

Carving up the Patchwise Transform: Towards a Filter Combination Model for Spatial Vision

Reference details:

Meese, T. S. & Georgeson, M. A. (2005) Carving up the patchwise transform: Towards a filter combination model for spatial vision. *Advances in Psychology Research*, Volume 34, pp 51-88, Nova Science Publishers, New York.

Please note, *Nova Science Publishers*, published this article without correcting the 90+ errors that they introduced during typesetting. This is the corrected proof of that article.

Tim S. Meese

Mark A. Georgeson

Neurosciences Research Institute, School of Life and Health Sciences, Aston University, Birmingham, B4 7ET, U.K.

Tel. 0121-264-4130

Fax. 0121-333-4220

e-mail: t.s.meese@aston.ac.uk m.a.georgeson@aston.ac.uk

Correspondence to Tim Meese.

Running Head: Towards a filter combination model

Key words: adaptive filtering, human vision, binding problem, segmentation, edge, zero-crossing, plaid, feature.

ABSTRACT

Local regions of the retinal image are processed by hypercolumns of quasi-linear filter-elements with overlapping receptive fields, each selective for different orientations and spatial scales of luminance contrast (Hubel & Wiesel, 1974, *J Comp Neurol*, **158**, 267-294). Because these filter-elements have fairly broad orientation- and spatial frequency-bandwidths, the presence of luminance contrast within a hypercolumn's receptive field typically stimulates several such mechanisms. Presumably, the visual system is faced with an inverse problem of determining how the overall distribution of activity within each hypercolumn should be bound and segmented into different neural assemblies so as not to confuse representations of multiple contours. Using subjective pattern matching tasks with stationary plaid stimuli, the present work aims to shed light on this parcelling operation. Results confirm and extend those of Georgeson & Meese (1997) (*Vision Research*, **37**, 3255-3271): binding can occur across orientation, but segmentation occurs across orientation when binding occurs across spatial frequency.

These and earlier results are discussed in the context of an image processing model which uses local image structure to switch gracefully from bound to segmented filter-elements by allowing each one to contribute to more than one assembly. This dynamic binding of filter-elements provides a good quantitative account of the changes in perceived structure that occur for stationary plaids under a variety of experimental conditions. The model also produces edge-maps (c.f. Marr & Hildreth, 1980; *Proc R Soc London B*, **207**, 187-217) that, unlike several other edge-detection algorithms, provide a good qualitative description of the spatial structure seen in a variety of plaids.

INTRODUCTION

In neural systems, binding and segmentation are the general problems of associating or dissociating the activities of two or more elements in a neural code. This is sometimes called the binding problem. This problem has many potential guises (e.g. see Treisman, 1996, Wolfe & Cave, 1999 and Howard, 2002, for reviews) and proposed solutions have been varied (e.g. von der Malsburg & Schneider, 1986; Morgan & Hotopf, 1989; Yen & Finkel, 1998). But how some of these solutions might be implemented and whether multiple solutions are used in human vision remains unclear (e.g. Gray, 1999; Shadlen & Movshon, 1999).

The reality of the binding problem is nicely illustrated by instances in which it fails (e.g. Wolfe & Cave, 1999; though see Howard, 2002 for some discussion). Moutoussis & Zeki (1997) found that for moving coloured targets and normal observers, false conjunctions can occur between the perceived position and perceived colour of the target, indicating a failure in the binding process. When attention is overloaded, Treisman and Schmidt (1982) found that for brief stimulus presentations, observers were sometimes confident but mistaken about the colour of a target object. Similar problems, but for much longer stimulus durations, have been reported in a patient with parieto-occipital lesions (Friedman-Hill et al, 1995; Robertson et al, 1997). For some amblyopic patients, elongated contours appear fragmented along their length (e.g. Hess, Field & Watt, 1990), suggesting that neural representations of local parts of the contour fail to be bound into a global whole. Another type of failure experienced by some strabismic patients is the inability to fuse the images from the two eyes into a whole, resulting in diplopia, or ‘double vision’.

Binding in the patchwise Fourier domain

When the binding problem involves population codes (Braddick et al, 1978; Meese & Georgeson, 1996a) it is ill-posed. This is because for any distribution of neural activity, the visual system must decide which parts of the distribution belong together, which belong apart and which, if any, should be allocated to multiple representations (Bregman, 1987; Meese & Georgeson, 1992). Presumably, there are numerous solutions to this problem, so what are the rules used by the visual system in guiding its decisions? In this paper we restrict our focus to illustrating and characterising the occurrence of these processes in the *patchwise Fourier domain*.

In mammalian foveal vision, the retinal image is processed by multiple quasi-linear spatial filters tuned for different orientations and spatial frequencies (e.g. Blakemore & Campbell, 1969; Watson, 1982). Each filter consists of a two-dimensional array of filter-elements (i.e. neurons), each processing a different region of the visual field. Thus, each region, or *patch*, of the image is examined by a collection of filter-elements with superimposed, spatially localised weighting functions selective for different spatial frequencies and orientations of luminance contrast (Robson, 1980). This early visual representation is sometimes described as the *patchwise Fourier transform*. The ability to detect ‘snake-like’ arrays of contours amongst noise (e.g. Field et al, 1993; Hess & Dakin, 1997) indicates that the visual system must solve the binding problem between different patches of this transform. In this chapter we elaborate on the notion that mammalian vision must also solve the binding problem *within* each patch of the transform (e.g. Engel, Konig & Singer 1991; Singer & Gray, 1995).

The specific problem is this. Within any single patch, activity might be due to the presence of either: (i) just a single (broad-band¹) object, for example, a checkerboard, or (ii) the superposition of two or more distinct visual features, such as an elongated contour crossed by a shadow boundary. Thus, at least in principle, the visual system faces the general problem of binding and segmenting the distribution of neural activity into one or more assemblies within each patch. In previous work on the perception of stationary plaids² (e.g. Georgeson, 1992; 1994; Meese & Freeman, 1995; Meese & Georgeson, 1992, 1996a, 1996b; Georgeson & Meese, 1997, 1999) we have reported many instances where the visual system appears to be performing a parcelling operation of this kind. For example, in several experiments, observers were presented with plaid stimuli made from pairs of overlapping luminance gratings of equal contrast but different orientations (e.g. $\pm 45^\circ$) and were required to respond by indicating which of two line-drawings best resembled the perceived spatial structure of the stimulus. In effect, these drawings (sketches) were the pattern of features revealed by zero-crossings³ (ZC) in the output of either: (a) the linear sum of all oriented filters or (b) two segmented assemblies of oriented filters,⁴ balanced around the orientations of the two plaid components. For the plaid stimulus described above, these two sketches look very different. The first sketch (the *compound* sketch) contains only vertical and horizontal contours, while the second one (the *component* sketch) contains contours only at the same left and right oblique orientations as the two stimulus components (e.g. see Meese & Georgeson, 1996a). When stimulus contrast was low, observers chose the component sketch, but when contrast was high, they typically chose the compound sketch (e.g. Meese & Freeman, 1995) because at low stimulus contrasts, plaid components look like overlapping gratings whereas at higher contrasts, they look like a blurred checkerboard (Meese & Freeman, 1995; Smith et al, 2001).

In another experiment (Meese & Georgeson, 1996a) it was found that while adapted to a vertical grating, a moderate contrast plaid stimulus retained its vertical and horizontal structure but appeared stretched horizontally. This finding is most

¹ i.e. a stimulus containing many Fourier components.

² A plaid is a stimulus made from the superposition of two or more sine-wave modulations of luminance contrast (gratings). The components might differ in spatial frequency, orientation, phase and contrast.

³ For a tutorial discussion of ZCs see Watt (1988).

⁴ The filter characteristics are of little importance here, though for a plaid stimulus with components separated by 90° , straight ZCs are found only if the orientation bandwidths are such that a filter does not respond to a component that is 45° or more, away from the filter’s preferred orientation.

easily understood if perceived structure is that which is carried by the inverse transform of the *full set* of oriented filter-elements within each patch. The response of some of the elements would be attenuated due to the selective effects of orientation adaptation, giving rise to the perceived distortion in the plaid's structure (for details see Meese & Georgeson, 1996a). On the other hand, in situations where the stimulus looks like two overlapping gratings (e.g. at very low contrasts; see Meese & Georgeson, [1996b] for a full list), we suggest that left and right oblique filters must be segmented so as to carry (almost) independent representations of the two overlapping structures. (A more detailed interpretation of our previous work is presented in Appendix A).

Thus, in sum, our preliminary proposal is that: (i) when a plaid looks like a blurred checkerboard there is binding between oriented filter-elements within each patch and a single neural assembly is created (though we shall refine this later), and (ii) when a plaid looks like a pair of overlapping gratings, segmentation occurs between two assemblies of filter-elements in each patch. We do not suppose, however, that these two states are discrete. One possibility that we elaborate upon later is that they are two ends of a continuum. Note that this general proposal avoids the issue of how, or even *if*, edge features are made explicit. Nevertheless, we illustrate the dynamic binding process that we propose with an edge-detection model that provides good qualitative and quantitative accounts of our data.

Aims

Previous experimentation with plaids suggests that the processes that govern binding and segmentation across oriented filters are flexible. For example, whereas a moderate contrast plaid looks like a checkerboard with blurred vertical and horizontal contours, the addition of a just detectable third-harmonic component in square-wave-phase with one of the fundamental components causes the plaid to segment perceptually into two overlapping gratings (Georgeson, 1994; Georgeson & Meese, 1997, 1999)—see figures 9a & 11a. The implication is that in the first case, there is summation between stimulus components and the oriented filters that process them, whereas in the second case, the same fundamental components and filters are treated independently (see Appendix A for further details). Certainly, as illustrated elsewhere (Georgeson & Meese, 1999), several other models, including some of those that extract ZCs (Marr & Hildreth, 1980; Watt & Morgan, 1986; Canny, 1986), do not process these plaid stimuli in the same way as human vision.

Our emphasis then, is on the need for vision to perform binding and segmentation within the patchwise Fourier transform prior to the feature (e.g. edge) extraction process. The experiment presented here further explores the conditions under which the addition of harmonic components leads to perceptual segmentation in plaids.

The present work has three main objectives. 1) To present further data on the subjective appearance of stationary plaid patterns. 2) To summarise our conclusions based on previous findings with plaids (Appendix A). 3) To present an image-processing model of spatial vision that can emulate several of our findings with plaids (Model Section).

Equipment and Methods

Any one stimulus component $I(x,y)$ is defined by its contrast (c), spatial frequency (f), orientation (θ) and phase (ϕ), related by the equation:

$$I(x, y) = L_0 \{1 + c \cdot \sin(ux = vy - \phi)\}$$

where $u = f \cdot \cos(\theta)$, $v = f \cdot \sin(\theta)$ and L_0 is mean luminance. A plaid is the sum of two or more such components at different orientations.

Stimuli were generated using an Innisfree Picasso Image Synthesizer with a frame rate of 240 Hz under the experimental control of an Acorn Archimedes 440 computer and were displayed on a Tektronix 608 oscilloscope with green phosphor (P31). The sinusoidal stimulus components were produced by modulating the z-axis of the oscilloscope at a spatial frequency determined by the manual setting of vernier dials controlling the two oscillator channels of the Picasso.

All plaid stimuli consisted of a base plaid with additional higher harmonics. The base components had spatial frequencies of 1 c/deg and were oriented at $\pm 45^\circ$ (referred to as a 'standard plaid'). The harmonic components were also oriented at $\pm 45^\circ$ and had spatial frequency of either 3 c/deg, 5 c/deg or 7 c/deg. The harmonic components could be summed with the base plaid in one of two relative phases: sine-phase (e.g. square-wave-phase for the third-harmonic, and negative sine-phase (e.g. triangle-wave-phase for the third-harmonic). This extends previous work (Georgeson & Meese, 1997), where only a single harmonic component was used and the range of conditions was less extensive.

Component orientation was under computer control and was achieved by the Picasso rotating the raster scan. Plaids were generated by temporally interleaving stimulus components and so required the raster scan to be rotated between frames. It took the software two frames to calculate the rotation and instruct the hardware, giving a picture refresh rate of 60 Hz for all plaid stimuli.

Component contrast was also under computer control and is reported as Michelson contrast of a single component in percent. The display field had a mean luminance of 17 cd/m² and was shown to be linear up to a contrast of about 50% by routine luminance and contrast calibration using a Photodyne digital photometer. Stimulus duration was 500 ms and the global phase of the stimulus was randomised from trial to trial. Brief stimulus presentations are particularly important because with prolonged stimulus exposure (as occurs when inspecting an image on the printed page for example), perceptual instability is often experienced. This is probably due to constructive and destructive interference between the printed image and its retinal afterimage whose relative positions on the retina fluctuate due to small eye-movements. This phenomenon has

sometimes been referred to as monocular rivalry and its deleterious effects were avoided in all of our experiments by using brief stimulus durations (Georgeson, 1984; Georgeson & Turner, 1985).

The circular display field had a diameter of 2.5 deg and the viewing distance was 228 cm.

Observers were the two authors (TSM and MAG), and two colleagues who were naive to the purpose of the experiments but were familiar with plaid stimuli (AL and TCAF). All observers had normal or corrected to normal vision and viewed the display binocularly with natural pupils and the aid of a chin and forehead rest.

Model simulations

The software for the model simulations reported in the Discussion was written in Turbo Pascal and executed on a Pentium PC.

Experiment: Effects of harmonic number and phase

The perceptual phenomenon outlined in the ‘aims’ section of the Introduction is explored parametrically in this experiment using a binary-choice sketch selection task to assess perceived structure.

Procedure

On each trial, a 1f base plaid always had component orientations of $\pm 45^\circ$, component contrast of 20% and was summed with a pair of n-th harmonic components. The harmonic number (n) varied between sessions and was either 3, 5 or 7 (i.e. third-, fifth- and seventh-harmonics).

