A Physiological Model for Motion-stereo Integration and a Unified Explanation of the Pulfrich-like Phenomena

Ning Qian[1] and *Richard A. Andersen*[2]

[1] Center for Neurobiology and Behavior
Columbia University
722 W. 168th Street
New York, NY 10032

[2] Division of Biology (216-76)
California Institute of Technology
Pasadena, CA 91125

Abbreviated title: Motion-stereo Integration and the Pulfrich Effects
Key words: Motion, Stereo, Pulfrich effect, Binocular disparity, Temporal delay, Temporal stretching

Send all correspondence to:

Dr. Ning Qian
Center for Neurobiology and Behavior
Columbia University
722 W. 168th Street
New York, NY 10032
212-960-2213 (phone)
212-960-2561 (fax)
nq6@columbia.edu (email)

Abstract

Many psychophysical and physiological experiments indicate that visual motion analysis and stereoscopic depth perception are processed together in the brain. However, little computational effort has been devoted to combining these two visual modalities into a common framework based on physiological mechanisms. We present such an integrated model in this paper. We have previously developed a physiologically realistic model for binocular disparity computation (Qian, 1994). Here we demonstrate that under some general and physiological assumptions, our stereo vision model can be combined naturally with motion energy models to achieve motion-stereo integration. The integrated model may be used to explain a wide range of experimental observations regarding motion-stereo interaction. As an example, we show that the model can provide a unified account of the classical Pulfrich effect (Morgan and Thompson, 1975) and the generalized Pulfrich phenomena to dynamic noise patterns (Tyler, 1974; Falk, 1980) and stroboscopic stimuli (Burr and Ross, 1979).

Introduction

Visual motion analysis and stereoscopic depth perception are among the most important and best studied of our visual functions. There is increasing evidence indicating that these two visual functions are closely related and are probably processed together in the brain. In primates, binocular convergence (and hence disparity tuning) and directional selectivity first appear in area V1 (Hubel and Wiesel, 1968; Poggio and Fischer, 1977). V1 cells project to area MT, where almost all neurons are directionally selective (Albright, 1984). Most MT cells are also tuned to binocular disparity (Maunsell and Van Essen, 1983; Bradley, Qian and Andersen, 1995). In fact, many individual V1 and MT neurons exhibit both motion and disparity tuning. These physiological properties are clearly reflected at the behavioral level: psychophysical experiments indicate that strong interaction exists between motion and stereoscopic depth perception. For example, the motion aftereffect is found to be contingent upon binocular disparity (Regan and Beverley, 1973; Anstis and Hassis, 1974). Disparity specific motion adaptation has been shown to significantly reduce motion direction ambiguity in rotating stimuli (Nawrot and Blake, 1989). Binocular disparity has also been found to facilitate transparent motion perception (Adelson and Movshon, 1982; Qian, Andersen and Adelson, 1994a).

In view of the close relationship between motion and stereo vision as revealed by both physiological and psychophysical experiments, it is surprising that little computational effort has been devoted to building unified models for these two visual modalities. Many computational models for biological motion processing (Reichardt, 1961; Hildreth, 1984; Watson and Ahumada, 1985; van Santen and Sperling, 1985; Adelson and Bergen, 1985; Heeger, 1987; Grzywacz and Yuille, 1990) and stereo vision (Marr and Poggio, 1976; Marr and Poggio, 1979; Prazdny, 1985; Pollard, Mayhew and Frisby, 1985; Sanger, 1988; Qian and Sejnowski, 1989; Yeshurun and Schwartz, 1989) have been proposed, but few dealt with these two visual functions at the same time. Although it is clear that at an abstract level both motion and stereo vision can be formulated as solving a correspondence problem (see Marr (1982), for example), this observation says little about how physiologically the two visual functions may be processed together by a population of cells with both motion and disparity tuning, and how the two modalities may interact with each other. In fact, the very notion of an explicit correspondence or matching is non-physiological (see the Discussion).

