Neural Encoding of Binocular Disparity: Energy Models, Position Shifts and Phase Shifts

David J. Fleet
Dept. Computing Science,
Queen's University,
Kingston, Canada, K7L 3N6

Hermann Wagner
Institut fuer Zoologie
Technische Universitaet Muenchen,
D-85748 Garching, Germany

David J. Heeger
Dept. Psychology
Stanford University
Stanford, CA 94305

Corresponding Author:
Prof. David Fleet
Dept. of Computing Science
Queen's University, Kingston, Canada, K7L 3N6
TEL. 613-545 6488
FAX: 613-545 6513

Running Head: Neural Encoding of Binocular Disparity

Key Words: stereopsis, binocular disparity, energy mechanism, phase shifts, and position shifts.

Acknowledgements: This work was supported by grants from NSERC Canada and from Queen's University to DJF, by a DFG grant to HW and DJF, and by an NIMH grant (MH50228), and an Alfred P. Sloan Research Fellowship to DJH.

Abstract

Neurophysiological data supports two models for the disparity selectivity of binocular simple and complex cells in primary visual cortex. These involve binocular combinations of monocular receptive fields that are shifted in retinal position (the position-shift model) or in phase (the phase-shift model) between the two eyes. This article presents a formal description and analysis of a binocular energy model with these forms of disparity selectivity. We propose how one might measure the relative contributions of phase- and position-shifts in simple and complex cells. The analysis also reveals ambiguities in disparity encoding that are inherent in these model neurons, suggesting a need for a second stage of processing. We propose that linear pooling the binocular responses across orientations and scales (spatial frequency) is capable of producing an unambiguous representation of disparity.

1 Introduction

Neurons sensitive to binocular disparity have been found in the visual cortex of many mammals and in the visual wulst of the owl, and are thought to play a significant role in stereopsis (Barlow *et al.*, 1967; Nikara *et al.*, 1968; Hubel & Wiesel, 1970; Clarke *et al.*, 1976; Pettigrew & Konishi, 1976; Poggio & Fischer, 1977; Fischer & Kruger, 1979; Ferster, 1981; Poggio & Talbot, 1981; Ohzawa & Freeman, 1986a,b; LeVay & Voigt, 1988; Ohzawa *et al.*, 1990; DeAngelis *et al.*, 1991; Wagner & Frost, 1993). A number of physiologists (Hubel & Wiesel, 1962; Pettigrew *et al.*, 1968; Pettigrew, 1972; Maske *et al.*, 1984; Poggio *et al.*, 1985; Wagner & Frost, 1993) have suggested that disparity might be encoded by a shift of receptive-field position. According to this *position-shift model*, disparity selective cells combine the outputs of similarly shaped, monocular receptive fields from different retinal positions in the left and right eyes. More recently, Ohzawa *et al.* (1990) and DeAngelis *et al.* (1991, 1995) suggested that disparity sensitivity might instead be a result of interocular phase shifts. In this *phase-shift model*, the centers of the left- and right-eye receptive fields coincide, but the arrangement of receptive field subregions are different. Simulations by Nomura *et al.* (1990) showed that phase-shift models can account for the disparity tuning of some V1 neurons, and an implementation by Qian (1994) showed that the disparity of random-dot stereograms can be correctly extracted when disparities are small.

This article presents a formal description and analysis of a binocular energy model. We examine the behavior of both position-shift and phase-shift models of disparity selectivity, as well as a hybrid of the two. Our analysis provides quantitative predictions of the models, and suggests how one might measure the relative contributions of phase and position shifts to the disparity selectivity of simple and complex cells. This analysis shows further that there are ambiguities in disparity encoding that are inherent in the position-shift and phase-shift models. The presence of these ambiguities suggests the need for a further stage of processing. We demonstrate that pooling the binocular energy responses across orientations and scales produces an unambiguous representation of disparity.

2 Models of Disparity Selectivity

There are two major classes of neurons in primary visual cortex (V1): simple cells and complex cells (Hubel & Wiesel, 1962). Both types are selective for stimulus position and orientation. They respond vigorously to stimuli of a preferred orientation, but less well or not at all to stimuli of other orientations. Many V1 neurons are also disparity selective.

Disparity-sensitive cells are often divided into four types: tuned-excitatory, tuned-inhibitory, near and far (Poggio & Fischer, 1977). Disparity selectivity in these different types might arise from different mechanisms (Poggio & Fischer, 1977; Ferster, 1981; but see Nomura *et al.*, 1990 for the opposite point of view). Tuned-inhibitory, near and far cells usually receive a strong excitatory input from one eye and an inhibitory input from the other eye (i.e., the monocular inputs are unbalanced), and most of them do not show binocular facilitation. Tuned-excitatory cells show a sharp response peak due to binocular facilitation, the responses at disparities flanking the peak are often inhibited, and they have balanced monocular inputs. This article concerns tuned-excitatory cells.

2.1 Linear Neurons and Energy Neurons

There is a long tradition of modeling simple cells as *linear neurons* (Hubel & Wiesel, 1962; Campbell *et al.*, 1968, 1969; Movshon *et al.*, 1978a; Ohzawa & Freeman, 1986a; Hamilton *et al.*, 1989). This model is attractive because a linear neuron can be characterized with a relatively small number of measurements.

Figure 1(A) shows a schematic diagram of a monocular linear neuron. A linear neuron's response is a weighted sum of stimulus intensities within a small region of the entire visual field, called the neuron's *receptive field*. In the illustration, the three ellipses depict subregions of the receptive field, one with positive weights (the unshaded ellipse), and two with negative weights (the shaded ellipses). The neuron is excited when a bright light is flashed in the positive subregion, and inhibited when a bright light is flashed in a negative subregion. Bright lights flashed simultaneously in both positive and negative subregions tend to cancel. The positive and negative weights are balanced so the neuron does not respond to blank stimuli. Rather, its response is proportional to stimulus contrast for patterned stimuli that vary in intensity over space.

Figure 1(B) depicts a binocular linear neuron. This neuron's response depends on a weighted sum of the stimulus intensities presented to both eyes. The left- and right-eye receptive fields are identical for the neuron depicted in the figure, but this need not be the case in general. Also, the left- and right-eye receptive fields of this linear neuron are in exact binocular correspondence as indicated by the small reference points below the weighting functions.

But there is a small problem with a linear model of simple cells. Linear neurons can have negative responses because they sum input intensities using both positive and negative weights, while extracellular responses (firing rates) of real neurons are by definition positive. Neurons with a high maintained firing rate could encode positive and negative values by responding either more

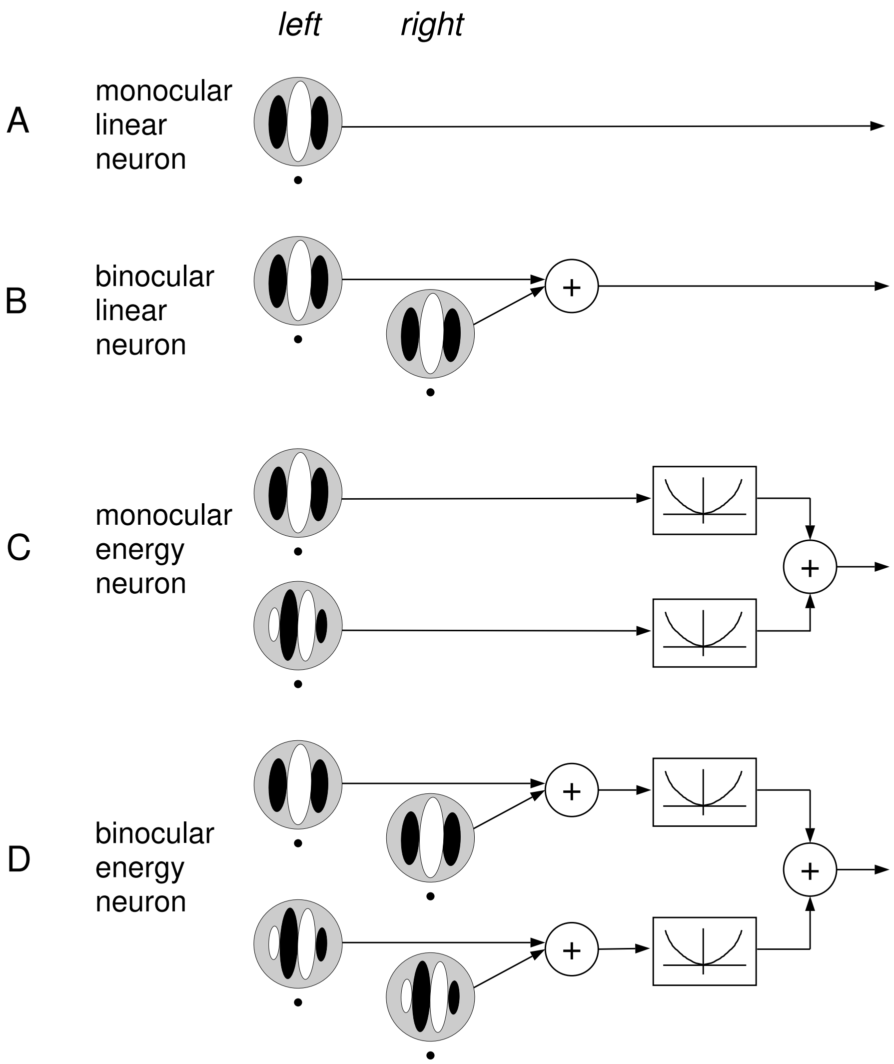

Figure 1: Receptive fields of model neurons. (A) A monocular, orientation selective, linear neuron. Its response depends on a weighted sum of the stimulus intensities within its receptive field. Shaded ellipses correspond to inhibitory subregions of the weighting function, and the unshaded ellipse corresponds to an excitatory subregion. (B) A binocular linear neuron's response depends on a weighted sum of the stimulus intensities presented to both eyes. The reference points (black dots) below the weighting functions indicate that the two weighting functions are in exact binocular correspondence. (C) A monocular energy neuron sums the squared responses of two monocular, linear neurons. The weighting functions of the two linear neurons are identical except for a 90° phase shift. (D) A binocular energy neuron sums the squared responses of two binocular, linear neurons. All four linear weighting functions are centered in exact (monocular and binocular) retinal correspondence.

or less than the maintained rate. But simple cells have very little maintained discharge. Instead, positive and negative values may be encoded by two neurons, one responsible for the positive part and one for the negative part. The two neurons are complements of one another; an excitatory subregion of one neuron's receptive field is aligned with an inhibitory subregion of the other neuron's receptive field. The response of each neuron is halfwave-rectified so that only one of the two neurons has a non-zero response at any given time. Simple cells are often characterized as halfwave-rectified linear neurons (e.g., Movshon *et al.*, 1978a; Heeger, 1992b).

Complex cells do not have discrete ON and OFF receptive field subregions, and have been modeled as *energy neurons* (Adelson and Bergen, 1985; Emerson *et al.*, 1992; Heeger, 1992b; Pollen & Ronner, 1983). An energy neuron sums the squared responses of a quadrature pair of linear neurons that are 90° out of phase, but with otherwise identical tuning properties (Fig. 1C). Equivalently, an energy neuron could sum the squared responses of four halfwave-rectified, linear neurons.

The monocular energy neuron depicted in Fig. 1(C) has one linear subunit that is even-symmetric (even phase) and another that is odd-symmetric (odd phase), but this is not necessary. The critical property is that the two subunits must be in quadrature phase (90° phase shift). Although simple cell weighting functions are not necessarily even- or odd-symmetric (Field & Tolhurst, 1986; Heggelund, 1986; Jones and Palmer, 1987), the receptive fields of adjacent simple cells tend to exhibit 90° or 180° phase relationships (Foster *et al.*, 1983; Liu *et al.*, 1992; Palmer and Davis, 1981; Pollen & Ronner, 1981). A local pool of simple cells thus provides the right combination of signals for an ideal energy neuron. Approximately the same behavior may be obtained by summing the squared responses of many linear neurons (or halfwave-rectified, linear neurons), regardless of their phase, but with receptive fields distributed over a local spatial region.

A binocular energy neuron (Ohzawa *et al.*, 1990) is depicted in Fig. 1(D). This neuron sums the squared responses of a quadrature pair of binocular linear neurons. This article is primarily concerned with the behavior of binocular energy neurons.

2.2 Disparity Selectivity: Position Shifts and Phase Shifts

Figure 2 depicts two ways that non-zero disparity preferences have been introduced in models of disparity selectivity. The neuron depicted in Fig. 2(A) is tuned for zero disparity because the locations of the two monocular receptive fields are in exact binocular correspondence (indicated relative to the reference points) and the two weighting functions are identical. In Fig. 2(B), the right eye's receptive field is shifted to the right. In Fig. 2(C), the right eye's subfield is shifted in phase by 90°. Both neurons in Fig. 2(B) and 2(C) are constructed to prefer uncrossed disparities; to evoke a maximal response, a visual feature (line, edge, grating) should be presented to the right eye in a position that is slightly shifted to the right.

Figure 3 depicts three binocular energy neurons. A non-zero disparity preference is introduced either by shifting the receptive field positions (Fig. 3B) or the receptive field phases (Fig. 3C).

