Masking exposes multiple global form mechanisms

Ben S. Webb
Visual Neuroscience Group, School of Psychology, University of Nottingham, Nottingham, United Kingdom

Neil W. Roach
Visual Neuroscience Group, School of Psychology, University of Nottingham, Nottingham, United Kingdom

Jon W. Peirce
Visual Neuroscience Group, School of Psychology, University of Nottingham, Nottingham, United Kingdom

Previous work suggests there are mechanisms at intermediate levels of visual processing specialized for the detection of radial and circular form. The evidence in favor of specialized global form mechanisms is derived from structure detection experiments that have told us very little about their bandwidth or number. To address these related questions, we examined the effects of configural backward masking on human observers’ ability to detect global structure in arrays with different spiral forms. Each array consisted of 100 Gabors randomly positioned within a circular annular window. Observers judged which of two sequentially presented Gabor arrays contained global structure. One array contained Gabors with random orientations; the other contained Gabors with a variable proportion of orientations coherent with a randomly chosen spiral pitch. At its offset, each array was immediately followed by a backward masking Gabor array with a fixed spiral pitch angle. When mask and test had the same spiral pitch, we found an approximately three-fold elevation of structure detection thresholds that was not explained by local orientation masking. The magnitude and breadth of tuning around each masking angle was predicted by a simple model consisting of at least eight detectors broadly tuned for different spiral forms.

Keywords: global form, shape, spiral pitch, orientation, backward masking


Introduction

Cortical encoding of visual form operates within a hierarchical network of reciprocally connected brain structures residing in the ventral stream (Felleman & Van Essen, 1991). It begins in striate cortex (V1), where neurons represent the retinal position, orientation and spatial scale of local lines and edges in the visual field (Hubel & Wiesel, 1962, 1968). At its conclusion in inferotemporal cortex, neurons are tuned to complex visual patterns and tolerant to changes in object size and position (Fujita, Tanaka, Ito, & Cheng, 1992; Gross, Rocha-Miranda, & Bender, 1972; Ito, Tamura, Fujita, & Tanaka, 1995; Logothetis, Pauls, & Poggio, 1995; Reddy & Kanwisher, 2006). A fundamental question that remains unanswered is exactly how intermediate levels of cortical processing in the ventral stream transform the local representation of visual form in V1 to a position and scale invariant object-based representation in inferotemporal cortex.

Sitting at an intermediate position in the cortical ventral stream, visual area V4 is a critical neural link between the local analysis of visual form in striate cortex and the complex pattern-based representations found at later stages of cortical processing. Damage to this cortical structure disrupts form and color perception (Gallant, Shoup, & Mazer, 2000; Merigan, 1996; Merigan & Pham, 1998; Schiller, 1995; Schiller & Lee, 1991). V4 neurons themselves respond to many different visual dimensions, including the angularity and curvature of object features (Pasupathy &Connor, 1999, 2002), modulation of gratings in polar and hyperbolic coordinates (Gallant, Braun, & Van Essen, 1993; Gallant, Connor, Rakshit, Lewis, & Van Essen, 1996), and geometric primitives (Kobatake & Tanaka, 1994). To encode these higher order visual dimensions, neurons at this intermediate cortical level must selectively group, or “pool,” orientation across space and spatial scale. From a computational standpoint, it may be beneficial to pool orientation in this way because it reduces redundancy in the neural representation of form. Yet there is little consensus on how such a pooling mechanism might work.

Human perceptual studies have started to make some progress towards this end. Glass patterns (Glass, 1969)—textural fields of geometrically transformed dot pairs—are most frequently used to probe grouping mechanisms. With these patterns, both local (within dipole) and global (between dipoles) grouping mechanisms can be manipulated. This work suggests that a narrow range of local filters, band-pass tuned for spatial frequency and orientation, encode the local structure in Glass patterns (Dakin, 1997a, 1997b; Prazdny, 1986; Zucker, 1985). Global structure, according to some psychophysical (Achtman, Hess, & Wang, 2003; Kelly, Bischof, Wong-Wylie, & Spetch, 2001; Kurki & Saarinen, 2004; Seu & Ferrera, Schiller, 1995; Schiller & Lee, 1991). V4 neurons themselves respond to many different visual dimensions, including the angularity and curvature of object features (Pasupathy & Connor, 1999, 2002), modulation of gratings in polar and hyperbolic coordinates (Gallant, Braun, & Van Essen, 1993; Gallant, Connor, Rakshit, Lewis, & Van Essen, 1996), and geometric primitives (Kobatake & Tanaka, 1994). To encode these higher order visual dimensions, neurons at this intermediate cortical level must selectively group, or “pool,” orientation across space and spatial scale. From a computational standpoint, it may be beneficial to pool orientation in this way because it reduces redundancy in the neural representation of form. Yet there is little consensus on how such a pooling mechanism might work.

