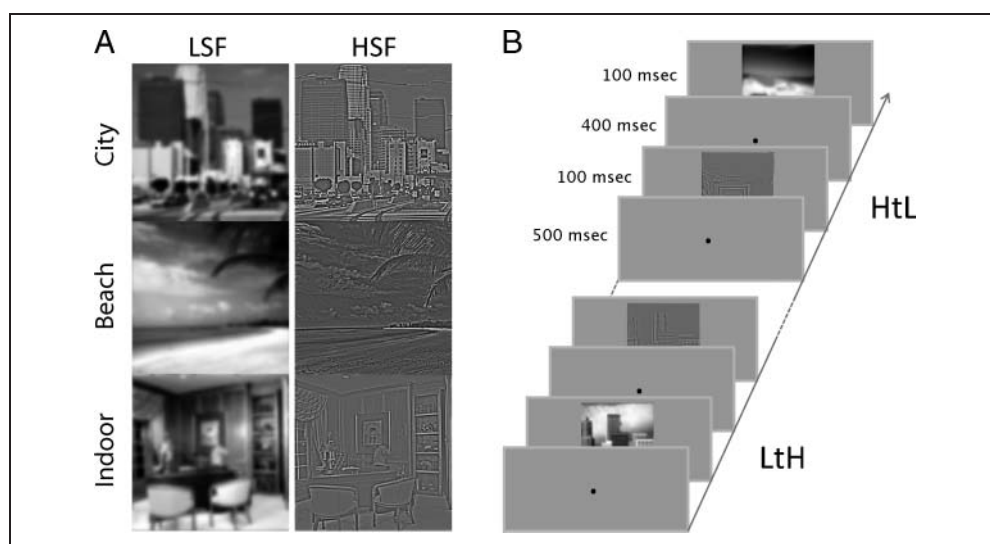
sequences by recording ERPs in the same participants during similar conditions. We performed detailed analysis of scalp topography (Michel et al., 2004) and estimation of sources (Grave de Peralta Menendez, Murray, Michel, Martuzzi, & Gonzalez Andino, 2004; Grave de Peralta Menendez, Gonzalez Andino, Lantz, Michel, & Landis, 2001) for ERPs across the different experimental conditions. Results revealed selective activations during LtH processing and converged with fMRI to highlight a distinct time course of neural responses to LSF and HSF cues across fronto-parietal and early visual areas. Our findings provide novel evidence in support of sequential SF analysis during visual recognition (Bar, 2003; Bullier, 2001) by showing that initial LSF extraction during LtH sequences may first enhance the activation of high-order areas and then retroactively modulate the subsequent HSF analysis in low-level areas as early as in V1.

METHODS

Participants

Eleven healthy male (age range = 20–38 years, mean ± *SD* age = 26.5 ± 5.5 years) volunteers participated to both fMRI and ERP sessions (approximately two months apart). All subjects were right-handed as assessed by the Edinburgh Inventory (Oldfield, 1971). They had normal or corrected-to-normal vision and no history of neurological disorders. They gave informed consent for the study according to the ethical regulation of the Geneva University Hospital.

Stimuli and Procedure

Stimuli were 54 black-and-white photographs (256 gray scales) of natural scenes classified in three distinct categories (18 cities, 18 beaches, and 18 indoors; all 4° of visual angle). For each scene, two types of images were created, one LSF and one HSF (Figure 1), using the image processing toolbox on MATLAB (Mathworks Inc., Sherborn, MA). These were obtained by multiplying the Fourier transform of the original images with Gaussian filters. The standard deviation of Gaussian filters is a function of the SF cutoff for a standard attenuation of 3 dB. We removed the SF content above 4 cycles/degree of visual angle (i.e., low-pass cutoff of 16 cycles per image) for LSF stimuli and below 6 cycles/degree (i.e., high-pass cutoff of 24 cycles per image) for HSF stimuli. The average energy level for LSF and HSF stimuli was equalized for each scene.[1] Overall, averaged stimuli luminance did not differ between LSF and HSF stimuli (118 and 120, respectively, on a 256 gray level scale), $F(1, 51) < 1$, or between cities, beaches, and indoors (116, 126, and 117, respectively), $F(2, 51) = 1.76$, $p = .18$.

Each experimental trial consisted of a brief sequence during which two SF-filtered images were displayed in rapid succession, with either an LSF image followed by an HSF image (LtH sequence) or an HSF image followed by an LSF image (HtL sequence). The two successive images were from the same category in half of the trials (city, beach, and indoor) and from different categories in the other half. Each trial began with a central fixation point presented for 500 msec, immediately followed by the first filtered image presented for 100 msec. Then the central fixation point reappeared for 400 msec, followed by the second filtered image, again presented for 100 msec. The average intertrial interval was 2 sec. Importantly, the two scenes in each sequence were presented with an interimage interval long enough (400 msec) to allow a complete processing of the first image (irrespective of SF content) and to avoid an overlap of brain responses to the two images during ERP recordings, although this was still short enough to be pooled into a single event during fMRI. These two conditions involved the exact same images for the same total exposure duration (600 msec) but differed by their relative temporal order (LtH or HtL), which was too close to produce distinct hemodynamic responses to each of the two images (LSF and HSF) during fMRI, yet sufficiently separated to record distinct evoked potentials in EEG. Participants were asked to decide whether the two scenes were from the same category (city, beach, or indoor). They were instructed to fixate the center of the screen during the whole trial and to respond as quickly and as accurately as possible by pressing one of two response buttons. Half of the subjects responded with their right index finger for "same" and right middle finger for "different" and vice versa for the other half of the subjects. This scene-matching task ensured that participants attended to the whole SF sequence and did not concentrate on a particular SF range only. Furthermore, because the magnocellular pathway (LSF information) originates in retina cells conveying predominant information from extrafoveal vision whereas the parvocellular pathway (HSF information) predominates for foveal vision (Martin & Grunert, 2003), leading to greater sensitivity to HSF in the fovea but greater sensitivity to LSF in the periphery (DeValois & DeValois, 1988; Robson & Graham, 1981), we also aimed at comparing any effect of presenting the first image in each sequence at peripheral versus central positions in the visual field. Therefore, participants performed fMRI sessions in which the first scene was either displayed in the center of the screen (CVF) or lateralized randomly in the left (LVF) or right (RVF) visual field. When lateralized, the inner and the outer edges of scenes subtended a visual angle of 2° and 6° off center along the horizontal axis. The central fixation point remained visible throughout the trials to keep gaze direction directed centrally during the whole sequence. Thus, by manipulating the location of the first scene, we could assess whether top–down processes during LtH sequence processing might depend on the parafoveal presentation of LSF information. Note that such sequences in which the two successive scenes did not overlap at the same location also minimized the use of retinotopic cues for the image matching task (see Figure 1). However, because fMRI results showed no

differential activation between LtH and HtL sequences according to the position of the first image (and to minimize experiment duration with sufficient trials in all conditions), the presentation of the first scene was restricted to the center of the screen during ERP study.

