From Coarse to Fine? Spatial and Temporal Dynamics of Cortical Face Processing

Valerie Goffaux[1,2], Judith Peters[2,3], Julie Haubrechts[4], Christine Schiltz[1], Bernadette Jansma[2] and Rainer Goebel[2]

[1]Educational Measurement and Applied Cognitive Science Unit and Faculté des Lettres, des Sciences Humaines, des Arts et des Sciences de l'Education, University of Luxembourg, L-7210 Walferdange, Luxembourg, [2]Department of Neurocognition, Maastricht University, 6229ER Maastricht, the Netherlands, [3]The Netherlands Institute for Neuroscience, Institute of the Royal Netherlands Academy of Arts and Sciences (KNAW), 1105-BA Amsterdam, the Netherlands and [4]Centre de Consultations Psychologiques Spécialisées, University of Louvain, 1348 Louvain-la-Neuve, Belgium

Address correspondence to Valerie Goffaux, Department of Neurocognition, Maastricht University, Universiteitssingel 40, 6229ER Maastricht, the Netherlands. Email: valerie.goffaux@maastrichtuniversity.nl.

Primary vision segregates information along 2 main dimensions: orientation and spatial frequency (SF). An important question is how this primary visual information is integrated to support high-level representations. It is generally assumed that the information carried by different SF is combined following a coarse-to-fine sequence. We directly addressed this assumption by investigating how the network of face-preferring cortical regions processes distinct SF over time. Face stimuli were flashed during 75, 150, or 300 ms and masked. They were filtered to preserve low SF (LSF), middle SF (MSF), or high SF (HSF). Most face-preferring regions robustly responded to coarse LSF, face information in early stages of visual processing (i.e., until 75 ms of exposure duration). LSF processing decayed as a function of exposure duration (mostly until 150 ms). In contrast, the processing of fine HSF, face information became more robust over time in the bilateral fusiform face regions and in the right occipital face area. The present evidence suggests the coarse-to-fine strategy as a plausible modus operandi in high-level visual cortex.

Keywords: coarse-to-fine, exposure duration, face perception, fMRI, spatial frequency

Introduction

Primary steps of human vision decompose the retinal input along 2 main dimensions: orientation and spatial frequencies (SF). This primary visual information is assumed to be combined in higher level visual regions located in inferior temporal cortex, yielding complex representations thought to underlie the perception of a rich and coherent environment. While there is extensive knowledge on the primary processing of SF in V1 (De Valois et al. 1982; Hess 2004), it is still not known how this primary visual information is integrated in higher level visual cortex.

A number of theoretical models assume that the visual system combines the information carried by different SF following a coarse-to-fine sequence (Marr 1982; Watt 1987; Bullier 2001; Bar 2007; see also Hochstein and Ahissar 2002). It is proposed that the coarse structure of a stimulus, which is carried by low SF (LSF), is processed before the fine details transmitted by high SF (HSF). For example, once the coarse structure of a face is detected, it would be used as an index into the fine facial structure. Such a strategy would be very efficient since the LSF structure provides a stable representation of the image before the noisier HSF structure is extracted. Electrophysiological evidence of such coarse-to-fine scenario has been reported in V1 (Bredfeldt and Ringach 2002; Mazer et al. 2002; Frazor et al.

2004). Moreover, coarse-to-fine temporal dynamics have been described with a variety of stimuli, ranging from lines, dots, and gratings (Musselwhite and Jeffreys 1985; Parker and Dutch 1987; Watt 1987; Hughes et al. 1996; Mihaylova et al. 1999) to complex stimuli such as faces (McCarthy et al. 1999; Halit et al. 2006; Vlamings et al. 2009) or natural scenes (Parker et al. 1992, 1997; Schyns and Oliva 1994; Peyrin et al. 2006). It has also been documented in other sensory modalities (Narayan et al. 2005; Sripati et al. 2006), suggesting that coarse-to-fine processing is a general property of signal processing in the brain (Allen and Freeman 2006; see Hegde 2008).

Within the visual domain, evidence for coarse-to-fine processing at high-level processing stages is however still lacking. The few past studies addressing coarse-to-fine processing in the human brain (Peyrin et al. 2010, 2005; Bar et al. 2006) did not explore the LSF over HSF processing precedence in high-level visual cortex (see Discussion). By manipulating exposure duration and SF content of filtered images, the present study investigated the differential contribution of SF during the build up of the visual representation of complex stimuli, for example, faces. Faces constitute an ideal visual category to tackle spatiotemporal dynamics of high-level vision. The ubiquity and social importance of faces in human life have pushed the visual system to adopt extremely fast and efficient strategies to extract face information. Moreover, several aspects suggest that face perception is more sensitive to SF than the visual processing of other complex visual categories (Biederman and Kalocsai 1997; Liu et al. 2000; Fiser et al. 2001; Goffaux et al. 2003; Collin et al. 2004; Yue et al. 2006; Williams et al. 2009). First, the integration of face cues into a global, so-called holistic, face representation relies on the processing of LSF face information (below 8 cycles per faces, cpf; Collishaw and Hole 2000; Goffaux and Rossion 2006; Goffaux 2009; but see Cheung et al. 2008). Second, the extraction of face identity relies on intermediate SF situated around 12 cpf (e.g., Gold et al. 1999; Nasanen 1999; Tanskanen et al. 2005). Finally, the analysis of face local details is based on HSF (above 32 cpf; Goffaux and Rossion 2006).

Human functional magnetic resonance imaging (fMRI) evidence portrays higher level visual cortex as a mosaic of category-preferring regions tuned to global object properties (Lerner et al. 2001). In particular, the fusiform face area (FFA) responds more robustly to faces than other object categories (Sergent et al. 1992; Kanwisher et al. 1997). The FFA, especially in the right hemisphere (right fusiform face area [rFFA]), is thought to represent the identity of faces based on the robust integration of local cues in a so-called holistic representation (Schiltz and Rossion 2006; Goffaux et al. 2009). However, how

primary visual information is combined to yield high-level face representations in the rFFA is an unanswered question.

Here, we compared the activation of face-preferring regions with faces that were filtered to selectively preserve LSF, middle SF (MSF), or HSF. These stimuli were presented either at 75, 150, or 300 ms and subsequently masked (Figure 1). We observe that the processing of face information in most face-preferring regions, especially in rFFA, initially relies on LSF; with increasing exposure time, face-preferring regions attenuate LSF processing in favor of HSF processing. Our findings thus indicate the existence of a coarse-to-fine sequence of SF processing in face-preferring cortical regions. The ventral lateral occipital complex (LOC), a general-purposed high-level visual region encoding complex shape properties with no preference for any given visual category, failed to reveal such a coarse-to-fine sequential processing, suggesting that this scenario selectively applies to high-level, category-preferring visual regions.

Methods

fMRI Acquisition

Thirteen adult subjects (normal or corrected-to-normal vision; mean age 26 ± 4, 4 males, 2 left handed; no history of neurological disease) performed 2 scanning sessions on different days (spread over 2 weeks, on average). In this paper, we report the results of 2 experiments, namely, the localizer and the SF experiments. The order of experiments and runs was counterbalanced across subjects.

Imaging was performed on a 3-T head scanner at the University of Maastricht (Allegra, Siemens Medical Systems) provided with standard head coil. T_2^*-weighted echo-planar imaging was performed using blood oxygen level–dependent (BOLD) contrast as an indirect marker of local neuronal activity.

In the localizer experiment, twenty-five 3.5-mm oblique coronal slices were acquired (no gap, time repetition [TR] = 1500 ms, time echo [TE] = 28 ms, flip angle [FA] = 67°, matrix size = 64 × 64, field of view [FOV] = 224 mm, in-plane resolution 3.5 × 3.5 mm). Each subject performed 2 localizer runs of 265 TRs each (approximately 400 s).

In the SF experiment, twenty-one 3.5-mm oblique coronal slices (no gap, TR = 1250 ms, TE = 28 ms, FA = 67°) were acquired. Each subject performed 4 experimental runs of 690 TRs each (approximately 862.5 s).