Human perceptual studies have started to make some progress towards this end. Glass patterns (Glass, 1969)—textural fields of geometrically transformed dot pairs—are most frequently used to probe grouping mechanisms. With these patterns, both local (within dipole) and global (between dipoles) grouping mechanisms can be manipulated. This work suggests that a narrow range of local filters, band-pass tuned for spatial frequency and orientation, encode the local structure in Glass patterns (Dakin, 1997a, 1997b; Prazdny, 1986; Zucker, 1985). Global structure, according to some psychophysical (Achtman, Hess, & Wang, 2003; Kelly, Bischof, Wong-Wylie, & Spetch, 2001; Kurki & Saarinen, 2004; Seu & Ferrera, Schiller, 1995; Schiller & Lee, 1991). V4 neurons themselves respond to many different visual dimensions, including the angularity and curvature of object features (Pasupathy & Connor, 1999, 2002), modulation of gratings in polar and hyperbolic coordinates (Gallant, Braun, & Van Essen, 1993; Gallant, Connor, Rakshit, Lewis, & Van Essen, 1996), and geometric primitives (Kobatake & Tanaka, 1994). To encode these higher order visual dimensions, neurons at this intermediate cortical level must selectively group, or “pool,” orientation across space and spatial scale. From a computational standpoint, it may be beneficial to pool orientation in this way because it reduces redundancy in the neural representation of form. Yet there is little consensus on how such a pooling mechanism might work.

Human perceptual studies have started to make some progress towards this end. Glass patterns (Glass, 1969)—textural fields of geometrically transformed dot pairs—are most frequently used to probe grouping mechanisms. With these patterns, both local (within dipole) and global (between dipoles) grouping mechanisms can be manipulated. This work suggests that a narrow range of local filters, band-pass tuned for spatial frequency and orientation, encode the local structure in Glass patterns (Dakin, 1997a, 1997b; Prazdny, 1986; Zucker, 1985). Global structure, according to some psychophysical (Achtman, Hess, & Wang, 2003; Kelly, Bischof, Wong-Wylie, & Spetch, 2001; Kurki & Saarinen, 2004; Seu & Ferrera, Schiller, 1995; Schiller & Lee, 1991). V4 neurons themselves respond to many different visual dimensions, including the angularity and curvature of object features (Pasupathy & Connor, 1999, 2002), modulation of gratings in polar and hyperbolic coordinates (Gallant, Braun, & Van Essen, 1993; Gallant, Connor, Rakshit, Lewis, & Van Essen, 1996), and geometric primitives (Kobatake & Tanaka, 1994). To encode these higher order visual dimensions, neurons at this intermediate cortical level must selectively group, or “pool,” orientation across space and spatial scale. From a computational standpoint, it may be beneficial to pool orientation in this way because it reduces redundancy in the neural representation of form. Yet there is little consensus on how such a pooling mechanism might work.

Human perceptual studies have started to make some progress towards this end. Glass patterns (Glass, 1969)—textural fields of geometrically transformed dot pairs—are most frequently used to probe grouping mechanisms. With these patterns, both local (within dipole) and global (between dipoles) grouping mechanisms can be manipulated. This work suggests that a narrow range of local filters, band-pass tuned for spatial frequency and orientation, encode the local structure in Glass patterns (Dakin, 1997a, 1997b; Prazdny, 1986; Zucker, 1985). Global structure, according to some psychophysical (Achtman, Hess, & Wang, 2003; Kelly, Bischof, Wong-Wylie, & Spetch, 2001; Kurki & Saarinen, 2004; Seu & Ferrera,
and target. Recent work suggests that there may be some surprisingly little work has utilized backward masking to specialized pooling mechanisms but the latter one does not colleagues, this suggests that the former two structures have mechanisms from structure detection thresholds alone. The reason being it is difficult to obtain unambiguous are necessary to encode myriad forms present in the visual world. The reason being it is difficult to obtain unambiguous estimates of the number and bandwidth of global form detectors. Nor have they told us much about the related question of how many detectors encoded early on in visual cortex. This supposition is supported by physiological work demonstrating that classical receptive field and surround mechanisms in V1 or V2 are incapable of encoding global structure in these complex patterns (Smith, Bair, & Movshon, 2002; Smith, Kohn, & Movshon, 2007). Moreover, there is circumstantial physiological and functional imaging evidence that further downstream in V4, neurons respond to radial and circular form. First, a subpopulation of macaque V4 neurons responds to families of gratings defined by their spiral pitch (Gallant et al., 1993, 1996). Second, many V4 neurons have radial shaped receptive fields (Pollen, Przybyszewski, Rubin, & Foote, 2002). Third, there are greater changes in the blood oxygenated level dependent signal in human V4 to circular and radial gratings than to Cartesian gratings (Wilkinson et al., 2000).