In this paper we present an integrated model of motion and stereopsis based on the receptive field properties of real visual cells. We have recently developed a physiologically realistic model for binocular disparity computation and for the first time demonstrated that broadly disparity-tuned units, modeled accurately after real binocular cells in the visual cortex, can effectively solve random dot stereograms (Qian, 1994). Here we demonstrate that under physiological assumptions our stereo vision model can be combined naturally with motion energy models (Watson and Ahumada, 1985; Adelson and Bergen, 1985; van Santen and Sperling, 1985) to achieve motion-stereo integration. As an application of the integrated model, we will show that the model can provide a unified explanation of the classical Pulfrich effect (Morgan and Thompson, 1975) and its more recent generalizations to dynamic noise patterns (Tyler, 1974; Falk, 1980) and stroboscopic stimuli (Burr and Ross, 1979). The explanation works equally well whether one assumes a temporal delay (Mansfield and Daugman, 1978; Lennie, 1981; Cynader, Gardner and Douglas, 1978; Carney, Paradiso and Freeman, 1989) or a temporal stretching (Kaufman and Palmer, 1990) in the neuronal responses accompanying a luminance reduction.

Stereo Vision

One possible strategy for constructing a unified model of motion and stereo vision is to examine existing models in these two categories and see if they can be combined naturally. There are physiologically plausible models for motion detection, namely the motion energy models[1] (Adelson and Bergen, 1985; Watson and Ahumada, 1985; van Santen and Sperling, 1985; Emerson et al., 1992). Until recently, most models of stereopsis, on the other hand, cannot be said to be truly biological. Some (Marr and Poggio, 1976; Marr and Poggio, 1979; Prazdny, 1985; Pollard et al., 1985; Qian and Sejnowski, 1989) require very sharply disparity-tuned units and use explicit matching of fine image features (see the Discussion). Others (Sanger, 1988; Yeshurun and Schwartz, 1989) contain certain mathematical operations (such as the explicit extraction of complex phases of stimuli) that are unlikely to be physiological.

We have recently proposed a physiologically realistic model for stereo vision (Qian, 1994). We briefly review the model in this section. Our model is based on the quantitative physiological studies of Freeman and coworkers (Freeman and Ohzawa, 1990; Ohzawa, DeAngelis and Freeman, 1990; DeAngelis, Ohzawa and Freeman, 1991). These investigators found that the left and right spatial receptive field profiles of a binocular simple cell in cat's primary visual cortex can be described by two Gabor functions with the same Gaussian envelopes but different phase parameters in the sinusoidal modulations. For horizontal disparity computation, only the horizontal dimension of cells' receptive fields is relevant. The left and right receptive field profiles of a simple cell centered at $x = 0$ are then given by:

$$f_l(x) = \exp(-\frac{x^2}{2\sigma^2})\cos(\omega_x^o x + \phi_l) \quad (1)$$

$$f_r(x) = \exp(-\frac{x^2}{2\sigma^2})\cos(\omega_x^o x + \phi_r) \quad (2)$$

where σ and ω_x^o are the Gaussian width and the preferred (angular) spatial frequency of the receptive fields; ϕ_l and ϕ_r are the left and right phase parameters.

Freeman and coworkers (Freeman and Ohzawa, 1990; Ohzawa et al., 1990) found that the response of a simple cell can be determined by first filtering, for each eye, the retinal image by the corresponding receptive field profile, and then adding the two contributions from the two eyes. They further showed that the response of a complex cell can be modeled by summing the squared outputs of a quadrature pair of such simple cells. Through mathematical analysis we found that under the assumption that stimulus disparity D is significantly smaller than the Gaussian width σ of the receptive fields, the response of a model complex cell to the disparity is given by (Qian, 1994):

$$r_c \approx c^2\ |\tilde{I}(\omega_x^o)|^2\ \cos^2(\frac{\Delta\phi}{2} - \frac{\omega_x^o D}{2}), \quad (3)$$

[1]Motion energy models were originally proposed based on human visual psychophysics (Adelson and Bergen, 1985; Watson and Ahumada, 1985; van Santen and Sperling, 1985). They were later found to describe the behavior of directionally selective cells in the primary visual cortex quite well (Emerson, Bergen and Adelson, 1992; Reid, Soodak and Shapley, 1987; Snowden, Treue, Erickson and Andersen, 1991).

where

$$\Delta\phi \equiv \phi_l - \phi_r \tag{4}$$

is the phase parameter difference between the left and right receptive fields, c is a constant, and $|\tilde{I}(\omega_x^o)|^2$ is the Fourier power of the stimulus patch (under the receptive field) at the preferred spatial frequency of the cell. According to equation 3, the cell's preferred disparity is determined by its receptive field parameters as: $D_{pref} \approx \Delta\phi/\omega_x^o$. Using this relationship we were able to compute disparity maps from random dot stereograms using a population of model complex cells without employing any non-physiological procedures such as explicit matching of fine stimulus features (Qian, 1994). Note that the periodic function in equation 3 is an approximation under small D; our simulations indicate that the side peaks of the cell's disparity tuning curves decay to zero as D increases. Also note that equation 3 was derived without assuming a specific functional form of the stimulus pattern. With explicit assumptions about the stimulus, accurate expressions of the complex cell responses for all D values may be derived (Zhu and Qian, 1996).

