Modulation of V1 Activity by Shape: Image-Statistics or Shape-Based Perception?

Serge O. Dumoulin and Robert F. Hess
McGill Vision Research Unit, Department of Ophthalmology, McGill University, Montreal, Canada

Dumoulin, Serge O. and Robert F. Hess. Modulation of V1 activity by shape: image-statistics or shape-based perception? *J Neurophysiol* 95: 3654–3664, 2006. First published March 1, 2006; doi:10.1152/jn.01156.2005. It is current dogma that neurons in primary visual cortex extract local edges from the scene from which later visual areas reconstruct more meaningful shapes. Recent neuroimaging studies, however, have shown V1 modulations by the degree of structure in the image (shape). These V1 modulations due to the level of shape coherence have been explained in one of two possible ways: due to changes in image statistics or shape-based perceptual influences from higher visual areas. Here we compare both hypotheses using stimuli composed of Gabor arrays constructed to form circular shapes that can be successively degraded by manipulating the orientations of individual Gabors while maintaining local and global statistics. In a first experiment, we confirm that V1 responses are inversely correlated with the degree of structure in the image. In a second experiment, stimulus predictions are compared based on the degree of circular shape or change in the image statistic varied (orientation variance) in the image. We find that these V1 modulations to shape change are correlated with low-level changes in orientation contrast rather than shape perception per se.

INTRODUCTION

One of the important roles of our visual system is to detect and segregate objects. Early visual neurons only process visual information in a small part of the total visual field. For example, in primary visual cortex (V1) local, oriented edges from the visual scene are extracted (Hubel and Wiesel 1959, 1962), and V1 has been considered as a bank of oriented filters (De Valois and De Valois 1988). These filters are the basis of shape perception from which later visual areas reconstruct more meaningful objects.

Recent human neuroimaging studies have identified a region in ventro-lateral occipital cortex involved in shape processing, which comprises several areas. These regions respond strongly to objects but not to textures, noise, or highly scrambled versions (Grill-Spector 2003; Kanwisher et al. 1997; Malach et al. 1995, 2002; Puce et al. 1996). At the level of V1, these structured and unstructured stimuli, used to define object-related regions, are supposed to be equated. However, recent studies have revealed modulation of V1 when comparing structured and unstructured stimuli (Grill-Spector et al. 1998; Lerner et al. 2001; Murray et al. 2002; Paradis et al. 2000; Rainer et al. 2001, 2002).

Why is V1 activity modulated by the degree of structure in the stimuli? Here we consider two possible explanations: one due to shape-based perceptual influences, the other due to processes related to changes in image statistics. The perceptual-based hypotheses proposes that higher-order areas feedback onto lower areas (e.g., V1), modulating their responses according a high-level perceptual hypothesis of the scene. The modulatory feedback could consist of subtracting out the perceptual hypothesis, thereby leaving low-level responses to signal the deviations from these hypotheses ("predictive coding") (Murray et al. 2002, 2004; Rao and Ballard 1999). Alternatively, higher areas may attenuate or amplify aspects of lower areas input consistent with a perceptual model, thus effectively sharpening the low-level response profiles ("efficient coding") (Murray et al. 2004; Seriès et al. 2003; Simoncelli and Olshausen 2001). Either explanation suggests that activity in lower visual areas will decrease when neurons in higher visual areas are able to account for more shape-based information in the scene.

The image-statistic explanation states that scrambling images perturbs low-level image statistics to which cells in V1 are sensitive. For example, image-scrambling by rearranging stimulus sections introduces new edges and changes in power-spectra (Rainer et al. 2002). Phase-scrambling preserves global but not local statistics, such as sparseness (Dakin et al. 2002; Olman et al. 2004). Even when global and local statistics have been equated, neuronal interactions beyond the classical receptive field, known to occur in early visual areas, may alter the functional magnetic resonance imaging (fMRI) signal. For example it is known that cells in V1 respond not only to the orientation information within its receptive field but also the orientation information outside the classical receptive field (Albright and Stoner 2002; Allman et al. 1985; Fitzpatrick 2000; Seriès et al. 2003). This kind of explanation may still involve feedback from higher areas but does not require a shape-based perceptual hypothesis of the scene and is solely based on the changes in image statistics. Furthermore, responses in higher visual areas may not necessarily be related to the low-level areas.

In this study, we assessed these two competing hypotheses for why the activity in area V1 can be modulated by global changes in object shape. We use tightly controlled narrowband stimuli composed of Gabors in arrangements designed to have identical local and global properties (Achtman et al. 2003). Only the orientations of the Gabors were varied to define circular shape with varying coherences. In agreement with a previous study using similar stimuli (Achtman et al. 2001), we confirm the observation that V1 responses vary inversely with

---

the amount of circular image structure. We extend these previous findings by creating random patterns that are not only matched in their local and global statistics but also have the same orientation differences between neighboring elements across the image. Using these specially constructed images, we compare predictions based on the preceding two competing explanations for modulation of V1 activity based on global shape: high-level predictive-coding versus low-level image statistics. The results suggest that V1 responses can be explained by low-level interactions beyond the classical receptive field.

METHODS

Subjects

Six experienced psychophysical observers were used as subjects (all male, mean age: 39, age range: 30–54). The subjects were instructed to fixate at a provided fixation point and trained prior to the scanning session to familiarize them with the task. All observers had normal or corrected-to-normal visual acuity. All studies were performed with the informed consent of the subjects and were approved by the Montréal Neurological Institute Research Ethics Committee.

Visual stimuli

The visual stimuli were generated in the MatLab programming environment using the PsychToobox (Brainard 1997; Pelli 1997) on a Macintosh G4 Powerbook, and displayed on a LCD projector (NEC Multisync MT820). The total visual display subtended 20°.

The stimuli consisted of oriented Gabors, i.e., a one-dimensional (1D) sine-wave enclosed in a two-dimensional (2D) Gaussian envelope ($\lambda = 0.2$ and $\sigma = 0.1°$), i.e., the spatial frequency content of the images was centered on 5 cycles/°. The stimulus was created of 625 Gabors (stimulus size: 20°, matrix size: 512 × 512). The positions of the Gabors were jittered around a square grid centered on the image matrix (grid distance: 0.8°, uniform jitter range: −0.4–0.4°). This approach creates an approximate uniform Gabor density. The actual displayed number of Gabors may be <625 because Gabors at the edge of the display may be jittered out of the image matrix. The contrast of each Gabor was randomly chosen from a uniform distribution (contrast range: 25–100%). The global orientation content was controlled to be roughly isotropic between $-\pi$ and π. Therefore the only difference between the stimulus conditions is the relative orientations between Gabors.

In experiment one, four different stimulus types were used (see Fig. 1). The Gabor array was organized to form one full circle approximately centered around the fixation. The coherence of the array was deteriorated at four levels (100, 25, 6.25, and 0%) by increasing the orientation jitter of individual Gabors to give rise to the four image categories (Achtman et al. 2003).