The evidence most frequently cited in favor of specialized circular and radial detectors is based on a finding, with one notable exception (Dakin & Bex, 2002), replicated many times: sensitivity to circular and radial structure is greater than to translational structure (Dakin & Bex, 2001; Kelly et al., 2001; Kurki & Saarinen, 2004; Pei et al., 2005; Seu & Ferrera, 2001; Wilson & Wilkinson, 1998; Wilson et al., 1997). According to Wilson and colleagues, this suggests that the former two structures have specialized pooling mechanisms but the latter one does not (Wilson & Wilkinson, 1998; Wilson et al., 1997). Yet structure detection experiments have told us very little about the bandwidth of global form detectors. Nor have they told us much about the related question of how many detectors are necessary to encode myriad forms present in the visual world. The reason being it is difficult to obtain unambiguous estimates of the number and bandwidth of global form mechanisms from structure detection thresholds alone.

To circumvent this potential ambiguity, we take a different approach and examine the effects of configural backward masking on human observers’ ability to detect global structure in stimulus arrays with different spiral forms. In backward masking, performance on a briefly presented target stimulus is impaired by the subsequent presentation of a visual mask. This technique has a long and successful history in uncovering the temporal dynamics of visual information processing (for reviews, see, Breitmeyer & Ögmen, 2006; Breitmeyer, 2007). Yet, surprisingly little work has utilized backward masking to probe complex spatial (shape) relationships between mask and target. Recent work suggests that there may be some mileage with this approach since backward complex pattern masking can have a profound and selective effect on shape perception (Habak, Wilkinson, & Wilson, 2006).

We exploit the shape selectivity of configural backward masking to probe the number and bandwidth of mechanisms tuned to different spiral forms. We find a selective elevation in structure detection thresholds when mask and test have the same spiral pitch angle. The magnitude and breadth of tuning around each masking angle are most easily explained by a simple model consisting of at least eight detectors each broadly tuned for spiral form.

### Methods

#### Subjects

Four observers with normal or corrected-to-normal visual acuity participated. Three were authors and one (C.V.H) was naive to the specific purposes of the experiment.

#### Stimuli

Gabor arrays (see Figure 1) were generated on an Apple Macintosh G5 using custom software written in Python (Peirce, 2007). We presented the stimulus arrays on a Vision Master Pro 454 monitor at a resolution of 1,024 × 768 pixels, refresh rate of 100 Hz, and viewing distance of 70 cm. Each array (generated anew prior to each trial) consisted of 100 Gabors randomly positioned inside an annular window (outer diameter 10°, inner diameter 1°) at non-overlapping positions on a uniform background (luminance 95 cd/m²). Each Gabor (Michelson contrast 0.9, sinusoidal carrier frequency 6 cycles/°, circular Gaussian envelope SD 0.166°) was in sine phase and assigned an orientation consistent with a spiral pitch angle (Figure 1a). Gabors that fell outside the window were redrawn inside at a random location.

#### Procedure

In a temporal two-alternative forced-choice task, observers judged which of two sequentially presented Gabor arrays contained global structure. On each trial, one array contained Gabors with random orientations; the other contained Gabors with a variable proportion of orientations coherent with a spiral pitch. Each test array was presented for 20 ms and immediately followed at its offset by a 100% coherent backward masking array of Gabors for 500 ms (ISI = 0). On each run, the spiral pitch of a test pattern was randomly chosen from 9 angles in the range 0–90 degrees and the backward masking pattern was presented with a fixed pitch angle. To control for
local orientation masking effects, we ran two experiments with (1) spatially non-overlapping mask and test elements and (2) randomly oriented masks elements, drawn from a uniform distribution spanning 180 degrees. The timing of both experiments was the same as above. For the non-overlapping experiment, mask and tests both had pitch angles that elicited the peak masking effect.

