



Properties of second-order spatial frequency channels

Michael S. Landy *, İpek Oruç

Department of Psychology and Center for Neural Science, New York University, 6 Washington Place, 8th floor, New York, NY 10003, USA

Received 14 March 2002; received in revised form 11 June 2002

Abstract

The segregation of texture patterns may be carried out by a set of linear spatial filters (to enhance one of the constituent textures), a nonlinearity (to convert the higher contrast of response to that constituent to a higher mean response), and finally subsequent (“second-order”) linear spatial filters (to provide a strong response to the texture-defined edge itself). In this paper, the properties of such second-order filters are characterized. Observers were required to detect or discriminate textures that were modulated between predominantly horizontally oriented and predominantly vertically oriented noise patterns. Spatial summation for these patterns reached asymptote for a stimulus size of 15×15 deg. Modulation contrast sensitivity was nearly flat over a five-octave range of spatial frequency, but was bandpass when stated as efficiency (relative to an idealized observer confronted with the same task). Increment threshold showed the improved performance with a sub-threshold pedestal seen in the “dipper effect”, but the typical Weber’s law behavior at higher pedestal contrasts was not observed at the highest pedestal modulation contrasts achievable with our stimuli. Sub-threshold summation experiments indicate that second-order filters have a moderate bandwidth.

© 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Spatial vision; Texture perception; Second-order processing; Contrast sensitivity

1. Introduction

The dominant view of spatial pattern detection and discrimination is that the visual system consists of a set of channels selective for a range of spatial frequency, temporal frequency, orientation, direction, color and so on. These selectivities are defined primarily by the linear filtering properties of the channel, but there is evidence for nonlinear channel processing as well. More recently, there has been interest in linear filtering that takes place following a nonlinear stage, which is known as “second-order” processing. In this paper, the spatial filtering properties of these second-order filters are investigated using techniques analogous to those originally used to define first-order spatial frequency channels. We begin by summarizing the kinds of studies used to investigate first-order channel properties as well as studies that have investigated the characteristics of second-order processing in vision.

1.1. First-order spatial frequency channels

Spatial frequency channels have been studied extensively in the last 30 or so years (De Valois & De Valois, 1990; Graham, 1989). Contrast sensitivity is bandpass in spatial frequency (for low temporal frequencies), but this contrast sensitivity function (CSF) is viewed as the envelope of a number of narrow-band, spatial-frequency-tuned mechanisms (Campbell & Robson, 1968).

The spatial frequency tuning of these underlying mechanisms or *channels* has been examined using several methods.

- *Adaptation:* When the observer is adapted to a particular spatial frequency grating, sensitivity is reduced for gratings that are close in spatial frequency, but not for those that are far removed in spatial frequency (Blakemore & Campbell, 1969; Pantle & Sekuler, 1968), indicating that a narrowband mechanism (or mechanisms) was desensitized by the adapter. In addition, the apparent size of gratings that are nearby in spatial frequency is altered (Blakemore & Sutton, 1969), suggesting that the response profile across narrowband channels has been altered by adaptation.

* Corresponding author. Tel.: +1-212-998-7857; fax: +1-212-995-4349.

E-mail address: landy@nyu.edu (M.S. Landy).

URL: <http://www.cns.nyu.edu/~msl>.

- *Summation*: When threshold contrast is determined for compound stimuli (sums of two or more gratings), a greater degree of summation is found for gratings similar in spatial frequency than for those with more distant spatial frequencies (Sachs, Nachmias, & Robson, 1971). This has been taken to indicate that similar gratings are summed within the same channel (that is, linearly combined by receptive fields sensitive to both summands), whereas dissimilar gratings only combine by *probability summation* across distinct analyzers (Campbell & Robson, 1968; Graham, 1980; Graham & Nachmias, 1971).
- *Masking*: Channel tuning has also been estimated using masking experiments, where it has been found that a narrowband masker will increase threshold for gratings similar in spatial frequency to the masker, but not for those with a spatial frequency far removed from that of the masker. This has been carried out with narrowband noise maskers (Henning, Hertz, & Hinton, 1977; Stromeyer III & Julesz, 1972) as well as gratings and other patterned maskers (Legge & Foley, 1980; Wilson, McFarlane, & Phillips, 1983).

Additional methods for estimating channel tuning (reviewed by Graham (1989)) include the manipulation of stimulus uncertainty and requiring subjects to detect as well as to identify the stimulus presented on each trial. The methods we have reviewed have not always provided the same estimates of channel bandwidth. Such inconsistencies may indicate a failure to note the degree to which probability summation across channels and across space contributes to the results (Graham, 1980).

When the observer's task is to discriminate two patterns differing in contrast, the *increment threshold* (the increment in contrast added to a pedestal to achieve threshold) at first decreases with increasing pedestal contrast, then begins to increase (the so-called "dipper function", Legge & Foley, 1980; Nachmias & Sansbury, 1974). Note that at higher contrasts the dipper function approximates a power function, but not one with a power equal to one (that is, Weber's law is typically not observed, Legge, 1981). There are two typical explanations for the shape of the dipper function. The first attributes the dipper function's shape to a static nonlinearity that follows the initial linear, spatial filter (Foley & Legge, 1981; Legge & Foley, 1980; Nachmias & Sansbury, 1974; Wilson, 1980). The initial fall in increment threshold is attributed to an accelerating contrast nonlinearity near threshold, and the later rise to a decelerating (or saturating) nonlinearity. An alternative explanation is that the shape of the contrast increment threshold function results from observer uncertainty (Pelli, 1985). Uncertainty about the signal causes the observer to monitor both relevant channels (those that are sensitive to the test pattern) and irrelevant channels (those that are completely insensitive to the test pattern).

At low contrasts, the irrelevant channels contribute noise that limits detection; at higher contrasts, the contribution of these irrelevant channels is increasingly dwarfed by the larger responses from the relevant channels.

1.2. Second-order processes in vision

In analyzing results from "pre-attentive" texture segregation experiments, many have tried to explain observer performance based on the pattern of outputs of spatial frequency channels to the textures to be discriminated (Landy & Graham, in press). A typical model begins with a set of linear, spatial filters, followed by a nonlinearity (so that regions with large response variability are mapped to large average response), followed by a second-order, linear, spatial filter (to enhance boundaries between regions with different average response strength). This linear–nonlinear–linear (LNL) model is referred to by Chubb and colleagues (Chubb, Econopouly, & Landy, 1994; Chubb & Landy, 1991) as the "back-pocket model" of texture segregation (Fig. 1), as it has become the default model that researchers in the field routinely "pull from their back pocket" to attempt to make sense of new texture segregation results. And, a large number of recent papers on texture segregation utilize this basic model structure (Bergen & Adelson, 1988; Bergen & Landy, 1991; Bovik, Clark, & Geisler, 1990; Caelli, 1985; Clark & Bovik, 1989; Fogel & Sagi, 1989; Graham, 1991, 1994; Knutsson & Granlund, 1983; Landy & Bergen, 1991; Malik & Perona, 1990; Sagi, 1990; Sutter, Beck, & Graham, 1989; Turner, 1986). In this paper, we will consider only the

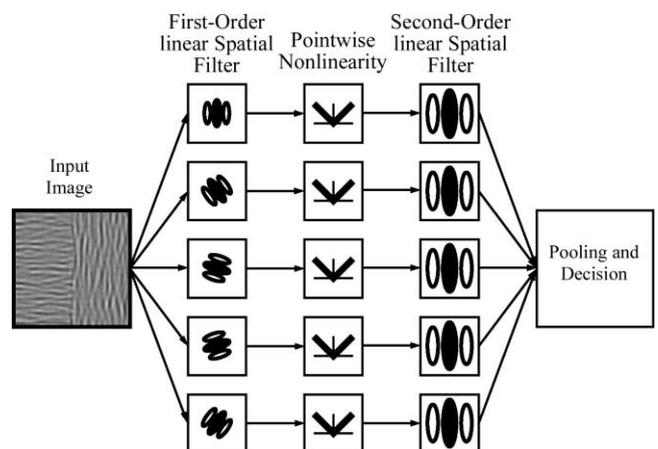


Fig. 1. The back-pocket model of texture segregation. The input image is passed through a set of linear spatial filters. A nonlinearity is applied to the filter outputs (here sketched as a pointwise nonlinearity, although it may also involve gain control, normalizing by the pooled outputs of several channels). A second-order, linear spatial filter is then used, for example to enhance the difference in responses to neighboring texture regions. Finally, the channel outputs are combined for performance predictions for particular experiments.

simplest form of LNL mechanism, one for which the second-order linear filter pools over rectified responses of first-order linear filters that are all identical in form to one another except for spatial position.

A linear spatial filter with both excitatory and inhibitory receptive field regions (i.e., a bandpass filter) can enhance luminance structure such as lines or edges in a stimulus. Analogously, the basic LNL structure allows back-pocket models to extract more general changes in structure in a stimulus such as edges across which there is a change of image contrast, average spatial scale or orientation content. This same LNL or second-order structure has seen wide application in the modeling of visual processes (see, e.g., Chubb, Olzak, & Derrington, 2001), including spatial vision (Olzak & Thomas, 1999), motion (Chubb & Sperling, 1988; Wilson, Ferrera, & Yo, 1992) and stereopsis (Hess & Wilcox, 1994) for similar purposes.

1.3. Back-pocket model channel properties

Having demonstrated that second-order processing is an important contributor to some forms of spatial discriminations, the next step is to delineate the properties of each stage (linear, nonlinear, and second-order linear) of a channel of the back-pocket model. Is the nonlinearity static? Dynamic? A gain control mechanism? If static, what is its functional form? If gain control, what serves as the gain normalization term? Are there multiple, second-order, linear mechanisms? What is the form of the second-order receptive field (shape, size, spatial frequency and orientation bandwidth)? Is the form of the second-order linear filter related in any particular way to that of the first-order linear filter that provides its input? The current paper concentrates on questions of the form and number of second-order, linear spatial filters. We next review a number of studies that have attempted to address each of these questions.

Early nonlinearity: Graham and colleagues (Graham, 1991; Graham, Beck, & Sutter, 1992; Graham & Sutter, 1998) have explored and rejected an early nonlinearity preceding the first stage of linear, spatial filtering as inconsistent with their texture-based region segregation experiments.

First stage linear, spatial filter: Graham and colleagues (Graham, 1994; Graham, Sutter, & Venkatesan, 1993) find that the first stage filter is both orientation- and spatial frequency-tuned, although the bandwidths appear to be wider than those used for computing first-order texture-based region segregations. Dakin and Mareschal (2000) also found the first stage filters to be orientation- and spatial frequency-tuned for a task of detecting a contrast modulation of a filtered noise carrier.

Static nonlinearity: Nearly all studies invoke some form of static nonlinearity following the first stage linear filter, whether or not they also include some form of

gain control. Chubb and colleagues (Chubb et al., 1994; Chubb & Landy, 1991) were able to isolate a single, small-scale texture channel and, by varying the first-order luminance statistics of two textures to be discriminated, were able to measure the static nonlinearity of that channel. Graham and Sutter (Graham, 1994; Graham & Sutter, 1998, 2000) found that it must be an accelerating/expansive nonlinearity for their region segregation task.