In order to measure perceived spatial structure, we used the sketch selection paradigm described by Georgeson (1992) and Meese and Freeman (1995). The basic idea is that observers make responses by selecting a sketch (see Fig 1) that best describes the perceived spatial structure of the test plaid. Sited to the left of the 608 oscilloscope and at approximately the same height, was a computer graphics screen, 24 cm wide and 19 cm tall, which was matched approximately in colour and luminance to the stimulus display. A pair of computer generated sketches was displayed on the graphics screen and could be selected by moving a mouse pointer into a box containing the required sketch. The response box into which each sketch was placed was chosen randomly (left or right) on each trial. When the required sketch was selected, the response was recorded by clicking a mouse button. Each sketch was matched in size to the circular field on the 608 display. The same pair of sketches was used for all trials. They were (i) a checkerboard sketch containing vertical and horizontal contours (Fig 1a) and (ii) a diamond sketch containing oblique contours at the same orientations as the stimulus components (Fig 1b). Observers were instructed to select the sketch that best represented the coarse global structure of the test plaid and no specific instructions were given relating to feature classifications such as edges, bars or lines. If perceptual segmentation took place across orientation, then we expected observers to make ‘diamond responses’ (Fig 1b), whereas if stimulus components were perceptually bound across orientation, then we expected observers to make ‘checkerboard responses’ (Fig 1a).

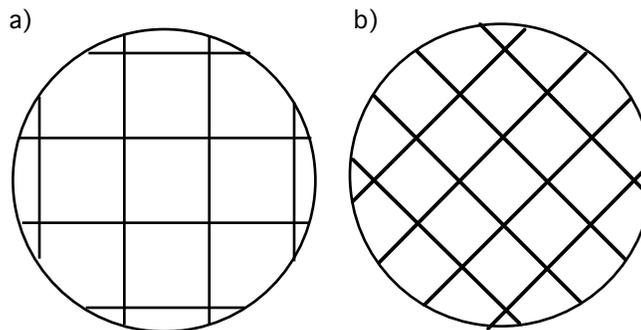


Fig 1. Response sketches used in Experiment 1. Observers inspected briefly presented plaid stimuli and had to select the sketch that best resembled its perceived spatial structure. (a) Compound Response Sketch (Checkerboard): ZCs in the output of a filter sensitive to both 1 c/deg stimulus components. (b) Component Response Sketch (Diamonds): ZCs were found independently for each 1f component and then superimposed.

Within a session (consisting of 180 or 200 trials), the experiment had a 2 x 9 (or 10) x 10 factorial design. There were two phases of n_f (sine-phase $[0^\circ]$ and negative sine-phase $[180^\circ]$); nine or ten contrast levels of n_f , spaced in equal log unit steps, but including one at 0%; and ten replications. Trials were performed in a pseudo-random order and observers performed one or two sessions for each n.

Experiment: Results and Discussion

As the general trends were similar for each observer, results were collapsed across observers. Fig 2 shows mean results from four observers (TSM, MAG, TCAF & AL), where the harmonic number (n) equals 3 (top panel), 5 (middle panel) and 7 (bottom panel). Filled symbols are for a phase of 0° and open symbols are for 180° . Data points show the percentage of ‘checkerboard’ responses (Fig 1a) as a function of n_f component contrast and curves are model fits described in a later section.

When the contrast of nf was zero percent (i.e. the stimulus consisted of the base plaid only), the percentage of checkerboard responses was close to 100% for all conditions, indicating that a checkerboard was seen reliably in this condition. However, as the contrast of the added harmonic was raised, responses increasingly tended away from checkerboards and towards diamonds for each condition (see *Intermediate responses* below). This implies that perceptual binding of components changed from across orientation to within orientation as harmonic contrast was increased. Note that when the third-harmonics were in square-wave phase (phase = 0°) this change in response type occurred at very low harmonic contrasts. This implies a shift in the orientation of *perceived* luminance contours by 45° (Fig 1); quite remarkable given that *physical* mean luminance contours are vertical and horizontal, regardless of the low-contrast harmonics (see Meese, 1999a for an illustration).

Harmonic number

As the harmonic number (n) was increased from 3 to 5, and then to 7, so generally more harmonic contrast was required for responses to change from checkerboard to diamonds. This is shown by the rightward progression of the data, moving down the panels in Fig 2.

Phase

Responses were largely invariant with relative phase for the fifth- and seventh-harmonics (middle and bottom panels), whereas for the third-harmonic ($3f$), relative phase had a large impact on response type. When the $3f$ component was in square-wave-phase (0°), very little $3f$ contrast was required for perceived structure to switch from checkerboards to diamonds. However, when $3f$ was in triangle-wave-phase (180°), about four times more contrast was required before observers changed their response type.

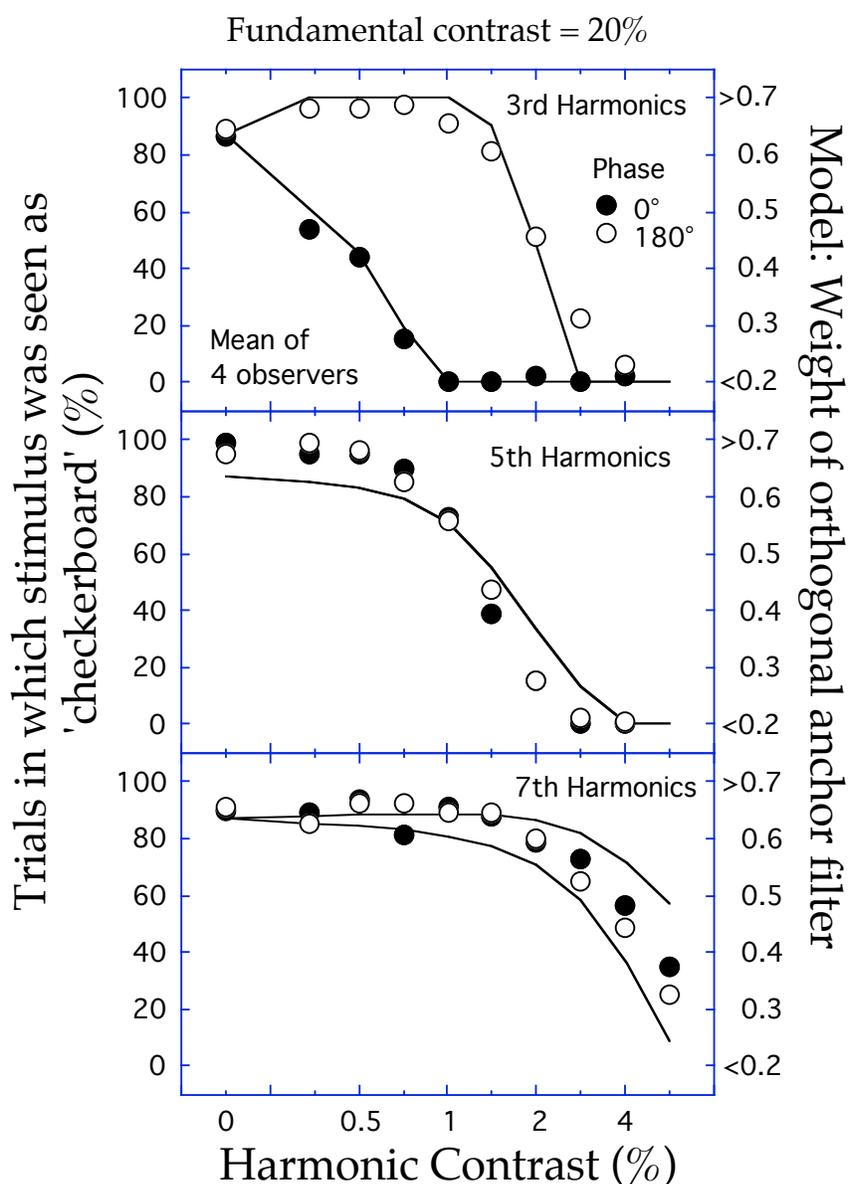


Fig 2: Percentage of ‘checkerboard’ responses (Fig 1a) as a function of the contrast of a pair of harmonics superimposed on a plaid. Means of four observers (two naive), for sine-phase (0° , filled symbols) and negative sine-phase (180° , open symbols). Spatial frequency (nf) of the harmonic components was 3f (top), 5f (middle) and 7f (bottom). Contrast, phase, and spatial frequency of the harmonics can all affect the perceived spatial structure of the test stimulus. The curves are for the model described in the text.

Intermediate responses

As is implied by Fig 2, and as was notable in the data from individual observers (not shown), some stimuli produced average responses split evenly between checkerboards and diamonds. There are at least two likely explanations for these intermediate responses. First, around a critical contrast level perception of the stimulus might be unstable, on some trials appearing to be from one category (e.g. checkerboard-like), while on other trials appearing to be from the other category (e.g. diamond-like). Second, the percept may in some sense fall between the two categories, causing observers to split their responses across trials. Inspection of the data alone cannot decide between these alternatives; both probably contribute to the data. However, we stress that we did not notice any instability of the stimulus (e.g. ‘monocular rivalry’) within a single trial. Presumably, stimulus stability was aided by the short presentation time which minimises interactions between the stimulus and negative retinal afterimages.

Summary of perception experiment

Here we have extended earlier work (Georgeson, 1994) and confirmed that the perception of contour orientation was greatly changed by adding low-contrast third-harmonic components to a standard plaid in square-wave-phase (Georgeson & Meese, 1999). Without the third-harmonics, the perceptually dominant orientation was vertical and horizontal, but with the third-harmonics, the orientations of the plaid’s fundamental components dominated. As elaborated upon in the modelling below, we

interpret this result to mean that the addition of appropriately oriented third-harmonic components can cause spatial filters to segment into two assemblies, each balanced around one of the two orientations of the plaid's components. We have also reported three new findings. First, when the third-harmonic is in triangle-wave phase, the component orientations eventually dominate when the harmonic contrast reaches 4%. Second, although the perceptual impact of the addition of the third-harmonic is phase sensitive, this is not so for the addition of fifth- or seventh-harmonics. Third, as the spatial frequency of the added harmonic is increased, then more harmonic contrast is required before the harmonic affects the perception of stimulus structure. To clarify, we illustrate this, as well as our more general conclusions about spatial filter combination, with an image processing model in the next section.

Model

Binding in the patchwise Fourier transform: An image processing model

In this section we describe an image processing model of spatial vision that emulates human perception of several of our plaid stimuli. Our aim is to demonstrate that the perception of spatial structure in stationary plaid patterns can be understood if appropriate binding and segmentation is performed amongst oriented basis-filters in the patchwise Fourier transform. The algorithm used in the present model is closely related to that presented by Meese (1999a), where nonlinear interactions allowed binding to propagate around a network of interconnected filters. The main difference between the models is that here, binding propagation between filters is replaced by analytic allocations of binding strengths. While the present model does little to address how the algorithm might be implemented in visual cortex beyond the initial stage of spatial filtering, it has the advantages of simplicity and speed. Thus, unlike the model of Meese (1999a), it is possible to perform model calculations for every pixel location in the stimulus. The limitations of the present model with respect to that of Meese (1999a) are discussed in later sections.

Model overview

In order to reveal stimulus features with different orientations at the same or similar locations in the image, Malik and Perona (1992) searched for local contrast-energy maxima in the orientation domain at each position in the image. We extend this idea to identify 'interesting' filter-elements across both orientation *and* spatial scale by performing a search for local maxima amongst contrast-energy filter-elements (Morrone & Burr, 1988) that sample local Fourier space; the filter-elements associated with these energy-peaks are referred to as *anchor-filter-elements*. While subsequent stages in computer-vision algorithms (e.g. Malik and Perona, 1992) typically treat these isolated filter-elements independently, psychophysical phenomena such as the two-dimensional tilt aftereffect (e.g. Meese & Georgeson, 1996a) suggest that assemblies of filter-elements must be involved. Available evidence provides little indication of how many filter-elements make an assembly, but for convenience we consider continuous distributions of discretely tuned filter-elements at each position in the image. Thus, all filter-elements at the same location in the image are bound to each of the anchor-filter-elements by assigning them multiple weights between and including zero and one. We call the filter-element created from the weighted sum of the filters in an assembly, a *synthetic filter-element*. Note, this arrangement means that the output of a single basis filter-element can appear in two (or more) subsequent processing streams (also see appendix A). This is an important feature of our model and is illustrated schematically in Fig 3. The model permits a pair of synthetic filters to be created on a continuum from one extreme, where they are uncorrelated (Fig 3a), to another, where they are fully correlated (Fig 3c). Intermediate points along the continuum represent intermediate levels of correlation (Fig 3b), an example of which is illustrated in the model output shown in Fig 10a.

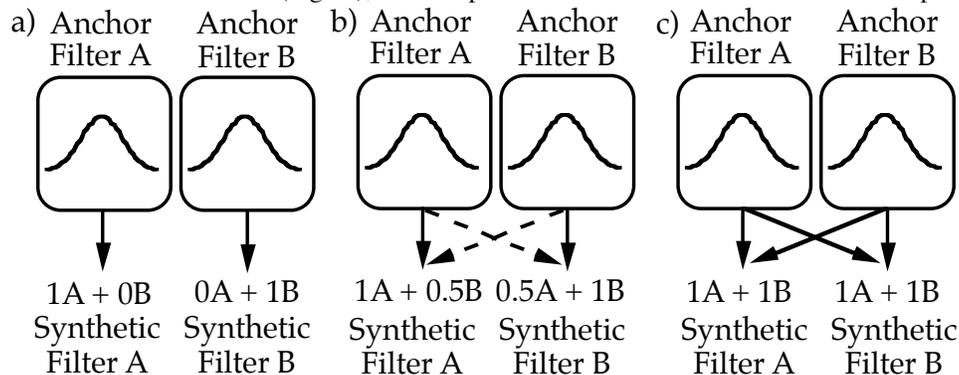


Fig 3. Schematic illustration of the model's continuum for filter combination (binding and segmentation). For simplicity, only a pair of spatial filter elements are shown, both of which have are anchor-filter elements (i.e. they are associated with a local maximum of contrast-energy within the patchwise Fourier transform). For the purpose of this illustration, the two anchor-filter-elements are assumed to be orthogonal, though this is not a necessary feature of the model. a) Segmentation. Each synthetic filter-element receives input from only its associated anchor-filter-element, resulting in no correlation between the two synthetic filters. b) Partial binding. In addition to the input in (a), each synthetic filter-element also receives attenuated input from the fellow anchor-filter-element, resulting in partial correlation between the two synthetic filter-elements. c) Complete binding. Each synthetic filter-element receives equal input from both anchor filter-elements, resulting in complete correlation between the two synthetic filter-elements.

