Mechanism independence for texture-modulation detection is consistent with a filter-rectify-filter mechanism

FREDERICK A.A. KINGDOM,[1] NICOLAAS PRINS,[1] AND ANTHONY HAYES[2]

[1]McGill Vision Research Unit, Department of Ophthalmology, McGill University, Montréal, Québec, Canada
[2]Department of Psychology, University of Hong Kong, Pokfulam Road, Hong Kong, China

Abstract

The ability of the visual system to detect stimuli that vary along dimensions other than luminance or color—"second-order" stimuli—has been of considerable interest in recent years. An important unresolved issue is whether different types of second-order stimuli are detected by a single, all purpose, mechanism, or by mechanisms that are specific to stimulus type. Using a conventional psychophysical paradigm, we show that for a class of second-order stimuli—textures sinusoidally modulated in orientation (OM), spatial frequency (FM), and contrast (CM)—the human visual system employs mechanisms that are selective to stimulus type. Whereas the addition of a subthreshold mask to a test pattern of the same stimulus type was found to facilitate the detection of the test, no facilitation was observed when mask and test were of different types, suggesting mechanism independence for the different types of stimulus. This finding raises the important question of whether mechanism independence is compatible with the well-known filter-rectify-filter (FRF) model of second-order stimulus detection, since FRF mechanisms, in principle, do not discriminate between stimulus types. We show that for all mask/test combinations except those with CM masks, the FRF mechanism giving the largest response to the test modulation is largely unaffected by subthreshold levels of a different stimulus-type mask. For this reason, we cannot rule out the possibility that FRF mechanisms mediate the detection of our stimuli. For combinations involving CM masks, however, we propose that a process of contrast normalization renders the test stimulus insensitive to the mask stimulus.

Keywords: Texture gradient detection, Second-order processing, Filter-rectify-filter model, Independent mechanisms

Introduction

Over the last 10 years, there has been much interest in how the visual system detects "second-order" stimuli (see, e.g. recent reviews by Wilson, 1999; Baker, 1999). A second-order stimulus is defined here as one that is modulated along a dimension other than luminance or color; for example, a texture edge that has unvarying mean luminance. Cortical neurones that respond to luminance contrast, such as simple cells, while implicitly carrying information about the presence of the texture edge, are unable to disentangle the edge information from that of the texture's fine-grain luminance detail. Instead, a nonlinear mechanism that smooths out the fine-grain detail while capturing the information about the edge is required for the texture-edge information to be represented explicitly.

Neurones with second-order properties have been isolated in a variety of visual areas in both cat (Zhou & Baker, 1994, 1996; Mareschal & Baker, 1998, 1999; Leventhal et al., 1998) and

monkey (Peterhans & von der Heydt, 1991; Olavarria et al., 1992; Grosof et al., 1993; Albright, 1995; Sary et al., 1995; Geesaman & Anderson, 1996; Chaudhuri & Albright, 1997; O'Keefe & Movshon, 1998). Moreover, psychophysical (Ledgeway & Smith, 1994; Nishida et al., 1997; Schofield & Georgeson, 1999), neurophysiological (reviewed by Baker, 1999), neuropsychological (Vaina & Cowey, 1996; Vaina et al., 1998), and brain-imaging studies (Smith et al., 1998; Wenderoth et al., 1999) suggest that first-order and second-order stimuli are processed by different mechanisms. These studies are complemented by analyses of natural images that reveal the importance of significant amounts of second-order structure in natural scenes; structure that is not always correlated with the scenes' first-order properties (e.g. Schofield, 2000; Geisler et al., 2001)

The filter-rectify-filter model

The "standard model" for processing second-order information is the filter-rectify-filter (FRF) model (Wilson et al., 1992; see reviews by Wilson, 1999; Baker, 1999), also termed a "second-order" (Cavanagh & Mather, 1989), "non-Fourier" (Chubb & Sperling, 1988; Wilson, 1999), "back-pocket" (Chubb & Landy, 1991) or "complex channel" model (Graham et al., 1992). The
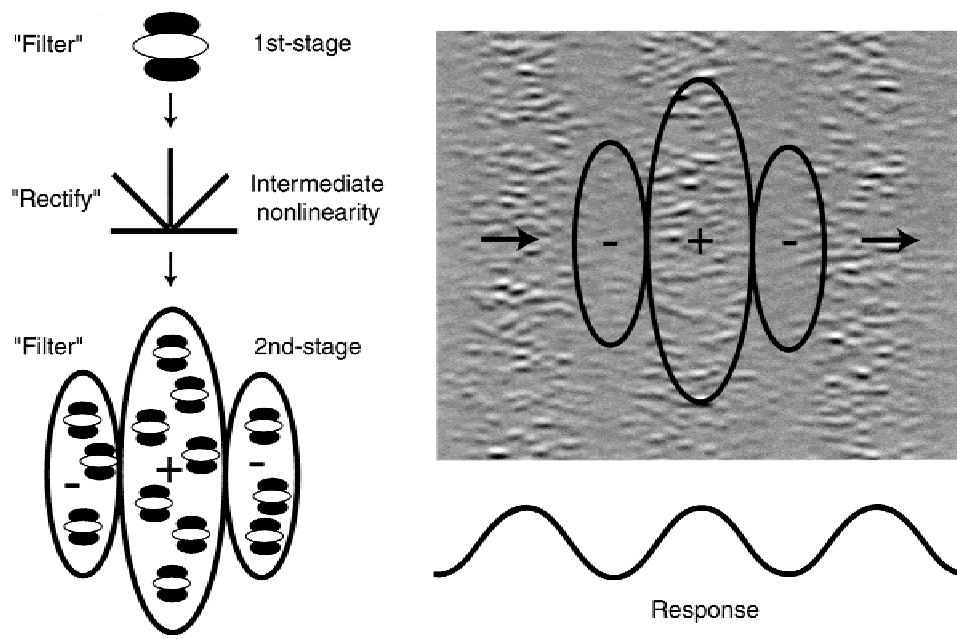
Fig. 1. Schematic representation of the filter-rectify-filter (FRF) model in response to a contrast-modulated stimulus, as shown. The "first-stage filters" are linear and narrowband, and convolve the input image at a number of spatial scales and orientations. The output is subject to a nonlinearity—typically full-wave rectification. The "second-stage filter" has a larger receptive field than the "first-stage filters," and sums the nonlinearly transformed outputs of the first-stage filters.

principle of the model is illustrated in Fig. 1. An image is first convolved with a bank of narrowband linear filters tuned to various spatial frequencies and orientations. With their simple-cell-like receptive fields, these "first-stage" filters together approximate a wavelet transform. The outputs of the first-stage filters are then subject to a nonlinearity, typically full-wave rectification, that corresponds to the summing of the half-wave rectified responses of pairs of ON-center and OFF-center filters (e.g. Schiller, 1982). The nonlinearity serves to make all responses positive, preventing their mutual cancellation when linearly pooled by the "second stage." The second-stage filters, which have larger receptive fields than their first-stage counterparts, but which in other ways are close analogues, sum the nonlinearly transformed first-stage outputs. In the model FRF mechanism illustrated by Fig. 1, both the first and the second stage are shown as being orientation and spatial frequency tuned. These model properties receive support from both single-unit recordings (e.g. Mareschal & Baker, 1998, 1999) and psychophysical studies that have conducted separate manipulation of the envelope and carrier information of various types of second-order stimuli (Sutter et al., 1995; Arsenault et al., 1999; Derrington & Ukkonen, 1999; Dakin & Mareschal, 2000; Kingdom & Keeble, 2000; Landy & Oruc, 2002). The differences between particular FRF models lie primarily in the form of the intermediate nonlinearity (half-wave rectification, full-wave rectification, squaring, etc.), and/or the manner in which the outputs of the second-stage filters with different carrier tunings are combined prior to the decision stage (e.g. in Landy & Bergen, 1991, second-stage filters with orthogonally tuned first-stage inputs are subtracted to produce an orientation-opponent signal).

If the second-stage filters of FRF mechanisms have excitatory and inhibitory receptive-field subregions, as shown in Fig. 1, they will respond to spatio-temporal variations in any number of different types of second-order stimuli. This inherent flexibility of FRF mechanisms is primarily a consequence of having an inter-

mediate nonlinearity which delivers an *energy* response to the second stage. If two parts of a stimulus comprise elements differing along any second-order dimension, for example in contrast, orientation, size, density, or direction-of-motion, then this will produce a differential output in all first-stage filters whose spectral tuning characteristics overlap with those of the stimulus. This differential response can then be detected by the appropriate second-stage filter. Thus, FRF mechanisms have been proposed to mediate effortless texture segmentation irrespective of whether the textures are defined by orientation (Malik & Perona, 1990; Bergen & Landy, 1991; Landy & Bergen, 1991) or by the arrangement of their elements (Graham et al., 1992; Sutter & Graham, 1995; Graham & Sutter, 1998). FRF mechanisms have been proposed to mediate the detection of textures that are smoothly modulated in contrast (Sutter et al., 1995), orientation (Kingdom & Keeble, 1996), and spatial frequency (Arsenault et al., 1999), and have also been proposed to mediate the detection of moving contrast-modulated patterns (Chubb & Sperling, 1988), illusory contours (Wilson, 1999), glass patterns (Wilson et al., 1997; Wilson & Wilkinson, 1998), and boundary curvature (Dobbins et al., 1989; Wilson & Richards, 1992).

