Contour integration and scale combination processes in visual edge detection

STEVEN C. DAKIN[†], ROBERT F. HESS
McGill Vision Research, Department of Ophthalmology, 687 Pine Avenue West, Montreal, Quebec, Canada H3A 1A1.

Contours in the natural visual environment consist mainly of *edges* which are spatially broadband and whose (cosinusoidal) components have arrival phases close to ±90°. Because early visual processing is thought to be based on a local Fourier description, the representation of edges requires two forms of filter combination: *scale integration* (filter combination across spatial frequency) and *contour integration* (filter combination across space). In order to determine how these two types of combination fit together, we determined spatial-frequency tuning for the detection of contours composed of broadband edge elements, alternating with narrow-band Gabor elements. A contour integration system operating independently at a number of spatial scales should be able to ignore the distracting influence of edge structure in such patterns. However subjects cannot ignore edge structure indicating that local phase-alignment across spatial scale is coded prior to, or concurrent with, contour integration. Moreover, unlike contours composed of Gabors, the bandwidth of local elements is important for edge integration; the coding of element bandwidth seems to be dependent on the phase alignment of features across spatial frequency.

Key words: Contour Grouping Edges Scale combination

INTRODUCTION

Spatial filtering has assumed a central role in our understanding of human visual processing. This approach supposes that early visual processing performs a local Fourier analysis of image structure, rather than constructing an image description based on local features such as edges. This view has received considerable support from neurophysiological study of cells in primary visual cortex which respond to the Fourier components of stimuli rather than to the presence of (subjective) edges (e.g. De Valois, De Valois & Yund, 1979). It is a consequence of this view that, having deconstructed the image, the visual system must combine local filter outputs in order to infer the presence of complex (and presumably interesting) image structure, such as extended contours. Although this reconstruction process is critical to our understanding of visual processing beyond V1 it remains poorly understood.

Given our knowledge of cortical physiology it is reasonable to characterize a cortical filter as being primarily defined by its sensitivity to the position, orientation, spatial frequency, and phase of a given stimulus. The visual system could potentially infer structure from combining across any or all of these attributes, but certain combinations are more sensible than others. We consider two broad classes of combination rule: *local* (operating at a single visual location) and *multi-local* (operating over multiple visual locations). The first class can be further subdivided into *phase-sensitive* and *phase-insensitive* combination. Several psychophysical studies have revealed rigid phase insensitive filter combination rules. Judgements based on a number of stimulus attributes (orientation, spatial frequency, contrast) do not operate independently from information at quite different orientations or

[†] To whom all correspondence should be addressed. Present address: Institute of Ophthalmology, 11-43 Bath Street, London EC1V 9EL (*Email:* s.c.dakin@ucl.ac.uk).

spatial frequencies (Olzak & Thomas, 1991; 1992). These results have led to the proposal of two combination rules, one across frequency and within orientation (mediating orientation judgements), the other across orientation and within spatial frequency (mediating spatial frequency judgements; Olzak & Thomas, 1992; Thomas & Olzak, 1996). Georgeson, Meese and co-workers have determined that summation of oriented filter outputs across orientation, prior to edge extraction, determines the appearance of static plaids (for review see Georgeson & Meese, 1997). For example, the stimulus depicted in Figure 1a appears as a blurred checkerboard, which is dominated by horizontal and vertical structure, even though its components are two low-frequency gratings at 45° and 135°.

Phase-sensitive local combination across scale is desirable in order to make explicit the presence of local phase alignments, such as edges. Natural images contain a disproportionate number of features with arrival phases around 90° (i.e. edges; Burr & Morrone, 1994). A number of computational schemes for edge detection exploit this by assessing the degree of local correlation between filter outputs at multiple scales (e.g. Marr & Hildreth, 1980; Marr, 1982; Canny, 1983; Torre & Poggio, 1986; Lowe, 1988; Morrone & Burr, 1988; Georgeson, 1992, 1994). This type of inter-scale combination appears to take precedence over other forms of combination, such as the inter-orientation linking operations associated with the perception of static plaids described above. Georgeson & Meese (1997) report that the addition of a 3f grating to one of the components of their plaid stimuli (e.g. Figure 1a) breaks the percept of the stimulus up into the components, if the f and $3f$ components are in square-wave phase.

Multi-local grouping principles predate those described, and are concerned with grouping of image structure *across space*. This idea originates with the Gestaltists (e.g. Wertheimer, 1938), who recognised a number of basic principles that the visual system adheres to in deriving image structure from spatially distributed features. However, certain fundamental principles of spatial grouping, such as proximity and contrast polarity, have been successfully recast in the terms of spatial filtering (e.g. Watt, 1988). This has prompted development of a paradigm for investigating contour integration which sets out to prevent observers from relying on grouping operations based on the output of single filters (Field, Hayes & Hess, 1993). Recently we have confirmed that (for foveal presentation at least) this paradigm does indeed probe grouping operations combining multiple filter outputs (Hess & Dakin, 1997). Studies using this “contour integration” paradigm have begun to tackle the grouping rules that determine association across space. Contour integration is largely phase-insensitive (Field, Hayes & Hess, 1997; Hess & Dakin, 1997) but is sensitive to local orientation (Field *et al*., 1993), with both contour smoothness and closure enhancing salience (Kovács & Julesz, 1993; Pettet, McKee & Grzywacz 1996). Recently, we have shown that contour integration is also tuned for the spatial frequency of components. Detection of contours composed of elements alternating between two spatial frequencies is optimal when the two populations are similar in spatial frequency and deteriorates with increasing difference between them (Dakin & Hess, 1998). We also reported that the maximum tolerable difference in spatial frequency between successive elements was inversely proportional to the curvature of the contour; i.e. subjects could tolerate more variability in spatial frequency within straight contours than within curved contours. We suggested that this could reflect a general bias towards structure of the sort that arises in natural scenes. A straight edge will produce a response from a bank of oriented filters that is more widely dispersed across spatial frequency than will a curved edge, and it is possible that the contour integration system is actively exploiting this via broader spatial tuning to straight contours. To put it another way, these data suggest that local scale combination rules (of the kind described in the preceding paragraph) could precede multi-local grouping processes such as contour integration. The purpose of this paper is to investigate if these local and multi-local combination processes do indeed fall into such a hierarchy.

GENERAL METHODS

Apparatus

A Macintosh 7500 microcomputer controlled stimulus presentation and the recording of subjects' responses. Stimuli were displayed on a Nanao Flexscan 6600 monochrome monitor, with a frame refresh rate of 75 Hz. 12-bit contrast accuracy was achieved by electronically combining the RGB outputs from the computer using a video attenuator (Pelli & Zhang, 1991). Luminance levels were linearised by combining using a look-up table derived using programs from the VideoToolbox package (Pelli, 1997), from which display routines were also derived. The screen was viewed binocularly at a distance of 98cm and had a mean background luminance of 48 cd / m^2.