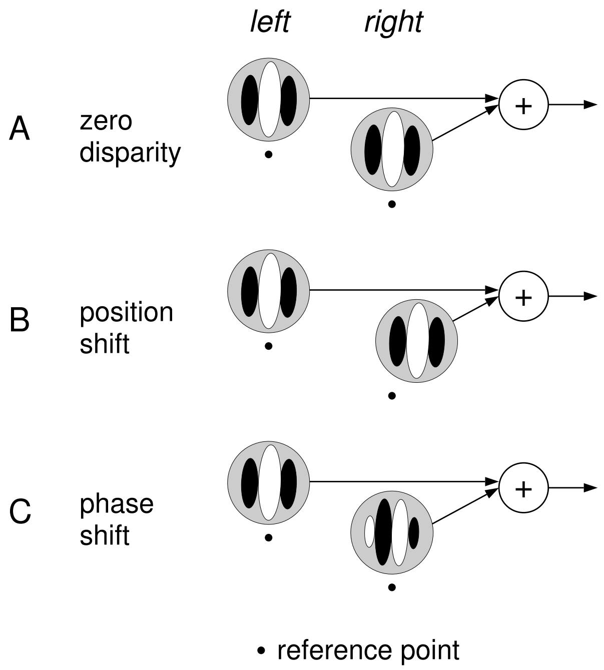

Figure 2: (A) A binocular linear neuron that prefers zero disparity. (B,C) Non-zero disparity preferences attained by shifting the position or phase of one of the monocular receptive fields.

3 Formalizing the Model

In order to examine the behavior of the model in detail, we derive formulas for their responses. A table of symbols (table 1) is provided to help the reader keep track of mathematical notation. We begin by concentrating on linear neurons and the zero-disparity energy-neuron, like that in Fig. 3(A), after which the position-shifts and phase-shifts are analyzed.

3.1 Spatial Arrays of Identical Linear Neurons: Convolution

Consider a spatial array of monocular linear (left-eye) neurons that are identical except for spatial location, so that their responses can be computed by convolving the stimulus with a linear filter. For one spatial dimension (ignoring time and the other spatial dimension), the responses $L(x)$ are given by the familiar convolution formula:

$$L(x) = \int f_l(\xi - x) I(\xi) d\xi \ , \tag{1}$$

where $I(\xi)$ is the stimulus intensity at each spatial position and $f_l(\xi)$ is a linear filter (i.e., the weighting function of a neuron), and x is the position of the receptive field center of each neuron.

For notational convenience, we use complex numbers to express the weighting functions and

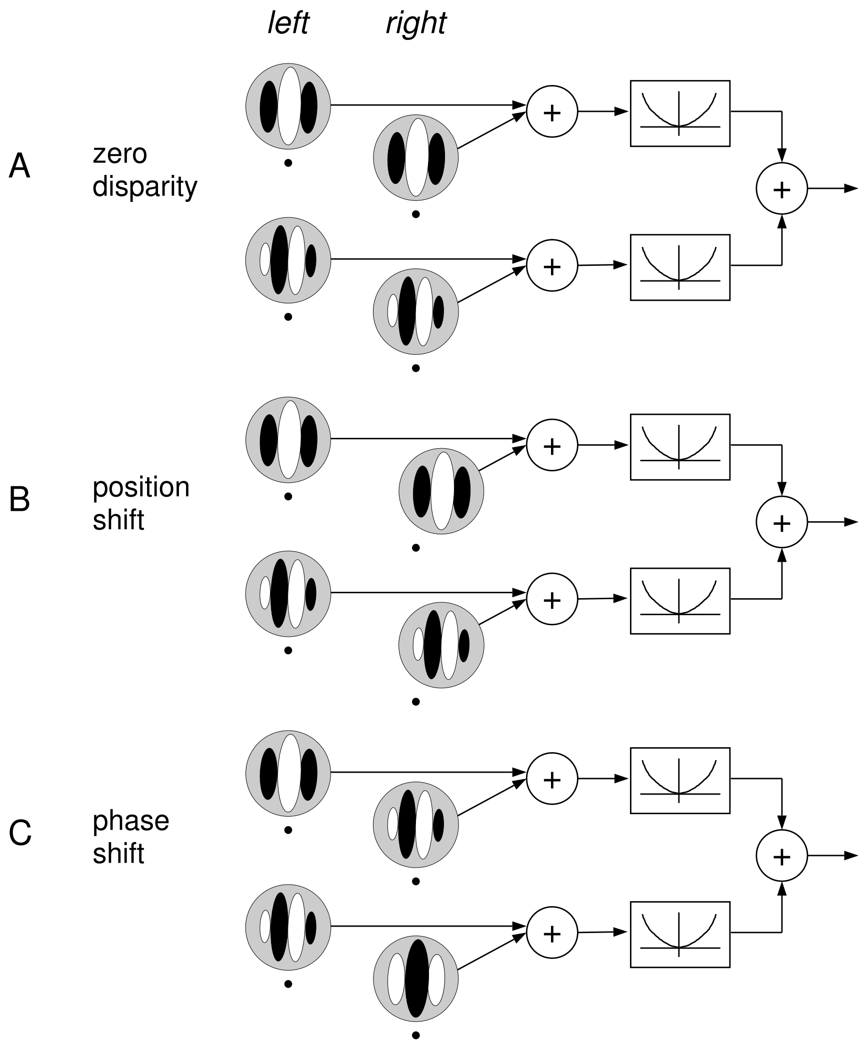

Figure 3: Disparity preferences of binocular energy neurons. (A) Zero disparity preference. (B) Non-zero disparity preference is introduced by shifting the positions of both right-eye receptive fields by the same amount (relative to the reference points). (C) Non-zero disparity preference is introduced by shifting the phases of both right-eye weighting functions by 90°.

Terms	Definitions
x	spatial or retinal position
$L(x)$, $R(x)$	complex-valued response of quadrature-pair, monocular, linear neurons in left- and right-eyes, at position x
$\mathrm{Re}[L(x)]$	real part of left monocular response (e.g. response of a linear neuron with an even-symmetric receptive field)
$\mathrm{Im}[L(x)]$	imaginary part of left monocular response (e.g. response of a linear neuron with an odd-symmetric receptive field)
$\rho_L(x)$	monocular (left-eye) amplitude signal (square-root of monocular energy)
$\phi_L(x)$	monocular (left-eye) phase signal
$k_L(x)$	left-eye instantaneous frequency at position x, equal to phase derivative $\phi'_L(x)$
s	receptive field position-shift
$\Delta\psi$	receptive field phase-shift
d	stimulus disparity
$\Delta\phi(x)$	interocular phase difference (equals $\phi_L(x) - \phi_R(x)$)
$E(x)$	binocular energy response at retinal position x
$E(x;\, s)$	response of binocular energy neuron with receptive-field position shift s
$E(x;\, \Delta\psi)$	response of binocular energy neuron with receptive-field phase-shift $\Delta\psi$
$E(x;\, s,\, \Delta\psi)$	response of binocular hybrid energy neuron with position-shift s and phase-shift $\Delta\psi$

Table 1: Symbol Table.

responses of a quadrature pair of linear neurons. For example, let f_l be a complex-valued weighting function in the left-eye like that in Fig. 1(C). In this example f_l consists of an even-symmetric weighting function that we call the real part of the complex-valued weighting function, and an odd-symmetric weighting function called the imaginary part. Similarly, let $L(x)$ denote the complex-valued response, where $\mathrm{Re}[L(x)]$ is the output of the real part of f_l, before the squaring step in the top half of Fig. 1(C), and $\mathrm{Im}[L(x)]$ is the output of the imaginary part, before the squaring step in the bottom half of Fig. 1(C).

3.2 Spatial Arrays of Identical Energy Neurons

Now consider a spatial array of binocular energy neurons that are identical to one another except for their receptive field locations. Let $L(x)$ and $R(x)$ be the complex-valued, monocular, linear responses for the left-eye and right-eye. The spatial array of energy responses, $E(x)$, can then be expressed as:

$$\begin{aligned} E(x) &= |L(x) + R(x)|^2 \\ &= (\mathrm{Re}[L(x)] + \mathrm{Re}[R(x)])^2 + (\mathrm{Im}[L(x)] + \mathrm{Im}[R(x)])^2 , \end{aligned} \tag{2}$$

where $(\mathrm{Re}[L(x)] + \mathrm{Re}[R(x)])^2$ corresponds to the top half of Fig. 1(D) and $(\mathrm{Im}[L(x)] + \mathrm{Im}[R(x)])^2$ corresponds to the bottom half of Fig. 1(D).

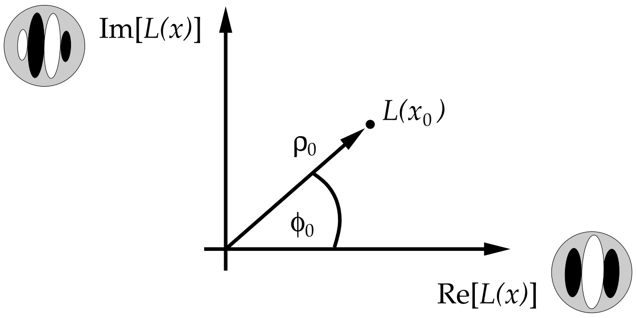

Figure 4: Illustration of the response of a monocular, quadrature-pair of linear neurons in terms of its real and imaginary parts, and its amplitude and phase. The real part (response of the even-phase, linear neuron) is indicated by the position along the horizontal axis. The imaginary part (response of the odd-phase, linear neuron) is indicated by position along the vertical axis. The amplitude is the radial distance from the origin, and phase is the angular coordinate.

In order to understand the binocular energy responses, we must introduce some additional terminology and notation, namely, we define the *monocular amplitude signal*, the *monocular phase signal*, the *interocular phase difference signal*, and the *instantaneous frequency signal*. These signals do not correspond directly to neural responses. Rather, they are implicit in the responses of a collection of model neurons.

We begin by expressing the monocular responses in polar coordinates, in terms of amplitude and phase. As illustrated in Fig. 4 we can write the left response as $L(x) = \rho_l(x)\, e^{i\,\phi_l(x)}$, where ρ_l^2 is the monocular energy,

$$\rho_l^2(x) = |L(x)|^2 \;=\; \mathrm{Re}[L(x)]^2 + \mathrm{Im}[L(x)]^2 \;, \tag{3}$$

and $\phi_l(x)$ is the phase angle of the complex-valued, monocular response, often written as

$$\phi_l(x) \;\;=\;\; \arg[L(x)] \;=\; \arctan(\mathrm{Im}[L(x)]/\mathrm{Re}[L(x)]). \tag{4}$$

The square root of the energy, $\rho_l(x)$, is called the monocular amplitude signal, and $\phi_l(x)$ is called the monocular phase signal. Note that the monocular amplitude and phase signals defined here are not the amplitude and phase spectra of a Fourier transform. Rather, they are a polar transformation on the responses of quadrature-pairs of linear neurons, at each retinal position. They are functions of spatial location x, not frequency.

Figure 5 shows an example. The stimulus in this case was a noise field in which the intensity varied randomly from one point to the next. The real part of the response is shown in Fig. 5(A); each point on this curve depicts the response of a single, monocular, linear neuron with an even-symmetric weighting function. The curve represents the responses of many linear neurons that are

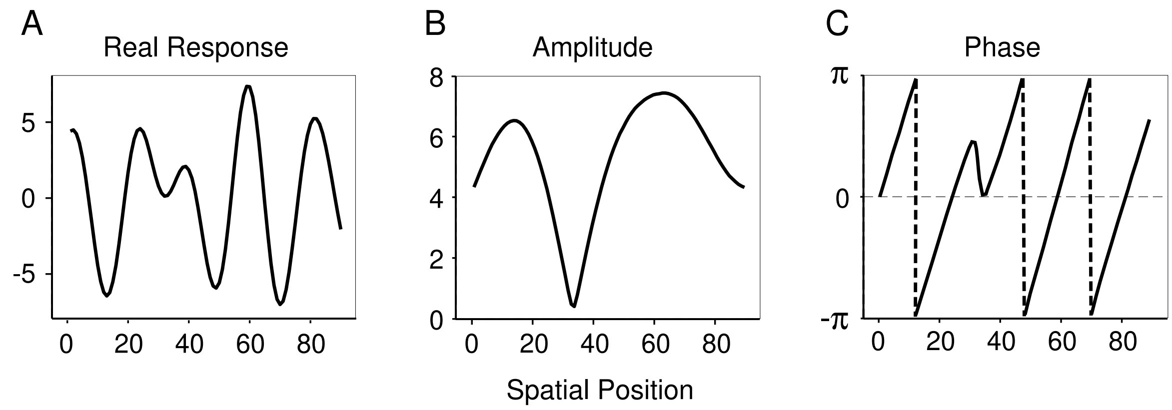

Figure 5: (A) Responses of a collection of monocular, linear neurons, all with even-symmetric weighting functions. The horizontal axis represents the spatial location of each successive linear neuron. (B) Monocular amplitude signal, computed as the square-root of the sum of squared linear neuron responses, the real parts of which are shown in (A). (C) Monocular phase signal, also computed from the linear responses. The amplitude signal typically changes slowly with x whereas the phase signal represents the fine structure of the responses. Phase is predominantly linear, and changes more rapidly as a function of spatial position. The vertical dashed lines reflect the fact that phase is uniquely defined between $-\pi$ and π radians (i.e., it wraps around every 360 degrees); these lines are *not* phase discontinuities. Small regions of unstable phase occur occasionally, like that shown near spatial location 35, and are discussed briefly in *Building Disparity Detectors*.

identical to one another except for their receptive field locations. Figure 5(B) shows the monocular amplitude signal. Each point on this curve corresponds to the square-root of the response of a monocular, energy neuron. These amplitudes were computed from the linear responses in Fig. 5(A) and the responses of a complementary set of odd-symmetric linear neurons. Figure 5(C) shows the monocular phase signal. The monocular amplitude signal usually changes slowly with x since the amplitude signal is the low-pass envelope of the convolution output. The phase signal, on the other hand, represents the fine structure of the responses.