The flexibility of FRF mechanisms may be considered as both an asset and an impediment to their usefulness. On the positive side, FRF mechanisms have the potential to act as general-purpose "salience" detectors (Lennie, 1998), alerting the visual system to the "where," (Prins & Mussap, 2001) if not necessarily the "what," of any spatio-temporal textural discontinuities in a scene. However, on the negative side, FRF mechanisms might prove incapable of providing the necessary information for the visual system to identify particular types of second-order stimuli.

The above considerations lead to the following question. Are different types of second-order stimuli detected by different mechanisms? And, if so, can these mechanisms be sensibly modeled as FRF mechanisms? In this paper, we have attempted to answer both
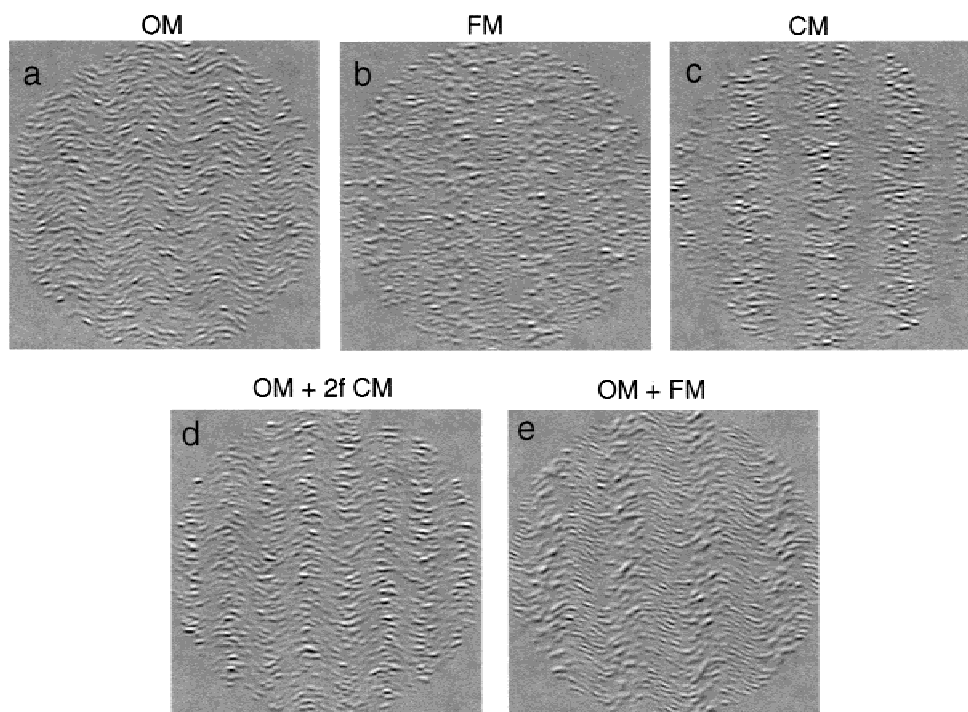
Fig. 2. Examples of sinusoidally modulated texture gratings used in the experiments. The top row (a–c) shows examples of gratings modulated in one dimension—orientation (OM), frequency (FM), and contrast (CM). The second row (d, e) shows examples of gratings modulated in two dimensions—OM + 2f CM, and OM + FM.

these questions for an important subclass of second-order stimuli—sinusoidally modulated texture gratings. We employed the following three types of texture grating modulation: contrast modulation (CM), orientation modulation (OM), and spatial-frequency modulation (FM). Examples of each are shown in Fig. 2. A brief report of the experimental findings of this study has been given in Kingdom and Hayes (2000).

Testing for mechanism independence

A conventional method for establishing whether two stimuli are detected by a common mechanism is to measure the amplitude threshold of one—the "test"—in the presence of various amplitudes of the other—the "mask." If subthreshold levels of the mask are found to lower the threshold for the test, resulting in the well known "dipper" function most commonly associated with contrast discrimination data (Campbell & Kulikowski, 1966; Foley & Legge, 1981; Legge & Kersten, 1983; Ross & Speed, 1991; Foley, 1994), then this suggests that mask and test are processed by the same mechanism.* On the other hand, the absence of a dipper indicates that mask and test are detected by different mechanisms.

---

*The exact cause of the dipper function is contentious, and is believed to be due either to a threshold (i.e. accelerating) nonlinearity (Foley & Legge, 1981), or uncertainty reduction in the mechanisms transducing the stimulus (Pelli, 1985). A mechanism with a purely linear transducer function would not, of course, produce a dipper, and therefore any mask+test combination in which the test was detected by a linear transducer would also not show a dipper, even if it were independently detected. It follows that for our experiment to be valid, the uni-modal conditions must show dippers; without these, any observed absence of a dipper in the cross-modal conditions would not constitute unequivocal evidence for independence.

This method has been used to test for mechanism independence in the detection of luminance- versus color-modulated patterns (Mullen & Losada, 1994), luminance- versus contrast-modulated patterns (Schofield & Georgeson, 1999), and bars versus edges (Burr et al., 1989). When the mask and test are of the same stimulus type, we refer to the resulting combination as a "unimodal" condition, whereas if mask and test are of a different stimulus type we refer to the resulting combination as a "cross-modal" condition.

Combining texture gratings to test the FRF model

Combining different types of texture grating to produce test+mask combinations is not a straightforward matter. The reason is that for the class of texture gratings employed here, different FRF mechanisms will respond optimally to different forms of test+mask combination. All our texture gratings were modulated with low amplitude about a common d.c. Gabor (Gaussian-windowed sinusoid) that was horizontally oriented with a spatial frequency of 4.0 cpd and a contrast of 33% (% amplitude/mean). Consider first a hypothetical FRF mechanism whose first-stage inputs were tuned to the d.c. orientation and spatial frequency of our stimuli. A schematic version of this mechanism is shown in Fig. 3, together with its response to one cycle of a CM (top) and OM (bottom) grating with the same d.c. Gabor. The FRF mechanism has been given an even-symmetric second-stage receptive-field profile, but the principle would apply equally well for other profiles. As the figure illustrates, the response of the mechanism to the CM grating is of the same spatial frequency and phase as the grating itself, with the maximum response when the filter is centered on Gabors with the highest contrast, and the minimum response when the filter is centered on Gabors with the lowest contrast. Consider now the
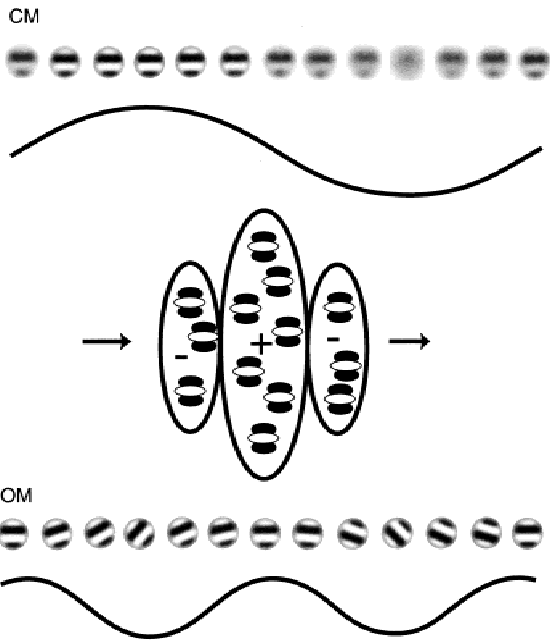
Fig. 3. Schematic diagram of a filter-rectify-filter (FRF) mechanism with horizontally oriented first-stage inputs, and its response to one cycle of a CM (top) and OM (bottom) grating, both modulated around a horizontal d.c. Gabor. Note the frequency-doubling in the FRF response to the OM grating.

response of the same mechanism to the single-cycle OM grating. In this case, the first-stage inputs of the FRF mechanism are least responsive not just to the trough of the grating, but also to its peak, and the peak filter response will occur when the filter is centered not on the peak, but the d.c. Because the OM waveform passes through its d.c. at twice the frequency of the grating itself, the response of the FRF mechanism will be frequency doubled and phase shifted by a quarter cycle.