In the second experiment, five different stimuli types were used (see Fig. 2). In two image categories, the Gabor array formed circular shapes, either 1 (A) or 10 (B) circles. Also a random array was presented to the subjects (C). Please note that A and C are similar to D and A of Fig. 1. The next two image types were random arrays where the local orientation gradient (or contrast) was constrained to be similar to the circular shapes (D and E, respectively). The notion of orientation contrast is similar to that of curvature, which has been shown to be encoded in some visual areas, such as macaque V4 (Pasupathy and Connor 1999, 2002). We will refer to these kinds of images as "flowfield." The orientations of Gabors in these flowfields was determined by a band-pass filtered random pixel array scaled to be isotropic between 0 and 360°. This array was then used as a look-up table to determine the Gabor's orientation depending on it's xy coordinates. The orientations of all Gabors and their relative distances were used to calculate the orientation gradient of the image. The peak frequency of the band-pass function was varied to match the
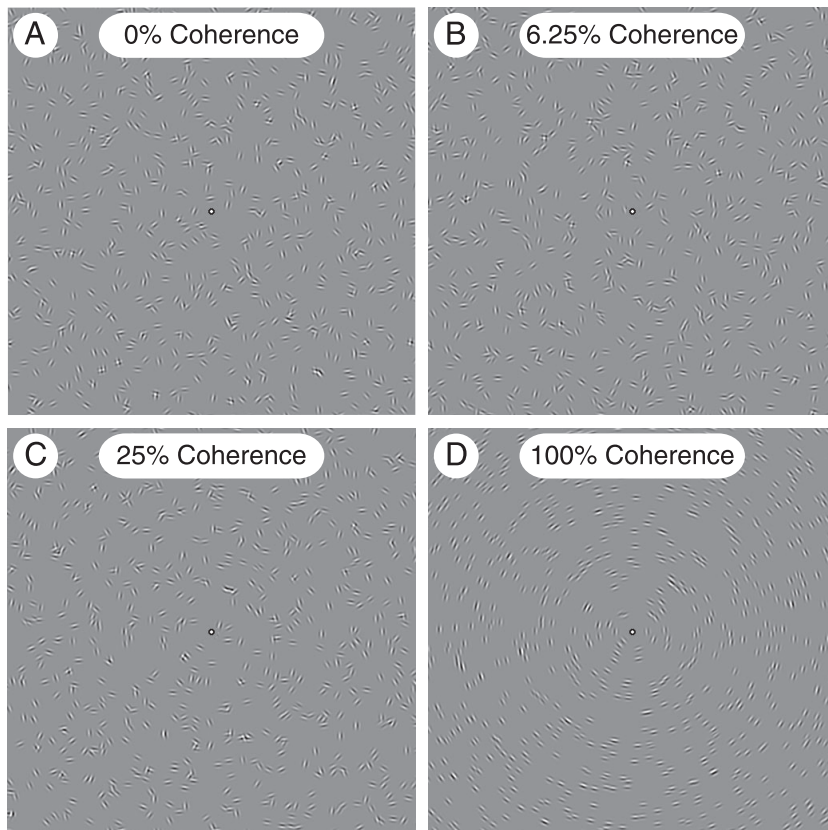

FIG. 1. Stimulus examples used in the coherence experiment. Coherence was decreased by increasing the orientation variance. Typically, global structure can be detected when about 10% of the elements are coherently oriented (Achtman et al. 2003). Here, coherence levels of 0, 6.25, 25, and 100% were used, and are shown in A–D, respectively.
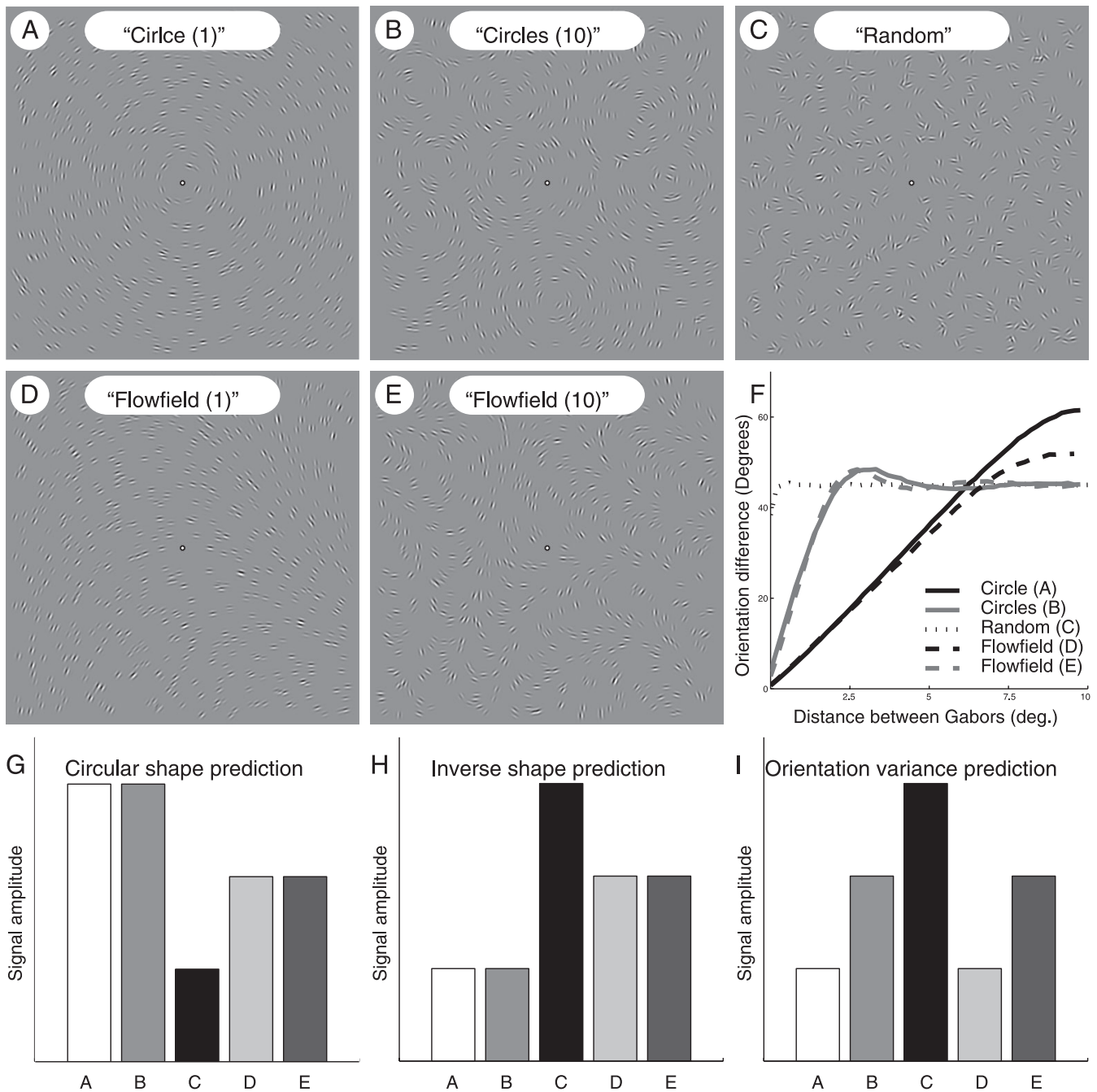
FIG. 2.   Stimulus examples used to probe the origin of V1 modulations to image structure. *A* and *B*: Gabor arrangements forming circular shapes, consisting of either 1 or 10 circles respectively. *C*: random array. *A* and *C* are similar to *D* and *A* in Fig. 1. *D* and *E*: examples from random "flowfield" patterns where the orientation gradient (or contrast) was matched to that of the circular shapes, i.e., *A* and *B*, respectively. The orientation difference, or orientation contrast, of neighboring elements of all 5 stimulus types is shown in *F*. *G–I*: relative ordering of hypothetical response profiles to these images. *G*: response profile of a higher visual area sensitive to shape. In this hypothesis, the responses are highest for the circular patterns and intermediate for the flowfield patterns. *H* and *I*: 2 hypothetical response profiles of V1. *H*: predicted profile in V1 based on the idea that higher visual areas (i.e., *G*) reduce the responses in V1 according to the amount that matches a high-level perceptual shape-based hypothesis of the scene. The proposed V1 response profile according to the orientation variance (*F*) is illustrated in *I*.

orientation gradient in the flowfields with the corresponding circular array. The orientation difference (or contrast) between two Gabor elements was computed by taking the absolute difference of their respective orientations. Due to the circular nature of orientation, the orientation difference was wrapped so that the minimum and maximum orientation difference between two elements is 0 and 90°, respectively. The orientation difference, as well as their respective distance, was computed for every possible element combination. The

average orientation difference as a function of element distance is shown in *F* for all five image categories.

In *G–I* of Fig. 2, fMRI signal change predictions are shown, the amplitudes of these predictions reflect the relative orderings of the signal amplitudes. The predictions in *G* and *H* are based on the amount of circular shape present in each image category where circular shape is the highest in the images containing circles and intermediate in the flowfield images. The circular shape prediction can

be formalized by a putative circular shape detector as suggested by Achtman et al. (2003). This putative shape detector is shown in Fig. 3*A*. The model sums the responses of individual elements tangential to the orientation of a circle, independent of element phase or position. The average response of this model to the images in Fig. 2, *A–E,* is shown in *B.* In this model, the responses are the highest for the images containing circles, the intermediate for the flowfields and the lowest for the random pattern. For more details on the model, see APPENDIX A. *H* indicates the reverse pattern of *G,* which might be hypothesized to occur in area V1, where the activity of lower areas is reduced according to the amount of grouping processes in higher regions (i.e., *G*). *I,* illustrates the fMRI amplitude prediction for areas that respond according to the amount of orientation contrast in the images (see *F*). The fMRI data will be correlated with these predictions.

