PERGAMON

Vision Research 43 (2003) 445–466

Vision Research

www.elsevier.com/locate/visres

Reading a population code: a multi-scale neural model for representing binocular disparity

Jeffrey J. Tsai *, Jonathan D. Victor

Department of Neurology and Neuroscience, Weill Medical College of Cornell University, 1300 York Avenue, New York, NY 10021, USA

Received 25 March 2002; received in revised form 11 September 2002

Abstract

Although binocular neurons in the primary visual cortex are sensitive to retinal disparity, their activity does not constitute an unambiguous disparity signal. A multi-spatial-scale neural model for disparity computation is developed to examine how population activity might be interpreted to overcome ambiguities at the single neuron level. The model incorporates a front end that encodes disparity by a family of complex cell-like energy units and a second stage that reads the population activity. Disparity is recovered by matching the population response to a set of canonical templates, derived from the mean response to white noise stimuli at a range of disparities. Model predictions are qualitatively consistent with a variety of psychophysical results in the literature, including the effects of spatial frequency on stereoacuity and bias in perceived depths, and the effect of standing disparity on increment thresholds. Model predictions are also consistent with data on qualitative appearance of complex stimuli, including depth averaging, transparency, and corrugation. The model also accounts for the non-linear interaction of disparities in compound grating stimuli. These results show that a template-match approach reduces ambiguities in individual and pooled neuronal responses, and allows for a broader range of percepts, consistent with psychophysics, than other models. Thus, the pattern of neural population activity across spatial scales is a better candidate for the neural correlate of depth perception than the activity of single neurons or the pooled activity of multiple neurons.

Keywords: Stereo vision; Population code; Spatial frequency pooling; Intersection-of-constraints; Disparity interaction; Depth transparency; Threshold

1. Introduction

Binocular neurons in the primary visual cortex (V1) are sensitive to disparity (Barlow, Blakemore, & Pettigrew, 1967; Ohzawa, DeAngelis, & Freeman, 1990; Poggio & Fischer, 1977). The activity of individual V1 neurons, however, does not immediately account for the perceptual representation of stimulus disparity. One reason for this is that the tuning of individual neurons is broad, and their responses are also influenced by other factors, such as contrast and spatial frequency. Nevertheless, one expects that precise information about disparity can be carried by a population of such neurons, even though a single neuron's activity is ambiguous. A second difficulty with equating the disparity tuning of individual neurons and the perceptual representation of disparity is that for some stimuli, there may be more than one disparity that is consistent with the retinal images. In this case, one expects that the population activity provides the substrate for the several candidate disparities, and may influence which disparity, or disparities, are perceived. In this paper, we ask how the activity of a population of binocular V1 neurons tuned to a range of spatial frequencies may contribute to the perceptual representation of disparity, by simulating the behavior of a physiologically based model and comparing the results of simulations to psychophysical studies.

Multiple band-limited spatial frequency channels contribute to stereopsis (Glennerster & Parker, 1997; Julesz & Miller, 1975; Prince, Eagle, & Rogers, 1998; but see Yang & Blake, 1991). One proposal for integration of responses across spatial scales is that large disparities

* Corresponding author. Tel.: +1-212-746-2343; fax: +1-212-746-8984.

E-mail address: jtsai@med.cornell.edu (J.J. Tsai).

0042-6989/03/$ - see front matter
PII: S0042-6989(02)00510-2

are processed by coarse channels and small disparities by fine channels—the so-called size-disparity correlation (Schor & Wood, 1983; Smallman & MacLeod, 1994). Alternatively, binocular matching may proceed in a coarse-to-fine fashion, with coarse scales constraining the range within which a binocular match is sought in finer scales, thereby reducing false matches (Marr & Poggio, 1979). Other possibilities include combination of disparity information across spatial scales, via a simple or weighted average (Mikaelian & Qian, 2000; Qian & Zhu, 1997). Fleet, Wagner, and Heeger (1996) suggested that the responses per se of binocular neurons be pooled over several scales before calculating the disparity (coincidence model). The above approaches represent different ways to read out a population code.

Some of the above possibilities are comparable to vector averaging, a common method of reading-out a population code (Georgopoulos, Schwartz, & Kettner, 1986; Salinas & Abbott, 1994). A variant of vector averaging, winner-take-all (WTA), has also been proposed as a way to represent motion direction (Salzman & Newsome, 1994; Shadlen, Britten, Newsome, & Movshon, 1996) and disparity (Qian, 1994). A rather different approach posits that the activity of the population as a whole is the irreducible representation of the parameter in question. Population activity is then read out by comparing the population activity to a set of canonical responses, or templates (Heeger, 1987; Lehky & Sejnowski, 1990), one for each value of the parameter. Such distributed representations can lead to behavior that is qualitatively different from averaging, WTA, or coincidence, since the domain of quantities represented by the population may be qualitatively different from those represented by individual neurons. Population codes read out by templates can represent percepts not directly present in the input domain (e.g., "white" (Lehky & Sejnowski, 1999)) and multiple values at a single point (e.g., motion transparency (Treue, Hol, & Rauber, 2000)). Importantly, the latter work also shows that the entire shape of the population response, not simply the number of peaks or the vector average, is important in determining perceived global motion directions (see also Zohary, Scase, & Braddick, 1996).

We begin with a physiologically based receptive field (RF) model for binocular neurons (Ohzawa et al., 1990; Qian, 1994). We adopt a template-matching scheme for reading the activity of a population of neurons that vary in spatial tuning and disparity tuning. While broadly similar in spirit to the model of Lehky and Sejnowski (1990), our approach differs in the front end of the model, the nature of the template, and the formulation of threshold. We compare our model's behavior with previously reported psychophysical results, including both threshold data for simple stimuli and appearance data for complex stimuli.

2. Methods

2.1. Overview of model

For the front end of our population model, we adopt the phase-difference RF model of Qian (1994) extended to two spatial dimensions (cf. Mikaelian & Qian, 2000). A phase-difference model is supported by quantitative studies of the RF of binocular neurons (DeAngelis, Ohzawa, & Freeman, 1991; Freeman & Ohzawa, 1990; Ohzawa, DeAngelis, & Freeman, 1996). The population contains several spatial frequency channels. Each spatial frequency channel contains a family of binocular neurons tuned to the same stimulus frequency and a range of inter-ocular phase differences. As previously proposed by Qian and colleagues (Qian, 1994; Qian & Zhu, 1997; Zhu & Qian, 1996), the depth represented by the neural population activity may be taken from the identity of the maximally responsive neuron in the population, or by a weighted averaging of depths thus identified across several spatial frequency channels (Mikaelian & Qian, 2000). These approaches (WTA or vector averaging) can result in loss of detail about the pattern of the population activity. We seek to make greater use of the information in the population activity by keeping track of the responses of neurons at each location, as a function of their spatial frequency tuning and inter-ocular phase difference. This overall response is read out by comparing to a set of defined templates, each of which corresponds to a unique depth. The template we use is the expected response of the phase-difference disparity energy model (Qian, 1994) to binocular white noise with a uniform disparity. It should be emphasized that we do not suggest that neurons in V1 actually perform template-matching calculations. Instead, the population activity in V1 may feed to a later computational stage where the depth signal is extracted. Or it may be that the specific pattern of activity per se constitutes the neural correlate of depth perception (Lehky & Sejnowski, 1999).

2.2. Front end

The RF model of Qian and co-workers (Qian, 1994; Qian & Zhu, 1997; Zhu & Qian, 1996) is based on quantitative studies of neurons in cat primary visual cortex by Freeman and co-workers (DeAngelis et al., 1991; Freeman & Ohzawa, 1990; Ohzawa et al., 1990, 1996). They found that for a typical binocular simple neuron, the two monocular RFs could be described by Gabor functions. Following the work of Qian et al., we use the following 2D Gabor filters to describe the left and right RF profile of a binocular simple cell:

$$f_{\mathrm{L}}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-(x^2/2\sigma_x^2)-(y^2/2\sigma_y^2)} \cos(\omega_0 x + \phi_{\mathrm{L}}),$$
$$f_{\mathrm{R}}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-(x^2/2\sigma_x^2)-(y^2/2\sigma_y^2)} \cos(\omega_0 x + \phi_{\mathrm{R}}), \quad (1)$$

where ω_0 is the peak preferred frequency, σ_x and σ_y are the horizontal and vertical Gaussian widths that vary with ω_0, and ϕ_{L} and ϕ_{R} are the left and right phase parameters respectively. To simplify calculations, we consider only neurons tuned to the vertical orientation because they are best suited for detecting horizontal disparity. This assumption is consistent with the finding that neurons tuned to vertical orientation, as a group, can encode a larger range of disparity than neurons tuned to horizontal orientation (Anzai, Ohzawa, & Freeman, 1999a; DeAngelis et al., 1991; Ohzawa et al., 1996). The two Gaussian widths together determine the aspect ratio of the filter. In our simulations, the aspect ratio is fixed at 2. Simple cells approximately compute the sum of the left and right images filtered by the respective RF (Anzai, Ohzawa, & Freeman, 1999b; Qian, 1994):

$$r_{\mathrm{s}} = L + R = f_{\mathrm{L}} \cdot I_{\mathrm{L}} + f_{\mathrm{R}} \cdot I_{\mathrm{R}}, \quad (2)$$

where L and R are the left and right eye contributions to the linear response, I_{R} and I_{L} are the right and left images, and the symbol '$\cdot$' denotes inner product operator (Fig. 1A).

The response r_{s} is dependent on stimulus phase and hence not a reliable signal of stimulus disparity (Ohzawa et al., 1990; Qian, 1994; Zhu & Qian, 1996). However, by summing the squared responses of a pair of such simple cells in quadrature phase, a disparity energy measure that is approximately phase-independent response can be obtained (Ohzawa et al., 1990; Qian, 1994). This quadrature pair construction is closely related to the energy model for motion detection (Adelson & Bergen, 1985; Watson & Ahumada, 1983):

$$r_{\mathrm{q}} = (L_1 + R_1)^2 + (L_2 + R_2)^2, \quad (3)$$

where the phase parameters of the underlying simple cells differ by 90°:

$$\phi_{2,\mathrm{L}} = \phi_{1,\mathrm{L}} + \pi/2 \quad \text{and} \quad \phi_{2,\mathrm{R}} = \phi_{1,\mathrm{R}} + \pi/2. \quad (4)$$

At each image location, we postulate a two-parameter family of neurons (Fig. 1B). One parameter is the interocular phase difference

$$\Delta\phi \equiv \phi_{\mathrm{R}} - \phi_{\mathrm{L}}. \quad (5)$$

The other parameter is the peak frequency of the Gabor filter, ω_0. All neurons have a fixed spatial frequency tuning bandwidth of 1.5 octaves, i.e., $\omega_0\sigma_x = 2\pi(0.39)$. Thus, this population of neurons together constitutes a distributed representation of stimulus disparity.

In the simulations, the input images are large enough so that they completely cover the RF of the model neurons. Unless otherwise noted, results shown are based on the responses of neurons whose RF center is located at the center of the images.

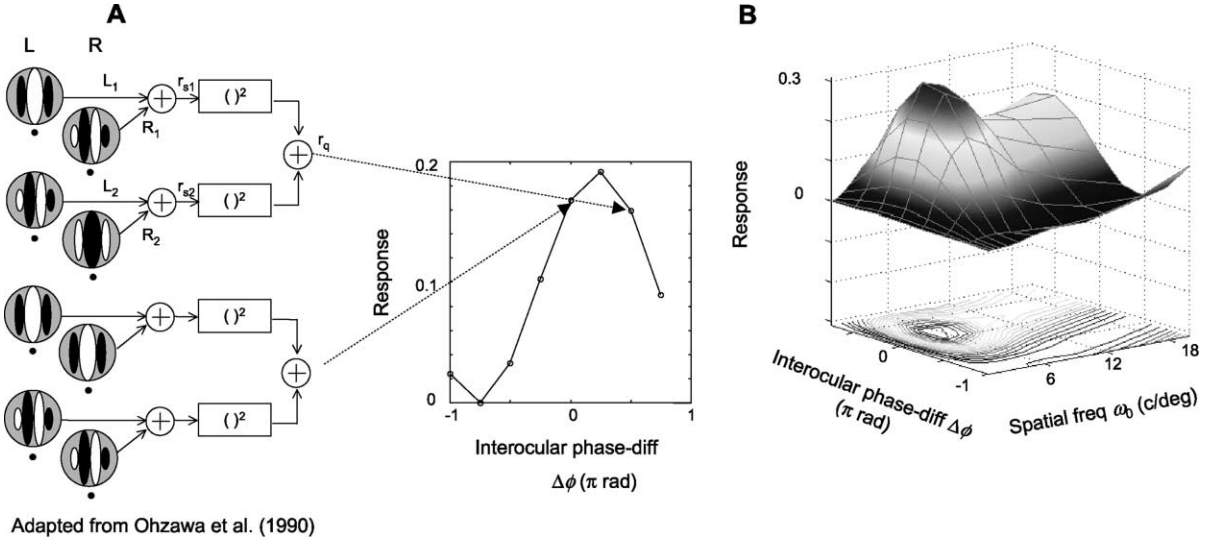

Fig. 1. A multi-spatial scale neural representation of binocular retinal disparity. (A) The front end of the model is based on the phase-difference RF model of binocular V1 neurons. Illustrated on the left are two disparity energy neurons tuned to inter-ocular phase differences ($\Delta\phi$) of 0 and $\pi/2$. The responses to an arbitrary stimulus of a family of energy neurons tuned to the same spatial frequency but different $\Delta\phi$'s are shown to the right. The responses are sinusoidal with respect to $\Delta\phi$. (B) At each visual location, families of energy neurons tuned to a range of spatial frequencies constitute a multi-scale population response. The spatial frequency shown here ranges from 0.625 to 20 c/deg, and is sampled and weighted equally at half-octave intervals. The phase difference $\Delta\phi$ is evenly sampled 8 times in $[-\pi, \pi)$. The bandwidth (spatial frequency range at half-height) of all neurons is fixed at 1.5 octaves.