The process by which weights are allocated to filter-elements was reverse engineered from the results of our experiments with plaids. For each anchor-filter-element, a copy of the response of a similar filter-element at a spatial scale three times higher is normalised by the response of the anchor-filter-element. If the normalised response is high, then the activity at the higher spatial scale is probably due to the presence of higher-harmonics in the stimulus. We suppose that this serves as a cue for broad-band oriented image structure, and so binding occurs across spatial-frequency at the expense of orientation. Note that this decision to combine filter-elements across scale avoids the problems of combining feature maps across multiple spatial scales (see Hildreth, 1983; Morrone et al, 1995; Elder & Zucker, 1996a, 1996b). If the normalised-response is low, however, then binding occurs across orientation at the expense of spatial frequency. Presumably, this strategy has more general utility than handling two-component plaids. One possibility is that the scheme helps localise the centres of objects (e.g. blobs and spots), where there is no dominant orientation structure, but much activity in oriented filters (Meese, 1999a).

In sum, our algorithm approximates the following heuristic: for each anchor-filter-element, bind broadly across orientation, unless activity in a third-harmonic filter-element indicates the presence of higher harmonics in the stimulus, in which case, bind across spatial frequency at the expense of orientation.

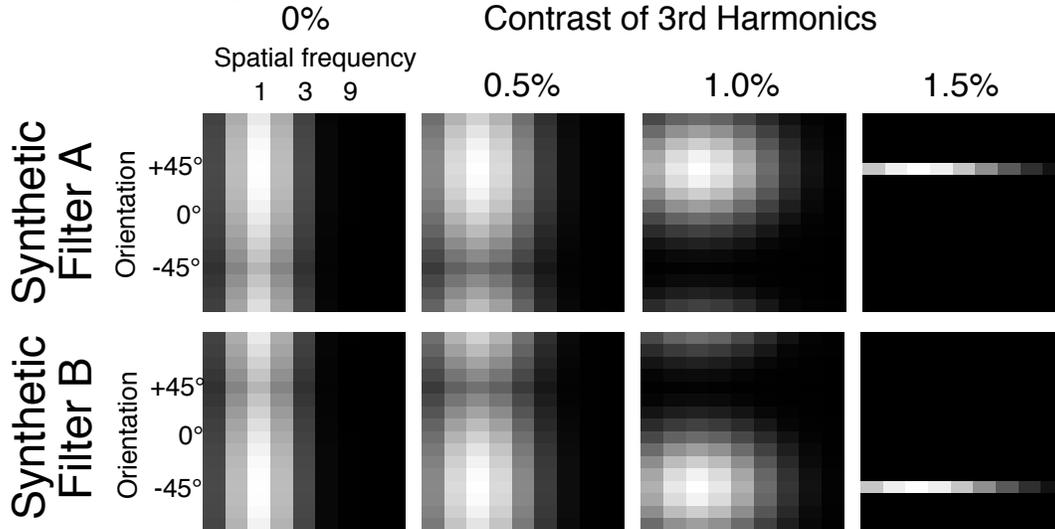


Fig 4. Basis-filter weights for two-component plaids (20% component contrast) with square-wave phase third-harmonics. The two rows are for two different synthetic filters and the four columns are for different contrasts of the third-harmonics. Within each panel, weights are plotted as a function of filter frequency and orientation. The calculations were performed at the zero-crossing of one of the plaid’s fundamental components. Black indicates a weight of zero and white indicates a weight of one. Grey levels are intermediate weights.

In the model reported here we do not address the neuronal implementation of binding, but simply allocate each cosine-phase linear filter-element with a set of weights (0-1) to associate it with each anchor-filter-element. To clarify, Fig 4 illustrates the distribution of weights (grey level) across filter-elements for each of a pair of synthetic filter-elements at a single location in the image. The stimulus was a two-component plaid with third-harmonic contrast of 0%, 0.5%, 1% and 1.5% from left to right in the figure (see figure legend for details). For each anchor-filter-element a *synthetic filter-element* is computed by weighted linear summation of the linear filter-elements (see next section for full details). In order to illustrate the consequences of these processes we create composite images by superimposing the output images for each synthetic filter quantised to two grey-levels. This representation tends to emphasise the location of luminance boundaries (edges), though a more elaborate scheme might be used to reveal additional feature classes such as bars and so forth (e.g. Morrone & Burr, 1988).

Model details

The spatial basis-filters chosen for the model are two-dimensional log Gabor functions defined in the spatial frequency domain by the following equation:

$$\log Gab2D(f, \theta) = \log Gab1D(f, \theta) \times orthFunc(f, \theta),$$

where (f, θ) are polar coordinates in the Fourier plane (spatial frequency [in c/deg] and orientation [in degrees]). The two terms on the right are defined as follows:

$$\log GablD(f, \theta) = \exp \left(\frac{- \left\{ \log_2 \left(\frac{f \cos(\theta - \theta_0)}{f_0} \right) \right\}^2}{2(0.424W)^2} \right),$$

where f_0 and θ_0 are the preferred spatial frequency and orientation of the filter and W is the filter's spatial frequency bandwidth (full-width at half-height, in octaves) and

$$orthFunc(f, \theta) = \exp \left(\frac{- \{ f \sin(\theta - \theta_0) \}^2}{2\eta^2} \right),$$

where

$$\eta = f_0 \sin(H) \sqrt{- \left(2 \ln \left\{ \frac{0.5}{\log 1D(f_0, \theta_0 + H)} \right\} \right)^{-1}},$$

where H is the filter's orientation bandwidth (half-width at half-height, in degrees).

Thus, the filters have modulation transfer functions that are the product of two one-dimensional Gaussian functions. One is defined along a radial log spatial frequency axis and the other is at right angles to this. The positions and dimensions of these filters (in the frequency domain) are defined entirely by four of the parameters above: f_0 , θ_0 , W & H . This class of filters is similar to those used by Field (1987) and Morrone & Burr (1992).

For the filters in the present model the orientation bandwidth (H) was always 20° and the spatial frequency bandwidth (W) was always 1.6 octaves. These filters provide a fair approximation to those observed both psychophysically and in single-cell recordings (e.g. DeValois & DeValois, 1990). For most of the simulations, filters were spaced every 11.25° in the orientation domain and every 0.8 octaves in the spatial frequency domain, though in later simulations these spacing were decreased to 5° and 0.2 octaves respectively. The lowest frequency filter was tuned (nominally) to 0.33 c/deg, and the highest to 28 c/deg.

Model filters with common tuning functions are arranged in triplets comprising a *complex-filter* and a quadrature pair of linear *simple-filters*. Simple-filters have zero-response to mean luminance and have output scaled to give unit response to their preferred sine-wave grating with matched phase and Michelson contrast of 100%. Complex-filters compute local contrast energy by taking the square-root of the sum of the squared outputs of sine-phase and a cosine-phase simple-filters (Morrone & Burr, 1988). Because of the way that these filters are used in the model, a useful intuition is to think of the complex-filters as 'control units', and the cosine-phase simple-filters as 'image-data units'.

The model operates as follows. First, at each position in the image a search for local maxima is performed amongst the full set of complex-filter-elements (this is similar to the algorithm described by Malik & Perona, [1992] and Freeman & Adelson, [1991]). If a filter-element identified this way has a response greater than t (where $t = 0.05$, to provide some protection against noise and spurious low energy features) it is labelled as an *anchor-filter-element*. Triplets of filter-elements are associated (bound) to anchor-filter-elements according to their proximity in the Fourier domain. The binding strength ($w_{a,i}$) of the i th triplet of filter-elements ($1 \leq i \leq N$, where N is the number of triplets of basis-filters in the model) to the a th anchor-filter-element (denoted $complex_\alpha$ where $1 \leq a \leq n$, and n is the number of anchor-filter-elements), at each position in the image (we ignore positional subscripts for clarity) is given by

$$w_{a,i} = \exp \left(\frac{- [\log_2(f_a / f_i)]^2}{2[\lambda.s_\alpha]^2} \right) \times \exp \left(\frac{- [\theta_a - \theta_i]^2}{2[\lambda.\sigma_\alpha]^2} \right). \quad (1)$$

Variables f_a and θ_a are the preferred spatial frequency and preferred orientation of $complex_\alpha$ f_i and θ_i are the preferred spatial frequency and preferred orientation of the i th triplet of filter-elements and $\lambda.s_\alpha$ and $\lambda.\sigma_\alpha$ determine the spread of binding in spatial frequency (in octaves) and orientation (in degrees) respectively. The values of these last three parameters (λ , s_α and σ_α) are image dependent and are determined as follows. First, a complex-filter-element that is selective for the same orientation as

$complex_a$ but tuned for a spatial frequency approximately three times higher (called the *third-harmonic filter-element* and denoted $complex3_a$) is identified⁵ and its response is normalised by the response of the anchor-filter-element to give:

$$norm3_a = resp(complex3_a)/resp(complex_a). \quad (2)$$

The Gaussian spread of binding in the orientation domain (σ_a) is inversely and bi-linearly related to $norm3_a$:

$$\sigma_a = -\mu \cdot norm3_a + \kappa; \text{ IF } \sigma_a < 0 \text{ THEN } \sigma_a = 0, \quad (3)$$

where κ and μ are free parameters of the model. The Gaussian spread of binding in the spatial frequency domain (s_a) is inversely related to that in the orientation domain by:

$$s_a = \gamma(\kappa - \sigma_a)/\kappa \quad (4)$$

where γ is a free parameter of the model. Note that when $\sigma_a = 0$ then $s_a = \gamma$. Thus, γ gives the maximum spread of binding across spatial frequency in octave units. This parameter has no affect on the quantitative fits of the model in this paper and consequently it received little attention. To provide qualitatively reasonable feature-maps, it was fixed at a value of 2. The values of σ_a and s_a are shown in Fig 5 as a function of $norm3_a$ for the parameter values used in the simulations (see Table 1).

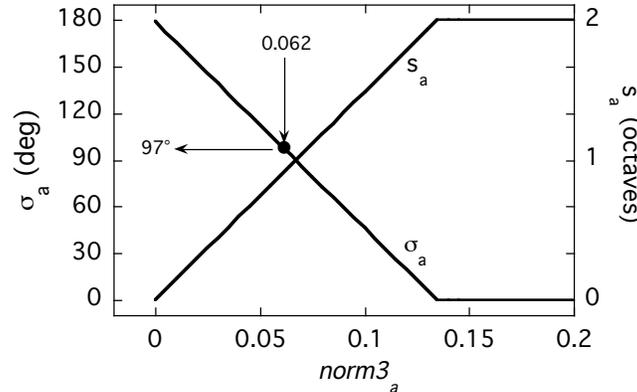


Fig 5. Values of σ_a and s_a as a function of $norm3_a$ for the parameter values used in the simulations (see Table 1). The circular symbol indicates the value of $norm3_a$ (0.062) for a standard plaid, for which it is constant across the whole image and for each of two anchor filters. In the model, σ_a was constrained to have a value of 97° for a standard plaid in order to produce qualitatively acceptable output. See text for details.

The parameter λ in Eq 1 is a nonlinear function of contrast-response, given by the sigmoidal logistic function

$$\lambda = 1/\{1 + \exp([T-resp(complex_a)]/S)\}, \quad (5)$$

where T and S are free parameters and describe the lateral position and spread of the function respectively. The purpose of Eq 5 is to limit the spread of binding when the contrast response of an anchor-filter-element is low (e.g. Meese & Georgeson, 1996b)—see forward to Fig 6. For moderate contrast and above, as in many of the simulations reported here, $\lambda \approx 1$, and has no impact on the behaviour of the model.

Feature map

An output image is created for each anchor-filter as follows. At each position in the image a synthetic filter-element is constructed by calculating a weighted sum of (phase-preserving) cosine-phase simple-filters where the weights are given by

⁵ For the stimuli used here, it was always possible to identify a third-harmonic filter. However, if anchor filters were identified in the highest spatial frequency bands, then there would be no third-harmonic filter. In such cases, a simple rule might be to assume a high response of a nominal third-harmonic filter.

Equation 1.

The output image for each synthetic-filter is quantised by setting all values above zero to white (1) and all those below zero to black (-1). Places in the image where the synthetic filter has no response, or is undefined because its anchor-filter-element is not a local maximum, are set to grey (0). The final feature-map is created by summing the synthetic filter images and normalising to the range -1 to 1. This procedure preserves the light-dark feature boundaries from each of the quantized synthetic-filter images.

Simulation results and parameter fitting

In addition to the filter parameters that were fixed as already described, four further parameters must be fixed to calculate $w_{a,i}$. These are: κ , μ , T & S . First, however, it was necessary to determine how to relate the model output ($w_{a,i}$) to the perceptual judgements about plaids. There are several possibilities, but for a standard plaid (spatial frequency = 1 c/deg and components oriented at $\pm 45^\circ$ with equal contrast), one is very straight forward. For this stimulus there are always two anchor-filter-elements at 1 c/deg with orientations of $\pm 45^\circ$, and for each one, $norm\beta_a = 0.062$ across the entire image. Thus, the spread of binding around orientation (and across spatial frequency) is constant across position and is equal for both anchor filters. This means that for both synthetic filters, the weights of the orthogonal anchor-filter-elements (e.g. in Fig 3, the weight of anchor-filter-element B for synthetic filter-element A) are constant across position. In this paper we describe this particular weight as the *binding strength* (β)⁶ and, as in a related model (Meese 1999a), use it to predict the extent of perceptual binding across orientation (i.e. checkerboard appearance of a plaid). Note that for a standard plaid it follows from Eq 1 that:

$$\beta = \exp\left(\frac{-90^2}{2[\lambda\sigma_a]^2}\right) \quad (6)$$

which means that β depends only upon the product $\lambda\sigma_a$. At moderate contrast, where $\lambda \approx 1$, β depends only upon σ_a (Eq 3).

The quantitative relation between binding strength and the percentage of checkerboard responses ($RESP_{check}$) is given by:

$$\begin{aligned} RESP_{check} &= 100 \cdot (\beta - \beta_{low}) / (\beta_{high} - \beta_{low}); \\ \text{IF } \beta < \beta_{low} &\text{ THEN } RESP_{check} = 0; \\ \text{IF } \beta > \beta_{high} &\text{ THEN } RESP_{check} = 100. \end{aligned} \quad (7)$$

The parameters β_{low} and β_{high} were determined by inspecting the feature maps and choosing values ($\beta_{low} = 0.25$; $\beta_{high} = 0.65$) that were judged as qualitatively suitable for the two behavioural extremes of the observers (see Figs 6 & 9c). For $\beta > \beta_{high}$, the feature map becomes even more checkerboard like (e.g. Fig 10c), whereas for $\beta < \beta_{low}$, the map more closely resembles the overlapping contours for a pair of gratings. We might instead have set $\beta_{low} = 0$ and $\beta_{high} = 1$, effectively obviating the need for these two parameters (Meese, 1999a). While this would have had the advantage of reducing the number of free parameters in the model, it would have compromised the appearance of the model's edge-maps at the behavioural extremes. For example, Meese & Georgeson (1996b) report that in the 'super-squares' effect (see later), plaids can appear even more checkerboard-like than usual. This suggests that observers are willing to classify a plaid as a checkerboard when binding is less than 100% between orthogonal anchor filters. Consequently, in the model we set $\beta_{high} < 1$. Similarly, we suppose that observers will classify plaids as overlapping gratings when a small amount of binding exists between anchor filters. Accordingly, $\beta_{low} > 0$.