Stimuli: Micropatterns

All stimuli were composed of micropatterns with luminance modulated by an isotropic Gaussian envelope multiplied by an oriented carrier function. For all micro-patterns the standard deviation of the Gaussian envelope was fixed at 9.4 arc min. The first class of micro-pattern, referred to as "Gabors", had a simple sinusoidal carrier and has the form:

$$G(x) = L_0 + L_0 C \sin\left(\frac{2\pi x_\theta}{\lambda} + \phi\right) \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad \textbf{(1)}$$

where L_0 is the mean luminance, C is the contrast, σ is the standard deviation of the Gaussian envelope, and λ and ϕ are the wavelength and phase, respectively, of the sinusoidal carrier. All stimuli were luminance balanced (i.e. ϕ=0°). x_θ refers to the x-coordinate rotated by the carrier angle θ

$$x_\theta = x\cos\theta + y\sin\theta \quad \textbf{(2)}$$

A second type of patch used a broadband carrier consisting of a "hard" (i.e. not anti-aliased), step-edge. These "edge" micro-patterns are defined by the function:

$$E(x,y,\theta) = L_0 + CL_0 \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \begin{cases} +1 & \text{if } x_\theta > 0 \\ -1 & \text{otherwise} \end{cases} \quad \textbf{(3)}$$

We also employed compound micropatterns composed of the sum of various combinations of odd-frequency Gabors (denoted, "f", "f+3f", "f+3f+5f", etc.) A square-wave is composed of:

$$\sin(f + \phi_1) + \frac{1}{3}\sin(3f + \phi_2) + \frac{1}{5}\sin(5f + \phi_3) \ldots \quad \textbf{(4)}$$

where the phase values ϕ are uniformly set to 0°. Similarly, micropatterns were composed of the Gaussian windowed sum of various numbers of 1 / f-amplitude modulated, phase-aligned sine-wave components. We denote these patterns "f+3f", "f+3f+5f" etc. In later experiments we also employed patches with phase randomised components (i.e. ϕ_1, ϕ_2, ϕ_3 etc. were independently randomised).

A final class of micro-pattern, used in Experiment 1, had a carrier composed of the sum of two perpendicular sinusoidal carriers:

$$\sin\left(\frac{2\pi x_\theta}{\lambda} + \phi\right) + \sin\left(\frac{2\pi x_{\theta+\pi/2}}{\lambda} + \phi\right) \quad \textbf{(5)}$$

These micro-patterns were used only in Experiment 1 and had components fixed at 2.69 cycles per degree.

In all of the experiments reported all classes of micropattern were scaled to maximise Michelson contrast, while retaining a mean luminance equal to the background (48 cd / m^2).

Stimuli: Paths

Details of the contour generation procedure are described in Field *et al.* (1993) and, unless stated otherwise, parameters were similar to those used in that study. Stimuli were composed of a series of micropatterns (the *path*) embedded in a background of randomly oriented distractors. Paths had a backbone of 8 invisible line segments joined at an angle uniformly distributed from $\alpha - 4^0$ to $\alpha + 4^0$. α, the change in orientation between subsequent micropatterns, is called the *path angle.* Micropatterns were located and oriented in accord with a line segment. The orientation of each line segment was ambiguous (within the range 0 - 360°), but traversing the path from one end to the other imposes a direction (and hence an unambiguous orientation) on each of the component line segments. Finally the path was checked to ensure that it neither intersected itself, nor looped back on itself and, if it did, it was regenerated.

The path was inserted into the display at a random location and the background was filled with randomly oriented Gabor elements. These distractor elements were positioned on a perturbed grid so that their average separation matched that of path elements. Thus, the path could not be located using a simple density cue but as an extra precaution against such a cue, uncued stimuli were generated in the same way except that a path composed of randomly oriented elements was inserted into the background. These paths are known to be undetectable by subjects (McIlhagga & Mullen, 1996; Hess & Field, 1995).

Paths were 8 elements long with a 20° path angle. These contours require linking between filters (Hess & Dakin, 1997). Their detection is also relatively narrowly tuned to the consistency of the spatial frequency of contour components (Dakin & Hess, 1998). In the tuning experiments reported here, the spatial frequency of simple Gabors was varied from 1.6 to 6.4 c.p.d. in quarter octave steps. This range was selected because we have previously shown, for the stimulus parameters used, that paths composed exclusively of Gabors at one of these carrier frequencies are approximately equally detectable (Dakin & Hess, 1998).

Procedure

The authors, both of whom are highly practiced at performing contour detection tasks, served as subjects. They performed a two alternative forced choice task: two images were sequentially presented (500ms presentation time, separated by a 1.0 second delay) and subjects were required to judge which interval contained a structured path. Graphed data show per cent correct performance over one hundred trials, and error bars show ±1 standard error.

EXPERIMENT 1 –DOES CONTOUR LINKING RELY ON FEATURES OR FOURIER STRUCTURE?

We indicated above that there was some evidence that grouping across scale seems to precede grouping across orientation. One might therefore conclude that contour integration must therefore be a relatively "high-level" process operating on spatial features that are the result of quite complex visual analysis. To put it another way, is it simply the similarity in appearance of local contour structure that determines strength of association across the visual field? We have reason to think that this is not the case, and that the grouping processes preceding (or concurrent with) contour integration are probably limited to inter-scale combination. We can certainly demonstrate the precedence of contour linking processes over other types of combination rules by using static plaids of the sort employed by Georgeson and Meese (Georgeson, 1992, 1994; Meese & Georgeson, 1996; Georgeson & Meese, 1997) and depicted in Figure 1a. Although this plaid is composed of the sum of two oblique gratings its appearance is dominated by non-Fourier "features": horizontal and vertical "checkerboard" structure (Georgeson, 1992). Which orientation does the contour integration system rely upon; features or Fourier structure?

We can answer this question using two facts: first, that the orientation of perceived features and Fourier structure in these patterns differs by 45°, and second, that contour integration is known to be difficult when element orientation differs substantially from local contour direction. We measured detectability of paths composed of plaids which had either their

Fourier orientation (Figure 1b) or their feature orientation (Figure 1c) aligned with the contour direction. Thus, in each case the orientation of the other component is tilted 45° away from the contour orientation. This component will be effectively useless for detecting the contour (which we confirmed by determining that contour detection is at chance for contours composed of Gabor elements rotated through 45° relative to the path orientation). Plaid paths were embedded in a field of randomly oriented plaid elements. Note that these textures are inherently more difficult than standard path stimuli of the sort used by Field *et al.* (1993) because there are now twice as many potential orientation matches for every element in the pattern. However, it is clear from the figure that it is the Fourier structure of contour elements, and not their spatial features, which determines the strength of association. Our experiment bore this out; for detection of 8 element long straight paths (500 ms presentation) the authors averaged 75% correct detection with co-aligned Fourier structure and 49% with co-aligned perceived local orientation.

This demonstrates that contour integration seems to operate on quite low-level Fourier descriptions. However we have seen that there is indirect evidence that scale combination might precede contour integration (Dakin & Hess, 1998). This hypothesis is clearly open to more direct testing by manipulating the local spatial frequency structure of contour elements. In this paper we do just that by measuring the detectability of contours composed of mixtures of broad-band elements (compound Gabors and edges) and narrow-band Gabor elements.