2.3. Decoding population activity

To decode the disparity represented by the population activity, the population response of the quadrature units is compared to a set of templates, one for each disparity. This comparison is done in a least-squares sense. The disparity that is represented by the population activity is taken to be the one that corresponds to a template that minimizes the mismatch. Note that, in contrast to a WTA or averaging approach, this scheme has the potential to support depth transparency, since the mismatch function (f below) may have multiple minima. A single set of templates is used for all types of stimuli presented: gratings, noise, transparent, or single depth. This generic nature of the templates reflects our goal to model bottom-up computations, independent of any a priori knowledge of the type of stimulus presented.

The templates are created by calculating the averaged response of the model delineated above to the ensemble of white noise stimuli with disparity D. The rationale for using spatial noise as a template is that we are interested in addressing the question of integration across spatial frequencies, and the extent to which disparity can be determined in the absence of form information. The choice of Gaussian white noise as a template is also analytically convenient, and not as arbitrary as it might first appear. If instead we had used a correlated Gaussian noise, similar results would have been obtained. This is because, given the relatively narrow bandwidth of the filters, changing the frequency content of the template is tantamount to changing the channel weighting function w_{ch} below. We take an approach similar to that of Zhu and Qian (1996), which is based on the autocorrelation properties of white noise and the properties of the quadrature phase. This leads to the expected response of the quadrature units to white noise stimuli:

$$r_{\mathrm{q,template}} = \frac{1}{2\pi\sigma_x\sigma_y}\left[1 + e^{-D^2/4\sigma_x^2}\cos(\omega_0 D - \Delta\phi)\right], \tag{6}$$

where D is the stimulus disparity. Although similar in spirit, our result differs slightly from that of Zhu and Qian (1996), in that we have retained small terms related to the product of the monocular responses in the quadrature pair and/or assumptions about the bandwidth of the monocular filters.

The template mismatch function is defined as:

$$f(D) = \sum_{\omega_0,\Delta\phi} w(\omega_0,\Delta\phi)\left|\tilde{r}_{\mathrm{q}},\mathrm{template}(\omega_0,\Delta\phi,D) - \tilde{r}_{\mathrm{q,stimulus}}(\omega_0,\Delta\phi)\right|^2, \tag{7}$$

where w is a weighting function, and $\tilde{r}_{\mathrm{q,template}}$ and $\tilde{r}_{\mathrm{q,stimulus}}$ are template and model responses, respectively. The latter quantities are normalized raw responses $r_{\mathrm{q,template}}$ and $r_{\mathrm{q,stimulus}}$, for each combination of preferred spatial frequency ω_0 and disparity D:

$$\tilde{r}_{\mathrm{q,template}}(\omega_0,\Delta\phi,D) = \frac{r_{\mathrm{q,template}}(\omega_0,\Delta\phi,D)}{\max_{\Delta\phi}(r_{\mathrm{q,template}}(\omega_0,\Delta\phi,D))}$$

and

$$\tilde{r}_{\mathrm{q,stimulus}}(\omega_0,\Delta\phi) = \frac{r_{\mathrm{q,stimulus}}(\omega_0,\Delta\phi)}{\max_{\Delta\phi}(r_{\mathrm{q,stimulus}}(\omega_0,\Delta\phi))}. \tag{8}$$

The normalization is to ensure that the mismatch function only compares the shape, not the magnitude, of the responses. This implies that the model is insensitive to the contrast of the input.

The weighting functions that we consider can be written as the product of two terms: one that is stimulus dependent, w_{stim}, and one that is stimulus independent, w_{ch}. The stimulus dependent term is the normalized disparity energy of the stimulus; that is,

$$w_{\mathrm{stim}} = \frac{r_{\mathrm{q,stimulus}}(\omega_0,\Delta\phi)}{\max_{\omega_0,\Delta\phi}(r_{\mathrm{q,stimulus}}(\omega_0,\Delta\phi))}. \tag{9}$$

This normalization with respect to stimulus energy is akin to response normalization proposed by Heeger (1992). Its purpose is to account for the spectral content of the stimulus. The stimulus independent term is a weighting of the spatial frequency channels. In the calculations below, we will consider two such weightings: a unit Gaussian function ($\mu = 3.5$ c/deg, $\sigma = 1.2$ c/deg) that approximates the distribution of peak frequency tuning of macaque striate cortical neurons (DeValois, Albrecht, & Thorell, 1982; Mikaelian & Qian, 2000), and equal weight per octave ("$1/f$ weighting"),

$$w_{\mathrm{ch}} = \omega_0^{-1} \Big/ \sum \omega_0^{-1}, \tag{10}$$

over a finite range of spatial frequencies ω_0 (two specific ranges will be considered: from 0.625 to 20 c/deg (5.0 octaves) and from 0.625 to 7 c/deg (3.5 octaves, a range comparable to that of Gaussian weighting)). For $1/f$ weighting, the spatial frequencies are sampled at intervals of half-octave. For Gaussian weighting, the sampling interval is 1.0 c/deg.

The mismatch function represents the part of stimulus population response unexplained by the template (Fig. 2). The depth encoded by the population is taken to be the value of D at which the mismatch function is minimized, which we denote by D_{min}. If two or more local minima are present with comparable values, they suggest the presence of multiple depths. This population read-out scheme also provides a prediction of stereoacuity (depth discrimination threshold), as follows. Assume that a constant noise N coexists with the mismatch function $f(x)$, which has a minimum at x_{min}. A reasonable notion of discrimination threshold is the distance Δx that x must move from x_{min} before $f(x_{\mathrm{min}} + \Delta x)$ be-

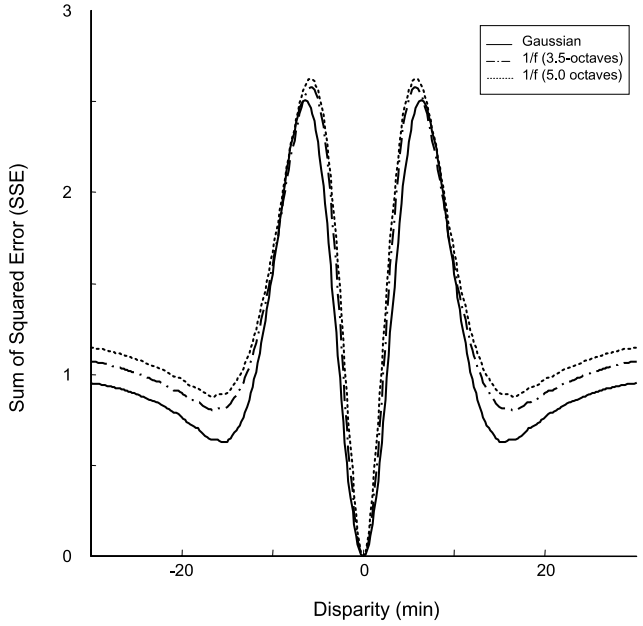

Fig. 2. Examples of template mismatch functions. The stimulus is a 3.75 c/deg vertical sinusoidal grating with zero disparity. The solid curve is calculated with a spatial frequency weighting (w_{ch}) that is proportional to the distribution of peak spatial frequency selectivity of neurons in macaque V1 and is approximated by a Gaussian distribution ($\mu = 3.5$ c/deg, $\sigma = 1.2$ c/deg; see text). The dotted curve is obtained with a uniform weighting of spatial frequency on a log scale ($1/f$ weighting) over the range 0.625–20 c/deg. The dashed-dot curve is obtained with an abbreviated $1/f$ weighting from 0.625 to 7 c/deg (i.e., a range comparable to the Gaussian weighting). For illustration purposes, the peak frequency weight is normalized to unity, so that the template mismatch functions will have approximately the same heights. The overall shape of the function is similar regardless of the weighting, reaching a minimum at the disparity of the stimulus.

comes distinguishable from $f(x_{\min})$ according to some criterion α, i.e., the value of Δx for which $f(x_{\min} + \Delta x) = f(x_{\min}) + \alpha N$. To determine Δx, we use the Taylor's expansion of f around $x_{\min}$:

$$f(x_{\min} + \Delta x) = f(x_{\min}) + f'(x_{\min})\Delta x + \tfrac{1}{2}f''(x_{\min})(\Delta x)^2 + \cdots \tag{11}$$

Since f has a minimum at $x_{\min}$, the derivative $f'(x_{\min})$ vanishes. Assuming that higher order terms are negligible and setting the left side of the equation equal to $f(x_{\min}) + \alpha N$, we can solve for Δx:

$$\Delta x = \sqrt{\frac{2\alpha N}{f''(x_{\min})}}. \tag{12}$$

Thus, the threshold is proportional to the reciprocal of the square root of the second derivative of the mismatch function at $D_{\min}$ (units: disparity). We therefore adopt this quantity as an index for comparison with psychophysical measures of depth discrimination threshold. This approach has the advantage of not requiring additional model parameters to generate predictions of relative thresholds for comparison between stimuli. A caveat is that the absolute magnitude of threshold cannot be directly compared between model and data because (1) threshold prediction of the model is dependent on the scaling of the template mismatch function, and (2) the model does not include sensory or decision-level noise sources or a decision criterion, which would contribute to psychophysical thresholds. For the same reason, model predictions of threshold can only be compared within, and not across, weighting schemes.

3. Results

We compare model predictions with two kinds of experimental data: (1) threshold phenomena, including spatial frequency dependence of stereoacuity (Legge & Gu, 1989; Schor & Wood, 1983; Smallman & MacLeod, 1994), dependence of depth increment threshold on disparity pedestal (Badcock & Schor, 1985; Blakemore, 1970; McKee, Levi, & Bowne, 1990; Rohaly & Wilson, 1993; Siderov & Harwerth, 1993b), and dependence of perceived match on spatial frequency (Schor & Wood, 1983), and (2) appearance of complex stimuli, including depth averaging and transparency (Parker & Yang, 1989; Rohaly & Wilson, 1994; Stevenson, Cormack, & Schor, 1989), and compound gratings in depth (Boothroyd & Blake, 1984). As detailed below, model predictions are in accord with experimental data for this wide variety of phenomena. The comparisons with threshold psychophysics primarily test the parameters chosen for the model's filters. The comparisons with appearance phenomena primarily test the notion that the population model can account for richer phenomena than WTA or vector averaging.

3.1. Dependence of stereoacuity on spatial frequency

In agreement with the "size-disparity correlation" hypothesis, psychophysical experiments have demonstrated that the range of disparity over which stereopsis is supported increases with the spatial period of a narrow-band stimulus (Schor & Wood, 1983; Schor, Wood, & Ogawa, 1984b; Smallman & MacLeod, 1994). Stereosensitivity increases with the spatial frequency of difference-of-Gaussian (DOG) stimuli, according to a constant 6 deg phase shift for frequencies less than 2.5–4 c/deg (Schor & Wood, 1983; Schor, Wood, & Ogawa, 1984a; Siderov & Harwerth, 1993a, 1995). Legge and Gu (1989) reported similar results using low-contrast sinusoidal grating stimuli: a linear increase in sensitivity as spatial frequency increases from 0.1 to ~3 c/deg. Other investigators have suggested that disparity tuning scales with spatial frequency up to even higher frequencies (Mckee, Glennerster, & Harris, 1996; Smallman & MacLeod, 1994). Taken together, the evidence

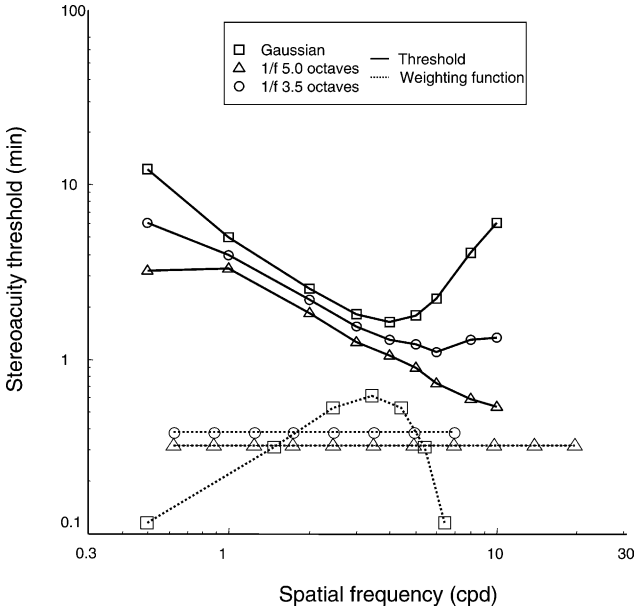

Fig. 3. Stereoacuity thresholds plotted as a function of spatial frequency of one-dimensional sine grating stimulus. Thresholds corresponding to three frequency weighting functions are shown displaced vertically for clarity (solid curves). The dotted curves at the bottom of the plot illustrate the relative magnitude of the frequency weighting functions. For the Gaussian (squares) and the 3.5-octave $1/f$ (circles) weighting, threshold plateaus or rises as the stimulus frequency exceeds the range implemented by the model. Threshold decreases monotonically with frequency for the 5.0-octave $1/f$ weighting (triangles). Slope of the best fitting line through the linear segment is −0.72, −0.81 and −0.98 for the 3.5-octave, 5.0-octave $1/f$, and Gaussian weighting respectively. The approximately reciprocal relationship between threshold and spatial frequency is consistent with psychophysical data.

for a linear relationship between stereo sensitivity (the reciprocal of disparity threshold) and spatial frequency below 2–4 c/deg is robust.

In the low-to-mid frequency regime, our model also demonstrates an approximately inverse relationship between threshold and spatial frequency. Fig. 3 (solid curves) shows the threshold as a function of spatial frequency of a sinusoidal grating stimulus, parametric in channel weighting functions (w_{ch}) (see Section 2.3; the dotted curves indicate the relative magnitude of the three weighting functions). In all three cases, threshold decreases with spatial frequency in the middle of the spatial frequency range. In the log–log plot shown, the slope of this relationship is approximately −1, indicating a reciprocal relationship between spatial frequency and threshold. This result is consistent with experimental findings. Legge and Gu (1989) found that the slope of threshold vs. frequency curve measured using sine grating stimuli ranged from −0.98 to −1.29 in their subjects. Schor and Wood (1983) found slopes close to unity using DOG stimuli. Siderov and Harwerth (1993a) reported a range from −0.72 to −0.89, also with DOGs. Our model, using either $1/f$ or Gaussian weighting, results in threshold functions with best-fit slope ranging from −0.72 to −0.98.