Fitting the model to Fig 6

With $\beta = \beta_{high} = 0.65$, the value of σ_a is constrained by Eq 6, leaving only T and S (Eq 5) as free parameters. These were fit by eye ($T = 0.007$; $S = 0.0275$) to previously published data (Meese & Georgeson, 1996b), which are replotted here in Fig 6. The data show the percentage of checkerboard sketch responses (Fig 1a) as a function of component contrast for a plaid with components oriented at $\pm 45^\circ$. Not surprisingly, with two free parameters the model provides a good account of the data.

This contrast response characteristic of the model is of no importance in fitting the model to Fig 2, but is necessary for fitting the model to data presented later in Fig 7.

⁶ In a different implementation of a related model, Meese (1999a) used this term more broadly to refer to the equivalent of $w_{a,i}$ in that paper.

Fitting the model to Fig 2

We now consider the parameters κ & μ (Eq 3) which determine how $norm3_a$ controls the changes in binding across orientation (σ_a). These two parameters represent only one degree of freedom (lets call it ζ) because of the constraint that at moderate contrast and above, $\beta = \beta_{high} = 0.65$ for a standard plaid (see Fig 6). This constraint is illustrated in more detail in Fig 5, which shows how σ_a depends upon $norm3_a$. The constraint is represented by the solid circular symbol and was derived as follows. For a standard plaid stimulus and model filters with spatial frequency bandwidth of 1.6 octaves, $norm3_a = 0.062$ from Eq 2. By rearranging Eq 6 and setting $\beta = 0.65$ (as required above), $\sigma_a = 97^\circ$. These values of $norm3_a$ and σ_a are the coordinates of the symbol plotted in Fig 5. As σ_a must pass through this point, the two parameters, κ & μ , which control the offset and slope of this function, represent only one degree of freedom.

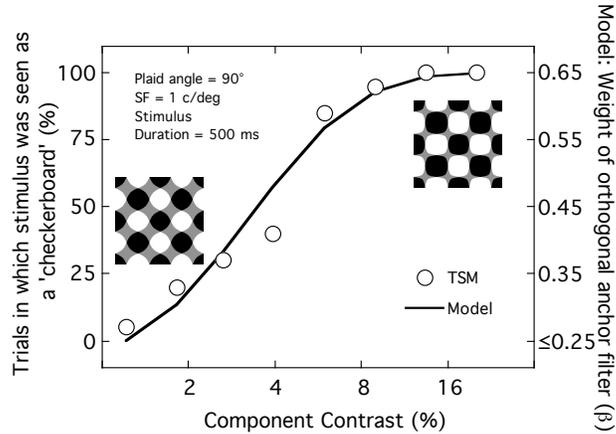


Fig 6. Checkerboard responses (Fig 1a) as a function of component contrast for a two-component plaid. To achieve the fit, two model parameters were set by eye: $T = 0.007$ and $S = 0.0275$. The insets show the feature maps produced by the model for the lowest and highest stimulus contrasts. The structure of these maps was determined from the settings of β_{low} and β_{high} . Data are replotted from Meese & Georgeson (1996b).

For some of the conditions in Fig 2, and at some locations in the image, the number of anchor-filters was greater than two. Nevertheless, the amplitude of these additional anchor-filters was much lower than those at 1 c/deg and $\pm 45^\circ$. Thus, as before, it seemed reasonable to use β as a predictor of behavioural response-type. For the stimulus used in Fig 6, however, β was constant across the entire image and so the issue of which part of the image should contribute to the quantitative analysis did not arise. For most images, including most of the stimuli used here (Fig 2), β is not constant across the image. The simplest solution would be to average β across the entire image. However, this leads to identical predictions for the two different phases of third-harmonic⁷ even though the model's qualitative feature maps are clearly very different (e.g. see forward to Fig 11). Another solution is to average β only over a region that is critical for the behavioural decision in the experiment. The most natural region to choose, and the one used in the model, is that around the ZCs in one of the fundamental components of the stimulus. (Note that, by symmetry, it is necessary to perform the analysis at only one orientation). The spatial extent over which β is averaged either side of the fundamental ZCs is a free parameter in the model (ϵ). We fixed this parameter at $\pm 12.5\%$, meaning that averaging took place over exactly half of the image.

The only remaining degree of freedom, ζ (recall that this represents the two model parameters: κ and μ), was manipulated to fit the entire data set in Fig 2 by eye, and gave $\kappa = 179$ and $\mu = 1330$. This was done with all other parameters fixed as before, except for β_{high} , which was reset to 0.71. This allowed the model to be fit to the same standard plaid in Fig 2 and Fig 6 with the same value of β (0.65). This change in β_{high} across experiments reflects a change in the percentage of 'checkerboard' responses made for the same stimulus in the different experiments—in Fig 6, there are 100% checkerboard

⁷ This happens for the following reason. The value of β depends upon $norm3_a$ (from Eq 3 & 6) and, consequently, the relative values of contrast energy for anchor-filters and third-harmonic-filters (Eq 2). The responses of these energy filters at different positions in the image are determined not only by the amplitude of the 1f and 3f stimulus components (at the same orientation as an anchor filter) but also their phase because the spatial frequency bandwidths of the filters are sufficiently broad for them to be sensitive to both components. However, reversing the phase of the 3f component (as in Fig 2) does not change the profile of the contrast energy function, but merely shifts its phase by half a cycle of the fundamental component. Thus, a process that averages over spatial location, as suggested above, would produce the same values of β for both phases of 3f component. This is clearly at odds with the data in Fig 2.

responses for the standard plaid with component contrast of 20%, whereas in the top panel of Fig 2 there are slightly fewer. This could be due to the use of different response criteria by the observers in the different experiments.

The model captures several aspects of the data extremely well. First, both the model and the data show a general drift to the right as the harmonic number is increased (compare between panels). In the model this happens because as harmonic number is increased, the activity in the third-harmonic filter decreases but the activity of the anchor-filter changes very little. It follows that $norm3_a$ decreases and so β increases. Second, even though the model contains no phase-based heuristic, the model describes the effect of phase for the third-harmonic extremely well. In the model, the effect occurs because of the relative phases of the fundamental and its harmonic component around the zero-crossings of the fundamentals. When the two components are in square-wave phase, they sum constructively, $norm3_a$ is high and so β is low, whereas when they are in triangle-wave phase, they sum destructively, $norm3_a$ is low when the third-harmonic contrast is low, and so β is high. Indeed, the destructive summation means that $norm3_a$ is even lower than it would be for a standard plaid. In the model this results in an increase in checkerboard responses and is consistent with the small effect in the data for the low contrast triangle-wave phase third-harmonics. Note, however, that as the contrast of the third-harmonic is further increased, $norm3_a$ eventually becomes sufficiently high for β to become low and the fundamental components segment.

For the fifth-harmonic the model shows no effect of phase, in good agreement with the data. In the model this occurs because the spatial averaging (ϵ) of β covers a greater range of relative phases between the fundamental and the harmonic component. For the seventh-harmonic the data suggest a very small effect of phase, which is consistent with the model (and in the opposite direction to that for the third-harmonic), though the model exaggerates the effect somewhat.

Fitting the model to Fig 7.

With all model parameters now fixed, and $\beta_{high} = 0.71$, we tested the model on another previously published data set (Meese & Georgeson, 1996b) on spatial aftereffects of adaptation. To simulate the aftereffect of adaptation, a subtractive model was used (Georgeson, 1985; Cannon & Fullenkamp, 1991). The response of each complex filter-element (r_i) to the test pattern is given by $r_i = s_i - c_i/\alpha$, where s_i is the response of the i -th complex filter-element to the test pattern without adaptation, c_i is the response of the same filter-element to the adapter, and α is a free parameter set to 20 by eye. Any filters with negative response were reset to zero.

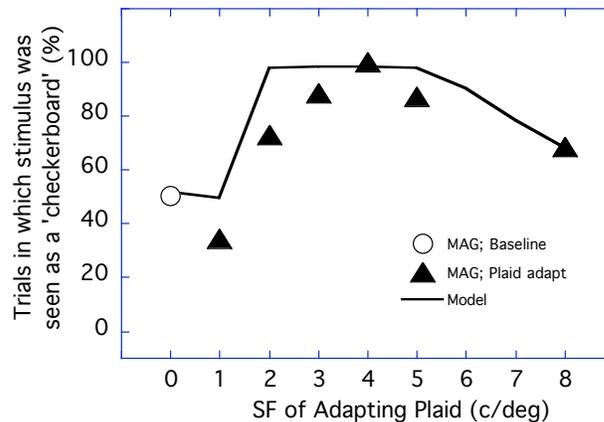


Fig 7. Percentage of checkerboard responses (collapsed over test plaid contrast) as a function of the spatial frequency of an adapting plaid. The adapting plaid component contrasts were 5% and their orientations matched those of the test plaid. At an adapt spatial frequency of around 3 c/deg, both model and data show a large increase in the checkerboard responses. Data are replotted from Meese & Georgeson (1996b).

In Fig 7, data and model are averaged over the contrast of a test plaid (i.e. the average of plots similar to Fig 6) for an experiment in which the observer adapted to a plaid with component contrasts of 5% and orientations matched to those in a test plaid.⁸ Both model and data show an adaptation aftereffect for perception of spatial structure that is tuned to an adapting spatial frequency of around 3 c/deg. Close to this spatial frequency, more checkerboard responses are made than usual. Note that the open symbol is for a control condition where the adapting pattern was of uniform mean luminance. In the model, the adaptation aftereffect occurs because adapting the third-harmonic filter-elements makes them less responsive to the test pattern.

⁸ For reasons that are of no concern here, the stimulus was a standard plaid rotated through 45°—i.e. it had component orientations of 0° and 90°.

This means that $norm\beta_a$ is lower and consequently, β is higher, than is usual for a two-component plaid. This results in more extensive binding around orientation than is usual for this stimulus (also see Fig 10).

For completeness, Fig 8 shows data and model (with no change in parameter values), for when the adapting component orientations were intermediate to those in the test pattern. The model shows little or no aftereffect of adaptation, whereas the data show that fewer checkerboard responses are made when the adapting spatial frequency is similar to that of the test plaid. Although the present model fails badly with the results from this experiment, the data are well described by the model of Meese (1999a). We will return to this limitation of the present model in the General Discussion.

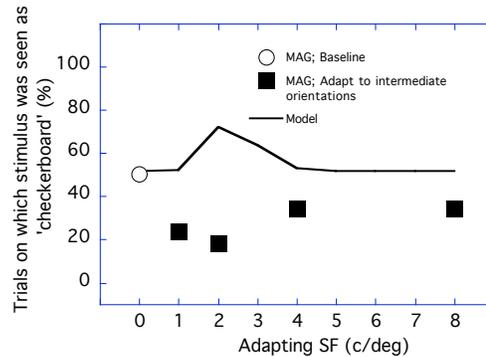


Fig 8. Percentage of checkerboard responses (collapsed over test plaid contrast) as a function of the spatial frequency of an adapting plaid. The adapting plaid component contrasts were 5% and their orientations were intermediate to those in a test plaid. The model does not capture the aftereffect of adaptation seen in the data. See text for further discussion. The data are replotted from Meese & Georgeson (1996b).

Summary of parameters

The model's parameters are summarised in Table 1. Note that in addition to the parameters discussed above, the spatial frequency bandwidth of the filters also turns out to be an important parameter. This is because it influences the level of activity in the third-harmonic filter and the impact of stimulus harmonics on Eq 2. This makes it important in controlling the lateral offset between the model predictions for the two different phases of the third-harmonic in Fig 2. In this respect, the model performs very well using a typical spatial frequency bandwidth of 1.6 octaves.

One final point that is not clear from Table 1, is that the tuning of the model's basis-filters ensured that there were basis-filters matched to the fundamental components of the plaid stimuli. (It is for this reason that the spacing of filters was decreased to 0.2 octaves and 5° towards the end of the next section). We do not characterise this arrangement as a free parameter but will return to this point in the General Discussion.

Parameters	Value	Figures in which the parameter represents a degree of freedom			Comments
		Fig 2	Fig 6	Fig 7	
Filters					
low sf	0.33				
high sf	28				
sf spacing	0.8 & 0.2 octaves				
orientation spacing	11.25° & 5°				
sf bandwidth	1.6 octaves	√		√	
orientation bandwidth	±20°			√	
Algorithm					
κ	179				Sets the maximum extent of binding across orientation
μ	1330				Controls the change from binding across orientation to binding across sf
ζ	n/a	√		√	This is the single degree of freedom represented by the above two parameters
γ	2				Sets the maximum extent of binding across sf
T	0.007		√	√	Threshold parameter for contrast control on binding
S	0.0275		√	√	Slope parameter for contrast control on binding
β_{low}	0.25	√	√	√	Used to scale model predictions to behavioural responses
β_{high}	0.65 & 0.71	√	√	√	Used to scale model predictions to behavioural responses
ε	±12.5%	√			Extent of spatial averaging for quantitative predictions
τ	0.05				Threshold for anchor-filter-elements
α	20			√	Adaptation parameter

Table 1. Summary of model parameters

Feature maps

In this section, we consider the feature maps produced by the model for a wide variety of plaid stimuli, some of which have been modelled quantitatively in the previous section.