MR Acquisition and Analysis

Stimuli were displayed using E-prime software (E-prime Psychology Software Tools Inc., Pittsburgh, PA) and projected onto a mirror mounted on the MRI head coil (visual angle ~15.2° × 11.4°). Each participant performed six sessions (two in which the first scene was displayed in the CVF and four in which the first scene was lateralized to either the LVF or the RVF). Each session consisted of 72 trials. This resulted in 108 trials for each SF sequence condition (LtH and HtL) of particular interest for the purpose of the current study. Trial onset was jittered with respect to scan repetition time (repetition time = 2.5 sec) to allow for better sampling of the hemodynamic response across the whole brain (Josephs & Henson, 1999). For each session, 18 null trials were also randomly intermixed with image sequence to provide an appropriate baseline measure (Friston, Zarahn, Josephs, Henson, & Dale, 1999). The order of the experimental trials was pseudorandom (i.e., no more than three consecutive trials of the same sequence type or visual hemifield for lateralized sessions), and the order of the experimental sessions was counterbalanced across participants. Eye position was recorded continuously during fMRI using an infrared eye tracker (ASL Model LRO 504, Applied Science Laboratories, Bedford, MA) in eight participants (no recordings were obtained in three others due to technical problems). Eye-tracking data allowed us to test for any systematic eye movements along the horizontal axis as a function of the first image location in SF sequences. To compute the mean horizontal eye position for each participant in each condition, we first removed blinks producing a loss of input data and then epoched the eye coordinates during the first image duration (100 msec), separately for each location in visual field (CVF, LVF, and RVF). Mean horizontal eye positions were analyzed by standard ANOVA for repeated measures with the factor Visual Field condition.

Whole-brain fMRI was performed using EPI on a 1.5-T whole-body INTERA system (Philips Medical Systems, Eugene, OR), equipped with a standard head coil configuration. The imaging volume was oriented parallel to the bicommissural (AC–PC) plane. Functional volumes composed of thirty 4-mm adjacent, axial slices were acquired using a gradient-echo-planar T2*-weighted sequence (repetition time = 2.5 sec, echo time = 40 msec, flip angle = 80°, matrix size = 128 × 128, field of view = 250 mm, in-plane voxel size = 2 × 2 mm). After discarding the four initial scans, a total of 180 scans were acquired for each participant in each experimental session (7.5 min each). Subsequent to the functional scans, a T1-weighted high-

resolution three-dimensional volume (130 adjacent, axial slices, 1.25 mm thickness; in-plane voxel size = 1 × 1 mm) was acquired. Data analysis was performed using the general linear model (Friston et al., 1999) for event-related designs in SPM2 (Wellcome Department of Imaging Neuroscience, London, UK, www.fil.ion.ucl.ac.uk/spm) implemented in MATLAB. Individual scans were realigned, time corrected, normalized to the Montreal Neurological Institute (MNI) space, and spatially smoothed by an 8-mm FWHM Gaussian kernel. Time series for each voxel was high-pass filtered (1/128 Hz cutoff) to remove low-frequency noise and signal drift.

Each trial sequence (for each condition) was modeled by convolving a delta function at the sequence onset with a canonical hemodynamic response function. Only sequences leading to correct responses were included. Six conditions of interest (LtH-CVF, LtH-LVF, LtH-RVF, HtL-CVF, HtL-LVF, and HtL-RVF) were modeled as six regressors convolved with a canonical hemodynamic response function. Movement parameters derived from realignment corrections were also entered in the design matrix as additional factors of no interest. Two-stage random-effect analyses were performed. Individual contrasts were created by comparing LtH and HtL sequence condition, irrespective of the visual field of presentation of the first image: LtH > HtL and HtL > LtH contrasts. At the second random-effect level, linear contrasts from all individual participants were analyzed using one-sample t tests. Clusters of activated voxels were then identified using an empirically defined threshold ($p < .005$ uncorrected for multiple comparisons, $T > 3.17$, cluster size ≥5 voxels, and all major peaks were significant at $p < .001$, see Table 1). To facilitate comparisons with other studies, we performed a transformation of MNI into Talairach and Tournoux (1988) coordinates using the MNI2TAL function (created by Matthew Brett, available at http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach).

ERP Acquisition and Analysis

Recording was carried out in an isolated, electrically shielded room. Participants were seated in the darkness 145 cm from the screen. The presentation of the first scene was restricted to the center of the screen during ERP study (see above). Subjects performed six experimental blocks, each lasting 15 min and containing 144 trials. This resulted in 432 trials for each SF sequence condition (LtH and HtL). A break of ~2 min was given after each block. Continuous EEG was acquired with a Geodesics Netamps system (Electrical Geodesics, Inc., Eugene, OR) from 123 scalp electrodes (impedances <50 kV; vertex reference; 500 Hz digitization; band-pass filtered 0.1–200 Hz). ERP epochs (from 200 msec prestimulus to 1.6 sec poststimulus onset) were separately averaged for each participant and each experimental condition. The 200-msec prestimulus epoch served as baseline. Only trials leading to correct responses were included. In addition to the rejection of EEG sweeps exceeding the amplitude of ±100 μV, the data were visually inspected to reject

Table 1. fMRI Correlates of the LtH Sequence Processing

Area	Side	x	y	z	t
LtH > HtL sequences					
Frontal Areas					
Middle frontal gyrus (FEF)	L	−36	0	53	5.28
Middle frontal gyrus (FEF)	R	39	11	41	4.39
IFG	L	−53	11	−3	4.30
Temporal Areas					
Posterior middle temporal gyrus	L	−59	−55	8	8.47
Posterior superior temporal gyrus	L	−53	−51	19	6.66
Superior temporal gyrus	R	62	−23	12	4.26
Middle temporal gyrus	R	59	−24	−11	3.78*
Parietal Areas					
Superior parietal lobule	L/R	−6	−41	63	3.78*
Inferior parietal lobule	R	48	−27	35	3.67*
Occipital Areas					
MOC	R	9	−93	5	4.17
Lateral/inferior occipital gyrus	L	−48	−81	18	3.80*
Thalamus	L	−12	−14	12	4.78

All peaks $p < .001$ uncorrected (random-effect analysis), except *$p < .005$.

epochs with blinks, eye movements, or other sources of transient noise. The mean number of accepted epochs per condition was 147 for LtH-R, 138 for LtH-UR, 138 for HtL-R, and 140 for HtL-UR. For each participant's ERP, the electrodes of the outermost circumference as well as artifact channels were excluded and interpolated to a standard 111-channel electrode array (two-dimensional spherical spline; Perrin, Pernier, Bertrand, Giard, & Echallier, 1987). ERPs were filtered off-line from 1 to 30 Hz, recalculated against the average reference and normalized to their mean global field power (Lehmann & Skrandies, 1980) before group averaging.

The main purpose of the ERP study was to search for differences in the temporal sequence of brain activity between LtH and HtL stimulation and to compare these activation patterns with fMRI results in the same subjects. We therefore restricted our analysis to differences in the electric potential fields (ERP maps) between the two conditions. Different potential field distributions on the scalp indicate different configuration of electric sources in the brain (Michel et al., 2004). To detect such topographic differences between conditions, we used a spatiotemporal pattern analysis approach as described in several previous ERP studies (for a review, see Michel et al., 2004). This seg-

mentation method is based on the observation that ERPs are characterized by a limited number of distinct topographical maps, each with a certain duration. These periods of stable topographies constitute "functional microstates" and have been interpreted as characterizing different steps of information processing (Michel et al., 2004; Lehmann & Skrandies, 1980).