Not surprisingly, when the stimuli do not lie within the frequency range represented by the population, stereoacuity is worse than predicted by extrapolation of this approximate linear relationship. This can be seen in the two curves corresponding to the channel weighting functions in which the highest spatial-frequency is limited to about 7 c/deg. The deviation from a linear trend at high frequencies has been reported (Legge & Gu, 1989; Schor & Wood, 1983; Siderov & Harwerth, 1993a, 1995). These investigators find that the threshold does not continue to decrease; instead above some frequency, it either plateaus or increases. This high-frequency knee in the threshold function has led to suggestions of either a high (Schor et al., 1984b) or low (Kontsevich & Tyler, 1994) frequency limit in the mechanisms subserving stereopsis. Our simulations suggest that a high-frequency drop-off in the representation of disparity signal can produce an upturn in the threshold function, but it does not rule out other explanations. Additionally, at high spatial frequencies, even a small amount of retinal jitter creates large monocular phase differences, causing increased binocular matching ambiguity in the case of periodic stimuli and hence poor stereoacuity (Legge & Gu, 1989). This may explain why the point of reversal of the threshold function occurs at a lower spatial frequency in human observers than in our model.

3.2. Dependence of depth increment threshold on standing disparity

Stereopsis allows fine depth localization (comparable to disparities in the hyperacuity range) when the target object is near the fixation plane (McKee et al., 1990). However, with increasing distance of targets from the fixation plane, depth increment threshold rapidly increases. Some investigators have reported that the threshold increases exponentially with standing disparity (Blakemore, 1970; Schumer & Julesz, 1984; Siderov & Harwerth, 1993b, 1995), while others have found the shape of the threshold function to have two segments: an exponential increase up to 20′ standing disparity, followed by a more gradual rise (Badcock & Schor, 1985; McKee et al., 1990; Rohaly & Wilson, 1993).

Fig. 4 shows our simulation results of threshold as a function of standing disparity of a Gabor patch stimulus, parametric in center frequency. The bandwidth of the Gabor patch is 1.75 octaves. Three channel weightings are used: $1/f$ weighting from 0.6 to 20 c/deg (Fig. 4A), $1/f$ weighting from 0.6 to 7 c/deg (Fig. 4B), and Gaussian weighting from 0.5 to 6.5 c/deg ($\mu = 3.5$ c/deg and $\sigma = 1.2$ c/deg) (Fig. 4C). The threshold function increases with disparity pedestal. We focus on the 2 c/deg stimulus results for comparison with experimental data. The regression slope of the increment threshold function

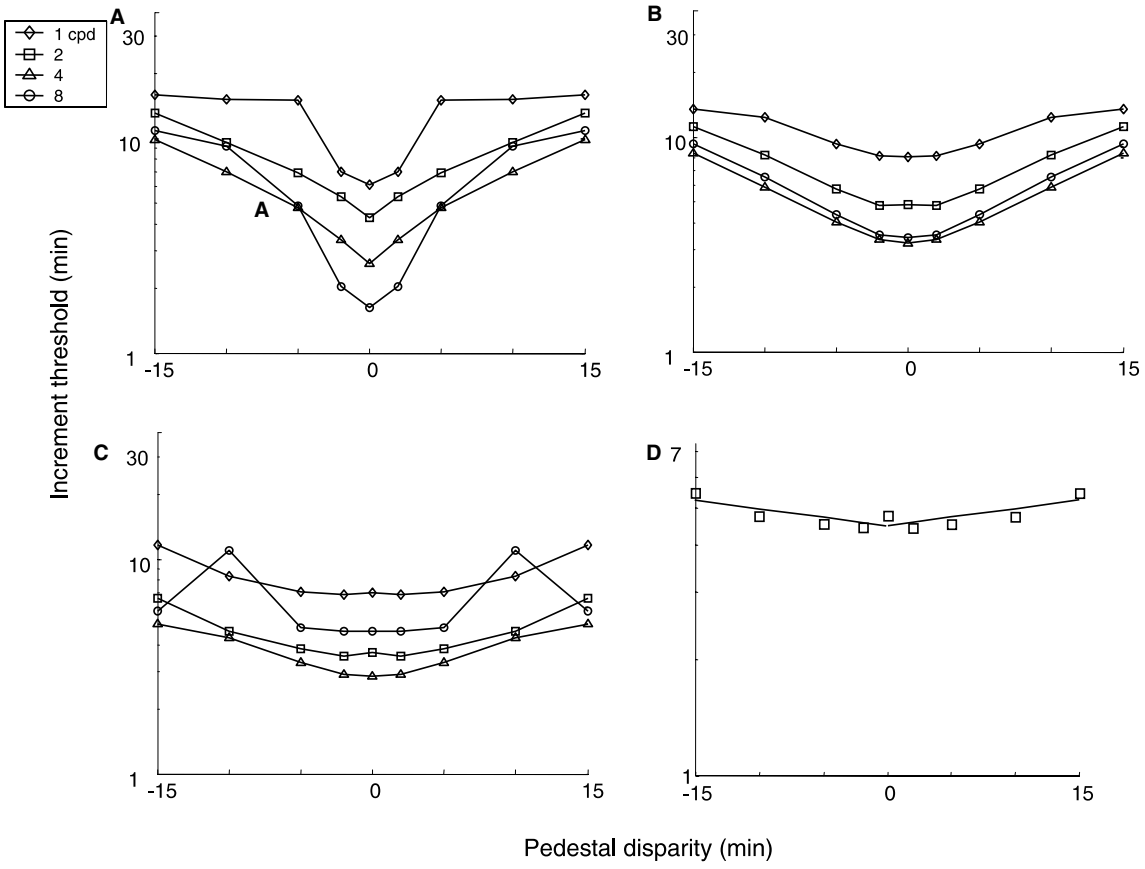

Fig. 4. Disparity increment thresholds plotted as a function of pedestal disparity of 1.75-octave one-dimensional Gabor patch stimuli, for a range of center frequencies (1 c/deg, diamonds; 2 c/deg, squares; 4 c/deg, triangles; 8 c/deg, circles). Three frequency channel weightings are illustrated: (A) 5.0-octave $1/f$, (B) 3.5-octave $1/f$, (C) Gaussian weighting. At the horopter, threshold increases with decreasing stimulus frequency. Away from the horopter, threshold increases as a function of pedestal disparity. Slopes (95% CI) of the best-fit lines (not shown) for the 2 c/deg stimulus are 0.07 (0.06–0.09), 0.06 (0.04–0.08), and 0.04 (0.02–0.06) for the three weightings, comparable to experimental data. (D) Thresholds calculated using a model that consists of only one channel with peak frequency of 2 c/deg. The stimulus is a 1.75-octave Gabor patch with center frequency of 2 c/deg. Note the change in y-axis scale and the much shallower slope, 0.01 (−0.005 to 0.03), compared to the multi-channel results in A–C.

from our simulations ranges from 0.04 to 0.07. In comparison, reported values have been variable; for example: Badcock and Schor (1985) found slopes ranging 0.08–0.12, or a 5- to 10-fold increase in threshold as pedestal increases from 0′ to 20′, using 1.75 octave DOG stimuli. Siderov and Harwerth (1993b, 1995), using the same stimuli but a paradigm that eliminated width cues, obtained slopes ranging 0.034–0.06 over a 60′ disparity range. Rohaly and Wilson (1993), using 1.0-octave D6 stimuli, found a slope of 0.17 or a 2-fold increase over a much smaller range (0–4′). While it is not clear why the numbers differ across studies, the results of our model lie within the range reported.

To the extent that the threshold functions can be fitted by straight lines, their slopes are not significantly different for Gabor patches of different frequencies (Fig. 4). This is in accord with the majority of reports: threshold functions have the same slope for narrowband stimuli above 0.15 c/deg (Badcock & Schor, 1985; Rohaly & Wilson, 1993; Siderov & Harwerth, 1993b, 1995; but see Smallman & MacLeod, 1997).

Previously, Qian and Zhu (1997) have suggested that a phase-difference RF model might account for the increase in increment threshold with pedestal. According to their explanation, fine-scale filters do not contribute to disparity computation at large disparities. If precise disparity information is encoded by fine-scale channels, their removal would result in a larger threshold. Fig. 4D shows thresholds calculated using only one frequency channel (tuned to 2 cpd) of our population model for a 2 cpd Gabor stimulus. The slope of the function, though slightly positive, is not significantly different from zero. Thus the pedestal effect in a single scale model is very small, compared to the robust effect in a multi-scale model and experimental data.

Our simulations confirm that a multi-scale phase-difference model can indeed account for the threshold function. The sensitivity of the filter is reduced as stimulus disparity increases beyond the envelope size of the filter. This is reflected in the decreased response modulation with respect to $\Delta\phi$ in the templates as disparity D increases (Eq. (6)). A template-match read-out

scheme can integrate the response across spatial scales to produce a result consistent with psychophysics.

3.3. Dependence of perceived depth on spatial frequency

Schor and Wood (1983) measured the disparity of a DOG pattern that matched the apparent depth of a thin bar. They found that, for suprathreshold disparities, the matching disparity of the DOG pattern equals that of the target when the frequency of the DOG is above 0.6 c/deg. Below this frequency, the DOG requires a greater disparity to match to the apparent depth of the bar; i.e., the DOG pattern appears closer to the fixation plane than the bar. Expressed as a ratio of the bar disparity to DOG disparity, apparent depth is most reduced for low spatial frequencies and small target disparities. Schor and Wood concluded that coarse spatial channels were tuned to large disparities and were less efficient in signaling small suprathreshold disparities than fine spatial channels. Subsequently, Schor and Howarth (1986) confirmed the uncrossed depth bias associated with low-spatial frequencies: crossed disparity was needed to make depth matches to targets with zero disparity. The bias is greatest for low contrast and long stimulus presentation, suggesting contrast fading of low-spatial-frequency stimuli contributes to the bias. However, even when the stimulus presentation is pulsed and the perceived contrast is equated, a residual depth error remains, indicating that contrast cannot account for the entire effect. These two studies suggest that there are opposing pictorial and binocular effects operating when the target has an uncrossed disparity. Low contrast and low spatial frequency of the stimulus cause a constant error in the uncrossed direction (Brown & Weisstein, 1988; Frisby & Mayhew, 1978; Schor & Howarth, 1986). These factors likely operate as pictorial cues to figure-ground segregation (O'Shea, Blackburn, & Ono, 1994; O'Shea, Govan, & Sekuler, 1997). On the other hand, the depth signal mediated by coarse spatial channels is less efficient than that by fine channels, leading to a decreased apparent depth, as was observed when depth matching experiments were done at high stimulus contrast and low spatial frequency (Schor & Wood, 1983). Therefore, monocular cues cannot account for the reduced apparent depth of low-spatial-frequency stimuli reported by Schor and Wood (1983), particularly for uncrossed disparities, although they may explain the greater reduction of apparent depth for *crossed* disparities.

Our model accounts for the binocular component of the interaction between spatial frequency and apparent depth, namely, the reduced apparent depth of a low-frequency stimulus compared to a high-frequency stimulus at high contrast. Fig. 5 shows the ratio of the disparity calculated by the model to the disparity physically present in the stimulus as a function of the disparity of 1.75 octave Gabor patterns, parametric in stimulus center spatial frequency. Results are based on the three frequency weighting functions described in Section 2.3. The general pattern of results is comparable to the findings of Schor and Wood (1983) using DOG patterns (their Fig. 3). If the stimulus was perceived at the veridical depth, then the disparity ratio curves would be flat with a value of one. Instead, for the lowest fre-

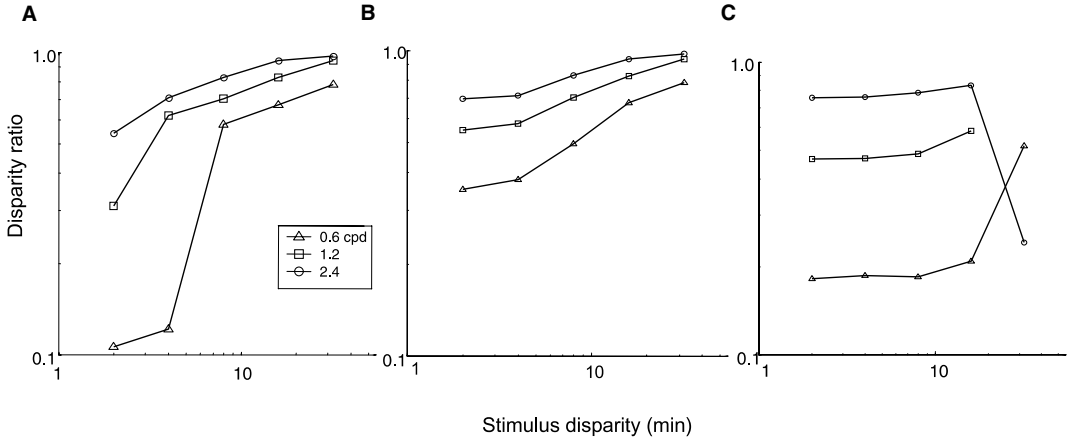

Fig. 5. Spatial frequency dependence of disparity bias. The ratio of the computed disparity to actual stimulus disparity is plotted against the stimulus disparity. The stimulus is a 1.75-octave one-dimensional Gabor patch (2.4 c/deg, circles; 1.2 c/deg, squares; 0.6 c/deg, triangles). (A) 5.0-octave $1/f$; (B) 3.5-octave $1/f$; (C) Gaussian frequency weighting. If there were no bias in the disparity calculation, then all ratio values would be 1. Instead, consistent with psychophysical depth-matching experiments, the model shows that at small disparities and low spatial frequencies, the apparent disparity is smaller than the actual stimulus disparity. The ratios approach unity for high frequency and large disparity. In agreement with psychophysical data, both $1/f$ weighting schemes demonstrate an upward trend in the slope of the disparity ratio curve as stimulus frequency decreases. In contrast, the Gaussian weighting does not show this pattern, and also produces a disparity value of incorrect sign for the 1.2 c/deg stimulus (data point not shown).

quency Gabor stimulus, there is the greatest reduction in perceived depth for small disparities. The disparity ratio approaches one as the stimulus disparity and frequency increase. The simulations using $1/f$ weighting of spatial frequency channels (Fig. 5A and B) capture another feature of the psychophysical data: the slope of the disparity ratio vs. stimulus disparity functions increases as the DOG spatial frequency decreases. In this respect, the Gaussian weighting (Fig. 5C) is less consistent with the data. It is also worth noting that model results are independent of the sign of the stimulus disparity, while psychophysical data for crossed disparity show greater bias effect than for uncrossed disparity. As discussed above, monocular cues may contribute the sign-dependent difference in the perceived depth-bias. The model addresses only binocular mechanisms and therefore is symmetrical with respect to the sign of disparity.