The different stimulus conditions were alternated in a block design (block duration 18 or 12 s, in *experiments 1* and *2,* respectively). Each condition (block) was repeated at least five times giving a total duration of ~6 min per scan. The stimuli were presented time-locked to the acquisition of fMRI time frames, i.e., every 3 s. To control for attention, the subjects continuously performed a two-interval forced-choice (2IFC) contrast-discrimination task. That is, a given stimulus presentation consisted of two intervals, both displaying a different image from the same condition either at full or reduced (×0.7) contrast. The subject indicated which interval contained the high-contrast stimulus. The contrast of individual Gabors in each pattern was varied therefore to perform the task the subjects were forced to attend to the entire— or at least a large part of the—image. Each image was presented for 500 ms, and the inter-stimulus interval was 500 ms. In the remaining 1.5 s, the subjects' responses were recorded. During mean luminance (blank) conditions an identical task was performed for the fixation dot. The subjects' performance was on average 75% correct.

The subjects were experienced psychophysical observers and instructed to fixate at the provided fixation point. However, to verify that the results are not due to eye movements elicited by the different stimulus conditions, we recorded the subjects' eye movements while performing an identical task outside the scanner using a 50-Hz video eyetracker (Cambridge Research Systems). Eye movements were measured during both stimulus and inter-stimulus intervals. An estimate of the maximum eye movement per stimulus was computed by taking the maximum difference between the eye positions during the stimulus presentation (25 estimates at 50 Hz) and the mean eye position during the subsequent inter-stimulus interval. We obtained 100 measurements for each stimulus type (see Fig. 2) per subject. The results are shown in Fig. 4 for three subjects. No differences in eye movements between the different conditions were observed (ANOVA $P \gg 0.05$).

FIG. 4. Average eye-movements and SD of 3 subjects. No differences between the different stimulus conditions were found.

Magnetic resonance imaging

The magnetic resonance images were acquired with a Siemens Sonata 1.5T MRI. The experiments were conducted with the subjects lying on their back with a surface coil (circularly polarized, receive only) centered over their occipital poles. Head position was fixed by means of a foam head rest and a bite bar.

Multislice T2*-weighted gradient echo (GE) echo-planar imaging (EPI) functional MR images (TR/TE: 3000/51 ms, flip angle: 90°, slices: 30 (contiguous), slice thickness: 4 mm) were acquired using a surface-coil (receive only) with a 64×64 acquisition matrix and a 256×256-mm rectangular field of view. The slices were taken parallel to the calcarine sulcus and covered the entire occipital and parietal lobes and large dorsal-posterior parts of the temporal and frontal lobes. One hundred and twenty-eight measurements (time frames) were acquired. Ten to 14 fMRI scans were performed in each session. T1-weighted anatomical MR images (aMRI) were acquired prior to the commencement of the functional scans. This aMRI utilized a three-dimensional (3D) GE sequence (TR: 22 ms, TE: 9.2ms, flip angle: 30°, 256×256-mm rFOV) and yielded 80 saggital images with a thickness of 2 mm.

In separate sessions, T1-weighted aMRI images were acquired with a head coil, also with a 3D GE sequence, yielding 160 saggital images comprising 1 mm^3 voxels. Identification of the visual areas was also performed in another separate session with identical parameters.

Processing of anatomical images

The anatomical MRI scans were corrected for intensity nonuniformity (Sled et al. 1998) and automatically registered (Collins et al. 1994) in a stereotaxic space (Talairach and Tournoux 1988). The surface-coil aMRI, taken with the functional images, was aligned with the head-coil aMRI, thereby allowing an alignment of the functional data with a head-coil MRI and subsequently stereotaxic space. This alignment was performed with an automated script combining correction for the intensity gradient in the surface-coil aMRI (Sled et al. 1998) and intra-subject registration (Collins et al. 1994). A validation of this method was described in a previous study (Dumoulin et al. 2000). The aMRIs were classified into gray-matter, white-matter, and cerebrospinal fluid (Zijdenbos et al. 2002), after which two cortical surfaces were automatically reconstructed at the inner and outer edge of the cortex (MacDonald et al. 2000). The surface normals of the cortical models were smoothed to produce an "unfolded" model of the cortical sheet (MacDonald et al. 2000). All processing steps were
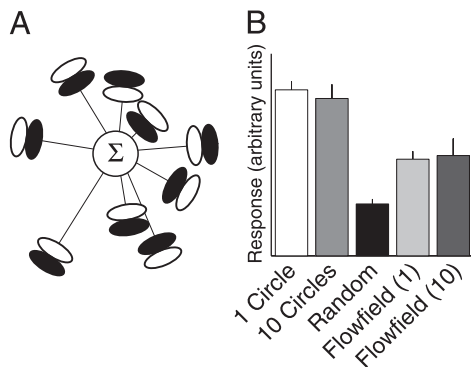
FIG. 3. Responses of putative circular shape model to the stimuli in Fig. 2. *A*: putative circular shape pattern detector proposed as by Achtman et al. (2003). The pattern detector pools across subunits with orientations tangential to a circle independent of their position. *B*: average responses and SDs ($n = 25$) of the model in *A* to the 5 stimulus conditions. The relative height of the response indicates the amount of circular shape in each stimulus.
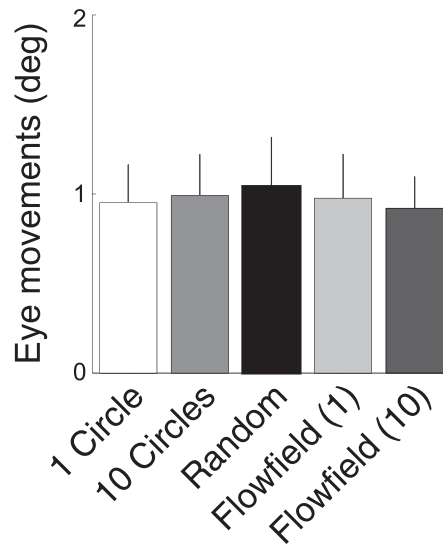
completely automatic and all the data are presented in a stereotaxic space (Collins et al. 1994; Talairach and Tournoux 1988).

Preprocessing of functional images

The first two time-frames of each functional run were discarded due to start-up magnetization transients in the data. All remaining time frames were blurred with an isotropic 3D Gaussian kernel (full-width-half-maximum: 6 mm) to attenuate high-frequency noise. The functional scans were corrected for subject motion within and between fMRI scans (Collins et al. 1994).

Identification of visual areas

Early visual cortical areas were identified using volumetric phase-encoded retinotopic mapping (COBRA package) (Dumoulin et al. 2003). By combining eccentricity and polar-angle phase maps with the anatomical MRI, the visual field signs of different visual areas could be segmented. Neighboring visual areas could be identified due to opposite field signs; i.e., V1, V2, V3/VP, V3a, and V4 (Dumoulin et al. 2003; Sereno et al. 1995). These visual areas were identified in each subject. Besides these retinotopic areas, another higher-level visual area in ventral occipital cortex (VO) was identified. This region was defined in the first experiment as a group of contiguous ventral occipital voxels beyond V4 that responded stronger to coherent stimuli than incoherent ones. We refrain from using the term LOC (lateral occipital complex) because LOC is typically defined by natural objects.

Statistical analysis

The fMRI data were analyzed using software developed by Worsley et al. (2002). The fMRI data were first converted to percent signal change. The statistical analysis is based on a linear model with correlated errors. The model incorporated the predictions of the hypotheses (see Fig. 2, *G–I*). Runs, sessions, and subjects were combined using a linear model with fixed effects and SDs taken from the previous analysis on individual runs. A random effects analysis was performed by first estimating the ratio of the random effects variance to the fixed effects variance, then regularizing this ratio by smoothing with a Gaussian filter. The amount of smoothing was chosen to achieve 100 effective degrees of freedom. The variance of the effect was then estimated by the smoothed ratio multiplied by the fixed effects variance to achieve higher degrees of freedom. The resulting *t*-statistical images were thresholded for peaks and cluster sizes using random field theory (Worsley et al. 1996).