Given that our texture gratings all have the same d.c., where one of the grating components is a CM grating, an FRF mechanism tuned to the d.c. would give its biggest response when the CM grating component was at twice the spatial frequency as the other (FM or OM), and phase shifted by 90 deg. An example combination OM+2fCM grating is shown in Fig. 2d, and the filtering principles behind this combination are illustrated in Fig. 4. We thus combined CM gratings with OM or FM gratings at double the frequency of the OM or FM gratings. When combining OM and FM gratings, on the other hand, we set the spatial frequencies to be equal; an example is shown in Fig. 2e.

It should be stressed that in justifying the use of frequency-doubled CM gratings in our cross-modal conditions, we are not claiming that if an FRF mechanism were to detect our stimuli, it must be the one whose carrier is centered on the d.c. Rather, we suggest that if it were d.c. centered, then its optimum stimulus would be a combination with a frequency-doubled CM component.

An FRF mechanism whose carriers were tuned to either the peak, the trough, or some value outside of the peak or trough, of
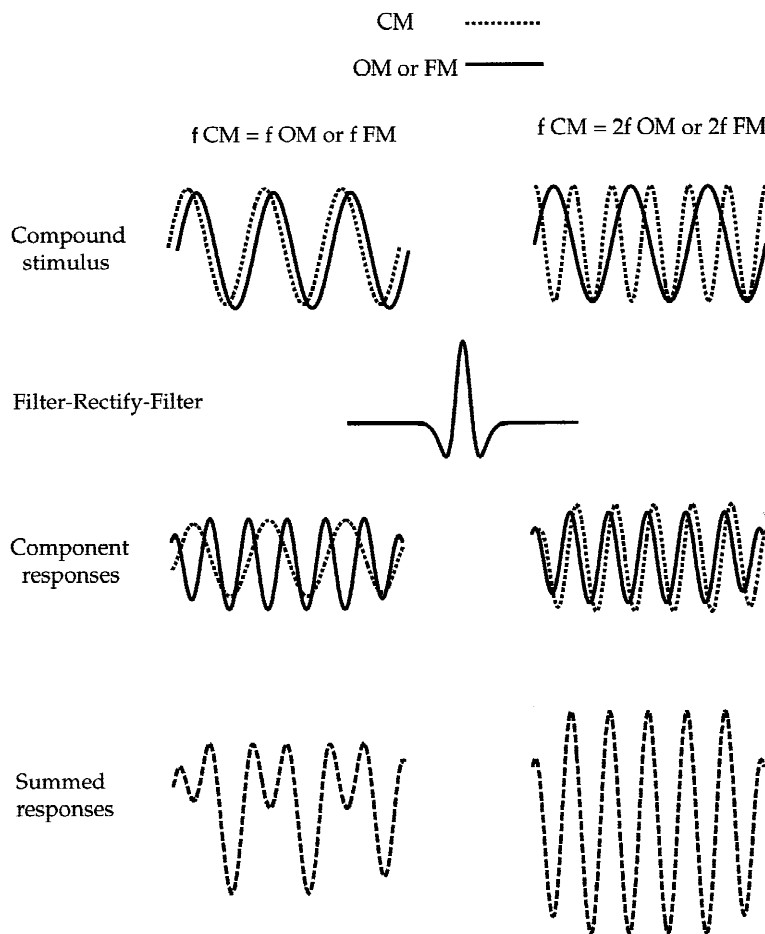

Fig. 4. The figure illustrates the optimum method of combining a CM with either an OM or FM grating in order to stimulate an FRF mechanism whose first-stage inputs are tuned to their common d.c. On the left the two component gratings have the same texture spatial frequency, but the frequency-doubling of the response to the OM/FM grating results in some partial cancellation of the two responses. On the right is shown how doubling the texture spatial frequency of the CM grating, and shifting its phase by 90 deg, brings the two responses into alignment, which when summed results in a relatively larger response.

the OM or FM waveform would, on the other hand, produce a response of the same spatial frequency as the waveform itself. Therefore, we also tested CM combination gratings with identical component spatial frequencies. Finally, we tested a variety of grating-component phase relationships.

Methods

Observers

Two of the authors, FK and AH, acted as observers. Both had normal, or corrected to normal, vision and were experienced psychophysical observers.

Stimuli

Generation

Example stimuli are shown in Fig. 2. The stimuli were generated by a VSG2/3 video-graphics card (Cambridge Research Systems, UK) and displayed on a BARCO Calibrator monitor. The displays were monochrome and were gamma corrected by selecting the appropriate 8-bit range (256 intensity levels) from 12-bit (4096 intensity levels) DACs (digital-to-analog converters) following calibration using a UDT photometer. Monitor mean luminance was 35 cd/m^2.

Gabor micropatterns

Gabors were generated by multiplying a sine function by a two-dimensional Gaussian envelope;

$$L(x, y) = M + A\cos[2\pi f_1(x\cos\theta - y\sin\theta) + \rho]$$
$$\times \exp[-(x^2 + y^2)/2\sigma^2]$$

with mean luminance, M; amplitude, A; luminance spatial frequency, f_1; orientation, θ phase, ρ; and σ the standard deviation of the Gaussian envelope, which was circularly symmetric. All Gabors had a mean luminance of 35 cd/m^2, a spatial-frequency bandwidth at half-height of 1.5 octaves, and a phase, ρ, of $\pi/2$ radians, making them all "odd-symmetric," and therefore d.c. balanced. In all stimuli, the Gabors were modulated in one or more of their properties (contrast, orientation, or spatial frequency) around a common d.c. Gabor with amplitude, A, of 33% of mean luminance (i.e. contrast 33%), spatial frequency of 4.0 cpd, and orientation of 90 deg (horizontal).

OM, FM, and CM gratings

Figs. 2a–2c show example stimuli. Each stimulus was a circular patch with a diameter of 8.6 deg visual angle at the viewing distance of 100 cm, and contained 2000 micropatterns. The position of each micropattern within the stimulus window was randomized. The orientations, spatial frequencies, and contrasts of the micropatterns were constrained in that they were made to vary sinusoidally along the horizontal axis of the display around the d.c. Gabor level. The amplitude of modulation, or texture amplitude, was determined by how much the Gabors changed in their properties throughout one complete cycle of orientation modulation. The CM gratings were modulated in contrast; for example, an amplitude of 5% would imply that the Gabors differed in contrast by an absolute value of 10% between the peak and trough of the waveform. The FM gratings were modulated in octave steps of spatial frequency; thus, an amplitude of 0.5 octaves would imply

that the Gabors varied in spatial frequency by a factor of 2 (1 octave) between peak and trough. The OM gratings were modulated in orientation; thus, an amplitude of 5 deg implied a 10-deg difference between peak and trough. For the FM gratings the bandwidth of the Gabors was held constant, and thus the Gabors "scaled up and down" as they traversed the display. One consequence of this is that the contrast power of our FM gratings was not constant, but varied sinusoidally, a fact that we will return to in the Discussion section. For all stimuli, when Gabors overlapped, their amplitudes, though not mean luminances, were added. Occasionally, the random positioning produced overlapping Gabors whose combined amplitudes exceeded the look-up-table of the display. This overrun was found to occur for about 2% of pixels. A simple clipping procedure removed these overflows/underflows. Texture spatial frequency was fixed throughout at 0.35 cpd, producing 3 cycles of orientation modulation across the stimulus. This texture spatial frequency was close to that producing peak sensitivity for orientation gratings made from similar Gabors to those used here (Kingdom et al., 1995). The phase of the texture modulation was randomized on each stimulus presentation.

Cross-modal conditions

Example combinations of texture modulations are shown in Figs. 2d and 2e. There are in principle two methods that one may use to combine different texture modulations when the textures are comprised of micropatterns. Either the same micropattern can carry both modulations, or different micropatterns can carry different modulations. We chose the same-micropattern option. However, we assume that our results would be qualitatively the same had we employed the different-micropattern option.

Procedure

A two-interval-forced-choice (2IFC) "two-up-one-down" staircase procedure was used to measure the threshold amplitude of texture modulation at the 70.7% correct level. On each trial two stimuli were presented; each for 200 ms, separated by a 500-ms interstimulus interval. The task for the observer was to decide which interval contained the test modulation. A tone signalled an incorrect response. The only difference between the two intervals was the amplitude of the test texture modulation, which in the comparison stimulus was zero. The staircase procedure increased or decreased the test amplitude by a factor of 1.25 on those trials where a change was required. The staircase was terminated after ten reversals, and the threshold amplitude was calculated as the geometric mean amplitude of the last eight reversals. For the CM test, thresholds were expressed as percent contrast modulation; for the FM test, thresholds were given in octaves; and for the OM test, thresholds were in degrees.