3.4. Stereoscopic depth transparency

How might a multi-scale model deal with stimuli in which Fourier components carry different disparity signals? Such stimuli may arise from transparent overlapping surfaces at different depths. Superposition of vertical sine gratings of different spatial frequencies, under certain conditions, appears transparent in depth (Boothroyd & Blake, 1984; Rohaly & Wilson, 1994). Superposition of two uncorrelated random-dot stereograms (RDSs) may also appear transparent in depth (Akerstrom & Todd, 1988; Prazdny, 1985; Weinshall, 1991). From the viewpoint of multi-scale analysis of disparity signals, these examples have an important difference. In a transparent compound grating, the discordant depth signals are carried in distinct frequency channels, but are spatially homogeneous. Conversely, a transparent RDS is homogeneous in the frequency domain, but depth signals are spatially inhomogeneous. That is, distinct disparity signals cannot be separated by the spatial frequencies that carry them. Thus, these two classes of stimuli are likely to elicit different neuronal population responses and different kinds of behavior from computational models.

3.4.1. Broadband case

When two independent RDSs with different disparities are summed, three percepts are possible: a single surface at the averaged disparity, a thickened slab, or two transparent surfaces at different depths (Parker & Yang, 1989; Stevenson et al., 1989). Generally, the single-surface percept occurs with the smallest disparity difference, and the transparent-surfaces percept occurs with the largest difference. In the latter percept, the disparity gradient between nearby points may be steep. Stereo models that assume smoothly varying fronto-parallel surfaces by imposing an inhibitory interaction between dissimilar disparities (e.g., Marr & Poggio, 1976) would likely not predict this behavior. Another class of models retains the facilitative interaction between nearby inter-ocular matches that satisfy a disparity gradient limit but eliminates the inhibitory interaction (Pollard, Mayhew, & Frisby, 1985; Prazdny, 1985). The PMF model assigns a unique disparity to each location in the stimulus. The collections of points at different disparities constitute overlapping depths. Nearby points may have dissimilar disparities as long as they are within a certain disparity gradient limit, a model parameter (Pollard & Frisby, 1990), but the value of this parameter required to account for transparency depends on the stimulus configuration (Weinshall, 1991), and disagrees with that found in psychophysical studies (Burt & Julesz, 1980). Finally, these models require the identification of image tokens, such as edges or zeros-crossings of second-derivatives, which are used for inter-ocular matching. However, stereopsis is possible with smoothly varying images without localized features (Arndt, Mallot, & Bulthoff, 1995). The visual system likely does not require explicit feature matching to solve the correspondence problem, as stereograms comprised of features finer than the foveal RF size are effective in generating depth perception (Qian & Andersen, 1997). In contrast, such images pose no difficulty to models such as the present one, which operate directly on the input. We will show that our model behaves similarly to human performance as the disparity difference between two random-dot surfaces increases. Moreover, the model's behavior with transparent grating stimuli is also in general agreement with psychophysics.

Since RDS stimuli are stochastic, in order to arrive at a consistent response over a patch of the stimulus, we consider the average response over the patch. For a sufficiently large patch at a single disparity, the averaged response approximates the ensemble mean given by Eq. (6), namely the template for that disparity. A transparent RDS consists of the sum of two uncorrelated RDSs with distinct disparities. Since the components of the RDS are uncorrelated, the mean population response to the transparent RDS is the sum of the mean responses to the components. Fig. 6 shows the mismatch functions obtained by comparing the template (using 5.0 octave, $1/f$ weighting) against the mean response to transparent RDSs of increasing disparities about the fixation plane. At disparity separation (ΔD) below $72''$, the function has a prominent minimum at zero. Parker and Yang (1989) found that, in two subjects, the maximum disparity that supports disparity averaging at the fixation plane is about 80 s. For ΔD of 84–$120''$, the model does not find a good match at any disparity, as indicated by the relatively broad minima of the mismatch function. In this regime, the threshold, 2.2–$4.2'$, is much greater than that ($0.43'$) obtained when disparity averaging is observed. Additionally, the "best" match is far from the disparity signals actually present in the stimulus. These findings

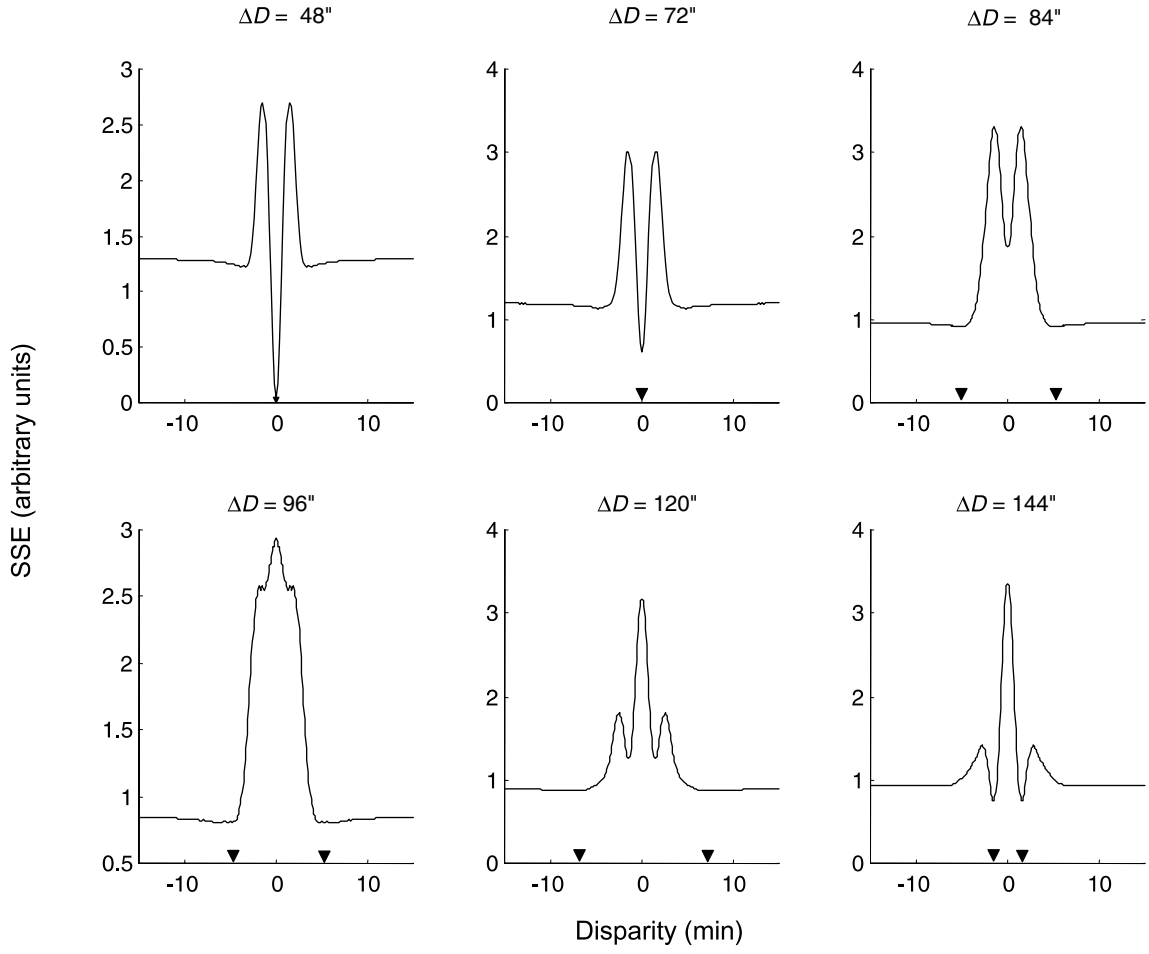

Fig. 6. Template mismatch functions for transparent RDSs of two surfaces at increasing disparity offset (ΔD). See text in Results for details of calculation. In this and subsequent figures, we only use the 5.0-octave $1/f$ weighting. For small disparity offset, the mismatch function has a sharply defined minimum (arrowhead), indicating a single surface. For an offset of between 72 and 84″, a transition occurs from a single minimum to two broad minima. The latter may correspond to a "thickened" appearance observed psychophysically. This is consistent with the range for depth averaging reported by Parker and Yang (1989). At larger offsets, a transition to two sharp minima occurs, corresponding to a percept of transparency. This is in line with the range of gap resolution thresholds reported by Stevenson, Cormack, and Schor (1991).

suggest that the model cannot find a good template match to a discrete number of unique disparities for this range of disparity separation, which corresponds to the reported "thickened slab" appearance. For ΔD above 120″, two well-defined minima appear in the mismatch function, which we interpret as the correlate of depth transparency. Stevenson et al. (1989) measured the sensitivity for detecting a gap in depth between two random-dot surfaces. They found thresholds ranging from 136″ to 351″ for three subjects. Thus, over a range of disparities, our model manifests three modes of behavior in a manner similar to human perception.

3.4.2. Narrow-band case

Rohaly and Wilson (1994) performed depth matching experiments using the sum of two vertical sine gratings of different spatial frequency and disparity. They found that the percept was a single surface at a depth between those of the two component gratings when the relative grating frequency difference was less than 3.5 octaves. Beyond this limit, two transparent surfaces were observed. We submit the same stimuli to our model. Fig. 7 illustrates the template mismatch functions associated with the sine gratings alone (top row) and sums of gratings at 1, 2, or 3 octaves apart (bottom row). In all three cases, the lower frequency grating has zero disparity while the higher frequency grating has a 2′ disparity. In each plot, the value of $D_{\min}$ is indicated next to the minimum. The functions for the compound gratings show clearly a global minimum at a disparity intermediate between the best-match disparities to the individual gratings alone. Interestingly, this disparity is biased toward that of the higher frequency grating, a tendency also observed in psychophysical data (Fig. 6 of Rohaly & Wilson, 1994).

The model also predicts transparency for 4-octave compound gratings when they are separated by a large disparity difference (Fig. 8). In this example, the low-

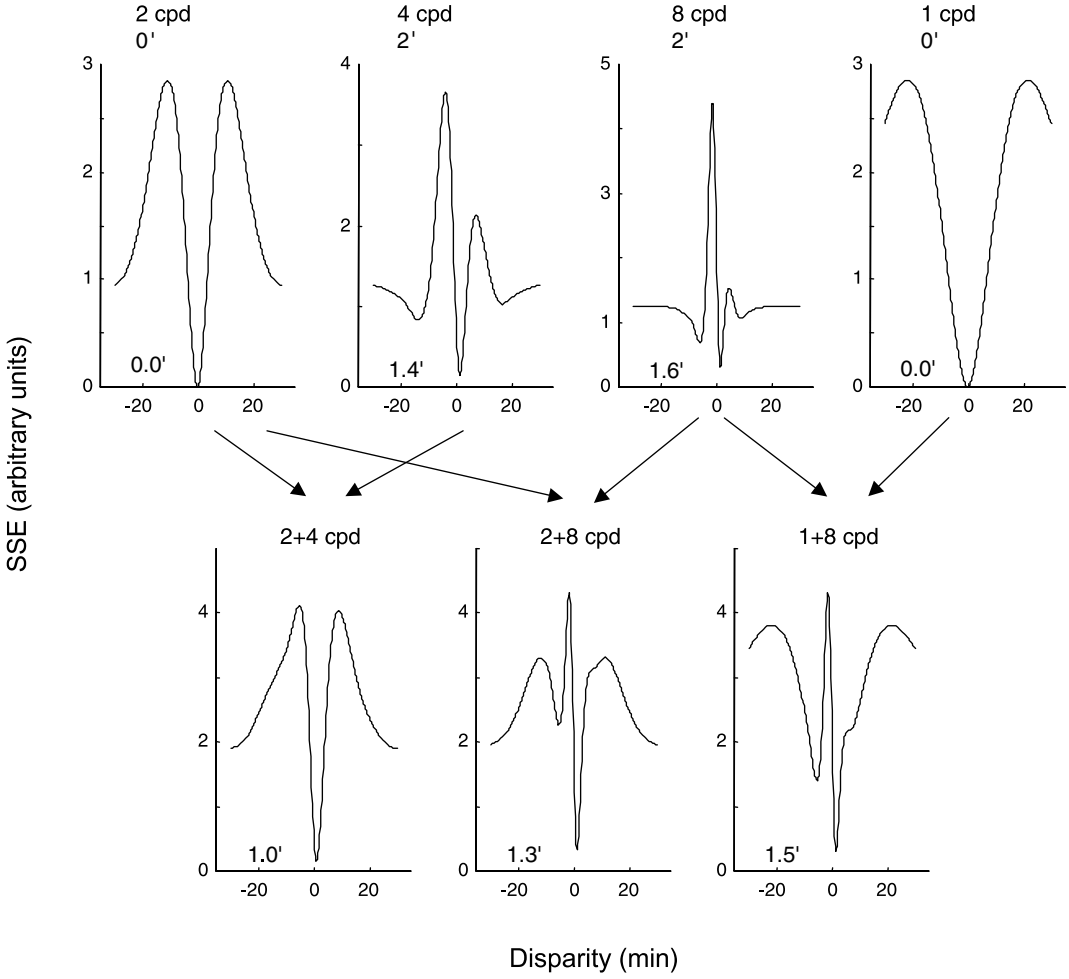

Fig. 7. Disparity averaging across spatial scales. Template mismatch functions for sine luminance grating stimuli of several spatial frequencies and disparities (top) and various pairwise sums in peaks-add phase relationship (bottom). The location $D_{\min}$ of the minimum of the mismatch function (in minutes) is indicated on each plot. As previously noted, the matched depth $D_{\min}$ of low-frequency stimuli is less than the veridical disparity (see Section 3.3 and Fig. 5). For frequencies that differ by less than three octaves, the disparity of the compound grating lies between those of the component gratings.

frequency grating has a disparity of 10′ while the high-frequency grating has disparity of 1′ (top row). The mismatch function for the 2 cpd grating has two comparable minima because its disparity is close to half of its spatial period, a condition that leads to ambiguous matching. Results for the compound stimuli (bottom row) show that the 2-octave compound grating has a mismatch function largely similar to the 2-cpd component, except for a small additional dip. Therefore, the model prediction is a single depth similar to that of the component grating. The mismatch function of the 4-octave compound grating has two minima that correspond closely to the disparities of the component gratings. These minima do not have equal offset from a "perfect match" (zero on the ordinate), which may reflect the relative strength of the disparity signals. Rohaly and Wilson reported variability amongst the subjects in the magnitude of their perceived depths. Only 1 out of 3 subjects was able to make reliable depth judgments in the 4.0-octave (transparency) condition. The model deviates from psychophysical data when the components are separated by 4.0 octaves and a small disparity difference. In this case, human subjects perceive transparency, while the model predicts a single disparity closer to that of the higher frequency component (not shown). These inconsistencies may be related to how the visual system interprets the template mismatch function when a global minimum is not clear-cut.