A high-resolution T_1-weighted anatomical data set encompassing the whole head was acquired in each session by means of a "modified driven equilibrium Fourier transform" sequence (TR = 2250 ms, TE = 26 ms, FA = 9°, matrix size = 256 × 256, FOV = 256 mm2, 192 slices, slice thickness = 1 mm, no gap, total run time= 8 min, 26 s).

Visual Stimulation

Visual stimuli were presented using Eprime 1.1 on a uniformly gray background. They were projected onto a translucent screen at the head
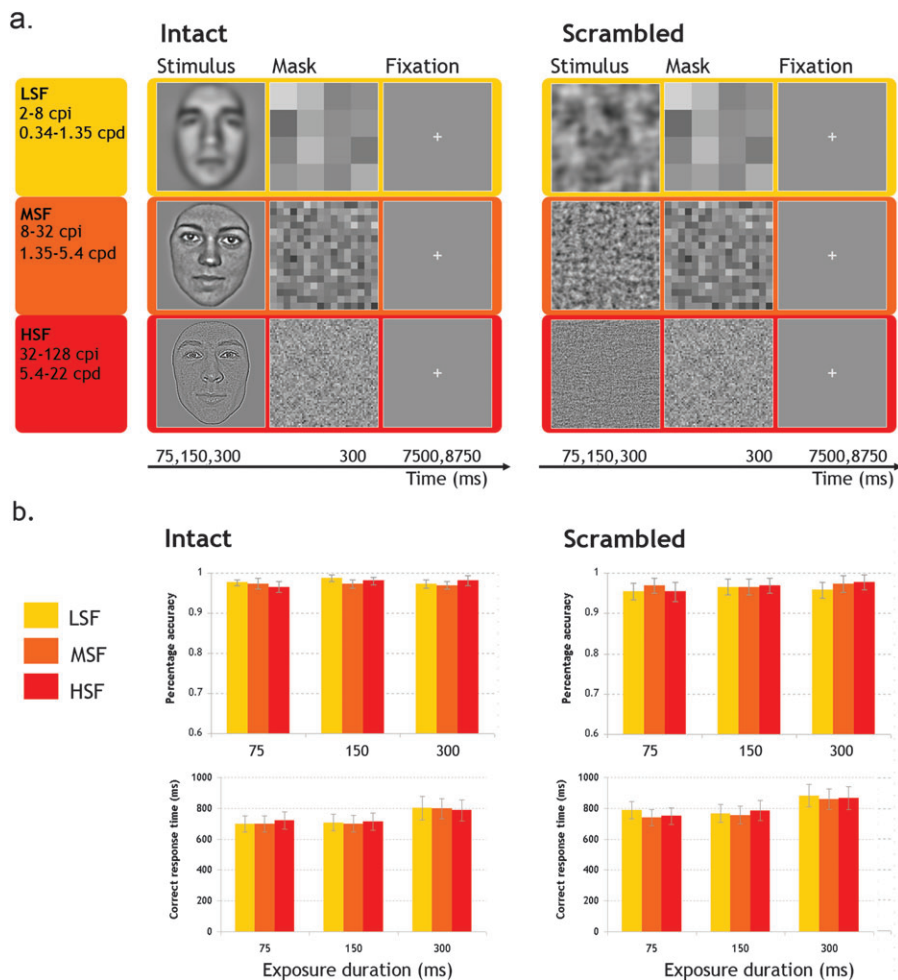
Figure 1. (*a*) LSF, MSF, and HSF faces were presented at 3 exposure durations, immediately followed by a Gaussian mask. The phase of face stimuli was either intact or scrambled in the Fourier domain. All conditions were equated for luminance, RMS contrast, and spectral composition. They were randomly interleaved within a run and subjects categorized each trial as an intact or a scrambled one. (*b*) Performance accuracy in intact-scrambled categorization was at ceiling and was not influenced by SF, exposure, or stimulus factors. In contrast, correct response times were shorter for intact than scrambled conditions, and significantly increased at 300-ms exposure duration compared with 75- and 150-ms exposure conditions.

of the scanner bore by means of a liquid crystal display projector and viewed by the subjects through a mirror placed within the radio frequency coil at a viewing distance of 57 cm. Stimulus size was 256 × 256 pixels. At a resolution of 1024 × 768 pixels, all stimuli subtended a visual angle of 5.8 × 5.8 degrees. Behavioral responses were collected during acquisition via a button box.

Face images were first normalized to obtain a global luminance with zero mean and a standard deviation (i.e., root mean square [RMS] contrast) equal to 1 using MatLab 7.5. Subsequently, filtered stimuli were generated by fast Fourier transforming the image and multiplying the Fourier energy with Gaussian filters. In the localizer experiment, stimuli were filtered using a broadband Gaussian filter (preserving information between 2 and 128 cycles per image, cpi, or 0.34–22 cycles per degree, cpd) in order to exclude SF below 2 cpi. In the main experiment (i.e., SF experiment), 2-octave-wide bandpass Gaussian filters were applied to the face images to filter the LSF (from 2 to 8 cpi or 0.34 to 1.35 cpd), MSF (from 8 to 32 cpi, 1.35 to 5.4 cpd), and HSF (from 32 to 128 cpi or 5.4 to 22 cpd; see Figure 1a).

In natural images such as face or scene pictures, amplitude typically decays as a function of SF. This decay obeys $1/f^\alpha$ with $0.7 < \alpha < 2$ (Field 1987; Tolhurst et al. 1992; see Figure 2). As a consequence, bandpass filters centered on lower versus higher ends of SF spectrum will pass information of high versus low energy, respectively. Since we were interested into BOLD modulations related to high-level processing of different SF ranges, we avoided this potential confound by attributing the same global luminance and RMS contrast to LSF, MSF, and HSF images (intact or scrambled). This control is necessary since RMS contrast has been shown to be the best index for perceived contrast in natural images (Bex and Makous 2002) and to largely drive neural activation in the visual cortex (Boynton et al. 1996). Without any control of this parameter, one cannot ascertain that all SF are equally visible to the observer, thus severely hampering conclusions about spatial scale processing per se.

Phase of the face images was scrambled in the Fourier domain via random permutation, a procedure known to preserve orientation content (Dakin et al. 2002). To substantiate this point, Figure 3 highlights the high similarity of SF and orientation spectra of stimulus images before and after phase scrambling.

After the inverse Fourier transform, the mean (i.e., the global luminance) and standard deviation (i.e., global RMS contrast) of each image were adjusted to match the average global luminance and RMS contrast of the original image set (Figure 2). This procedure is conventionally used to warrant equal global luminance and RMS contrast values across SF conditions (e.g., Vlamings et al. 2009). Luminance (intact LSF: 0.52 ± 0.00003; intact MSF: 0.52 ± 0.00004; intact HSF: 0.52 ± 0.00001; scrambled LSF: 0.52 ± 0; scrambled MSF: 0.52 ±0; scrambled HSF: 0.52 ± 0) and contrast values (intact LSF: 0.1 ± 0; intact MSF: 0.1 ± 0.009; intact HSF: 0.09 ± 0.002; scrambled LSF: 0.1 ± 0; scrambled MSF: 0.1 ± 0; scrambled HSF: 0.1 ± 0) were highly similar between the stimulus and SF conditions of the SF experiment and barely varied within conditions, indicating the high efficiency of our equalization procedure. Figure 2 further illustrates that equalization does not alter SF spectral

envelope. A 2-pixel light gray border surrounded all stimuli to minimize global shape differences between intact and scrambled stimuli.

A localizer run comprised 16-s blocks of 20 gray-scale images: intact faces, intact cars, scrambled faces, or scrambled cars. Face pictures used in the localizer runs were not shown during the experimental runs. Within a block, each stimulus appeared during 600 ms at a random x,y position (±10 pixels away from screen center), followed by a blank screen of 200 ms. During each block, subjects performed a one-back matching task. Blocks were interleaved with 15 s of fixation pauses. There were 3 blocks per condition per run.