Gain control: A number of studies indicate the existence of a gain control mechanism such that the output of a first stage linear filter followed by its static nonlinearity is then normalized by a pooled response from other filters (Graham, 1991; Graham et al., 1992; Graham & Sutter, 1996, 1998, 2000; Olzak & Thomas, 1996, 1999; Olzak & Wickens, 1997). Graham and Sutter (2000) suggest that this gain control is inhibition between channels rather than normalization of responses over space within a channel. Olzak and Thomas (1996, 1999) have studied a variety of spatial pattern discriminations, resulting in a model that includes two forms of stereotyped normalization and pooling, either across spatial frequency (for finer orientation judgments) or across orientation (for improved spatial frequency discrimination).

Second-order modulation contrast sensitivity: Jamar and Koenderink (1985) measured the sensitivity to contrast modulations of white and bandpass noise gratings (that is, one-dimensional patterns) and found this form of second-order modulation contrast sensitivity to be modestly lowpass in form. Sutter, Sperling, and Chubb (1995) measured the second-order modulation CSFs for the detection of contrast modulations of isotropic, bandpass-filtered carrier noise patterns. They found a modulation CSF that was nearly flat, although it showed a modest bandpass shape (for a similar result, see also Schofield & Georgeson, 1999). Kingdom, Keeble, and Moulden (1995) measured a modulation CSF for the detection of a sinusoidal modulation of orientation across a texture. Their modulation CSF was also quite flat over a large range, although with more of a tendency to be lowpass.

Second-order channel bandwidth: Several studies have examined whether the second-order modulation CSF is an envelope over multiple, underlying, second-order mechanisms, as in the first-order case. Kwan and Regan (1998) used orientation-modulated texture and an adaptation paradigm. Adaptation to second-order texture modulation increased thresholds, and the effect was orientation-tuned, indicating that second-order channels were orientation-tuned. Arsenault, Wilkinson, and Kingdom (1999) used a second-order masking paradigm with spatial frequency-modulated noise textures. They found the masking effect to be spatial frequency- and orientation-tuned. But, unlike other studies cited here, they found no evidence of orientation tuning for the

carrier. However, this lack of orientation tuning for the carrier is also consistent with a model with multiple mechanisms, each of which has an orientation-tuned first-order linear filter, but which as a group span the range of possible orientations (rather than a single mechanism with an initial linear filter that is not orientation-tuned). Kingdom and Keeble (1996) compared orientation modulation detection for sine wave, square wave, and missing fundamental modulators. In contrast to the other studies, their results were consistent with the use of only a single, second-order channel. Finally, the region segregation tasks used by Graham and Sutter (1998, 2000) require that the second-order filter be orientation-tuned.

Note that for the type of second-order mechanisms considered here (that use a single form of first-order linear filter in all locations), a frequency or orientation modulation of a carrier texture is equivalent to a contrast modulation. This is also true of the stimuli used in the current study. The authors of the above-cited studies may have had more complex mechanisms in mind, with first-order receptive fields that varied in frequency or orientation at different locations (e.g., matched filters for their stimuli). Such more complex mechanisms are probably unnecessary to explain their results.

Relationship between first and second stage filters: Many studies have found that texture discrimination is scale invariant (Dakin & Mareschal, 2000; Kingdom et al., 1995; Sutter et al., 1995). That is, if viewing distance is altered, the results do not change. Kingdom and Keeble (1999) argue that these results imply a link between the scale of the first- and second-order filters. Graham (1994) suggests the ratio of scales is about 3–4, whereas Sagi (1990) suggests a far larger ratio of 6–18, and Kingdom et al. (1995) suggest a ratio of 40–50 using orientation-modulated textures. Several studies suggest the preferred orientations of the first and second stage filters are correlated. Most studies indicate the preferred orientations of the first and second stage filters tend to be aligned, although perpendicular orientations may be favored as well (Dakin & Mareschal, 2000; Dakin, Williams, & Hess, 1999; Graham & Wolfson, 2001; Wolfson & Landy, 1995).

Finally, a number of recent studies involve tasks that require observers to analyze the local orientation structure of a texture per se, rather than using orientation to examine texture segregation (Dakin et al., 1999; Kingdom & Keeble, 2000; Keeble, Kingdom, & Morgan, 1997). Some of these studies claim their results are incompatible with the LNL framework (Dakin et al., 1999; Kingdom & Keeble, 2000).

1.4. *Motivation and preview*

In this paper, we describe experiments intended to elucidate the properties of the second-order, linear,

spatial filters underlying texture segregation and texture modulation detection. Our study differs from those that precede it in a number of ways. First, we use a different type of stimulus. Second, we have used a number of new experimental manipulations. Third, we have compared our results to simulations of idealized back-pocket models using the same stimuli.

To investigate the form and number of first-order filters, the experiments summarized earlier used sine wave gratings and other simple patterns as the stimuli for detection and discrimination experiments. Several investigators, including ourselves, have had the idea to investigate the second-order filters by constructing a stimulus that, after processing by the first-order filters and subsequent nonlinearity, delivers a sine wave grating (or other simple pattern) as the input to the second-order linear spatial filters. If this were possible, then one could study these filters using the same techniques summarized in Section 1.1. To accomplish this, several researchers have used the intended pattern as a “modulator”, and have assumed that the first two stages of processing successfully demodulate the stimulus. For example, Sutter et al. (1995) use Gabor patches as modulators of a bandpass, isotropic noise carrier. This is modulation in the traditional engineering sense: the modulator waveform is multiplied by the carrier waveform to produce the stimulus (this multiplication is performed in the contrast domain). Kingdom et al. (1995) use a different form of modulation entirely. Their modulation patterns (typically sine wave gratings) are used to modulate the average orientation of a texture consisting of line segments or small Gabor patches.

We are specifically interested in the sort of second-order mechanisms that are required for the analysis of purely texture-defined patterns as in, e.g., studies of texture segregation. As such, it was also our purpose to choose stimuli that would be relatively easy to analyze theoretically but that, at the same time, could not be detected using mechanisms designed purely for the analysis of local contrast. Local contrast can be estimated using second-order mechanisms with isotropic first-stage linear filtering. On the other hand, many stimuli used to study texture segregation involve no change in local contrast across a texture-defined boundary, and oriented, first-order linear filters are required to detect such edges. There is a large literature on the detection of contrast modulations (some of which is reviewed above), and Dakin and Mareschal (2000) suggested that it would be simpler if identical mechanisms were used for such tasks as well as for texture tasks. Their evidence of orientation tuning for the carrier in contrast modulation detection is consistent with this idea. However, to ensure that our tasks probed the types of mechanisms used for texture segregation, we eschewed the use of contrast modulation in favor of a purely textural modulation.

Our choice of stimuli was inspired by the work of Watson and Eckert (1994) concerning second-order mechanisms of motion analysis. In their study, a vertical, sinusoidal modulator was used to mix two different isotropically filtered, bandpass noise textures: one moving upward, and a second moving downward. At the peaks of the modulator, the stimulus would have a majority of upward-moving energy, and at the valleys, the primary component was downward. At the zero crossings, the stimulus was simply the average of the two components, with equal upward and downward motion energy. Modulation contrast sensitivity was found to be bandpass and scale invariant, and a second-order filter was estimated. We have developed an analogous stimulus for the modulation of static texture (Fig. 2). The modulator pattern is used to modulate between a ver-

tically and a horizontally oriented noise pattern. Later, we argue that this stimulus is substantially easier to relate to current models than some others that have been used.

We used these second-order texture-modulated stimuli in a series of experiments analogous to several classic experiments used to understand first-order, spatial frequency channels. These include estimates of spatial pooling, modulation contrast sensitivity and channel bandwidth, as well as a measurement of the increment threshold curve. In addition, we compared the performance of the human observers with that of an idealized observer (essentially, a simulation of a back-pocket model performing the same task with the same stimuli), which required us to revise our conclusions concerning second-order modulation contrast sensitivity.

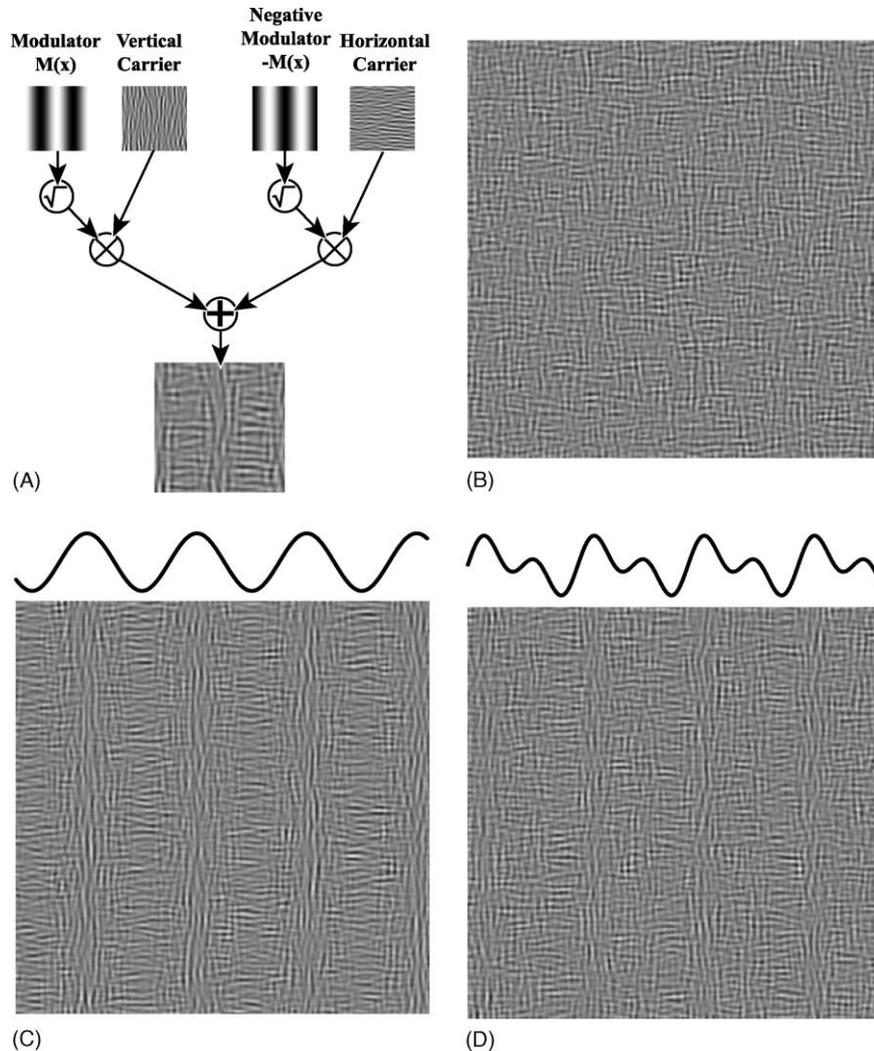


Fig. 2. Example stimuli. (A) Stimuli were constructed from vertically and horizontally oriented filtered noise textures (the “carrier”), combined as directed by a lower frequency modulator. (B) Stimulus based on a zero-amplitude modulator. The texture is a uniform plaid throughout. (C) Stimulus based on a 100% contrast sine wave modulator. (D) Stimulus based on a 100% contrast modulator that is the sum of two sine waves. The modulator is shown above the stimuli in (C) and (D).