In any given experimental session, only one type of mask-test combination was tested, and observers were made aware of the combination. The order of mask amplitudes, however, was randomized within each session.

Results

Fig. 5 shows the results when mask and test were constructed from the same type of texture grating. This experiment can be thought of as measuring the increment threshold in amplitude as a function of pedestal amplitude. The horizontal dotted line shows the amplitude threshold in the absence of a mask. The vertical dotted line shows the point where the mask reaches its own threshold which, since mask and test are the same type of grating, is the same value as the
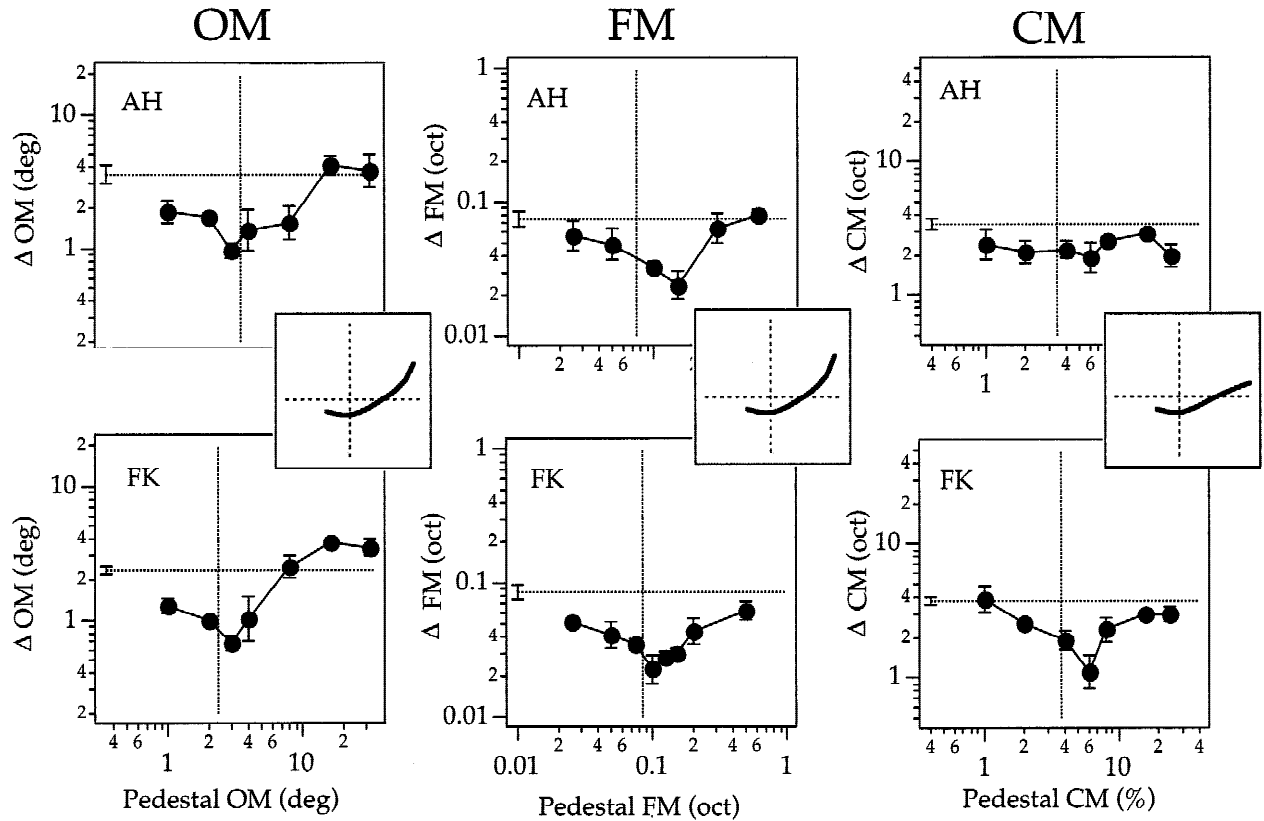
Fig. 5. Results for the unimodal conditions. Threshold increment amplitudes are plotted as a function of pedestal amplitude for the three types of texture grating, and for two subjects. Note the units of measurement: deg for OM; octaves for FM, and % for CM. The horizontal and vertical dotted lines are thresholds for the test and mask when presented alone (same values in the unimodal conditions). Insets are the predictions from an FRF model whose first-stage inputs were chosen to maximize the differential response to the two half-cycles of texture modulation, and whose output was subject to a positively accelerating, followed by decelerating nonlinearity (see text for details).

zero-mask test threshold. In all figures, as the mask amplitude increases from zero, thresholds for detecting the increment decline until just beyond the point where the threshold for the mask is reached. These results serve as a necessary precursor to the main results of the study, demonstrating that a dipper is characteristic of the texture gratings we used.

Fig. 6 shows the results when test and mask were different types of texture grating, but with the same modulation spatial frequency and phase. Test-versus-mask functions for all combinations of OM, FM, and CM gratings are shown. To enable both observers' data to be presented on the same graphs, both ordinate and abscissa values have been normalized to detection threshold; in the case of the ordinate values, normalization is to the test threshold in the absence of a mask; in the case of the abscissa values, normalization is to the threshold for the mask alone (both shown as unity). In none of the graphs is there any hint of a systematic dipper function in the region where the mask is subthreshold. In one graph, FK's FM-mask/CM-test, there is some facilitation, but only at suprathreshold mask amplitudes. Such an effect has been found with other types of mask/test combinations (Mullen & Losada, 1994), and is generally believed to be due to a reduction in positional uncertainty: with the mask modulation clearly visible, the observer "knows where to look" when comparing the test at different stimulus locations. There is also a hint of subthreshold masking in AH's FM-mask/CM-test data, which

raises the question as to whether in this condition facilitation might be observed at some other phase relationship between test and mask. We looked at the effect of phase in a separate experiment.

Fig. 7 shows results for four phase relationships for same modulation-frequency mask and test conditions. For these conditions two mask amplitudes only were used—one when the mask was zero amplitude (horizontal dashed line), and one when the mask amplitude was set to just above its own detection threshold (the four points on each graph). Again, there is no evidence of significant facilitation at any of the four phase relationships tested.

Finally, Fig. 8 shows the results for the frequency-doubled CM conditions. Again, there is no evidence of subthreshold facilitation.

Modeling

Our results strongly suggest that our OM, FM, and CM textures are detected by independent mechanisms. Is this compatible with an FRF approach? A mechanism whose carrier was tuned to the d.c. orientation and spatial frequency of our stimuli would presumably give a univariant response to all three types of texture grating, and we would therefore expect, but did not find, subthreshold facilitation in our cross-modal conditions. FRF mechanisms that have carriers tuned to the peak, or trough, orientation, and spatial frequency would also give univariant responses to suitable combinations of FM, OM, and CM pairs. Again, we found no evidence
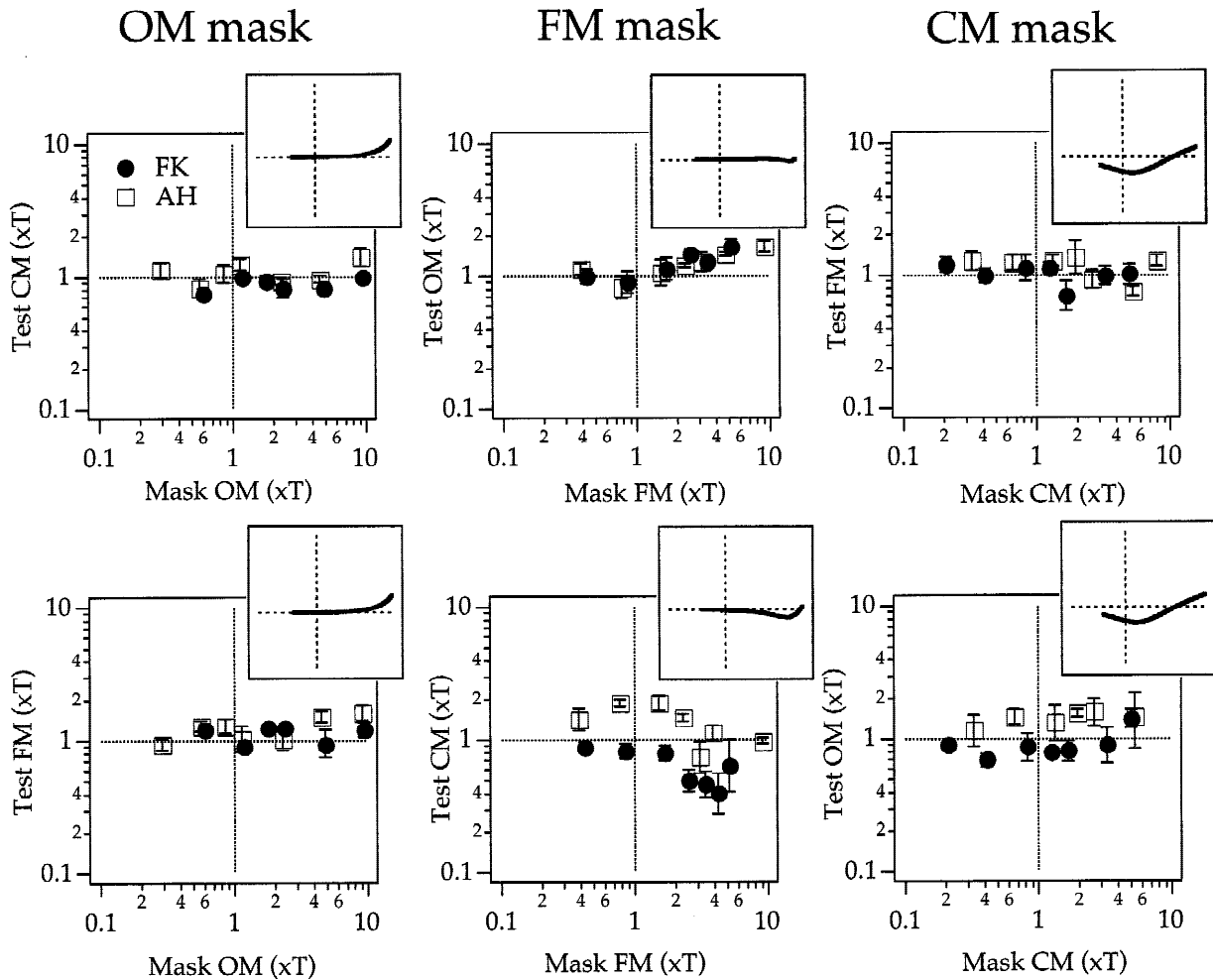
Fig. 6. Results for the cross-modal conditions. The two subjects' data are presented together on each graph. Both ordinate and abscissa values have been normalized to detection threshold (xT) and are therefore unitless. Insets are predictions for the same model as in Fig. 5. Note that the absence of a dipper in all graphs is predicted by the model in all except the CM mask conditions.