**Data analysis**

For the unmasked and each masking condition, observers completed 6 runs of 315 trials. Data were expressed as the proportion of trials on which subject correctly identified the Gabor array containing global structure as a function of the proportion of orientations coherent with a given pitch angle. To estimate structure detection thresholds, we fitted these data with a function (Weibull, 1951) of the form:

\[
y = 1 - 0.5 \left(2^{-1(s/a)^\beta}}\right),
\]

where \(y\) is the proportion of correct trials, \(a\) is an estimate of threshold at 75% correct, and \(\beta\) is related to the slope of the function.

To estimate the bandwidth of each spiral pitch mechanism, we fitted a Gaussian to normalized orientation...
coherence thresholds (Th) for each masking condition. This function has the form:

$$\text{th} = M + \text{Amp} \exp(-0.5((x - \mu)/\sigma)^2),$$

where M is the baseline, Amp is the amplitude, and $\sigma$ is the bandwidth of each masking function.

We modeled observers’ performance on the structure detection task on a trial-by-trial basis using different numbers of mechanisms tuned along the spiral pitch dimension ($-90$ to $+90^\circ$). In each case, the normalized response of the $i$th mechanism to a Gabor array with pitch angle $\theta$ and coherence $c$ was defined by a Gaussian function of the form:

$$R_i(\theta, c) = c e^{-\frac{(\theta-\theta_i)^2}{2\sigma^2}},$$

where $\theta_i$ is the mechanism’s preferred pitch angle and $\sigma$ is the standard deviation of the tuning curve. Separate Monte Carlo simulations were run with (i) two mechanisms with preferences for radial (0 deg) and circular (90 deg) structure; (ii) four mechanisms (0, 45, 90, and 135 deg); and (iii) eight mechanisms (0, 22.5, 45, 67.5, 90, 112.5, 135, and 147.5 deg).

Physiological studies in monkeys have demonstrated that impairments of target discrimination induced by backwards masking coincide with changes in neuronal responses to the target stimulus (Kovács, Vogels, & Orban, 1995; Rolls, Toveé, & Panzeri, 1999). Consistent with popular “interruption” theories (Kahneman, 1968), backwards masking both attenuates firing rate and reduces the temporal interval over which neurons respond to a previously presented stimulus. The classic theoretical alternative to “interruption” is “integration,” whereby masking is thought to impair performance due to temporal integration of responses to stimulus and mask (Felsten & Wasserman, 1980; Kahneman, 1968). In separate simulations, we modeled the effect of masking as either a selective reduction of mechanism responses to each stimulus array (interruption) or as weighted summation of stimulus and mask responses (integration). As both approaches produced qualitatively similar results, only the results for the interruption method are reported here.

The influence of masking was proportional to the detector’s response to the masking array:

$$R_{\text{Masked}} = R(\theta_{\text{Stim}}, c_{\text{Stim}}) - w R(\theta_{\text{Mask}}, c_{\text{Mask}}),$$

where $w$ is a weighting constant that sets the overall magnitude of the masking effect (fixed at 0.35). In rare instances where the weighted response to the mask exceeded that to the stimulus, a normalized response of zero was assigned.

The mean number of spikes ($N$) elicited in response to a stimulus was given by:

$$N_i = N_{\text{rest}} + (N_{\text{peak}} - N_{\text{rest}}) R_i,$$

where $N_{\text{rest}}$ and $N_{\text{peak}}$ are resting (fixed at 5) and peak (fixed at 20) responses, respectively. The probability of $n$ spikes from a mechanism on a given stimulus presentation was defined by a Poisson probability density function with mean value $N$, such that:

$$p(n|N) = \frac{e^{-N}N^n}{n!}.$$

To decode the response of a set of $I$ detectors, the log likelihood of each potential stimulus was calculated (see Jazayeri & Movshon, 2006):

$$\log L(\theta) = \sum_{i=1}^{I} n_i \log R_i(\theta) - \sum_{i=1}^{I} R_i(\theta) - \sum_{i=1}^{I} \log(n_i!).$$

Model responses on each trial were driven by the maximum stimulus likelihood across two intervals of the 2IFC.