Compound gratings with small spatial frequency differences can produce a percept other than averaging and transparency. Boothroyd and Blake (1984) showed that compound grating patterns consisting of components with different disparities separated by less than 1.0 octave were perceived as a corrugated surface rather than a single plane at an average depth. The nature of the multiple-depth percept in this case is qualitatively different from that of a transparent compound grating in which the components are widely separated in spatial

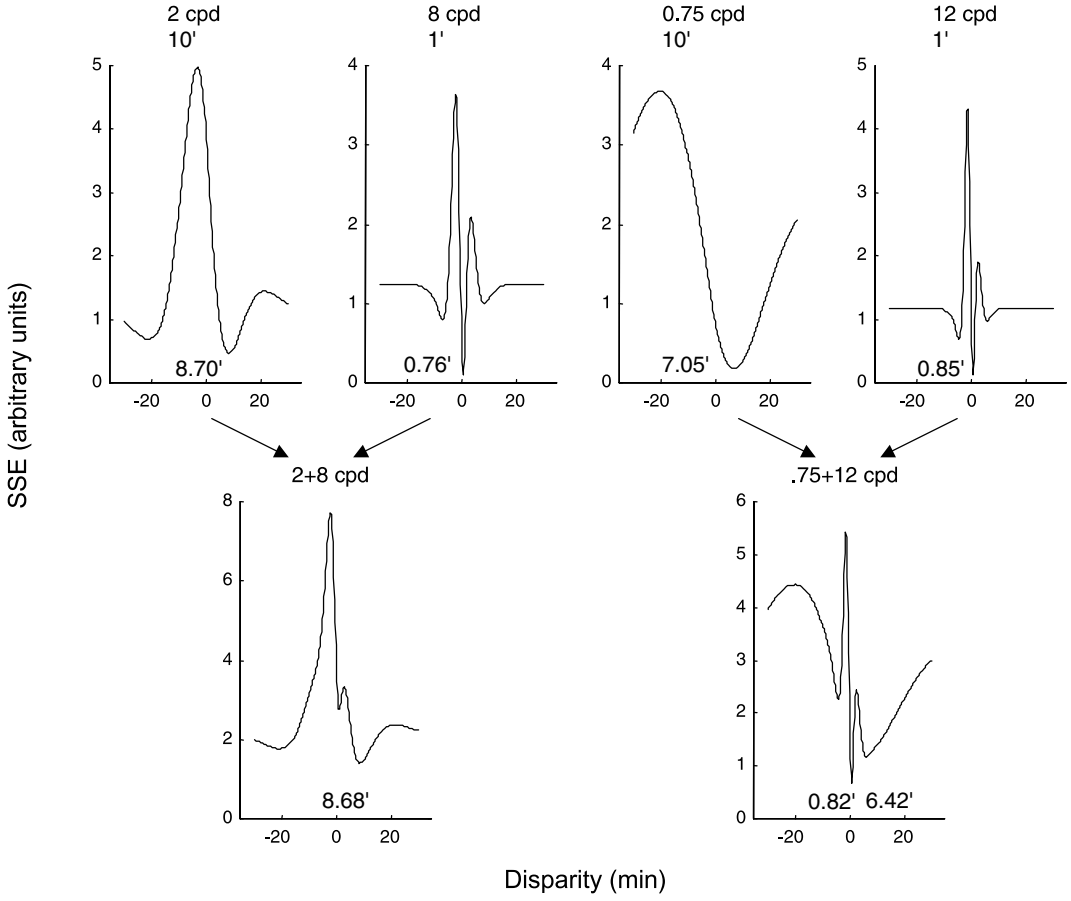

Fig. 8. Transparency in broadband sum-of-sinusoid stimuli. Template mismatch functions for sinusoid and sum-of-sinusoid stimuli are shown, along with the location of the local minima of the function indicated next to them. Stimulus configuration and plotting conventions as in Fig. 7 but with a larger disparity difference and spatial frequency range of the component gratings. For the 2-octave compound grating (bottom left), the function is similar to that of the 2-cpd grating. For the 4-octave stimulus (bottom right), two comparable minima are seen.

frequency. In the latter, planes at different depths appear continuous and overlapping; that is, multiple depths are perceived at the same spatial position. In the former, different portions of the visible edges of the compound waveform, corresponding to high- and low-contrast regions, mediate the multiple depths (Fig. 9A). Our model identifies two predominant disparities at distinct locations in the stimulus (Fig. 9B). The high-contrast regions have a disparity closer to the fixation plane than the low-contrast regions, and the disparity is between that of the component gratings.

In summary, the model accounts for the spatial frequency dependence of the perceived depths of compound gratings: for stimuli whose components have a small frequency difference, the interaction between components can generate periodically varying depth, but a unique depth at each spatial position (corrugation). With an intermediate difference, the percept is a single flat surface. With a large difference, the percept is transparency. The model also predicts the appearance of stimuli composed of sums of RDSs with different disparities. As it stands, the model does not predict precisely when local minima are interpreted as multiple depths, versus when only the global minimum is perceived as a single depth. What the model does predict is the qualitative change in percept from a single depth to transparency in a family of narrow-band stimuli, and the three qualitative percepts seen (above) for broadband stimuli.

3.5. Disparity interaction in compound gratings

Yet another percept occurs when the component gratings have the same disparity. Boothroyd and Blake (1984) measured the perceived depth of such sum-of-sinusoids stimuli, in which the components have a fixed phase relationship. These waveforms have energy at two harmonics, but have distinct contrast edges at a "beat" frequency (a frequency equal to the difference of the two fundamentals). For a periodic stimulus, disparities greater than a half-cycle are "wrapped-around" and perceived as being within one half-cycle from the fixation plane (Boothroyd & Blake, 1984; McKee & Mitchison, 1988). Since the beat frequency is less than

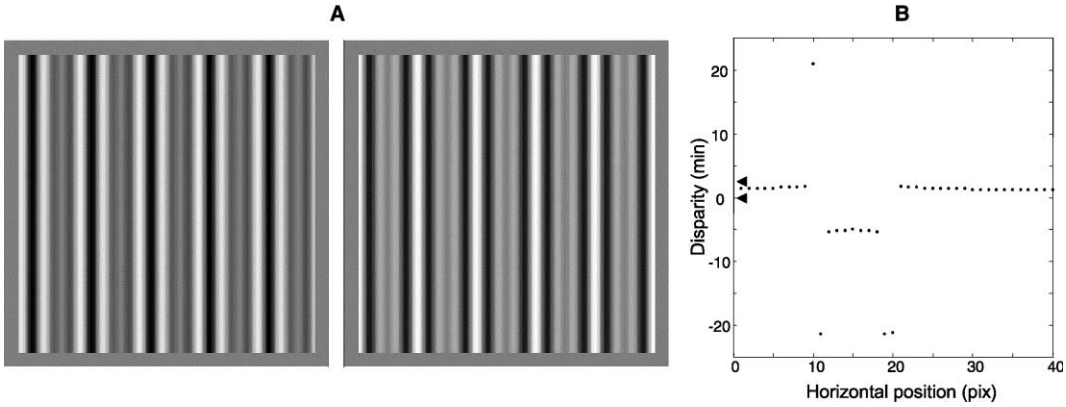

Fig. 9. Compound gratings with different component disparities. (A) Stereo pair of a 4.5 c/deg grating with 3′ disparity added to a 3.0 c/deg grating with 0′ disparity. The resulting waveform is perceived as corrugated in depth—with the high-contrast region appearing at one depth and the low-contrast region at another (adapted from Boothroyd & Blake, 1984). Neither depth corresponds to the disparity of the components. (B) The disparity calculated at each pixel location. The abscissa represents one period of the compound grating waveform. Two predominant disparities may be seen: +1.5′ and −5.5′. Consistent with psychophysics, the calculated disparities differ from the disparity of the components (arrowheads). The disparity of the high-contrast regions is between those of the two component gratings. The disparity magnitude of the low-contrast regions is greater than that of the high-contrast regions.

either of the two fundamentals, its spatial period is longer than that of either. Consequently, with a compound grating stimulus, it is possible to introduce disparities greater than a full period of either of the fundamentals. Boothroyd and Blake asked the question: when a stimulus has conflicting disparity information carried by its components and by the visible edges, which one is perceived? They found that subjects perceived the disparity mediated by the contrast edges of the compound grating (corresponding to the beat frequency) if the relative contrast of the fundamentals is equal. Similar experiments using sum-of-sinusoid waveforms have been done to study motion perception and show that perceived direction can follow either the "beat" pattern or the Fourier components (e.g., Hammett, Ledgeway, & Smith, 1993). Separate mechanisms (e.g., "long-range" vs. "short-range") have been proposed to account for the perception based on contrast edges and on Fourier motion energy (Braddick, 1980). Here we show that our model provides a unified account for depth perception mediated by contrast edges of compound gratings and by component gratings.

Following Boothroyd and Blake, we performed simulations with 6 and 9 c/deg sine gratings and their sum. When the disparity of these stimuli is small, the perceived depth is unambiguous and corresponds to the nominal disparity. The simulations are consistent with this finding. When the disparity is large, the perceived depth of the compound grating results from an interaction of the disparity of the components. Fig. 10 shows the population response (*top*) to the three waveforms each presented with a 14′ disparity and the corresponding template mismatch functions (*bottom*). The 6 c/deg grating, presented alone, has a best-match disparity of 3.7′, corresponding to the nominal 14′ disparity because of its 10′ periodicity. Likewise, the best-match disparity for the 9 c/deg grating is 0.5′, corresponding to the nominal 14′ disparity because of its 6.67′ periodicity. The best-match disparity for the compound grating is −5.9′, corresponding to its nominal 14′ disparity (because of a periodicity of 20′). Our model results are in accord with the findings of Boothroyd and Blake [1] (their Figs. 9 and 10). Significantly, the perceived disparity of the compound grating is not a weighted average of the disparities of its fundamentals, but rather is equivalent to the disparity of the beat frequency. This beat frequency coincides with the periodicity of edge-like elements in the compound grating stimulus, which cannot be detected by linear mechanisms. The perceived disparity may be arrived at in several ways: (1) recovery of non-linear features followed by matching (e.g. Wilson, Ferrera, & Yo, 1992), (2) explicit intersection-of-constraint (IOC) calculation (Adelson & Movshon, 1982), or (3) examination of activity across spatial scales without an explicit IOC calculation. Our approach follows the third alternative and closely parallels template-match models for motion extraction (Heeger, 1987; Heeger & Simoncelli, 1993; Victor & Conte, 1994). These models are distinguished from IOC models in that they may be non-veridical, and do not require accurate

[1] Boothroyd and Blake commented that under their experimental conditions, crossed disparities were "less stable and harder to match" than uncrossed disparities. Perhaps for this reason, the compound grating was consistently matched to an uncrossed disparity exceeding the half-cycle limit (e.g., 14′), while the model predicts a crossed disparity (e.g., −5.9′). The perceptual bias towards uncrossed disparities was found to be subject to bistability and dependent on exposure time (Blake, personal communication).