A *k*-means cluster analysis (Pascual-Marqui, Michel, & Lehmann, 1995) was applied to the group-averaged ERP of the two conditions to identify the different segments of stable map configurations and their topography during these periods (i.e., segmentation maps). The optimal number of segmentation maps was determined by cross-validation (Michel et al., 2004; Picton et al., 2000; Pascual-Marqui et al., 1995; Lehmann, 1987). Once these maps were determined, their timing and sequence were determined in the individual ERPs of each subject and statistically compared between the two conditions. To do this, we compared each segmentation map (defined by group data) with the moment-by-moment map of the individual participants' ERPs in each condition by strength-independent spatial correlation (Michel et al., 2004). That is, for each time point of the individual participant's ERPs, the scalp topography was compared with all segmentation maps and labeled according to that with which it best correlated. From this spatial fitting procedure, we could then determine the total amount of time a given topography was observed for each condition in each participant. These values, which represent the frequency with which a given segmentation map was observed within a given period for each experimental condition, were then subjected to a repeated measures ANOVA using condition and segmentation maps as within-subjects factors. Thus, the fitting results could determine if the ERP from a given condition was more consistently described by one segmentation map versus another and therefore if different generator configurations better accounted for the particular experimental conditions in the particular period.

As a final step, we estimated the possible neural sources in the brain that might give rise to each of the segmentation maps, using a distributed linear inverse solution. The inverse matrices applied here were based on a Local Auto-Regressive Average (LAURA) model of the unknown current density in the brain (Grave de Peralta Menendez et al., 2004). This source localization procedure relies on electromagnetic laws that describe the activity at one point in the brain in dependency of the activity at neighboring points and uses local autoregressive averaging to describe these dependencies. Because LAURA belongs to the class of distributed inverse solutions, it is capable of dealing with multiple simultaneously active sources with a priori unknown location (Grave de Peralta & Gonzalez Andino, 2002). A realistic head model was used with a solution space of 4024 nodes, selected from a $6 \times 6 \times 6$-mm grid equally distributed within the gray matter of the average brain provided by the MNI. The procedure was implemented using the CARTOOL software by Denis Brunet (http://brainmapping.unige.ch/Cartool.php). Statistical analysis of the LAURA

source estimations was performed in the following manner: First, the above analyses of ERPs were used to define a period when stable topographies were observed within each condition and also when these topographies significantly differed between conditions (Michel et al., 2004). Next, ERP data from this period were averaged across time to generate a single time point of data for each subject and condition. The LAURA inverse solution for these data (11 subjects × 2 conditions) was estimated for each of the 4024 nodes in the source space. Paired t tests were then calculated for each node in the inverse solution space using across-subjects variance. To directly compare fMRI to source estimation results, we considered nodes with $p < .005$ uncorrected, $t(10) > 3.58$, as significant.

RESULTS

Behavioral Performance

Response accuracy and RTs for the matching task during fMRI and ERP experiments were assessed by repeated measure ANOVAs. Correct responses were faster for sequences with two scenes belonging to the same than different categories, fMRI, 603 and 660 msec, respectively, $F(1, 10) = 9.42$, $p < .02$; ERPs, 548 and 584 msec, respectively, $F(1, 10) = 8.55$, $p < .02$. There was no significant difference

in accuracy between same and different categories, fMRI, 95.8% and 97.7%, respectively, $F(1, 10) = 1.56$, $p = .24$; ERPs, 96.3% and 97%, respectively, $F(1, 10) = 2.52$, $p = .14$. Furthermore, there was no significant difference between LtH and HtL sequences for accuracy, fMRI, 96.6% and 96.3%, respectively, $F(1, 10) < 1$; ERPs, 96.3% and 97%, respectively, $F(1, 10) = 2.52$, $p = .14$, and RTs, fMRI, 636 and 627 msec, respectively, $F(1, 10) = 2.60$, $p = .14$; ERPs, 566 and 565 msec, respectively, $F(1, 10) < 1$; there was also no main effect of the visual field of the first image during fMRI for accuracy and RTs, $F(1, 10) < 1$, and no interaction between these factors.[2]

fMRI Experiment

Functional images were analyzed by statistical parametric mapping (SPM2) using the general linear model applied at each voxel across the whole brain. We first identified areas selectively engaged in a "coarse-to-fine" analysis by comparing LtH and HtL sequences, irrespective of the visual field of the first image in the sequence ([LtH-CVF + LtH-LVF + LtH-RVF] > [HtL-CVF + HtL-LVF + HtL-RVF]). Although these two sequences included the same images presented for the same short time interval (below the temporal resolution of fMRI), we found stronger neural responses to LtH than HtL in a distributed network of cortical regions (Table 1

Figure 2. Cortical activations in the fMRI experiment. (A and B) The coarse-to-fine sequence (LtH) relative to the inverse (HtL) sequence of the same images (one LSF and one HSF in each case) produced greater activation in several cortical regions including (1) bilateral FEFs, (2) left IFG, (3) left posterior middle and superior temporal cortex, (4) right inferior parietal lobule and right middle and superior gyri, and (5) right MOC. (C) Average parameter estimates of the fMRI responses (betas) are plotted for the left FEF (mean $x\ y\ z = -36\ 0\ 53$), right FEF (mean $x\ y\ z = 39\ 11\ 41$), and right MOC (mean $x\ y\ z = 9\ -93\ 5$), demonstrating a general enhancement of responses to LtH relative to HtL sequences, irrespective of the visual field of the first image (RVF vs. LVF). Despite a trend for larger fMRI responses when the first scene was lateralized, there was no significant interaction between sequence and first image position (see text).
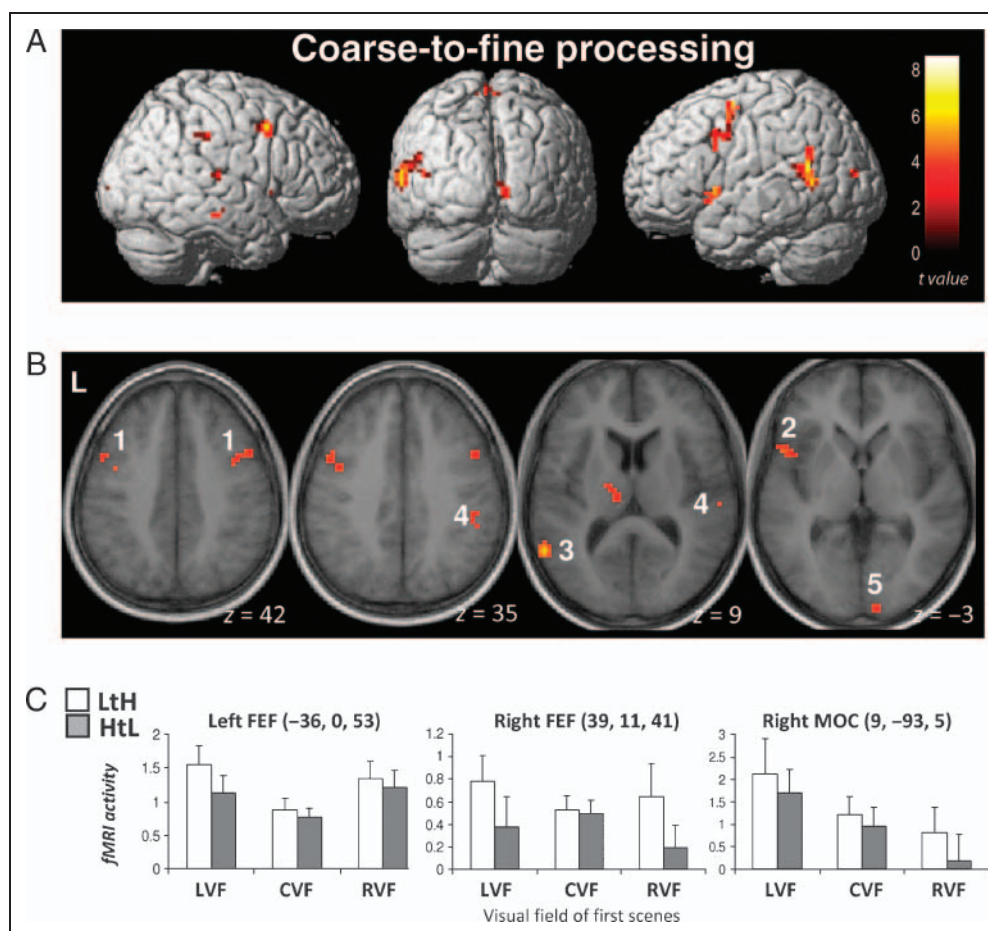
and Figure 2). These included bilateral areas in middle frontal gyri (BA 6/8; Figure 2B) whose coordinates, right peak $x\ y\ z$, 39 11 41, $t(10) = 4.39$, and left peak $x\ y\ z$, −36 0 53, $t(10) = 5.28$, closely matched those of the medial FEF (Blanke et al., 2000; Paus, 1996). As the FEF receives direct projection from the dorsal occipito-parietal stream (Bullier, Schall, & Morel, 1996) and shows short latencies of neuronal firing after stimulus onset (in monkeys, Schall, 2002; in humans, Foxe & Simpson, 2002), these results suggest that this region may be preferentially recruited by the initial processing of LSF in LtH sequences (rather than by the same images in HtL sequences). Because FEF is involved in oculomotor control (Schall, 2002; Paus, 1996), it might be argued that differential recruitment of FEF during LtH processing could potentially reflect eye movements. However, the first scene presentation was limited to 100 msec, and participants were instructed to maintain central fixation during the whole trial, thus preventing any substantial effect of overt eye movements on visual inputs. Nevertheless, we directly tested whether FEF was more activated during sequences with peripheral as compared with central presentations of the first image. A repeated measure ANOVA on parameter estimates of event-related responses extracted from the right and left FEF clusters showed no significant interaction between the visual field of the first image (CVF, RVF, or LVF) and the SF sequence type, right FEF, $F(2, 20) = 1.39$, $p > .27$, left FEF, $F(2, 20) = 1.20$, $p > .32$ (Figure 2C). In addition, eye-tracking data confirmed that the mean eye-position along the horizontal axis did not differ as a function of the first image location in visual field (CVF, LVF, and RVF) neither during the first image presentation (100 msec), $F(2, 14) < 1$, $p > .95$, nor from the onset of the first image until the appearance of the second image (500 msec), $F(2, 14) < 1$, $p > .87$.