**Results**

We began by measuring detection thresholds for different spiral forms. From Weibull fits to these data, we estimated detection thresholds at 75% correct for each spiral pitch angle. Figure 2 (white circles) shows the results of this analysis for four observers as a function of spiral pitch angle. For two observers (CVH and JWP), thresholds are lowest for circular patterns and roughly equivalent for radial and intermediate spiral angles. This sensitivity profile is consistent with previous work using Glass patterns (Kelly et al., 2001; Kurki & Saarinen, 2004; Seu & Ferrera, 2001; Wilson & Wilkinson, 1998; Wilson et al., 1997) and Gabor arrays (Achtman et al., 2003), showing that thresholds for circular patterns are moderately lower than they are for radial patterns. For the other two observers (BSW and NWR), except for a slight elevation at intermediate spiral angles, thresholds are roughly equivalent across the continuum. This profile is different to that seen with Glass patterns, which typically shows a threshold advantage for circular over radial patterns and substantial elevation of thresholds at intermediate spiral angles (Seu & Ferrera, 2001). The extensive practice that BSW and NWR had with the stimulus set may have decreased the threshold elevation at
intermediate spiral angles. Consistent with this view, previous Glass pattern studies have shown that practice reduces structure detection thresholds (Dakin & Bex, 2002) and more experienced observers have a flatter sensitivity profile across the global form continuum (Seu & Ferrera, 2001).

As we go on to show in our modeling, it is difficult to unambiguously estimate the number and bandwidth of the underlying form mechanisms from structure detection thresholds alone. Therefore, to probe the underlying mechanisms, we measured the effects on structure detection thresholds of backward masking with Gabor arrays at different spiral pitch angles. Figure 2 (colored circles) shows the masking functions for individual observers. Colored arrows indicate the spiral pitch angle of each backward mask. For all observers, it is clear that when mask and test arrays had the same spiral pitch angle there was an approximately two to three-fold elevation in structure detection thresholds. The magnitude and tuning width of the effect was similar for all masking angles, suggesting that detectors with similar tuning properties underlie the encoding of spiral form. This point is reinforced in Figure 3a, which shows the masking functions normalized to individual structure detection thresholds and averaged across observers. Each masking function is well characterized by Gaussian functions (Equation 1) of similar amplitude (Amp range: 1.54–2.88) and bandwidth (1 SD, range: 14–20.71°).

Because we have used a fixed mask angle in each experimental run, this may have caused some cumulative adaptation of the mechanism tuned to the masking stimulus. To ensure that any adaptation effects were balanced across conditions, we re-ran the main experiment with three masking angles (0°, 45°, 90°) in a fully randomized structure. The magnitude and tuning of the masking functions were qualitatively the same as the original experiment (data not shown), ruling out an explanation of our results based on adaptation.

Because there was some local spatial overlap between mask and test Gabors, it is possible that local orientation masking (e.g. Phillips & Wilson, 1984) at early levels of cortical processing contributes to the magnitude of the
masking effect. To examine this possibility, we tested the same subjects on two additional control experiments. In both controls, the design was the same as the original masking experiment except that we used test patterns with spiral pitch angles that elicited the peak masking effect. To control for local orientation masking, we used (1) spatially non-overlapping mask and test Gabor elements and (2) masks with randomly oriented Gabors.

The results of these experiments are shown in Figure 3b. Colored circles show the peak effect with non-overlapping mask and test elements. The threshold elevation is similar for all masking angles, yet, to our surprise, was larger than the peak effects we found when mask and test elements were partially overlapping. Nonetheless, the magnitudes of these masking effects are of the same order as the first experiment. Coupled with the result that randomly oriented masking had negligible effects on structure detection thresholds (black circles in Figure 3b), it is difficult to reconcile our results with local orientation masking.