With the amplitude and phase signals we can now simplify the expression for binocular energy. Substituting $\rho_l(x)\, e^{i\,\phi_l(x)}$ and $\rho_r(x)\, e^{i\,\phi_r(x)}$ for $L(x)$ and $R(x)$ in equation (2) gives:

$$E(x) \;=\; \rho_l^2(x) + \rho_r^2(x) + 2\rho_l(x)\rho_r(x)\,\cos(\Delta\phi), \tag{5}$$

where $\Delta\phi = \phi_l(x) - \phi_r(x)$ is called the *interocular phase difference*. Using this notation, it is clear that a binocular energy neuron generates a response that is the sum of the three terms: the two monocular energies, ρ_l^2 and ρ_r^2, and a term that is a cosinusoidal function of the interocular phase difference. The binocular energy response is independent of the monocular phases.

Note that throughout this exposition (and indeed, throughout this entire article) we assume that there are collections of neurons centered at each spatial position x, and that the responses of the linear neurons can be expressed collectively as a convolution operation. Alternatively, we could have defined $L(x)$, R(x) and $E(x)$ by fixing x to $x_\circ$, and varying the spatial location of the stimulus. This corresponds more closely to the usual preparation of single-cell neurophysiology in which one records from a single neuron while drifting a visual stimulus across its receptive field. This is important to remember, although it is often not critical to the arguments that follow.

3.3 Instantaneous Frequency

How does the binocular energy response depend on binocular disparity? If one varies disparity by slightly shifting the stimulus position in one eye, the monocular amplitude and phase signals of that eye are also shifted. The shift in the amplitude signal typically has negligible impact on the binocular energy response because the amplitude signal changes slowly with x (see Fig. 5B). Thus, the binocular energy response modulates with disparity based mainly on the interocular phase difference.

How does the interocular phase difference depend on disparity? Note in Fig. 5(C) that the monocular phase signal increases approximately linearly with spatial position x. For small changes in disparity, the interocular phase difference depends on the slope of this curve. If the phase signal is rising quickly then a small disparity (a small shift of the phase signal in one eye) will result in a large interocular phase difference. The derivative of the monocular phase signal is therefore critical, and it is often referred to as the *instantaneous spatial frequency* (Papoulis, 1965):

$$k_l(x) \;=\; \frac{d\,\phi_l(x)}{dx}. \tag{6}$$

Note that instantaneous frequency is not the same as the usual Fourier frequency, which has no explicit dependence on spatial position.

To help clarify this definition of instantaneous frequency, recall that the frequency of a sinusoid is the inverse of its wavelength, and its phase changes linearly from $-\pi$ to π over one wavelength. If the signal is $\cos(kx)$, then the phase signal is $\phi(x) = kx$, and the instantaneous frequency (the phase derivative) is k, constant at all positions x. A more interesting example is shown in Fig. 6. The frequency of this sinusoidal signal is constant on the left half, and then increases linearly as one moves further to the right. As frequency increases, the wavelength decreases and the phase signal begins to cycle more quickly between $-\pi$ and π; the instantaneous frequency captures this variation.

For more general inputs, the responses of a spatial array of linear neurons will modulate sinusoidally in small spatial neighborhoods. The instantaneous frequency in each neighborhood will be different. This local variation is what distinguishes instantaneous frequency from global Fourier frequency, which does not specify what frequency content is predominant in each neighborhood. Finally, from the quasi-linearity of monocular phase in Fig. 5(C), one can see that instantaneous frequency usually varies slowly across space. It also remains close to the preferred spatial frequency of the underlying linear neurons.

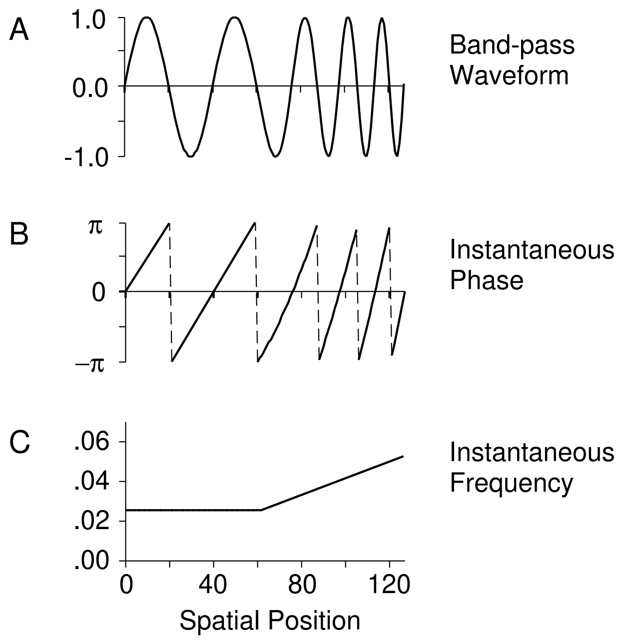

Figure 6: (A) Quasi-sinusoidal signal, the frequency of which varies with spatial location; the frequency is constant from locations 0 to 60, afterwhich it increases linearly. (B) The corresponding phase signal. (C) Instantaneous frequency, which is the derivative of the phase signal. When stimulus frequency is constant, the phase signal increases linearly and the instantaneous frequency is constant. When frequency increases, so does the slope of the phase signal, which is the instantaneous frequency.

3.4 Interocular Phase Difference and Instantaneous Frequency

As mentioned above, the binocular energy response depends on the interocular phase difference, and the phase difference depends on instantaneous frequency. Here, we formalize the relationship between interocular phase difference and instantaneous frequency.

For now, assume that the right-eye stimulus $I_r(x)$ is a shifted version of the left-eye stimulus $I_l(x)$: i.e., $I_r(x) = I_l(x - d)$, where the d is the disparity. When disparity d is positive $I_r(x)$ must be shifted to the left to match $I_l(x)$. With these inputs, the right monocular responses $R(x)$ are a shifted version of the left monocular responses $L(x)$, i.e., $R(x) = L(x - d)$. Similarly, the phase signals satisfy $\phi_r(x) = \phi_l(x - d)$. From this, one can reexpress the interocular phase difference using a Taylor series of $\phi_l(x - d)$:

$$\begin{aligned} \Delta\phi(x; d) &= \phi_l(x) - \phi_r(x) \\ &= \phi_l(x) - \phi_l(x - d) \\ &= d\,\phi_l'(x) + O[d^2] \end{aligned} \tag{7}$$

where $O[d^2]$ denotes all terms of second-order, i.e., d^2 and higher. In words, the interocular phase difference is (approximately) proportional to the product of disparity and the instantaneous frequency.

Combining this with equation (5) gives us a useful characterization of a binocular energy neuron. As the disparity is increased slightly above zero, the binocular energy response decreases as the cosine of disparity times instantaneous frequency, $\cos(d\phi')$. When the disparity is zero, the interocular phase difference is also zero. Zero disparity, therefore, produces a peak, $\cos(0) = 1$, in the binocular energy response. If the disparity becomes too large (more than half a wavelength of the preferred spatial frequency), second- and higher-order terms in the Taylor expansion become significant, and the approximation breaks down. In this case, a response peak in the binocular energy would no longer indicate a disparity of zero. These false peaks are discussed at length below in *False Energy Peaks*.

3.5 Position-Shift Model

We now introduce position shifts between the left and right monocular subfields of a binocular energy neuron to obtain nonzero disparity preferences. Toward this end, consider a binocular energy neuron whose right monocular subfield is shifted by a distance s compared to the retinal position of the left monocular subfield. From the formulation in equation (5), the energy response becomes

$$E(x;\, s) \;=\; \rho_l^2(x) + \rho_r^2(x+s) + 2\rho_l(x)\rho_r(x+s)\,\cos(\phi_l(x) - \phi_r(x+s))\,. \tag{8}$$

If we further assume that the right and left input signals were shifted versions of one another with disparity d, as above, then this equation becomes

$$E(x;\, s) \;=\; \rho_l^2(x) + \rho_l^2(x+s-d) + 2\rho_l(x)\rho_l(x+s-d)\,\cos(\phi_l(x) - \phi_l(x+s-d))\,. \tag{9}$$

Finally, if the receptive-field position-shift s is close to the disparity d, so that $\phi_l(x+s-d)$ is well approximated by its first-order Taylor series, i.e., $\phi_l(x+s-d) \approx \phi_l(x) + (s-d)k_l(x)$, where $k_l(x) = \phi_l'(x)$, then the binocular energy response simplifies to

$$\begin{aligned} E(x;\, s) &\approx \rho_l^2(x) + \rho_l^2(x) + 2\rho_l(x)\rho_l(x)\,\cos(k_l(x)\,(d-s)) \\ &= 2\,\rho_l^2(x)\,\left[1 + \cos(k_l(x)\,(d-s))\right]. \end{aligned} \tag{10}$$

This approximation makes use of the fact that instantaneous frequency changes slowly through space. It also relies on the fact that the amplitude signal is expected to change slowly with x, so that $\rho_l(x+s-d) \approx \rho_l(x)$.

The position-shift model posits that there is a population of energy neurons with different receptive-field position-shifts. The continuous binocular energy function $E(x;\, s)$ at each spatial position is therefore sampled at different values of s. One can see from equation (10) that the binocular energy function has a peak when the position shift s equals the stimulus disparity d. Therefore, the position shift of a binocular energy neuron is also the *preferred disparity* of that neuron.

The stimulus disparity may not exactly agree with the preferred disparity of any one neuron in the population. To find the peak in the binocular energy function, we must interpolate between the samples. This interpolation can be done exactly if there are enough samples (enough position shifts). From the analysis above we know that the energy response has a cosinusoidal shape at the peak. The frequency of this cosine is equal to the instantaneous frequency. This limits the rate of modulation of the binocular energy response to the same band of frequencies to which the linear neurons are responsive. This also tells us how finely one should sample $E(x;\, s)$, that is, how many and how closely spaced the different position-shifts must be so that the responses of linear neurons and energy neurons can be interpolated.

3.6 Phase-Shift Model

The phase-shift model involves differently shaped, monocular receptive fields in the left and right eyes, but centered at the same retinal position. Let $\Delta\psi$ denote a phase shift between the left- and right-eye weighting functions, i.e., $f_r(x) = f_l(x)\, e^{i\Delta\psi}$. In order to analyze the binocular energy behavior assume, as above, that the left and right input signals are shifted versions of one another with a disparity of d. The left and right linear responses are then related as follows:

$$\begin{aligned} R(x) &= e^{i\Delta\psi}\, L(x-d), \\ &= \rho_l(x-d)\, e^{i[\phi_l(x-d)+\Delta\psi]}. \end{aligned}$$

The phase difference $\Delta\phi(x)$ now has the form

$$\Delta\phi(x) = \phi_l(x) - \phi_l(x-d) - \Delta\psi\,.$$

If we use a Taylor series expansion as above, and assume that d is small (less than half a wavelength of the instantaneous frequency), then the phase difference can be approximated by

$$\phi_l(x) - \phi_r(x) \approx d\, k_l(x) - \Delta\psi$$

where $k_l(x) = \phi_l'(x)$ is the instantaneous frequency. Then the binocular energy response can be approximated by

$$\begin{aligned} E(x;\, \Delta\psi) &\approx \rho_l^2(x) + \rho_l^2(x) + 2\rho_l(x)\rho_l(x)\, \cos(d\, k_l(x) - \Delta\psi) \\ &= 2\,\rho_l^2(x)\, \left[1 + \cos(d\, k_l(x) - \Delta\psi)\right] \end{aligned} \tag{11}$$

The phase-shift model, like the position-shift model, posits that there is a population of binocular energy neurons. In this case, however, it is a population of neurons with different receptive-field phase-shifts $\Delta\psi$ so that the binocular energy response function $E(x;\, \Delta\psi)$ is sampled at each spatial position x and with several phase shifts $\Delta\psi$. This binocular energy function, like its analog in the position-shift model, has a peak for a particular shift. The peak in $E(x;\, \Delta\psi)$ occurs when the receptive-field phase-shift is equal to the product of stimulus disparity and instantaneous frequency

$$\Delta\psi \;=\; d\, k_l(x)\,. \tag{12}$$

For disparities d close to this value the response falls off cosinusoidally, with a frequency of $k_l(x)$.

Comparing equations (10) and (11), one can see that position-shifts and phase-shifts (s and $\Delta\psi$) have different effects on binocular energy responses. Unlike the position-shift model, a peak in the energy response of the phase-shift model does not completely constrain the disparity. It would be inaccurate to say that a phase-shifted binocular energy neuron has a unique preferred disparity. Rather, disparity is equal to the interocular phase difference divided by instantaneous frequency, quantities that depend both on the neuron's tuning properties and the stimulus. To estimate disparity using the phase-shift model, therefore, we must first estimate the instantaneous frequency. One could measure the instantaneous frequency as the derivative of phase, as in equation (6), or one could assume that the instantaneous frequency is equal to the neuron's preferred spatial frequency. The accuracy of this latter approximation depends on the expected distribution of possible instantaneous frequencies, which depends on the spatial frequency content of the stimulus and the bandwidth of the linear weighting functions. This is discussed in detail in *Frequency Uncertainty*.

Despite this difference between position-shifted and phase-shifted energy neurons, note that both exhibit a clear preferred disparity when tested with conventional bar stimuli. For bar stimuli, the instantaneous frequency will stay close to the neuron's preferred frequency. But, unlike the position-shift energy neuron, a phase-shift neuron's disparity tuning curve will not be symmetric about the central peak, though the central peak will be clearly evident.

3.7 Hybrid Model

One restriction on phase-shifted energy neurons stems from the fact that phase shifts are only unique between $-\pi$ and π. When combined with a restricted spatial frequency bandwidth, this means that for any one spatial frequency band, there is a limited range of disparities that one could hope to detect. The upper limits are reached as the phase-shift approaches $\pm\pi$ (i.e., half a wavelength) and the instantaneous frequency approaches the lowest spatial frequencies to which the neurons are responsive. This limitation of the phase-shift model is particularly restrictive for neurons tuned for high spatial frequencies. This leads us to consider a third model, a hybrid of the previous two, that would allow one to extend the range of disparities that phase-shifted energy neurons might detect.