The SF experiment was a slow event-related design comprising 18 different conditions: SF (LSF, MSF, HSF) × exposure (75, 150, 300 ms) and stimulus (intact, scrambled). All conditions were randomly interleaved within a run. There were 5 trials per condition per run and there were 4 runs in total, giving a total of 20 trials per condition. The start of a trial was announced by a transiently brighter fixation cross (average duration: 1685 ms). Either an intact or a scrambled face then appeared during 75, 150, or 300 ms, immediately followed by a Gaussian noise mask (duration: 300 ms; 256 × 256 pixels) to eliminate any retinally persisting image of the stimulus and to limit processing time to exposure duration (Keysers and Perrett 2002). To maximize masking, the SF content of the mask was adjusted to fit stimulus center SF: square size of 64 × 64 pixels were used in LSF conditions (i.e., 4 cpi in a 256 × 256 pixel image), square size of 16 × 16 pixels in MSF conditions (i.e., 16 cpi), and square size of 4 × 4 pixels in HSF conditions (i.e., 64 cpi). Intact and scrambled conditions were matched for luminance, RMS contrast as well as spectral composition; they were also matched with respect to mask since different Gaussian masks were paired with different faces but were identical across intact and scrambled conditions. Our findings, which mostly rely on intact–scrambled comparisons across SF and exposure duration, thus cannot be due to divergent masking parameters. The mask was followed by a long fixation pause (8.125 s on average). Subjects had to perform an intact versus scrambled categorization task by pressing 1 of 2 buttons with their right index or middle fingers. Within a run, a given face appeared in both intact and scrambled version. Over the 4 runs, all faces were equally often presented in LSF, MSF, or HSF range. However, to avoid face-priming effects across SF, a given face appeared in only one SF range within a run.

Localizer Behavioral Performance

In the localizer experiment, hits and correct rejections of the one-back sensitivity were combined to compute standard sensitivity estimate (d') individually. One-back sensitivity was high, in all conditions (intact faces: 3.9 ± 0.17; intact cars: 3.55 ± 0.23; scrambled faces: 3.25 ± 0.16; scrambled cars: 3.08 ± 0.22) but was significantly affected by category (faces vs. cars; $F_{1,11} = 7.07$, $P < 0.03$) and stimulus (intact vs. scrambled; $F_{1,11} = 12.02$, $P < 0.007$) as subjects performed less accurately for cars than faces and for scrambled than intact stimuli. There was no significant difference between faces and cars conditions when intact and scrambled conditions were considered separately (Ps > 0.07).
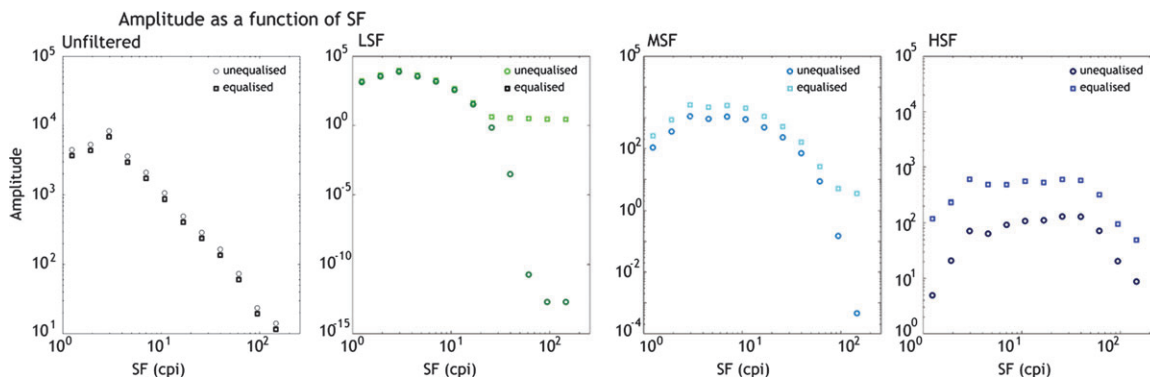
Figure 2. Amplitude spectrum as a function of SF in unfiltered, LSF, MSF, and HSF stimuli, before and after luminance and RMS contrast have been equalized. Note that luminance and contrast equalization did not alter spectral envelope.

fMRI Data Analyses

Functional and anatomical images were analyzed using BrainVoyager QX (version 1.10, Brain Innovation). The first 4 volumes were skipped to avoid T_1 saturation effect. Functional runs then underwent several preprocessing steps: correction of interslice scan time differences, linear trend removal, temporal high-pass filtering (to remove frequencies lower than 3 cycles per time course), smoothing with a Gaussian kernel of 6-mm full width at half maximum, and correction for interscan head motion (translation and rotation of functional volumes to align them to a reference volume). Anatomical and functional data were spatially normalized to Talairach coordinate system (Talairach and Tournoux 1988) with a resolution of $3 \times 3 \times 3$ mm using sinc interpolation.

Individual regions of interest (ROIs) were isolated based on 2 localizer scans. The fMRI signal in the localizer runs was analyzed using single-participant general linear model. The predictor time courses for stimulation blocks were constructed as box-car functions filtered through a linear model indirectly relating neural activity and BOLD response (Boynton et al. 1996). For anatomical reference, the statistical maps were overlaid on Talairach-normalized individual anatomical volumes. The areas responding preferentially to faces were defined independently for each participant by the (intact faces – intact cars) contrast. Significant voxel clusters on individual t maps were selected as ROIs for further analysis. Face-preferring voxel clusters were located in bilateral middle fusiform gyri (rFFA and left fusiform face area [lFFA]; selected at a q[false discovery rate, FDR] < 0.01), superior temporal sulci (STS; q[FDR] < 0.01), anterior inferotemporal cortex (AIT; q[FDR] < 0.05), and right inferior occipital gyrus (the right Occipital Face Area [rOFA]; q[FDR] < 0.01). The left occipetal face area (lOFA) was only found in 6 out of the 13 subjects and was discarded from the analyses. Right- and left-lateralized AIT activation foci were only found
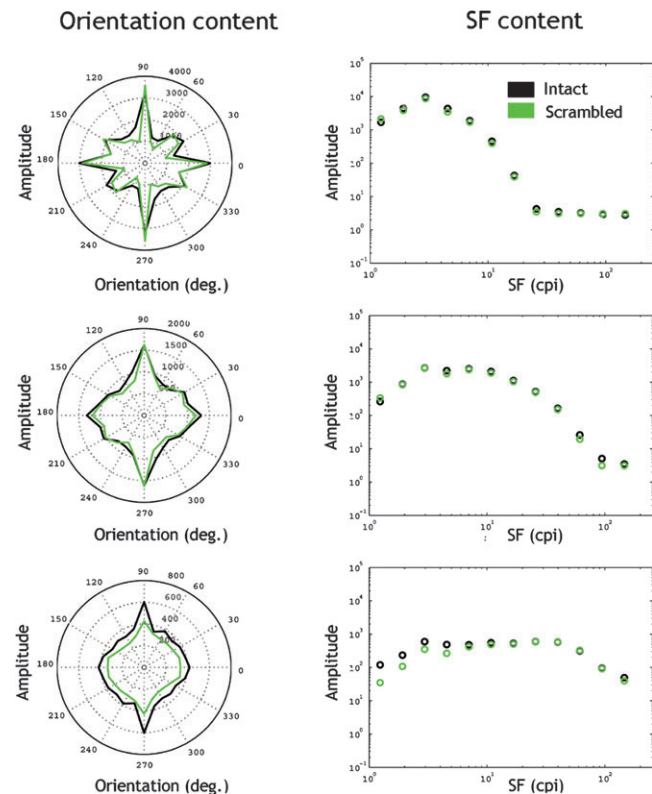
in 9 and 7 subjects, respectively, and were consequently collapsed in subjects showing bilateral foci (7 out of 9 subjects). We localized ventral LOC in both hemispheres in all the subjects using the contrast (intact cars – scrambled cars) at a P(Bonferroni) < 0.001. To ascertain that the LOC ROIs also process face information, individual z-scored beta weights from rLOC and lLOC were extracted in each condition of the localizer experiment and submitted to a repeated-measure analysis of variance (ANOVA) with stimulus (intact, scrambled) and category (face, car) as factors. Afterward, post hoc Fisher's least significant difference (LSD) tests were used to compare conditions 2×2. We found a significant intact–scrambled difference for each category (Ps < 0.0002).