The volume-of-interest analysis (VOI) of the identified visual areas (V1 to VO) was done in an identical fashion (Worsley et al. 2002).

These visual areas were identified in each subject. Prior to the statistical analysis, the time-series were converted to percent blood-oxygen-level-dependent (BOLD) signal change, and all the time series of voxels responding to all stimuli within a VOI (left and right hemispheres) were averaged together, with exclusion of voxels displaying artifacts. Because the time series were converted to percent BOLD signal change prior to the analysis, the effect size of the linear model (β) is also in percent signal change. The effects sizes and their SDs, averaged across all subjects, of each condition relative to the overall mean of the time-series are plotted in the Figs. 6 and 7.

RESULTS

In global shape discrimination for these kinds of patterns, sensitivity is the highest for circular shape (Achtman et al. 2003; Kovács and Julesz 1993; Levi and Klein 2000; Wilson et al. 1997). More general, concentric shape processing has been proposed as an important aspect of intermediate shape processing (Gallant et al. 1993, 1996; Wilkinson et al. 2000; Wilson and Wilkinson 1998; Wilson et al. 1997). Orientation jitter is the primary array parameter that decreased shape perception significantly (Achtman et al. 2003). Therefore in this fMRI study, the patterns consisted of circles and the orientation of the individual elements were altered to modulate perceived shape perception.

In the first experiment, we aimed to replicate the observation that V1 is modulated by the degree of shape with our stimulus. Shape was manipulated by varying the coherence, i.e., amount of orientation jitter, as shown in Fig. 5. FMRI responses in higher visual areas correlate with the coherence levels. The region that responded stronger to coherent as opposed to incoherent stimuli in ventral occipital cortex beyond area V4 in Fig. 5 was defined as area VO (ventral occipital). This area VO was defined in every subject separately. Not surprisingly in the VOI analysis (Fig. 6), this area is found to respond stronger to coherent than incoherent stimuli. In area VO, the response to three images with degraded coherence (Fig. 2, *A–C*) is similar. This response profile may correspond to the perceived organization of the images, which, in absence of a forced-choice shape judgment as in Achtman et al. (2003), is most apparent only in the most coherent image (*D*). It may be of interest to note that this region may in part correspond to area LOC (lateral occipital complex), which is defined by a stronger response to whole than scrambled objects (Kanwisher et al. 1996; Malach et al. 1995). But the regions are not identical
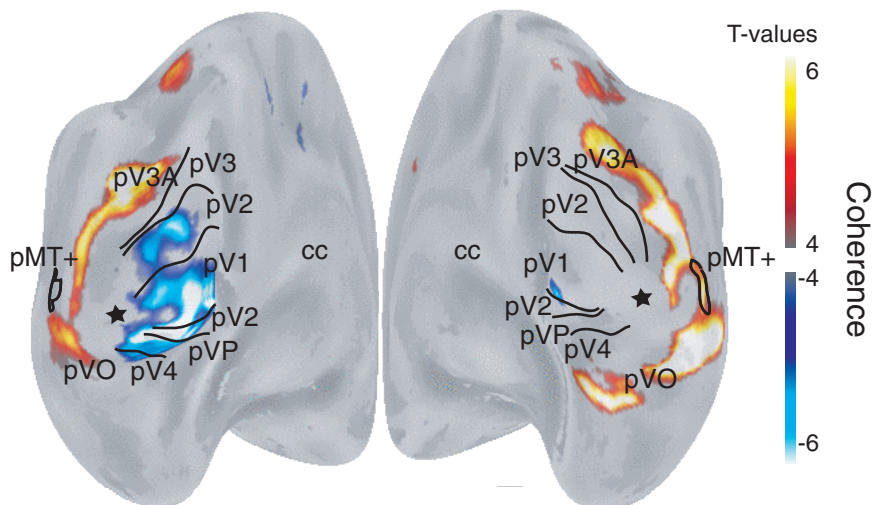
FIG. 5. Average *t*-statistical maps (4 subjects) displayed on their unfolded average cortical surfaces (*t* = 5.2 corresponds to *P* = 0.05 corrected for multiple comparisons). Functional magnetic resonance imaging (fMRI) activation, which covaried with the coherence level, is shown. Oblique posterior-medial views are shown of the left and right hemisphere. These views reveal all differential activations. The average borders of the visual areas are drawn (—), thus describing the probabilistic (*P*) location of the visual areas. The foveal representation is indicated (★). As can be seen from these *t*-statistical maps, later visual areas increase their responses with increasing coherence, whereas early visual areas (V1) respond in an inverse fashion.
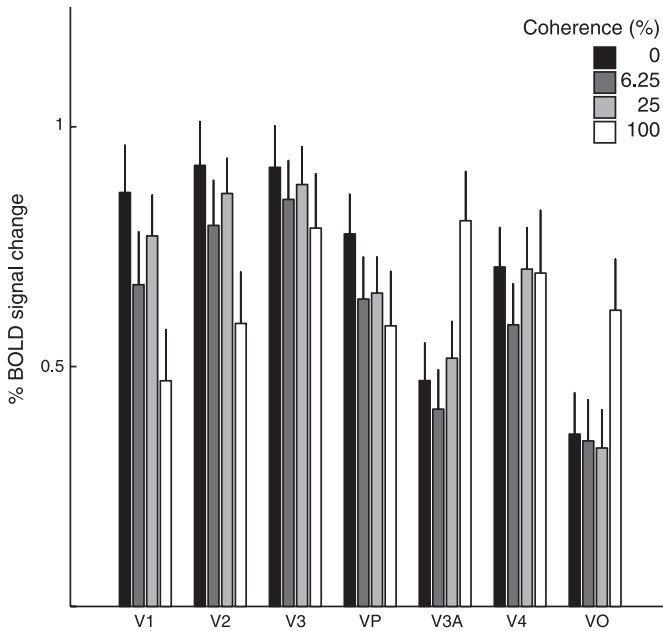
FIG. 6. Average blood-oxygen-level-dependent (BOLD) signal amplitudes and SDs elicited stimuli of different coherence levels are plotted for the identified visual areas. The bar graph reveals an inverse dependence of shape coherence in V1 (negative linear trend and a positive trend with shape coherence in area VO).

TABLE 1. *t-values and corresponding P values for the identified brain regions for a correlation with the coherence of the presented images*

Area	*t*-Value
V1	−2.4 (0.07)
V2	−2.2 (0.11)
V3	−0.6 (>0.7)
VP	−1.1 (>0.7)
V3A	2.7 (0.01)
V4	0.3 (>0.7)
VO	2.1 (0.14)

The *P* values in parentheses were Bonferroni corrected for the number of identified areas (7).

In the second experiment, we address the question of why V1 is inversely correlated with image structure by constructing images (unlike those in Fig. 1) for which perceived shape is not highly correlated with the change in the image statistics. For example, take the two pairs of images illustrated in Fig. 2, *A* and *D,* and *B* and *E,* they have the same orientation variance (see curves in Fig. 2*F*) and yet they differ in their perceived shape content (i.e., Fig. 2, *A* and *B,* depict circular shapes, whereas *D* and *E* only depict amorphic flow fields). Using stimuli such as these it is possible to distinguish between these two competing hypotheses by assessing how well neuroimaging signals correlate to specific predictions (Fig. 2. *G–I*) based on the to five image categories shown in Fig. 2. Three predictions are given; one based on perceived shape (Fig. 2*G*), one based on the inverse of perceived shape (Fig. 2*H*), and one based on orientational variance in the image (Fig. 2*I*).