In this hybrid model, binocular energy neurons have both a phase-shift of $\Delta\psi$ and a position-shift of s. With the same analysis used above, when the input disparity d is close to $s + [\Delta\psi / k(x)]$, where $k(x)$ is the instantaneous frequency, the energy response is given by

$$E(x;\, s, \Delta\psi) \;\approx\; \rho_l^2(x) + \rho_l^2(x) + 2\rho_l(x)\rho_l(x)\, \cos(k_l(x)(d-s) - \Delta\psi)$$

$$= \quad 2\,\rho_l^2(x)\;[1+\cos(k_l(x)(d-s)-\Delta\psi)] \tag{13}$$

This hybrid binocular energy response function, $E(x;\, s, \Delta\psi)$, depends cosinusoidally on both the position shift s and the phase shift $\Delta\psi$. The hybrid model posits that there is a population of binocular energy neurons with different receptive-field position- and phase-shifts so that $E(x;\, s, \Delta\psi)$ is sampled at each spatial position x, with several position shifts s, and with several different phase shifts $\Delta\psi$. We discuss in *Building Disparity Detectors* how this collection of neural responses can be pooled to estimate disparity unambiguously.

4 Model Predictions and Comparisons with Existing Data

The position-shift model involves binocular combinations of monocular receptive fields of similar shape at different retinal positions, while the phase-shift model combines monocular receptive fields with different shapes from the same retinal location. Only when both are tuned to a disparity of zero are they strictly equivalent. The next sections review neurophysiological evidence for position shifts and phase shifts. We also present predictions of the two "pure" models and the hybrid model.

4.1 Distribution of Preferred Disparities

In attempting to measure the range of preferred disparities caution must be taken because eyes tend to drift and rotate under anaesthesia. To control for this, Hubel & Wiesel (1970) introduced the reference-cell method, in which a binocular cell is recorded for an extended period to find the disparity that elicits a maximal response. A second electrode is used to record from other neurons. By adjusting disparity settings to maintain the maximal response from the reference cell, one can track eye movements. Interestingly, it is not necessary to track eye drift in the owl, as their eye movements are negligible (Steinbach & Mooney, 1973; du Lac *et al.*, 1990).

If a broad distribution of preferred disparities is found in a sample of neurons, relative to their preferred spatial frequencies, then one can infer that position shifts occur. In the cat, early reports gave a range of $\pm 3^\circ$ for the distribution of preferred disparities (Barlow *et al.*, 1967). Later studies using a reference-cell method found that the range of preferred disparities of tuned-excitatory cells in area 17 is less than 1° for eccentricities up to 8° (Ferster, 1981; LeVay & Voigt, 1988). In the owl, the range of preferred disparities was found to be $\pm 2.5^\circ$ (Pettigrew, 1979). In anesthetized monkeys, cells with preferred disparities up to $30'$ were documented in V2 (Hubel & Wiesel, 1970). Studies on awake, behaving monkeys seldomly found preferred disparities greater than $12'$ (crossed or uncrossed) for eccentricities within 2 degrees of the fovea (Poggio & Fischer, 1977). One would expect that cells in the parafoveal region might have larger preferred disparities, but we are aware of no quantitative data regarding this issue. In the monkey, near and far cells often respond maximally at the largest disparities that have been tested (up to 1°) (Poggio & Fischer, 1977). Near and far cells of cats cover a range of at least $\pm 5^\circ$ of disparity (Ferster, 1981; LeVay & Voigt, 1988).

Unfortunately, spatial frequency tuning has usually not been measured along with disparity tuning. However, data from Ohzawa and Freeman (1986a,b) suggest that the range of preferred spatial frequencies in disparity-sensitive cells is similar to the overall range of preferred spatial frequencies in cat area 17. If we assume the same in the monkey, with foveal simple and complex cells having preferred spatial frequencies between 1 and 10 cpd (DeValois *et al.*, 1982), then one can indirectly conclude that in monkeys, cats, and owls the preferred disparities cover a range that is larger than one period of the typical spatial frequency preference. This suggests that position-shifts do occur, but it does not rule out an additional contribution from phase-shifts.

4.2 Monocular Receptive-Field Shape

The results reviewed so far imply that there are position-shifts. However, receptive-field shape was only crudely characterized in these studies. To determine whether there are phase-shifts, a more elaborate method is required. One method is to directly examine the shapes of the monocular receptive fields using white noise stimuli and reverse-correlation procedures. Ohzawa *et al.* (1991) and DeAngelis *et al.* (1991, 1995) applied this method to simple cells in cat area 17. They then fitted Gabor functions to the monocular receptive fields and used the phase of the fitted Gabor functions as a measure of receptive field shape. They found that the monocular receptive field shapes of binocular cells are often different. Moreover, the differences depend on orientation; cells tuned to horizontal orientations have similar receptive field shapes, while cells tuned to near vertical orientations exhibit a wide range of phase shifts (from 0° to 180°). While these data show that phase shifts exist, the existence of additional positional-shifts cannot be excluded.

This reverse-correlation procedure works well for simple cells, the monocular responses of which depend strongly on the position of a stimulus within the receptive field. More sophisticated procedures, analyzing higher-order kernels of the white noise responses, would be needed to determine the monocular receptive field properties underlying disparity selectivity of complex cells.

4.3 Binocular Dependence on Spatial Frequency

Another way to examine the mechanism of disparity selectivity is to measure the dependence of disparity-tuning curves on the spatial frequency of a sine-grating stimulus. Several studies have shown that simple and complex cells respond with sinusoidal disparity-tuning curves when presented with drifting sinusoidal gratings (Ohzawa & Freeman, 1986a, 1986b; Hammond, 1991; Wagner & Frost, 1994). The binocular linear neurons and energy neurons behave similarly. We next describe the behavior of the position-shift, phase-shift and hybrid models with sinusoidal stimuli. The goal is to describe experimental predictions for each model. Although we concentrate on the predictions for binocular energy neurons, these predictions also hold for binocular linear neurons.

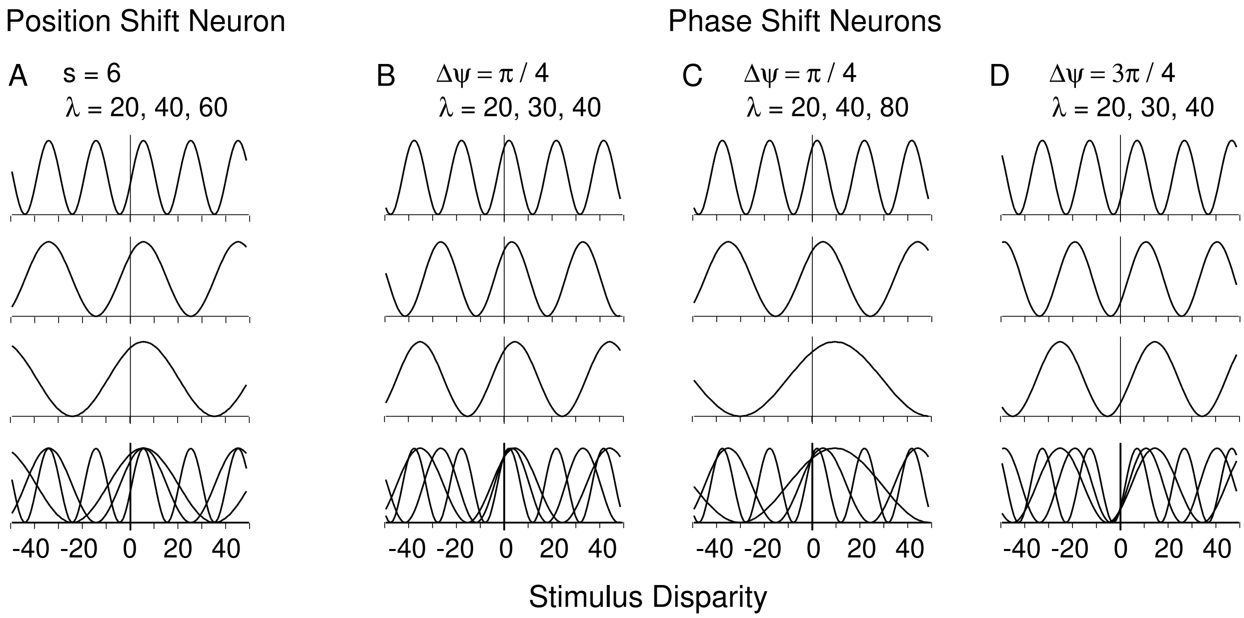

Figure 7: Disparity tuning curves of energy neurons when presented with drifting sinusoidal gratings. (A) A position-shift energy neuron with a preferred disparity of $s = 6$, and input wavelengths spanning 1.5 octaves. The bottom panel shows the superposition of 3 tuning curves, where the peak at the preferred disparity occurs at all wavelengths. (B, C, and D) Phase-shifted energy neuron responses with different phase-shifts, and different frequency ranges. With small phase shifts and a small frequency range (as in B) only a slight, but systematic, frequency dependence is evident. With a larger frequency range (as in C) or a larger phase-shift (as in D), the dependence of peak disparity on frequency is more evident. Phase-shifted energy neurons do not have a unique preferred disparity.

Position-Shift Neuron

Let the left and right stimuli be sinusoidal gratings with spatial frequency k_0 and disparity d:

$$I_l(x) \;=\; \sin(k_0 x) \quad , \qquad I_r(x) \;=\; \sin(k_0(x-d)). \tag{14}$$

For a spatial array of linear neurons we get an array of responses that vary sinusoidally as a function of retinal position. When there is no receptive-field phase-shift, i.e., when $\Delta\psi = 0$, it follows from the analysis in *Formalizing The Model* that the binocular energy responses are

$$E(x;\, s) \;=\; 2\,\rho_l^2\;[1+\cos(k_0(d-s))]. \tag{15}$$

The monocular energy, ρ_l^2, is constant for sine-grating stimuli, independent of spatial position. The instantaneous frequency is also constant and equal to the stimulus frequency k_0. Peaks in the energy response occur whenever the cosine term is equal to one. This happens when the disparity satisfies

$k_0(s-d) = n2\pi$, for integer values of n, i.e., when

$$d \;=\; s + \frac{n2\pi}{k_0}. \tag{16}$$

Because n can be any integer, peaks occur periodically, spaced by one wavelength of the stimulus frequency. One of these peaks always occurs when the disparity equals the position-shift s, independent of the frequency of the input (Fig. 7A).

Phase-Shift Neuron

When there is a receptive-field phase-shift $\Delta\psi$ but no position-shift the energy response reduces to

$$E(x;\,\Delta\psi) \;=\; 2\,\rho_L^2\;[1+\cos(d\,k_0-\Delta\psi)] \tag{17}$$

Peaks in the energy response, $E(x;\,\Delta\psi)$, now occur when $d\,k_0 - \Delta\psi = n2\pi$, or equivalently, when

$$d \;=\; \frac{\Delta\psi}{k_0} + \frac{n2\pi}{k_0} \tag{18}$$

As illustrated in Fig. 7(B)-(D), this means that the neuron's "disparity preference" depends, in a systematic way, on the stimulus frequency. The distribution of "disparity preferences" depends on the neuron's spatial frequency tuning bandwidth, and on the neuron's interocular phase shift.

Thus, one way to discriminate the two "pure" models is to measure disparity tuning curves for sine-grating stimuli with different spatial frequencies. For the position-shift model, peaks in the tuning curves occur at the same disparity (position shift) for all frequencies. For the phase-shift model, peaks in the tuning curves will occur when the two gratings have a certain phase difference (but a fixed disparity) for all frequencies. Data of this sort have been obtained for the owl (Wagner & Frost, 1993, 1994), and are more consistent with the position-shift model.

Hybrid Neuron

With a hybrid energy neuron, following equation (13), the energy response is given by

$$E(x;\,s,\Delta\psi) \;=\; 2\,\rho_l^2\;[1+\cos(k_0(d-s)-\Delta\psi)] \tag{19}$$

It is a cosinusoidal function of disparity, with a frequency k_0, and a phase-offset of $k_0 s + \Delta\psi$. The phase offset is a linear function of frequency; the slope is the position-shift and the intercept is the phase-shift. Therefore, to measure both the phase-shift and the position-shift of a single neuron, one can record disparity tuning curves with drifting sinusoidal gratings of different spatial frequencies k_j. A cosinusoidal function with frequency k_j can then be fit to each tuning curve, from which a phase offset, denoted by Ω_j, is obtained (see Fig. 8A). One then plots the phase offset as a function

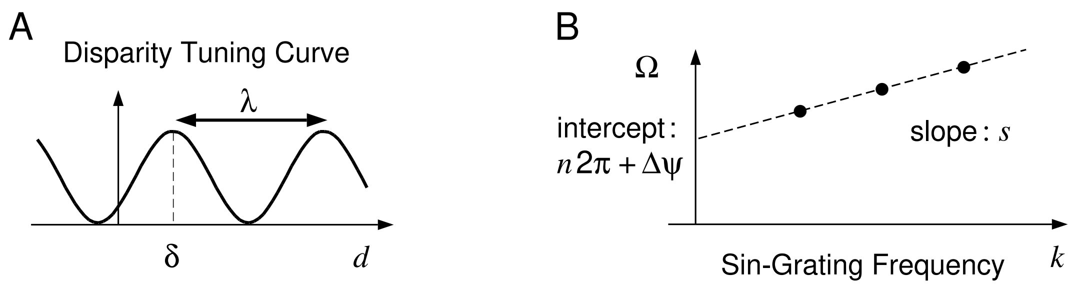

Figure 8: (A) Cosinusoidal disparity tuning curve of a hybrid energy neuron. With the peak closest to the origin (zero disparity) at δ and an input spatial wavelength of λ, the phase-offset of the tuning curve is $\Omega = 2\pi\delta/\lambda$. (B) This phase-offset is a linear function of frequency $k = 2\pi/\lambda$, the slope and intercept of which give the position-shift and the phase-shift of the monocular receptive fields.

of spatial frequency as depicted in Fig. 8(B), and fits a linear function to the data. The slope of the linear fit gives the positional shift s. The intercept of the linear fit on the vertical axis will be equal to the phase-shift plus an integer multiple of 2π. Because $\Delta\psi$ is only unique within $-\pi$ and π, one can find the phase-shift from the intercept by adding whatever multiple of 2π is required to bring the result into the range $(-\pi, \pi]$.