In addition, LtH sequences produced greater activation in several other areas, including the left inferior frontal gyrus (IFG, BA 45/47) and left TPJ (encompassing the posterior middle and superior temporal gyrus, MTG/STG, BA 21/37), plus the right inferior parietal lobule, the right middle and superior temporal gyri, the bilateral superior parietal lobule, and the left thalamus (Table 1 and Figure 2). Most critically, we also found a stronger response of the medial occipital cortex (MOC) to LtH as compared with HtL sequences, including the right primary visual cortex (V1, $x\ y\ z$, 9 −93 5), $t(10) = 4.17$, $p < .001$ (Figure 2B). This activation could not be attributed to the retinotopic projection of the visual information only, because the visual field of presentation of the first image did not interact with the sequence type, $F(1, 10) < 1$, $p > .74$ (Figure 2C) despite a main effect of visual field (LVF > RVF), $F(1, 10) = 13.26$, $p < .005$.

On the other hand, the opposite contrast ([HtL-CVF + HtL-LVF + HtL-RVF] > [LtH-CVF + LtH-LVF + LtH-RVF]) showed a greater response to HtL than LtH sequence only in a few brain areas within the ventral visual stream, including parahippocampal, $x\ y\ z$, 21 −38 −3, $t(10) = 6.92$, and $x\ y\ z$, −21 −21 −17, $t(10) = 4.24$), and bilateral temporal cortex, $x\ y\ z$, −53 −44 −8, $t(10) = 4.95$, and $x\ y\ z$, 33

−49 8, $t(10) = 4.87$. This result suggests that for HtL sequences, HSF and LSF might converge only in higher level visual areas of the ventral stream within the inferior temporal cortex. This is consistent with a predominant recruitment of the HSF-dependent recognition processes along ventral temporal areas during HtL sequences, without enhanced top–down influences from prefrontal and temporo-parietal areas of the dorsal stream as seen in the LtH condition.

Note finally that our main analysis was performed by collapsing across all related (same) versus unrelated (different) categories in the stimulus sequence because this factor was irrelevant and orthogonal to the main question of interest in our study (i.e., LtH and HtL order in the sequence). However, ANOVAs conducted on parameter estimates extracted from all ROIs showed no significant main effect or interaction involving the same/different image factor. Nevertheless, for completeness, we also directly tested for the effect of stimulus repetition (i.e., priming) by comparing fMRI responses to sequences with the same versus different image categories (same trials > different trials) for both LtH and HtL. Results showed a very different pattern compared with the main findings above. Repetition-related decreases were selectively observed in several areas in the inferior occipito-temporal cortex, LtH peak $x\ y\ z$, 56 −38 −3, $t(10) = 4.62$, HtL peak $x\ y\ z$, −39 −50 −8, $t(10) = 4.19$, consistent with repetition-priming studies (e.g., Eger, Schyns, & Kleinschmidt, 2004; Vuilleumier, Henson, Driver, & Dolan, 2002; Grill-Spector et al., 1999) and with behavioral priming observed in RTs (see above).

Altogether, these fMRI results demonstrate enhanced activity in both early visual cortex and fronto-parietal areas when LSF visual inputs precede HSF inputs (LtH), relative to a reverse sequence (HtL) of the same images. Next, to assess the exact dynamics of activation in these regions during the course of different SF sequences (LtH relative to HtL), we recorded ERPs in the same participants using the same task, during a separate session with a high-density EEG system.

ERP Experiment

Analysis of ERP Topography

To compare ERPs with the fMRI data, we performed a spatiotemporal cluster analysis that determined time segments during which significantly different potential field distributions (ERP maps) were evoked by each image type (LSF or HSF) in the two sequence conditions (Michel et al., 2004; Lehmann, 1987); (Pascual-Marqui et al., 1995). We then estimated brain sources where electrical activity significantly differed during these periods. Although the cluster analysis was conducted over the whole ERP time period (Michel et al., 2004) to identify distinct map configurations across trials, our statistical comparisons between conditions focused on two separate periods of 400 msec time locked to the onset of the first and second images, respectively.