2. Methods

2.1. Stimuli

Stimuli were constructed from two filtered noise textures, one horizontally and one vertically oriented. These textures constituted the *carrier* waveform. A lower frequency *modulator* waveform was used, effectively, to vary between the two noise carriers (Fig. 2A).

The carrier textures were generated from white noise by applying a linear filter in the Fourier transform domain, followed by the inverse Fourier transform. The stimuli were viewed from several viewing distances, which shifts the range of spatial frequencies. For the moment, we will describe the stimuli as viewed from the nearest viewing distance, which was used in most of the experiments. In spatial terms, the linear filter was a Gabor function in cosine phase, resulting in a pair of real, Gaussian distributions in the Fourier domain, centered on ± 4 cpd. The Gaussians were separable functions of f_x (horizontal spatial frequency) and f_y (vertical spatial frequency), with standard deviations adjusted so that the resulting noise had a full bandwidth at half-height of one octave and 30° . Each stimulus was constructed using a vertically oriented noise image N_V and a horizontally oriented noise image N_H .

Each stimulus was computed by combining N_V and N_H with a modulator function (Fig. 2A). All modulators $M(x)$ were vertically oriented sine wave gratings or sums of gratings. Following Watson and Eckert (1994), we treated the noise images as contrast images (i.e., mean luminance equals a value of zero), and a stimulus L was defined as

$$m_1(x) = \left[\frac{1}{2}(1 + cM(x)) \right]^{1/2}$$

$$m_2(x) = \left[\frac{1}{2}(1 - cM(x)) \right]^{1/2} = (1 - m_1^2(x))^{1/2} \quad (1)$$

$$L(x, y) = L_0[1 + A(m_1(x)N_V(x, y) + m_2(x)N_H(x, y))],$$

where c was the modulation contrast (a number between 0 and 1 that was varied across trials to determine modulation detection thresholds), A was a fixed amplitude set so that stimulus peak contrast used the full range of available luminance values (a very small number of pixel values were clipped at 0 or 255), and L_0 was the mean luminance of the display. The modulating functions m_i ranged from zero to one. Example stimuli are shown in Fig. 2B–D. As you can see in Fig. 2C, at the peaks of the modulator $M(x)$, the stimulus consisted primarily of vertically oriented noise, and at the valleys, the stimulus was mostly horizontal.

As stated above, we sought stimulus modulations that would be invisible to mechanisms that merely code local contrast and its variation across the stimulus. The definition of “contrast” appropriate for such stimuli is a matter of some contention, but the most widely em-

ployed such measure for noise stimuli is contrast energy. The square root in Eq. (1) ensured that expected contrast energy did not vary across the stimulus as the weighted sum of the two independent, random noise images was computed. However, at high modulation contrasts the square root does result in distortion of the modulator.

Contrast modulation can result in low-frequency distortion products visible to first-order mechanisms. To minimize such distortion, the modulator frequency was never higher than one-half of the carrier noise peak spatial frequency. However, as stated above, the square root operation in Eq. (1) results in higher frequency distortion products of the modulator, which can result in additional first-order stimulus artifacts. We examined this possibility by comparing phase-randomized versions of our modulated stimuli (e.g. a phase randomized version of Fig. 2C) with our unmodulated stimuli (Fig. 2B). Visible difference between them can only be attributed to first-order artifacts, as the phase randomization destroys the texture modulation structure. For a carrier spatial frequency of 4 cpd and a modulator spatial frequency of 2 cpd, these artifacts were visible for modulation contrasts as low as 0.3–0.4. For a 1 cpd modulator, these artifacts were barely, if at all visible at modulation contrasts of 0.8–1.0, and were never visible for a 0.5 cpd modulator. The visible artifacts corresponded to stimuli whose power spectra were, in fact, discriminable from those of our unmodulated stimuli. In general, these artifacts are visible at modulation contrast levels that are too high to account for our results, much as was found by Dakin and Mareschal (2000), who ran a control experiment using phase-randomized stimuli to check for such artifacts.

Stimuli were computed using the HIPS image processing software (Landy, Cohen, & Sperling, 1984). For each condition, 20 stimuli were computed for a modulation contrast $c = 0$, and 10 stimuli were computed for each of 12 nonzero modulation contrast levels in equal logarithmic steps from 0.08 to 1.0. Modulator spatial phase was chosen randomly for each stimulus.

Stimuli were displayed using a computer equipped with a CRS VSG 2/3 frame buffer. For most experiments, stimuli were 500×500 pixels. The experiments were carried out over several years, using three different monitors. In almost all cases, viewing distance was adjusted so that those 500×500 pixels subtended 15×15 deg. The setups are summarized in Table 1. All displays ran at or above 60 Hz, and used linearized lookup tables. For the spatial pooling experiments, other stimulus sizes were used. For the modulation contrast sensitivity experiments used to examine scale invariance (the first monitor setup), experiments were also carried out at distances of 78 and 157 cm so as to increase the stimulus spatial frequencies by factors of 2 and 4, and similarly reduce the retinal image size. Mean luminance was

Table 1
Viewing conditions for each setup

Monitor	Resolution	Viewing distance(s) (cm)	Subjects/Tasks
Mitsubishi Diamond Scan HL6605	1312 × 997	38, 78 and 157	MSL & KNW; ELA (modulation contrast sensitivity)
Mitsubishi Diamond Scan HL6605	832 × 590	62	IO (modulation contrast sensitivity)
Dell Trinitron D1025TM	800 × 600	68	ELA (increment threshold)
Nanao Flexscan 9070U	800 × 591	68	IO (increment threshold, summation), ELA (summation)

comparable in all setups; it was 25 cd/m² for the Mitsubishi and 35 cd/m² for the Dell monitor.

2.2. Procedure

The subject's task was either to detect the modulation or to discriminate its modulation contrast using a 2-interval, 2-alternative forced choice procedure. Trials were blocked by condition, so that a single modulator $M(x)$ was used in a given block. Each trial consisted of a 500 ms cue, 250 ms stimulus interval, 500 ms blank interval and 250 ms stimulus interval. In one interval, chosen randomly on each trial, the modulation contrast was a fixed value c_0 and in the other, the modulation contrast $c = c_0 + \Delta c$ was variable. The screen remained blank (at mean luminance) until subjects indicated by a keypress the interval perceived to have higher modulation contrast, which began the subsequent trial. Auditory feedback was provided after each trial. The task in most experiments was modulation detection. That is, $c_0 = 0$, so that the fixed, lower modulation contrast stimulus was, in fact, unmodulated "plaid noise" (Fig. 2B). For the increment threshold experiments c_0 took on nonzero values.

Blocks of trials consisted of either 100 or 200 trials in a given condition. That is, in a given block c_0 , the spatial frequency, stimulus size, and other variables were fixed, and only Δc was varied. The modulation contrast increment Δc was controlled by two interleaved staircases each running for 50 or 100 trials: a 1-up-2-down staircase (converging to a probability correct of 0.71) and a 1-up-3-down staircase (converging to a probability correct of 0.79). In any given experiment, a set of blocks, one per condition, was run in random order, and at least two such sets of blocks of trials were run.

Reported thresholds were calculated by fitting log-normal distributions (corrected for chance) to the data by maximum likelihood. Standard errors were computed by a bootstrap method (Efron & Tibshirani, 1993), as follows. Having estimated the underlying form of a given psychometric function, we simulated an observer governed by that psychometric function. This simulated observer performed in exactly the same experimental design (same number of blocks of each type of staircase) a large number of times. The resulting datasets were fit in exactly the same way, and the stan-

dard deviation of the resulting estimates of threshold was used as our estimate of the standard error.

2.3. Subjects

We report the results of four subjects across all the experiments. All had normal or corrected-to-normal vision.

3. Results

In this section the results of the detection and discrimination experiments are described. We begin with a brief control experiment in spatial summation to determine the stimulus size used for most of the subsequent experiments. This is followed by experiments in modulation contrast sensitivity, increment threshold, and sub-threshold summation. In the following section the results will be compared to those of a specific computational model for an idealized (i.e., nearly optimal) observer.

3.1. Second-order spatial summation

Our first experiment looked at the degree to which stimulus size affected thresholds. In typical first-order contrast detection experiments, threshold improves rapidly with increasing stimulus size for small sizes, reflecting summation within a single mechanism. For larger sizes, however, improvement is more gradual, and this is generally attributed to probability summation across mechanisms (Graham, 1980). Fig. 3 shows the data for two spatial summation experiments. Fig. 3A shows threshold as a function of the stimulus width for a fixed height of 7.5 deg. For all subjects, spatial summation has reached asymptote by a width of 15 deg pretty much independent of the modulator spatial frequency. Fig. 3B shows threshold as a function of the stimulus height for a fixed width of 15 deg. Spatial summation appears to be pretty much complete by a height of 15 deg. Thus, to be able to use a constant stimulus size independent of spatial frequency, in all subsequent experiments a stimulus size of 500 × 500 pixels was used, corresponding to a stimulus size of 15 × 15 deg at the standard viewing distance.

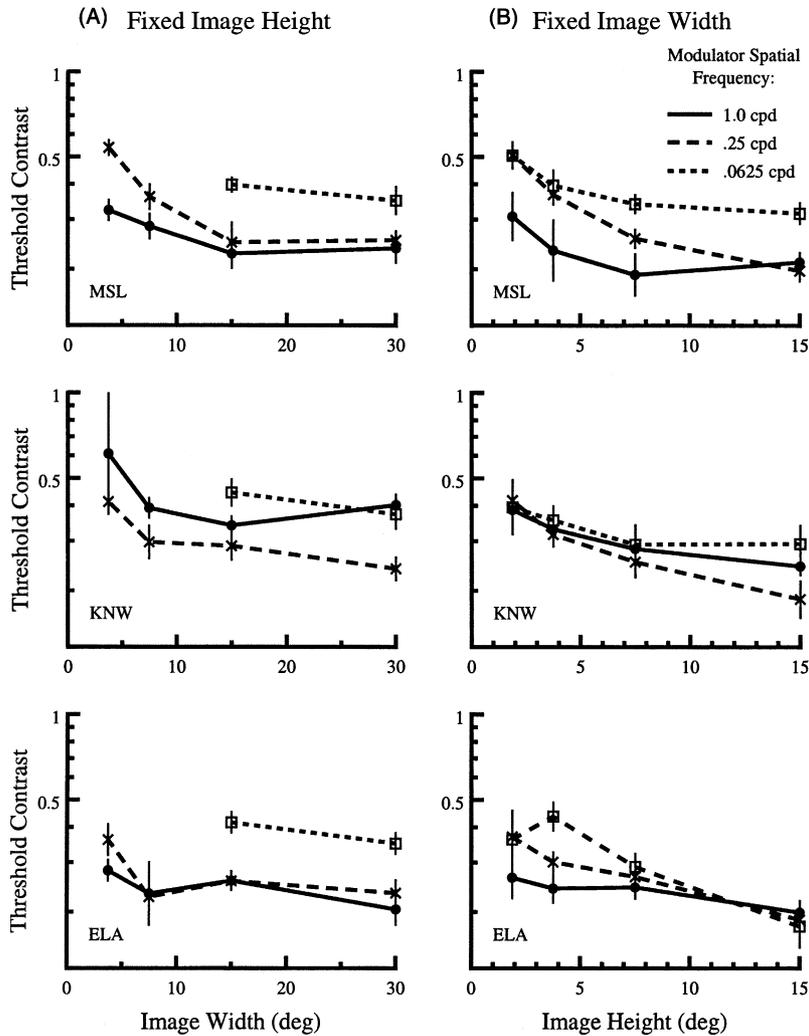


Fig. 3. Spatial summation. (A) Modulation contrast threshold as a function of stimulus width. Stimulus height was fixed at 7.5 deg. (B) Modulation contrast threshold as a function of stimulus height. Stimulus width was fixed at 15 deg. Improvement in performance with increased spatial extent asymptotes at a size of 15×15 deg.