Similar methods have been used to explore the encoding of interaural time differences in the auditory system of cats (Yin & Kuwada, 1983). In the visual system this method can be used to measure position- and phase-shifts in simple and complex cells, without requiring that the monocular receptive fields shapes be accurately localized or described. It is necessary however that disparity be defined with stabilized eyes or a reference cell.

5 Encoding of Disparity

The next major issue to address is the way in which disparity might be inferred from the responses of binocular energy neurons. Ohzawa *et al.* (1990), DeAngelis *et al.* (1991), and Qian (1994) have suggested that disparity is encoded directly by peaks in the responses of phase-shifted binocular energy neurons, and therefore called them ideal disparity detectors. We show here that disparity is not directly encoded by peaks in the responses of binocular energy neurons. Rather, binocular energy responses are ambiguous; for both the position- and phase-shift models false peaks are inevitable, and for the phase-shift model there is a further uncertainty about disparity.

5.1 False Energy Peaks

For binocular energy-neurons, response peaks may occur even when the input disparity is outside the range of disparities to which the neuron was thought to respond. One example of this is given above in *Binocular Dependence on Spatial Frequency*, where, for sinusoidal grating stimuli, both position-shifted and phase-shifted energy neurons respond periodically as a function of disparity. Peaks in their disparity tuning curves occur every wavelength.

Of course this is a contrived example because the stimulus is periodic. But the problem of false peaks is more general than this, lying not with the periodicity of the stimulus, but with the quasi-periodicity of the responses of the underlying linear weighting functions. False peaks occur for almost all stimuli, including white noise. In fact, some of the false peaks will be significantly larger in magnitude than the peak at the correct disparity.

To illustrate this, Fig. 9 shows responses of two binocular energy neurons. The left column shows the responses of an energy neuron with a disparity preference of zero. The right column shows responses of an energy neuron with a position shift of zero and a phase shift of $\pi/2$. The top 4 panels in each column (Figs. 9(A) and (B)) show responses, as a function of stimulus disparity, for individual samples of white noise. The bottom panels (Figs. 9(C) and (D)) show the average responses, averaged over many samples of white noise. False peaks are clearly evident for individual trials (Figs. 9(A) and (B)). But when responses are averaged over many stimulus presentations, one obtains tuning curves that do not necessarily show the false peaks.

In typical single-cell recordings, stimuli are swept over the receptive field of the cell, with different disparities on different trials. This produces a two-dimensional response surface, with disparity on one axis, and time (or spatial position) on the other. Conventional disparity-tuning curves plot average firing rate (computed by averaging across time). The simulations in Figs. 9(A) and (B) show the instantaneous simulated responses of a single neuron to different noise inputs; equivalently, each plot can be viewed as a slice through the two-dimensional response surface at widely separated times. The disparity tuning curves in Figs. 9(C) and (D) are the average responses over time, and therefore correspond more closely to the conventional data analysis. In other words, the methods used in physiology experiments will often hide the presence of false peaks.

Therefore, one may only see false peaks in disparity-tuning curves for very short stimulus presentations, or when one analyzes the spike train by computing average firing rate over very short time intervals. When conventional bar stimuli are used to measure disparity tuning curves, the false peaks are sometimes evident as small side-lobes that flank the central peak (LeVay & Voigt, 1988). These are expected to occur in model neurons when the bandwidth is reasonably small (e.g., less than 1.5 octaves). Larger false peaks occur with richer stimuli like textured surfaces, random-dot stereograms, and white noise. Only when stimulus disparities are known to be small, as in the implementation reported by Qian (1994), can these false peaks be ignored.

False peaks occur because the linear neurons are spatial frequency selective. Recall that a peak in the binocular energy response occurs when the interocular phase difference $\Delta\phi(x)$ is zero; i.e., whenever the monocular phase signals from left and right linear neurons are the same. Because of the spatial frequency selectivity of the linear weighting functions, the monocular phase signals cycle between $-\pi$ and π as a function of spatial position. For example, if the left phase signal at x_0 is $\phi_l(x_0)$, then one can expect the phase at the nearby position $\phi_l(x_0 + \lambda)$ to be almost the same,

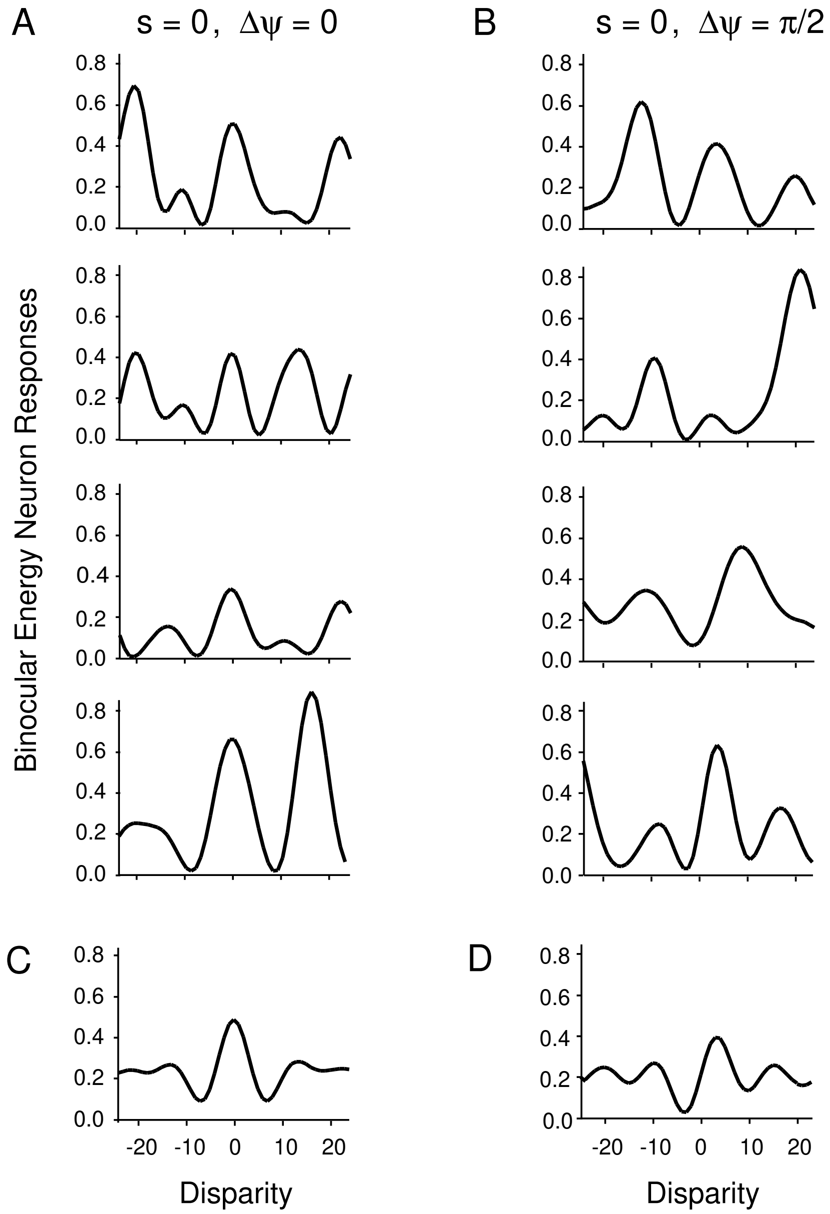

Figure 9: Disparity tuning curves of binocular energy neurons for white noise stimuli. The top 4 panels in each column (A and B) show responses to individual samples of white noise. The bottom panels (C and D) show the average response over many samples of white noise, or equivalently, the average response as the white-noise stimuli are swept across the neuron's receptive field. (A, C) A binocular energy neuron tuned for zero disparity, with a preferred wavelength of 16. One can see false peaks in (A) with approximately that periodicity. (B, D) Responses of a binocular energy neuron with the same frequency tuning, and a phase-shift of $\Delta\psi = \pi/2$, i.e., tuned for disparities equal to $\Delta\psi/k_0 = 4$. False peaks are evident in (B) with approximately the same periodicity. The average responses over many stimulus presentations (C and D) do not show the false peaks.

where λ is one wavelength of the preferred spatial frequency. Thus the phase signal in one eye will usually equal the phase signal at several spatial locations in the other eye. Within a population of binocular energy neurons with different disparity preferences, peaks occur whenever the left and right monocular phase signals have similar values; there will in general be more than one stimulus disparity for which this occurs. The distribution of these false peaks will depend on the frequency tuning of the linear neurons, with false peaks occurring approximately every wavelength on average.

Binocular energy neurons produce peak responses at their putative preferred disparities more often than at other disparities. But some of the false peaks will be larger than the response at the preferred disparity. To explain this, one can see from equation (5) that the magnitude of an energy peak depends on the monocular energies. False peaks are larger than the correct peak whenever the monocular energies at the false peaks are larger than at the correct matching position.

5.2 Frequency Uncertainty

A further problem for phase-shift neurons is caused by the dependence of response peaks on instantaneous frequency. Given a peak response of a phase-shifted energy neuron, the disparity is equal to the phase-shift divided by the instantaneous frequency. One might suppose that instantaneous frequency is encoded by another population of neurons, and then used to compute disparity. Alternatively, one might assume that the instantaneous frequency is almost equal to the neuron's preferred spatial frequency. The latter approach, although simpler, introduces uncertainty about the instantaneous frequency, and therefore about the disparity.

To examine the extent of this uncertainty we derived an expression for the probability density function for the resulting disparity estimates. We calculated these probability densities for binocular energy neurons with Gabor weighting functions of various bandwidths and various interocular phase shifts, responding to mean-zero white Gaussian noise stimuli. Detailed formulae are given in the appendix, and Fig. 10 provides an example. Fig. 10(A) shows the density function for instantaneous frequency of Gabor-filtered white noise (solid curve). The bandwidth of the Gabor function in this example is 1.0 octave. Figure 10(B) shows the behavior of the mean and standard deviation of disparity estimates as functions of the phase-shift. The solid line is the true disparity.

The conclusions one can draw from this analysis and our simulations are straightforward. First, the uncertainty in disparity increases with the filter bandwidth because a larger bandwidth yields a broader distribution of instantaneous frequencies. Second, the uncertainty also grows with the magnitude of the phase-shift; one can say less about disparity from a peak response of an energy neuron as the receptive field phase-shift increases.

6 Building Disparity Detectors

Energy neurons respond quasi-periodically as a function of disparity, depending on the stimulus and the neuron's spatial frequency tuning. Also, phase-shifted energy neurons do not have unique

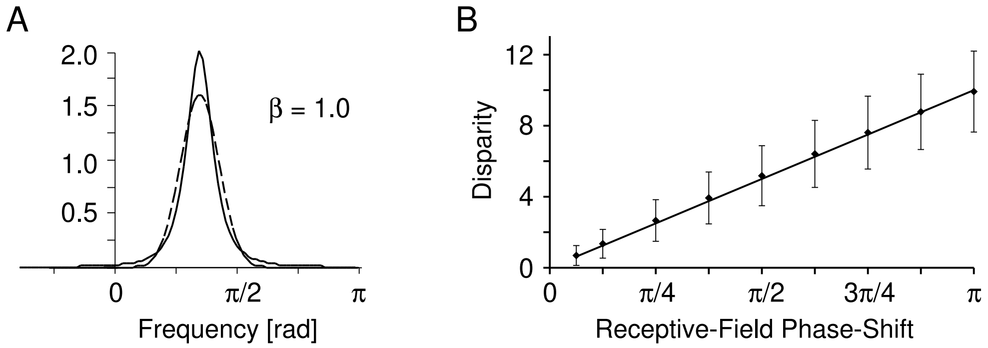

Figure 10: (A) Probability density function for instantaneous frequency of Gabor-filtered white noise (solid curve), with the Gabor power spectrum (dashed curve). The bandwidth, defined by one standard deviation of the Gabor amplitude spectrum, is 1.0 octave. (B) Means and standard deviations of disparity estimates as functions of the phase-shift. The solid line is the true disparity.

preferred disparities. Disparity could be computed from the phase-shift with the instantaneous frequency, but this would necessitate a population of neurons to encode instantaneous frequency, and a divisive nonlinearity.

Here we propose that energy-neuron responses can be pooled linearly (i.e., summed) over several scales and orientations, and in local spatial neighborhoods, in order to build disparity detectors. The theoretical basis for the pooling is the theory of phase-correlation (Fleet, 1994), and the analysis of phase stability (Fleet & Jepson, 1993). Consider a population of energy neurons at a single retinal location, with the same preferred spatial frequency and orientation, but with different disparity preferences. That neuron tuned for the stimulus disparity will respond vigorously. But other neurons in the population might also respond with false peaks. Now consider other neurons with different spatial frequency and orientation preferences. Neurons in these other populations that are tuned for the stimulus disparity will again respond vigorously, and again there will be other neurons with large false peak responses. Thus, when we pool the responses across different scales and orientations, the large responses near the correct disparity sum to produce an even larger peak. Conversely, since the false peaks are distributed over a broad range of different disparities, the false peaks in the response of one neuron will cancel with local minima in the responses of others.