Our analysis revealed that nine distinct field topographies could be identified to describe ERPs during both sequences (Maps 1–9; Figure 3A–C) and that LtH and HtL conditions elicited different maps over three specific time segments. During the period after the first image onset (LSF for LtH sequences or HSF for HtL sequences), three similar field topographies were observed from 0 to ~140 msec and after ~220 msec (Maps 1, 2, and 4) irrespective of the image type, but the processing of HSF elicited two additional maps between 140 and 190 msec (Maps 3 and 5) that were not seen for LSF. Statistical analysis (based on fitting these maps to individual data during different time windows, see METHODS) indicated that LSF processing (relative to HSF) was dominated by Maps 2 and 4 (from 83 msec onward), $F(1, 10) = 20.44$, $p < .002$, whereas Maps 3 and 5 were selectively present during HSF rather than LSF processing, Map 3, 142–154 msec, $F(1, 10) = 6.36$, $p < .05$; Map 5, 154–194 msec, $F(1, 10) = 12.30$, $p < .006$. Furthermore, Map 4 appeared earlier in LtH than HtL sequences (150 and 194 msec, respectively; see Figure 3A and B), suggesting a more rapid

access to this processing step for LSF than HSF stimuli, consistent with faster transmission of LSF inputs through magnocellular channels (Van Essen & DeYoe, 1995).

The second period where topographies differed between LtH and HtL sequences occurred during the first 100 msec after the second image onset (Figure 3A and B). Three different maps (Maps 1, 6, and 7) were identified for this period, with HSF processing selectively associated with Map 6 in LtH relative to HtL sequences (from 12 to 96 msec), as confirmed by statistical analysis based on individual fitting, $F(1, 10) = 11.46$, $p < .007$. By contrast, LSF processing in this period was predominantly associated with Map 1, a map that was already observed during the initial processing of the first image for both LtH and HtL sequences (see above). Therefore, processing of HSF information differed (very early postonset) depending of whether it was preceded by an LSF stimulus or not, whereas the preceding stimulus did not influence the processing of LSF stimuli.

Finally, a third period of differences between conditions occurred ~100–250 msec after the onset of the second image. This period was characterized by an additional ERP
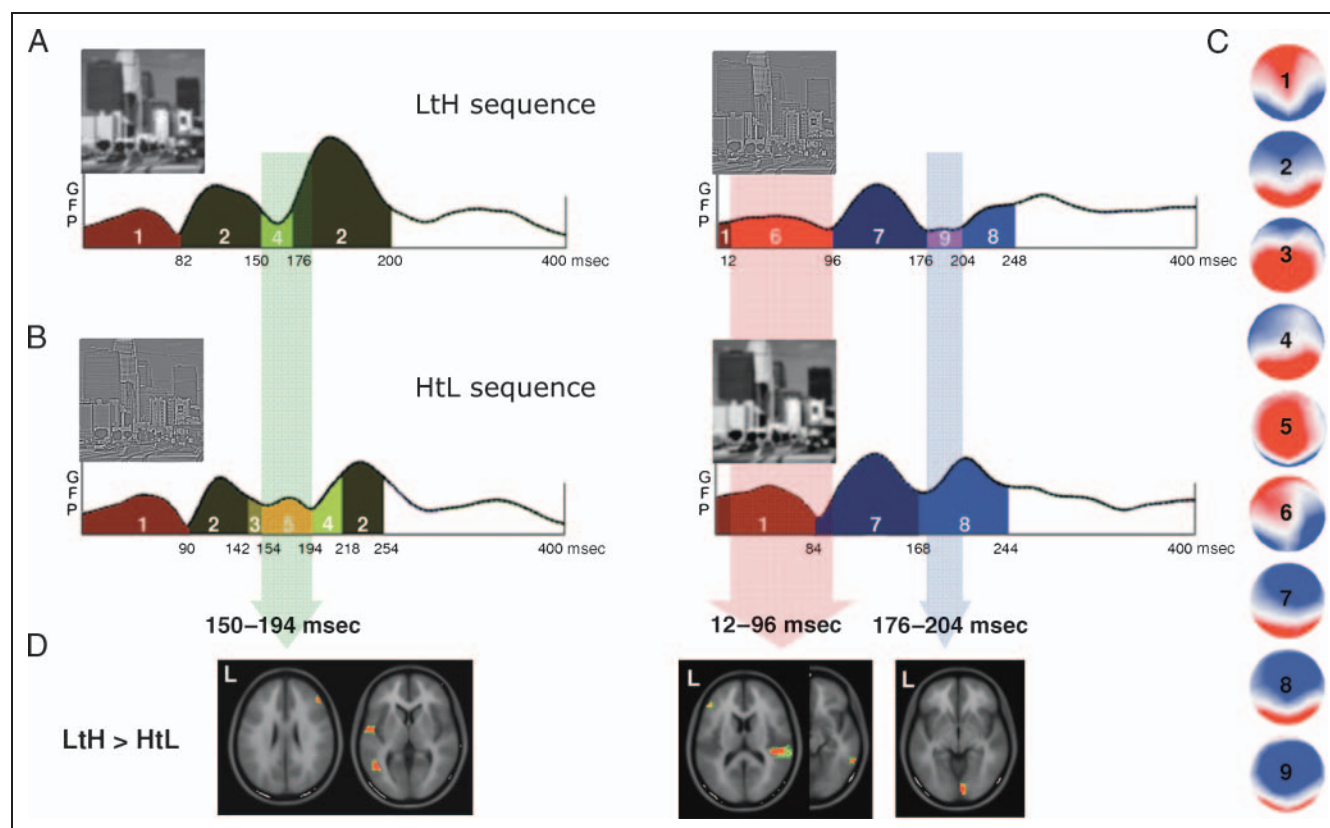
Figure 3. Results of the ERP topography and source analyses. Global field power and segmentation of electrical field maps for the group-averaged ERPs in the (A) LtH sequence and (B) HtL sequence. (C) A series of nine distinct topographical maps was identified to account for the succession of electric field configurations recorded during each type of sequence. Statistical analysis revealed that scalp topography significantly differed between LtH and HtL processing during three different time periods (shaded color). (D) LAURA-distributed linear source solution analyses were applied to these three time periods that significantly differed between LtH and HtL sequences. Statistical t-maps comparing the inverse solution for each sequence revealed greater activity for LtH than HtL in the left inferior frontal, left temporal, and right frontal sources (150–194 msec) during the processing of the first image (LSF) and in the right temporo-parietal and left frontal sources (12–96 msec) as well as in the right occipital cortex (170–200 msec) after the onset of the second image (HSF). Cluster of activated nodes were identified based on the intensity of individual responses, $p < .005$, $t(10) > 3.58$.

configuration (Map 9) that occurred exclusively for the processing of HSF in the second image of LtH sequences (from 176 to 204 msec), $F(1, 10) = 5.07$, $p < .05$.

Analysis in the Source Space

To localize the likely electrical generators underlying each topographic configuration, we applied a distributed linear inverse solution (Michel et al., 2004; Lehmann, 1987) using a LAURA model (Grave de Peralta Menendez et al., 2001, 2004) for the LtH and HtL conditions separately and then calculated paired t tests between these conditions in the three-dimensional solution space across subjects. LAURA-distributed linear source solution analyses were applied to the three time periods that significantly differed between LtH and HtL sequences. Because our ERP topography analysis (see above) showed that the onset of Map 4 occurred earlier for LSF processing in LtH sequences (150 msec) than for HSF processing in HtL sequences (194 msec), we first sought to identify the sources that may account for the more rapid LSF extraction during this period (150–194 msec). The statistical analysis in source space during the 150–194 msec after onset of the first image revealed stronger activation of the left IFG, the left posterior temporal cortex, and the right anterior frontal cortex during LtH than during HtL sequences (Figure 3D).