Talairach coordinates of ROIs were consistent with previous studies (see Table 1).

We extracted individual z-scored beta weights from these individual ROIs for each condition of the SF experiment. Beta weights were subjected to a repeated-measure ANOVA with stimulus (intact, scrambled), SF (LSF, MSF, HSF), and exposure duration (75, 150, 300 ms) as factors. Post hoc Fisher's LSD tests were used to compare conditions 2×2.

Scrambled conditions were used as control conditions, from which no face representation can be extracted despite identical luminance, RMS contrast and SF spectrum (Figure 3). To gain more insight in high-level visual processing, we compared ROI activation in intact and scrambled conditions in SF and exposure conditions in 2 ways. First, since all face-preferring ROIs, but the lOFA, responded more strongly to intact faces than scrambled faces shown in the SF experiment, we ran separate ANOVAs for intact and scrambled conditions with SF (low, middle, high) and exposure duration (75, 150, 300 ms) as within-subject factors. Second, we directly compared ROI activation in intact and scrambled conditions using planned comparisons. We estimated the magnitude of this difference using partial eta squared (partial η^2). Partial η^2 quantifies the percentage of variance explained by a given factor (here, stimulus) when excluding the contribution of intersubject variance. Partial η^2 was used to estimate the percentage of BOLD variance related to the processing of face information across SF and time while avoiding unwarranted computations of face-related activation (Baker et al. 2007; Simmons et al. 2007).

Additionally to these ROI analyses, we performed a random-effects (RFX) whole-brain analysis by computing (intact face – scrambled face) contrasts for each SF and duration (see Supplementary Data 2). We restricted this analysis to the subspace of all subjects' brain resulting from intersecting the scanned functional volumes.

Results

In a slow event-related design, subjects viewed intact and scrambled faces that were filtered to preserve only LSF, MSF, or HSF. Intact and scrambled faces were presented at 3 different exposure durations (75, 150, 300 ms), immediately followed by a Gaussian mask (see Figure 1*a*). They performed an easy intact–scrambled categorization task, which yielded comparable accuracy across SF conditions. This allowed us to avoid potential confounds (e.g., attentional, decisional, and/or motor

Figure 3. Left: Fourier amplitude is plotted as a function of orientation, revealing the similar orientation content across intact and scrambled conditions in each SF conditions, separately. These plots are based on a single measurement, so not taking into account the lack of a set of continuous orientation vectors in the Fourier domain (e.g., Hansen and Essock 2004). Right: Fourier amplitude plotted as a function of SF. Note the high similarity between intact and scrambled spectra.

Table 1

Mean Talairach coordinates (standard errors are shown in italics) of face-preferring and ventral LOC voxel clusters

	N	Mean x		Mean y		Mean z		No. of voxels	
rFFA	12	39	±1	–42	±2	–19	±1	517	±155
lFFA	12	–39	±1	–44	±2	–19	±1	346	±108
rOFA	9	34	±2	–75	±3	–9	±2	418	±195
rSTS	10	51	±1	–46	±3	5	±1	840	±219
lSTS	10	–52	±2	–52	±2	8	±2	713	±255
rAIT	9	37	±3	8	±3	–22	±2	454	±273
lAIT	7	–34	±2	3	±4	–25	±1	71	±35
rLOC	13	39	±1	–71	±1	–12	±1	1795	±458
lLOC	13	–38	±1	–75	±1	–12	±1	604	±167
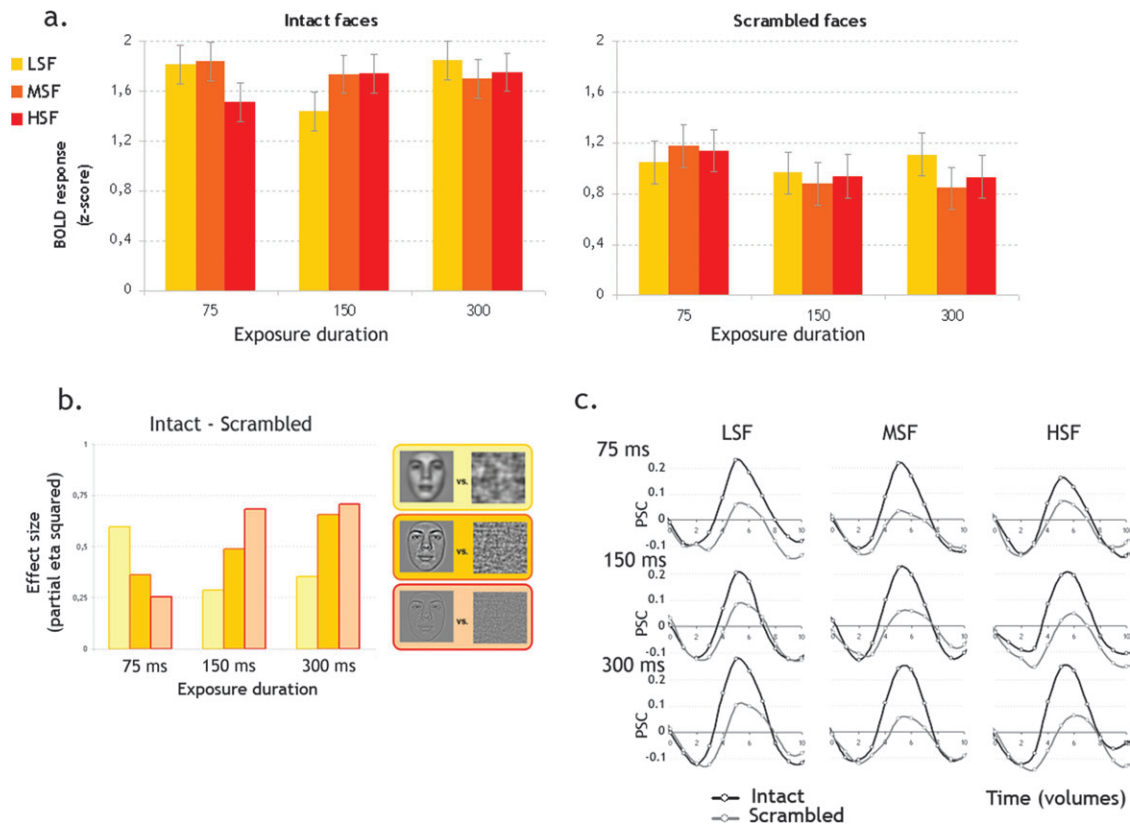
Figure 4. Average BOLD activity in the rFFA. (a) Normalized beta weights in the rFFA (bars 5 mean intrasubject variance). (b) Effect size of the difference between intact and scrambled faces in separate SF and exposure duration conditions. (c) Grand averaged event-related time course of intact and scrambled face processing in the rFFA. Time courses are expressed in percent signal change relative to fixation baseline activity (baseline interval: from 2 to ±2 TR around preparatory cue onset). The activity time courses shown on (c) reflect the findings based on the beta weights.

load) due to task difficulty. Performance accuracy in intact–scrambled categorization was at ceiling and was not influenced by SF, exposure, or stimulus factors (Figure 1b). Correct response times (computed with respect to stimulus onset) were shorter for intact than scrambled conditions ($F_{1,12} = 11.2$, $P < 0.006$), and they significantly increased at 300-ms exposure duration compared with 75- and 150-ms exposure conditions ($F_{2,24} = 18.3$, $P < 0.0001$).

Furthermore, all conditions were randomly interleaved within a run, ruling out SF differences in terms of cognitive strategies as alternative accounts of our findings. In addition, all conditions were perfectly matched with respect to masking parameters and physical properties of the stimuli (i.e., luminance, RMS contrast, orientation, and SF composition, see Methods) such that our findings are not influenced by low-level visual processing differences and therefore can be related to the high-level processing of face information.