Pooling over different spatial frequencies is critical for attenuating false peaks. Recall that the expected interval between false peaks is approximately the wavelength of an energy neuron's preferred spatial frequency (see Fig. 9). Thus, the false peaks at different frequencies occur at different disparities. Pooling across enough scales yields a prominent peak only at the stimulus disparity. This pooling embodies some of the advantages of the coarse-to-fine algorithms typical in computer vision (Marr & Poggio, 1979), but without the sequential (one scale at a time) processing. It is more like a coincidence strategy, where the peaks coincide through scale at the correct disparity. Transparency may also occur, where more than one significant peak remains after pooling.

Pooling over orientation also helps to boost the correct peak and attenuate false peaks. With textured stimuli (e.g., textured surfaces, random dot stereograms, or white noise), where false peaks are prominent, neurons tuned to different orientations provide nearly independent responses. Therefore, false peaks are expected to occur at different disparities.

Localized spatial pooling plays a somewhat different role. Normally, the amplitude and instantaneous frequency signals are low-pass, and we expect interocular phase differences to also change slowly, at least for slowly changing disparity. Under these normal circumstances, energy responses are correlated over local spatial positions and spatial pooling does very little. However, there are some circumstances when the monocular phase signal is unstable (fluctuates rapidly from one spatial position to the next) and it is particularly sensitive to small distortions and scale changes between the left- and right-eye views (Fleet & Jepson, 1993). In these cases, the interocular phase difference is an unreliable measure of binocular disparity. Spatial pooling will tend to attenuate the rapid fluctuation of the phase signals and suppress unreliable energy responses (Fleet, 1994).

The binocular energy neurons described above in *Models of Disparity Selectivity* sum the squared responses of a quadrature pair of linear neurons (Figs. 1C and D). With spatial pooling, however, a binocular energy neuron computes a local spatial average of the quadrature pair of linear neurons. Emerson *et al.* (1992) found that spatial pooling of this sort was needed to explain complex cell responses. This also agrees for the larger extent of complex cell receptive fields. For the simulations reported below we used a Gaussian spatial weighting function so that the energy neuron receptive fields were 50% larger than those of the linear neurons.

Pooling of Position-Shifted Energy Neurons

Pooling is straightforward for the position-shift model. In our simulations, neurons with the same preferred disparity are summed over four spatial frequency bands (octave bandwidths and octave spacing) and over three orientation bands.[1] Figure 11 shows an example simulation result. The top row shows responses of one neuron at each spatial frequency, after pooling over space and orientation. The stimulus was white noise. Although pooling over space and orientation helps to attenuate the false peaks, they still occur. When these responses are summed across scale, however, the ambiguity is essentially removed. The remaining peak occurs at the intended disparity ($d = 4$ in this case).

Figure 12 summarizes the results of 300 simulations, using different samples of white noise on each trial. The histograms in Fig. 12 show the percentage of times, of the 300 trials, that the peak response occurred at each disparity. Figure 12(A) shows the histogram for a single binocular energy neuron with a position shift of 4 pixels, tuned to a wavelength of 8 pixels. For only 18% of the trials did the peak response fall within one pixel of the intended "preferred" disparity. False peaks dominated the other 82% of the trials. Figure 12(B) shows the result after pooling over orientation

[1] To simulate the pooling over three orientations at each scale, we summed three statistically independent responses at each scale, assuming that neighboring orientation-tuned operators are largely non-overlapping in their tuning. This allowed us to perform the simulations with 1d signals, avoiding the computational expense of two spatial dimensions.

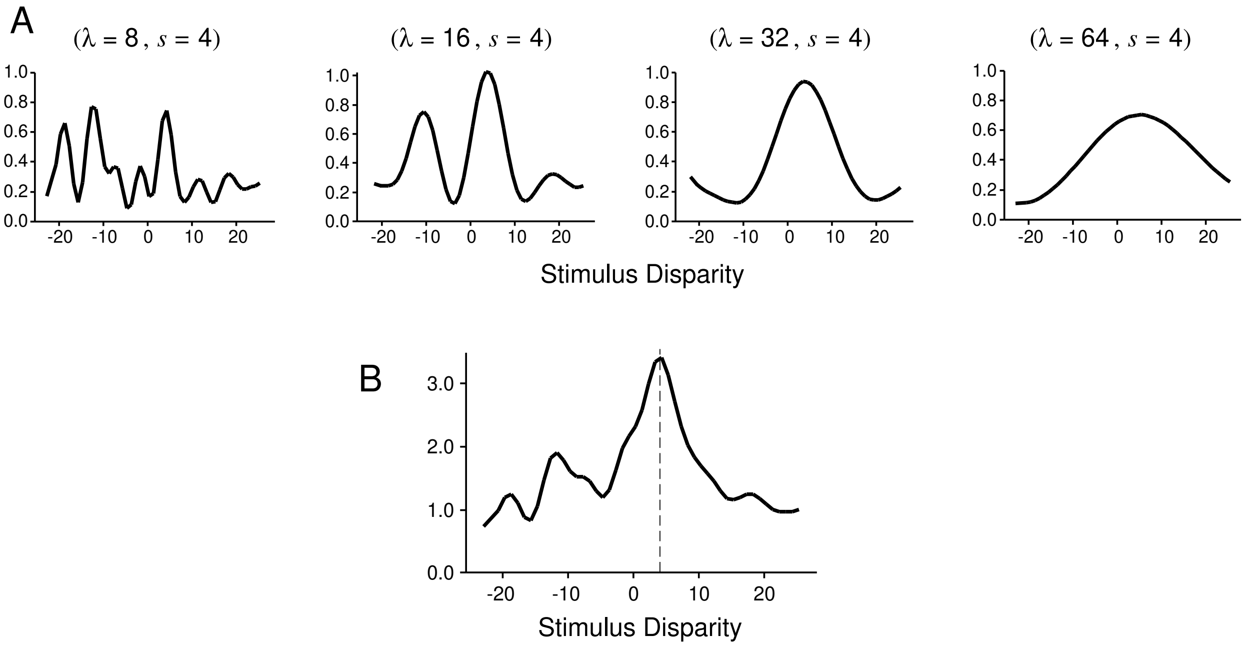

Figure 11: (A) Responses of position-shifted energy-neurons to white noise stimuli. Each curve is the response, as a function of stimulus disparity, after pooling over local space and over three orientation bands. Different panels correspond to different spatial frequency bands. The preferred wavelengths of the four spatial frequency bands are given; each has a preferred disparity of 4 pixels. (B) Response after summing the four panels in (A). When peaks coincide across scales the pooled response sums constructively, and at other disparities the response peaks and troughs cancel. The coarse-scales are particularly useful in isolating the appropriate peak, and the fine scales help make it sharper.

and spatial position. Figure 12(C) shows the result after pooling over scale as well. The concentration of peaks at one disparity shows that we have successfully constructed a disparity detector.

Pooling of Phase-Shifted or Hybrid Energy Neurons

With phase-shifted energy neurons response peaks do not uniquely determine the disparity unless the instantaneous frequency is also known. We posit that one might assume that the instantaneous frequency is equal to a neuron's preferred spatial frequency, and that response pooling can alleviate the problem of frequency uncertainty, in addition to attenuating false peaks.

As above, we sum the responses of energy neurons that we expect to produce a peak at or near one particular disparity. Assuming that instantaneous frequency is close to the neuron's preferred frequency, we pool responses using the following rule: let the disparity represented by a phase-shifted energy neuron equal $\Delta\psi\,\lambda/(2\pi)$, where $\Delta\psi$ is the neuron's receptive-field phase-shift and λ is the wavelength of the neuron's preferred spatial frequency. For example, if the phase shift is $\pi/2$ and the preferred spatial frequency is 1 cycle/degree (cpd), then the neuron represents a disparity of 1/4

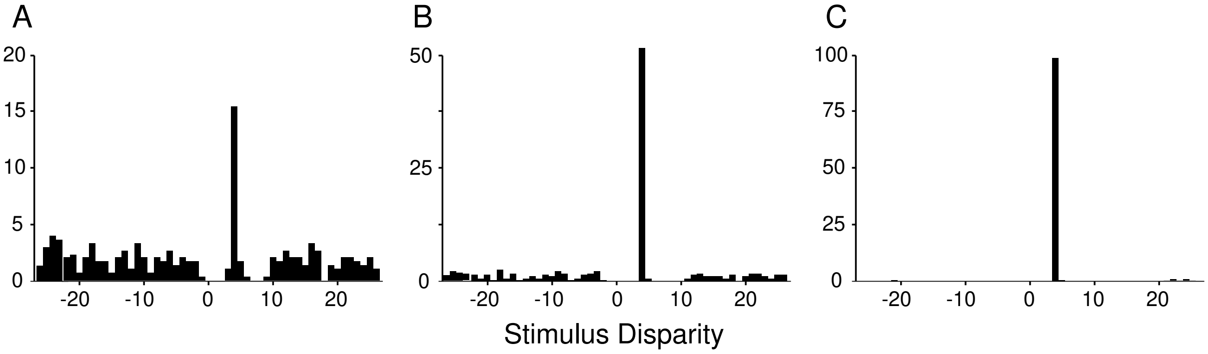

Figure 12: Simulations like that in Fig. 11 were run 300 times with statistically independent white noise stimuli. Position-shifted energy neurons were tuned for a disparity of 4 pixels. Each histogram plots the percentage of times that the peak response occurred at each disparity. (A) Responses of a single energy neuron shows a wide range of false peaks, with only 18% of the peaks falling within one pixel of the correct disparity. (B) Pooling over orientation and space yields a sharper concentration of peaks, but still only 52% of the peaks within one pixel of the correct disparity. (C) After pooling over scale, orientation, and space, 99% of the peaks lie at the correct disparity.

degree. Likewise, if the phase shift is π and the preferred spatial frequency is 2 cpd, then the neuron again represents a disparity of 1/4 degree. The responses of these two neurons could be pooled.

An example simulation result is shown in Fig. 13. One can see that the pooling largely eliminates the false peaks. It also enhances the peak near the intended disparity. Although false peaks still exist at individual scales they are attenuated when pooled across scale.

Of course, one consequence of assuming that the instantaneous frequency is equal to the preferred spatial frequency is that peaks do not occur exactly at the correct stimulus disparity. However, as one pools over larger spatial neighborhoods, and especially over different orientations, the distribution of peak locations should be centered upon the correct disparity. One might conclude that a greater amount of spatial pooling is therefore needed for the phase-shift model than for the position-shift model. However, for the simulations here we used the same spatial pooling in both simulations. One can see, as a consequence, that the peak is not as sharply defined in Fig. 13(B) as it is in Fig. 11(B).

We also carried out 300 simulations of this phase-shift pooling, using statistically independent samples of white noise on each trial. The results are summarized in Fig. 14. Each histogram in Fig. 14 shows the percentage of times the peak occurred at each disparity. Figure 14(A) shows the histogram of response peaks for the finest scale, after pooling over orientation and space. The histogram shows two concentrations of peaks because the preferred disparity (4 pixels) corresponds to a phase shift of π at this scale. Whenever the instantaneous frequency is higher than the neuron's preferred frequency, then a disparity of 4 pixels will be more than half a wavelength, leading to a disparity estimate of similar magnitude but in the opposite direction (a form of aliasing).

Figure 14(B) shows the result of pooling over space, orientation, and scale. The largest con-

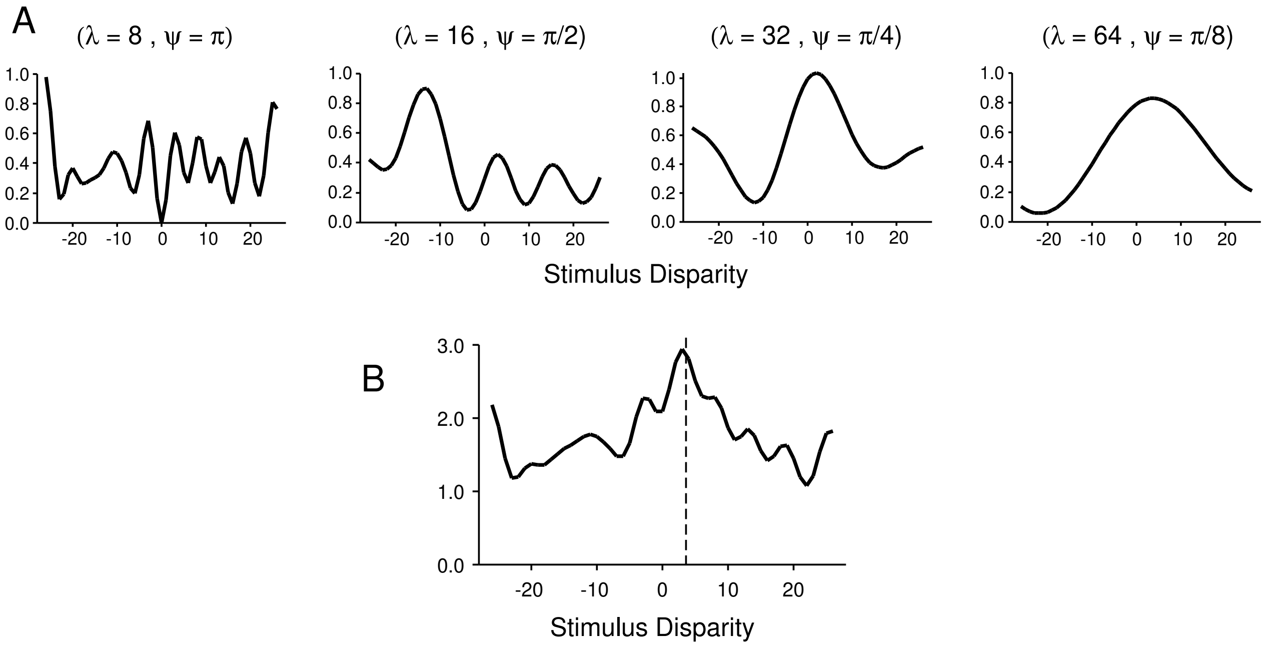

Figure 13: Same format as Fig. 11. (A) Responses from phase-shifted energy-neurons tuned to white noise stimuli. The receptive-field phase-shifts and the preferred wavelengths of the four spatial frequency bands are given. For each spatial frequency, the "preferred" disparity (phase-shift divided by preferred wavelength) was 4 pixels. (B) Sum of the four responses in (A).

centration of responses is at the intended disparity (4 pixels in this case), so the problems of false peaks and frequency uncertainty are largely eliminated. This histogram is, however, not as sharp as the one we obtained from the position-shift model (Fig. 12C), and it is slightly biased toward a disparity less than 4 pixels. The reason for this, as discussed above, is that a relatively few number of energy neurons are being summed, so that uncertainties about the instantaneous frequencies in a small number of neurons has a noticable effect in each simulation. When we use a larger amount of spatial pooling in the model, then the peak in the histograms, as well as a the peak in the disparity tuning curves (Fig. 14B), are sharper and less biased.