3.2. Second-order modulation contrast sensitivity

Luminance modulation (i.e., first-order) contrast sensitivity is bandpass with a moderate drop in sensitivity at low spatial frequencies, and steeper drop at high spatial frequencies. As we shall see shortly, second-order modulation contrast sensitivity is quite different. As a preview, the modulator used to generate Fig. 4 increases in spatial frequency from left to right, and increases in modulation contrast from top to bottom. It is a second-order version of the luminance frequency sweep grating described by Campbell and Robson (1964) that appears, e.g., in Ratliff (1965, p. 156). The level at which the texture stripes may no longer be discerned is approximately the same across the figure, indicating that second-order modulation contrast sensitivity is relatively independent of spatial frequency.

Measurements of modulation contrast sensitivity taken at the standard viewing distance are shown for

four subjects in Fig. 5A. As was already clear from the demonstration in Fig. 4, although there is a clear hint of a bandpass sensitivity profile, second-order modulation contrast sensitivity is remarkably flat over a five-octave range.

It has often been claimed that mid- and high-level visual processes exhibit scale invariance. That is, visual performance will not change if the size of the stimulus is scaled up or down, with all stimulus features scaled in proportion. For example, Parish and Sperling (1991) found that letter recognition for letters embedded in visual noise is unaffected by changes in the viewing distance to the monitor over a 32:1 range. Similar results have been found for texture segregation (Landy & Bergen, 1991) and reading (Legge, Pelli, Rubin, & Schleske, 1985). The above second-order modulation contrast sensitivity measurements were repeated using the same stimuli on the monitor, viewed from two additional viewing distances. The increased viewing dis-

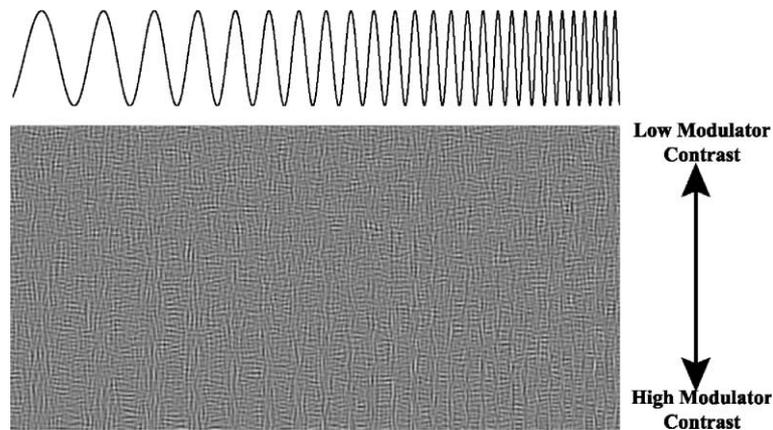


Fig. 4. Second-order sweep grating. This stimulus is constructed analogously to the other stimuli discussed in this paper. In this case, the modulator $M(x,y)$, shown above the stimulus, increases in spatial frequency from left to right and in contrast from top to bottom. Large values of the modulator result in a local texture dominated by vertically oriented noise, and small values by horizontally oriented noise. Notice that the modulation reaches detection threshold at about the same modulation contrast level across the frequency range.

tance reduced the overall stimulus size (in retinal subtense) and increased both the carrier and modulator spatial frequencies by the same factor. Results from this experiment are plotted as a function of the modulator retinal spatial frequency in Fig. 5B. In Fig. 5C the same results are shown shifted horizontally so that performance with the same stimuli on the monitor is plotted at the same abscissa position. The abscissa becomes “carrier cycles per modulator cycle” (a value that does not change with viewing distance). Observer performance is unaffected by changes in viewing distance over this 4:1 range. The modulation CSFs, when plotted in scale-invariant terms in Fig. 5C, overlap marginally better than when plotted in retinal terms (Fig. 5B).

Our modulation contrast sensitivity results may be compared directly with several other second-order modulation contrast sensitivity studies. Sutter et al. (1995) asked subjects to detect contrast modulations of bandpass-filtered, isotropic noise patterns. The modulation targets were Gabor patches with a constant bandwidth in octaves (i.e., a constant number of cycles). They found modulation sensitivity to be bandpass and scale invariant (for carrier frequencies ranging from 2 to 16 cpd), with a peak sensitivity at about 16 carrier cycles per modulator cycle (cf. Fig. 5C), and a drop-off by a factor of 4 in sensitivity as modulation spatial frequency increased to 2 carrier cycles per modulator cycle. Their demonstration of scale invariance was more convincing than ours, as the modulation CSF curves for different viewing distances only coincided after the requisite horizontal shift. They discuss the possibility that the high-frequency drop in sensitivity was due to the use of stimuli of constant bandwidth, so that higher-frequency modulators were smaller and lower energy. But, they claim that little changed in a control experiment that used stimuli of constant retinal size. Thus, their results

differ from ours qualitatively, indicating perhaps that contrast modulations of isotropic texture are detected by a different mechanism or mechanisms than the modulation of orientation content that we employed.

Dakin and Mareschal (2000) also examined contrast modulation sensitivity on various noise carriers (both isotropic and oriented). In most conditions, they held the modulator spatial frequency constant and varied the carrier spatial frequency. Only two modulator spatial frequencies were used in these experiments, varying by a factor of 2. Thus, their results are not directly comparable to ours or to those of Sutter et al. (1995). In particular, the carrier spatial frequency was varied over a very large range over which the first-order sensitivity to the carrier would be expected to vary a great deal. And, to directly compare their results to ours would entail an assumption of scale invariance over this large range. They found (when plotted in comparable terms to ours) either a modestly lowpass modulation CSF (for isotropic carrier noise and for carrier noise oriented parallel to the modulator), or a bandcut modulation CSF for carrier noise perpendicular to the modulator. They also found scale invariance across the two modulator spatial frequencies they used.

Kingdom et al. (1995) used stimuli consisting of randomly placed texture elements (either line segments or Gabor patches) that varied sinusoidally in orientation across the texture. They measured threshold orientation modulation as a function of modulator spatial frequency. They also found a bandpass modulation CSF (although the drop-off from the peak sensitivity was more modest for most subjects) and evidence of scale invariance. In later work (Kingdom & Keeble, 1997) they found that scale-invariance results from a link between the preferred spatial frequency of second-order linear spatial filters and the preferred spatial frequency of the underlying first-order spatial filters.

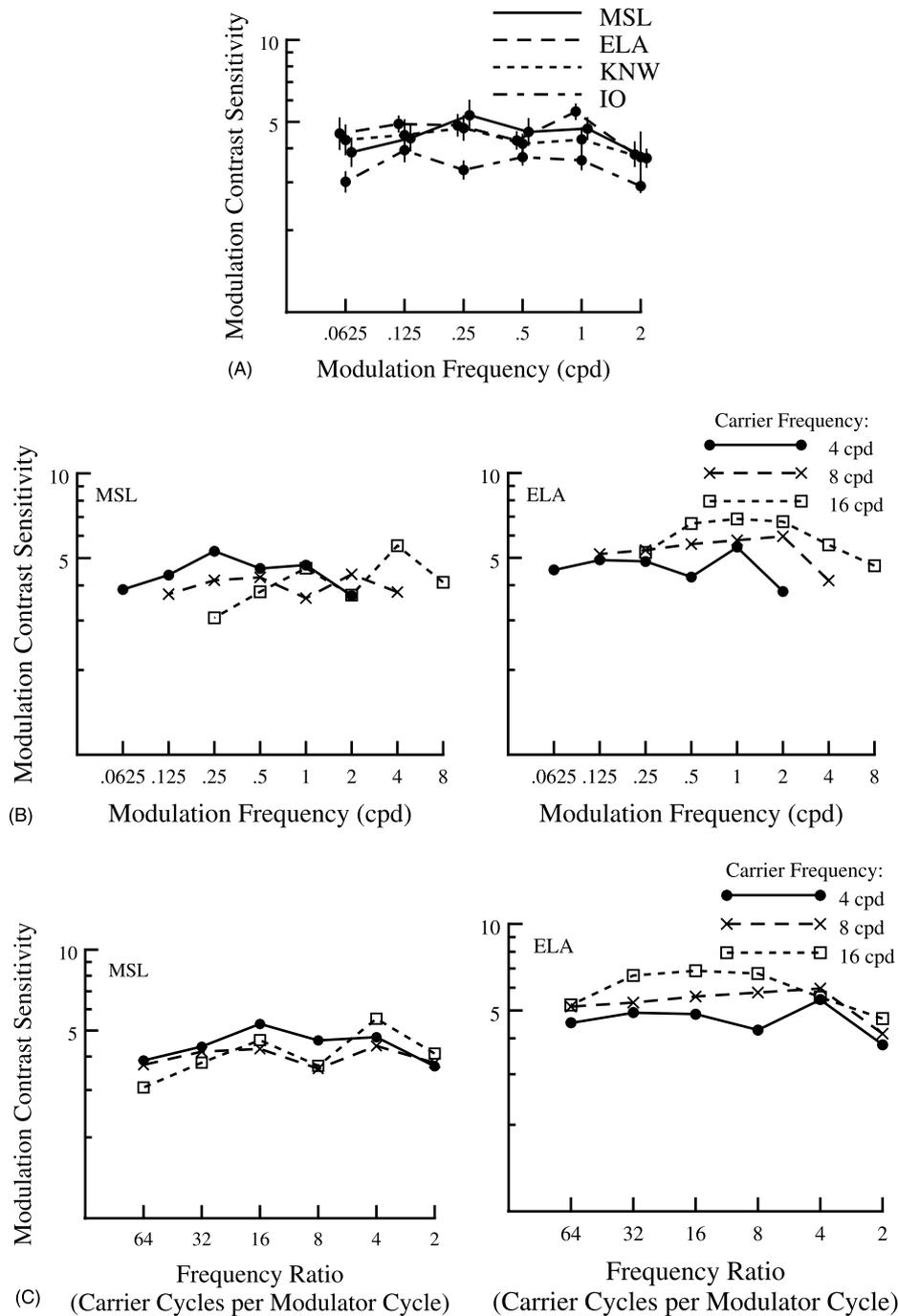


Fig. 5. Second-order modulation contrast sensitivity. (A) Modulation contrast sensitivity is shown as a function of spatial frequency for four subjects. Datasets are offset horizontally for clarity. Measured modulation contrast sensitivity is nearly flat across a five-octave range. (B) Modulation contrast sensitivity with varied viewing distance. Stimuli on the monitor were identical; a change in viewing distance varied carrier and modulator spatial frequency by the same factor. Carrier frequency was 4, 8 or 16 cpd. (C) The same data in B shifted horizontally so that identical stimuli on the monitor are plotted together. The data demonstrate scale invariance, although the test is extremely weak given the nearly flat modulation CSF.