Note that topography analysis also indicated two additional maps elicited by HSF relative to LSF processing between 140 and 190 msec (Maps 3 and 5). Source estimation for this period revealed stronger activity in right superior parietal cortex during the 142- to 154-msec interval (Map 3) and in ACC during the 154- to 194-msec interval (Map 5), both specifically evoked by the processing of HSF information in HtL sequences but not by the processing of LSF information in LtH.

ERP topography analysis also indicated distinct map configurations during the processing of the second image (i.e., HSF information in LtH sequences condition and LSF information in HtL), arising in the 12- to 96-msec (Map 6) and in the 176- to 204-msec (Map 9) periods after onset. Statistical analysis in source space revealed that these ERP topography differences were explained by greater right temporo-parietal and left frontal activity during the first period (Map 6) and by right medial occipital sources during the second period (Map 9). These sources were remarkably consistent with our fMRI results for LtH versus HtL sequences (see Figure 2B) but in addition revealed that the preferential activation in frontal and left temporo-parietal areas in LtH was associated with the first image, whereas the preferential activation of early occipital cortex was associated with the second image.

DISCUSSION

By combining fMRI and ERPs, we could track brain activity while participants classified visual scenes presented in brief sequences of two successive pictures that contained complementary SF contents, such that the same LSF or HSF stimuli were seen across trials but in different succession (forming either an LtH or an HtL sequence). Although this procedure was obviously not physiological, it allowed us to experimentally "mimic" the sequential processing of SF inputs postulated by visual recognition models (Hegde, 2008; Bullier, 2001; Schyns & Oliva, 1994) and to systematically assess neural responses to LSF and HSF information presented in different processing order.

Convergent results from fMRI and ERPs revealed specific activations during LtH sequences relative to HtL that were highly consistent with a "coarse-to-fine" advantage and feedback modulation from higher order areas on primary visual cortex in this condition. On the one hand, fMRI showed selective increases to LtH in early occipital areas, together with frontal and temporo-parietal areas including FEF and TPJ. On the other hand, ERP topography and source analyses highlighted a similar network of cortical areas but could additionally determine a differential time course of activation in these regions, involving either LSF or HSF images in the different sequences: higher order areas in frontal and temporo-parietal regions responded more to LSF stimuli when presented first, whereas occipital visual cortex responded more to HSF presented after LSF. Taken together, these combined imaging data converge to suggest that top–down effects arising from the higher order areas might precede and enhance neural activity in early visual cortices, as we discuss below in detail.

When contrasting LtH to HtL sequences, in which the same images were shown over a 600-msec duration but in opposite order, our fMRI results revealed bilateral increases in the middle frontal gyri, overlapping with coordinates reported for FEF (Blanke et al., 2000; Paus, 1996). These findings indicate that FEF might be preferentially engaged during "coarse-to-fine" analysis of natural scenes. Such increase was observed only when processing LtH sequences, not during the reverse HtL sequences with identical images. Although FEF activity can also be related to eye movements (Schall, 2002; Paus, 1996), this was unlikely here because the first image was presented in the periphery for 100 msec only, unpredictably on either the right or the left side, such that saccade could not arise prior to the second central picture and systematically differ between sequences. Moreover, statistical analysis showed no significant effect of the lateralization of the first image on neural activity in both right and left FEF. Our fMRI results therefore provide new evidence for a sensitivity of FEF to the SF content of visual scenes. In addition, our source estimation of ERPs suggested that stronger right FEF activity in LtH sequences arose during the 140- to 160-msec period after the onset of LSF information in the first image. In keeping with neurophysiology data in monkeys and humans showing fast visual responses in FEF (Foxe & Simpson, 2002; Schall, 2002), we conclude that rapid LSF inputs to this region might serve to facilitate subsequent visual processing within early visual areas (e.g., V1) and ventral

occipito-temporal stream (e.g., V4, IT) prior to any eye movements, through feedback mechanisms sending spatially organized information about the current visual input (Hamker, 2005; Moore, Tolias, & Schiller, 1998). On the other hand, the additional involvement of right parietal and anterior cingulate areas during the processing of HSF in HtL sequences in the 140- to 190-msec period suggests that scene recognition required greater attentional and monitoring resources when HSF were presented first in sequences (Posner & Petersen, 1990).

Secondly, our fMRI data also showed increased activity in the left pFC and left middle temporal cortex during LtH as compared with HtL sequences. This finding adds support to the recent proposal (Bar et al., 2006; Bar, 2003) that the LSF content of a visual image may be rapidly projected through magnocellular pathways from early visual areas to ventrolateral regions in PFC (e.g., OFC; see Bar et al., 2006), where stored knowledge might be activated and used to generate predictions about the most likely interpretations of the visual input. The result of this computation in PFC would then be fed back to extrastriate areas in temporal cortex where it might be integrated with ongoing bottom–up analysis. Consistent with this view, our ERP data revealed a selective activation of left frontal as well as temporal sources during the 140- to 160-msec period after the onset of the first (LSF) image in LtH sequences. These findings converge with data from a combined MEG and fMRI study of Bar et al. (2006), who found that the OFC is strongly activated by LSF (relative to HSF) images and that this activation occurs at early latencies (~130 msec), although in our experiment, this early frontal activation involved more lateral areas (i.e., left IFG, coordinates $x\ y\ z$, $-53\ 11\ -3$) that did not overlap with the left OFC peak (coordinates $x\ y\ z$, $-21\ 21\ 19$) reported by Bar et al. (2006). However, the ventrolateral PFC, including IFG, as well as middle temporal regions are known to be crucially implicated during the retrieval of semantic concepts related to visual stimuli (Freedman, Riesenhuber, Poggio, & Miller, 2001; Wagner, Pare-Blagoev, Clark, & Poldrack, 2001; Mummery, Patterson, Hodges, & Price, 1998) and contribute to the maintenance of visual stimulus information in visual awareness or working memory (Fletcher & Henson, 2001). An activation of semantic processes in these areas could be particularly relevant in our scene-matching task as this required both extracting and maintaining information about the category of the two successive visual scenes. OFC activation might potentially be related to other aspects of visual recognition based on memory or motivational or affective associations (see Barrett & Bar, 2009) rather than semantic representations. Thus, although our results provide new evidence for rapid processing and precedence of LSF inputs in frontal and temporal semantic networks, further research is needed to determine whether different task demands might recruit different frontal regions. In addition, our ERP data also revealed a second activation in the left PCF during the 12- to 96-msec period after onset of the second (HSF) image, specific to LtH sequences. Whereas the first left PFC and left temporal sources in ERP data may reflect bottom–up processing of the first scene, we hypothesize that the second left PFC source observed at the onset of the second scene may correspond to the origin of top–down influences that could subsequently constrain the perceptual analysis of HSF images.

Thirdly, we found that the right TPJ, including the posterior–superior temporal cortex and the inferior parietal lobule, showed increased fMRI activity during LtH relative to HtL processing. ERP data did not only confirm an activation in the right TPJ but further demonstrated that it arose at the onset of the second (HSF) image (similar to left PFC). The TPJ is known to be critically involved in orienting and sustaining attention to visual scenes, including selective allocation to global and local information in hierarchical forms (Yamaguchi, Yamagata, & Kobayashi, 2000; Fink et al., 1996; Robertson, Lamb, & Knight, 1988) or to LSF and HSF content of natural scenes (Peyrin, Baciu, Segebarth, & Marendaz, 2004). This region may thus also contribute to top–down modulation over perceptual processes taking place in lower visual areas.