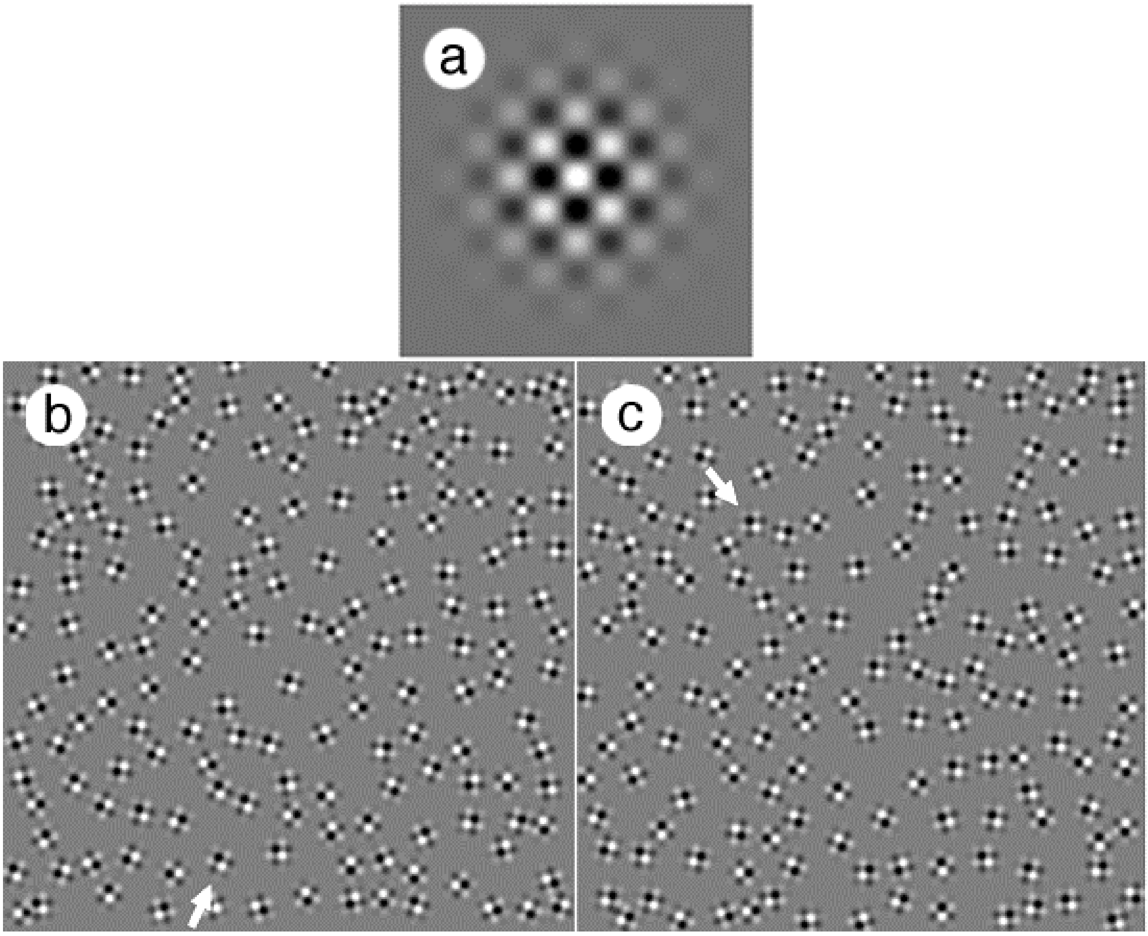

Figure 1. (a) A Gaussian windowed plaid which, although composed of the sum of two oblique gratings, produces a percept of horizontal and vertical "checkerboard" structure. (b) Contours composed of plaid elements with local contour orientation aligned with either (b) one of the grating components or (c) the perceived checkerboard structure. The detectability of contours composed of plaids clearly depends on the Fourier components of the plaids and not their appearance.

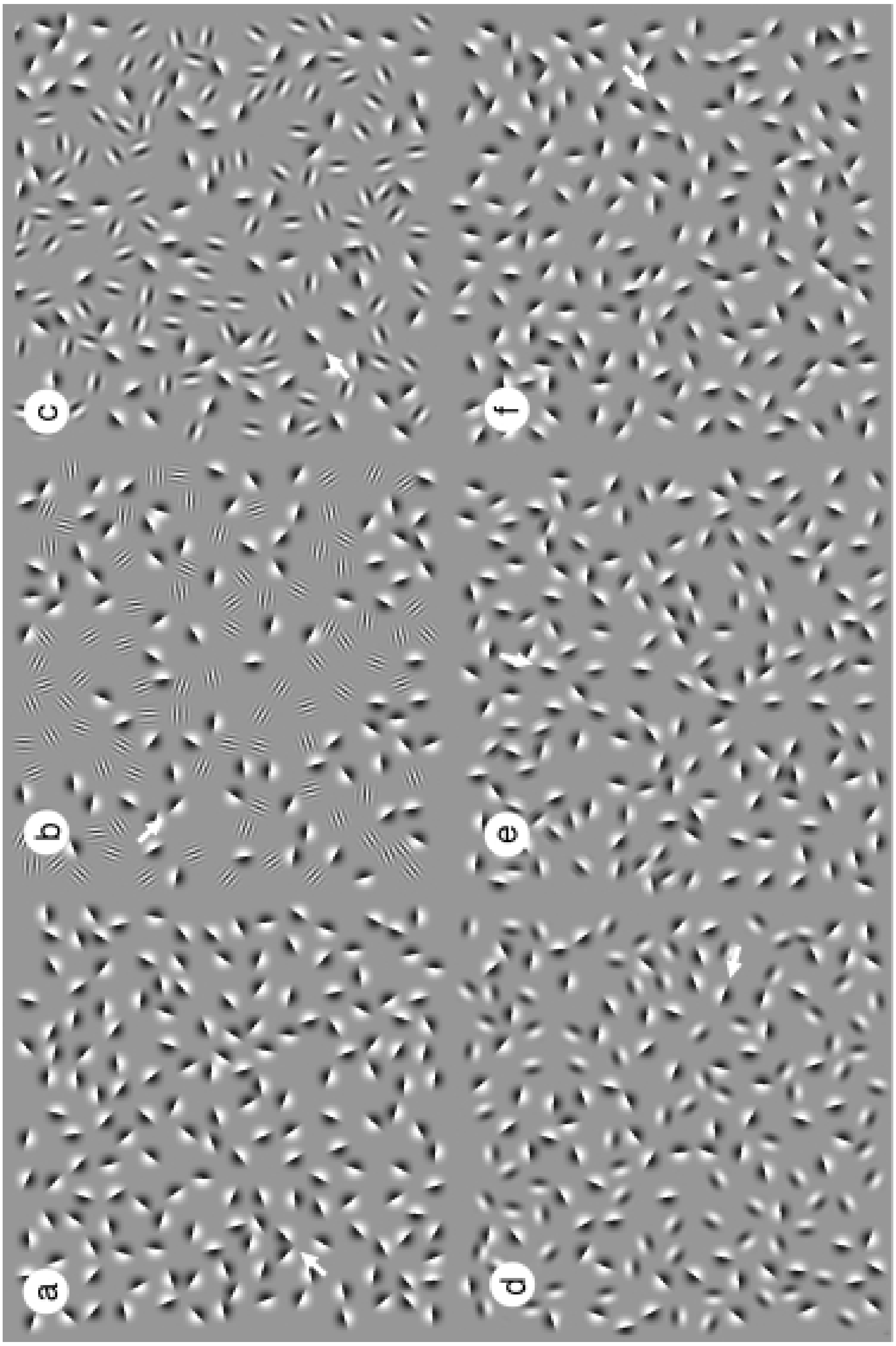

Figure 2. (a) Stimuli from Experiment 2. (a) Shows a path composed of local edge elements with a path angle of 20 degrees. (b-c) Shows the effect of interleaving edge elements with Gabors with carrier spatial frequencies of (b) 6.4, (c) 2.69 and (d) 1.6 cpd. Paths are more detectable with mid-scale Gabors (c). (e,f) show edges interleaved with (e) "f+3f" and (f) "f+3f+5f" micro-patterns. Notice that path detection is facilitated by the addition of phase-aligned structure

Experiment 2 - Detection of Contours Containing Local Edge Structure

This experiment assessed the detectability of paths composed of locally broadband edges. In order to determine observers' ability to integrate contours based on the individual spatial frequency components of edge stimuli, we measured the detectability of contours composed of edge elements alternating with simple Gabors or with compound Gabors in square-wave phase ("f+3f" patterns). Examples of the stimuli are shown in Figure 2. Details of the procedure and micropattern generation are given in the General Methods section above.