The result of a VOI analysis in the identified visual areas is shown in Fig. 7. The percent signal change pattern elicited by
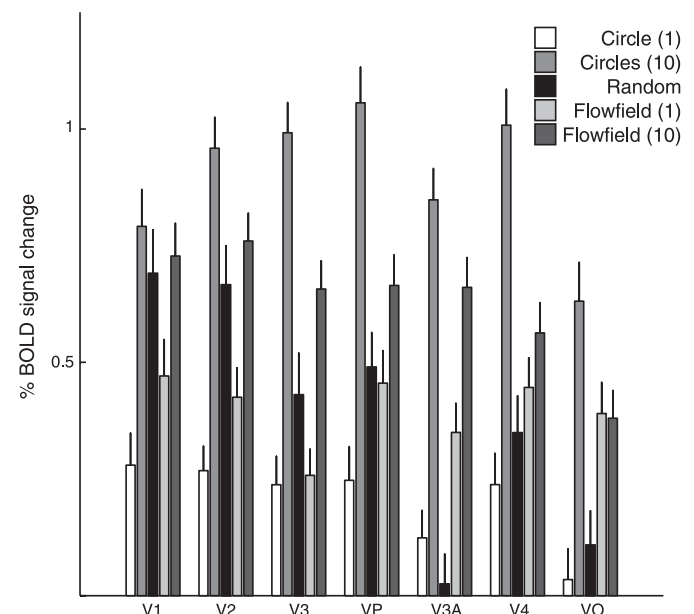
because the stimuli we use are not the same as those typically used to define LOC.

In early visual areas, however, an inverse dependence is seen. In the group stereotaxic analysis (Fig. 5), this inverse dependence reached significance only in the left hemisphere. We do not attach any significance to this apparent hemispheric difference, because, first, it is not consistent across subjects. Second, this difference may have biased by noise fluctuation, surface-coil placement, and an increased overlap of V1 in the left hemisphere (which is typically larger) in stereotaxic space for this particular group of subjects. Last, we cannot attribute these left-right differences to sensory processes per se. More specifically, even though the task was designed to encourage the subject to attend to the entire stimulus, attention to only a part of the stimulus (e.g., left or right) may still yield acceptable results. Such a constant attentional bias would not affect the comparison between different conditions but could explain the apparent hemispheric difference.

This inverse relationship is further illustrated in a VOI analysis (see Fig. 6 and Table 1), after visual areas were identified in each subject using phase-encoded mapping technology (Dumoulin et al. 2003). V1 modulation with circular shape perception has been shown in a previous study (Achtman et al. 2001) and is consistent with other studies reporting modulation of early visual cortex by the degree of structure in the image (Grill-Spector et al. 1998; Lerner et al. 2001; Malach et al. 1995; Murray et al. 2002; Rainer et al. 2001, 2002). Thus we confirm the observation that activity in primary visual cortex is modulated by the degree of image structure using a sparse stimulus that is matched in both its local and global statistics. This inverse relationship with shape coherence is in accordance with shape-based perceptual hypotheses as well as with the orientation variance in the image because shape and orientation variance are covary in these images.

FIG. 7. Average BOLD signal amplitudes and SDs elicited by the different stimuli (see Fig. 2, *G–I*) are plotted for the identified visual areas. In area V1, the bar graph reveals a profile that correlated significantly with the low-level orientation-contrast prediction but that did not correspond inversely with the high-level shape prediction. These 2 correlations were significantly different. Response profiles in higher visual areas (V3A, V4, and VO) corresponded more to the shape-predictions than the orientation-contrast prediction (see Fig. 2*G*), although this was only significant in area V3A.

TABLE 2. *t-values and corresponding P values for the identified brain regions for a correlation with the orientation-contrast prediction, the circular shape prediction, and their difference*

Area	*t*-Value		
	Orientation	Shape	Difference
V1	3.4 (<0.01)	1.5 (0.53)	2.7 (0.03)
V2	6.3 (<0.01)	4.4 (<0.01)	2.4 (0.07)
V3	6.7 (<0.01)	8.1 (<0.01)	0.5 (>0.7)
VP	5.6 (<0.01)	4.7 (<0.01)	0.5 (>0.7)
V3A	3.2 (<0.01)	6.1 (<0.01)	−2.7 (0.03)
V4	5.1 (<0.01)	5.1 (<0.01)	−1.1 (>0.7)
VO	1.6 (0.44)	2.7 (0.3)	−1.1 (>0.7)

The *P* values in parentheses were Bonferroni corrected for the number of identified areas (7).

the different conditions in V1 is consistent with the low-level orientation variance hypothesis but not with the high-level shape prediction. These two correlations were significantly different (see Table 2). These results show that primary visual cortex is not inversely correlated with the shape prediction. This differs from the previous experiment where we varied the coherence. The lack of inverse correlation suggests that the reason the primary visual cortex responds more to incoherent stimuli is not because they have less shape but because of the way local orientation is varying across space.

Responses in later visual areas (V3A, V4 and VO – the latter defined in the 1st experiment) corresponded more to the high-level shape predictions although the difference between the orientation variance prediction and shape hypothesis was only significant in area V3A (Table 2). These results show that V1 and higher visual areas do not have an inverse response profile as suggested by the predictive coding hypothesis.

In the preceding analysis, the data were fit with predetermined models, and the results will depend on these models. It may be informative to look at the correlation between V1 and other visual areas. This analysis is shown in Fig. 8. The data were normalized to remove any overall differences in percentage signal change between the areas. This was done by divid-

ing the signal changes within each area by the sum of all conditions of that area. Thus the sum of all conditions in each area will be 100%. In the first experiment (*A*), the responses to the different coherence levels are shown. It reveals the inverse relationship between V1 and higher visual areas as shown by other studies (Achtman et al. 2001; Grill-Spector et al. 1998; Lerner et al. 2001; Murray et al. 2002; Paradis et al. 2000; Rainer et al. 2001, 2002). This inverse correlation is predicted by both the shape-based perceptual and image-statistic models. In the second experiment (*B*), this inverse correlation between V1 and higher visual areas is not present. The result in the second experiment is inconsistent with shape-based perceptual models.

DISCUSSION

We describe modulation of primary visual cortex by varying degrees of structure in images (shape) consistent with previous studies (Achtman et al. 2001; Grill-Spector et al. 1998; Lerner et al. 2001; Murray et al. 2002; Paradis et al. 2000; Rainer et al. 2001, 2002). Because our stimuli have identical global and local properties, these V1 modulations cannot be explained by changes in powerspectra (Rainer et al. 2002) or sparseness (Olman et al. 2004). The pattern of percent BOLD signal change in V1 to the different image categories is distinct from that of higher cortical areas. V1 responses are more consistent with the amount of orientation variance, whereas response profiles in later visual areas correspond more to the amount of circular shape. In the perceptual models, early and late visual areas are inversely correlated. Therefore the response profiles in V1 elicited by our stimuli are inconsistent with the higher-order perceptual models suggested by Murray et al. (2002). These signal changes in V1, however, do correlate more with the orientation contrast between the image elements, indicating that interactions beyond the classical receptive field known to occur in V1 (Allman et al. 1985; Fitzpatrick 2000) may mediate these fMRI signals.

Our proposal is supported by reports that the relationship between V1 and higher cortical areas is not always inversely
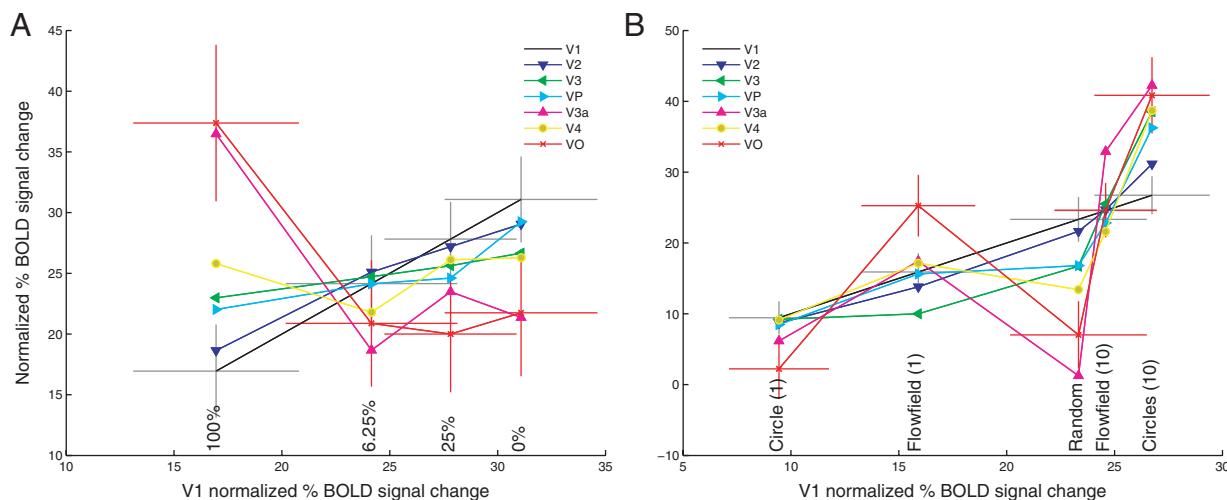