for subthreshold facilitation between any combination of texture modulations, at any relative phase. However, what of FRF mechanisms whose carriers are not centered on either the d.c., peak or trough of the stimuli, but which lie outside the range of designated orientations/spatial frequencies in the stimulus? Two of us (Prins & Kingdom, 2002) have recently obtained evidence, using an adaptation paradigm, that OM gratings are detected by mechanisms whose carrier is tuned to orientations that lie at some distance beyond the range of designated orientations in the stimulus. Similarly, the spatial frequency of the carrier of the mechanism that detects FM gratings lies at some distance outside the range of designated spatial frequencies. Prins and Kingdom also showed that this is exactly what would be expected if the visual system chose the FRF mechanism that produced the biggest *differential* response to the peak and trough of the stimuli. This leads to the tantalizing possibility that in our cross-modal conditions, the optimal FRF mechanism for detecting the test might be different from that for detecting the pedestal, thereby accounting for the lack of facilitation in the cross-modal conditions and suggesting that our results are compatible with an FRF model. We therefore explored in some detail the possibility that FRF mechanisms may mediate the detection of the stimuli in the present study.

All the modeling described in this section involved searching through all possible combinations of the orientation and spatial frequency of the carrier of a hypothetical FRF mechanism, to find that combination producing the biggest difference in output to the pedestal-only (PO) and pedestal-plus-test (PPT) intervals of a particular condition, according to a number of criteria that will be described below. It is important to emphasize that this modeling is not intended to provide precise quantitative predictions of the data, because the iterative procedures that would be necessary to do so would be far too time consuming, and are in any event not necessary to make the point we wish to make. Our primary concern is whether we can find a plausible implementation of an FRF mechanism that correctly predicts the presence of facilitation in the unimodal conditions, and the absence of it in the cross-modal conditions. If we are able to find such a mechanism, then we must conclude that we are unable to reject the possibility that FRF mechanisms underlie the detection of the texture gratings in our study.

The modeling was performed entirely in Fourier space. To estimate the stimulus amplitude spectra, we used the average amplitude spectra of multiple samples of textures generated using the same micropatterns as employed in the stimuli. Full details are
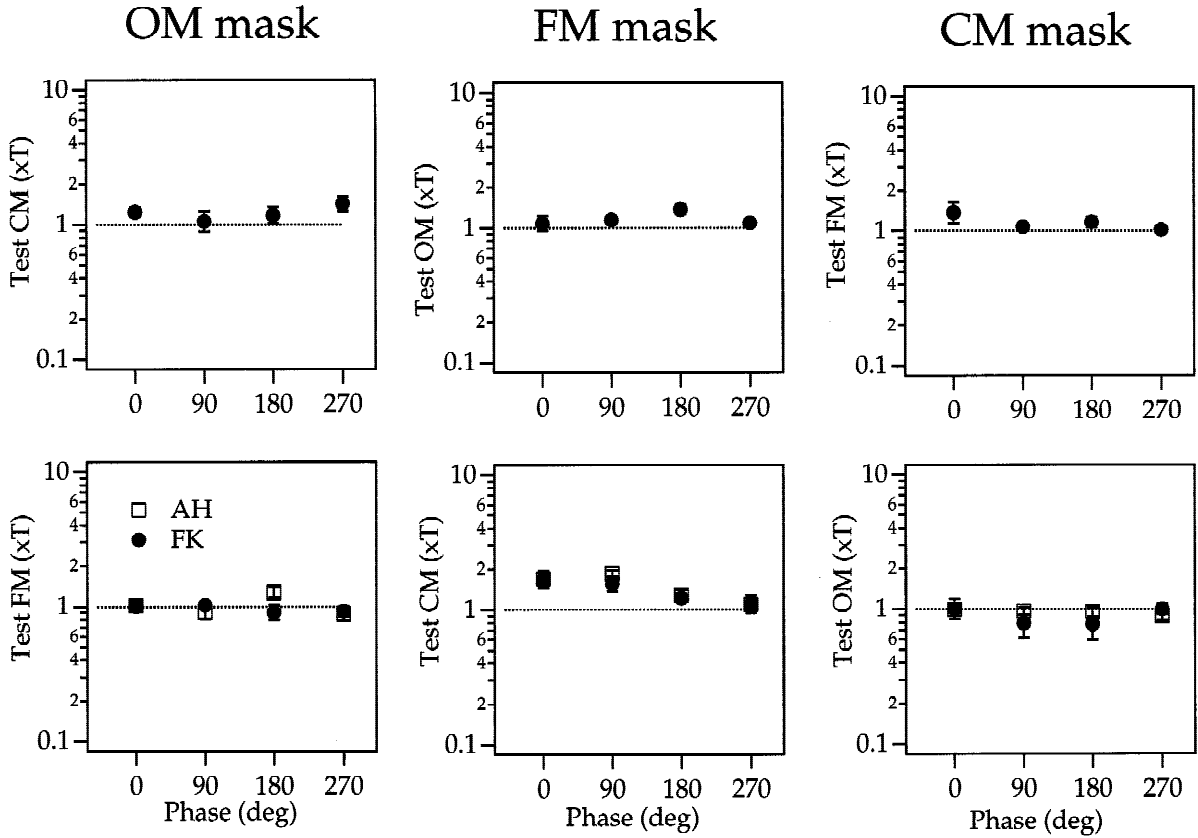
Fig. 7. Effect of relative phase in the cross-modal conditions. The mask was set to just above its own detection threshold, and test thresholds measured for four phase relationships between mask and test. The horizontal dotted line shows (normalized) thresholds for the test alone.

provided in Appendix A of Prins and Kingdom (2002).† In using the modeled average difference in amplitude spectra, rather than a measure normalized to the stochastic variability of the spectra, we have assumed that any differences between conditions in the variability of the FRF responses will not have a significant effect on the overall pattern of results. We modeled the carrier filters as Gabors with an orientation bandwidth equal to 30 deg (full-width at half-height) and a spatial-frequency bandwidth equal to 1.5 octaves. The activation of an FRF mechanism was modeled as the difference in the activation of the carrier filters to the peak and trough of the stimulus. In other words, we employed the simplifying assumption of a square-wave texture modulation and an ideal shape of the second-stage filter; that is, the excitatory region falls entirely within one texture region and the inhibitory surround falls entirely within the neighboring region.