7 Discussion

To understand the neural basis for stereoscopic vision one must address several issues including the form of binocular interaction in simple and complex cells, the basis for their disparity selectivity, and the way in which they encode disparity. This article examines a model of binocular interaction based on binocular linear neurons and binocular energy neurons. Disparity selectivity of the model neurons arises from a combination of position-shifts and phase-shifts between the monocular subfields

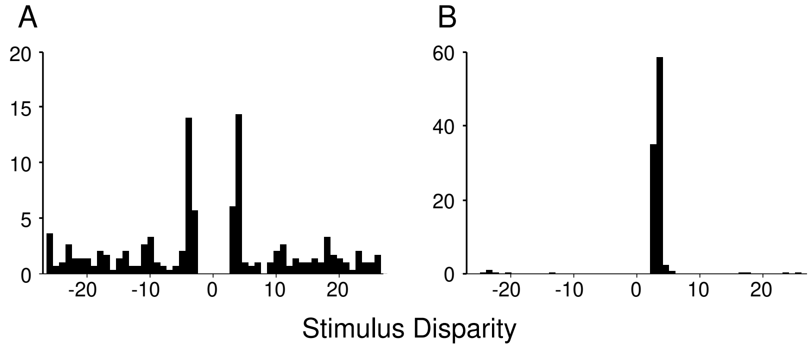

Figure 14: Simulations like that in Fig. 13 were run 300 times with statistically independent white noise stimuli. Phase-shifted energy neurons were designed to have a preferred disparity of 4 pixels. (A) After pooling over orientation and space, the histogram shows two distinct concentrations of response peaks. As explained in the text, this is the result of a receptive field phase-shift of π, which leads to significant aliasing, where responses appear at -4 instead of 4, as the preferred wavelength of the neurons is 8 pixels. Only 23% of the peaks fall within one pixel of 4, the correct disparity. (B) After pooling over scale, orientation, and space, 97% of the peaks lie within one pixel of the correct disparity. The problems of false peaks and frequency uncertainty are largely eliminated.

of binocular receptive fields. Position- and phase-shifts have different quantitative properties, and it is argued that both likely contribute to the disparity selectivity of cells in V1. The relative contribution of position- and phase-shift can be inferred by measuring disparity-tuning curves using drifting sinusoidal grating stimuli with several different spatial frequencies.

Position- and phase-shifted binocular energy neurons are not ideal disparity detectors; they do not respond to a unique, narrow range of disparities. Instead, they respond quasi-periodically as a function of disparity. False response peaks occur with a periodicity that depends on the spatial frequency selectivity of the neurons. Only when disparities are kept sufficiently small, as in the implementation reported by Qian (1994), can these false peaks in response be ignored. Our analysis shows that these false peaks should be evident only when one uses richly textured stimuli, such as random noise, and only for brief stimulus presentations.

In order to construct disparity detectors from position-shifted and/or phase-shifted energy neurons, we propose that energy-neuron responses are pooled linearly across several scales and orientations, and in local spatial neighborhoods. Pooled responses will not exhibit false peaks, but they will exhibit much broader orientation and spatial frequency specificity.

Related Computational Frameworks

Although the computational framework developed here is based on an energy mechanism, many of our results do not depend critically on this particular form of binocular interaction. The main predictions (see *Binocular Dependence on Spatial Frequency*) of the energy model, with position-shifts and/or phase-shifts, remain valid for several alternative computational frameworks. The significance of false response peaks and frequency uncertainty (see *Encoding Binocular Disparity*) also remain.

For example, one might replace the squaring nonlinearity in the energy model by fullwave-rectification (Pollen & Ronner, 1983), or a higher-order nonlinearity with an exponent greater than 2 (Albrecht & Hamilton, 1984; Sclar *et al.*, 1990; Albrecht & Geisler, 1991). This would affect the quantitative nature of the binocular energy response. However, for drifting sinusoidal grating stimuli, the response magnitude would remain a periodic function of stimulus disparity with peaks at the same disparities as the energy model. Moreover, the disparity response curves would still be phase-shifted so that one could determine the relative contributions of position-shift and phase-shift. The false peaks, although sharper or broader with different exponents, would still exist.

Interocular cross-correlation provides another plausible computational framework for modeling binocular interaction (Mallot *et al.*, 1995). It has also been used to model binaural properties of cells in the inferior colliculus of the cat for encoding interaural time differences (Yin & Kuwada, 1984).

One way to build a cross-correlation model is to use quadrature pairs of monocular receptive fields, with position and/or phase shifts between the left- and right-eye receptive fields. Binocular neurons would then compute a sum (over space, orientation, and multiple scales) of the product of the left and right monocular responses. Interestingly, one can show that this model is almost exactly the same as the energy model. Using the complex notation introduced in *Spatial Arrays of Identical Energy Neurons*, the energy response at x_0 (with a position-shift s and spatial pooling with a weighting function $W(x)$) is given by:

$$\begin{aligned} E(x_0;\, s) &= \int W(x-x_0)\, |L(x) \;+\; R(x-s)|^2 \; dx \\ &= \int W(x-x_0)\, (|L(x)|^2 \;+\; |R(x-s)|^2 \;+\; \mathrm{Re}[L(x) R^*(x-s)]) \; dx \;, \end{aligned} \tag{20}$$

The cross-correlation of $L(x)$ and $R(x)$ (computed as the sum of cross-correlations of the real and imaginary parts) is given by:

$$\begin{aligned} C(x_0;\, s) &= \int W(x-x_0)\, \mathrm{Re}[L(x)]\, \mathrm{Re}[R(x-s)] \; dx + \int W(x-x_0)\, \mathrm{Im}[L(x)]\, \mathrm{Im}[R(x-s)] \; dx \;. \\ &= \int W(x-x_0)\, \mathrm{Re}[L(x) R^*(x-s)] \; dx \;. \end{aligned} \tag{21}$$

The similarity of the two models is evident by comparing equations (20) and (21). The energy response is equal to the sum of the cross-correlation and the two monocular energies. The empirical methods for determining the relative contributions of position-shifts and phase-shifts would therefore remain the same. The pooling (over space, orientation, and scale) in the cross-correlation framework reduces the prevalence of false peaks, as it does for the binocular energy model.

There is, however, an important difference between the energy and cross-correlation models. The energy model predicts that complex cell responses modulate about a baseline equal to the sum of the monocular energies. The cross-correlation model does not predict a stimulus-dependent baseline.

An alternative cross-correlation model might use only even-symmetric receptive fields, for example, instead of quadrature pairs. This would require more extensive spatial pooling; otherwise, responses would depend significantly on the position of the stimulus within the receptive field.

A problem with both of these cross-correlation models is that they are inconsistent with the behavior of simple cells. The initial binocular interaction is multiplicative in the cross-correlation framework. By contrast, the initial binocular interaction is additive in the energy model, consistent with the predominantly linear behavior of simple cells (Ohzawa & Freeman, 1986a).

Behavioral Relevance of Position Shifts

Humans fuse and extract depth information when disparities are larger than half a wavelength of the frequencies present (Blake & Wilson, 1991). With difference-of-Gaussian (DOG) stimuli, Schor *et al.* (1984) found that for high frequency DOGs (with central frequencies higher than 2.5 cpd, up to 10 cpd), the upper fusion limit remains approximately constant at about $10'$ (Schor *et al.*, 1984). For frequencies above 3 cpd, a $10'$ disparity is greater than half a wavelength. With similar stimuli, Schor *et al.* (1984) found that upper depth limits can be as much as 5 times greater than upper fusion limits. For 10 cpd DOGs, they found upper depth limits close to $50'$ which corresponds to more than 8 cycles. Smallman and MacLeod (1994) reached a similar conclusion using stimuli at contrast threshold. A purely phase-based model, at least for high spatial frequencies, will not account for this performance. Position-shifts must also be present.

Unfortunately, quantitative neurophysiological data on the dependence of disparity tuning on scale has not been published. A related issue is how position- and phase-shifts contribute to the disparity selectivity of neurons at different scales.

Behavioral Relevance of False Peaks

We have argued that complex cells, modelled as binocular energy neurons, are not disparity detectors, in part owing to the existence of false peaks. At present there is little published data that shows false peaks in the responses of simple and complex cells (but they are evident in the data of Wagner & Frost, 1994). Many experiments use simple stimuli that are either sparse (e.g., a single bar) so that false matches do not occur, or periodic so that false matches can be attributed to ambiguity in the stimulus. Others report only the average responses as stimuli are swept through a cell's receptive field. To reveal false peaks one must test a wide range of disparities with rich stimuli like random noise, without averaging over long stimulus presentations. It is also useful to know a cell's preferred spatial frequency, to predict the approximate disparities at which false peaks are likely to occur.

False peaks are problematic mainly when disparities are large. Moreover large disparities occur regularly under normal viewing. For example, consider an observer with eyes 6 cm apart, fixated on a target 1 m away. A point 90 cm away from the observer has a crossed disparity of about $23'$, while a point 110 cm away has an uncrossed disparity of about $18'$. Points at 50 cm and 2 meters have disparities larger than 2°. A disparity of $15'$ is more than half a wavelength for all frequencies above 2 cpd. Whether or not we fuse or perceive depth from such large disparities, they can certainly be expected to produce false peaks in binocular energy neurons.

Multiple Scales and Pooling

The model developed in this article has two stages, with binocular linear and energy neurons in the first stage, and a second stage that pools the energy responses over space, scale, and orientation. These stages complement one another. The band-pass nature of the first stage significantly reduces frequency uncertainty for the phase-shift neurons, and allows for fine spatial and disparity resolution. The subsequent pooling then reduces the adverse effects of the false peaks that are inevitable with narrow-band signals.

Psychophysical data also supports the scale-specificity of binocular interaction. The frequency content of narrow-band stimuli in the two eyes must overlap for stereopsis to occur (Julesz, 1971; Mayhew & Frisby, 1976). Narrow-band noise, if two or more octaves away from a band-pass filtered random-dot stereogram, has little affect on stereopsis (Julesz & Miller, 1975). Also, upper depth and fusion limits change with the scale of the stimulus (Schor et al., 1984; Smallman & MacLeod, 1994).

To date there are two main theories of multiscale interaction. The first, popularized by Marr and Poggio (1979), is commonly referred to as a coarse-to-fine control structure, in which disparity estimates are first computed at coarse scales (low spatial frequencies). Once obtained, they are used as initial guesses for finer-scale matching. Marr and Poggio (1979) suggested that coarse-scale disparity estimates are used to drive vergence eye movements, thereby shifting the left and right views into closer alignment and allowing for a fine-scale match. Others have suggested that the "shifting" might be done neurally. Regardless of how the shifting is done, recent psychophysical data is inconsistent with this strict coarse-to-fine sequential model (McKee & Mitchison, 1988; Mowforth *et al.*, 1981; Mallot *et al.*, 1993; Smallman, 1995).

A second form of multiscale interaction, advocated in this article, is a form of coincidence model. When multiple scales and/or orientations produce strong responses at similar disparities then they support one another. When peaks at several scales coincide then they sum to produce a stronger peak. The false peaks at one scale will cancel with response minima at the other scales.

The majority of V1 cells, including binocular cells, are selective for orientation and spatial frequency, consistent with the first stage of our model. Complex cells have larger receptive fields than simple cells on average, indicating that complex cells perform some amount of local spatial pooling. But simple and complex cells appear to have similar bandwidths and frequency selectivities (DeValois *et al.*, 1982; Movshon *et al.*, 1978b; but see Hammond & Fothergill, 1994), and therefore there is little evidence for pooling over scale or orientation in V1. Indeed, we know of no physiological evidence for pooling over scale and orientation in the visual system to build disparity detectors.

In the barn owl auditory system, by contrast, one transformation from the central to the external nucleus of the inferior colliculus is the convergence of frequency channels (Kundsen & Konishi, 1978; Knudsen, 1984). There is a well-defined map of interaural time differences in this nucleus and it is thought that pooling over frequency helps to reduce ambiguities in this representation of interaural time differences (Wagner *et al.*, 1987). It is possible that there is no analogous map of disparity in the visual system. Rather, since the hypothesized pooling is a simple linear summation, it could be accomplished concurrently with later stages of processing. For example, an oculomotor neuron involved in vergence eye movements might pool over scale indirectly from a population of binocular energy neurons. The pooled signals would not be evident in responses in sensory neurons.