Our hypothesis is as follows. If local edge detection (i.e. filter combination across scale) comes *after* contour integration then subjects should be able to ignore the distracting influence of local edge structure on this task. This will manifest itself as generally good performance with little dependence on the spatial frequency of the narrow-band elements. If on the other hand, edge detection *precedes* contour detection, edge structure will interfere with contour integration and performance will be reduced.

Results

Subjects' performance with stimuli composed exclusively of local edges (data are indicated by the dashed horizontal line in Figure 3) is similar to their performance with 20° paths composed of narrow-band Gabors. However, when detection performance is measured using paths composed of alternating edge and Gabor elements (data are shown by circles in upper panels of Figure 3) there is a notable drop in peak performance from around 90% to 77%. Edges contain fractally weighted components, which should equally excite a range of neural mechanisms (Field, 1987). If the contour integration system operates independently, at multiple spatial scales, one should be able to focus on the output of the filter that is relevant to the task at hand. In this case tuning should be broad, and should peak at similar levels to performance with a contour composed of only Gabors (around 91% for contour elements at 3.2 cpd; Dakin and Hess, 1998). That tuning is narrow and peaks at a lower level indicates that contour integration does not operate in this manner.

To investigate if it is the absence of phase aligned structure in narrowband elements that leads to the observed drop in performance, we ran the same experiment with contours composed of edges alternating with (a) "f+3f" and (b) "f+3f+5f" patterns. Examples are shown in Figure 2 e,f. The parameters of the components of the micropatterns were selected so that the patches' log-average spatial frequencies matched the spatial frequencies tested in the first condition. The log-average of a set of n measurements is defined as:

$$\bar{f} = \exp\left(\frac{1}{n}\sum_{i=1}^{N}\ln(f_i)\right) \qquad \textbf{(6)}$$

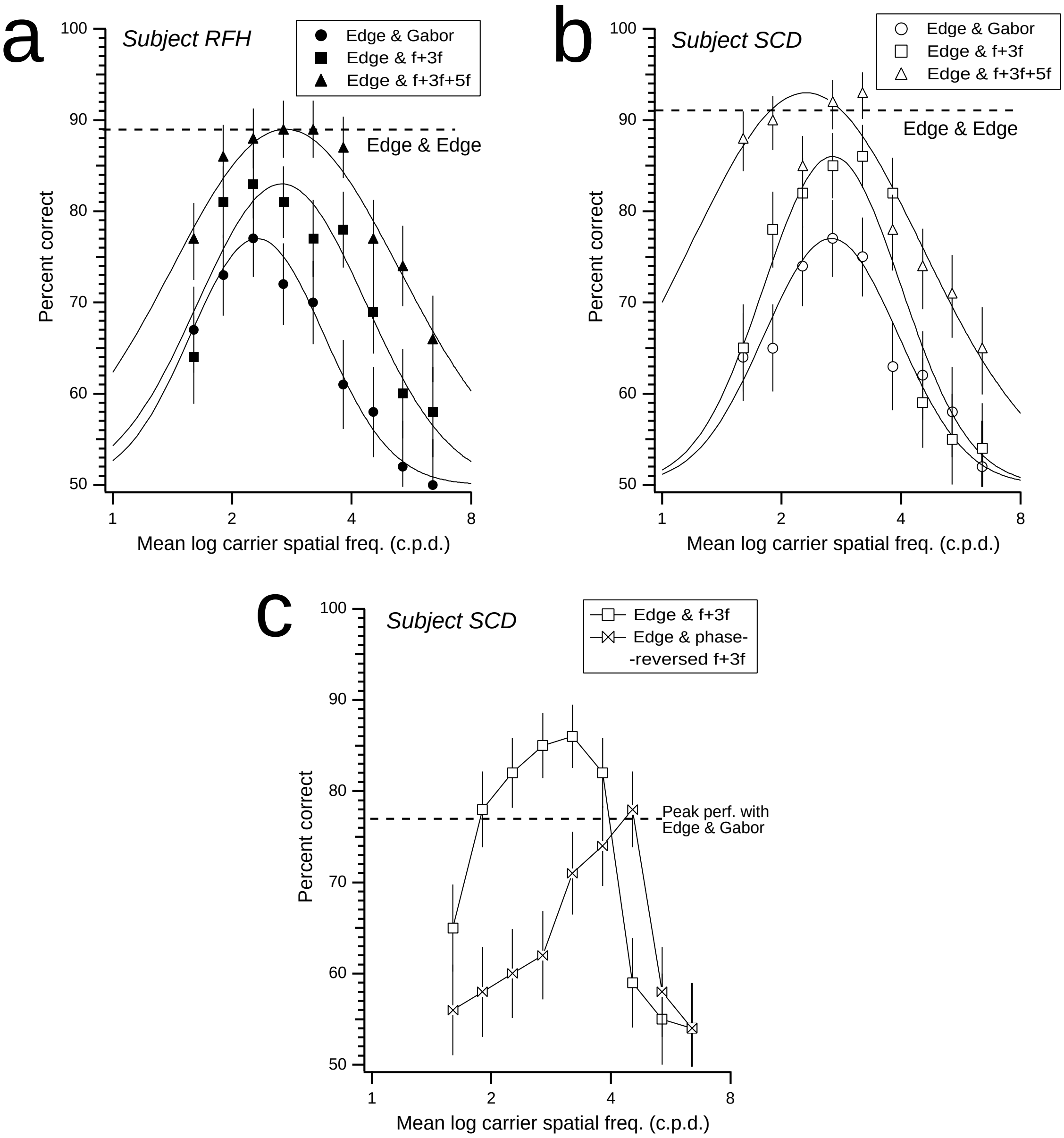

Figure 3. Upper panels show spatial frequency tuning for the detection of paths composed of alternating "edge" and "f", "f+3f" or "f+3f+5f" Gabors patterns. Percent correct performance is plotted as a function of the log-average spatial frequency of the Gabor or compound micro-patterns. Data from RFH and SCD are represented as filled and open symbols respectively, fits are Gaussians (operating on log-transformed data; Dakin and Hess, 1998). The dashed line shows performance with contours composed only of edge elements. Performance in all conditions peaks at spatial frequencies of 2-3 c.p.d. Peak performance improves from the "f" condition (mean of 76.5%), to "f+3f"(83.5%), but only approaches "edge-edge" performance (90%) for the "f+3f+5f" condition (89%). (c) shows a replot of data from SCD for the Edge versus "f+3f" compared to data from a control condition where the phase of the low frequency component of the "f+3f" patch is reversed. Performance now approaches that for contours composed of interleaved edges and simple Gabors, suggesting that the advantage conveyed by the presence of additional frequency components depends on their degree of phase alignment.

Specifically, the fundamental frequencies used (the “f” components) were 0.93, 1.14, 1.28, 1.55, 1.85, 2.19, 2.63, 3.1, and 3.68 c.p.d. for the “f+3f” condition, and 0.71, 0.88, 1.00, 1.20, 1.43, 1.70, 2.02, 2.42, and 2.86 c.p.d. for the “f+3f+5f” conditions. Performance improves under these conditions (square and triangle symbols in the upper two tuning curves in Figure 3) peaking at similar carrier spatial frequencies as the edge-Gabor condition and, in the case of the edge-”f+3f+5f” condition, at peak levels similar to the edge-edge condition. These data suggest that the edge-structure used by the contour integration process is effectively limited to frequency components spanning approximately 2.3 octaves.