Importantly, our assumption of top–down influences on the early processing of HSF images in LtH sequences was supported by converging evidence from our ERP topography analysis. First, the initial neural response (≤100 msec) to the first image was associated with the same ERP topography (Map 1; see Figure 3) in both the LtH and the HtL sequences (despite different SF content), which very probably reflected purely bottom–up, low-level visual processing. The exact same topography was also observed during the initial (100 msec) processing of the second image (LSF) in HtL sequences, but it was rapidly replaced by a distinct topography elicited by HSF during the same time window in LtH sequences (Map 6 from 12 to 96 msec, see Figure 3). This striking dissociation between the different SF images as a function of their presentation order suggests that the initial processing of HSF scenes differed depending of whether it was preceded by an LSF scene or not. Importantly, note that similar ERP topographies were observed for the first LSF and HSF images from ~220 msec postonset until the appearance of the second image (i.e., 500 msec after the first image) and that a duration of 100 msec has already been shown to be sufficient to categorize SF-filtered scenes (Peyrin, Chauvin, Chokron, & Marendaz, 2003), making it very unlikely that the different ERP topographies observed during the initial processing of the second image were due to a late effect in the first image processing.

Taken together, these new fMRI and ERP results suggest a dynamic sequential activation of extended brain networks for visual scene recognition. During an LtH sequence, LSF information rapidly engages high-order areas in fronto-parietal cortex. The computation performed in these areas may then project back to lower visual areas and ventral stream areas, so as to guide the subsequent analysis of HSF information. Cortical areas possibly receiving feedback from this network may extend along the whole ventral visual stream, including not only the inferior and lateral

temporal cortex (e.g., V4; Hamker, 2005; Moore et al., 1998; or fusiform gyrus; Bar et al., 2006) but also the earliest cortical visual areas (such as V1; Hupe et al., 2001). Indeed, a major finding of our study was the preferential activation of MOC during LtH sequences, as revealed conjointly by fMRI and ERP source estimation. Further, this differential activation arose ~170–200 msec after onset of the second (HSF) scene (but not when the HSF scene was presented first in HtL sequences). Note that these occipital sources in ERP data could not be attributed to the processing of HSF per se but specifically reflected HSF processing following LSF because we did not found similar occipital activation (neither in fMRI nor in ERPs) when contrasting HSF to LSF using the first images in the reverse sequence. Importantly, the time course of occipital activation followed all other sources in frontal and parietal areas. This result provided the first direct evidence in support of models (Bullier, 2001), proposing that the human primary visual cortex might operate as a "display" or "blackboard" of visual inputs on which higher order areas can exert modulatory influences to promote the selection of critical information required for scene recognition and to guide further processing into the ventral visual stream.

Although coarse-to-fine processing may constitute the dominant mode of functioning for the human visual system, this does not preclude some flexibility in the extraction of spatial frequencies depending on task demands (Schyns & Oliva, 1997). In a previous fMRI study using the same experimental paradigm as here (Peyrin et al., 2005), we found a relative difference in hemispheric dominance during the processing of LtH versus HtL sequences, with greater activation of the inferior temporal cortex in the right hemisphere for LtH (peak coordinates x y z, 53 −53 −2) but in the left hemisphere for HtL (peak coordinates x y z, −39 −56 −5). These findings suggest that both types of sequence processing may coexist in the visual system, but primarily modulating higher order stages along the ventral visual stream, and each predominating in one hemisphere. Our previous study directly compared the right and the left sides by contrasting "flipped" to original "unflipped" scan images, whereas the current study used a more conventional whole-brain analysis, which makes the results of these two studies hard to compare. Importantly, the direct interhemispheric comparison method (Peyrin et al., 2005) allowed us to cancel out any main effect due to a spatial frequency bias and thus to determine a relative hemispheric specialization irrespective of the more general coarse-to-fine processing course.[3]

In sum, our combined fMRI and ERP study allowed us to identify specific neural substrates for top–down processes during "coarse-to-fine" perception of natural scenes in humans. Our results demonstrate that low-pass signals (conveyed by fast magnocellular channels) can rapidly activate high-order areas, providing spatial (via FEFs) and semantic information (via left PFC and temporal areas) as well as attentional signals (via TPJ) that altogether may promote ongoing perceptual organization and categorization of the

visual input. This first coarse analysis might be refined by further processing of high-pass signals (conveyed more slowly by the parvocellular channels) in visual cortices. For this purpose, feedback from the first low-pass computations could be "retro-injected" back into lower level areas, including the primary visual cortex, to guide the high-pass analysis and select the relevant finer details necessary for recognition and categorization. These results provide critical support to recent models of vision (Hegde, 2008; Bar, 2003; Bullier, 2001) and illustrate how integrating fMRI and advanced EEG techniques can now allow a precise delineation of dynamic neural events underlying human perception and cognition.

Notes

1. The energy level for LSF and HSF stimuli was equalized for each scene as follows: If $LSF(i, j)$ and $HSF(i, j)$ represent the value of the pixel at position (i, j) of the low and the high-pass filtered images of a scene, respectively, their energies are given by $E_{LSF} = \sum_{i,j} LSF(i,j)^2$ and $E_{HSF} = \sum_{i,j} HSF(i,j)^2$. The average energy between LSF and HSF stimuli is then given by $E_{AVR} = (E_{LSF} + E_{HSF}) / 2$. The stimuli are then normalized by the average energy, $LSF_{norm}(i, j) = LSF(i,j)E_{AVR} / E_{LSF}$ and $HSF_{norm}(i, j) = HSF(i,j)E_{AVR} / E_{HSF}$.

2. The present behavioral results showed that both HtL and LtH sequences were rapidly and accurately recognized. According to the hypothesis of a preferential coarse-to-fine processing, one might expect that scene categorization would be faster and/or easier for LtH than HtL sequences. The lack of behavioral effects in the present experiments might be due to the long interimage interval (400 msec) used because of ERP constraints. We therefore conducted an additional behavioral experiment in 10 other male participants who performed the same matching task except that the two successive scenes were now presented for 100 msec each in the screen center, without any interimage interval. Although the task was more difficult without an interimage interval, there was no significant difference in accuracy between LtH and HtL sequences (86% and 84% correct, respectively), $F(1, 9) = 3.90, p = .08$. However, correct responses were faster for LtH than HtL, irrespective of the same/different condition (735 and 752 msec, respectively), $F(1, 9) = 5.93, p < .04$. These results support the hypothesis of a coarse-to-fine advantage for this set of stimuli and matching task.

3. To verify the results of Peyrin et al. (2005), we also applied the direct interhemispheric comparison method on the current fMRI data. Results confirmed a greater activation of inferior temporal

cortex in the right than left hemisphere during LtH visual analysis (peak coordinates x y z, 53 −53 −2), $t(10) = 6.38$, but greater activation in the left than right hemisphere during HtL visual analysis (peak coordinates x y z, −50 −70 3), $t(10) = 8.27$.

REFERENCES

Badcock, J. C., Whitworth, F. A., Badcock, D. R., & Lovegrove, W. J. (1990). Low-frequency filtering and the processing of local-global stimuli. *Perception, 19,* 617–629.

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

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top–down facilitation of visual recognition. *Proceedings of the National Academy of Sciences, U.S.A., 103,* 449–454.

Bar, M., Tootell, R., Schacter, D., Greve, D., Fischl, B., Mendola, J., et al. (2001). Cortical mechanisms of explicit visual object recognition. *Neuron, 29,* 529–535.