Temporal Aspects and Response Normalization

The model, as presented in this article, ignores several significant aspects of neural responses, namely, temporal properties and response normalization. These issues remain topics for further exploration, as do issues concerning fusion, binocular rivalry, and occlusion.

For example, the monocular linear neurons should have spatiotemporal linear weighting functions, possible selective for direction and/or temporal frequency. Also, the binocular energy neurons should pool over time as well as orientation, scale, and space. Pooling the binocular energies over time would help to attenuate false peaks like the pooling over space.

Another problem with the current model is the fact that V1 cell responses saturate at high contrasts. To explain response saturation and other violations of the linear/energy models, we and others have recently proposed a new model of V1 cell responses called the *normalization* model (Robson *et al.*, 1991; Albrecht & Geisler, 1991; Heeger, 1991, 1992a, 1993; Carandini & Heeger, 1994). We are extending the binocular linear and energy models to include response normalization. Our preliminary simulation results (Fleet *et al.*, 1995) indicate that appropriate normalization can account for a large body of data including the observed invariance with respect to interocular contrast differences (Freeman & Ohzawa, 1990; Ohzawa & Freeman, 1993). The response normalization also helps to attenuate the false peaks.

References

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

Albrecht, D.G. & Hamilton, D.B. (1982) Striate cortex of monkey and cat: Contrast response function. *J. Neurophysiology*, 48, 217–237

Albrecht, D.G. & Geisler, W.S. (1991) Motion sensitivity and the contrast-response function of simple cells in the visual cortex. *Visual Neuroscience*, 7, 531-546

Barlow, H.B., Blakemore, C., & Pettigrew, J.D. (1967) The neural mechanism of binocular depth discrimination. *J. Physiology*, 193, 327-42.

Blake, R. & Wilson, H.R. (1991) Neural models of stereoscopic vision. *Trends in Neurosciences*, 14(10), 445-52.

Broman, H. (1981) The instantaneous frequency of a Gaussian signal: The one-dimensional density function. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 29, 108–111

Campbell, F.W., Cleland, B.G., Cooper, G.F. & Enroth-Cugell, C. (1968) The angular selectivity of visual cortical cells to moving gratings. *J. Physiology*, 198, 237–250

Campbell, F.W., Cooper, G.F. & Enroth-Cugell, C. (1969) The spatial selectivity of visual cells of the cat. *J. Physiology*, 203, 223–235

Carandini M. & Heeger, D.J. (1994) Summation and Division by Neurons in Primate Visual Cortex. *Science*, 264, 1333–1336

Clarke, P.G.H., Donaldson, I.M.L. & Whitteridge, D. (1976) Binocular visual mechanisms in cortical areas I and II of the sheep. *J. Physiology*, 256, 509–526.

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.

DeAngelis, G.C., Ohzawa, I., & Freeman, R. (1995) Neuronal mechanisms underlying stereopsis: how do simple cells in the visual cortex encode binocular disparity? *Perception*, 24, 3–32.

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

Emerson, R.C., Bergen, J.R. & Adelson, E.H. (1992) Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research*, 32, 203–218

Ferster D. (1981) A comparison of binocular depth mechanisms in areas 17 and 18 of the cat visual cortex. *J. Physiology*, 311, 623-55.

Field, D.J. & Tolhurst, D.J. (1986) The structure and symmetry of simple-cell receptive field profiles in the cat's visual cortex. *Proceedings of the Royal Society of London B*, 228 , 379–400

Fischer, B. & Kruger, J. (1979) Disparity tuning and binocularity of single neurons in cat visual cortex. *Experimental Brain Research* 35, 1–8.

Fleet, D.J. (1994) Disparity from local weighted phase-correlation. *Proc. IEEE International Conference on Systems, Man, & Cybernetics*, San Antonio, October, 48–56

Fleet, D.J., Heeger, D.J. & Wagner, H. (1995) Computational model of binocular disparity. *Investigative Ophthalmology & Visual Science Supplement*, 36, 365

Fleet, D.J. & Jepson, A.D. (1993) Stability of phase information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15, 1253–1268

Fleet, D.J., Jepson, A.D. & Jenkin, M. (1991) Phase-based disparity measurement. *CVGIP: Image Understanding*, 53, 198-210

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

Foster, K.H., Gaska, J.P., Marcelja, S. & Pollen, D.A. (1983) Phase relationships between adjacent simple cells in the feline visual cortex. *J. Physiology*, 345, 22

Hamilton, D.B., Albrecht, D.G. & Geisler, W.S. (1989) Visual cotical receptive field in monkey and cat: Spatial and temporal phase transfer function. *Vision Research*, 29, 1285–1308

Hammond, P. (1991) Binocular phase specificity of striate cortical neurons. *Experimental Brain Research*, 87, 615–23.

Hammond, P. & Fothergill, L.K. (1994) Cat striate cortex: monocular and binocular comparisons of spatial-frequency selectivity. *Anais da Academica Brasileira de Ciencias*, 66, 95-113

Heeger, D.J. (1991) Nonlinear model of neural responses in cat visual cortex. In Landy, M. & Movshon, J.A. (Eds), *Computational Models of Visual Processing* (119–133), MIT Press, Cambridge, MA

Heeger, D.J. (1992a) Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9, 181–197.

Heeger, D.J. (1992b) Half-squaring in responses of cat striate cells. *Visual Neuroscience*, 9, 427–443.

Heeger, D.J. (1993) Modeling simple cell direction selectivity with normalized, half-squared, linear operators. *J. Neurophysiology*, 70, 1885–1898

Heggelund, P. (1986) Quantitative studies of the discharge fields of single cells in cat striate cortex. *J. Physiology*, 373, 277–292

Hubel, D.H. & Wiesel, T.N. (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiology*, 160, 106-154.

Hubel, D.H. & Wiesel, T.N. (1970) Stereoscopic vision in macaque monkey. *Nature*, 225, 41-42.

Jones, J.P. & Palmer, L.A. (1987) The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J. Neurophysiology*, 58, 1187–1211

Julesz, B. (1971) *Foundations of Cyclopean Perception*, University of Chicago Press, Chicago

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

Knudsen, E. (1984) Synthesis of a neural map of auditory space in the owl. In Edelman, G., Gall, W. and Cowan, W. (Eds), *Dynamic Aspects of Neocortical Function* (375-396), John Wiley and Sons, New York.

Knudsen, E. & Konishi, M. (1978) Space and frequency are represented separately in the auditory midbrain of the owl. *J. Neurophysiolgy*, 41, 870-884

LeVay, S. & Voigt, T. (1988) Ocular dominance and disparity coding in cat visual cortex. *Visual Neuroscience*, 1, 395-414.

Liu, A., Gaska, J.P., Jacobson, L.D. & Pollen, D.A. (1992) Interneuronal interaction between members of quadrature phase and anti-phase pairs in the cat's visual cortex. *Vision Research*, 32, 1193–1198

Mallot, H.A., Arndt, P.A. & Bülthoff, H.H. (1995) Matching versus correlation in human stereopsis. *Biological Cybernetics*, 72, 279-293

Mallot, H.A., Dartsch, S. & Arndt, P.A. (1993) Human stereo vision does not always proceed from coarse to fine. *Perception*, 22, 112

Marr, D. & Poggio, T. (1979) A computational theory of human stereo vision. *Proceedings of the Royal Society of London B* 207, 187-217

Maske, R., Yamane, S. & Bishop, P.O. (1984) Binocular simple cells for local stereopsis: comparison of receptive field organizations for the two eyes. *Vision Research*, 24, 1921–1929.

Mayhew, J.E.W. & Frisby, J.P. (1976) Rivalrous texture stereograms. *Nature*, 264, 53-56

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

Movshon, J.A., Thompson, I.D. & Tolhurst, D.J. (1978a) Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J. Physiology*, 283, 53–77

Movshon, J.A., Thompson, I.D. & Tolhurst, D.J. (1978b) Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. *J. Physiology*, 283, 101–120

Mowforth, P., Mayhew, J.E.W. & Frisby, J.P. (1981) Vergence eye movements made in response to spatial-frequency-filtered random-dot stereograms. *Perception*, 10, 299-304

Nikara, T., Bishop, P.O. & Pettigrew, J.D. (1968) Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex. *Experimental Brain Research*, 6, 353-372

Nomura, M., Matsumoto, G. & Fujiwara, S. (1990) A binocular model for the simple cell. *Biological Cybernetics* 63, 237-242

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

Ohzawa, I. & Freeman, R.D. (1986a) The binocular organization of simple cells in the cat's visual cortex. *J. Neurophysiology*, 56, 221-42.

Ohzawa, I. & Freeman, R.D. (1986b) The binocular organization of complex cells in the cat's visual cortex. *J. Neurophysiology*, 56, 243-59.

Ohzawa, I. & Freeman, R.D. (1994) Monocular and binocular mechanisms of contrast gain control. In Lawton, T.A. (Ed), *Computational Vision Based on Neurobiology, SPIE Proceedings*, Vol. 2054

Palmer, L.A. & Davis, T.L. (1981) Receptive-field structure in cat striate cortex. *J. Neurophysiology*, 46, 260–276

Papoulis, A. (1965) *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York

Pettigrew, J.D. (1972) The neurophysiology of binocular vision. *Scientific American* August, 84-95.

Pettigrew, J.D. (1979) Binocular visual processing in the owl's telencephalon. *Proceedings of the Royal Society of London B*, 204, 435-454.

Pettigrew, J.D. & Konishi, M. (1976) Neurons selective for orientation and binocular disparity in the visual Wulst of the barn owl (Tyto alba). *Science*, 193, 675-678.

Pettigrew, J.D., Nikara, T. & Bishop, P.O. (1968) Binocular interaction on single units in cat striate cortex: simultaneous stimulation by single moving slit with receptive fields in correspondence. *Experimental Brain Research*, 6, 391-410.

Pollen, D.A. & Ronner, S. (1981) Phase relationships between adjacent simple cells in the visual cortex. *Science*, 212, 1409–1411

Pollen, D.A. & Ronner, S. (1983) Visual cortical neurons as localized spatial frequency filters. *IEEE Transactions on Systems, Man, and Cybernetics*, 13, 907–916

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

Poggio, G.F., Motter, B.C., Squatrito, S. & Trotter, Y. (1985) Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms. *Vision Research*, 25, 397-406.

Poggio, G.F. & Poggio, T. (1984) The analysis of stereopsis. *Ann. Rev. Neurosci.*, 7, 379-412.

Poggio, G.F. & Talbot, W.H. (1981) Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey. *J. Physiology*, 315, 469-492.

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

Robson, J.G., DeAngelis, G.C., Ohzawa, I. & Freeman, R.D. (1991) Cross-orientation inhibition in cat cortical cells originates from within the receptive field. *Investigative Ophthalmology & Visual Science Supplement*, 32, 429

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

Sclar G., Maunsell, J.H.R., & Lennie, P. (1990) Coding of image contrast in central visual pathways of the macaque monkey. *Vision Research*, 30, 1–10

Smallman, H.S. (1995) Fine-to-coarse scale disambiguation in stereopsis. *Vision Research*, 34, 2971-2982

Smallman, H.S. & MacLeod, D.I.A. (1994) Size-disparity correlation in stereopsis at contrast threshold. *J. Optical Society of America A*, 11, 2169-2183

Steinbach, M.J. & Money, K.E. (1973) Eye movements of the owl. *Vision Research*, 13: 889-891.

Wagner, H. & Frost B.J (1993) Disparity-sensitive cells in the owl have a characteristic disparity. *Nature*, 364, 796-798.

Wagner, H. & Frost, B.J. (1994) Binocular responses of neurons in the barn owl's visual Wulst. *J. Comparative Physiology*, A 174, 661-670.

Wagner, H., Takahashi, T. & Konishi, M. (1987) Representation of interaural time difference in the central nucleus of the barn owl's inferior colliculus. *J. Neuroscience*, 7, 3105-3116

Yin, T.C.T. & Kuwada, S. (1984) Neuronal mechanisms of binarual interaction. In Edelman, G., Gall, W. and Cowan, W. (Eds) *Dynamic Aspects of Neocortical Function* (385-430), John Wiley and Sons, New York.

A Density Function for Disparity from Phase-Shifts

We derive a probability density function for disparity. We assume the phase-shift is given, and that instantaneous frequency is approximated by the frequency tuning of the underlying neuron. The key relation is $d = \Delta\phi/k$, for disparity d, instantaneous frequency k, and the phase-shift $\Delta\phi$.

The probability density function for instantaneous frequency is available in analytic form if we assume the stimulus is mean-zero white Gaussian noise. We assume for convenience that the power spectrum of the linear neuron's weighting function, denoted by $P(\omega)$, integrates to 1, with spectral mean μ and variance σ^2 given by

$$\mu = \int |\omega| P(\omega) d\omega \quad , \qquad \sigma^2 = \int \omega^2 P(\omega) d\omega \; - \; \mu^2 \; .$$

Under these conditions the instantaneous frequency is a random variable, and its probability density function can be shown to be (Broman 1981):

$$f_k(k) \;=\; \frac{\sigma^2}{2((k-\mu)^2 + \sigma^2)^{3/2}} \tag{22}$$

Figure 10(A) shows the probability density function for the instantaneous frequency of the output of a Gabor filter with a bandwidth of 1.0 octave. This distribution closely resembles the power spectrum of the Gabor weighting function, but with somewhat longer tails. This behavior generalizes to Gabor filters of other frequencies and bandwidths of interest.

Given the density function for instantaneous frequency, one can use the relation $d = \Delta\phi/k$ to derive the following density function for disparity:

$$f_d(d) \;=\; \frac{\Delta\phi}{d^2} \, f_k\left(\frac{\Delta\phi}{d}\right) \; . \tag{23}$$

The behavior of the mean and standard deviation of this distribution, for several different phase shifts is shown in Fig. 10(B).