In order to confirm that it is the presence of phase-aligned inter-scale structure which improves integration of edges with other broadband orientation elements we performed a control experiment. We tested if the presence of multi-scale information that is misaligned in phase would also improve performance. Contours were composed of edges and “f+3f” elements (with spatial frequencies of 1.85 and 5.55 c.p.d.) where the fundamental component was phase-reversed. Results for one subject (SCD), shown as bow-tie symbols in the lower panel of Figure 3, indicate that performance is poor compared to the edge vs. phase-aligned “f+3f” condition. Peak performance is now similar to the edge vs. Gabor condition suggesting that phase misalignment can break the linking process that operates between orientation components at different scales. The x-position of the peak, at around 4.5 cpd., is shifted towards the high frequency component, which dominates the appearance of such micropatterns. Thus our data indicate that it is lack of phase aligned structure across scale in the Gabor patches that leads to poor performance with edge-Gabor contours.

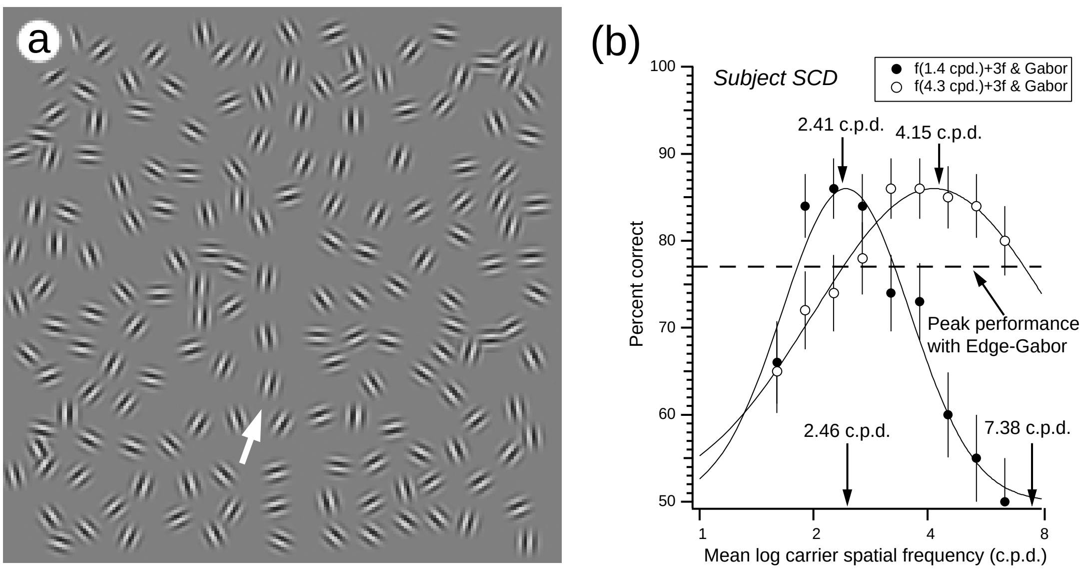

Figure 4. (a) Path stimuli composed of “f+3f” micropatterns (f=4.26 c.p.d. under experimental viewing conditions) interleaved with Gabors with carriers at 4.5 c.p.d.(b) Detection of “f+3f”-Gabor paths as a function of the spatial frequency of the simple Gabor, at two different frequencies of the fundamental of the compound micropatterns. Notice that the leftmost tuning function peaks at the log-mean spatial frequency of the “f+3f” patterns, but peaking of the rightmost curve nearer the fundamental frequency when the “3f” component becomes fine.

What is it about an edge’s inter-scale structure that prevents it from being effectively linked to narrow-band elements? We sought to determine if it is the presence of *any* phase alignment across scale or if, alternatively, it is that edges present alignment across a *continuum* of scales. To test this hypothesis we measured spatial tuning for “f+3f” vs. Gabor patterns using “f+3f” patterns with two different fundamental spatial frequencies (1.42 and 4.26 c.p.d.). An “f+3f” compound pattern contains phase-aligned structure but this structure is not present over the continuous range of scales that it is in an edge micro-pattern. Results for subject SCD are shown in Figure 4b (the other subject showed a similar pattern of results). Performance is better with “f+3f”-Gabors than with edge-Gabor contours, indicating that the contour linking process is sensitive both to the presence of a phase alignment and to the scale range over which that alignment extends. Taken together with data from the previous condition, this experiment suggests that contour integration cannot ignore local edge structure and therefore does not operate independently within individual spatial frequency bands.

EXPERIMENT 3 – BANDWIDTH & EDGE DETECTION

There is a second point of interest in the data shown in Figure 4. While tuning for the “f+3f” patterns with a low frequency “f” component (filled symbols) is centred on the log-mean frequency (as expected), tuning for micropatterns with a higher frequency “f” component (open symbols), is centred nearer the fundamental of the compound patterns (4.26 cpd.). This indicates that there is little or no contribution of the high frequency component when the fundamental is at the higher frequency tested. Examination of an example stimulus from this condition, with a Gabor near the optimal spatial frequency for this task (Figure 4a), reveals why the 3f component has so little influence. Patches in Figure 4a are composed of Gabors and f+3f compound patterns, but it seems that their similar *bandwidth* plays a significant role in the strength of their grouping. It is therefore possible that the observed tuning for edge-Gabors has more to do with differences in bandwidth of micro-patterns than with differences in the amount of local phase alignment. From this perspective, improved grouping for “f+3f” patches could be due to their being less periodic (i.e. less cycles of their carrier are visible) than spatially narrow-band Gabor micro-patterns. Are changes in element periodicity/bandwidth interfering with grouping processes along the contour?

To examine the importance of bandwidth in the edge linking tasks we measured tuning for paths composed of broad-band edges alternating with (aperiodic) spatially band-limited edges (Figure 5a) which, appear as a windowed pair of bright and dark bars regardless of the spatial scale. These patterns were composed of a first derivative of a Gaussian in the x-direction multiplied by a Gaussian in the y-direction:

$$f'(x,y) = -\frac{x}{\sigma_x^2}\exp\left(\frac{x^2}{2\sigma_x^2}+\frac{y^2}{2\sigma_y^2}\right) \tag{6}$$

Such micropatterns are spatially band-limited but have a much broader spatial bandwidth than Gabor micropatterns. As illustrated in Figure 5a, they are similar in appearance to Gabor patterns whose envelopes have been truncated in a direction perpendicular to the carrier leaving only one duty cycle of their carriers visible.

In the following experiment we employed band-limited edges with the same peak spatial frequencies as used previously (i.e. 1.6-6.4 c.p.d. in quarter octaves steps).