FIG. 8. Normalized percent BOLD signal change in the identified visual areas is plotted as a function of the BOLD signal change in V1 for the 2 experiments (*A* and *B*), respectively. The image categories are identified. Standard deviations are only shown for areas V1 and VO, they are not shown for the other areas to avoid clutter in the figure. In the 1st experiment (*A*), the inverse relationship is confirmed between V1 and higher visual areas. However, in the 2nd experiment (*B*), the inverse correlation does not exist between V1 and higher visual areas as would have been predicted by a shape-based perceptual model.

correlated with the degree of image structure (Rainer et al. 2001, 2002). Although the intermediate phase-blended results of Rainer et al. (2001) may be due to an artifact (Dakin et al. 2002) this would not affect the extremes, i.e., intact or phase-scrambled images. The finding that the fMRI responses to changes in image structure can elicit stronger or weaker responses than higher visual areas is not easily explained by a hypothesis based on the perceptual shape-based information. On the other hand, an image-statistic based explanation could account for these changes. For example, scrambling the images by randomizing the phases of Fourier components (Rainer et al. 2001) or blocking of stimulus parameters (Rainer et al. 2002) will affect several statistical attributes of the images in various ways. Consequently the net responses of the early visual areas sensitive to these attributes may either increase or decrease depending on the attributes involved and the neuronal sensitivity to these attributes.

Although we used synthetically generated images, higher visual areas were correlated with the degree of circular shape in the images. This finding is in agreement with the suggestion that processing circular shape is an intermediate stage in object recognition (Gallant et al. 1993, 1996; Wilkinson et al. 2000; Wilson and Wilkinson 1998; Wilson et al. 1997). It is also consistent with a texture segregation study where the textures were defined by oriented lines (Kastner et al. 2000).

Some deviations from our predictions were observed (see Figs. 2, 7, and 8). First in V1, the main deviation from the orientation-variance hypothesis is that the fMRI signal amplitudes were similar for the "random" image array as for the 10-circles array and its corresponding flowfield. If the responses were based on orientation contrast in the images, the responses to the "random" image array should have been stronger (see Fig. 2, F and I). However, this discrepancy may be explained by a maximal (saturation) response elicited by the orientation difference. More specifically, the orientation difference varies as a function of element distance (see Fig. 2F), therefore the net response based on orientation difference will depend on the distance over which this computation is done. For instance, orientation differences between elements in random images, 10-circle images and its corresponding flowfield are similar (maximal) for distances larger than 2.5°. Thus the larger the maximum distance over which the orientation difference is computed will result in more similar (and maximal) responses to the random images, the ten circle image array and its corresponding flowfield. On the other hand—or simultaneously—other processes besides orientation contrast may alter the fMRI signals. For instance, contour-integration (Field et al. 1993; Hess et al. 2003) would boost the responses to all image categories except the "random" image arrays.

The second deviation from our predictions was a decreased response to the image containing only one circle as compared with its corresponding flowfield. This deviation may be attributed to temporal aspects in our stimulus paradigm, which were not considered in our prediction. More specifically, for a particular location in visual space the orientations presented over time were random except when the stimuli consisted of one circle. For example, within blocks with stimuli constructed of one circle, similar orientations will be presented to a given spatial location and the overall shape will be identical, even though the center of the circle was jittered slightly. This may result in lower fMRI signals due to low-level orientation

specific (known to occur in V1) (e.g., Müller et al. 1999) or high-level shape specific adaptation or repetition-suppression (known to occur in higher visual areas) (e.g., Grill-Spector et al. 2006). The latter is supported by our model (see Fig. A1), where only one hypothetical detector unit is able to capture all the information in all images of the one-circle image category.

The last deviation from our models is that the 10-circles image category elicits the strongest response in virtually all visual areas. Why do all visual areas respond the strongest to the 10-circles images? A possible explained may be provided by a combination of the above-mentioned arguments. The 10-circles images contain a large degree of orientation contrast and numerous contours; taken together (or separately) they may provide an explanation why these images elicit the strongest response in lower areas such as V1. Our putative circular shape detector indicates that one (1) unit would be sufficient to code all the information in all the one-circle images, whereas all other images would require responses from a large range of units. This may provide an explanation why higher visual areas sensitive to circular shape respond the strongest to the ten-circles images, where circular shape processing is proposed to be important.

Attentional modulation can substantially affect neuroimaging responses in visual cortex, including V1 (Brefczynski and DeYoe 1999; Gandhi et al. 1999; Martinez et al. 1999; Somers et al. 1999) and higher-order object processing areas (Murray and Wojciulik 2004), and could potentially confound the interpretation of the results. Therefore in functional imaging it is crucial to control for attention. This was achieved by our contrast discrimination task, which was identical for each condition. Importantly, this task focused the subjects' attention on the images, which may increase both the gain and specificity of the neural population representing the image attributes (Murray and Wojciulik 2004). In addition, our distinct activation pattern in striate and extra-striate cortex (see Fig. 7) cannot be explained by an overall attentional boost of a particular image category. Therefore we believe that sensory, rather than attentional, processes are underlying our results.

The V1 modulations due to orientation contrast may be attributed to processes such as surround suppression and facilitation and may be considered a low-level phenomenon of perceptual grouping (Murray et al. 2004). In particular, antagonistic surrounds are a characteristic part of most neuronal tuning properties. This surround is relevant to the tuning properties of neurons for direction, velocity of motion, orientation, spatial frequency, and phase (Allman et al. 1985; Fitzpatrick 2000). Surround suppression, i.e., a response reduction due to a surrounding stimulus, has been found to decrease fMRI signals when using narrowband stimuli (Williams et al. 2003; Zenger-Landolt and Heeger 2003) and broadband stimulus patches (Kastner et al. 2001). Consistent with our study, Williams et al. (2003) described stronger decreases of the fMRI signal when the surround consisted of a similar orientation.

Our interpretation differs from Murray et al. (2002), who proposed a perceptual shape-based feedback mechanism for explaining V1 modulations to image structure. Indeed V1 activity can be modulated by high-order mechanisms such as attention (Brefczynski and DeYoe 1999; Gandhi et al. 1999; Martinez et al. 1999; Somers et al. 1999), rivalry (Blake and Logothetis 2002; Tong 2003), and V1 activity has been shown to correlate with perception (Ress and Heeger 2003). Alternatively, the stimuli of Murray et al. (2002) were broadband, and even though, they

controlled for number of line terminations and parallel lines, it is not clear that these stimuli were equated in terms of low-level statistics known to affect the fMRI response, such as local orientation contrast (Williams et al. 2003; Zenger-Landolt and Heeger 2003), local phase contrast (Williams et al. 2003), powerspectra (Rainer et al. 2002), or sparseness (Olman et al. 2004). Any combination of these attributes may have confounded their interpretation.

On the other hand, Murray et al. (2002) described a bistable rivalrous stimulus whose percept alternates between a complete and disrupted shape. V1 activations correlated with the subject's perception rather than the stimulus statistics. This bistable stimulus provides support for perception-based modulation of V1 activity. Another possible explanation for the perceptual V1 modulation proposes that V1 gates what information reaches higher-level areas (Blake and Logothetis 2002; Tong 2003) and the rivalry percept could primarily result from noise sources in bottom up processes (Ress and Heeger 2003). Nevertheless, we are not arguing against perception-based activity in V1.