It can also be shown that our stereo algorithm can be extended to a more general class of receptive field profiles than the Gabor functions (Qian and Zhu, 1995). Specifically, we found that equation 3 can be derived under the general assumption that the frequency tuning of the receptive field profiles is much sharper than the frequency spectrum of the input stimulus, and that there is a phase difference $\Delta\phi$ between the left and right receptive field profiles.

Motion-stereo Integration

Since the quadrature pair construction of model binocular complex cells (Ohzawa et al., 1990; Qian, 1994) is rather similar to that used previously in motion energy models (Adelson and Bergen, 1985; Watson and Ahumada, 1985; van Santen and Sperling, 1985), our stereo algorithm can be combined naturally with motion energy models into a unified framework. We have previously demonstrated that such an integration can be achieved by using the following binocular three-dimensional (3D) spatiotemporal Gabor filters [2] (Qian, 1994):

$$f_l(x,y,t) = \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} - \frac{t^2}{2\sigma_t^2}\right)\cos(\omega_x^o x + \omega_y^o y + \omega_t^o t + \phi_l), \tag{5}$$

$$f_r(x,y,t) = \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} - \frac{t^2}{2\sigma_t^2}\right)\cos(\omega_x^o x + \omega_y^o y + \omega_t^o t + \phi_r). \tag{6}$$

where σ's and ω's determine the sizes and the preferred frequencies along the spatial and temporal dimensions of the receptive fields, and ϕ_l and ϕ_r are again the phase parameters. Note that without the dependence on the vertical spatial coordinate y and time t, these equations will be reduced to equations 1 and 2 for disparity computation discussed in the previous section. If, on the other hand, the phase parameters are omitted, the filters will become the standard Gabor functions with an orientation in the spatiotemporal space that have been used for motion computation (Adelson and Bergen, 1985; Heeger, 1987; Grzywacz and Yuille, 1990). One therefore expects that when these two elements are put together in

[2]More generally, the σ's in the Gaussian may be replaced by a 3×3 covariance matrix.

these equations as simple cell receptive field profiles, they can be used to construct model complex cells with simultaneous disparity and motion selectivity. Our previous analysis (Qian, 1994) and simulations (Qian, Andersen and Adelson, 1994b) confirm that this is indeed the case.

There is, however, one major problem with this formulation: while the spatial receptive fields of cortical simple cells can be modeled accurately by Gabor functions (Marcelja, 1980; Daugman, 1985; Jones and Palmer, 1987; Freeman and Ohzawa, 1990), the temporal responses of the cells are clearly not Gabor-like[3] (DeAngelis, Ohzawa and Freeman, 1993). The integrated model we developed using spatiotemporal Gabor filters is therefore not completely physiologically realistic. We now present a more general result demonstrating that our previous approach can be readily extended to encompass the realistic spatiotemporal receptive field properties found in the brain.

Let the left and the right receptive field profiles of a binocular simple cell be denoted by $f_l(x,\ y,\ t)$ and $f_r(x,\ y,\ t)$. Under the assumptions that both of these receptive fields are well tuned around the same spatial frequencies $(\omega_x^o,\ \omega_y^o)$, and that the main difference between the two receptive field profiles is a phase difference $\Delta\phi$, it can be shown that the complex cell response, constructed from a quadrature pair of such simple cells, to a moving stimulus of disparity D and image velocity $(V_x,\ V_y)$ is given by (see the Appendix):

$$r_c \approx c^2\ |\tilde{I}(\omega_x^o, \omega_y^o)|^2\ \cos^2(\frac{\Delta\phi}{2} - \frac{\omega_x^o D}{2})\ [\iint_0^\infty |\tilde{f}_l(\omega_x,\ \omega_y,\ \omega_t')| d\omega_x d\omega_y]^2, \tag{7}$$

where

$$\omega_t' = -\omega_x V_x - \omega_y V_y \tag{8}$$

is the familiar motion constraint (Watson and Ahumada, 1983), $|\tilde{f}_l(\omega_x,\ \omega_y,\ \omega_t)|$ is the Fourier amplitude of the left receptive field profile, and $|\tilde{I}(\omega_x^o, \omega_y^o)|^2$ is simply the Fourier power of the stimulus patch at the cell's preferred frequencies.