In summary, our study and those we have just reviewed have some quantitative differences as to the exact form of the second-order modulation CSF. They all found only modest second-order tuning. And, wherever tested, second-order scale invariance was found to hold.

3.3. Second-order increment threshold

Fig. 6A shows the results of a second-order increment threshold experiment. In this experiment, observers were required to identify the interval containing the larger

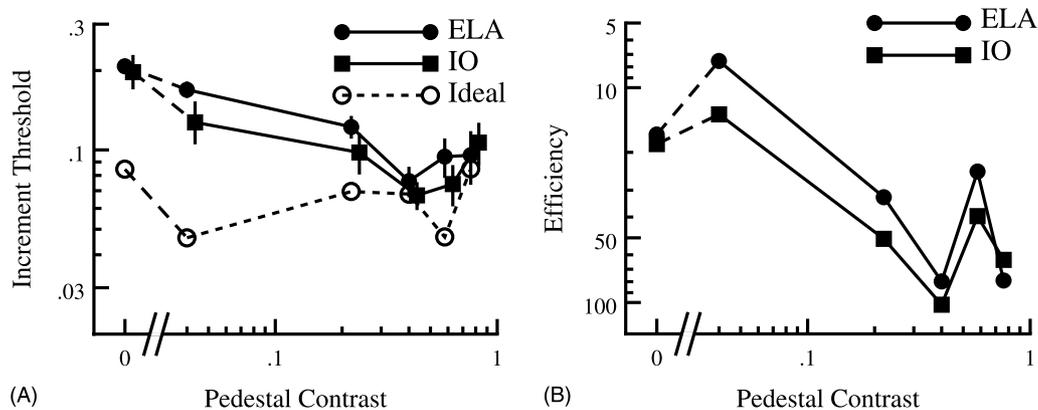


Fig. 6. Increment thresholds. (A) The increment in modulation required for threshold discrimination is plotted as a function of the pedestal modulation contrast for two subjects and our idealized observer. Discrimination improves with increasing pedestal modulation contrast, and shows signs of developing a Weber's Law region at the point where the available modulation contrast range is exhausted. (B) Increment threshold efficiency of the same two observers. Note that the ordinate is reversed so that the curves are of the same form as in (A). That is, better performance is plotted lower, as is true for plotted increment thresholds.

modulation amplitude. Modulation spatial frequency was fixed at 0.5 cpd. Trials were blocked by pedestal modulation contrast, and staircases controlled the modulation contrast increment. Increment threshold decreases with increasing pedestal modulation contrast until the pedestal modulation contrast is about double absolute modulation contrast threshold. For higher modulation contrast pedestals, increment threshold appears to begin to increase, at which point we run out of modulation contrast range.

These results are similar to the "dipper function" found in analogous first-order (luminance contrast modulation) results. First, there is an improvement in increment threshold for sub-threshold pedestals. Second, the amount of improvement is less than that which would indicate the existence of a fixed absolute threshold (i.e., "pedestal + threshold increment = constant"). Third, the bottom of the dipper appears to be near absolute modulation contrast threshold. Thus, the initial portion of the curve is consistent with the typical explanations for the improvement in sensitivity for increment luminance contrast detection (an accelerating nonlinearity or a reduction in stimulus uncertainty with increasing pedestal modulation contrast). We cannot ascertain whether Weber's Law holds for above-threshold stimuli for lack of modulation contrast range.

3.4. Second-order channels: summation experiments

The flat second-order modulation contrast sensitivity displayed in Fig. 5 is traditionally used as the first step toward identifying the underlying mechanism used to perform detection tasks. If the modulation contrast sensitivities we measured were indeed those of a single mechanism used regardless of the spatial frequency to be detected, then the inverse transform would yield the

receptive field profile of the second-order linear filter. The modulation CSF does not constrain the relative phase of the components. But, assuming cosine phase, for example, would yield a receptive field that was small and monophasic (that is, almost a delta function). This seems an untenable result.

An alternative explanation is that the flat modulation CSF is the envelope overlying multiple, second-order spatial frequency channels, analogous to what has been found for luminance contrast detection. In the case of luminance contrast, one popular method for determining the spatial frequency sensitivity of the mechanisms responsible for detection is sub-threshold summation (Graham & Nachmias, 1971). Fig. 7A illustrates the logic behind a typical summation experiment.

In a summation experiment, thresholds are determined for two sine wave gratings differing in spatial frequency. Then, thresholds are determined for stimuli consisting of sums of these two gratings. In any given threshold determination, the ratio of the modulation contrasts of the two summand gratings (i.e., the waveform shape) is held fixed, and the overall modulation contrast is varied to determine threshold for this pattern. In the "summation square" illustrated in Fig. 7A, summation stimuli are plotted relative to the thresholds for the individual grating stimuli, and a single summation threshold determination is carried out using stimuli lying on a line through the origin (i.e., the zero modulation contrast stimulus).

Consider first the situation where the two summands differ greatly in spatial frequency, so that no common mechanism is sensitive to both gratings. Instead, threshold is determined by two mechanisms, those that are most sensitive to each of the individual summands.

In the simplest case, the stimulus is only detected when at least one grating reaches its individual threshold. This

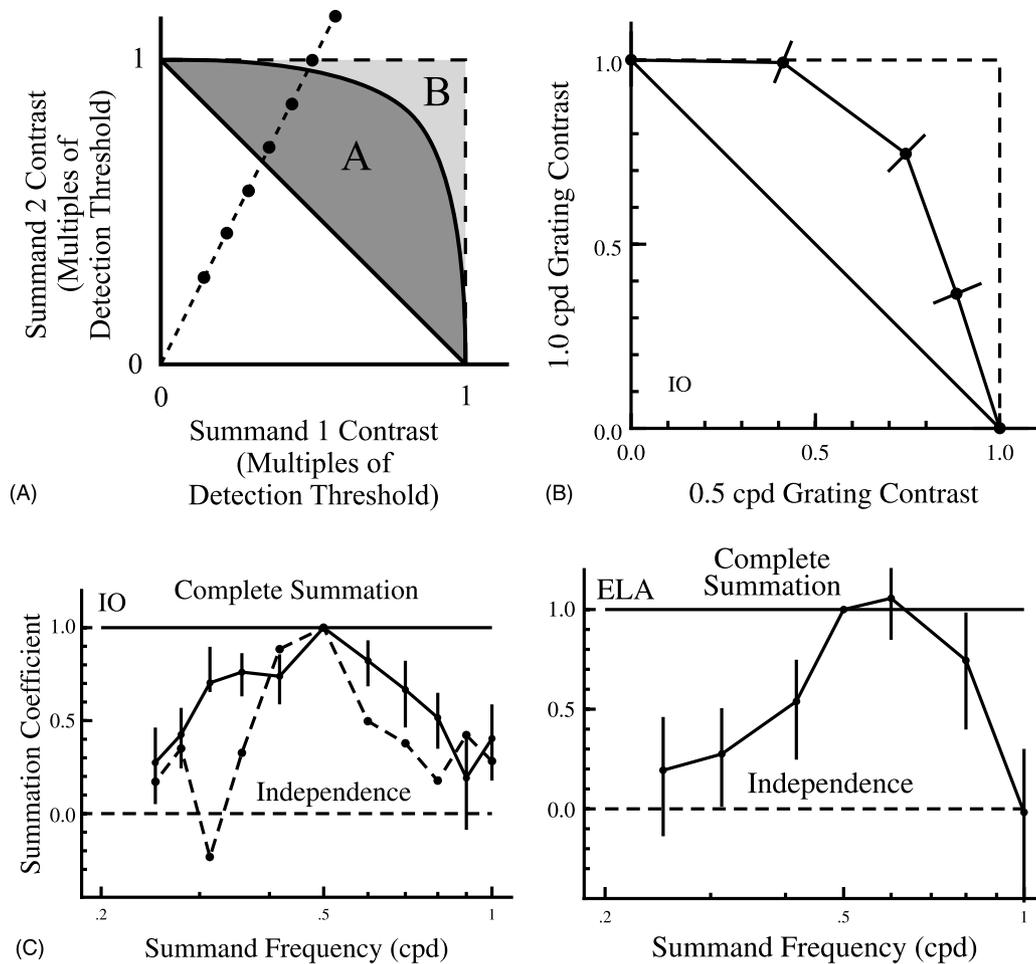


Fig. 7. Summation experiments. (A) Theoretical summation results and the summation coefficient. Summation stimuli are characterized by the amplitudes of the two sine wave grating summands, here plotted as multiples of their individual modulation contrast thresholds. Threshold experiments are carried out by varying modulation amplitude alone for a fixed waveform (i.e., stimuli having a fixed ratio of summand amplitudes, thus lying on a line through the origin such as the dotted line shown here). Complete summation within a single detector results in thresholds lying along the negative diagonal (solid line). Independent hard thresholds (“first to the post”) are indicated by dashed lines. Data typically lie on a curve between these extremes. The degree of summation is calculated as the fraction of the area between these extremes lying outside the curve (i.e. $B/(A + B)$). (B) Sample summation data. The summation coefficient is 0.4. (C) Summation data. In each data set, one summand’s spatial frequency is fixed at 0.5 cpd and the other is varied across blocks of trials. Error bars are derived by bootstrap simulations. For each summation square, new sample values of the five constituent thresholds are drawn from the estimated sampling distributions of each. These are then plotted in a summation square (with axes rescaled based on the new, single-summand thresholds) and a summation coefficient is calculated. This process is repeated 5000 times. The error bars in the figure represent the 5th and 95th percentiles of the resulting distribution of summation coefficients. The dashed curve for subject IO is the summation coefficient calculated for the idealized observer model (described in Section 4.3) for the same conditions. The function is substantially narrower than that of the human observers.

“first to the post” behavior results in the threshold detection contour shown as the dashed lines in Fig. 7A.