Standard plaid

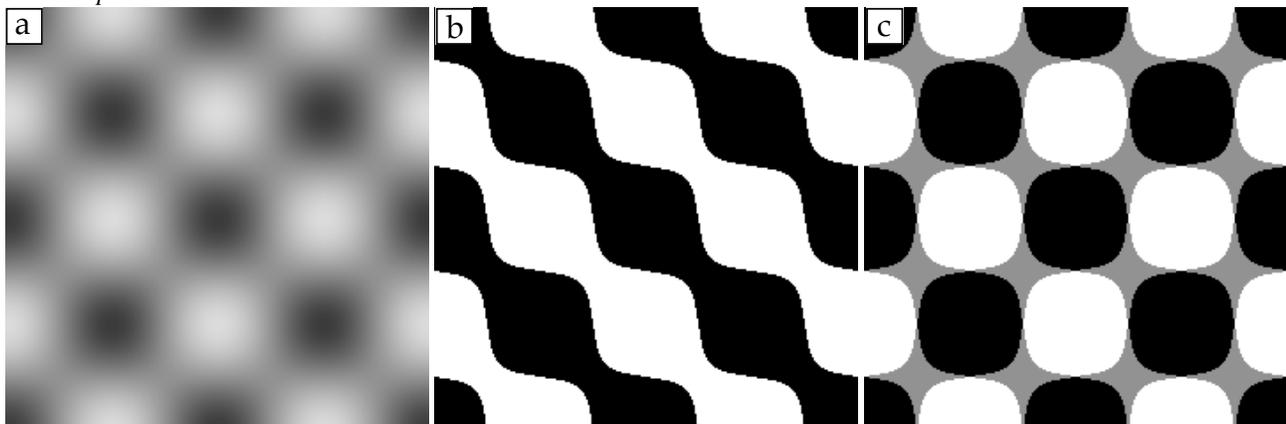


Fig 9. (a) A standard plaid stimulus with component contrast of 20%. (b) Feature map produced for one of the two synthetic filters. The feature map for the other synthetic filter looks like a mirror reversal of (b). (c) Feature map produced by the model in response to (a); this is the normalised sum of the feature maps produced by the two synthetic filters.

Fig 9 is for a standard plaid (component contrast = 20%; nominal component spatial frequency = 1 c/deg) and shows,

the grey-level stimulus (a), the quantised response image for one of the synthetic filters (b) and the normalised response image (the feature map) for the superposition of the model's two quantised synthetic filter outputs (c). This feature map captures the key aspects of the perceived structure of this two-component plaid stimulus extremely well. The overall rectilinear blocking that dominates the perception of the stimulus is evident, but the existence of residual oblique structure is also acknowledged.

Note that this represents a development from the preliminary proposal stated in the Introduction (and in earlier work), where we have suggested that the checkerboard-like appearance of a standard plaid is mediated by a *single* assembly of visual neurons (i.e. a single spatial filter). Fig 9 illustrates that the output of a *pair of highly correlated* spatial filters performs a similar job.

Super-squares

The correlation between information carried by the two synthetic filters in Fig 9c would be increased still further if binding around orientation were increased for each of the anchor-filters. This was achieved in Fig 10c by desensitising the third-harmonic filters according to the simulated aftereffects of adaptation to a 3 c/deg standard plaid (with component contrast of 5%; Fig 10b). For a standard plaid stimulus with component contrast of 4%, the model output is shown in Fig 10a and Fig 10c pre- and post- adaptation respectively. (See above for details of the adaptation simulation). The effect of adaptation was to enhance the checkerboard appearance of the feature-map considerably: Fig 10c looks much more checkerboard-like than the unadapted model output for both moderate and high contrast test plaids (Figs 10a & 9c). This is a good emulation of the 'super-square' effect (Meese & Georgeson, 1996b), modeled quantitatively in Fig 7. Note also that the feature map in Fig 10a appears intermediate to the two response categories (Fig 1), consistent with the behavioural data for a plaid with component contrast of 4% (Fig 6).

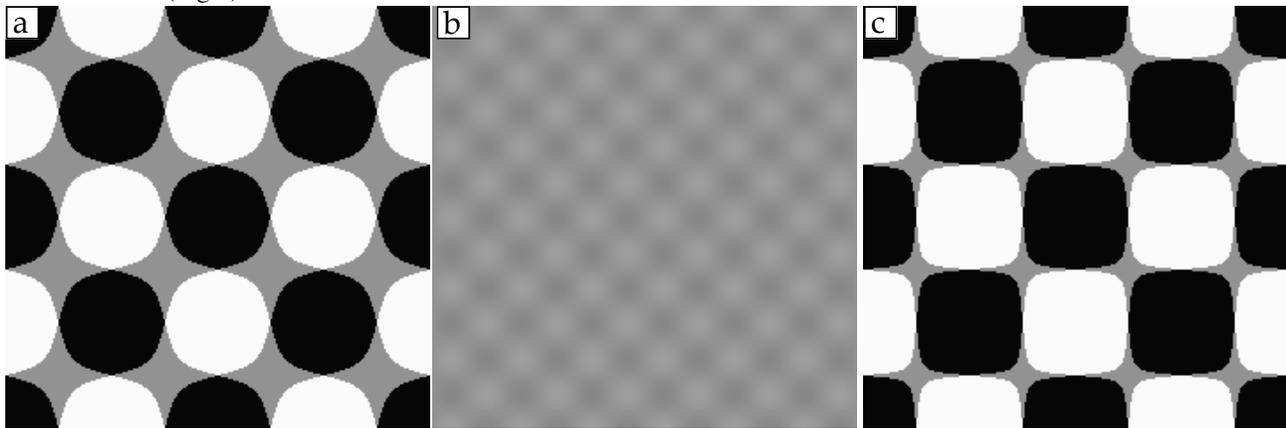


Fig 10. a) Feature map for the standard plaid in Fig 9a, but with component contrast of 4%. (b) Adapting pattern: a standard plaid with spatial frequency three times higher than the test plaid (Fig 9a) and component contrasts of 5%. (c) Feature map for the same stimulus as in (a) but after simulated adaptation to (b). This is the 'super-squares' effect. Note that the rectilinear structure is even more pronounced than it is for the higher contrast test plaid in Fig 9c.

Third-harmonics

The correlation between the information carried by the two synthetic filters can be reduced by lessening the extent to which binding takes place across orientation. This was achieved in Fig 11a where the standard plaid from Fig 9a is summed with a pair of 1% contrast, third-harmonic components in square-wave-phase. Inspection of this stimulus should confirm the results of Fig 2 (top panel), where observers indicated that this type of plaid looked like a pair of overlapping left- and right-oblique contours. This is in good agreement with the model's feature map shown in Fig 11c. In particular, notice the slight bulging of the light and dark diamond shapes in Fig 11a compared to the smaller grey ones. This perceptual effect is mimicked in the feature map and arises because the two synthetic filters are not completely independent: combination of oriented basis-filters results in some residual cross-talk between the synthetic filters (see Fig 4)⁹. In Fig 11b, the phase of the third-harmonic is

⁹ Small shifts in perceived feature locations have been explained in terms of a pointwise compressive luminance nonlinearity at an early stage of the visual system. Such nonlinearities cause features to be seen shifted towards the dark side of the boundary (Georgeson & Freeman, 1997). However, the situation here is very different. In Fig 11a, the 'black' and the 'white' diamonds appear larger than the intermediate grey diamonds. In fact, the perceived feature locations are shifted away from light and dark sides, to the grey side. This cannot be explained by a compressive luminance nonlinearity.

changed to triangle-wave phase, and the output is shown in Fig 11d. In this case the model no longer segments the stimulus across orientation, at least, not at the critical regions of the stimulus close to the fundamental ZCs. It is not until the contrast of the third-harmonic is substantially increased that this happens for both the feature map (not shown) and the quantitative fit to the data (Fig 2).

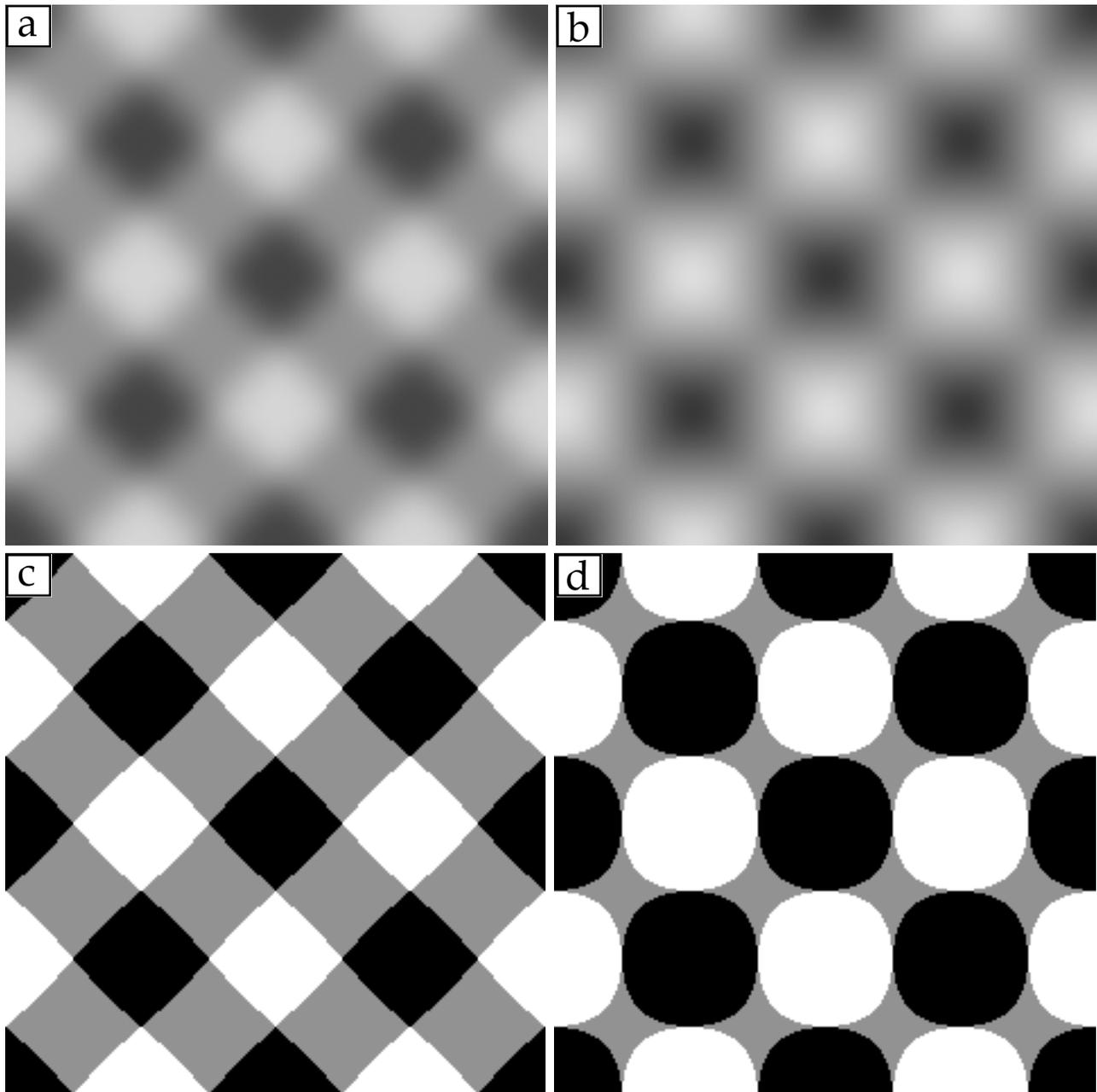


Fig 11. (a) A standard plaid plus low contrast third-harmonic components in square-wave phase (phase = 0°). (b) The same as (a) but with harmonics in triangle-wave phase (phase = 180°). (c & d) Model feature-maps for (a) and (b) respectively.

Three-component plaids

The model's success is certainly not limited to the plaid patterns for which quantitative modeling has been performed. Fig 12 shows the model's feature maps for a collection of three-component plaids and further exemplifies that the structure seen in plaids does not depend upon the component orientations. For example, in Fig 12a the component orientations are 0° and $\pm 60^\circ$, though the stimulus looks like a set of almost circular blobs (Georgeson, 1992, 1998). Changing the phase from 90° to 0° (Fig 12b), reveals contours at the component orientations, though in both cases, the model produces an appropriate feature map by binding broadly around orientation for each of three anchor filters (Figs 12d,e).

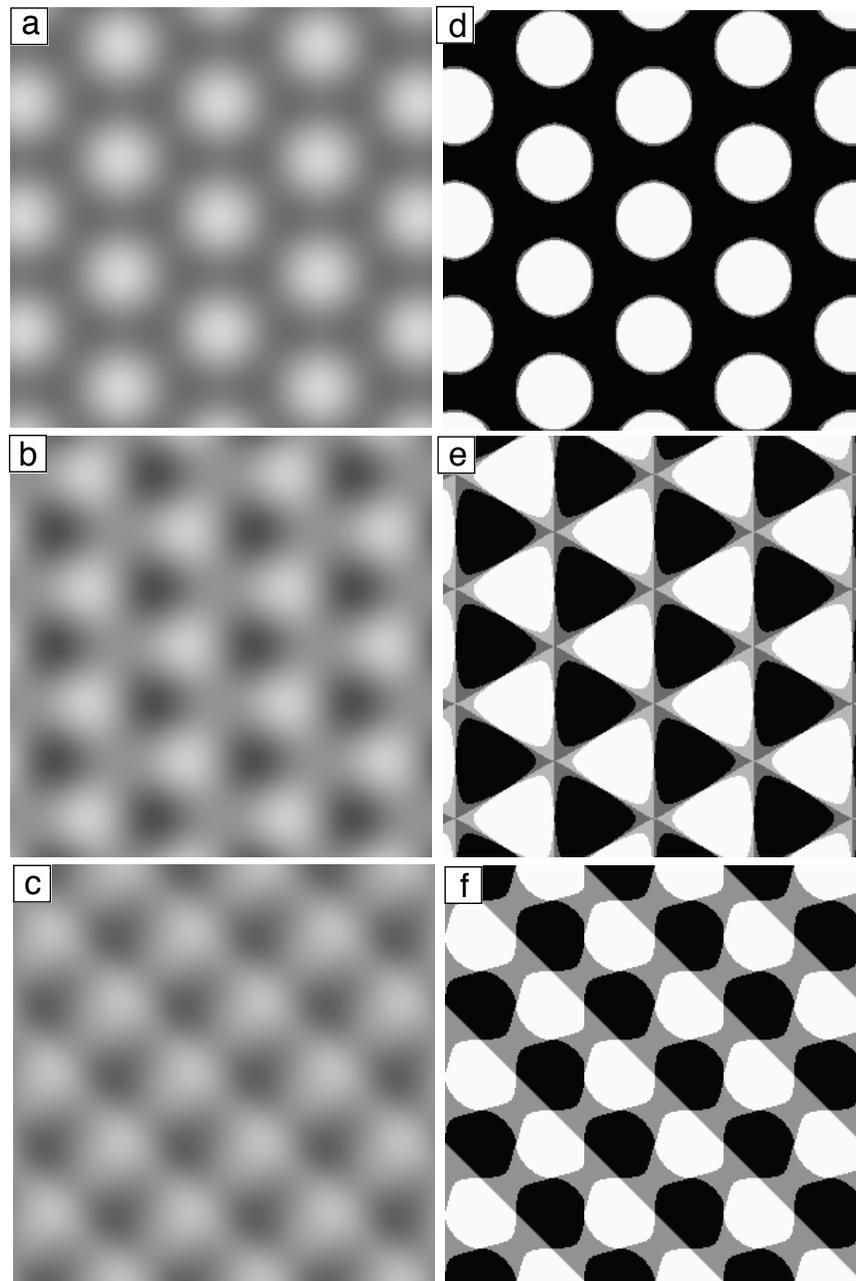


Fig 12. Three-component plaids. (a) Three components with common spatial frequency and contrast oriented at 0° and $\pm 60^\circ$ with phase = 90° . (b) The same as (a) except that phase = 0° . (c) Two component standard plaid with component contrast of 20% plus left oblique second-harmonic component (phase = 0°) with component contrast of 4%. (d,e,f). Feature maps for (a,b,c) respectively.