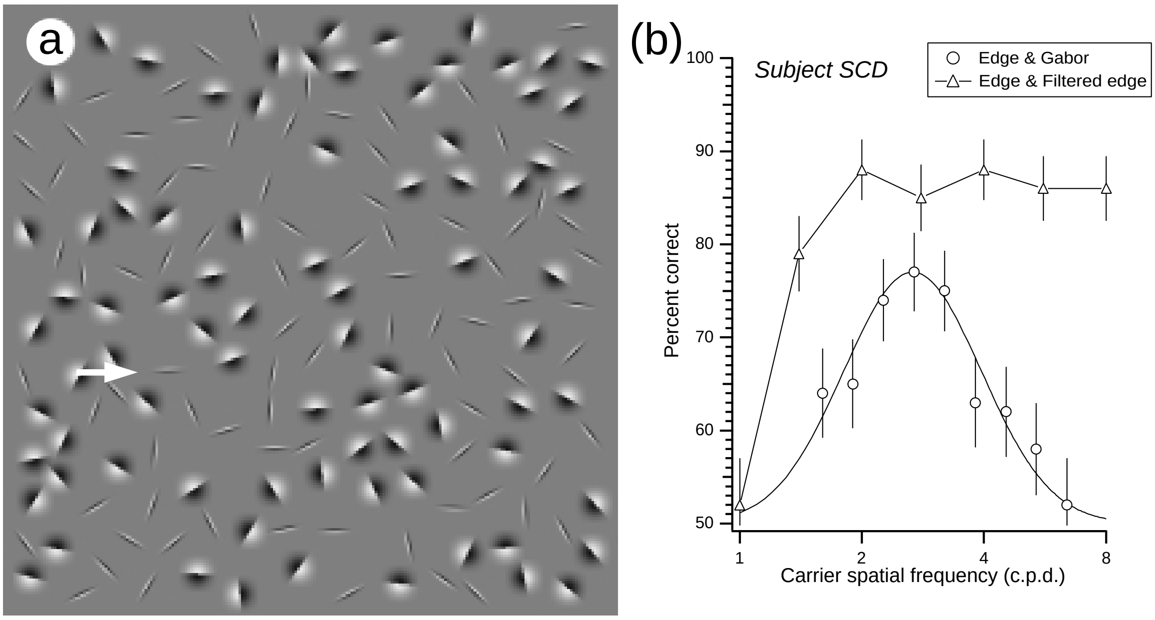

Figure 5. (a) Path stimulus composed of edges alternating with high-frequency (8 c.p.d.) first-derivative edges. Subjects are able to integrate such mixtures. (b) Compares detection of edge-Gabor paths (circles) to edge-filtered-edge paths (triangles). The latter produce high-pass tuning with peak performance approaching that for edge-edge contours (around 92% for this subject).

Inspection of the example in Figure 5a, which shows a high frequency filtered edge (8 c.p.d.) alternating with broad-band edges, demonstrates that linking to high frequency elements is easier in this condition than with edge-Gabor stimuli (compare to Figure 2b). Results (Figure 5b) confirm this. Tuning is now high-pass indicating that subjects have access to a much wider range of spatial scales from the edge for linking.

This difference seems likely to be due to the reduced periodicity of the luminance component of first-derivative patches, compared to Gabors. For Gabors, this reduction in periodicity equates to an increase in bandwidth (or, alternatively, an increase in envelope-size in the direction perpendicular to the carrier). Previously we have reported that bandwidth alternation along the contour has little effect on detection of paths composed of Gabor elements, but rather that it was spatial frequency which was the main determinant of association strength between contour elements (Dakin & Hess, 1998). These data question that conclusion; that edge tokens can be matched to filtered versions of themselves suggests that there may be a role for micro-pattern bandwidth in edge-integration.

In order to address this issue we performed a final experiment to determine if it is the broadband nature of edges, rather than locally phase-aligned structure, which determines their poor association with Gabors. We used micropatterns composed of “f+3f+5f” compound Gabors which, when components are phase-aligned, produce a strong perceptual match to local edge stimuli (Experiment 2). However in this experiment we phase randomised the components so that elements had similar spatial bandwidth to edges but did not contain phase alignments. We then measured detection of paths composed solely of these elements, and of paths containing mixtures of these random-phase compound elements with ordinary Gabors.

Examples of the stimuli are shown in Figure 6a,b and results from two subjects in Figure 6c,d. Graphs compare performance for edge-Gabor paths (circles) with detection of paths composed of Gabors interleaved with phase-scrambled compound elements. Data indicate that the latter are clearly easier to detect. This is an interesting and quite counterintuitive finding: the stimuli used in this experiment are very similar to the edge-Gabor stimuli, except that the local phase structure of the edges has been scrambled. That introducing a large amount of variability in the structure of the micropatterns should improve performance can only indicate that it is the phase alignment of edge elements, and not their bandwidth, that disrupts association of contour elements. This resolves the apparent discrepancy with our previous finding that bandwidth is not critical for linking Gabor elements. Bandwidth is only an issue for contour integration in the presence of significant local phase alignment across spatial frequency.