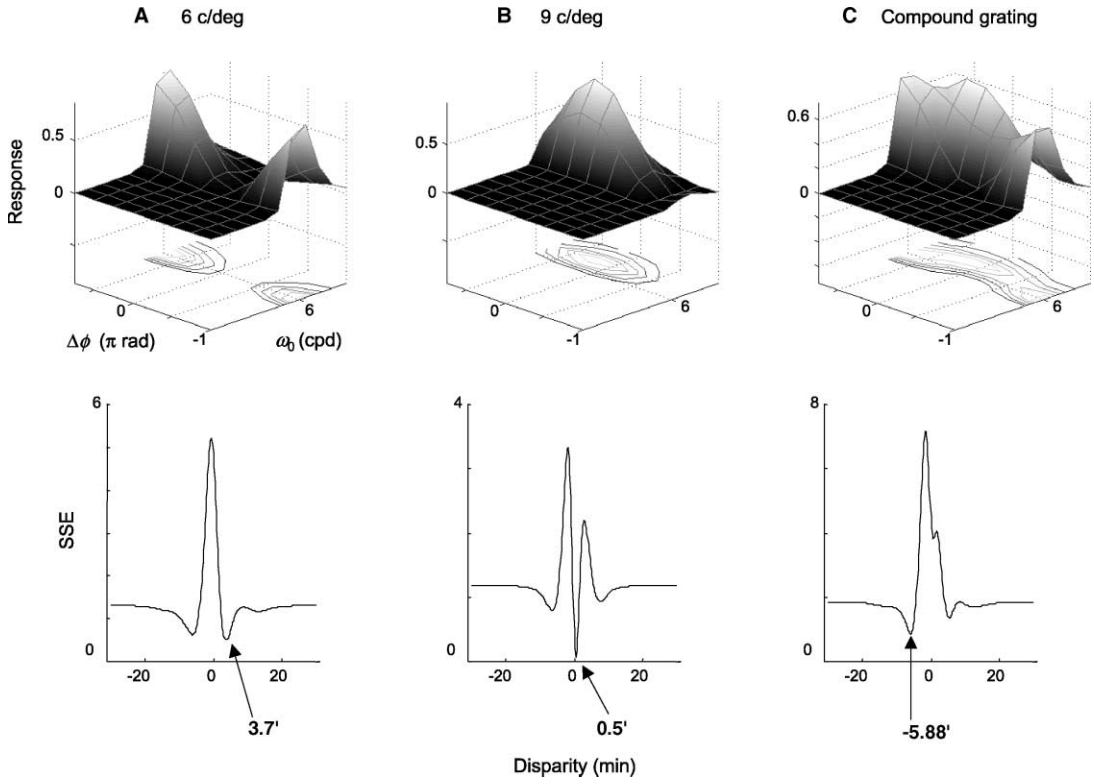

Fig. 10. Disparity interactions in compound gratings with identical component disparities. For single sine gratings (A and B), there are multiple valid binocular matches due to the stimulus periodicity, but the perceived depth is unambiguous and corresponds to the smallest disparity from the horopter. Summing two grating of similar spatial frequency can yield a resulting perceived disparity that is *not* a weighted average of the component disparities, but the smallest disparity consistent with both component gratings (C). This disparity necessarily corresponds to the disparity of the beat formed by the compound grating. *Top*, Multi-scale population response to 1-D sinusoids and sum of sinusoids. All stimuli had a physical disparity of 14′. *Bottom*, Template mismatch functions corresponding to the responses. The location of the minimum of the function is indicated by arrows. The model results are consistent with perception.

processing of the stimulus components (Victor & Conte, 1994).

There are two loci of non-linearity in our model: the squaring that follows linear filtering at the front end, and the pattern matching at the decoding stage. Which is responsible for depth perception mediated by contrast edges? Distortion products resulting from an early non-linearity can introduce Fourier energy at the beat frequency (Burton, 1973). This kind of non-linearity cannot be responsible for the model's behavior for compound-grating stimuli consisting of widely different spatial frequencies (because of the bandwidth of the elements in the model's front end), but it may play a role when the frequencies of the two components are similar. If this early non-linearity were the basis for depth perception in this case, one would expect that its removal would abolish or distort the corresponding percept. We test this hypothesis (Fig. 11) by submitting the sum of the population responses to the fundamentals directly to the population read-out stage of our model. This step removes interactions between the fundamentals that may result from a front end non-linearity. A difference in the fine details of the population response (cf. Fig. 10) suggests that non-linear aspects of the RF do contribute to the population response. However, the template mismatch function is largely similar to the one derived from the full response and the best-match disparity remains unchanged. Thus, the basis for depth perception corresponding to matching of image features lies in the template-matching stage, not in an early non-linearity.

4. Discussion

Our model makes predictions consistent with a wide range of psychophysical data. For sine grating stimuli, the inverse relationship between stereoacuity thresholds and spatial frequency (Fig. 3) is dependent on the width of the disparity tuning curve, not on the multi-scale response per se. On the other hand, the relationship between depth increment thresholds and disparity pedestal (Fig. 4) depends critically on how neural activity is combined across spatial scales, in addition to the size-disparity correlation of individual units. The model also

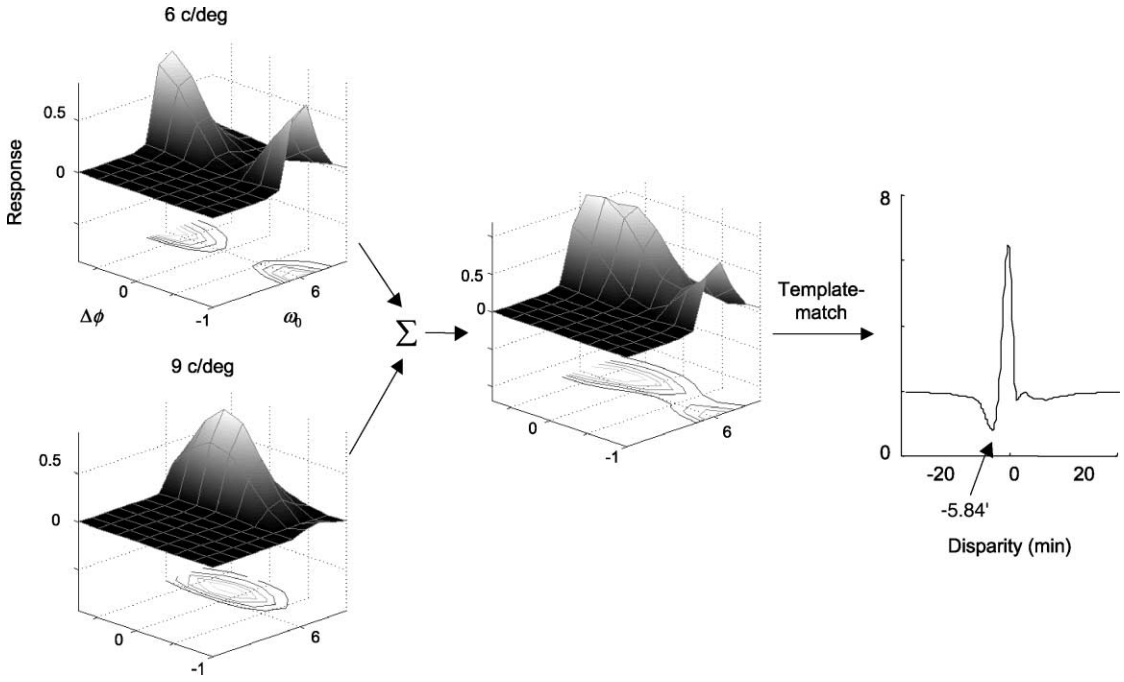

Fig. 11. Role of early non-linearity in perception of depth of narrow-band compound gratings. Responses to the 6 and 9 c/deg gratings are summed and passed to the decoding stage, bypassing the non-linear interaction between the gratings (due to the squaring non-linearity at the front end of the model). This manipulation does not change the location of the minimum of the template mismatch function. Therefore, the grating interactions in our model do not depend on the detection of the beat frequency via an early non-linearity. Instead, they depend on the integration of information across spatial scales in the read-out of the population response.

reproduces the observed decrease in apparent depths of low-spatial-frequency stimuli (Fig. 5).

Our model also makes predictions for broadband stimuli comprised of multiple, not necessarily concordant, disparity values. The responses to the sum of two independently generated sets of RDS depend on the difference in their disparity in a manner consistent with the observed changes from singleness of depth to transparency (Fig. 6). In addition, the model accounts for the variety of percepts elicited by sum of sine waveforms including corrugation, averaging, and transparency (Figs. 7–9). Thus transparency may be represented across spatial frequency (for narrow band stimuli such as gratings) or spatial location (for broad band stimuli such as RDS). Recently, simulations using our model showed that it can also account for the psychophysical finding that paired dots separated by disparity gradients greater than 1 can produce a robust appearance of transparent surfaces (McKee & Verghese, 2002). Finally, the model accounts for the perceived depth corresponding to the beat frequency in compound gratings (Figs. 10 and 11).

This variety of stereoscopic phenomena is not readily accounted for by other models for disparity representation, including coarse-to-fine, coincidence, WTA, vector averaging, and another template-matching model (Lehky & Sejnowski, 1990). We compare and contrast these models next.

4.1. Comparison to other models for decoding multi-scale population activity

While it is clear that stereopsis makes use of visual information across spatial scales, it is unclear how this information is combined. Coarse-to-fine (Marr & Poggio, 1979) and coincidence (Fleet et al., 1996) models seek to eliminate false binocular matches and false peaks in the disparity response functions. The present model achieves this goal, but differs with those previous models by avoiding sequential processing or a collapse of information across spatial scales. Disparity processing is not strictly uni-directional: psychophysical experiments have shown that information in fine scales can disambiguate information in coarse scales (Mallot, Gillner, & Arndt, 1996; Smallman, 1995). The coincidence model does not predict a dependence of transparency on the details of stimulus spatial frequency because activity is pooled across frequencies. In contrast, under a template-match model, responses to gratings of similar frequency overlap and interact while those of sufficiently different frequency remain segregated in the response space and do not average. In addition, in the case of mixed-disparity RDSs in which the component disparities are symmetric around the horopter, the response of the frequency channels peaks at an inter-ocular phase difference of zero at low spatial frequencies, but bifurcates to become doubly peaked at $\Delta\phi = \pm\pi$ at high

frequencies. Thus, each channel signals either a single disparity at the horopter or two disparities at $\pm\pi/\omega_0$ depending on its frequency range. When pooled under the coincidence model, these signals do not yield two values, as would be required for transparent percepts. The template-match model, however, makes correct predictions dependent on the disparities of the constituent RDSs. Finally, it is difficult for the coincidence model to account for the non-linear interaction between sine gratings with the same disparity but different frequency since the model combines responses across scales in a linear fashion.

We now compare the template-match model for decoding population activity to WTA and vector averaging models. First, averaging and WTA models do not immediately account for the relationship between stereoacuity and stimulus spatial frequency and for the pedestal effect on threshold because these models typically address the point of subjective equality (PSE), but do not provide a prediction of threshold data. Second, a WTA read-out scheme would completely eliminate the bias in disparity estimation because the filter that matches the stimulus in frequency also generates the greatest response. On the other hand, a model that averages the responses of the filters across spatial scales will likely produce similar biases as the template-match model as a consequence of off-peak frequency biases. Third, depth transparency serves as a litmus test for stereo models. WTA and vector averaging models inherently cannot predict multiple disparities at a single visual location. Recovery of transparency, on the other hand, is possible from the shape of the multi-scale population response (particularly the orientation of the spatial frequency-phase difference response in the case of phase-shift model). Fourth, the depth percept derived from a narrow-band compound grating may be neither an average nor a weighted mixture of the component disparities; however, it can be explained by the best match to the overall population response.

Our template-based approach to interpreting population activity differs from that of Lehky and Sejnowski (1990) in the front end of the model, the nature of the template, and the formulation of threshold. Lehky and Sejnowski used a family of DOG functions to emulate the disparity tuning function of three classes of neurons: near, tuned, and far. The function parameters are such that the tuned neurons have narrow widths while the near and far neurons have large widths. This specific arrangement of tuning widths is crucial to their model because the broader functions account for the increase in threshold at large disparities. Physiological data since 1990 suggest that, for a physiologically-based model, this arrangement needs to be modified. First, neurons broadly tuned to zero disparity exist, consistent with both phase- and position-shift mechanisms of encoding disparity. Second, position-shift mechanism allows for the existence of narrow tuning functions away from zero disparity. Finally, a continuum of disparity tuning rather than discrete classes is found—also reflected by the simulation result of Lehky and Sejnowski that a fairly large number of neurons (17–200) is needed to account for threshold data.

A second difference between their population model and the current one is the nature of the template used, stemming from the difference in the front end. Since Lehky and Sejnowski model deals with a one-variable function (i.e. disparity tuning curves), their template is the response of the family of DOGs at any particular disparity. Their model, therefore, does not consider the information represented across multiple scales. The absence of details about the multi-scale response, particularly the shape, predisposes to difficulties encountered by the coincidence model, as discussed above, including predictions for phenomena that depend on frequency differences and the non-linear interaction of grating disparities.

A third difference is the formulation of disparity discrimination threshold. Lehky and Sejnowski included Poisson noise in their model and used the signal/noise ratio to determine thresholds. Our approach is equivalent to the addition of a constant noise. We choose not to incorporate noise or a decision process explicitly in our model because pertinent physiological parameters are lacking. Furthermore, we are mainly interested in the comparison of relative thresholds between stimuli of the same type, and not in the absolute magnitudes of the thresholds. Details of the noise modeling and decision process would affect the absolute, but not the relative magnitudes of thresholds. The consistency of simulations with psychophysical data suggests that this is a valid assumption.

4.2. Phase vs. position-shift RF models

There are two general strategies for encoding disparity: phase difference and position shift (Anzai et al., 1999a; Cumming & DeAngelis, 2001; Fleet et al., 1996; Zhu & Qian, 1996). Studies in anesthetized cats suggest that disparity is mainly encoded via RF phase differences (Anzai, Ohzawa, & Freeman, 1997, 1999a). A recent study in awake monkeys found both phase and position encoding, but the phase mechanism encoded a larger range (by 25%) of disparities than did position shifts (Prince, Cumming, & Parker, 2002). This margin is smaller than that (60%) found in cats (Anzai et al., 1999a). These recent results suggest that both phase and position shift mechanisms contribute importantly to disparity encoding. For many stereo tasks, the underlying computation is equally well served by either phase- or position-shift scheme for encoding disparity (Mikaelian & Qian, 2000; Qian & Zhu, 1997; Qian & Andersen, 1997). This computational equivalence lar-

gely holds for our population model as well. Here, we focus on the distinctions between the implications of these schemes for our model.

One difference between the two mechanisms is that within a frequency channel, unambiguous encoding of retinal disparity by RF phase difference is limited to a range of $\pm$ half cycle of the preferred spatial period. The position-shift strategy, in principle, is not bound by such a limit, a feature that should be advantageous to the visual system. In practice, however, studies showed that V1 neurons, regardless of their mechanism for encoding disparity, generally have a range of preferred disparity within the half-cycle limit (Anzai et al., 1999a; Prince et al., 2002). Thus, both phase and position mechanisms of encoding disparity lend support to a correlation between the spatial frequency and preferred disparity.

Psychophysical data also support a size-disparity correlation; however, the range of disparity observed is greater than the half-cycle limit (Prince & Eagle, 1999; Schor & Wood, 1983; Schor et al., 1984b; but see Smallman & MacLeod, 1994). How can this discrepancy between the initial stage of disparity encoding and the perceived depth be explained? One possibility is that subsequent processing of the disparity encoded by the front end detectors allows for extraction of this information. Our model can account for a disparity greater than the half-cycle limits in a sum of sinusoidal luminance gratings (see Section 3.5). Activity in the coarser spatial channels whose half cycle limit is not exceeded by the stimulus disparity helps disambiguate activity in the finer channels (see also Fig. 12, which demonstrates how 0 and 12′ disparities can be discriminated based on the activity in channels centered near 5 cpd). Because our model does not require a serial processing of scales, it may account for data showing that observers can detect disparities exceeding the half cycle limit in a duration too short to allow vergence movement (Rohaly & Wilson, 1993; Smallman & MacLeod, 1997; Wilson, Blake, & Halpern, 1991).