In Fig 12c, a second-harmonic component is added to the left oblique component of a standard plaid. In this case, the model segments orientations at the stimulus regions where the second-harmonic is in phase with the fundamental, producing the fairly straight left oblique contours in Fig 12f. However, half a cycle of the fundamental further on, the fundamental and second-harmonic are in anti-phase. This causes the model to bind filters across orientation and produces the wavy left oblique contours in Fig 12e. In general, the model produces a good structural description of all of the stimuli in Fig 12 as well as many other plaid stimuli that we have tested.

Two-component plaids

A particularly striking feature of the model is its ability to change gracefully between binding and segmentation of stimulus components. This is shown in figures 13 & 14 where a set of two-component plaids investigated by Georgeson (1998) are shown. In both figures, the spatial frequency ratio of the two components increases up the figure, while the orientation difference increases across the figure. Broadly speaking, Georgeson (1998) found that the perception of the plaids in Fig 13 changes from compound structures at the bottom of the figure to overlapping structures at the top of the figure. The character

of this result is very well captured by the model as shown by the feature maps in Fig 14.

While good feature maps were produced for the implementation of the model described above (Meese, 1999c), we found that some of the feature maps could be improved slightly by arranging that basis-filters were matched in preferred spatial frequency and orientation to each of the stimulus components. For Fig 14 we approximated this by placing basis-filters every 5° in the orientation domain and every 0.2 octaves in the spatial frequency domain. This considerable increase in the number of model filters introduced further over-sampling of the image by the basis-filters but is not inconsistent with neurophysiological evidence (e.g. DeValois & DeValois, 1990). The main improvement in the feature maps was the occurrence of bulging in the right oblique contours in the second row of Fig 14. Without the additional filters these were straight, and inconsistent with the structure seen in Fig 13 (Meese, 1999c).

These changes to the model have no impact on any of the earlier quantitative analyses.

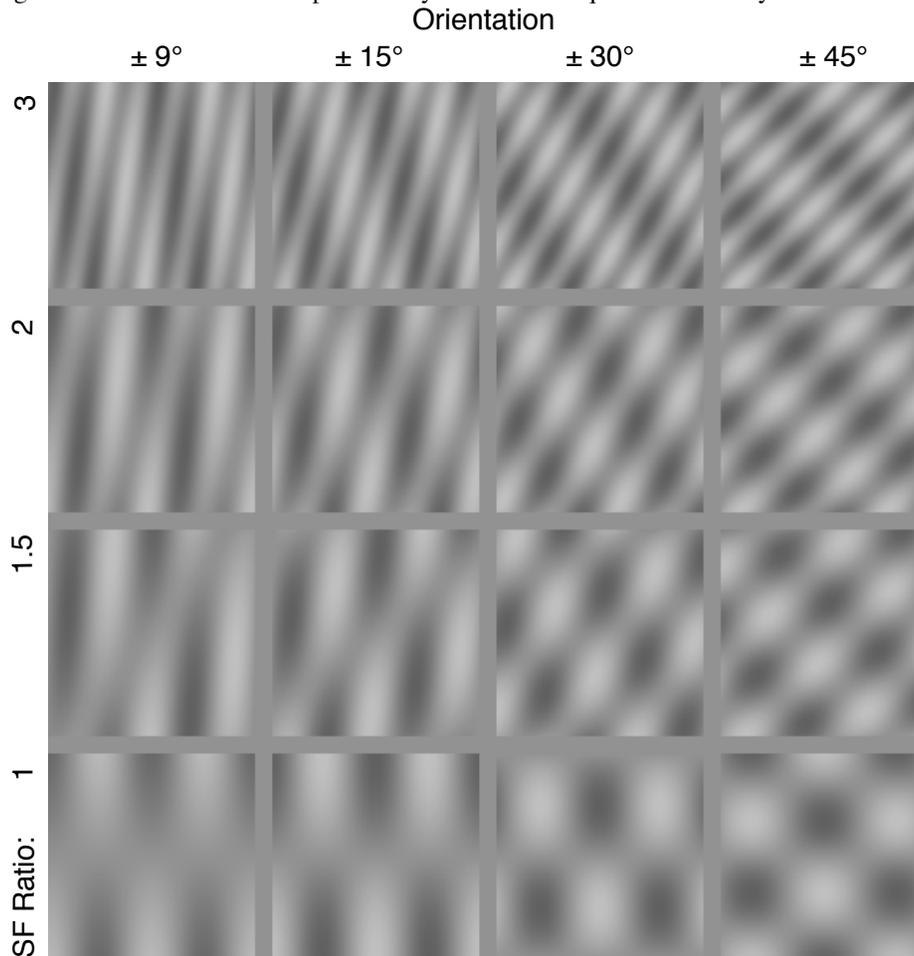


Fig 13. Two-component plaids used in the experiments performed by Georgeson (1998). Spatial frequency ratio increases up the figure and orientation difference increases across the figure.

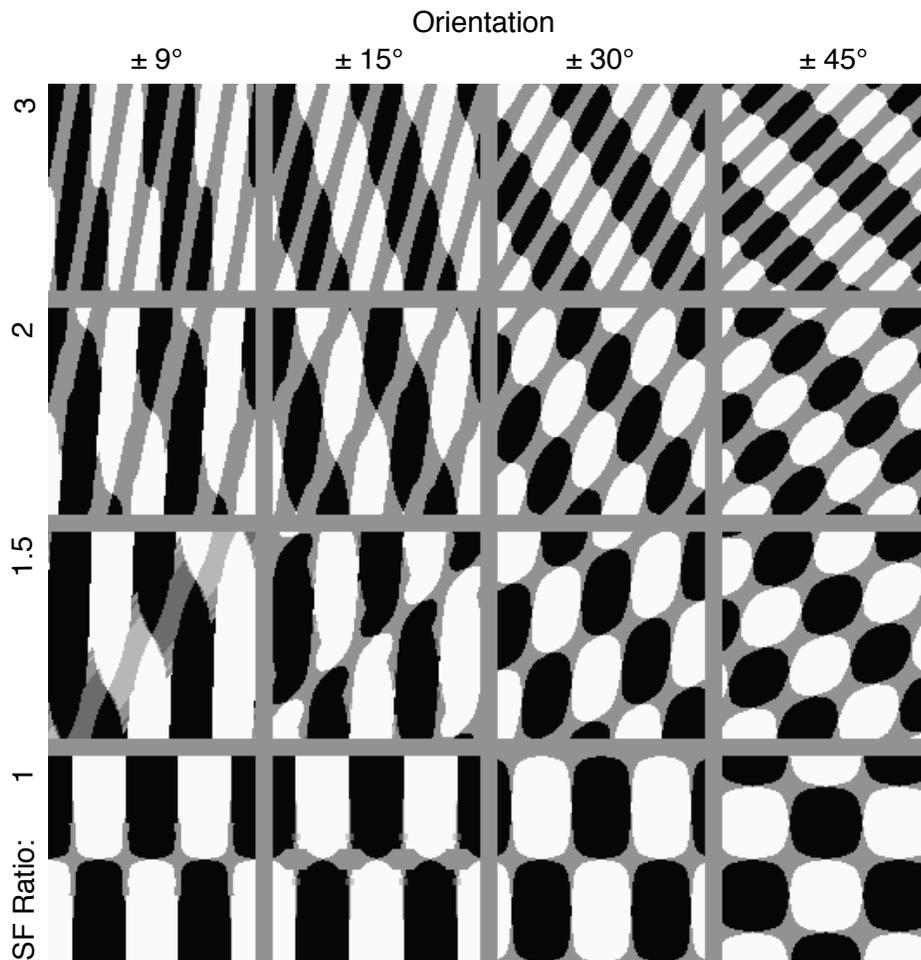


Fig 14. Feature maps for the stimuli in fig 13.

General Discussion

Limitations of the model

We have presented an image processing model in the spirit of an attempt at demonstrating how spatial vision might lawfully execute the binding and segmentation of spatial filters in a stimulus dependent way. For reasons of expediency, we have illustrated the results of this process using a method similar to plotting the zero-crossings in the weighted linear sum of bound filter-elements at each position in the image. While this procedure may be an oversimplification, we are nevertheless struck by how well it describes the general structure of stationary plaids.

The model's algorithm was guided by our experimental work (see above), though the particular implementation that we have used here has known limitations. For example, as illustrated in Fig 8, the model does not describe the aftereffect of adapting to component orientations that are intermediate to those in a test plaid (Meese & Georgeson, 1996a, 1996b). In this case, the data suggest that binding across orientation also depends upon the activities of intermediate oriented filters: an effect that we have previously described in terms of the *bridge hypothesis* (Meese & Georgeson, 1996b; Georgeson & Meese, 1997). We might have built this requirement into the present model but were reluctant to do so for the cost of additional parameters and complexity. A second weakness is that while the model does a good job of segmenting standard plaids when harmonics are added at both orientations (e.g. Fig 11), it fares less well when a square-wave phase third-harmonic is added at only one orientation (not shown). In this case, the model correctly segments the anchor-filter whose orientation is the same as the harmonic's, but the orthogonal anchor-filter is bound across orientation and produces a synthetic filter map similar to that in Fig 9b. This is inconsistent with the perception of this stimulus, which looks like overlapping segmented gratings (Georgeson & Meese, 1999).

We suspect that these model weaknesses are to do with implementation details. Unlike the model above, the brain cannot use a high-level programming language to implement Equation 1. Instead, the control of binding and segmentation is presumably governed by some kind of network of interconnections between neural units selective for different spatial frequencies and orientations. A network of inter-unit links that overcomes both of the limitations described above has been developed and presented by Meese (1999a, 1999b).

Finally, because of the functional role of the anchor- and third-harmonic filters there is an implicit assumption that the model contains basis-filters with preferred orientations and spatial frequencies that are similar to each of the components in

the stimulus. Thus, in general, the model is not scale invariant. This means that if the stimuli were scaled by any factor, other than that of the frequency spacing of the filters, the model predictions would change from those shown in Fig 2. However, we see this as an implementation issue and not a critical failure of the model. In the model, the problem might be lessened by either increasing the number of basis-filters (see the previous subsection), or overcome by using ‘deformable’ basis-filters (e.g. Manduchi, Perona & Shy, 1997). On the other hand, it remains to be seen how well our restricted set of basis-filters performs on more naturalistic stimuli. Such stimuli are less ‘peaked’ in the Fourier domain (Field, 1987) and so perhaps less sensitive to the precise selectivities of basis-filters.

No doubt, some of our simplifying assumptions, such as the shape of the functions in Fig 5 and the use of Gaussian functions in Eq 1, may also ultimately require revision.

Comparisons with other edge-detection models

Like Marr & Hildreth’s (1980) edge-detector, our own model marks edges at the zero-crossings in the output of even-symmetric filters (Georgeson, 1992), though this does not preclude the use of the centroid of filter responses, as in the MIRAGE algorithm of Watt & Morgan (1985). Similar to MIRAGE, filters are summed across spatial frequency and so the problem of combining feature maps across spatial scale (Hildreth, 1983) is avoided. However, unlike in MIRAGE, the combination process is flexible. Furthermore, to achieve good feature maps for plaids we have not found it necessary to include the rectification stage of MIRAGE. For many plaids, the circular filtering used by Marr & Hildreth (1982) and in MIRAGE (Watt, 1988) causes features to be marked incorrectly (e.g. Meese, 1999a; Georgeson & Meese, 1999). For these reasons, and those discussed in Georgeson & Meese (1999), oriented filters and their flexible combination appear to be necessary features of ZC-based (or equivalent) models of edge-detection in plaids.

In a similar way to the ‘energy models’ of Morrone & Burr (1988), Malik and Perrona (1992) and Freeman & Adelson (1991), our own model computes contrast energy at each position in the image. However, these other energy models do not see appropriate features in sine-wave gratings, compound ($f+3f$) gratings (e.g. Georgeson & Freeman, 1997) and (some) plaids. Including luminance nonlinearities before filtering can improve these energy-models for some stimuli, but causes them to fail for different stimuli that produce sine-wave profiles *after* the luminance nonlinearity (Meese, 1999a). Furthermore, it is not that there is something ‘improper’ about the ‘blurred’ luminance boundaries in gratings and plaids that causes their features to be invisible to energy models. This is because a feature is seen by these models at the centre of a single half-cycle of a grating that is flanked by luminance plateaus. Thus, while the computation of contrast energy is an important feature of our model, unlike the energy models above, we also use phase-preserving linear filters, which correctly mark features in plaids and 1D stimuli (Georgeson & Freeman, 1997; Barbieri & Georgeson, 1998).

In a similar way to the edge-detection algorithms of Elder & Zucker (1996) and Morrone, Navangione & Burr (1995), our own model derives synthetic filter characteristics from local image properties. However, neither of the above models, nor the Canny (1986) edge detector, can represent more than one contour (i.e. overlapping contours) at each image location. The ability to do this is an important feature of both our model and those of Malik & Perona (1992) and Freeman & Adelson (1991).

Finally, in contrast to the models above, our model uses populations of filter-elements (model visual neurons) that are bound and segmented in a manner consistent with population coding. This neural coding strategy is useful for resolving response ambiguity and is thought to underlie several adaptation aftereffects (e.g. Coltheart, 1971; Braddick et al, 1978; Mather & Harris, 1998; Meese & Georgeson, 1996a).