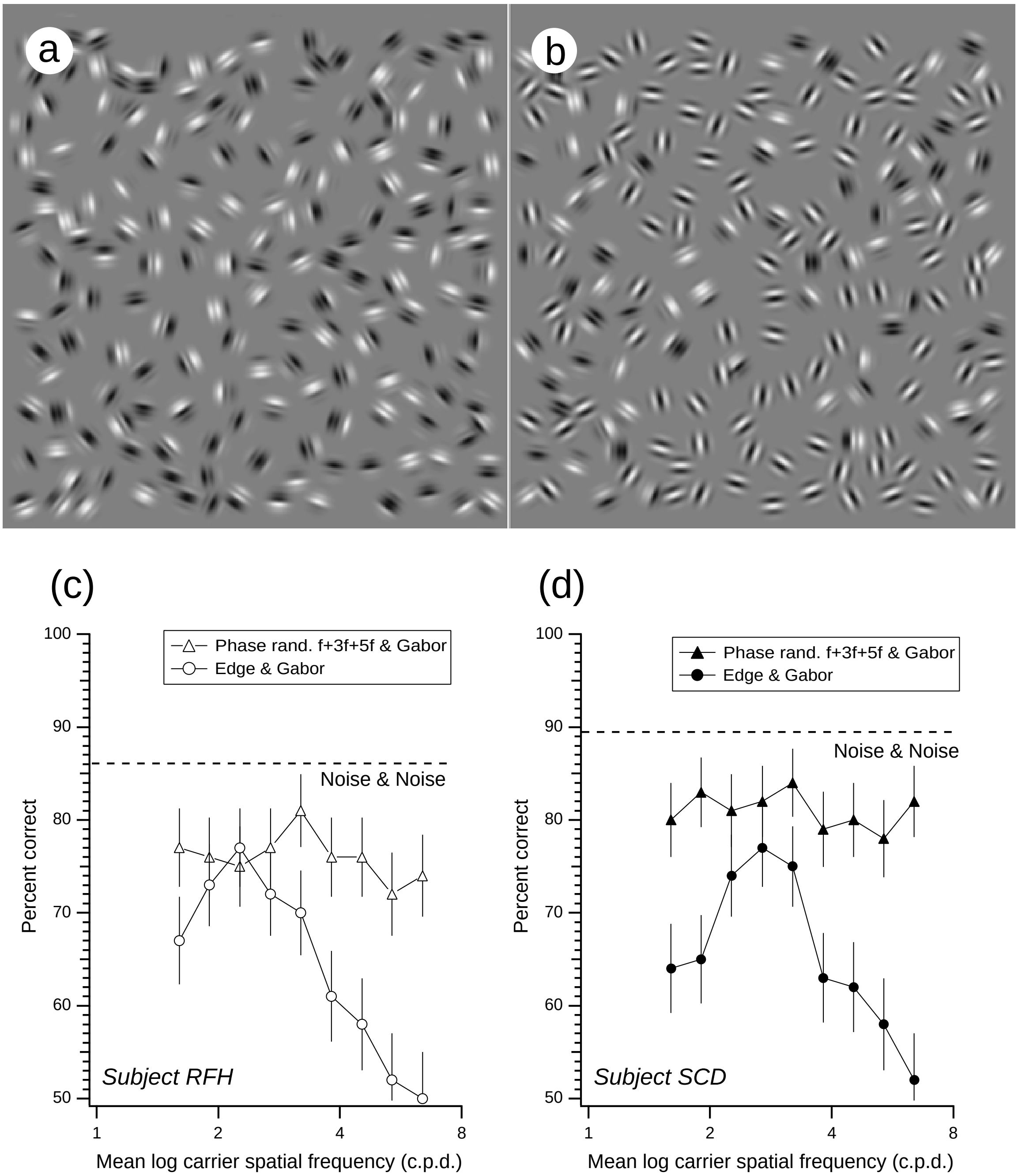

Figure 6 (a) Path stimulus composed of phase randomised "f+3f+5f" compound Gabor micro-patterns, (b) intermixed with Gabors at 3.2 c.p.d. (c) Detection of randomized compounds and Gabor paths for RFH and (d) SCD. Notice that performance is superior to the identical case where components of the broadband stimulus are phase-aligned (circles).

GENERAL DISCUSSION

To summarise, we have shown the following:

- Contour linking operates on relatively low-level Fourier descriptions; Fourier energy, is a better predictor of the strength of association between two contour elements than their appearance.

- Spatial tuning for paths composed of edges and Gabors is band-pass, and peak performance is poor. Contour integration does not appear to operate independently within spatial frequency bands.

- Spatial tuning for paths composed of edges and filtered edges is high-pass, and peak performance is good. Element bandwidth is relevant to the spatial integration of edge structure.

- Phase scrambling the components of edges actually improves subjects' ability to link them to Gabors and suggests that the importance of element bandwidth for contour integration is linked to the degree of phase alignment of local orientation structure.

We now consider the implication of these findings both for models of contour integration and for the role of scale combination in early visual processing.

Phase tracking: Implicit or explicit?

Block-quantized images are hard to recognise (Harmon & Julesz, 1973). Disrupting the phase relationship between the high-frequency components (introduced by blocking), and the low-frequency components (of the original image) facilitates recognition. This demonstrates that the phenomenon, rather than being the result of masking, is due to the continuity of phase information across scale (Canny, 1983; Morrone & Burr, 1983, 1997; Hayes, 1989). This suggests that the visual system could be "tracking" across locally contiguous phase structure, to the finest spatial scale producing significant local filter activity. However, a purely local and automatic version of such a "phase tracking" scheme will always descend to the finest scale at a given location. In terms of our stimuli it would therefore predict high-pass tuning for all contours which include edge elements, rather than the band-pass spatial tuning we observed for edge-Gabor contours in Experiment 2, which peaked at around 2-3 c.p.d. This suggests that phase-tracking must take local context into account, i.e. operate on a contour-by-contour basis, which poses a correspondence problem between scales. The alternative is that phase-tracking is not performed as an explicit matching process, but is implicit in the operation of orientation detectors that are tuned to local phase-tracks (i.e. edge detectors). The latter scheme (which we return to below) is consistent with an emerging consensus that human vision generally relies on filtering operations to solve correspondence problems (e.g. in the motion domain, Adelson & Bergen, 1985).

Early filter combination

Rigid, early inter-scale combination is a notable characteristic of the MIRAGE scheme for encoding pattern using spatial filters (Watt & Morgan, 1985; Morgan & Watt, 1997) that sets it apart from other schemes, such as the "local energy" model (Morrone & Burr, 1988). Although aspects of this scheme have been extended from one- to two-dimensions (such as the image description stage; Watt, 1991), the problem of 2D filter combination has not. The work reported here is consistent with MIRAGE-style filter combination operating within orientational channels, prior to the type of filter combination across orientation that presumably subserves detailed shape analysis.

A general hybrid model

Our results imply that contour integration can operate between narrow band detectors when phase alignments are not present or between broad band detectors utilizing phase alignments. One plausible reason for this seemingly complicated finding is that there may be competition between these two detector classes at an early stage of processing prior to contour integration. This would follow if one assumes that the goal of early visual processing is to produce a sparse code (Olshausen & Field, 1996, 1997) i.e. to maintain activity of as few cells as possible. It would therefore not seem unreasonable if there were locally competitive interactions(winner-take-all or soft-maximisation) between cells. In the case of edges, "hard-wired" edge detecting neurons may inhibit more narrowly tuned cells to produce a more economic code. Why code an edge with an array of spatial filters when a single edge-detecting neuron will do? This code would then form the input to the linking process. If both the tuning of detector neurons and the interactive links between them are determined by the visual diet, the fact that contours in natural scenes are unlikely to vary greatly in bandwidth along their length may lead to a lack of association between edge detectors and more narrowly tuned neurons. This would explain the observed dichotomy in contour integration between phase-aligned and phase-insensitive structure.

Neural encoding of edges: a role for side stopping?

There are a number of ways in which one could endow neurons with the ability to selectively respond to edges. For example, it is known that complex cells respond to the difference frequency of compound sinusoidal stimuli which would make them candidates for coding local phase alignments (Pollen, Gaska & Jacobson, 1988). Alternatively, many striate cortical cells are tuned not only for the length of stimuli falling in their receptive fields (“end-stopped” cells), but also for their width (Bishop, Coombs & Henry, 1973). The response of such “side stopped” cells to gratings decreases when the number of bars falling in their receptive field exceeds some optimal value. Such inhibition is orientation-tuned and has been observed in both cat (Bishop, et al, 1973; Blakemore & Tobin, 1972; Maffei & Fiorentini, 1976; Fries, Albus & Creutzfeldt, 1977; Nelson & Frost, 1978) and monkey (De Valois, Thorell & Albrecht, 1985; Born & Tootell, 1991). Born & Tootell (1991) report that 70% of cells in inter-blob regions of layers 2 and 3 of macaque striate cortex show side stopping. These authors have suggested that the role of side-stopping in cortical cells may be to remove the influence of textured regions on later processing and to retain only orientation information arising from contours. Given that layers 2 and 3 form the major projections to areas which are responsible for the processing of form (i.e. V2 and beyond) this would be an appropriate site to isolate contour information from texture.

Born & Tootell (1991) also point out that an illusion first demonstrated by Galli & Zama (1931) (reproduced in Figure 7a) illustrates this dissociation of texture from form. It is difficult to see the hexagon embedded in Figure 7a because the oblique gratings engulf the contour defining it. Instead we see a partially occluded square; an interpretation that is consistent with a dissociation of contour from texture. Clearly another way of interpreting the “texture versus contour” dichotomy is “narrow versus broad-band orientation”, and this inability to link across them is consistent with our findings for contours composed of edges mixed with Gabors. At present it is not known if cells in layers 2 and 3 of primary visual cortex respond selectively to edge structure, but if they do it could be that side-stopping is simply a cortically efficient way of encoding local bandwidth in the context of contour integration.