To produce a dipper function we introduced a simple transform based on that proposed by Legge and Foley (1980). The transform was of the form

$$F = |r|^{\alpha}/(|r|^2 + \beta^2).$$

F is the output of the nonlinear transducer, r the input to the transducer (i.e. the output of the modeled mechanism), and α and β are constants. We used the empirical thresholds for FK's unimodal OM condition to estimate α and β. The activation of the model to pure OM modulations is, at least up to depths of modulation of several times the modulation threshold, a near-linear function of the depth of modulation when expressed in degrees. Hence, we allowed r to be proportional to the depth of modulation expressed in degrees, that is, $r = \gamma a$, where γ is an estimated scaling factor and a is the depth of modulation in degrees. The resulting values were $\alpha = 2.659$, $\beta = 0.601$, and $\gamma = 0.248$. The difference threshold in the output of the transducer required to distinguish the PPT from the PO interval was estimated to be 0.313. Model predictions for the different pedestal-test combinations were then generated using these values.

Three different scenarios were used to generate model predictions. In the first scenario, the threshold was determined by the single FRF mechanism showing the greatest difference in activation between the PO and PPT interval. In the second scenario, the threshold was determined by the difference in activation between the FRF mechanism showing the highest activation in the PO interval and the FRF mechanism showing the highest activation in the PPT interval (these mechanisms may or may not be the same). In the third scenario, the threshold was determined by the difference in activation of the single FRF mechanism that was most sensitive to the test alone (i.e. without the mask). For all three scenarios, it was assumed that in order to reach threshold the difference in activation between the PO and PPT interval had to reach a value corresponding to that estimated from FK's unimodal OM data (i.e. 0.313; see above).

---

†An important difference however between Prins and Kingdom (2002) and the modeling here is that here the density of Gabors in the FM stimuli was held constant and did not co-vary with spatial frequency.
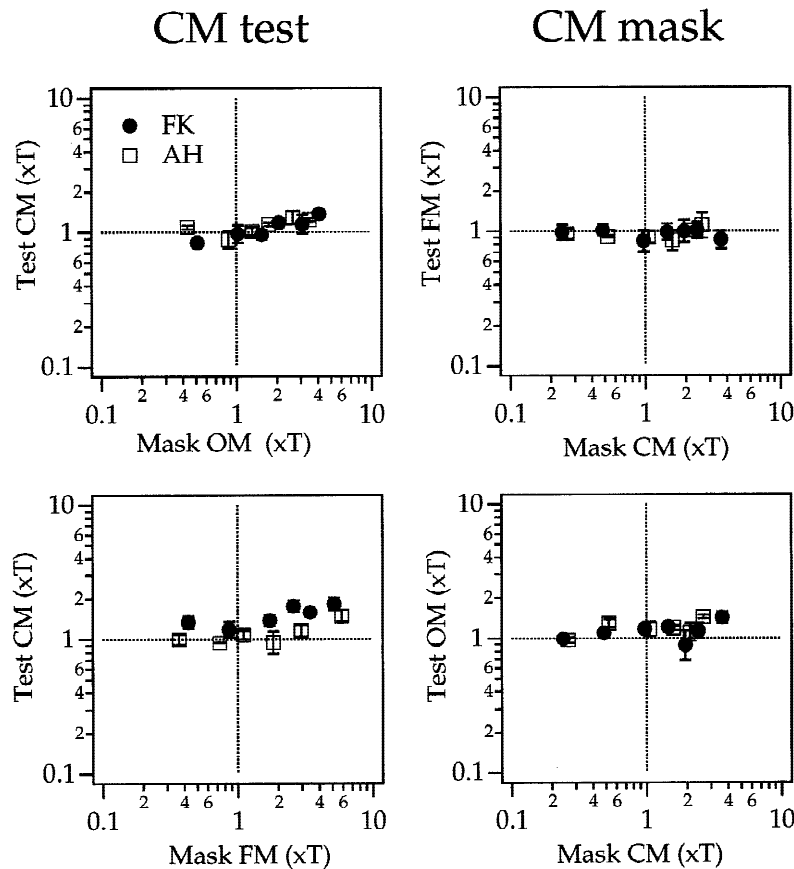
Fig. 8. Results for the CM-frequency-doubled conditions. On the left are shown the results when the test was CM and the mask either OM (top) or FM (bottom); on the right are results when the mask was CM, and the test either FM (top) or OM (bottom).

Of the three model scenarios, the one producing the best overall fit to the data, as revealed by visual inspection of the curves, was the last of these, namely the FRF mechanism most sensitive to the test alone. To avoid unnecessary detail, we will restrict our discussion to the results of this model, which we refer to as the best-FRF-for-test-alone model. One might legitimately question the plausibility of a model in which observers selected a channel most responsive to a stimulus that was only present in its pure form in one of the conditions; that is, when the mask was at zero. However, it seems to us possible that the observers learnt which channel was appropriate for each type of texture modulation, given the extensive amount of exposure they received to the three types of texture modulation during numerous practice sessions. The model predictions are shown as inset graphs alongside the data in Figs. 5 and 6. The insets show the behavior of the model for all three unimodal conditions, and a representative subset of the cross-modal conditions. The predicted dipper in the unimodal conditions is almost entirely a consequence of the accelerating nonlinearity in the mechanism's transducer function. For the cross-modal conditions, no subthreshold facilitation is predicted for the conditions where the mask is either OM or FM. To understand why, consider the condition in which the mask is FM and test OM. The FRF mechanism most sensitive to the test has a carrier tuned closely to the d.c. spatial frequency of the stimulus, but a preferred carrier orientation about 30 deg off from the orientation of the d.c. (see Prins & Kingdom, 2002). The key here is that the output of this off-orientation FRF mechanism is only slightly affected by low amplitudes of the mask. The small elevation of thresholds at low mask amplitudes is a consequence of the frequency charac-

teristics of the texture, which becomes less and less optimal as mask amplitude is increased. Eventually, the spatial frequency of the texture elements moves outside of the range of sensitivity of the carrier of the FRF mechanism, and thresholds rapidly rise. Thus, at mask amplitudes well above threshold, the model does not perform well. If the test is a CM grating in the cross-modal conditions, the lack of facilitation is a consequence of the fact that near-threshold levels of the OM or FM mask do not sufficiently activate an FRF mechanism tuned to the d.c. Gabors in the stimulus, at least as compared to the activation of the off-orientation and off-frequency FRF mechanisms responsible for mask detection.

None of the models reviewed correctly predicted the lack of facilitation we found in the cross-modal conditions where the mask was CM. We will return to the implications of this fact in the Discussion.

Discussion

Whereas in the unimodal conditions we found clear dipper functions in the subthreshold mask regions, we found none in any of the cross-modal conditions. This strongly suggests that our OM, FM, and CM textures were detected by independent mechanisms. The question that arises is whether this result can be explained in terms of a FRF model, which as we pointed out in the Introduction is ostensibly univariant with respect to second-order stimulus type. The modeling exercise showed that the pattern of psychophysical data was for the most part explicable if one supposed that the visual system selects the FRF mechanism that is optimal for the

differential response between texture regions. Only the cross-modal data using CM masks could not be explained by such a model; for these conditions the model predicted subthreshold facilitation. The importance of these findings is that it means we cannot rule out the possibility that FRF mechanisms mediated the detection of our stimuli; a different conclusion to the one that two of us initially drew from the data (Kingdom & Hayes, 2000).

We are in a position to put forward as a candidate for the detection of texture gratings a model in which the visual system selects the FRF mechanism most sensitive to the test stimulus. It is possible that this selectivity requires attention, in which case one might usefully consider the FRF channels mediating texture grating detection as steerable filters that are under attentional control. But this account appears to ignore the implications that arise from the cross-modal conditions involving CM masks, where, in all the models we ran, subthreshold facilitation was predicted. A parsimonious explanation is that a subset of FRF channels have their inputs contrast normalized (Heeger, 1991; Landy & Bergen, 1991), and that this subset mediates the detection of OM and FM gratings. Note that the putative contrast normalization mechanism would not be necessary to discount the inherent variation in the contrast *power* of our FM gratings that was noted earlier (because the density of Gabors was fixed throughout the FM waveform, the higher spatial frequency part of the waveform had less contrast power than the lower spatial frequency part of the waveform). For FRF mechanisms whose center and surround were fed by the same first-stage filters with the same input density, as assumed here, the variation in contrast power in the FM grating would not in itself produce a response.