Spatial frequency

The experiments reported here were performed with a fundamental frequency of 1 c/deg. Elsewhere (Meese, 1993; Meese & Freeman, 1995; Georgeson & Meese, 1999) we have performed experiments over a greater range of frequencies (0.6 c/deg to 8 c/deg). We have found that at moderate to high spatial frequencies (≥ 4 c/deg), perceptual judgments of plaids become more difficult. For example, a global impression of pattern texture tends to dominate over the fine local pattern structure. This observation is difficult to quantify given the subjective nature of our procedures but suggests that further work is required to determine whether high and low spatial frequency filters make different contributions to perception.

Multiple spatial channels

Our general proposal includes a set of spatially tuned basis-filters consistent with those implied by adaptation and masking studies (e.g. Blakemore and Campbell, 1969; Wilson et al, 1983; Phillips & Wilson, 1984). Subthreshold summation studies suggest that components in the frequency ratio 1:3 (Graham & Nachmias, 1971) and those 90° apart (Georgeson & Shackleton, 1994; Meese & Williams, 2000) are detected independently. These widely accepted features of early spatial vision are entirely consistent with our model. In our plaid experiments, the low-frequency components were always well above threshold in conditions where perceptual binding occurred across spatial frequency. In fact, interactions across spatial frequencies have also been implied by adaptation and simultaneous masking studies when one of the components is above threshold (Klein & Stromeyer 1980; Tolhurst & Barfield, 1978; Hess & Pointer, 1987). In experiments reported elsewhere (e.g. Meese & Georgeson, 1996b) we have also shown that perceptual binding across orientation occurs only when both components are above threshold. Only at and around threshold do pairs of oriented components appear to be treated independently (e.g. Phillips & Wilson, 1984; Georgeson & Shackleton, 1994; Meese & Williams, 2000). Two- and three-component discrimination and uncertainty experiments also indicate a lack of independence above threshold (Olzak & Thomas, 1991, 1992; Thomas & Olzak, 1996; Olzak & Wickens, 1997; Meese, 1995b). Thomas & Olzak (1996) conclude from their

results that "...judgments about spatial frequency and orientation are mediated by mechanisms that combine signals across orientations and across spatial frequency bands, respectively." Wilson & Wilkinson (1997) concurred with this, concluding that human vision probably operates somewhere between the extremes of (i) the parallel, independent multiple channels model of the 1970's and (ii) the hard-wired MIRAGE model of Watt & Morgan (1985; also see Morgan & Watt, 1997). A flexible scheme like the one that we propose would seem to possess the necessary ingredients.

Performance measures and perception

While performance measures (see above) provide objective support for our conclusions based on subjective pattern matching, they are no substitute in investigations of *perception*. For example, forced-choice discrimination experiments may reveal constraints on the location of the limiting source of noise (e.g. after the combination of spatial filters) for a given task, but cannot imply anything about the perceptual qualities of the stimulus. Furthermore, some evidence suggests that certain visual discriminations may be performed successfully in the absence of perceptual awareness (Milner & Goodale, 1995; Kolb & Braun, 1995; though see Morgan, Mason & Solomon, 1997 for a failure to replicate). Indeed, if human vision contains parallel streams for action and perception, as suggested by Milner & Goodale (1995), then standard forced-choice experiments are agnostic as to which of these pathways mediates performance (Georgeson, 1997b).

A further cautionary note about performance data emerges from some of the specifics of the present model. As mentioned in the 'model details' section, our model consists of (at least) two types of units: cosine-phase, linear 'data filters' (like simple cells), and nonlinear, phase-insensitive 'control filters' (like complex cells). In essence, the control filters interact (e.g. to calculate the normalised response of the third-harmonic filter), and determine how the 'data filters' should be routed (i.e. bound and segmented) to later processing streams. Presumably, performance based experiments reflect *any* change in the nervous-system that is available to the decision making process that directs the behavioural response. Consequently, it is unclear that such experiments can distinguish between the two functionally very different streams of processing (data and control) that are proposed in the present model, or, for that matter, any other information streams that the visual system might possess. In this respect, performance experiments are a blunt instrument and while they may reveal the properties and characteristics of visual mechanisms it is unclear what information-processing role is being performed by these mechanisms for visual perception. On the other hand, the explicit perceptual requirements of our plaid experiments ensure that our own data reflect the properties of brain processes relating to conscious visual perception: data obtained on the way things look pertain to the processes that mediate the way things look.

Summary

The model in this paper performs binding and segmentation of spatial filter-elements in the patchwise Fourier transform such that they are: (i) bound into highly correlated assemblies, (ii) segmented into weakly correlated or uncorrelated assemblies or (iii) only partially segmented, with some filter-elements contributing to each of a pair of assemblies. From these experiments and others (Georgeson, 1992, 1997a; Meese & Freeman, 1995; Georgeson & Meese, 1997; Meese & Georgeson, 1996b) has emerged the view that, at least at low spatial frequencies, an important role of first-order spatial filters in human vision is to segment information in different regions of the local Fourier domain as a precursor to representing spatial structure. In the model that we propose, this process is guided by comparing the activity within common orientation bands and nearby spatial frequency bands with a ratio of 1:3.

Acknowledgements

Some of these results were first presented in abstract form by Meese (1999b, 1999c). The work was supported by a University of Bristol scholarship, Wellcome Trust pilot study grant 048076/Z/96/Z and EPSRC project grant GR/M10854 to TSM, and BBSRC project grant S06946 to MAG & TSM.

We would like to thank Brian Roberts for alerting us to the work of Albert Bregman.

REFERENCES

- Barbieri, G. S. & Georgeson, M. A. (1998) Perceived location of bars and edges in broad-band 1-D images: experiment and computational model. *Perception*, 27 supp, 71.
- Blakemore, C. & Campbell, F. W. (1969) On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology*, 203, 237-260.
- Blakemore, C. & Sutton, P. (1969) Size adaptation: A new aftereffect. *Science*, 166, 245-247.
- Braddick, O., Campbell, F. W. & Atkinson, J. (1978) Channels in Vision: Basic Aspects. in Handbook of Sensory Physiology: Perception. Held, R., Leibowitz, H. W. & Teuber, H. (Eds), Springer-Verlag, Berlin.
- Bregman, A. S. (1987) The Meaning of Duplex Perception: Sounds of Transparent Objects. In The Psychophysics of Speech Perception. Schouten, M. E. H. (Ed), Nijhoff, Dordrecht, The Netherlands.
- Cannon, M. W. & Fullenkamp, S. C. (1991). A transducer model for contrast perception. *Vision Research*, 31, 983-998.
- Canny, J. (1986) A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8, 679-697
- Coltheart, M. (1971) Visual feature-analyzers and aftereffects of tilt and curvature. *Psychological Review*, 78, 114-121.
- Derrington, A. M. & Henning, D. R. (1989) Some observations on the masking effects of two-dimensional stimuli. *Vision Research*, 29, 241-246.
- DeValois, R. L., & DeValois, K. K. (1990) Spatial Vision, Oxford.
- Elder, J. H. & Zucker, S. W. (1996a) Local scale control for edge detection and blur estimation. In Lecture Notes in Computer Science, B. Buxton & R. Cipolla (Eds.), 1064, 57-69.
- Elder, J. H. & Zucker, S. W. (1996b) Scale space localization, blur, and contour-based image coding. CVPR, June 1996, San Francisco.
- Engel, A. K., Koenig, P. & Singer, W. (1991b) Direct physiological evidence for scene segmentation by temporal coding. *Proc. Nat. Acad. Sci. USA.*, 88, 9136-9140.
- 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, 173-193.
- 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 A*, 4, 2379-2394.
- Freeman, W. T. & Adelson, E. H. (1991) The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, 891-906.
- Friedman-Hill, S. R., Robertson, L. C. & Treisman, A. (1995) Parietal contributions to visual feature binding: Evidence from a patient with bilateral lesions. *Science*, 299, 853-855.
- Georgeson, M. A. (1984) Eye movements, afterimages and monocular rivalry. *Vision Research*, 10, 1311-1319.
- Georgeson, M. A. (1985) The effect of spatial adaptation on perceived contrast. *Spatial Vision*, 1, 103-112.
- Georgeson, M. A. (1990) Human vision combines oriented filters to compute edges. *Perception*, 19, 354.
- Georgeson, M. A. (1992) Human vision combines oriented filters to compute edges. *Proc. R. Soc. Lond. B*, 249, 235-245.
- Georgeson, M. A. (1994) From filters to features: location, orientation, contrast and blur. In G. R. Bock & J. A. Goode (eds.) Higher order processing in the visual system (Ciba Foundation Symposium 184), pp 147-165. Chichester: Wiley.
- Georgeson M. A. (1998) Edge-finding in human vision: a multi-stage model based on the perceived structure of plaids. *Image & Vision Computing*, 16, 389-405.

- Georgeson, M. A. (1997) Vision & Action: You ain't seen nothin' yet. *Perception*, 26, 1-6
- Georgeson, M. A. & Freeman, T. C. A. (1997) Perceived location of bars and edges in one-dimensional images: Computational models and human vision. *Vision Research*, 37, 127-142.
- Georgeson, M. A. & Meese, T. S. (1996) Perceived structure of plaids implies variable combination of oriented filters in edge-finding, in: *Human Vision and Electronic Imaging, Proc SPIE*, B. E. Rogowitz and J. P. Allbech (Eds) **2657**, 175-189.
- Georgeson, M. A. & Meese, T. S. (1997) Perception of stationary plaids: the role of spatial filters in edge analysis. *Vision Research*, 37, 3255-3271.
- Georgeson, M. A. & Meese, T. S. (1999) Adaptive filtering in spatial vision: Evidence from feature marking in plaids. *Perception*, 28, 687-702.
- Georgeson, M. A. & Shackleton, T. M. (1994) Perceived contrast of gratings and plaids: non-linear summation across oriented filters. *Vision Research*, 34, 1061-1075.
- Georgeson, M. A. & Turner, R. S. E. (1985) Afterimages of sinusoidal, square-wave and compound gratings. *Vision Research*, 25, 1709-1720.
- Graham, N. & Nachmias, J. (1971) Detection of grating patterns containing two spatial frequencies: A comparison of single-channel and multiple-channels models. *Vision Research*, 11, 251-259.
- Gray, C. M. (1999) The temporal correlation hypothesis of visual feature integration: still alive and well. *Neuron*, 24, 31-47.
- Hess, R. F. & Dakin, (1997) Absence of contour linking in peripheral vision. *Nature*, 390, 602-604.
- Hess, R. F. Field, D. J. & Watt, R. J. (1990) The puzzle of amblyopia. In *Vision: Coding and Efficiency*. Blakemore, C. (Ed), CUP, 267-280.
- Hess, R. F. & Pointer, J. S. (1987) Evidence for spatially local computations underlying discrimination of periodic patterns in fovea and periphery. *Vision Research*, 27, 1343-1360.
- Hildreth, E. (1983) The detection of intensity changes by computer and biological vision systems. *Computer Vision Graphics and Image Processing*, 22, 1-27.
- Howard, I. (2002) *Seeing in Depth: Basic Mechanisms*, I Porteus, Toronto.
- Klein, S. & Stromeyer, C. F. (1980) On inhibition between spatial frequency channels: Adaptation to complex gratings. *Vision Research*, 20, 459-466.
- Kolb, F. C. & Braun, J. (1995) Blindsight in normal observers. *Nature*, 377, 336-338.
- Hubel, D. H. & Wiesel, T. N. (1974), *Journal of Comparative Neurology*, 158, 267-294
- Malik, J. & Perona, P. (1992) Finding boundaries in images. in H. Wechsler (Ed.) *Neural Networks for Perception Vol 1: Human & Machine Perception*. Academic Press: London.
- von der Malsburg, C. & Scheider, W. (1986) A neural cocktail-party processor, *Biological Cybernetics*, 54, 29-40.
- Marr, D. (1982) *Vision*. Freeman: San Francisco.
- Marr, D. & Hildreth, E. (1980) Theory of edge detection. *Proc. R. Soc. Lond. B.*, 207, 187-217.
- Mather, G. & Harris, J. (1998) 'Theoretical models of the motion aftereffect' in *The Motion Aftereffect: A Modern Perspective*. Mather, G., Verstraten, F. & Anstis, S. (Eds). pp157-185. The MIT Press, Massachusetts.
- Meese, T. S. (1993) Feature coding in human pattern vision. PhD Thesis, Bristol, UK.
- Meese, T. S. (1995a) Using the standard staircase to measure the point of subjective equality: A guide based on computer simulations. *Perception and Psychophysics*, 57, 267-281.

- Meese, T. S. (1995b) Phase-reversal discrimination in one and two dimensions: Performance is limited by spatial repetition, not spatial frequency content. *Vision Research*, 35, 2157-2167.
- Meese, T. S. (1999a) A model of human pattern vision: Association fields for adaptive filters. *Spatial Vision*, 12, 363-394.
- Meese, T. S. (1999b) A model for perception of structure in stationary plaids using adaptive filters. [ARVO Abstract]. *Invest Ophthalmol Vis Sci*, 40, s350.
- Meese, T. S. (1999c) Adaptive spatial filtering prior to edge analysis: A nonlinear image processing model for the perception of stationary plaids. *Perception*. [ECVP Abstract], 28 supp, 102.
- Meese, T. S. & Freeman, T. C. A. (1995) Edge Computation in human vision: Anisotropy in the combining of oriented filters. *Perception*, 24, 603-622.
- Meese, T. S. & Georgeson, M. A. (1991) Edge computation in human vision: What controls the combining of oriented filters? [ECVP abstract], *Perception*, 20, 81.
- Meese, T. S. & Georgeson, M. A. (1992) Perceived structure of 3-component plaids: Combining and sharing of oriented filter outputs. *Perception*, 21, supp-2, 91 - 92.
- Meese, T. S. & Georgeson, M. A. (1996a) The tilt aftereffect in plaids and gratings: channel codes, local signs and 'patchwise' transforms. *Vision Research*, 36, 1421-1438.
- Meese, T. S. & Georgeson, M. A. (1996b) Spatial filter combination in human pattern vision: Channel interactions revealed by adaptation. *Perception*, 25, 255-277.
- Meese, T. S. & Williams, C. B. (2000) Probability summation for multiple patches of luminance modulation. *Vision Research*, 40, 2101-2113
- Milner, A. D. & Goodale, M. A. (1995) *The Visual Brain in Action*. Oxford.
- Morgan M. J. and Hotopf W. H. N. (1989) Perceived diagonals in grids and lattices. *Vision Research*, 29, 1005-1015.
- Morgan, M. J., Mason, J. S. & Solomon, J. A. (1997) Blindsight in normal observers? *Nature*, 385, 401-402.
- Morgan, M. J. & Watt, R. J. (1997) The combination of filters in early spatial vision: a retrospective analysis of the MIRAGE algorithm. *Perception*, 26, 1089-1100.
- Morrone, M. C. & Burr, D. C. (1988) Feature detection in human vision: A phase-dependent energy model. *Proc R. Soc Lond. B*. 235, 221-245.
- Morrone, M. C. & Burr, D. C. (1992) A model of human feature detection based on matched filters. in *Robots and Biological Systems*. Dario, P. & Sandini, G. (Eds.).
- Morrone, M. C., Navangione, A. & Burr, D. C. (1995) An adaptive approach to scale selection for line and edge detection. *Pattern Recognition Letters*, 16, 667-677.
- Moutoussis, K. & Zeki, S. (1997) A direct demonstration of perceptual asynchrony in vision. *Proc R. Soc. Lond. B*, 264, 393-399.
- 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.
- Olzak, L. A. & Wickens, T. (1997) Complex pattern discriminations: orientation information is integrated across spatial scale; spatial frequency and contrast information is not. *Perception*, 26, 1101-1120.
- Phillips, G. C. & Wilson, H. R. (1984) Orientation bandwidths of spatial mechanisms measured by masking. *Journal of the Optical Society America A*, 1, 226-232.