An alternative is that two independent mechanisms are used, but each can contribute to detection in a noisy fashion even with sub-threshold stimulation. This may be modeled by probability summation wherein the probability of detection increases with increasing summand modulation contrast and the compound stimulus is detected if either summand is detected (e.g., Graham, 1980). Or, it may be modeled by an observer that sums the noisy output of multiple detectors using a vector length combination rule (Quick, 1974). Either approach predicts a threshold detection contour that is a smooth

curve like that shown in Fig. 7A, lying within the first-to-the-post contour with a shape determined by the psychometric function shape for the components or by the particular vector length metric chosen.

If the two summands are relatively close in spatial frequency, they may both contribute to the response of a single detector. If they are in a spatial phase relationship so as to maximally stimulate the same detector, then linear spatial summation results (the solid negative diagonal in Fig. 7A). The locus of threshold stimuli depends on the relative phases of the summands in a complicated manner when both stimulate the same detector (see Graham & Nachmias, 1971, for details).

We carried out such a second-order summation experiment using modulators that were sums of two sine wave gratings. An example stimulus is shown in Fig. 2D. One summand was 0.5 cpd; the other varied in spatial frequency. The two gratings were in sine phase. The absolute phase was randomized over a 2 deg range (one cycle of the 0.5 cpd summand). The spatial frequencies of the summands were never too close to one another. Thus, the resulting beat frequency was always sufficiently high to ensure that the observer could see the high modulation contrast portion of the beat within the 15 deg stimulus window. For each observer, modulation detection thresholds were determined for each constituent grating alone. Then, for every pair of gratings, thresholds were determined for summation waveforms corresponding to angles of 22.5, 45 and 67.5 deg in the summation square (an example dataset is shown in Fig. 7B).

By the above logic, a high degree of summation corresponds to a summation contour near to the negative diagonal (the solid diagonal line in Fig. 7A) and poor summation to the dashed square. The detailed behavior between those extremes depends on the details of the underlying model of channel behavior and detector combination rule (probability summation versus signal detection theory, peak detection versus vector length, etc.). To obviate difficulties of interpretation, we summarize the results of each summation experiment with a simple, nonparametric statistic we call the summation coefficient. It is defined as $B/(A+B)$, where A and B are the areas below and above the summation contour illustrated in Fig. 7A. A summation coefficient of zero corresponds to complete independence (the dashed lines), and a value of one corresponds to the summation contour (the solid, negative diagonal line).

Fig. 7C shows the summation coefficient as a function of the spatial frequency of the variable summand. The summation coefficient has a value of one at a summand frequency of 0.5 cpd (the spatial frequency of the fixed summand) by definition. The summation coefficient drops substantially for spatial frequencies an octave above or below the fixed summand. Thus, the flat second-order modulation CSF of Fig. 5 may well be the envelope over multiple, second-order spatial frequency channels, each with bandwidth of an octave or so, much as in the first-order, luminance contrast case. This conclusion is in agreement with the results of Arsenault et al. (1999) and Schofield and Georgeson (1999), but not those of Kingdom and Keeble (1996).

4. Idealized observer modeling

When confronted with an array of results such as those just presented, it is generally useful to compare

observer performance to that of an ideal observer. For example, by applying an ideal observer at various stages along the visual pathway, one can learn what aspects of stimulus information are lost at each stage of processing (Geisler, 1989). A clear example is first-order (luminance) contrast sensitivity, for which poor foveal performance at high spatial frequencies may be attributed to losses in the optics, whereas performance in the periphery or at low spatial frequencies is worse than that of an ideal observer at the level of the cone quantum catch, so these losses must be attributed to neural processing (Banks & Bennett, 1988; Banks, Sekuler, & Anderson, 1991). Thus, ideal observer analysis can help one discover whether a pattern of performance data is the result of the inherent information content of the stimuli, or whether it indicates something more profound about the nature of visual processing.

The use of modulated stimuli for studying second-order pathways carries with it an implicit assumption that the early visual stages successfully demodulate the stimulus. And, for the results to characterize a second-order filter, the stimulus should isolate a single second-order channel. Further, the stimulus noise at the input of the second-order channel (after demodulation) should be independent of the modulator used, so that signal strength of the modulator corresponds in a simple way to the signal-to-noise ratio at the input of the linear filter under study.

Previous attempts to measure second-order modulation contrast sensitivity violate one or more of these assumptions. Sutter et al. (1995) measured the detectability of Gabor patches used to modulate the contrast of isotropic, bandpass-filtered noise. This is a stimulus that would be successfully demodulated by any channel with a first-order linear spatial filter matched in spatial frequency to that of the carrier noise, regardless of whether that filter were orientation tuned or not, and the orientation preference of that filter. Thus, it is possible that the modulator Gabor would be delivered to multiple second-order filters, and so one would need to model the manner in which those signals are combined.

The stimulus used by Kingdom et al. (1995) is even more complicated to analyze. They used the modulator to control the local orientation in a texture. Thus, a sine wave modulator would result in a stimulus, e.g., that sinusoidally modulated across the pattern from horizontal to diagonal upward (at the peak of the modulator), back to horizontal, and then diagonal downward (at the trough). Consider a pathway with first-order filter tuned to the average orientation (i.e., horizontal). After the succeeding nonlinearity, the response will be maximal at portions of the pattern that are horizontal, and less strong as the pattern deviates from horizontal (in either direction). Thus, the response reaches a maximum at the zero-crossings of the modulator pattern

(where the texture is horizontally oriented), and a minimum for both the peaks and troughs of the modulator. In other words, the response that serves as the input to the second-order linear filter is a frequency-doubled version of the modulator. A pathway with a filter tuned to the most clockwise-rotated texture element (say, the diagonal upward ones) would not be frequency doubled. But, the identity of this pathway will depend on the degree of orientation modulation, which is precisely the value varied in their experiments. Thus, this is a stimulus that is exceedingly difficult to analyze in terms of the back-pocket model.

We chose our stimulus for a simple characterization in terms of the back-pocket model. The first-order (carrier) content of the stimulus is bandpass in the spatial frequency domain so that off-frequency-tuned first-order channels are uninformative (as with the stimuli of Sutter et al. (1995), and those of Kingdom et al. (1995) when Gabor patches are used as texture elements). The orientation content is varied between vertical and horizontal only, so that the only relevant channels are those tuned to one or the other of those orientations (and they provide comparable, although phase-inverted, demodulated output to the second-order filter). As a result, the stimulus nearly completely isolates a single first-order linear-filter stage, allowing us to concentrate on the second-order linear filter(s) that it feeds.

But, we do not know that the noise at the input to the second-order filter is independent of the particular modulator used. “Noise” is that characteristic of the input to the second-order linear filter unrelated to and independent of modulation contrast. The amplitude and spectral characteristics of this noise limit detection of the modulator. The carrier patterns are filtered noise patterns and thus are random from trial to trial, adding random noise to the demodulated pattern. If that noise is additive, independent of modulation pattern or contrast, and has a flat spatial frequency spectrum, then the results of our experiments can be used to interpret the form and number of second-order filters in much the same way as was done for first-order spatial filters. But, the carrier-induced noise is unlikely to have all these characteristics. For example, the nonlinearities required for demodulation produce distortion products resulting from an interaction between the carrier and modulator. Hence, our results cannot be interpreted without reference to a model of the processing of these stimuli.

The most desirable such model is an ideal observer. Our human observers were faced with equally likely alternatives and received feedback as to whether their responses were correct. The appropriate ideal observer chooses the interval on each trial that has maximum likelihood. Unfortunately, for such complex stimulus construction as used here, the ideal observer is compu-

tationally formidable. As such, we have settled on comparing our observers’ performance to an “idealized” observer, that is, to a computational model that although not ideal, is likely to be close in performance to the ideal and that is relatively easy to compute.

Our idealized observer processes a stimulus in three stages. The first two stages are identical to the first two stages of the back-pocket model. First, the stimulus is filtered to emphasize one of the two constituent textures (in our case, the vertical texture is emphasized). As we know the bandpass characteristics (in spatial frequency and orientation) that were used to generate the textures, the same exact linear filter is used in the model. This produces regions of high contrast where the texture is predominantly vertical, and low contrast where it is horizontal. A pointwise nonlinearity (x^2) is then applied to demodulate the signal, resulting in a large average response in the regions of vertical texture. We use an even nonlinearity so that both strong positive and negative responses of the first-order linear filter are used. Other even nonlinearities (e.g., fullwave rectification) would likely have given similar results. The resulting image is indeed similar to the modulator that gave rise to the texture pattern, although quite noisy (Fig. 8A).

Finally, the resulting image is cross-correlated with the known signal (which is always fixed except for contrast and spatial phase in any block of trials). Since our patterns are one-dimensional, we first average the values across each column of the demodulated image, resulting in a one-dimensional function. This function is cross correlated with the known signal (with the DC removed) at every possible phase, and the maximum cross correlation is calculated. Our stimulus window size is not an integral multiple of the spatial period of our stimuli, so we are forced to use explicit cross correlation rather than, say, the power in the Fourier spectrum. Also, since the expected signals were clipped at the edges of the stimulus window, they had a nonzero mean (that is, a DC component) that had to be removed as it interfered with our idealized observer’s performance.

The idealized observer’s maximum cross-correlation value was computed for each of the stimulus images used in our experiments. Consider, for example, the experiment used to determine modulation contrast sensitivity at 0.5 cpd. The stimulus set consisted of 20 stimuli with zero modulation contrast, and 10 stimuli at each of 12 nonzero modulation contrast levels. For a given modulation contrast level c , the idealized observer’s performance level was estimated as the proportion of pairs of zero modulation contrast and c modulation contrast stimuli for which the maximum cross-correlation was larger for the c modulation contrast stimulus (out of the 200 possible such pairs). A psychometric function was thus constructed for the idealized observer, and thresholds and other statistics

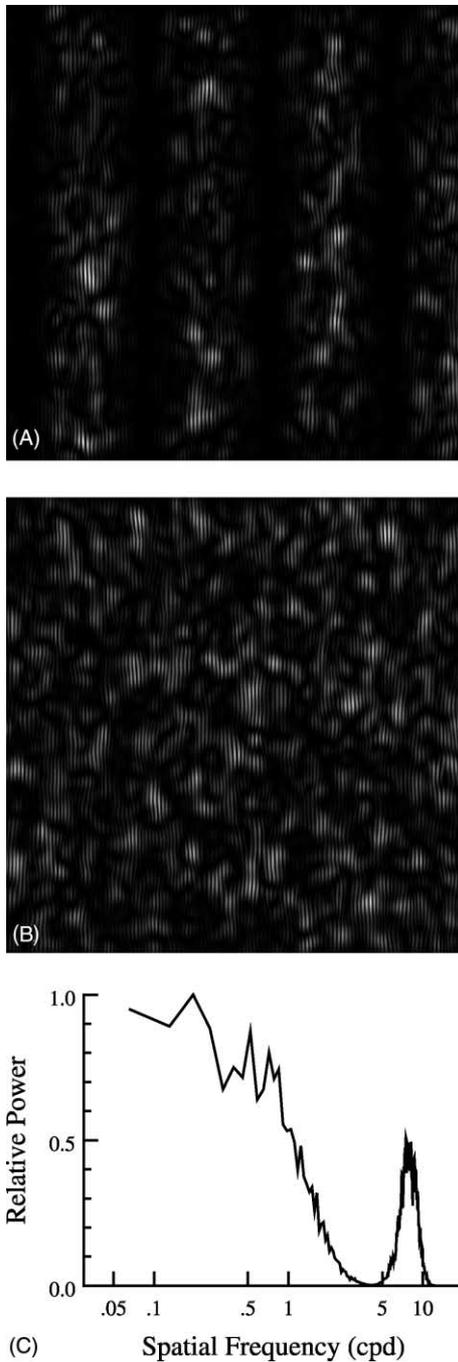


Fig. 8. Idealized observer behavior. (A) The idealized observer applies a linear filter tuned to the vertically oriented texture used to generate the stimuli, followed by a pointwise nonlinearity (x^2). Here is the result of those two operations applied to the stimulus shown in Fig. 2C, resulting in a noisy demodulation of the stimulus. (B) The analogous image for the zero modulation contrast stimulus of Fig. 2B. (C) The Fourier spectrum of the vertically oriented components was averaged over 20 zero modulation contrast stimuli including (B) and normalized to a peak value of 1. Although not shown here, note that the standard deviation of the spectral power across these same 20 images follows the same function of spatial frequency as the average power.

were calculated in exactly the same way as for the human observers.