Fig. 9 illustrates the hypothetical FRF units that we suggest may mediate the detection of the OM, FM, and CM gratings used in the present study, along with schematic descriptions of the stimuli. Note how the carrier inputs for the OM and FM detectors have preferred orientations and spatial frequencies that are very different from those designated in the stimulus. Note also that for these detectors there is a stage of contrast normalization instantiated by a process of local divisive inhibition.
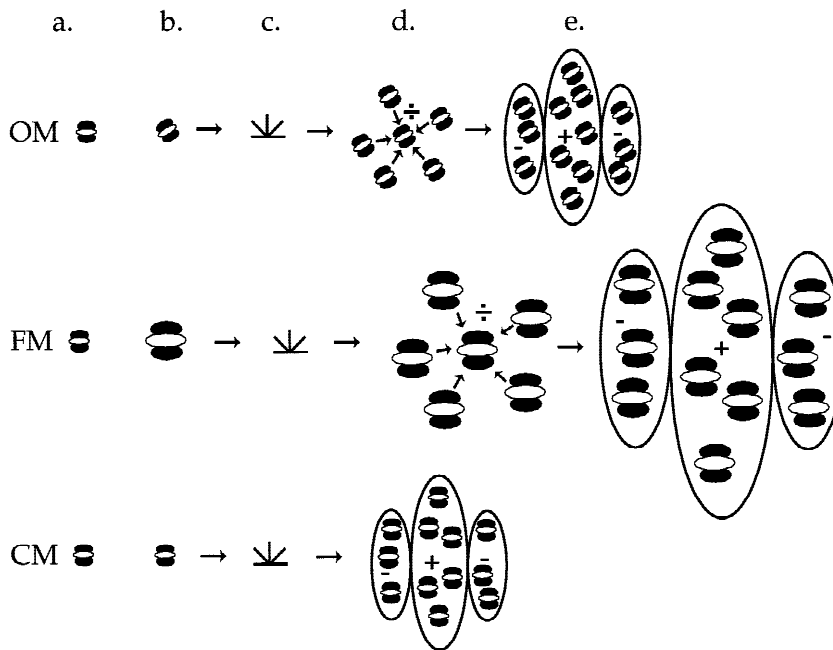
Our findings reinforce the idea that the human visual system preserves the signals from FRF mechanisms with different carriers at least up to the stage of texture detection. Later stages may well combine the signals from different FRF carriers, especially for making judgements about suprathreshold textural discontinuities in which more than one type of texture type is present (e.g. see Landy & Kojima, 2001). Our results also do not rule out the possibility that the first-stage inputs to some FRF mechanisms are organized in an orientation-opponent manner, as proposed by Landy and Bergen (1991).

Features versus energy?

Although the FRF channels illustrated in Fig. 9 represent the most parsimonious explanation of our results, our data do not provide direct proof of them. An alternative to the FRF explanation is one where the visual system provides some explication of the local feature content of textures prior to the detection of any spatial variations therein. This is not a new idea—it was the basis of thinking behind Julesz's (Julesz, 1975, 1981; Julesz & Chang, 1979; Julesz et al., 1973) and Beck's (Beck, 1966, 1982; Beck et al., 1986) early ideas on texture segmentation, and is central to Marr's (1982) ideas on the primal sketch. Given Marr's assumptions, local parts of the image might be initially labeled according to their feature content; that is, their orientation, size, luminance polarity, color, contrast, etc. The nonlinear mechanisms that are necessary to make such feature properties explicit are by no means understood, but they almost certainly require some form of comparison between the outputs of subsets of spatio-temporally tuned first-stage filters. For example, if the feature of interest was orientation, the relative outputs of the range of filters tuned to different orientations could be used. The second-stage filters would then detect spatial variations in *feature* content. Although a feature-based account cannot be ruled out, it is however the less parsimonious explanation. Moreover, it accords less well with Marr's principle of "least commitment," in that the results can be ex-

Fig. 9. Model of texture-modulation detection consistent with the results. For the top figure: (a) type of texture modulation (OM) and d.c. stimulus Gabor; (b) first-stage input filters of FRF mechanism most sensitive to OM; (c) rectification; (d) contrast normalization; and (e) second-stage filter. Middle and bottom figures follow the same sequence except there is no contrast-normalization stage for CM. For the OM test gratings (top), the most active FRF mechanism has first-stage inputs that are 30 deg off horizontal, and as a result will only very weakly respond to FM masks with horizontal d.c. Gabors. For the FM test gratings (middle), the most active FRF mechanism has first-stage inputs an octave below that of the d.c Gabor, and will only very weakly respond to low-amplitude OM masks. Insensitivity of both the OM- and FM-sensitive mechanisms to the CM masks is achieved by contrast normalization. For the CM test gratings (bottom), the most active FRF mechanism will have first-stage inputs tuned to the d.c Gabor. These will be largely unaffected by OM and FM masks.

plained without the need to invoke a stage of processing which makes explicit the feature content of the stimulus.

Why mechanism independence?

The mechanism independence revealed by our data suggests that even at threshold amplitudes we should be able to identify the type of texture modulation present, and this has been recently confirmed (Prins & Kingdom, 2003). Why would the visual system want to employ feature-specific mechanisms to detect modulated textures rather than general purpose salience detectors? While perhaps useful for the rapid segmentation of textures, or moving textured objects, a general purpose salience detector is inappropriate for providing information about the shapes of textured surfaces, for this depends not only on "where" the variation takes place, but on "what" it is that is varying. Gradients in certain features, such as orientation, appear to be particularly reliable cues to surface shape (Knill, 1998, 2001), and it is therefore no accident that orientation gradients turn out to be a more powerful perceptual cue to surface shape than density, or size (Li & Zaidi, 2000).

References

ALBRIGHT, T.D. (1995). Form-cue invariant motion processing in primate visual cortex. *Science* **255**, 1141–1143.

ARSENAULT, S.A., WILKINSON, F. & KINGDOM, F.A.A. (1999). Modulation frequency and orientation tuning of second-order texture mechanisms. *Journal of the Optical Society of America A* **16**, 427–435.

BAKER, C.L., JR. (1999). Central neural mechanisms for detecting second-order motion. *Current Opinion in Neurobiology* **9**, 461–466.

BECK, J. (1966). Effect of orientation and shape similarity on perceptual grouping. *Perception and Psychophysics* **1**, 300–302

BECK, J. (1982). Textural segmentation. In *Organization and Representation in Perception*, ed. BECK, J., pp. 285–317. London: Lawrence Erlbaum.

BECK, J., SUTTER, A. & IVRY, R. (1986). Spatial frequency channels and perceptual grouping in texture segregation. *Computer Vision, Graphics and Image Processing* **37**, 299–325.

BERGEN, J.R. & LANDY, M.S. (1991). Computational modelling of visual texture segregation. In *Computational Models of Visual Processing*, ed. LANDY, M.S. & Movshon, J.A., pp. 253–271. Cambridge, Massachusetts: The MIT Press.

BURR, D.C., MORRONE, M.C. & SPINELLI, D. (1989). Evidence for edge and bar detectors in human vision. *Vision Research* **29**, 419–431.

CAMPBELL, F.W. & KULIKOWSKI, J.J. (1966). Orientation selectivity of the human visual system. *Journal of Physiology* (London) **187**, 437–445.

CAVANAGH, P. & MATHER, G. (1989). Motion: The long and short of it. *Spatial Vision* **4**, 103–129.

CHAUDHURI, A. & ALBRIGHT, T.D. (1997). Neuronal responses to edges defined by luminance vs. temporal texture in macaque area V1. *Visual Neuroscience* **14**, 949–962.

CHUBB, C. & SPERLING, G. (1988). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A* **5**, 1986–2007.

CHUBB, C. & LANDY, M.S. (1991). Orthogonal distribution analysis: A new approach to the study of texture perception. In *Computational Models of Visual Processing*, ed. LANDY, M.S. & MOVSHON, J.A., pp. 291–301. Cambridge, Massachusetts: The MIT Press.

DAKIN, S.C. & MARESCHAL, I. (2000). Sensitivity to contrast modulation depends on carrier spatial frequency and orientation. *Vision Research* **40**, 311–329.

DERRINGTON, A.M. & UKKONEN, O.I. (1999). Second-order motion discrimination by feature-tracking. *Vision Research* **36**, 1907–1918.

DOBBINS, A., ZUCKER, S.W. & CYNADER, M.S. (1989). Endstopping and curvature. *Vision Research* **29**, 1371–1387.

FOLEY, J.M. (1994). Human luminance pattern-vision mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America A* **11**, 1710–1719.

FOLEY, J.M. & LEGGE, G.E. (1981). Contrast detection and near-threshold discrimination in human vision. *Vision Research* **21**, 1041–1053.

GEESAMAN, B.J. & ANDERSON, R.A. (1996). The analysis of complex motion patterns by form/cue invariant MSTd neurons. *Journal of Neuroscience* **16**, 4716–4732.

GEISLER, W.S., PERRY, J.S., SUPER, B.J. & GALLOGLY, D.P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research* **41**, 711–724.

GRAHAM, N. & SUTTER, A. (1998). Spatial summation in simple (Fourier) and complex (no-Fourier) texture channels. *Vision Research* **38**, 231–257.

GRAHAM, N., BECK, J. & SUTTER, A. (1992). Nonlinear processes in spatial-frequency channel models of perceived segregation: Effects of sign and amount of contrast. *Vision Research* **32**, 719–743.