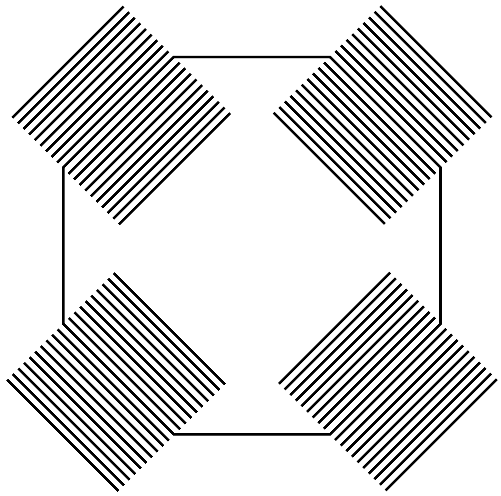

Figure 7. It is difficult to tell that this figure contains a hexagon defined by a single continuous contour. Instead one tends to perceive a square whose corners are occluded by four gratings. (After Born & Tootell (1991) and Galli & Zama (1931))

REFERENCES

Adelson, E. H. & Bergen, J. R. (1985). Spatio-temporal energy models for the perception of motion. *Journal of the Optical Society of America A, 2*, 284-299.

Bishop, P., Coombs, J. & Henry, G. (1973). Receptive fields of simple cells in the cat striate cortex. *Journal of Physiology, London, 231*, 31-60.

Blakemore, C. & Tobin, E. R. (1972). Lateral inhibition between orientation detectors in the cat's visual cortex. *Experimental Brain Research, 15*, 439-440.

Born, R. T. & Tootell, B. H. (1991). Single unit and 2-deoxyglucose studies of side-inhibition in macaque visual cortex. *Proceedings of the National Academy of Science, 88*, 7071-7075.

Burr, D. C. & Morrone, M. C. (1994). The role of features in structuring visual images. In M. J. Morgan (Ed) *Higher-order processing in the visual system: CIBA Foundation Symposium 184* (pp. 129-141). Chichester, UK: John Wiley and Sons.

Canny, J. F. (1983). Finding edges and lines in images. Boston, MA: MIT AI Laboratory Technical Report.

Dakin, S. C. & Hess, R. F. (1998). Spatial-frequency tuning of visual contour integration. *Journal of the Optical Society of America, 15*.

De Valois, K. K., De Valois, R. L. & Yund, E. W. (1979). Responses of striate cortex cells to grating and checkerboard patterns. *Journal of Physiology (London), 291*, 483-505.

De Valois, R. L., Thorell, L. G. & Albrecht, D. G. (1985). Periodicity of striate-cortex-cell receptive fields. *Journal of the Optical Society of America, A2*, 1115-1123.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America, A4*, 2379-2394.

Field, D. J., Hayes, A. & Hess, R. F. (1993). Contour integration by the human visual system: evidence for a local "association field". *Vision Research, 33*, 173-93.

Field, D. J., Hayes, A. & Hess, R. F. (1997). The role of phase and contrast polarity in contour integration. *Investigative Ophthalmology and Visual Science, 38*, S999.

Fries, W., Albus, K. & Creutzfeldt, O. D. (1977). Effects of interacting visual patterns on single cell responses in cats striate cortex. *Vision Research, 17*, 1001-1008.

Galli, A. & Zama, A. (1931). *Zeitschrift Psychologie, 123*, 308-348.

Georgeson, M. (1992). Human vision combines oriented filters to compute edges. *Proceedings of the Royal Society of London, B249*, 235-245.

Georgeson, M. A. (1994). From filters to features: location, orientation, contrast and blur. In M. J. Morgan (Ed) *Higher-order processing in the visual system: CIBA Foundation Symposium 184* (pp. 147-163). Chichester, UK: John Wiley and Sons.

Georgeson, M. A. & Meese, T. S. (1997). Perception of stationary plaids: The role of spatial filters in edge analysis. *Vision Research, 37,* 3255-3271.

Harmon, L. D. & Julesz, B. (1973). Masking in visual recognition: Effects of two dimensional filtered noise. *Science, 180*, 1194-1197.

Hayes, A. (1989). Representation by images restricted in resolution and intensity range. In *Department of Psychology*: University of Western Australia.

Hess, R. F. & Dakin, S. C. (1997). Absence of contour linking in peripheral vision. *Nature, 390*, 602-604.

Hess, R. F. & Field, D. J. (1995). Contour integration across depth. *Vision Research, 35,* 1699-1711.

Kovács, I. & Julesz, B. (1993). A closed curve is much more than an incomplete one: effect of closure in figure-ground segmentation. *Proceedings of the National Academy of Sciences of the United States of America, 90,* 7495-7.

Lowe, D. G. (1988). Organization of smooth image curves at multiple spatial scales. In *Proceedings of the IEEE International Conference on Computer Vision.*, pp. 558-567. Los Alamitos, CA: The IEEE Computer Society Press.

Maffei, L. & Fiorentini, A. (1976). The unresponsive regions of visual cortical receptive fields. *Vision Research, 16,* 1131-1139.

Marr, D. (1982). *Vision*. San Francisco, California.: Freeman.

Marr, D. & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London, B207,* 187-217.

McIlhagga, W. H. & Mullen, K. T. (1996). Contour integration with colour and luminance contrast. *Vision Research, 36,* 1265-1279.

Meese, T. S. & Georgeson, M. A. (1996). Spatial filter combination in human pattern vision: Channel interactions revealed by adaptation. *Perception, 25,* 255-278.

Morgan, M. J. & Watt, R. J. (1997). The combination of filters in early spatial vision: a retrospective analysis of the MIRAGE model. *Perception, 26,* 1073-1088.

Morrone, M. C. & Burr, D. C. (1983). Added noise restores recognition of coarse quantized images. *Nature, 305,* 226-228.

Morrone, M. C. & Burr, D. C. (1988). Feature detection in human vision: a phase-dependent energy model. *Proceedings of the Royal Society of London, B235,* 221-245.

Morrone, M. C. & Burr, D. C. (1997). Capture and transparency in coarse quantized images. *Vision Research, 37,* 2609-2629.

Nelson, J. I. & Frost, B. J. (1978). Orientation-selective inhibition from beyond the classic visual receptive field. *Brain Research, 139,* 359-365.

Olshausen, B. A. & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, 381,* 607-609.

Olshausen, B. A. & Field, D. J. (1997). Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision Research, 37,* 3311-3325.

Olzak, L. A. & Thomas, J. P. (1991). When orthogonal orientations are not processed independently. *Vision Research, 31,* 51-57.

Olzak, L. A. & Thomas, J. P. (1992). Configural effects constrain Fourier models of pattern discrimination. *Vision Research, 32,* 1885-1898.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming number into movies. *Spatial Vision, 10*, 437-442.

Pettet, M. W., McKee, S. P. & Grzywacz, N. M. (1996). Smoothness constrains long-range interactions mediating contour-detection. *Investigative Ophthalmology and Visual Science, 37,* 4368.

Pollen, D. A., Gaska, J. P. & Jacobson, L. D. (1988). Responses of simple and complex cells to compound sine-wave gratings. *Vision Research, 28,* 25-39.

Thomas, J. P. & Olzak, L. A. (1996). Uncertainty experiments support the roles of second-order mechanisms in spatial frequency and orientation discriminations. *Journal of the Optical Society of America, 13,* 689-696.

Torre, V. & Poggio, T. A. (1986). On edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 8,* 147-163.

Watt, R. J. (1988). *Visual processing: Computational, psychophysical and cognitive research.* London: Lawrence Erlbaum Associates Ltd.

Watt, R. J. (1991). *Understanding Vision*. London: Academic Press.

Watt, R. J. & Morgan, M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Research, 25*, 1661-1674.

Wertheimer, M. (1938). *Laws of organisation in perceptual forms*. London: Harcourt, Brace and Jovanovich.

Acknowledgments. This research was supported by MRC grant MT108-18 to R.F. Hess. We are indebted to David Field and Stephane Rainville for helpful discussions of this project.