- Robertson, L., Treisman, A., Friedman-Hill, S. & Grabowecky, M. (1997) The interaction of spatial and object pathways: Evidence from Balint's syndrome. *Journal of Cognitive Neuroscience*, 9, 295-317.
- Robson, J. G. (1980) Neural images: The physiological basis of spatial vision. In *Visual Coding and Adaptability*, Harris, C. S. (Ed), LEA, 177-214.
- Shadlen, M. N. & Movshon, J. A. (1999) Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron*, 24, 67-77.
- Singer, W. & Gray, C. M. (1995) Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.*, 18, 555-586.
- Smith, S., Wenderoth, P. & van der Zwan, R. (2001) Orientation processing mechanisms revealed by the plaid tilt illusion. *Vision Research*, 41, 483-494.
- 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 A*, 13, 689-696.
- Tolhurst, D. J. & Barfield, L. P. (1978) Interactions between spatial frequency channels. *Vision Research*, 18, 951-958.
- Treisman, A. (1996) The binding problem. *Current Opinion in Neurobiology*, 6, 171-178.
- Treisman, A. & Schmidt, H. (1982) Illusory conjunctions in the perception of objects, *Cog Psychol*, 14, 107-141.
- Watson, A. B. (1982) Summation of grating patches indicates many types of detector at one retinal location. *Vision Research*, 22, 17-25.
- Watt, R. J. (1988) *Visual processing: computational, psychological and cognitive research*. East Sussex: Lawrence Erlbaum Associates Ltd.
- Watt, R. J. & Morgan, M. J. (1985) A theory of the primitive spatial code in human vision. *Vision Research*, 25, 1661-1674.
- Wilson, H. R. McFarlane, D. K. & Phillips (1983) Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research*, 23, 873-882.
- Wilson, H. R. & Wilkinson, F. (1997) Evolving concepts of spatial channels in vision: from independence to nonlinear interactions. *Perception*, 26, 939-960.
- Wolfe, J. E. & Cave, K. R (1999) The psychophysical evidence for a binding problem in human vision. *Neuron*, 24, 11-17.
- Yen, S. -C. & Finkel, L. H. (1998) Extraction of perceptually salient contours by striate cortical networks. *Vision Research*, 38, 719-741.

Appendix A

The perception of plaids suggests combination of spatial filters

Experiments on perceived spatial structure in plaids (see Georgeson & Meese, 1997 for a review) suggest that in human vision, spatial frequency and orientation tuned filters are bound or segmented in an image dependent way. By ‘binding’, we mean the process by which filters are associated and assigned a weight w ($0 \leq w \leq 1$) in a single assembly prior to summation. If the filtering and summation were linear, this would generate another linear filter (Georgeson, 1992; Freeman & Adelson, 1991). By ‘segmentation’, we mean parallel output streams of spatial filters *after* summation.

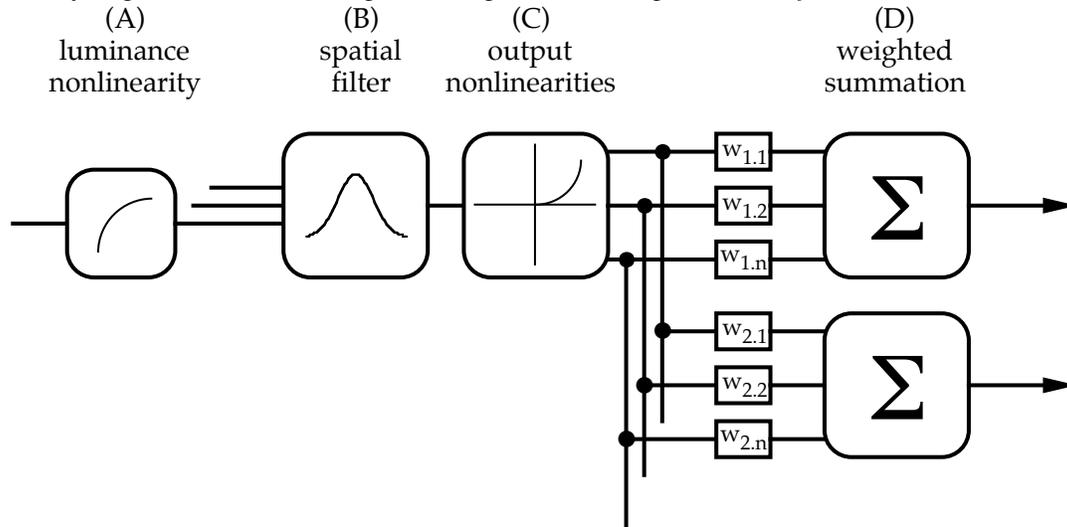


Fig A1. Model of early spatial vision. (A) Luminance nonlinearity for each sample point in the retinal image. (B) Quasi-linear spatial basis-filters constructed by inflexible weighting and summing of multiple inputs from (A). (C) Spatial filter output nonlinearities. (D). Flexible weighted summation of spatial basis-filters to create one or more synthetic filters. Note that in principle, the flexible weighting at (D) means that the multiple outputs could be either independent or (partially) correlated. We argue that only stages (B) and (D) are needed to understand our results on the perception of plaids and that stages (A) and (C) are insufficient to account for our results. (See text for details).

Fig A1 is a general scheme showing a series of putative stages in early vision. The first stage (A) is a pointwise luminance nonlinearity (compressive or expansive). This is followed by inflexible weighted spatial summation between different points in the retinal image (B). If it were not for the initial luminance nonlinearity (A) these mechanisms (B) would be linear filters. As it is, we refer to them as quasi-linear filters. This stage is followed by a second nonlinear stage (C), which encompasses plausible output nonlinearities such as thresholding, contrast compression or expansion, and full- or half-wave rectification. Stages A to C are consistent with much evidence on spatial vision (e.g. Graham, 1989; DeValois & DeValois, 1990). Finally, a nonlinear stage (D) consists of a decision process where multiple input stages (A-C) are selected and weighted prior to linear summation. In general, the same or different weights for the same or different filters can appear on multiple output lines at D. Aside from the *alternative view* below, we argue that stage D is essential for understanding our catalogue of results on the perception of spatial structure in stationary plaid stimuli, and that the nonlinearities A and C are insufficient by themselves to account for our results. The image processing model presented in this paper does not include a luminance nonlinearity at A or a filter-output nonlinearity at C.

Interpretation of the perception of plaid stimuli

At low component contrasts (e.g. 2%), a briefly presented plaid stimulus made from sine-wave components at $\pm 45^\circ$ (a *standard plaid*) tends to look like two overlapping gratings, suggesting that the components are processed independently. However, at higher contrasts, the stimulus looks like a blurred checkerboard containing vertical and horizontal edges (Georgeson, 1990). Why might this be so? One possibility is that at higher contrasts, perceived structure is mediated by vertical and horizontal labelled-filters at B, which respond to vertical and horizontal distortion products (DP) introduced by the initial luminance nonlinearity (A). For example, treating the nonlinearity at A as the sum of a linear component and a quadratic component, the spatial frequency of the DPs introduced by the quadratic term for a standard plaid are $1.414f$ and have vertical and horizontal orientations (Derrington & Henning, 1989), regardless of whether the coefficient of the quadratic component is positive (expansive nonlinearity) or negative (compressive nonlinearity). However, we can reject this explanation for the following reason. After adapting to a vertical grating with spatial frequency either the same ($1f$), or one octave lower ($0.5f$) than the fundamental spatial frequency ($1f$) of a standard test plaid, the test plaid appears checkerboard-like but stretched horizontally (Meese & Georgeson, 1996a). If the perceived structure of plaids were due to the presence of DPs then this aftereffect would not be expected. This is because the spatial frequency of the vertical adapter ($0.5f$ or $1f$) is lower than that of the putative vertical DP ($1.414f$), and so, according to conventional accounts of spatial aftereffects (Blakemore & Sutton, 1971), the vertical DP would be expected to have a higher perceived spatial frequency post-adaptation. If this were so,

then presumably the plaid should appear *compressed* horizontally, not stretched. DPs and orientation-labelled linear filters do not provide a direct account of perceived structure in plaids.

Consider now the response of stage B to a standard plaid. Could the response of a *single* filter at this stage mediate the checkerboard appearance of a standard plaid? First consider a filter that is sensitive to only *one* of the stimulus components (i.e. a filter oriented around $+45^\circ$ or -45°) and assume that DPs introduced by stage A are negligible. It follows that the output of any such filter would have iso-response contours at the same orientation as the stimulus component to which it is sensitive. Output nonlinearities such as thresholding, compression and rectification transform iso-responses in exactly the same way at each place in the image. Thus, the nonlinearity at stage C could not transform the oblique orientations of iso-response contours in the filter under discussion and therefore could not account for the checkerboard-like appearance of a standard plaid. Assume now that the DPs are severe. As mentioned above, the largest DPs will be vertical and horizontal and will have spatial frequency of 1.414f. If the bandwidths of the oblique filters under discussion are sufficiently narrow they will not respond to these DPs and their presence is of no consequence. If the bandwidths are broad and presumably, approximately symmetric, then the oblique filters will respond to a single oblique stimulus component, plus *both* of the DPs. This situation is shown in Fig A2, where, from left to right, the contrast-response to each of the DPs is (i) the same, (ii) one half, and (iii) one fifth, of the response to a right oblique stimulus component. The presence of the DPs means that in general, the orientation of iso-response contours are not the same as the stimulus component orientation (right oblique). However, it is clear from Fig A2, that vertical and horizontal iso-response contours do not predominate, as they do in the perception of a standard plaid.

Thus, we have rejected explanations of the checkerboard appearance of a standard plaid: (i) in terms of spatial filters sensitive to DPs, and (ii) in terms of spatial filters sensitive to only one of the stimulus components, both with and without the presence of DPs. This means that there must be at least one spatial mechanism that is sensitive to *both* components of the stimulus. The only plausible candidates at stage B are circular filters and vertical and horizontal filters with sufficiently broad orientation bandwidths. However, because of their symmetrical relationship with a vertical and/or horizontal adapting gratings, none of these filters could mediate spatial distortions such as the two-dimensional tilt aftereffect described earlier (Meese & Georgeson, 1996a). Note that the presence of the nonlinear stage at C has no impact on this argument—output nonlinearities cannot introduce anisotropies.

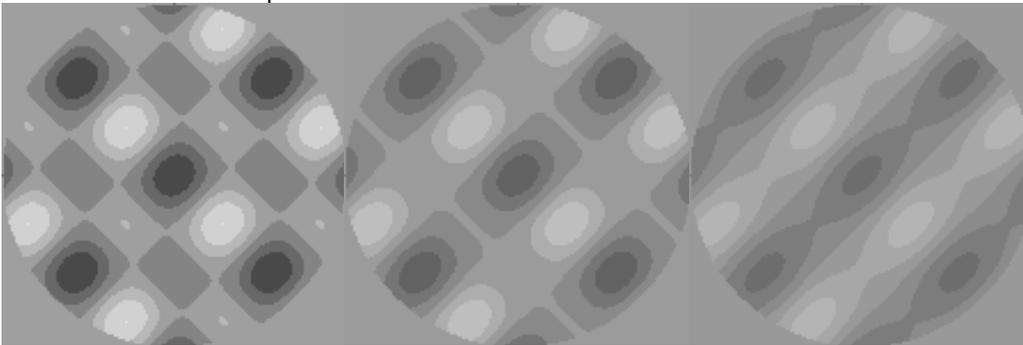


Fig A2. Iso-response contours to a right oblique 1f component (2.5 cycles per image), with the addition of vertical and horizontal distortion products (DPs) at 1.414f. From left to right the response to each DP is (i) the same, (ii) one half and (iii) one fifth of the response to the oblique stimulus component.

Because, (i) as argued above, there must be a filter that is sensitive to both components in a plaid, and (ii) the above arguments rule out the positioning of this filter at stage B, this filter must be created by summing together two or more oriented filters (B), at stage D. In sum, stage A and stage C are immaterial to our interpretation—most of our results require only stages B and D. However: (1) a complete understanding of the two-dimensional tilt-aftereffect may require additional processes (see Meese & Georgeson 1996a for details), and (2) implementation details may require some parallelism between B & C that is not shown in Fig A1. This would allow the outputs of both phase-preserving filters (e.g. simple-cells) and contrast-energy filters (e.g. complex-cells) to be inspected at stage D (see Discussion).

Alternative view

It has been suggested to us that some of our results with plaids could be understood in terms of a static filter at stage B, followed by some unspecified nonlinear feature extraction process that emulates adaptive filtering by revealing features consistent with either bound or segmented stimulus components. We cannot rule out this possibility, though it is difficult to see how such a process could account for the two-dimensional tilt aftereffect of Meese & Georgeson (1996a). We suggest, therefore, that for the present at least, our earlier account is the more parsimonious.