In this paper, we contrasted perceptual shape-based hypotheses with a stimulus-based orientation variance prediction. Although, orientation contrast could be computed in V1 using a strictly feed-forward model (Seriès et al. 2003), this may also be implemented by feedback from other areas (Hupé et al. 1998). This feedback differs from the perception-based models because it depends solely on stimulus layout and no internal representation or shape-based hypothesis of the scene is required.

In conclusion, we propose that decreased activity in V1 with increasing image structure does not necessarily reflect high-order perceptions of the scene but may be explained by signal changes in low-level image statistics generally correlated with shape. These modulations of V1, whether they be due to purely feedforward or feedforward/feedback influences, may reflect the first steps in reconstructing shape information from the local V1 neuronal receptive fields.

APPENDIX A

The circular shape detector model schematics are illustrated in Fig. 3A. The model has three stages. These stages are illustrated in Fig. A1,
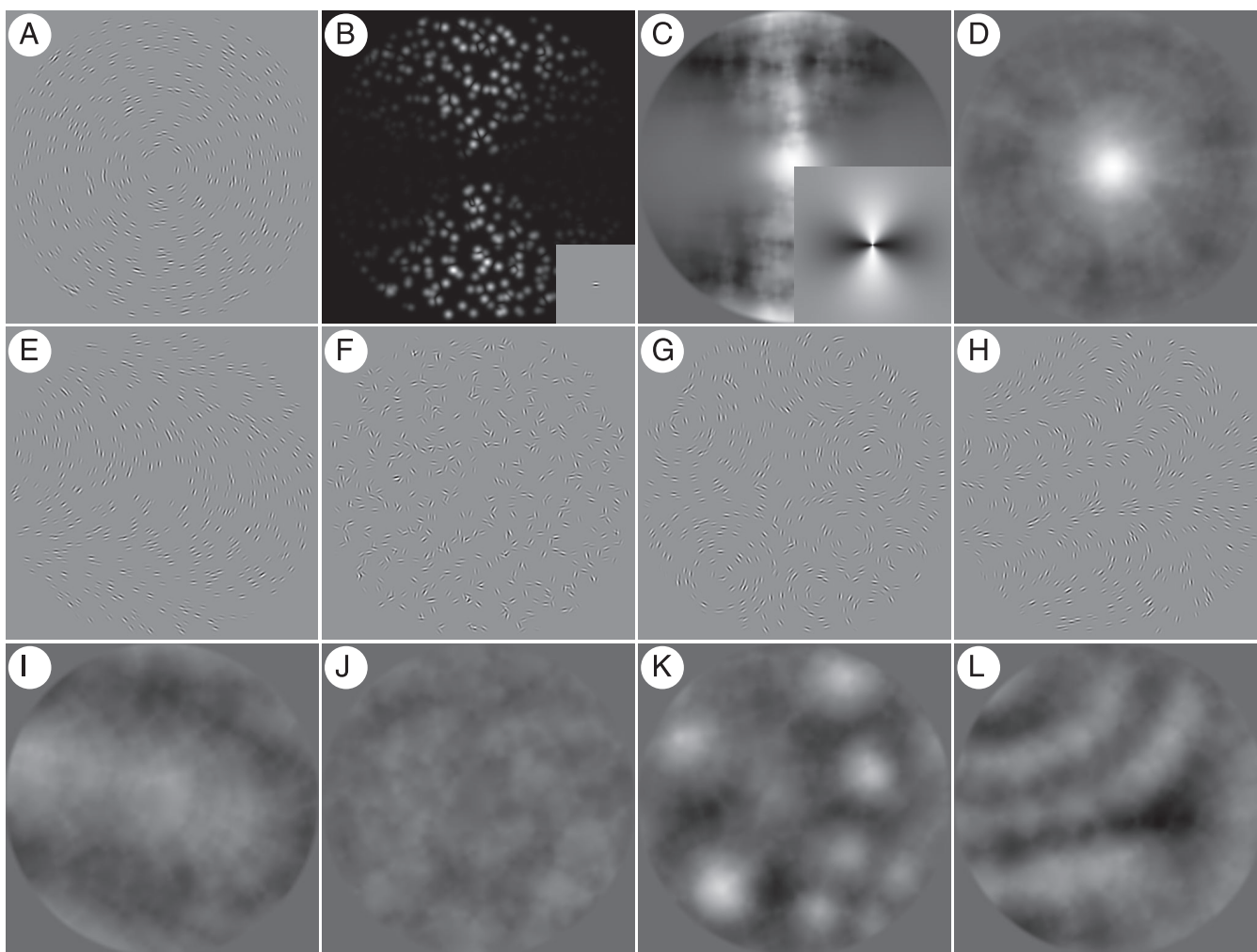

FIG. A1. Illustration of a putative circular shape detector. Different stages are shown in A–D. A: original image. B: local orientation energy computed by convolution of the original image (A) with the filter in the *inset* of B. The 2nd stage shows the resulting image computed by convolution of the local maxima of B with the filter in the *inset* of C. This stage mimics the behavior of a facilitative and suppressive surround. The final image (D) is computed by repeating these 2 steps in B and C for different orientations of the filters (*insets* in B and C). Results of other image categories (E–H) are shown in I–L. In the resultant images, the local maxima correspond to the center locations of circles. Images in D and I–L are scaled identically.

B–D. The original image is shown in *A. B* illustrates the first stage where the energy of an oriented filter is extracted. These filters were Gabors identical to the Gabor elements that make up the pattern. The particular filter used in *B* is shown in the *inset.* The local maxima in this image are convolved with a radial sinusoid enclosed in a Gaussian envelope (sigma = 2°, *C*). This radial sinusoid is the inverse of the "association field" described in contour integration (Field et al. 1993; Hess et al. 2003). The radial sinusoid is shown in scale in the *inset* of *C.* The radial sinusoid was oriented with the maxima orthogonal to the orientation of the initial filter (*inset* in *B*). In essence, the second stage mimics a surround that enhances elements orthogonal to the initial filters' orientation and suppresses in the parallel orientation. These two steps were repeated for different orientations of the two stages (16 orientations in total) and summed. The resultant image is shown in *D.* Examples of the other image categories (*E–H*) are shown in panels *I–L,* respectively.

The pixel intensities of the output images (*D, I–L,* which are identically scaled) correspond to the response of a putative circle detector centered on that pixel or, in other words, the probability of a circle centered on each pixel location. The center of circles can be identified by local maxima in the final images, see for example *D* and *K.* In random patterns, the excitatory and inhibitory influences cancel each other out, see for example *J.* In flowfields local intermediate enhancing effects can be seen (*I* and *L*) that are not as localized as in images containing circles. To summarize the amount of circular shape in each image, all values larger than zero were summed in the resultant images (*D* and *I–L*), giving one value per image. The mean amount of circular shape and SD is shown in Fig. 3*B.* A total of 25 images for each category was used.

REFERENCES

Achtman RL, Dumoulin SO, Ledgeway T, and Hess RF. FMRI activation of striate and extra-striate areas by circumferential gratings and phase-scrambled patterns of varying contrast (Abstract). *Invest Ophthalmol Vis Sci* 42: S407, 2001.

Achtman RL, Hess RF, and Wang Y-Z. Sensitivity for global shape detection. *J Vision* 3: 616–624, 2003.

Albright TD and Stoner, GR. Contextual influences on visual processing. *Annu Rev Neurosci* 25: 339–379, 2002.

Allman J, Miezin F, and McGuinness E. Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Annu Rev Neurosci* 8: 407–430, 1985.

Blake R and Logothetis N. Visual competition. *Nat Rev Neurosci* 3: 1–11, 2002.

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

Brefczynski JA and DeYoe EA. A physiological correlate of the "spotlight" of visual attention. *Nat Neurosci* 2: 370–374, 1999.

Collins DL, Neelin P, Peters TM, and Evans AC. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J Comput Assist Tomogr* 18: 192–205, 1994.

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

De Valois RL and De Valois KK. *Spatial Vision.* Oxford, UK: Oxford Univ. Press, 1988.

Dumoulin SO, Bittar RG, Kabani NJ, Baker CL Jr, Le Goualher G, Pike GB, and Evans AC. A new anatomical landmark for reliable identification of human area V5/MT: a quantitative analysis of sulcal patterning. *Cereb Cortex* 10: 454–463, 2000.