Another difference between position and phase-shift based encoding bears upon the simulation results for disparity bias (see Section 3.3). The phase-shift mechanism has an inherent ambiguity in the relationship between phase and disparity, which contributes to an error in off-frequency channels (Fleet et al., 1996; Qian & Andersen, 1997; Zhu & Qian, 1996). The error in the off-frequency channels underlies the bias observed in low-spatial frequency sinusoidal luminance gratings. The reason for the direction of bias (towards the fixation plane) is that a grating stimulus activates more channels of higher off-frequencies than of lower off-frequencies because the bandwidth of the channels increases with frequency. The reason for the smaller bias in the high-frequency range is that a ceiling built into the model frequency spectrum limits the amount of off-frequency bias encountered in this range. A narrower filter bandwidth (such as that found at higher spatial frequencies in physiological data) would similarly reduce the amount of bias produced by the model. The position-shift scheme, on the other hand, does not face this uncertainty between phase and disparity. The observed bias would be difficult to explain in a pure position-shift model.

A final difference between the phase- and position-shift schemes is that only the former can account for the appearance of the beat in compound gratings. Templates generated from response of position-shift binocular filters to white noise stimuli differ from those of phase-shift filters in that the latter has oriented "ridges" in the response space (e.g. Fig. 12B), while the former resembles that in Fig. 12A. As a consequence, the position-shift model would not predict an interaction between grating disparities other than weighted averaging.

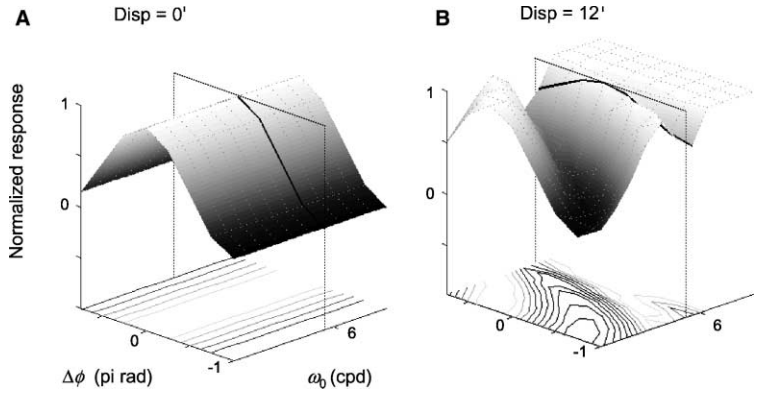

Fig. 12. The templates corresponding to disparity values of 0′ (A) and 12′ (B). The relatively narrow tuning of striate neurons means that peaks in the disparity tuning responses of a neuron occurs approximately at intervals equal to its preferred spatial period. For example, in the frequency channel tuned to 5 c/deg, the response peak occur at the inter-ocular phase $\Delta\phi = 0$ for stimulus disparities of 0 and 12′. The position of the response peak of this frequency channel, taken alone, cannot distinguish between these two disparities—even for non-periodic stimuli. However, the population response over multiple frequency channels is quite different for these two disparities.

4.3. False peaks in disparity tuning curves

Interpretation of disparity tuning curves of a neuron is ambiguous for both periodic and non-periodic stimuli. The visual system must select a single percept from multiple equally compatible ones in the former, and eliminate the selection of a percept that is inconsistent with the retinal image in the latter. Two aspects of our model contribute to this process. First, within a spatial channel, the template has a limited range of disparity to which it is sensitive, due to the finite envelope of the RF. Second, the read-out mechanism combines information across scales. The former is important in the case of periodic stimuli, while the latter is applicable to both types of ambiguity.

Simple sine-wave grating stereograms, in principle, are consistent with multiple disparities since shifting the grating by multiples of its spatial period results in the same stimulus pattern. In practice, however, observers report the perceived depth to be the one closest to the fixation plane (Boothroyd & Blake, 1984; McKee & Mitchison, 1988). Previously, stereo models have accounted for this tendency toward zero disparity by adopting a weighted sum of the output at each disparity and either adding a bias toward zero disparity (Prince & Eagle, 2000) or omitting the potentially ambiguous phase differences from consideration altogether (Zhu & Qian, 1996). Here we show that the population model can also account for this phenomenon, without requiring an ad hoc weighting function.

Obtained by a similar approach used to derive the response to white noise (Eq. (6)), the response of a quadrature unit to a grating stimulus is:

$$\begin{aligned} r_{\mathrm{q}} &= \tfrac{1}{2}e^{-\sigma_x^2(\omega_{\mathrm{s}}+\omega_0)^2} + \tfrac{1}{2}e^{-\sigma_x^2(\omega_{\mathrm{s}}-\omega_0)^2} + \tfrac{1}{2}e^{-\sigma_x^2(\omega_{\mathrm{s}}^2+\omega_0^2)} \\ &\quad \times \cos(2\phi_{\mathrm{s}} + \omega_{\mathrm{s}}D)[\cos(\omega_{\mathrm{s}}D) + \cos(\Delta\phi)] \\ &\quad + \tfrac{1}{2}e^{-\sigma_x^2(\omega_{\mathrm{s}}+\omega_0)^2}\cos(\omega_{\mathrm{s}}D - \Delta\phi) + \tfrac{1}{2}e^{-\sigma_x^2(\omega_{\mathrm{s}}-\omega_0)^2} \\ &\quad \times \cos(\omega_{\mathrm{s}}D - \Delta\phi) \\ &\approx \tfrac{1}{2}e^{-\sigma_x^2(\omega_{\mathrm{s}}-\omega_0)^2}[1 + \cos(\omega_{\mathrm{s}}D + \Delta\phi)], \end{aligned} \tag{13}$$

where ω_{s} and D are the spatial frequency and disparity of the grating respectively, and ω_0, σ_x and $\Delta\phi$ are intrinsic parameters of the model neuron. For the parameter values used in our simulation, terms containing $e^{-\sigma_x^2(\omega_{\mathrm{s}}+\omega_0)^2}$ are negligibly small (these terms are omitted by Zhu & Qian (1996)). Thus peaks in the response occur at $D = (2\pi n + \Delta\phi)/\omega_{\mathrm{s}}$.

The read-out stage of the model compares the disparity tuning (Eq. (13)) to the template given by Eq. (6). While the disparity tuning is periodic, the template mismatch function generated by the model has a well-defined minimum. To see how the read-out algorithm contributes to the solution of the false-peaks problem, first we focus on the spatial frequency channel whose center frequency matches that of the stimulus. While the disparity tuning curve is sinusoidal with respect to disparity D, the amplitude of modulation of the *template* decreases as disparity increases. In the limit of large D, the template response approaches a constant and is thus a poor match to the population response. Under the phase-difference model, the finite disparity range of each scale leads to a preference for the smallest percept consistent with the stimulus, without an explicitly built-in bias for small disparities. This result may also be achieved via a position-shift model, provided that the preferred disparities at each scale are proportional to the spatial scale.

Next we consider the how a multi-scale model contributes to the disambiguation of a grating stimulus. A grating stimulus activates all of the spatial frequency channels whose bandwidth overlaps the stimulus frequency ω_{s}. The peak response of all channels occurs at an inter-ocular phase difference of $\Delta\phi = \omega_{\mathrm{s}}D$ (Eq. (13)), while the template peaks at $\Delta\phi = \omega_0 D$ (Eq. (6)). Thus there is a phase offset between the peak of the channel response and that of the template response. This offset is proportional to the product of the disparity D and the deviation of the channel's center frequency from the stimulus frequency. For large D, the peaks in the template response at off-frequencies occur at significantly different positions from those in the population response, resulting in a poor match. Put another way, while the peak response of the predominant channel ($\omega_0 = \omega_{\mathrm{s}}$) is consistent with multiple disparities $(2\pi n + \Delta\phi)/\omega_{\mathrm{s}}$, only one disparity, namely $\Delta\phi/\omega_{\mathrm{s}}$ minimizes the phase offset in *all* channels. This is illustrated by an example in Fig. 12, which shows the normalized template responses corresponding to two disparities, 0 and 12′. Consider the response of the channel centered at 5 c/deg: the peak occurs at $\Delta\phi = 0$ for both disparity values. It is not possible to separate them based on the location of the peak in this channel alone. However, by taking into consideration neighboring channels, it is clear that the pattern of the population response is distinctive for the two disparity values. This multi-scale read-out works in concert with the single scale disparity limit in damping responses to false peaks. Importantly, the former selects the percept most compatible with a single stimulus disparity across spatial scales, while the latter mechanism indiscriminately favors the smallest disparity. The relative contribution of these two influences is controlled by the range of frequencies present in the model (a wider range accentuates the former process) and the bandwidth of the typical neuron (a wider bandwidth accentuates the latter process).

The problem of false peaks in the disparity tuning curve is not limited to periodic stimuli, but rather results from the quasi-periodicity of the response of the narrowly tuned front end filters and affects both phase- and

position-shift stereo models in their single-channel formulation (Fleet et al., 1996; Qian, 1994). In any instance of the noise stimulus, false peaks may be larger than the peak at the correct disparity (Fleet et al., 1996; Qian & Zhu, 1997). Thus, a simple WTA or vector-averaging algorithm cannot identify the correct disparity for either spectrally broad (noise) or narrow (grating) stimuli. Both the coincidence model (Fleet et al., 1996) and our model overcome this ambiguity by combining information across spatial scales.

4.4. Extensions of the model

There are a number of respects in which the present model is incomplete, but in which there is an evident strategy for extensions. The present model is a local model, limited to a family of neurons with RF centers at one spatial location. Consequently, it cannot account for perceptual phenomena that involve long-range interaction such as "depth capture" (Ramachandran & Cavanagh, 1985), depth interpolation (Collett, 1985; Mitchison & McKee, 1985), and attractive and repulsive interaction in line stereograms (Westheimer, 1986) and RDS (Stevenson et al., 1991). It is natural to extend the model by pooling disparity energy response over a local spatial neighborhood (Fleet et al., 1996; Qian & Zhu, 1997). Mikaelian and Qian (2000) have shown that such a modification can account for depth repulsion and attraction in simple line stereograms.

Hess and Wilcox (1994) and others have studied a class of stimuli that rely on higher order image characteristics to generate stereoscopic depth. In these "non-Fourier" or "second-order" stimuli, disparity information is carried by inter-ocular correlations in the *envelope* of contrast modulation, analogous to the "second-order" motion stimuli introduced by Chubb and Sperling (1988). Such stimuli fail to excite elementary Reichardt (1961)-type motion detectors, but nevertheless elicit a strong percept of motion, presumably due to motion of a contrast envelope or another second-order feature of the stimulus. The current model cannot account for the depth percept elicited by second-order stimuli, because the key step for disparity computation operates on (signed) image contrast (Qian & Zhu, 1997), not its envelope. However, a front end non-linearity, such as occurs in the responses of retinal Y-type ganglion cells (Enroth-Cugell & Robston, 1966; Shapley, 1990; Shapley & Perry, 1986), would allow for the detection of the contrast envelope along with first-order features, as is the case with motion perception (Wilson et al., 1992; Taub, Victor, & Conte, 1997). This would enable processing of second-order disparity by the mechanism we have proposed, though of course a separate pathway specifically devoted to second-order disparity cannot be excluded.

While depth transparency can be qualitatively demonstrated by our model, the specific psychophysical parameters do not agree completely with the data. For example, a greater disparity offset than reported in the literature is needed to produce transparency in sum-of-sinusoids stimuli. One possibility is that cross-channel disparity inhibition at the front end of the model, as suggested by Rohaly and Wilson (1994), sharpens the disparity tuning of a linear model and leads to greater depth resolving power. Alternatively, a template that is more narrowly tuned to inter-ocular phase difference can also result in better depth resolution. This may arise as a consequence of inhibitory interactions between neighboring disparity channels (Akerstrom & Todd, 1988; Cormack, Stevenson, & Schor, 1993; Stevenson, Cormack, Schor, & Tyler, 1992). In its current form, our model does not implement explicit interactions between disparity channels (although such can arise indirectly from the template-match read-out stage). We hypothesize that such interactions are present, analogous to those that may contribute to motion transparency (Qian, Andersen, & Adelson, 1994). However, the details of the interactions remain an open question.

5. Conclusions

We focused on the representation and processing of retinal disparity, a fundamental aspect of stereoscopic depth perception. An understanding of the effect of disparity on the activity of individual binocular V1 neurons is well established, by both experimental studies of their RF structure and theoretical studies. The main goal of this paper is to examine how the activity of a population of binocular V1 neurons at multiple spatial scales may contribute to the representation of disparity. Our approach is in the same category as the "spectrum" method that Lehky and Sejnowski (1990, 1999) have applied to depth and color vision, and models for distributed representations of motion (Heeger, 1987; Heeger & Simoncelli, 1993; Treue et al., 2000; Victor & Conte, 1994).

Our model makes predictions consistent with a variety of psychophysical data, including disparity thresholds and biases from veridicality measured with simple stimuli, and the qualitative features of the depth percepts elicited by complex stimuli. The ability of the model to account for data that other models including WTA, vector averaging, coarse-to-fine, and coincidence could not account for suggests that the pattern of population neural activity across spatial scales is a better correlate for the perceptual representation of disparity than the activity of single neurons or the pooled activity of multiple neurons.

Acknowledgements

This work was supported by National Institute of Health Grant EY7977. J.J.T. was supported by the Cornell/Rockefeller/Sloan-Kettering MD–PhD Program. We thank Ning Qian for generously giving us the computer code of his model. We thank Mary Conte for helpful suggestions on the manuscript.

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A—Optics Image Science and Vision, 2*, 284–299.

Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature, 300*, 523–525.