4.1. Modulation contrast sensitivity

Fig. 9A shows the second-order modulation contrast sensitivity of our idealized observer. The function, unlike that of our human observers, is not flat. In fact, it shows increased sensitivity for both the lowest and highest spatial frequencies tested. In Fig. 9B we plot the average observer efficiency relative to the idealized observer. This is the square of the ratio of the average human observer sensitivity to that of the idealized observer. It is the percentage of the modulation contrast power that the idealized observer requires to achieve the performance of the average human observer. Two points are notable about this plot. First, since the performance of the idealized observer is not independent of spatial frequency (unlike the human observer modulation CSFs in Fig. 5), efficiency turns out to be a band-pass function of spatial frequency. That is, despite the flat modulation CSFs we measured, the human observers appear to utilize the mid-range of modulator frequencies more effectively than the highest and lowest spatial frequencies tested. Second, for those mid-range frequencies, the observer efficiency is somewhere between 10% and 20%. To the extent that our idealized observer closely approximates that of the maximum likelihood ideal observer (something we are unable to check), these efficiencies are reasonably similar to those seen for other mid-level vision tasks (such as letter discrimination in noise; Parish & Sperling, 1991).

Why is the idealized observer's performance dependent on modulator spatial frequency? Fig. 8B shows the idealized observer's response to a zero modulation contrast stimulus after filtering and pointwise nonlinearity. Note that this noisy image has evident modulations of contrast across the pattern at relatively low spatial frequencies. These low-frequency patterns are in response to local contrast modulations present in the vertical noise pattern used to generate the stimuli. Kovács and Fehér (1997) pointed out that bandpass filtered noise patterns contain low frequency, second-order contrast modulations that depend on the spatial frequency- and orientation-bandwidth of the noise pattern in a predictable way. These contrast modulations are analogous to the contrast envelopes of sums of sine wave grating patterns with similar spatial frequency (i.e., "beat" patterns) except that they are irregular. This modulation pattern acts like a noise mask through which the observer is required to detect the grating stimulus (after demodulation).

Fig. 8C shows the spectral power for vertically oriented components averaged over 20 filtered and squared images like that in Fig. 8B. The spectral power has a local peak like that in Fig. 8B. The spectral power has a local peak at 8 cpd, which shows the frequency doubling of the 4 cpd carrier noise, visible in Fig. 8B, due to the even nonlinearity we employ. The spectrum also shows a lowpass component with a steep spectral slope above

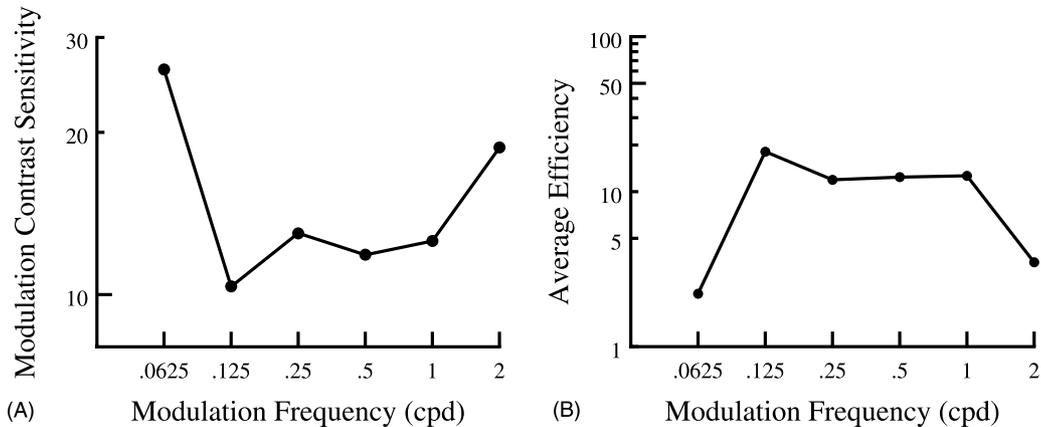


Fig. 9. Idealized observer modulation CSF and efficiency. (A) The second-order modulation contrast sensitivity function of an idealized observer that demodulates the stimulus by use of a linear filter followed by squaring, and then cross-correlates the stimuli with the target modulator grating. It chooses the interval with the larger peak correlation value. Note the different range of the ordinate compared to Fig. 5. Unlike the human observers (Fig. 5A), the idealized observer modulation CSF is not flat. (B) The average efficiency, i.e., the square of the ratio between the average sensitivity of the four observers in Fig. 5A and idealized observer sensitivity. The efficiency is bandpass.

0.8 cpd. That is, above 0.8 cpd the noise power decreases, allowing the idealized observer's performance to improve at 2 cpd. However, there is nothing in this noise spectrum to indicate why the idealized observer also performs relatively well with the 0.0625 cpd stimuli. This latter performance must therefore result from other interactions between the modulator and this contrast noise or other aspects of the stimulus such as edge effects from the hard-edged stimulus window.

4.2. Increment threshold

The variability in the model's response to a particular class of stimuli is generally proportional to the average response. We have seen this in the model output statistics (i.e., the statistics of the maximum correlation value across the set of stimuli used in the experiments, as a function of modulation contrast). It is also clear from the statistics of the contrast modulation noise in the idealized observer output after filtering and pointwise nonlinearity. Fig. 8C shows the average spectral power of the response to 20 unmodulated texture images, but the standard deviation of the spectral power has the same form. These characteristics lead to an idealized observer increment threshold function that, although a bit noisy over the small sample of images used, is essentially flat (Fig. 6A, dotted line). Relative to our idealized observer, as a result, efficiency improves (although noisily) as pedestal modulation contrast increases (Fig. 6B). Interestingly, this efficiency approaches 100% for the higher pedestal values. This is a strong clue that our idealized observer is, in fact, sub-ideal.

4.3. Summation

The interpretation of the results of sub-threshold summation experiments, outlined above, rests on the assumption of an observer who uses the best detector (in this case, second-order receptive field profile) among a fixed set of detectors to which s/he has access. What would we predict for an ideal observer, that is, an observer that could tailor the second-order filter to the particular stimulus used in any given block of trials? Such an observer would construct a different template for each condition in the summation experiment (corresponding to the particular mix of the two summands displayed). When the two summands were identical, then by definition perfect summation would result. For any other combination of summands, and assuming a flat noise spectrum limiting performance, performance is a function of signal power. We use orthogonal (i.e., uncorrelated) summands. Hence, for any such pair of summands, one would predict thresholds for different directions in the summation square to have equal vector length (equal power), resulting in a circular threshold contour. Such a threshold contour results in a summation coefficient of 0.43. By this logic, one would predict ideal summation of 0.43 independent of the frequency of the variable summand (except when it is identical to the fixed summand).

Fig. 7C (left-hand panel) shows the summation coefficient as a function of summand spatial frequency for our idealized observer as the dashed curve. The individual estimates were computed using the stimuli that appeared in the analogous summation experiment for observer IO. Since the ideal observer did not have the same thresholds for the single sine wave components as IO had, when the axes were scaled as multiples of

threshold, the summation conditions for the idealized observer were no longer located at angles of 22.5, 45 and 67 deg in the idealized observer's summation plots. The idealized observer does indeed show a narrower bandwidth than the human observers in this task, although not infinitely narrow as predicted. In any case, the difference between the summation plots for the idealized observer and our human observers is reassuring. It indicates that the summation experiment does indeed indicate something about how observers perform this detection task, and not simply that the information content of the summation stimuli has this property. Thus, we are on reasonably firm ground in interpreting the results of the summation experiments as indications that second-order filters used to accomplish that task have a bandwidth of an octave or so.

5. Discussion

We have described a series of experiments intended to characterize the properties of the second-order linear filters used to detect modulations in visual texture. We found second-order modulation contrast sensitivity to be nearly flat and scale invariant, although the modulation CSF was bandpass when converted to efficiency relative to our idealized observer. The other groups that have measured second-order modulation contrast sensitivity (Kingdom et al., 1995; Sutter et al., 1995) have found a bandpass function. This difference between our results and theirs surely stems from the different types of stimuli used. We pointed out that our choice of stimuli may be easier to interpret in the context of the back-pocket model. The increment threshold function showed an improvement with increasing pedestal modulation contrast that could not be attributed to stimulus information content (i.e., that was not evident in the results of the idealized observer). Finally, sub-threshold summation experiments indicated a moderate channel bandwidth that was substantially wider than would have been found had the observer used a matched filter in each condition. In summary, the results indicate that second-order modulation detection and, by extension, other spatial, second-order tasks such as texture segregation, may be carried out by multiple, second-order, spatial frequency-tuned mechanisms analogous in their operation to those that have been well established in the first-order, luminance contrast case.

Acknowledgements

This work was supported by NIH grant EY08266. We thank Beau Watson for the initial spark that led to this work. We thank Cristina Ternes, Kevin Willenson and Leanne Chukoskie with help and observation along

the way, and Tony Movshon for suggesting the non-parametric summation coefficient. We also thank Sabina Wolfson, Sergei Gepshtein, Larry Maloney and two anonymous reviewers for comments on the manuscript. Some of these results were previously discussed at annual meetings of the Association for Research on Vision and Ophthalmology (Chukoskie & Landy, 1997; Landy, 1996; Oruç & Landy, 2000) and the Optical Society of America (Landy & Ternes, 1995). Future research might try to shed light on other filtering properties of these second-order filters. Are these filters fixed, or are they programmable by the observer (with some constraints, as a matched filter is clearly not achievable by observers in our task) based on the task at hand?