Dumoulin SO, Hoge RD, Baker CL Jr, Hess RF, Achtman RL, and Evans A C. Automatic volumetric segmentation of human visual retinotopic cortex. *Neuroimage* 18: 576–587, 2003.

Field DJ, Hayes A, and Hess RF. Contour integration by the human visual system: evidence for a local "association field. " *Vision Res* 33: 173–193, 1993.

Fitzpatrick D. Seeing beyond the receptive field in primary visual cortex. *Curr Opin Neurobiol* 10: 438–443, 2000.

Gallant JL, Braun J, and Van Essen DC. Selectivity for polar, hyperbolic, and cartesian gratings in macaque visual cortex. *Science* 259: 100–103, 1993.

Gallant JL, Connor CE, Rakshit S, Lewis JW, and Van Essen DC. Neural responses to polar, hyperbolic, and cartesian gratings in area V4 of the macaque monkey. *J Neurophysiol* 76: 2718–2739, 1996.

Gandhi SP, Heeger DJ, and Boynton GM. Spatial attention affects brain activity in human primary visual cortex. *Proc Natl Acad Sci USA* 96: 3314–3319, 1999.

Grill-Spector K. The neural basis of object perception. *Curr Opin Neurobiol* 13: 1–8, 2003.

Grill-Spector K, Henson R, and Martin A. Repetition and the brain: neural models of stimulus-specific effects. *Trends Neurosci* 10: 14–23, 2006.

Grill-Spector K, Kushnir T, Hendler T, Shimon, Itzchak EY, and Malach R. A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Hum Brain Mapp* 6: 316–328, 1998.

Hess RF, Hayes A, and Field DJ. Contour integration and cortical processing. *J Physiol* 97: 105–119, 2003.

Hubel DH and Wiesel TN. Receptive fields of single neurons in the cat's striate cortex. *J Physiol* 148: 574–591, 1959.

Hubel DH and Wiesel TN. Receptive fields, binocular interactions, and functional architecture in the cat's visual cortex. *J Physiol* 160: 106–154, 1962.

Hupé JM, James AC, Payne BR, Lomber SG, Girard P, and Bullier J. Cortical feedback improves discrimination between figure and background by V1, V2, and V3 neurons. *Nature* 394: 784–787, 1998.

Kanwisher N, Chun M, McDermott J, and Ledden P. Functional imaging of human visual recognition. *Cogn Brain Res* 5: 55–67, 1996.

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

Kastner S, De Weerd P, Pinsk MA, Elizondo MI, Desimone R, and Ungerleider LG. Modulation of sensory suppression: Implications for receptive field sizes in the human visual cortex. *J Neurophysiol* 86: 1398–1411, 2001.

Kastner S, De Weerd P, and Ungerleider LG. Texture segregation in the human visual cortex: a functional MRI study. *J Neurophysiol* 83: 2453–2457, 2000.

Kovács I, and Julesz B. A closed curve is much more than an incomplete one: effect of closure in figure-ground segmentation. *Proc Natl Acad Sci USA* 90: 7495–7497, 1993.

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

Levi DM and Klein SA. Seeing circles: what limits shape perception? *Vision Res* 40: 2329–2339, 2000.

MacDonald D, Kabani N, Avis D, and Evans AC. Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *Neuroimage* 12: 340–356, 2000.

Malach R, Levy I, and Hasson U. The topography of high-order human object areas. *Trends Neurosci* 6: 176–184, 2002.

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

Martinez A, Anllo-Vento L, Sereno MI, Frank LR, Buxton RB, Dubowitz DJ, Wong EC, Hinrichs H, Heinze HJ, and Hillyard SA. Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nat Neurosci* 2: 364–369, 1999.

Müller JR, Metha AB, Krauskopf J, and Lennie P. Rapid adaptation in visual cortex to the structure of images. *Science* 285: 1405–1408, 1999.

Murray SO, Kersten D, Olshausen BA, Schrater P, and Woods DL. Shape perception reduces activity in human primary visual cortex. *Proc Natl Acad Sci USA* 99: 15164–15169, 2002.

Murray SO, Schrater P, and Kersten D. Perceptual grouping and the interactions between visual cortical areas. *Neural Networks* 17: 695–705, 2004.

Murray SO and Wojciulik E. Attention increases neural selectivity in the human lateral occipital complex. *Nat Neurosci* 7: 70–74, 2004.

Olman CA, Ugurbil K, and Kersten D. The role of feature density in determining V1 BOLD fMRI sensitivity to spatial phase structure (Abstract). *J Vision* 4: 544a, 2004.

Paradis AL, Cornilleau-Pérès V, Droulez J, Van de Moortele PF, Lobel E, Berthoz A, Le Bihan D, and Poline J. Visual perception of motion and 3-d structure from motion: an fMRI study. *Cereb Cortex* 10: 772–783, 2000.

Pasupathy A and Connor CE. Responses to contour features in macaque area V4. *J Neurophysiol* 82: 2490–2502, 1999.

Pasupathy A and Connor CE. Population coding of shape in are V4. *Nat Neurosci* 5: 1252–1254, 2002.

Pelli DG. The Videotoolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis* 10: 437–442, 1997.

Puce A, Allison T, Asgari M, Gore JC, and McCarthy G. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *J Neurosci* 16: 5205–5215, 1996.

Rainer G, Augath M, Trinath T, and Logothetis NK. Nonmonotonic noise tuning of BOLD fMRI signal to natural images in the visual cortex of the anesthetized monkey. *Curr Biol* 11: 846–854, 2001.

Rainer G, Augath M, Trinath T, and Logothetis NK. The effect of image scrambling on visual cortical BOLD activity in the anesthetized monkey. *Neuroimage* 16: 607–616, 2002.

Rao RPN and Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2: 79–87, 1999.

Ress D and Heeger DJ. Neuronal correlates of perception in early visual cortex. *Nat Neurosci* 6: 414–420, 2003.

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, and Tootell RBH. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268: 889–893, 1995.

Seriès P, Lorenceau J, and Frégnac Y. The "silent" surround of V1 receptive fields: theory and experiments. *J Physiol* 97: 453–474, 2003.

Simoncelli EP and Olshausen BA. Natural image statistics and neural representation. *Annu Rev Neurosci* 24: 1193–1216, 2001.

Sled JG, Zijdenbos AP, and Evans AC. A non-parametric method for automatic correction of intensity non-uniformity in MRI data. *IEEE Trans Med Imag* 17: 87–97, 1998.

Somers DC, Dale AM, Seiffert AE, and Tootell RBH. Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. *Proc Natl Acad Sci USA* 96: 1663–8, 1999.

Talairach, J. and Tournoux, P. *Co-Planar Stereotaxic Atlas of the Human Brain.* New York: Thieme, 1988.

Tong F. Primary visual cortex and visual awareness. *Nat Rev Neurosci* 4: 219–229, 2003.

Wilkinson F, Wilson TWJHR, Gati JS, Menon RS, and Goodale MA. An fMRI study of the selective activation of human extrastriate form vision areas by radial and concentric gratings. *Curr Biol* 10: 1455–1458, 2000.

Williams AL, Singh KD, and Smith AT. Surround modulation measured with functional MRI in the human visual cortex. *J Neurophysiol* 89: 525–533, 2003.

Wilson HR and Wilkinson F. Detection of global structure in glass patterns: implications for form vision. *Vision Res* 38, 1998.

Wilson HR, Wilkinson F, and Asaad W. Concentric orientation summation in human form vision. *Vision Res* 37: 2325–2330, 1997.

Worsley KJ, Liao C, Aston J, Petre V, Duncan GH, Morales F, and Evans AC. A general statistical analysis for fMRI data. *Neuroimage* 15: 1–15, 2002.

Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, and Evans AC. A unified statistical approach for determining significant signals in images of cerebral activation. *Hum Brain Mapp* 4: 58–73, 1996.

Zenger-Landolt B and Heeger DJ. Response suppression in V1 agrees with psychophysics of surround masking. *J Neurosci* 23: 6884–6893, 2003.

Zijdenbos AP, Forghani R, and Evans AC. Automatic pipeline analysis of 3D MRI data for clinical trials: application to multiple sclerosis. *IEEE Trans Med Imag* 21: 1280–1291, 2002.