Coarse-to-Fine Processing in the rFFA

Individual rFFAs were defined based on an independent localizer and a standard comparison of activations between faces and cars (see Methods). The omnibus ANOVA revealed a significant main effect of stimulus as intact faces induced larger rFFA activity than scrambled faces ($F_{1,11} = 18.2$, $P < 0.001$; Figure 4a).

In intact conditions, exposure duration significantly interacted with SF ($F_{4,44} = 2.8$, $P < 0.03$). Hence, HSF faces induced weaker

response than LSF and MSF faces ($Ps < 0.05$) at 75 ms of exposure. However, this pattern reversed for 150-ms exposure as the weakest activation was observed for LSF as compared with MSF and HSF faces ($Ps < 0.05$). In contrast, there was no difference between SF with the 300-ms-long stimuli.

These findings indicate different temporal dynamics of SF processing in rFFA. While LSF processing was initially strong and attenuated at 150 ms of exposure, HSF processing increased with exposure time. Polynomial contrasts showed a quadratic trend for activations induced by LSF stimuli across time ($P < 0.02$), confirming the strong attenuation at intermediate exposure duration. In contrast, a linear trend was found for HSF processing over exposure duration ($P < 0.04$). Importantly, none of these trends were significant in scrambled conditions ($Ps > 0.2$), suggesting that they specifically relate to the processing of complex and structured face information (Figure 4a).

In order to estimate the magnitude of BOLD response related to the processing of complex face information, we directly compared intact and scrambled conditions, in each SF and exposure condition, via planned comparisons (see Supplementary Table 1 and Figure 4b) and computed the effect size (partial η^2; see Methods) of this difference.

When stimuli were presented for 75 ms, the intact–scrambled difference was significant in LSF and MSF ($P < 0.002$ and $P < 0.03$, respectively) but not in HSF ($P = 0.08$; see Figure 4b). Even though significant in both LSF and MSF, intact–scrambled difference of activation was almost twice as large in

LSF (60% of rFFA signal variance) as in MSF (36%). At 150 ms, this pattern strikingly reversed as the intact–scrambled difference was significant in MSF and HSF ($Ps < 0.008$) but not in LSF ($P = 0.06$). Effect sizes reveal that HSF face processing explained 68% of rFFA signal variance, while signal variance related to MSF face processing was approximately 48%. The contribution of LSF at 150 ms was marginal and half as strong as in the 75-ms duration condition. After an exposure of 300 ms, intact–scrambled difference was significant in every SF (LSF: $P < 0.03$; MSF: $P < 0.0008$; HSF: $P < 0.0003$). Yet, MSF and HSF each accounted for twice a larger rFFA response variance than LSF.

These results indicate that the contribution of SF to rFFA face processing dynamically changes over time. At the shortest exposure duration, the processing of face information is strongest in LSF. At longer exposures, LSF processing decreases, whereas face processing in MSF and HSF gets more robust. The use of scrambled controls allows us to conclude that the bias observed in SF processing over time is related to high-level representations, here faces, and not to more general or low-level aspects of SF processing (see also averaged time course of rFFA activity; Figure 4*c*).

Processing of SF over Time in Other Face-preferring Regions

The above analyses focused on rFFA, which is the main cortical site assumed to be involved in the holistic processing of face identity (Schiltz and Rossion 2006). Yet, besides rFFA, other face-preferring regions have proven essential for normal face perception (Haxby et al. 2000; Rossion et al. 2003). Using the same "faces minus cars" contrast as for rFFA, face-preferring regions were individually localized in the left FFA (lFFA) as well as in bilateral STS, OFA, and AIT (Kriegeskorte et al. 2007; Rajimehr et al. 2009; see Methods). Since left OFA failed to show a significantly larger response to intact than scrambled faces in the SF experiment, it was discarded from the subsequent analyses. Full statistical analyses are presented in Supplementary Data 1.

The processing of LSF face information engaged most face-preferring regions (see Supplementary Table 1, Fig. 3*a*) at short exposure duration. At longer exposure durations (150 and 300 ms), the LSF intact–scrambled differential response was only significant in lFFA (in addition to above-mentioned rFFA). Though significant, the bilateral FFA response to LSF face information was weaker at 150 and 300 ms than at 75-ms exposure durations. These results support the coarse-to-fine hypothesis of visual processing in FFA, which assumes that LSF processing decays over time, in favor of finer-grained processing. Our results indicate that LSF face processing mainly decayed from 75 to 150 ms in bilateral FFA and largely stabilized after 150 ms of processing. Interestingly, rOFA did not engage in the processing of face information based on LSF, at any exposure duration.

Neural activation related to MSF face processing was robust in bilateral FFA and rOFA, at all durations ($Ps < 0.02$; Figure 5*a*). The temporal dynamics of MSF processing in these regions was mixed. In rFFA, MSF processing steadily increased from 75 to 300 ms, suggesting the progressive accumulation of face identity cues over time in this region. In the lFFA, MSF processing mainly increased from 150 to 300 ms of exposure. In contrast, MSF processing decreased from 75 to 150 ms of exposure in rOFA.
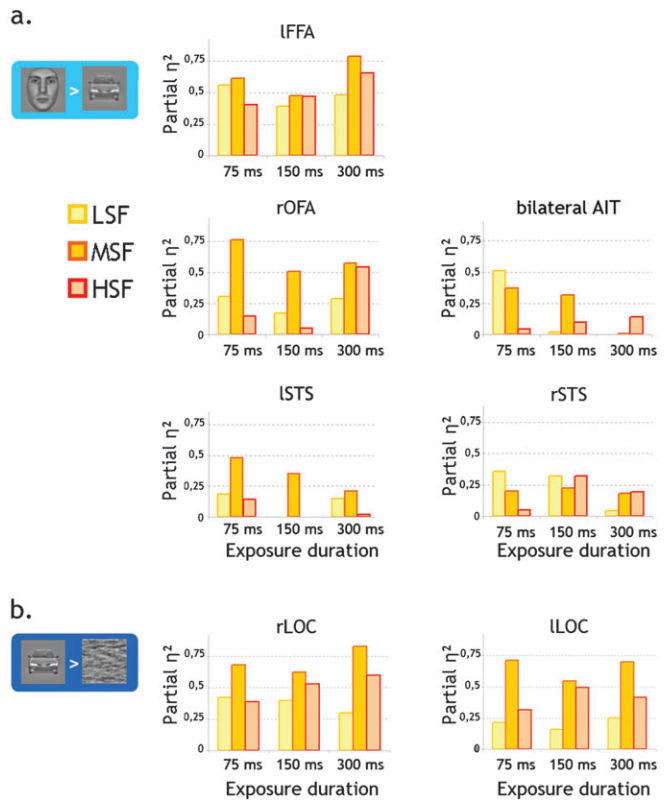
Figure 5. Effect size plots in (a) face-preferring regions (lFFA, rOFA, rSTS, lSTS, bilateral AIT) and (b) object-preferring regions (right and left ventral LOC).

In contrast to LSF and MSF, activations to HSF faces mainly spread across face-preferring regions over time; at 75 ms, the intact–scrambled difference was only significant in lFFA; at 150 ms, the intact–scrambled differential response extended to rFFA (see above); and at 300 ms, it was significant also in rOFA. Effect size estimates suggest that HSF processing temporal dynamics differed across these regions. In rFFA, the processing of HSF face content became more robust from 75 to 150 ms of exposure duration, whereas it mainly strengthened from 150 to 300 ms of exposure duration in rOFA and lFFA.

Bilateral AIT failed to show a coarse-to-fine profile over time. Actually, the intact–scrambled contrast was only significant for brief LSF stimuli. Intact and scrambled conditions did not differ in any other condition. This finding indicates that anterior face-preferring clusters of the ventral pathway are mostly responsive to brief and coarse input. As for left STS, it mainly activated to short MSF stimuli; it did not reveal any trend for coarse-to-fine processing dynamics.