The simplest account of these data is that the pattern of tuning is mediated by detectors at intermediate levels of visual processing that pool local orientation information across space. To determine how many detectors are required to encode multiple spiral forms, we modeled observers’ performance on the structure detection task (for details, see Methods). Each model consisted of a different number of detectors spaced evenly along the spiral form dimension (Figure 4a). We first simulated unmasked performance for each model using a range of different tuning bandwidths. As illustrated in Figure 4b, different combinations of bandwidth and number of detectors produced a variety of performance profiles. However, by assigning an appropriate bandwidth, predictions approximating observers’ mean unmasked performance could be obtained for each model. Critically, the fact that these different models can produce near metameric predictions illustrates that it is not possible to unambiguously estimate the number and bandwidth of spiral form mechanisms from structure detection thresholds alone.

To overcome this problem, we next simulated masked performance using tuning bandwidths that provided the best prediction of the unmasked threshold data (σ = 29, 27, and 12 deg for 2, 4, and 8 detectors, respectively). The results of these simulations are shown in Figure 4c. Clearly, predictions of the two cardinal detectors model do not provide an accurate description of the masking functions. In particular, the model erroneously predicts a broadening of masking functions at intermediate mask pitch angles (e.g., 45 deg). Overall, the cardinal model predictions provide a poorer fit of the empirical data set than a simple straight line through the mean ($R^2 = 0.39$). The model consisting of four detectors provided a closer approximation of these data ($R^2 = 0.42$), although some broadening of masking functions remains at angles falling between detector preferences (22.5 and 67.5 deg). The model consisting of eight detectors accounted for the largest proportion of variance in the pattern of tuning we observed across the spiral form dimension ($R^2 = 0.78$).

We also considered alternative models consisting of multiple circular and radial detectors with different bandwidths. However, we unable to find a combination of bandwidths that accurately predicted the pattern of tuning observed at multiple spiral pitch angles.

Discussion

We have shown that backward masking selectively elevates thresholds for detecting visual patterns with different spiral pitches. Spatially segregating the local elements in the mask and test patterns did not reduce the magnitude of the masking effect, ruling out a contribution from local orientation masking. Simulations with a model consisting of eight broadly tuned detectors accurately predicted the magnitude and breadth of tuning around each masking angle.

We chose to probe the neural machinery underlying global form perception with backward masking rather
than adaptation because, with structured patterns, the latter has perceptual consequences undesirable for our current purpose. Clifford and Weston (2005) demonstrated that adapting to a globally structured Glass pattern causes an unstructured pattern to acquire illusory structure that is orthogonal to the adaptor (Clifford & Weston, 2005). This illusory aftereffect, which we have reproduced with the Gabor arrays we used here (data not shown), would have contaminated observers’ structure detection judgments.

Our results suggest that backward masking may actually be a very useful tool for uncovering the nature of intermediate form mechanisms. In a typical masking experiment, if the mask resembles the test and is presented soon after it, the visibility of the test is severely reduced. The subjective effects of backward masking we report here, however, were not a simple reduction in the visibility of spiral patterns, but rather a change in their perceived global structure. This subjective impression reported by all observers suggests that we were masking at an intermediate level of form processing encoding the global structure in spiral patterns.

It has proven very difficult to determine the characteristics of the detectors and grouping algorithms used at intermediate levels of form processing. Previous psychophysical work suggests that there are mechanisms at this level, specialized for the detection of radial and circular form (Achtman et al., 2003; Kelly et al., 2001; Kurki & Saarinen, 2004; Seu & Ferrera, 2001; Wilson & Wilkinson, 1998; Wilson et al., 1997). These detectors are constructed from three stages of processing: early linear, oriented-filtering, followed by rectification and then further filtering by detectors that respond selectively to radial and circular contours (Wilson & Wilkinson, 1998; Wilson et al., 1997). Our results confirm that human observers are adept at detecting structure in patterns drawn from the spiral form dimension. Moreover, we demonstrate that backward

Figure 4. Simulated unmasked and masked performance on the structure detection task. (a) Schematic representation of the spiral form detectors used in the simulations. Separate polar plots display example tuning functions with 2, 4, and 8 detectors. (b) Simulated unmasked performance for each model using a range of different tuning bandwidths. For comparison, gray circles show experimental data averaged across observers. (c) Simulated masked performance with tuning bandwidths that provided the best prediction of the unmasked threshold data. Notation is the same as Figure 2.
masking selectively impairs this ability. However, our results do not conform to the notion that local orientation is grouped solely by cardinal mechanisms tuned to detect circular and radial form. We needed to draw on responses from at least eight detectors in our modeling to explain accurately the magnitude and breadth of the masking functions.