Equation 7 indicates that a single step of quadrature pair construction generates a model complex cell tuned to both motion and binocular disparity. The $\Delta\phi$ dependent cosine term determines the cell's disparity tuning just as in equation 3; it reaches maximum when D is equal to $\Delta\phi/\omega_x^o$. The last term determines the cell's motion sensitivity via spatiotemporal frequency selectivity just as in motion energy models[4] (Adelson and Bergen, 1985; Watson and Ahumada, 1985; Heeger, 1987; Grzywacz and Yuille, 1990). Note that the response of an individual complex cell confounds motion and stereo information. However, a population of cells with a wide range of parameters can form a distributed coding of both types of

[3]Real temporal response functions are typically skewed with an envelope having a longer decay time than rise time while the envelopes of Gabor functions are symmetrical Gaussian functions. Also, unlike Gabor functions, zero-crossing intervals of real temporal responses are not equally spaced.

[4]Here is an intuitive explanation of why the last term in equation 7 gives the cell motion selectivity. Since we assume that the receptive fields are well tuned to spatiotemporal frequencies $(\omega_x^o, \omega_y^o, \omega_t^o)$, the Fourier transform of the left receptive field, $\tilde{f}_l(\omega_x,\ \omega_y,\ \omega_t)$, has significant power only in a window centered around the point $(\omega_x^o, \omega_y^o, \omega_t^o)$ in the frequency space. The magnitude of the last term in equation 7 depends on whether the motion constraint plane defined by equation 8 goes through this window. As the image velocity $(V_x,\ V_y)$ changes, the constraint plane will be tilted in different orientations, thus changing the value of the last term in equation 7.

information simultaneously. For disparity computation, we can look at the responses of a family of cells with identical ω_x^o, ω_y^o and ω_t^o but different $\Delta\phi$ (Qian, 1994). Similarly, for velocity field computation we can use a family of cells with constant $\Delta\phi$, but different ω_x^o, ω_y^o and ω_t^o (Watson and Ahumada, 1985; Heeger, 1987; Grzywacz and Yuille, 1990). By holding $\Delta\phi$ at different values, one could estimate velocity fields at different depth planes.

We conclude that our rather general assumptions about a cell's frequency tuning and the phase relationship between the left and right receptive fields ensure that the cell is tuned to both disparity and motion. These assumptions are satisfied by the receptive field profiles of real cells in the visual cortex (Freeman and Ohzawa, 1990; Ohzawa et al., 1990; DeAngelis et al., 1993). Furthermore, our analysis shows how a population of these cells may be used to extract both motion and disparity information in the stimulus. We have previously applied a special version of the above general model to explain our psychophysical and physiological observations of disparity-specific motion suppression (Qian et al., 1994a; Qian et al., 1994b; Qian and Andersen, 1994; Bradley et al., 1995). We now show that the model can be used to account for a family of psychophysical observations related to the Pulfrich effect (Morgan and Thompson, 1975).

The Pulfrich Effects

The classical Pulfrich effect refers to the observation that a pendulum oscillating back and forth in the frontal parallel plane appears to move along an elliptical path in depth when a neutral density filter is placed in front of one of the two eyes (Morgan and Thompson, 1975) (see Fig. 1). It is known that by reducing the amount of light reaching the covered retina, the filter causes a temporal delay in the neuronal transmission from that retina to the cortex (Mansfield and Daugman, 1978; Lennie, 1981; Cynader et al., 1978; Carney et al., 1989). The standard explanation of this effect is that since the pendulum is moving, the temporal delay for the covered eye corresponds to a spatial displacement of the pendulum, which produces a disparity between the two eyes and therefore a shift in depth. This interpretation becomes problematic, however, when it is observed that the Pulfrich depth effect is present even with dynamic noise patterns (Tyler, 1974; Falk, 1980), since there is no coherent motion in these patterns to convert a temporal delay into a spatial disparity. It was further discovered that the effect is still present when a stroboscopic stimulus is used such that the two eyes never see an apparently moving target at the same time (Burr and Ross, 1979) and therefore no conventionally defined spatial disparity exists. It has been suggested that more than one mechanism may be responsible for these phenomena (Ross, 1974; Poggio and Poggio, 1984). Our mathematical analyses and computer simulations indicates that all of the above observations can be explained in a unified way by our integrated model.