Coarse-to-fine models of vision predict that processing resources dedicated to the processing of LSF input initially dominate but then progressively decrease, while they become increasingly devoted to the processing of finer spatial scales over processing time. Our findings largely corroborate this view as most face-preferring regions disclosed coarse-to-fine temporal dynamics (see Supplementary Table 1). Neural activity to LSF was strong in early stages of visual processing but decayed as a function of time (mostly until 150 ms of processing). Moreover, the processing of HSF face information strengthened at different temporal intervals depending on the

region. In contrast, neural responses to MSF face information were strong in bilateral FFA and rOFA, already at the shortest exposure duration.

These findings were confirmed by a whole-brain analysis of intact–scrambled differential activations (see Supplementary Data 2).

No Coarse-to-Fine Processing in Ventral LOC

Do the spatiotemporal processing dynamics observed in the face-preferring network, especially in rFFA, apply to high-level, noncategory-preferring, visual regions? To answer this question, the ventral LOC was localized using an "intact cars minus scrambled cars" contrast in each individual subject (see Methods). This region is a more general-purpose high-level visual area as it responds to any shape with no preference for a given category (Malach et al. 1995). As a matter of fact, there was no difference of activation between intact faces and cars in bilateral LOC ($Ps > 0.2$).

As expected, lLOC and rLOC responded more strongly to intact than scrambled faces in the SF experiment (rLOC: $F_{1,12} = 27.3$, $P < 0.0002$; lLOC: $F_{1,12} = 22.63$, $P < 0.0005$; see Figure 5b). Both regions were largely driven by MSF and HSF at any exposure duration. In contrast to face-preferring regions, there was no larger BOLD response to LSF than to HSF in initial stages of processing (see Supplementary Table 1).

In the face-preferring network, both the whole-brain and the ROIs analyses revealed that distinct SF were processed at different time points during the processing of face information (see Supplementary Figure 1 and Supplementary Table 2). Precisely, LSF processing was initially strong but was progressively attenuated, while BOLD responses to HSF face information increased over time. In contrast, MSF processing was robust at all durations in most face-preferring regions. Importantly, LSF and HSF spatiotemporal dynamics did not generalize to the adjacent LOC regions, which are engaged in general aspects of object encoding. However, a marked advantage for processing MSF information was observed in LOC at all durations, indicating that the large response to MSF is a general trait of high-level visual processing.

Discussion

The present study shows, for the first time, that the human brain regions responsible for high-level face representations rely on different SF over time. The temporal dynamics of SF processing were coarse to fine in most face-preferring regions.

Coarse-to-fine models of visual processing propose that LSF are extracted mainly in the first stages of visual processing. Accordingly, all face-preferring regions (but the rOFA) robustly responded to LSF in early stages of visual processing (until 75 ms of exposure duration), and this response decayed over time (mostly until 150 ms of processing). Coarse-to-fine models further suggest that visual processing becomes finer grained over processing time and increasingly relies on the processing of HSF information. Indeed, the processing of HSF face information got more robust over time in bilateral FFA and rOFA. In contrast, MSF face processing was strong in bilateral FFA and rOFA, already at the shortest exposure duration. Neural activity related to MSF face processing increased over time in bilateral FFA (though in different temporal intervals in the 2 hemispheres), while it decreased in rOFA.

Interestingly, these spatiotemporal processing dynamics revealed in face-preferring cortex were not observed in LOC, a high-level visual region showing no visual object category preference. This suggests that coarse-to-fine processing is a special signature of category-preferring brain regions (but see below). This would agree with Bar's theoretical framework, which proposes that inferences generated in prefrontal cortex based on early LSF input are sent back to high-level/category-preferring regions of the ventral pathway to guide visual processing (e.g., Bar 2007). In contrast, the HSF content of a scene is thought to be processed in posterior visual regions, which projects on category-preferring regions of the ventral pathway. Accordingly, LOC may not belong to the coarse-to-fine network of visual processing and may rather engage in the slow encoding of fine image content. As a matter of fact, bilateral LOC responded more robustly to MSF and HSF than to LSF, irrespective of exposure duration.

Past fMRI studies investigated coarse-to-fine processing dynamics using nonface stimuli such as scenes and objects. In a recent combined fMRI and event-related potentials study, Peyrin et al. (2010) presented SF-filtered natural scenes sequences. Sequences followed either a coarse-to-fine (i.e., LSF-to-HSF) or a fine-to-coarse (i.e., HSF-to-LSF) order. They showed that coarse-to-fine sequences induce an initial increase of activity in prefrontal cortex, followed by enhanced occipital responses to HSF. However, it is unclear from this and previous studies by the same authors (Peyrin et al. 2005) whether the scene-preferring high-level regions situated in parahippocampal gyrus (Epstein et al. 1999) would also show a coarse-to-fine dynamic over processing time. Indeed, scene-preferring regions were not explored by Peyrin and colleagues.

Studies by Bar and colleagues also addressed the question of coarse-to-fine processing in the human brain. One study of this group (Kveraga et al. 2007) suggested that prefrontal regions, thought to facilitate visual processing via feedback, receive visual input from primary visual cortex very rapidly after stimulus onset via M pathway. Counter intuitively, however, the authors reported larger prefrontal deactivations to HSF than LSF stimuli (Bar et al. 2006). Bar and colleagues mostly explored temporal dynamics in prefrontal regions; they did not address whether activations in object-preferring regions follow a coarse-to-fine temporal dynamic. More generally, the findings of Bar et al. do not provide unequivocal evidence of coarse-to-fine processing in the human brain, for several reasons (see Hegde 2008). First, Bar's framework relies on the unwarranted assumption that M pathway selectively carries LSF information; however, this assumption is not supported by the literature (Kaplan 2004). Moreover, luminance and contrast largely differed between LSF and HSF stimuli used by Bar et al. (2006), whereas it was highly similar between unfiltered and LSF stimuli. The differential activations observed across SF in prefrontal cortex may thus be due to these differences in input properties rather than spatial scale *per se*.

The present study reports evidence for coarse-to-fine processing in high-level visual face-preferring regions while strictly equating stimulus and cognitive properties across SF conditions. Coarse-to-fine strategy may apply more to the processing of faces than other object categories, for several reasons. First, behavioral and fMRI evidence jointly indicate that face processing is more largely dependent on SF than object processing (Collin et al. 2004; Yue et al. 2006; Williams et al. 2009). It has been suggested that especially for faces, the

SF-dependent representations generated in primary visual cortex are kept segregated at high-level processing stages (Biederman 1987). Second, and in relation to the previous point, previous publications showed that holistic processing relies on the processing of LSF face information (Collishaw and Hole 2000; Goffaux et al. 2003, 2005; Goffaux and Rossion 2006; Goffaux 2009; though see Cheung et al. 2008). Holistic processing emerges very early during face processing (Richler et al. 2009; see also Singer and Sheinberg 2006). It plays a key role in, and is highly specific for, face perception. When holistic processing is disrupted, face recognition is dramatically impaired (Sergent and Signoret 1992; Barton et al. 2002; though see Konar et al. 2010). Schiltz and Rossion (2006) showed that holistic face representations emerge in high-level face-preferring visual cortex and especially in the rFFA. We speculate that the early and strong rFFA responses to LSF face information observed in the present study may serve the generation of holistic face representations. However, further research is needed to support this proposal. The key contribution of LSF to early face perception is also indicated by the observation that the human N170, that is an electrophysiological component known to be stronger in response to faces than other visual categories (Rossion et al. 2000), is stronger in response to LSF faces than HSF faces (Goffaux et al. 2003; Flevaris et al. 2008). Another aspect that likely favors the coarse-to-fine strategy for faces is related to development. Faces are ubiquitous in human visual environment since the first minutes of life and newborns show an exceptional ability to discriminate faces. Due to the immaturity of their visual system, newborns individuate faces mainly based on LSF (de Heering et al. 2008). The predominance of LSF-based face processing early in life may contribute to the importance of this band of information for the early processing stages in adulthood (Le Grand et al. 2001).

Given that face perception is more affected by SF content than the processing of other visual categories (see above), it is unclear whether our findings can be generalized to other high-level, category-preferring regions. However, since coarse-to-fine processing has been evidenced with simple stimuli (e.g., Watt 1987; Bredfeldt and Ringach 2002), and complex visual stimuli like natural scenes (e.g., Peyrin et al. 2010), one might speculate that it generalizes to low- and high-processing levels of vision. Nevertheless, further research is necessary to tackle this issue.