An alternative hypothesis in the psychophysical literature is that image-based statistics rather than specialized detectors are used to group local orientation into textural surfaces (Dakin, 1999; Dakin & Watt, 1997). The evidence indicates that first moment (mean orientation) and second moment (orientation variance) statistics can be exploited in global orientation judgments. At maximum coherence, the distributions of orientations in the spiral pitch patterns we have used here have exactly the same moment statistics. Therefore, it is possible in principle that observers’ could have exploited these image-based cues to discriminate structured test patterns from randomly oriented stimuli. However, it is unlikely that our tuned masking functions could be explained within this framework, since disruption of this process ought to impair structure detection judgments irrespective of spiral pitch.

That spatially non-overlapping mask and test elements did not reduce the magnitude of the masking effect and randomly oriented mask elements had no effect on structure detection encourages the view that whatever mechanism is mediating our results resides at an intermediate level of visual processing, like V4. Receptive fields in this region are well suited to the encoding of spiral form because they integrate orientation over spatial extensive areas of the visual field (Desimone & Schein, 1987) and respond selectively to non-Cartesian gratings (Gallant et al., 1993, 1996). The spiral pitch dimension is one of many intermediate form dimensions encoded by neurons at intermediate levels of cortical visual processing (David, Hayden, & Gallant, 2006; Desimone & Schein, 1987; Freiwald, Tsoa, Tootell, & Livingstone, 2004; Gallant et al., 1993, 1996; Kobatake & Tanaka, 1994; Pasupathy & Connor, 1999, 2002; Pollen et al., 2002). Because of the practical difficulties of exploring a high dimensional space, it is unclear if neurons at intermediate levels simultaneously represent multiple form dimensions or actively reduce the stimulus space down into a more compact representation. Based on the belief that form representations are feature based (Biederman, 1987; Marr & Nishihara, 1978), one widely explored possibility is that neurons in the ventral pathway encode the orientation changes (curves, corners, angles, etc.) in visual patterns in terms of higher order contour derivatives (Pasupathy & Connor, 1999, 2002). According to this model, curvature and spiral pitch would be second- and third-order derivatives, respectively (Connor, Brincat, & Pasupathy, 2007). Framed within the terms of this scheme, our results suggest that third-order derivatives like spiral form are encoded by the distributed activity of multiple broadly tuned detectors.

An intriguing alternative hypothesis describes intermediate neural representations of form in terms of the joint orientation and spatial frequency power spectrum of patterns, independently of spatial phase (David et al., 2006). The spectral receptive field model is a direct extension of the notion that V1 is performing a local patch wise Fourier analysis of the visual image (De Valois & De Valois, 1990). It explains many aspects of form selectivity in V4, including tuning for the angularity and curvature of object features (Pasupathy & Connor, 1999, 2002) and modulation of gratings in polar and hyperbolic coordinates (Gallant et al., 1993, 1996). Moreover, it accounts for V4 neurons’ broad orientation and spatial frequency tuning and bi-modal orientation tuning (David et al., 2006), characteristics that are not present in earlier visual cortical areas. Despite the comprehensive nature of this model, our results are not well described by a mechanism tuned to the orientation and spatial frequency power spectrum. A spectral receptive field would not be able to distinguish between the narrowband Gabor arrays with different spiral pitches used here because the power spectrum of each pattern is very similar. We have confirmed this in some further analysis not shown here. This may illustrate a general limitation of the notion that intermediate levels of visual processing are performing a frequency-based analysis of the visual image in a fashion analogous to that performed in early visual cortical areas.

By systematically manipulating the local and global spatial relationships between mask and test elements, we were able to probe the underlying mechanisms at intermediate levels of form processing. These mechanisms are not specialized for the detection of particular patterns but rather broadly tuned to the global configuration of a range of patterns. The approach we have taken here illustrates the utility of backward masking as a tool for uncovering the properties of detectors and grouping algorithms underlying form processing.

Acknowledgments

B.S.W, N.W.R, and J.W.P are funded by the Leverhulme Trust, Wellcome Trust, and BBSRC, respectively.

Commercial relationships: none.
Corresponding author: Ben S. Webb.
Email: bsw@psychology.nottingham.ac.uk.
Address: Visual Neuroscience Group, School of Psychology, University of Nottingham, Nottingham, UK, NG7 2RD.

References


Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cerebral Cortex, 1, 1–47. [PubMed] [Article]