GROSOF, D.H., SHAPLEY, R.M. & HAWKEN, M.J. (1993). Macaque V1 neurons can signal 'illusory' contours. *Nature* **365**, 550–552.

HEEGER, D.J. (1991). Nonlinear model of neural responses in cat visual cortex. In *Computational Models of Visual Processing*, ed. LANDY, M.S. & MOVSHON, A.J., pp. 119–133. Cambridge, Massachusetts: The MIT Press.

JULESZ, B. (1975). Experiments in the visual perception of texture. *Scientific American* **232**, 34–43.

JULESZ, B. (1981). Textons, the elements of texture perception and their interactions. *Nature* **290**, 91–97.

JULESZ, B. & CHANG, J. (1979). Symmetry perception and spatial-frequency channels. *Perception* **8**, 711–718.

JULESZ, B., FRISCH, H.L., GILBERT, E.N. & SHEPP, L.A. (1973). Inability of humans to discriminate between textures that agree in second-order statistics—revisited. *Perception* **2**, 391–405.

KINGDOM, F.A.A. & KEEBLE, D.R.T. (1996). A linear systems approach to the detection of both abrupt and smooth spatial variations in orientation-defined textures. *Vision Research* **36**, 409–420.

KINGDOM, F.A.A. & HAYES, A. (2000). Mechanism independence for texture-property modulation. *Investigative Ophthalmology and Visual Science* (Suppl.), **40**, 317.

KINGDOM, F.A.A. & KEEBLE, D.R.T. (2000). Luminance spatial frequency differences facilitate the segmentation of superimposed textures. *Vision Research* **40**, 1077–1087.

KINGDOM, F.A.A., KEEBLE, D. & MOULDEN, B. (1995). Sensitivity to orientation modulation in micropattern-based textures. *Vision Research* **35**, 79–91.

KNILL, D.C. (1998). Discrimination of planar surface slant from texture: Human and ideal observers compared. *Vision Research* **38**, 1683–1711.

KNILL, D.C. (2001). Contour into texture: Information content of surface contours and texture flow. *Journal of the Optical Society of America A* **18**, 12–35.

LANDY, M.S. & BERGEN, J.R. (1991). Texture segregation and orientation gradient. *Vision Research* **31**, 679–691.

LANDY, M.S. & KOJIMA, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America A* **18**, 2307–2320.

LANDY, M.S. & ORUC, I. (2002). Properties of second-order spatial frequency channels. *Vision Research* **42**, 2311–2329.

LEDGEWAY, T. & SMITH, A.T. (1994). Evidence for separate motion-detecting mechanisms for first- and second-order motion in human vision. *Vision Research* **34**, 2727–2740.

LEGGE, G.E. & FOLEY, J.M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America* **70**, 1458–1471.

LEGGE, G.E. & KERSTEN, D. (1983). Light and dark bars; contrast discrimination. *Vision Research* **23**, 473–483.

LENNIE, P. (1998). Single units and visual cortical organization. *Perception* **27**, 889–935.

LEVENTHAL, A.G., WANG, Y., SCHMOLESKY, M.T. & ZHOU, Y. (1998). Neural correlates of boundary perception. *Visual Neuroscience* **15**, 1107–1118.

LI, A. & ZAIDI, Q. (2000). Perception of three-dimensional shape from texture is based on patterns of local orientation. *Vision Research* **40**, 217–242.

MALIK, J. & PERONA, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A* **7**, 923–932.

MARESCHAL, I. & BAKER, C.L., JR. (1998). A cortical locus for the processing of contrast-defined contours. *Nature Neuroscience* **1**, 150–154.

MARESCHAL, I. & BAKER, C.L., JR. (1999). Cortical processing of second-order motion. *Visual Neuroscience* **16**, 1–14.

MARR, D. (1982). *Vision*. New York: W.H. Freeman and Company.

MULLEN K.T. & LOSADA M.A. (1994). Evidence for separate pathways for color and luminance detection mechanisms. *Journal of the Optical Society of America A* **11**, 3136–3151.

NISHIDA, S., LEDGEWAY, T. & EDWARDS, M. (1997). Dual multiple-scale processing for motion in the human visual system. *Vision Research* **37**, 2685–2698.

O'KEEFE, L.P. & MOVSHON, J.A. (1998). Processing of first- and second-order motion signals by neurons in area MT of the macaque monkey. *Visual Neuroscience* **15**, 305–317.

OLAVARRIA, J.F., DEYOE, E.A., KNIERIM, J.J., FOX, J.M. & VAN ESSEN, D.C. (1992). Neural responses to visual texture patterns in middle temporal area of the macaque monkey. *Journal of Neuroscience* **68**, 164–181.

PELLI, D.G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A* **2**, 1508–1532.

PETERHANS, E. & VON DER HEYDT, R. (1991). Subjective contours—bridging the gap between psychophysics and physiology. *Trends in Neurosciences* **14**, 112–119.

PRINS, N. & MUSSAP, A.J. (2001). Adaptation reveals a neural code for the visual location of orientation change. *Perception* **30**, 669–680.

PRINS, N. & KINGDOM, F.A.A. (2002). Orientation- and frequency-modulated textures at low depths of modulation are processed by off-orientation and off-frequency texture mechanisms. *Vision Research* **42**, 705–713.

PRINS, N. & KINGDOM, F.A.A. (2003). Detection and discrimination of texture modulations defined by orientation, frequency and contrast. *Journal of the Optical Society of America A* **20**, 401–410.

ROSS, J. & SPEED, H.D. (1991). Contrast adaptation and contrast masking in human vision. *Proceedings of the Royal Society B* (London) **246**, 61–69.

SARY, G., VOGELS, R., KOVACS, G. & ORBAN, G.A. (1995). Responses of monkey inferior temporal neurons to luminance-, motion-, and texture-defined gratings. *Journal of Neurophysiology* **73**, 1341–1354.

SCHOFIELD, A.J. (2000). What does second-order vision see in an image? *Perception* **29**, 1071–1086.

SCHOFIELD, A.J. & GEORGESON, M.A. (1999). Sensitivity to modulations of luminance and contrast in visual white noise: Separate mechanisms with similar behaviour. *Vision Research* **39**, 2697–2716.

SCHILLER, P.H. (1982). Central connections of the retinal ON and OFF pathways, *Nature* **297**, 580–583.

SMITH, A.T., GREENLEE, M.W., SINGH, K.D., KRAEMER, F.M. & HENNIG, J. (1998). The processing of first- and second-order motion in human visual cortex assessed by functional magnetic resonance imaging (fMIR). *Journal of Neuroscience* **18**, 3816–3830.

SUTTER, A. & GRAHAM, N. (1995). Investigating simple and complex mechanisms in texture segregation using speed-accuracy trade-off method. *Vision Research* **35**, 2825–2843.

SUTTER, A., SPERLING, G. & CHUBB, C. (1995). Measuring the spatial frequency selectivity of second-order texture mechanisms. *Vision Research* **35**, 915–924.

VAINA, L.M. & COWEY, A. (1996). Impairment of the perception of second order motion but not first order motion in a patient with unilateral focal brain damage. *Proceedings of the Royal Society B* (London) **263**, 1225–1232.

VAINA, L.M., MAKRIS, N., KENNEDY, D. & COWEY, A. (1998). The selective impairment of the perception of first-order motion by unilateral cortical brain damage. *Visual Neuroscience* **15**, 333–348.

WENDEROTH, P. WATSON, J.D., EGAN, G.F., TOCHON-DANGUY, H.J. & O'KEEFE, G.J. (1999). Second order components of moving plaids activate extrastriate cortex: A positron emission tomography study. *Neuroimage* **9**, 227–234.

WILSON, H. (1999). Non-Fourier cortical processes in texture, form, and motion perception. *Cerebral Cortex* **13**, 445–477.

WILSON, H.R. & RICHARDS, W.A. (1992). Mechanisms of contour curvature discrimination. *Journal of the Optical Society of America A* **6**, 106–115.

WILSON, H.R. & WILKINSON, F. (1998). Detection of global structure in glass patterns: Implications for form vision. *Vision Research* **38**, 2933–2947.

WILSON, H.R., FERRERA, V.P. & YO, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience* **9**, 79–97.

WILSON H.R. & WILKINSON, F. & ASAAD, W. (1997). Concentric orientation summation in human form vision. *Vision Research* **37**, 2325–2330.

ZHOU, Y.X. & BAKER, C.L., JR. (1994). Envelope-responsive neurons in areas 17 and 18 of cat. *Journal of Neurophysiology* **72**, 2134–250.

ZHOU, Y.X. & BAKER, C.L., JR. (1996). Spatial properties of envelope-responsive cells in area 17 and 18 neurons of the cat. *Journal of Neurophysiology* **75**, 1038–1050.