Our results resolve the empirical divergence between the past fMRI explorations of SF processing in face-preferring cortical regions. While some papers reported overall larger BOLD responses to HSF than LSF (Vuilleumier et al. 2003; Eger et al. 2004; Iidaka et al. 2004), others observed no BOLD response difference between LSF and HSF (Gauthier et al. 2005). Despite the pervasive assumption that SF processing is time-dependent, the potential role of exposure duration was not addressed in any of these earlier studies. The strong initial response to LSF face information has thus likely been hindered in the past studies, which used long exposure duration (>200 ms).

Another important new finding relates to the large cortical response measured in some face-preferring regions in response to MSF face information, already at the shortest exposures. This finding is without precedent since no fMRI study explored cortical processing of face MSF information so far. In contrast to LSF, robust MSF responses were also observed early in bilateral LOC. They may thus reflect the general peak of human visual acuity centered at intermediate SF (De Valois et al. 1974; Tanskanen et al. 2005).

Our finding that BOLD responses to faces depend on different SF over time suggests that face-preferring cells tune to different SF ranges of face information. Accordingly, our whole-brain analyses (Supplementary Data 2) indicate that different voxel clusters respond to distinct ranges of SF, suggesting that SF are segregated until high-level stages of face processing. This is further supported by electrophysiological studies in the monkey brain, showing that face-preferring cells located in the inferotemporal cortex are sensitive to SF (Rolls et al. 1985; Bermudez et al. 2009).

The present evidence suggests the coarse-to-fine strategy as a plausible modus operandi in high-level visual cortex. Because LSF are processed earlier than—and independently from—HSF, they may be used for an initial coarse segmentation of the stimulus, to be later refined by the slower accumulation of HSF information. This is further supported by electrophysiological evidence in the monkey brain that inferotemporal cells respond to the global, coarse image structure before encoding local, fine information (Sugase et al. 1999; Sripati and Olson 2009). By revealing the spatial and temporal dynamics in high-level visual cortex dedicated to face perception, the present study opens a new avenue for investigating the composition of high-level visual representations in the human brain (see Hegde 2008).

Supplementary Material

Supplementary data can be found at: http://www.cercor.oxfordjournals.org/.

References

Allen EA, Freeman RD. 2006. Dynamic spatial processing originates in early visual pathways. J Neurosci. 26:11763-11774.

Baker CI, Hutchison TL, Kanwisher N. 2007. Does the fusiform face area contain subregions highly selective for nonfaces? Nat Neurosci. 10:3-4.

Bar M. 2007. The proactive brain: using analogies and associations to generate predictions. Trends Cogn Sci. 11:280-289.

Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Dale AM, Hamalainen MS, Marinkovic K, Schacter DL, Rosen BR, et al. 2006. Top-down facilitation of visual recognition. Proc Natl Acad Sci U S A. 103:449-454.

Barton JJ, Press DZ, Keenan JP, O'Connor M. 2002. Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. Neurology. 58:71-78.

Bermudez MA, Vicente AF, Romero MC, Perez R, Gonzalez F. 2009. Spatial frequency components influence cell activity in the inferotemporal cortex. Vis Neurosci. 26:421-428.

Bex PJ, Makous W. 2002. Spatial frequency, phase, and the contrast of natural images. J Opt Soc Am A Opt Image Sci Vis. 19:1096-1106.

Biederman I. 1987. Recognition-by-components: a theory of human image understanding. Psychol Rev. 94:115-147.

Biederman I, Kalocsai P. 1997. Neurocomputational bases of object and face recognition. Philos Trans R Soc Lond B Biol Sci. 352:1203-1219.

Boynton GM, Engel SA, Glover GH, Heeger DJ. 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci. 16:4207-4221.

Bredfeldt CE, Ringach DL. 2002. Dynamics of spatial frequency tuning in macaque V1. J Neurosci. 22:1976-1984.

Bullier J. 2001. Integrated model of visual processing. Brain Res Rev. 36:96-107.

Cheung OS, Richler JJ, Palmeri TJ, Gauthier I. 2008. Revisiting the role of spatial frequencies in the holistic processing of faces. J Exp Psychol Hum Percept Perform. 34:1327-1336.

Collin CA, Liu CH, Troje NF, McMullen PA, Chaudhuri A. 2004. Face recognition is affected by similarity in spatial frequency range to a greater degree than within-category object recognition. J Exp Psychol Hum Percept Perform. 30:975-987.

Collishaw SM, Hole GJ. 2000. Featural and configurational processes in the recognition of faces of different familiarity. Perception. 29:893-909.

Dakin SC, Hess RF, Ledgeway T, Achtman RL. 2002. What causes non-monotonic tuning of fMRI response to noisy images? Curr Biol. 12:R476-R477.; author reply R478.

de Heering A, Turati C, Rossion B, Bulf H, Goffaux V, Simion F. 2008. Newborns' face recognition is based on spatial frequencies below 0.5 cycles per degree. Cognition. 106:444-454.

De Valois RL, Albrecht DG, Thorell LG. 1982. Spatial frequency selectivity of cells in macaque visual cortex. Vision Res. 22:545-559.

De Valois RL, Morgan HC, Snodderly DM. 1974. Psychophysical studies of monkey vision—III. Spatial luminance contrast sensitivity tests of macaque and human observers. Vision Res. 14:75-81.

Eger E, Schyns PG, Kleinschmidt A. 2004. Scale invariant adaptation in fusiform face-responsive regions. Neuroimage. 22:232-242.

Epstein R, Harris A, Stanley D, Kanwisher N. 1999. The parahippocampal place area: recognition, navigation, or encoding? Neuron. 23:115-125.

Field DJ. 1987. Relations between the statistics of natural images and the response properties of cortical cells. J Opt Soc Am A. 4:2379-2394.

Fiser J, Subramaniam S, Biederman I. 2001. Size tuning in the absence of spatial frequency tuning in object recognition. Vision Res. 41:1931-1950.

Flevaris AV, Robertson LC, Bentin S. 2008. Using spatial frequency scales for processing face features and face configuration: an ERP analysis. Brain Res Rev. 1194:100-109.

Frazor RA, Albrecht DG, Geisler WS, Crane AM. 2004. Visual cortex neurons of monkeys and cats: temporal dynamics of the spatial frequency response function. J Neurophysiol. 91:2607-2627.

Gauthier I, Curby KM, Skudlarski P, Epstein RA. 2005. Individual differences in FFA activity suggest independent processing at different spatial scales. Cogn Affect Behav Neurosci. 5:222-234.

Goffaux V. 2009. Spatial interactions in upright and inverted faces: re-exploration of spatial scale influence. Vision Res. 49:774-781.

Goffaux V, Gauthier I, Rossion B. 2003. Spatial scale contribution to early visual differences between face and object processing. Brain Res Cogn Brain Res. 16:416-424.

Goffaux V, Hault B, Michel C, Vuong QC, Rossion B. 2005. The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. Perception. 34:77-86.

Goffaux V, Rossion B. 2006. Faces are "spatial"—holistic face perception is supported by low spatial frequencies. J Exp Psychol Hum Percept Perform. 32:1023-1039.

Goffaux V, Rossion B, Sorger B, Schiltz C, Goebel R. 2009. Face inversion disrupts the perception of vertical relations between features in the right human occipito-temporal cortex. J Neuropsychol. 3:45-67.

Gold J, Bennett PJ, Sekuler AB. 1999. Identification of band-pass filtered letters and faces by human and ideal observers. Vision Res. 39:3537-3560.

Halit H, de Haan M, Schyns PG, Johnson MH. 2006. Is high-spatial frequency information used in the early stages of face detection? Brain Res. 1117:154-161.

Hansen BC, Essock EA. 2004. A horizontal bias in human visual processing of orientation and its correspondence to the structural components of natural scenes. J Vis. 10:1044-60.

Haxby JV, Hoffman EA, Gobbini MI. 2000. The distributed human neural system for face perception. Trends Cogn Sci. 4:223-233.