Barrett, L. F., & Bar, M. (2009). See it with feeling: Affective predictions during object perception. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 364,* 1325–1334.

Biederman, I. (1995). Visual object recognition. In S. F. Kosslyn & D. N. Osherson (Eds.), *An invitation to cognitive science* (pp. 121–165). Cambridge, MA: MIT Press.

Blanke, O., Spinelli, L., Thut, G., Michel, C. M., Perrig, S., Landis, T., et al. (2000). Location of the human frontal eye field as defined by electrical cortical stimulation: Anatomical, functional and electrophysiological characteristics. *NeuroReport, 11,* 1907–1913.

Breitmeyer, B. G. (1975). Simple reaction time as a measure of the temporal response properties of transient and sustained channels. *Vision Research, 15,* 1411–1412.

Bullier, J. (2001). Integrated model of visual processing. *Brain Research, Brain Research Reviews, 36,* 96–107.

Bullier, J., Schall, J. D., & Morel, A. (1996). Functional streams in occipito-frontal connections in the monkey. *Behavioural Brain Research, 76,* 89–97.

DeValois, R. L., & DeValois, K. K. (1988). *Spatial vision.* New York: Oxford University Press.

Eger, E., Schyns, P. G., & Kleinschmidt, A. (2004). Scale invariant adaptation in fusiform face-responsive regions. *Neuroimage, 22,* 232–242.

Fink, G. R., Halligan, P. W., Marshall, J. C., Frith, C. D., Frackowiak, R. S., & Dolan, R. J. (1996). Where in the brain does visual attention select the forest and the trees? *Nature, 382,* 626–628.

Fletcher, P. C., & Henson, R. N. (2001). Frontal lobes and human memory: Insights from functional neuroimaging. *Brain, 124,* 849–881.

Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans A framework for defining "early" visual processing. *Experimental Brain Research, 142,* 139–150.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science, 291,* 312–316.

Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N., & Dale, A. M. (1999). Stochastic designs in event-related fMRI. *Neuroimage, 10,* 607–619.

Grave de Peralta Menendez, R., Gonzalez Andino, S., Lantz, G., Michel, C. M., & Landis, T. (2001). Noninvasive localization of electromagnetic epileptic activity. I. Method descriptions and simulations. *Brain Topography, 14,* 131–137.

Grave de Peralta Menendez, R., Murray, M. M., Michel, C. M., Martuzzi, R., & Gonzalez Andino, S. L. (2004). Electrical neuroimaging based on biophysical constraints. *Neuroimage, 21,* 527–539.

Grave de Peralta, R., & Gonzalez Andino, S. (2002). Comparison of algorithms for the localization of focal sources: Evaluation with simulated data and analysis of experimental data. *International Journal of Bioelectromagnetism, 4.*

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron, 24,* 187–203.

Hamker, F. H. (2005). The reentry hypothesis: The putative interaction of the frontal eye field, ventrolateral prefrontal cortex, and areas V4, IT for attention and eye movement. *Cerebral Cortex, 15,* 431–447.

Hegde, J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Progress in Neurobiology, 84,* 405–439.

Hupe, J. M., James, A. C., Girard, P., Lomber, S. G., Payne, B. R., & Bullier, J. (2001). Feedback connections act on the early part of the responses in monkey visual cortex. *Journal of Neurophysiology, 85,* 134–145.

Josephs, O., & Henson, R. N. (1999). Event-related functional magnetic resonance imaging: Modelling, inference and optimization. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 354,* 1215–1228.

Lamb, M. R., & Yund, E. W. (1993). The role of spatial frequency in the processing of hierarchically organized stimuli. *Perception & Psychophysics, 54,* 773–784.

Lehmann, D. (1987). Principles of spatial analysis. In A. S. Gevins & A. Remond (Eds.), *Handbook of electroencephalography and clinical neurophysiology. Methods of analysis of brain electrical and magnetic signals* (Vol. 1, pp. 309–354). Amsterdam: Elsevier.

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology, 48,* 609–621.

Martin, P. R., & Grunert, U. (2003). Ganglion cells in mammalian retinae. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (pp. 410–421). Cambridge, MA: MIT Press.

Michel, C. M., Murray, M. M., Lantz, G., Gonzalez, S., Spinelli, L., & Grave de Peralta, R. (2004). EEG source imaging. *Clinical Neurophysiology, 115,* 2195–2222.

Moore, T., Tolias, A. S., & Schiller, P. H. (1998). Visual representations during saccadic eye movements. *Proceedings of the National Academy of Sciences, U.S.A., 95,* 8981–8984.

Mummery, C. J., Patterson, K., Hodges, J. R., & Price, C. J. (1998). Functional neuroanatomy of the semantic system: Divisible by what? *Journal of Cognitive Neuroscience, 10,* 766–777.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9,* 353–383.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9,* 97–113.

Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. *Perception, 21,* 147–160.

Pascual-Marqui, R. D., Michel, C. M., & Lehmann, D. (1995). Segmentation of brain electrical activity into microstates: Model estimation and validation. *IEEE Transactions on Biomedical Engineering, 42,* 658–665.

Paus, T. (1996). Location and function of the human frontal eye-field: A selective review. *Neuropsychologia, 34,* 475–483.

Perrin, F., Pernier, J., Bertrand, O., Giard, M. H., & Echallier, J. F. (1987). Mapping of scalp potentials by surface spline interpolation. *Electroencephalography and Clinical Neurophysiology, 66,* 75–81.

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., Chauvin, A., Chokron, S., & Marendaz, C. (2003). Hemispheric specialization for spatial frequency processing in the analysis of natural scenes. *Brain and Cognition, 53,* 278–282.

Peyrin, C., Schwartz, S., Seghier, M., Michel, C., Landis, T., & Vuilleumier, P. (2005). Hemispheric specialization of human inferior temporal cortex during coarse-to-fine and fine-to-coarse analysis of natural visual scenes. *Neuroimage, 28,* 464–473.

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., Jr., et al. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology, 37,* 127–152.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13,* 25–42.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2,* 1019–1025.

Robertson, L. C., Lamb, M. R., & Knight, R. T. (1988). Effects of lesions of temporal-parietal junction on perceptual and attentional processing in humans. *Journal of Neuroscience, 8,* 3757–3769.

Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research, 21,* 409–418.

Rotshtein, P., Vuilleumier, P., Winston, J., Driver, J., & Dolan, R. (2007). Distinct and convergent visual processing of high and low spatial frequency information in faces. *Cerebral Cortex, 17,* 2713–2724.

Schall, J. D. (2002). The neural selection and control of saccades by the frontal eye field. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 357,* 1073–1082.

Schulman, G., Sulivan, M., Gisch, K., & Sadoka, W. (1986). The role of spatial frequency channels in the perception of local and global structure. *Perception, 15,* 259–273.

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.

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.

Van Essen, D. C., & DeYoe, E. A. (1995). Concurrent processing in the primate visual cortex. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 383–400). Cambridge: Bradford Book.

Vuilleumier, P., Henson, R. N., Driver, J., & Dolan, R. J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nature Neuroscience, 5,* 491–499.

Wagner, A. D., Pare-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. *Neuron, 31,* 329–338.

Yamaguchi, S., Yamagata, S., & Kobayashi, S. (2000). Cerebral asymmetry of the "top–down" allocation of attention to global and local features. *Journal of Neuroscience, 20,* RC72.