Akerstrom, R. A., & Todd, J. T. (1988). The perception of stereoscopic transparency. *Perception & Psychophysics, 44*, 421–432.

Anzai, A., Ohzawa, I., & Freeman, R. D. (1997). Neural mechanisms underlying binocular fusion and stereopsis: position vs. phase. *Proceedings of the National Academy of Sciences of the United States of America, 94*, 5438–5443.

Anzai, A., Ohzawa, I., & Freeman, R. D. (1999a). Neural mechanisms for encoding binocular disparity: receptive field position versus phase. *Journal of Neurophysiology, 82*, 874–890.

Anzai, A., Ohzawa, I., & Freeman, R. D. (1999b). Neural mechanisms for processing binocular information I. Simple cells. *Journal of Neurophysiology, 82*, 891–908.

Arndt, P. A., Mallot, H. A., & Bulthoff, H. H. (1995). Human stereovision without localized image features. *Biological Cybernetics, 72*, 279–293.

Badcock, D. R., & Schor, C. M. (1985). Depth-increment detection function for individual spatial channels. *Journal of the Optical Society of America A, 2*, 1211–1216.

Barlow, H. B., Blakemore, C., & Pettigrew, J. D. (1967). The neural mechanisms of binocular depth discrimination. *Journal of Physiology, 193*, 327–342.

Blakemore, C. (1970). Range and scope of binocular depth discrimination in man. *Journal of Physiology-London, 211*, 599–622.

Boothroyd, K., & Blake, R. (1984). Stereopsis from disparity of complex grating patterns. *Vision Research, 24*, 1205–1222.

Braddick, O. J. (1980). Low-level and high-level processes in apparent motion. *Philosophical Transactions of the Royal Society of London Series B—Biological Sciences, 290*, 137–151.

Brown, J. M., & Weisstein, N. (1988). A spatial frequency effect on perceived depth. *Perception & Psychophysics, 44*, 157–166.

Burt, P., & Julesz, B. (1980). A disparity gradient limit for binocular fusion. *Science, 208*, 615–617.

Burton, G. J. (1973). Evidence for non-linear response processes in the human visual system from measurements on the thresholds of spatial beat frequencies. *Vision Research, 13*, 1211–1225.

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.

Collett, T. S. (1985). Extrapolating and interpolating surfaces in depth. *Proceedings of the Royal Society of London Series B—Biological Sciences, 224*, 43–56.

Cormack, L. K., Stevenson, S. B., & Schor, C. M. (1993). Disparity-tuned channels of the human visual system. *Visual Neuroscience, 10*, 585–596.

Cumming, B. G., & DeAngelis, G. C. (2001). The physiology of stereopsis. *Annual Review of Neuroscience, 24*, 203–238.

DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1991). Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature, 352*, 156–159.

DeValois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research, 22*, 545–559.

Enroth-Cugell, C., & Robston, J. G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology, 187*, 517–552.

Fleet, D. J., Wagner, H., & Heeger, D. J. (1996). Neural encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision Research, 36*, 1839–1857.

Freeman, R. D., & Ohzawa, I. (1990). On the neurophysiological organization of binocular vision. *Vision Research, 30*, 1661–1676.

Frisby, J. P., & Mayhew, J. E. (1978). The relationship between apparent depth and disparity in rivalrous-texture stereograms. *Perception, 7*, 661–678.

Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science, 233*, 1416–1419.

Glennerster, A., & Parker, A. J. (1997). Computing stereo channels from masking data. *Vision Research, 37*, 2143–2152.

Hammett, S. T., Ledgeway, T., & Smith, A. T. (1993). Transparent motion from feature- and luminance-based processes. *Vision Research, 33*, 1119–1122.

Heeger, D. J. (1987). Model for the extraction of image flow. *Journal of the Optical Society of America A—Optics Image Science and Vision, 4*, 1455–1471.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9*, 181–198.

Heeger, D. J., & Simoncelli, E. P. (1993). Model of visual motion sensing. In L. Harris & J. M. Jenkins (Eds.), *Spatial vision in humans and robots* (pp. 367–392). New York: Cambridge University Press.

Hess, R. F., & Wilcox, L. M. (1994). Linear and non-linear filtering in stereopsis. *Vision Research, 34*, 2431–2438.

Julesz, B., & Miller, J. E. (1975). Independent spatial-frequency-tuned channels in binocular fusion and rivalry. *Perception, 4*, 125–143.

Kontsevich, L. L., & Tyler, C. W. (1994). Analysis of stereothresholds for stimuli below 2.5 c/deg. *Vision Research, 34*, 2317–2329.

Legge, G. E., & Gu, Y. C. (1989). Stereopsis and Contrast. *Vision Research, 29*, 989–1004.

Lehky, S. R., & Sejnowski, T. J. (1990). Neural model of stereoacuity and depth interpolation based on a distributed representation of stereo disparity. *Journal of Neuroscience, 10*, 2281–2299.

Lehky, S. R., & Sejnowski, T. J. (1999). Seeing white: qualia in the context of decoding population codes. *Neural Computation, 11*, 1261–1280.

Mallot, H. A., Gillner, S., & Arndt, P. A. (1996). Is correspondence search in human stereo vision a coarse-to-fine process? *Biological Cybernetics, 74*, 95–106.

Marr, D., & Poggio, T. (1976). Cooperative computation of stereo disparity. *Science, 194*, 283–287.

Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London Series B—Biological Sciences, 204*, 301–328.

Mckee, S. P., Glennerster, A., Harris, J. M. (1996). Dot spacing affects disparity tuning in a visual search task. Spatial Scale Interactions in Vision and Eye Movement Control, Durham UK (Conference abstract).

McKee, S. P., Levi, D. M., & Bowne, S. F. (1990). The imprecision of stereopsis. *Vision Research, 30*, 1763–1779.

McKee, S. P., & Mitchison, G. J. (1988). The role of retinal correspondence in stereoscopic matching. *Vision Research, 28*, 1001–1012.

McKee, S. P., & Verghese, P. (2002). Stereo transparency and the disparity gradient limit. *Vision Research*, 1963–1977.

Mikaelian, S., & Qian, N. (2000). A physiologically-based explanation of disparity attraction and repulsion. *Vision Research, 40*, 2999–3016.

Mitchison, G. J., & McKee, S. P. (1985). Interpolation in stereoscopic matching. *Nature, 315*, 402–404.

Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science, 249*, 1037–1041.

Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1996). Encoding of binocular disparity by simple cells in the cat's visual cortex. *Journal of Neurophysiology, 75*, 1779–1805.

O'Shea, R. P., Blackburn, S. G., & Ono, H. (1994). Contrast as a depth cue. *Vision Research, 34*, 1595–1604.

O'Shea, R. P., Govan, D. G., & Sekuler, R. (1997). Blur and contrast as pictorial depth cues. *Perception, 26*, 599–612.

Parker, A. J., & Yang, Y. (1989). Spatial properties of disparity pooling in human stereo vision. *Vision Research, 29*, 1525–1538.

Poggio, G. F., & Fischer, B. (1977). Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *Journal of Neurophysiology, 40*, 1392–1405.

Pollard, S. B., & Frisby, J. P. (1990). Transparency and the uniqueness constraint in human and computer stereo vision. *Nature, 347*, 553–556.

Pollard, S. B., Mayhew, J. E. W., & Frisby, J. P. (1985). PMF—a stereo correspondence algorithm using a disparity gradient limit. *Perception, 14*, 449–470.

Prazdny, K. (1985). Detection of binocular disparities. *Biological Cybernetics, 52*, 93–99.

Prince, S. J., Cumming, B. G., & Parker, A. J. (2002). Range and mechanism of encoding of horizontal disparity in macaque V1. *Journal of Neurophysiology, 87*, 209–221.

Prince, S. J., & Eagle, R. A. (1999). Size-disparity correlation in human binocular depth perception. *Proceedings of the Royal Society of London Series B—Biological Sciences, 266*, 1361–1365.

Prince, S. J., & Eagle, R. A. (2000). Weighted directional energy model of human stereo correspondence. *Vision Research, 40*, 1143–1155.

Prince, S. J., Eagle, R. A., & Rogers, B. J. (1998). Contrast masking reveals spatial-frequency channels in stereopsis. *Perception, 27*, 1345–1355.

Qian, N. (1994). Computing stereo disparity and motion with known binocular cell properties. *Neural Computation, 6*, 390–404.

Qian, N., & Andersen, R. A. (1997). A physiological model for motion–stereo integration and a unified explanation of Pulfrich-like phenomena. *Vision Research, 37*, 1683–1698.

Qian, N., Andersen, R. A., & Adelson, E. H. (1994). Transparent motion perception as detection of unbalanced motion signals. I. Psychophysics. *Journal of Neuroscience, 14*, 7357–7366.

Qian, N., & Zhu, Y. (1997). Physiological computation of binocular disparity. *Vision Research, 37*, 1811–1827.

Ramachandran, V. S., & Cavanagh, P. (1985). Subjective contours capture stereopsis. *Nature, 317*, 527–530.

Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In W. A. Rosenbluth (Ed.), *Sensory communication* (pp. 303–317). New York: Wiley.

Rohaly, A. M., & Wilson, H. R. (1993). Nature of coarse-to-fine constraints on binocular fusion. *Journal of the Optical Society of America A—Optics Image Science and Vision, 10*, 2433–2441.

Rohaly, A. M., & Wilson, H. R. (1994). Disparity averaging across spatial scales. *Vision Research, 34*, 1315–1325.

Salinas, E., & Abbott, L. F. (1994). Vector reconstruction from firing rates. *Journal of Computational Neuroscience, 1*, 89–107.

Salzman, C. D., & Newsome, W. T. (1994). Neural mechanisms for forming a perceptual decision. *Science, 264*, 231–237.

Schor, C. M., & Howarth, P. A. (1986). Suprathreshold stereo-depth matches as a function of contrast and spatial frequency. *Perception, 15*, 249–258.

Schor, C. M., & Wood, I. (1983). Disparity range for local stereopsis as a function of luminance spatial-frequency. *Vision Research, 23*, 1649–1654.

Schor, C. M., Wood, I. C., & Ogawa, J. (1984a). Spatial tuning of static and dynamic local stereopsis. *Vision Research, 24*, 573–578.

Schor, C., Wood, I., & Ogawa, J. (1984b). Binocular sensory fusion is limited by spatial-resolution. *Vision Research, 24*, 661–665.

Schumer, R. A., & Julesz, B. (1984). Binocular disparity modulation sensitivity to disparities offset from the plane of fixation. *Vision Research, 24*, 533–542.

Shadlen, M. N., Britten, K. H., Newsome, W. T., & Movshon, J. A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *Journal of Neuroscience, 16*, 1486–1510.

Shapley, R. (1990). Visual sensitivity and parallel retinocortical channels. *Annual Review Psychology, 41*, 635–658.

Shapley, R., & Perry, V. H. (1986). Cat and monkey retinal ganglion cells and their visual functional roles. *TINS, 9*, 229–235.

Siderov, J., & Harwerth, R. S. (1993a). Effects of the spatial-frequency of test and reference stimuli on stereo-thresholds. *Vision Research, 33*, 1545–1551.

Siderov, J., & Harwerth, R. S. (1993b). Precision of stereoscopic depth-perception from double images. *Vision Research, 33*, 1553–1560.

Siderov, J., & Harwerth, R. S. (1995). Stereopsis, spatial-frequency and retinal eccentricity. *Vision Research, 35*, 2329–2337.

Smallman, H. S. (1995). Fine-to-coarse scale disambiguation in stereopsis. *Vision Research, 35*, 1047–1060.

Smallman, H. S., & MacLeod, D. I. (1994). Size-disparity correlation in stereopsis at contrast threshold. *Journal of the Optical Society of America A—Optics Image Science and Vision, 11*, 2169–2183.

Smallman, H. S., & MacLeod, D. I. (1997). Spatial scale interactions in stereo sensitivity and the neural representation of binocular disparity. *Perception, 26*, 977–994.

Stevenson, S. B., Cormack, L. K., & Schor, C. M. (1989). Hyperacuity, superresolution and gap resolution in human stereopsis. *Vision Research, 29*, 1597–1605.

Stevenson, S. B., Cormack, L. K., & Schor, C. M. (1991). Depth attraction and repulsion in random dot stereograms. *Vision Research, 31*, 805–813.

Stevenson, S. B., Cormack, L. K., Schor, C. M., & Tyler, C. W. (1992). Disparity tuning in mechanisms of human stereopsis. *Vision Research, 32*, 1685–1694.

Taub, E., Victor, J. D., & Conte, M. M. (1997). Nonlinear preprocessing in short-range motion. *Vision Research, 37*, 1459–1477.

Treue, S., Hol, K., & Rauber, H. J. (2000). Seeing multiple directions of motion-physiology and psychophysics. *Nature Neuroscience, 3*, 270–276.

Victor, J. D., & Conte, M. M. (1994). Investigation of a patient with severely impaired direction discrimination: evidence against the intersection-of-constraints model. *Vision Research, 34*, 267–277.

Watson, A. B., & Ahumada, A. J. (1983). A look at motion in the frequency domain. NASA Tech. Memo, TM-84352.

Weinshall, D. (1991). Seeing "ghost" planes in stereo vision. *Vision Research, 31*, 1731–1748.

Westheimer, G. (1986). Spatial interaction in the domain of disparity signals in human stereoscopic vision. *Journal of Physiology, 370*, 619–629.

Wilson, H. R., Blake, R., & Halpern, D. L. (1991). Coarse spatial scales constrain the range of binocular fusion on fine scales. *Journal*

of the Optical Society of America A—Optics Image Science and Vision, 8, 229–236.

Wilson, H. R., Ferrera, V. P., & Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience, 9*, 79–98.

Yang, Y., & Blake, R. (1991). Spatial-frequency tuning of human stereopsis. *Vision Research, 31*, 1177–1189.

Zhu, Y. D., & Qian, N. (1996). Binocular receptive field models, disparity tuning, and characteristic disparity. *Neural Computation, 8*, 1611–1641.

Zohary, E., Scase, M. O., & Braddick, O. J. (1996). Integration across directions in dynamic random dot displays: vector summation or winner take all? *Vision Research, 36*, 2321–2331.