References

- Arsenault, A. S., 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.
- Banks, M. S., & Bennett, P. J. (1988). Optical and photoreceptor immaturities limit the spatial and chromatic vision of human neonates. *Journal of the Optical Society of America A*, *5*, 2059–2079.
- Banks, M. S., Sekuler, A. B., & Anderson, S. J. (1991). Peripheral spatial vision: limits imposed by optics, photoreceptors, and receptor pooling. *Journal of the Optical Society of America A*, *8*, 1775–1787.
- Bergen, J. R., & Adelson, E. H. (1988). Early vision and texture perception. *Nature*, *333*, 363–364.
- Bergen, J. R., & Landy, M. S. (1991). Computational modeling of visual texture segregation. In M. S. Landy, & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 253–271). Cambridge, MA: MIT Press.
- 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 (London)*, *203*, 237–260.
- Blakemore, C., & Sutton, P. (1969). Size adaptation: a new aftereffect. *Science*, *166*, 245–247.
- Bovik, A. C., Clark, M., & Geisler, W. S. (1990). Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *12*, 55–73.
- Caelli, T. (1985). Three processing characteristics of visual texture segmentation. *Spatial Vision*, *1*, 19–30.
- Campbell, F. W., & Robson, J. G. (1964). Application of Fourier analysis to the modulation response of the eye. *Journal of the Optical Society of America*, *54*, 581A.
- Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology (London)*, *197*, 551–566.
- Chubb, C., Econopouly, J., & Landy, M. S. (1994). Histogram contrast analysis and the visual segregation of iid textures. *Journal of the Optical Society of America A*, *11*, 2350–2374.
- Chubb, C., & Landy, M. S. (1991). Orthogonal distribution analysis: a new approach to the study of texture perception. In M. S. Landy, & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 291–301). Cambridge, MA: MIT Press.
- Chubb, C., Olzak, L., & Derrington, A. (2001). Second-order processes in vision: introduction. *Journal of the Optical Society of America A*, *18*, 2175–2178.

- 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, 1986–2006.
- Chukoskie, L., & Landy, M. S. (1997). 2nd-order summation experiments indicate multiple 2nd-order channels. *Investigative Ophthalmology & Vision Science Supplement*, 38, S2.
- Clark, M., & Bovik, A. C. (1989). Experiments in segmenting texton patterns using localized spatial filters. *Pattern Recognition*, 22, 707–717.
- Dakin, S. C., & Mareschal, I. (2000). Sensitivity to contrast modulation depends on carrier spatial frequency and orientation. *Vision Research*, 40, 311–329.
- Dakin, S. C., Williams, C. B., & Hess, R. F. (1999). The interaction of first- and second-order cues to orientation. *Vision Research*, 39, 2867–2884.
- De Valois, R. L., & De Valois, K. K. (1990). *Spatial vision*. New York: Oxford University Press.
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- Fogel, I., & Sagi, D. (1989). Gabor filters as texture discriminator. *Biological Cybernetics*, 61, 103–113.
- Foley, J. M., & Legge, G. E. (1981). Contrast discrimination and near-threshold discrimination in human vision. *Vision Research*, 21, 1041–1053.
- Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discrimination. *Psychoanalytic Review*, 96, 1–71.
- Graham, N. (1980). Spatial frequency channels in human vision: detecting edges without edge detectors. In C. Harris (Ed.), *Visual coding and adaptability* (pp. 215–262). Hillsdale, NJ: Erlbaum.
- Graham, N. V. S. (1989). *Visual pattern analyzers*. New York: Oxford University Press.
- Graham, N. (1991). Complex channels, early local nonlinearities, and normalization in texture segregation. In M. S. Landy, & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 273–290). Cambridge, MA: MIT Press.
- Graham, N. (1994). Non-linearities in texture segregation. In G. R. Bock, J. A. Goode (Eds.), *CIBA Foundation Symposium (Vol. 184, pp. 309–329)*. New York: Wiley.
- Graham, N., Beck, J., & Sutter, A. (1992). Nonlinear processes in spatial-frequency channel models of perceived texture segregation: Effects of sign and amount of contrast. *Vision Research*, 32, 719–743.
- 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.
- Graham, N., & Sutter, A. (1996). Effect of spatial scale and background luminance on the intensive and spatial nonlinearities in texture segregation. *Vision Research*, 36, 1371–1390.
- Graham, N., & Sutter, A. (1998). Spatial summation in simple (Fourier) and complex (non-Fourier) texture channels. *Vision Research*, 38, 231–257.
- Graham, N., & Sutter, A. (2000). Normalization: contrast-gain control in simple (Fourier) and complex (non-Fourier) pathways of pattern vision. *Vision Research*, 40, 2737–2761.
- Graham, N., Sutter, A., & Venkatesan, C. (1993). Spatial-frequency- and orientation-selectivity of simple and complex channels in region segmentation. *Vision Research*, 33, 1893–1911.
- Graham, N., & Wolfson, S. S. (2001). A note about preferred orientations at the first and second stages of complex (second-order) texture channels. *Journal of the Optical Society of America A*, 18, 2273–2281.
- Henning, G. B., Hertz, B. G., & Hinton, J. L. (1977). Effects of different hypothetical detection mechanisms on the shape of spatial-frequency filters inferred from masking experiments: I. noise masks. *Journal of the Optical Society of America*, 71, 574–581.
- Hess, R. F., & Wilcox, L. M. (1994). Linear and non-linear filtering in stereopsis. *Vision Research*, 34, 2431–2438.
- Jamar, J. H. T., & Koenderink, J. J. (1985). Contrast detection and detection of contrast modulation for noise gratings. *Vision Research*, 25, 511–521.
- Keeble, D. R. T., Kingdom, F. A. A., & Morgan, M. J. (1997). The orientational resolution of human texture perception. *Vision Research*, 37, 2993–3007.
- 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., & Keeble, D. R. T. (1997). The mechanism for scale invariance in orientation-defined textures. *Investigative Ophthalmology & Vision Science Supplement*, 38, S636.
- Kingdom, F. A. A., & Keeble, D. R. T. (1999). On the mechanism for scale invariance in orientation-defined textures. *Vision Research*, 39, 1477–1489.
- 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. R. T., & Moulden, B. (1995). Sensitivity to orientation modulation in micropattern-based texture. *Vision Research*, 35, 79–91.
- Knutsson, H., & Granlund, G. H. (1983). Texture analysis using two-dimensional quadrature filters. In *1983 IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management* (pp. 206–213). Silver Spring, MD: IEEE Computer Society.
- Kovács, I., & Fehér, A. (1997). Non-Fourier information in bandpass noise patterns. *Vision Research*, 37, 1167–1175.
- Kwan, L., & Regan, D. (1998). Orientation-tuned spatial filters for texture-defined form. *Vision Research*, 38, 3849–3855.
- Landy, M. S. (1996). 2nd-order contrast discrimination. *Investigative Ophthalmology & Vision Science Supplement*, 37, S1147.
- Landy, M. S., & Bergen, J. R. (1991). Texture segregation and orientation gradient. *Vision Research*, 31, 679–691.
- Landy, M. S., Cohen, Y., & Sperling, G. (1984). Hips: a unix-based image processing system. *Computer Vision Graphics and Image Processing*, 25, 331–347.
- Landy, M. S., & Graham, N. (in press). Visual perception of texture. In L. M. Chalupa, J. S. Werner (Eds.), *The visual neurosciences*. Cambridge, MA: MIT Press.
- Landy, M. S., & Ternes, C. M. (1995). 2nd-order spatial contrast sensitivity. In 1995 OSA Annual Meeting/ILS-XI Program (p. 64). Washington, DC: Optical Society of America.
- Legge, G. E. (1981). A power law for contrast discrimination. *Vision Research*, 21, 457–467.
- 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., Pelli, D. G., Rubin, G. S., & Schleske, M. M. (1985). Psychophysics of reading—I. normal vision. *Vision Research*, 25, 239–252.
- Malik, J., & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A*, 7, 923–932.
- Nachmias, J., & Sansbury, R. V. (1974). Grating contrast: discrimination may be better than detection. *Vision Research*, 14, 1039–1042.
- Olzak, L. A., & Thomas, J. P. (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.
- Olzak, L. A., & Thomas, J. P. (1999). Neural recoding in human pattern vision: model and mechanisms. *Vision Research*, 39, 231–256.
- Olzak, L. A., & Wickens, T. D. (1997). Discrimination of complex patterns: orientation information is integrated across spatial scale; spatial-frequency and contrast information are not. *Perception*, 26, 1101–1120.

- Oruç, İ., & Landy, M. S. (2000). 2nd-order summation experiments indicate narrow 2nd-order channel bandwidth. *Investigative Ophthalmology & Vision Science Supplement*, 41, S805.
- Pantle, A., & Sekuler, R. (1968). Size-detecting mechanisms in human vision. *Science*, 162, 1146–1148.
- Parish, D. H., & Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Research*, 31, 1399–1415.
- Pelli, D. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A*, 2, 1508–1532.
- Quick, R. F., Jr. (1974). A vector-magnitude model of contrast detection. *Kybernetik*, 16, 65–67.
- Ratliff, F. (1965). *Mach bands: quantitative studies on neural networks in the retina*. San Francisco: Holden-Day.
- Sachs, M. B., Nachmias, J., & Robson, J. G. (1971). Spatial-frequency channels in human vision. *Journal of the Optical Society of America*, 61, 1176–1186.
- Sagi, D. (1990). Detection of an orientation singularity in Gabor textures: Effect of signal density and spatial-frequency. *Vision Research*, 30, 1377–1388.
- Schofield, A. J., & Georgeson, M. A. (1999). Sensitivity to modulations of luminance and contrast in visual white noise: separate mechanisms with similar behavior. *Vision Research*, 39, 2697–2716.
- Stromeyer III, C. F., & Julesz, B. (1972). Spatial-frequency masking in vision: Critical bands and spread of masking. *Journal of the Optical Society of America*, 62, 1221–1232.
- Sutter, A., Beck, J., & Graham, N. (1989). Contrast and spatial variables in texture segregation: Testing a simple spatial-frequency channels model. *Perception & Psychophysics*, 46, 312–332.
- Sutter, A., Sperling, G., & Chubb, C. (1995). Measuring the spatial frequency selectivity of second-order texture mechanisms. *Vision Research*, 35, 915–924.
- Turner, M. R. (1986). Texture discrimination by Gabor functions. *Biological Cybernetics*, 55, 71–82.
- Watson, A. B., & Eckert, M. P. (1994). Motion-contrast sensitivity: visibility of motion gradients of various spatial frequencies. *Journal of the Optical Society of America*, 11, 496–505.
- Wilson, H. R. (1980). A transducer function for threshold and suprathreshold human vision. *Biological Cybernetics*, 38, 171–178.
- Wilson, H. R., Ferrera, V. P., & Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Vision Neuroscience*, 9, 79–97.
- Wilson, H. R., McFarlane, D. K., & Phillips, G. C. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research*, 23, 873–882.
- Wolfson, S. S., & Landy, M. S. (1995). Discrimination of orientation-defined texture edges. *Vision Research*, 35, 2863–2877.