Hegde J. 2008. Time course of visual perception: coarse-to-fine processing and beyond. Prog Neurobiol. 84:405-439.

Hess RF. 2004. Spatial scale in visual processing. In: Chalupa LM, Werner JS, editors. The visual neurosciences. Cambridge (MA): MIT Press. p. 1043-1059.

Hochstein S, Ahissar M. 2002. View from the top: hierarchies and reverse hierarchies in the visual system. Neuron. 36:791-804.

Hughes HC, Nozawa G, Kitterle F. 1996. Global precedence, spatial frequency channels, and the statistics of natural images. J Cogn Neurosci. 8:197-230.

Iidaka T, Yamashita K, Kashikura K, Yonekura Y. 2004. Spatial frequency of visual image modulates neural responses in the temporo-occipital lobe. An investigation with event-related fMRI. Cogn Brain Res. 18:196-204.

Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci. 17:4302-4311.

Kaplan E. 2004. The M, P and K pathways of the primate visual system. In: Chalupa LM, Werner JS, editors. The visual neurosciences. Cambridge (MA): MIT Press. p. 1043-1059.

Keysers C, Perrett DI. 2002. Visual masking and RSVP reveal neural competition. Trends Cogn Sci. 6:120-125.

Konar Y, Bennett PJ, Sekuler AB. 2010. Holistic processing is not correlated with face-identification accuracy. Psychol Sci. 21:38-43.

Kriegeskorte N, Formisano E, Sorger B, Goebel R. 2007. Individual faces elicit distinct response patterns in human anterior temporal cortex. Proc Natl Acad Sci U S A. 104:20600-20605.

Kveraga K, Boshyan J, Bar M. 2007. Magnocellular projections as the trigger of top-down facilitation in recognition. J Neurosci. 27:13232-13240.

Le Grand R, Mondloch CJ, Maurer D, Brent HP. 2001. Neuroperception. Early visual experience and face processing. Nature. 410:890.

Lerner Y, Hendler T, Ben-Bashat D, Harel M, Malach R. 2001. A hierarchical axis of object processing stages in the human visual cortex. Cereb Cortex. 11:287-297.

Liu CH, Collin CA, Rainville SJ, Chaudhuri A. 2000. The effects of spatial frequency overlap on face recognition. J Exp Psychol Hum Percept Perform. 26:956-979.

Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB. 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc Natl Acad Sci U S A. 92:8135-8139.

Marr D. 1982. Vision: a computational investigation into the human representation and processing of visual information. San Francisco (CA): Freeman.

Mazer JA, Vinje WE, McDermott J, Schiller PH, Gallant JL. 2002. Spatial frequency and orientation tuning dynamics in area V1. Proc Natl Acad Sci U S A. 99:1645-1650.

McCarthy G, Puce A, Belger A, Allison T. 1999. Electrophysiological studies of human face perception. II: response properties of face-specific potentials generated in occipitotemporal cortex. Cereb Cortex. 9:431-444.

Mihaylova M, Stomonyakov V, Vassilev A. 1999. Peripheral and central delay in processing high spatial frequencies: reaction time and VEP latency studies. Vision Res. 39:699-705.

Musselwhite MJ, Jeffreys DA. 1985. The influence of spatial frequency on the reaction times and evoked potentials recorded to grating pattern stimuli. Vision Res. 25:1545-1555.

Narayan R, Ergun A, Sen K. 2005. Delayed inhibition in cortical receptive fields and the discrimination of complex stimuli. J Neurophysiol. 94:2970-2975.

Nasanen R. 1999. Spatial frequency bandwidth used in the recognition of facial images. Vision Res. 39:3824-3833.

Parker DM, Dutch S. 1987. Perceptual latency and spatial frequency. Vision Res. 27:1279-1283.

Parker DM, Lishman JR, Hughes J. 1992. Temporal integration of spatially filtered visual images. Perception. 21:147-160.

Parker DM, Lishman JR, Hughes J. 1997. Evidence for the view that temporospatial integration in vision is temporally anisotropic. Perception. 26:1169-1180.

Peyrin C, Mermillod M, Chokron S, Marendaz C. 2006. Effect of temporal constraints on hemispheric asymmetries during spatial frequency processing. Brain Cogn. 62:214-220.

Peyrin C, Michel CM, Schwartz S, Thut G, Seghier M, Landis T, Marendaz C, Vuilleumier P. 2010. The neural substrates and timing of top-down processes during coarse-to-fine categorization of visual scenes: a combined fMRI and ERP study. J Cogn Neurosci.

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.

Rajimehr R, Young JC, Tootell RB. 2009. An anterior temporal face patch in human cortex, predicted by macaque maps. Proc Natl Acad Sci U S A. 106:1995-2000.

Richler JJ, Mack ML, Gauthier I, Palmeri TJ. 2009. Holistic processing of faces happens at a glance. Vision Res. 49:2856-2861.

Rolls ET, Baylis GC, Leonard CM. 1985. Role of low and high spatial frequencies in the face-selective responses of neurons in the cortex in the superior temporal sulcus in the monkey. Vision Res. 25:1021-1035.

Rossion B, Caldara R, Seghier M, Schuller AM, Lazeyras F, Mayer E. 2003. A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. Brain. 126:2381-2395.

Rossion B, Gauthier I, Tarr MJ, Despland P, Bruyer R, Linotte S, Crommelinck M. 2000. The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. Neuroreport. 11:69-74.

Schiltz C, Rossion B. 2006. Faces are represented holistically in the human occipito-temporal cortex. Neuroimage. 32:1385-1394.

Schyns PG, Oliva A. 1994. From blobs to boundary edges: evidence for time and spatial scale dependent scene recognition. Psychol Sci. 5:195-200.

Sergent J, Ohta S, MacDonald B. 1992. Functional neuroanatomy of face and object processing. A positron emission tomography study. Brain. 115(1):15-36.

Sergent J, Signoret JL. 1992. Varieties of functional deficits in prosopagnosia. Cereb Cortex. 2:375-388.

Simmons WK, Bellgowan PS, Martin A. 2007. Measuring selectivity in fMRI data. Nat Neurosci. 10:4-5.

Singer JM, Sheinberg DL. 2006. Holistic processing unites face parts across time. Vision Res. 46:1838-1847.

Sripati AP, Olson CR. 2009. Representing the forest before the trees: a global advantage effect in monkey inferotemporal cortex. J Neurosci. 29:7788-7796.

Sripati AP, Yoshioka T, Denchev P, Hsiao SS, Johnson KO. 2006. Spatiotemporal receptive fields of peripheral afferents and cortical area 3b and 1 neurons in the primate somatosensory system. J Neurosci. 26:2101-2114.

Sugase Y, Yamane S, Ueno S, Kawano K. 1999. Global and fine information coded by single neurons in the temporal visual cortex. Nature. 400:869-873.

Talairach G, Tournoux P. 1988. Co-planar stereotaxic atlas of the human brain. New York: Thieme.

Tanskanen T, Nasanen R, Montez T, Paallysaho J, Hari R. 2005. Face recognition and cortical responses show similar sensitivity to noise spatial frequency. Cereb Cortex. 15:526-534.

Tolhurst DJ, Tadmor Y, Chao T. 1992. Amplitude spectra of natural images. Ophthal Physiol Opt. 12:229-232.

Vlamings PH, Goffaux V, Kemner C. 2009. Is the early modulation of brain activity by fearful facial expressions primarily mediated by coarse low spatial frequency information? J Vis. 9(12):11-13.

Vuilleumier P, Armony JL, Driver J, Dolan RJ. 2003. Distinct spatial frequency sensitivities for processing faces and emotional expressions. Nat Neurosci. 6:624-631.

Watt RJ. 1987. Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. J Opt Soc Am A. 4:2006-2021.

Williams NR, Willenbockel V, Gauthier I. 2009. Sensitivity to spatial frequency and orientation content is not specific to face perception. Vision Res. 49:2353-2362.

Yue X, Tjan BS, Biederman I. 2006. What makes faces special? Vision Res. 46:3802-3811.