Pulfrich's pendulum

We first consider the original Pulfrich effect on an oscillating pendulum. Unlike the standard explanation discussed above, we believe that the central issue is how a population of *neurons* with both motion and disparity selectivity would treat a temporal delay along one of the two ocular pathways as a binocular disparity. Consider the case where a neutral density filter is placed in front of the right eye and it introduces a temporal delay of Δt in the response of the right receptive field of binocular cells in area V1 (Carney et al., 1989; Gardner, Douglas and Cynader, 1985). For a pendulum with velocity (V_x, V_y) and with zero disparity, the complex cell response, constructed from a quadrature pair of simple cells well tuned to spatiotemporal frequencies $(\omega_x^o,\ \omega_y^o,\ \omega_t^o)$ and with a phase parameter difference $\Delta\phi$ between the left and right receptive fields, can be shown to be (see the Appendix):

$$r_c \approx c^2\ |\tilde{I}(\omega_x^o,\ \omega_y^o)|^2\ \cos^2(\frac{\Delta\phi}{2} - \frac{\omega_t^o \Delta t}{2})\ [\iint_0^\infty |\tilde{f}_l(\omega_x,\ \omega_y,\ \omega_t')| d\omega_x d\omega_y]^2. \tag{9}$$

where ω_t' is a function of the pendulum velocity (V_x, V_y) and is given by the motion constraint equation 8. As we discussed above, the $\Delta\phi$ dependent cosine term determines the disparity tuning of the cell. We conclude, by comparing equation 9 with equation 7, that for a complex cell with preferred horizontal spatial frequency ω_x^o and temporal frequency ω_t^o, the effect of an interocular time delay Δt is equivalent to a binocular disparity of [5]

$$d \approx \frac{\omega_t^o}{\omega_x^o} \Delta t. \tag{10}$$

In other words, the complex cell will respond to a interocular time delay as if there were a real binocular disparity in the input stimulus. For the family of cells with different $\Delta\phi$'s that code the disparity of a stimulus (Qian, 1994), they would not be able to tell whether their pattern of activity is caused by an actual binocular disparity or an interocular time delay. The ratio of the two preferred frequencies in equation 10 is approximately equal to the preferred horizontal velocity of the cell (Watson and Ahumada, 1983). Cells with different preferred velocity will therefore treat a given time delay as different equivalent disparities. It is reasonable to assume that the perception is determined by the equivalent disparities of the most responsive cells in the population. As the oscillating pendulum is going through different velocities, different groups of cells with appropriate preferred temporal to spatial frequency ratios will be maximally activated, generating different perceived depths according to equation 10. In particular, for the two opposite directions of motion of the pendulum, cells tuned to the opposite directions (and thus with opposite signs of ω_t^o) will be optimally activated, generating disparities of opposite signs. Finally, when the neutral density filter is used to cover the left eye instead of the right eye, the left ocular input to a binocular cell will be lagged behind the right input and this is equivalent to having a negative time delay Δt in equation 10. Consequently, the pendulum will appear to rotate in the opposite direction in depth. These results explain the observed behavior of Pulfrich's pendulum.

We have also performed some computer simulations for verifying our mathematical analyses. We ignore the unimportant vertical spatial dimension and consider only the horizontal spatial dimension and time dimension in the simulations. An example of our simulation is shown in Fig. 2, where an oscillating pendulum with trajectory

$$x = \frac{400}{\pi} \sin(\frac{\pi}{400} t) \tag{11}$$

[5] Note that equation 10 can be obtained very easily under the special case of using 3D spatiotemporal Gabor filters (equations 5 and 6) as receptive field profiles. Our derivation shown here is much more general.

is considered. The units of both space x and time t are pixels. The maximum velocity of the pendulum is therefore 1 space pixel per time pixel. The spatiotemporal representation of the pendulum in one full cycle is shown in Fig. 2a. The pendulum first swings to the right (positive x direction), it then reverses direction and moves to the left, and finally it moves to the right again. A periodic boundary condition is used along the time axis (i.e., the x–t plot of a full stimulus period wraps around in time so that the stimulus is equivalent to one that extends to infinite times) in the simulation to eliminate the "edge" effect. The left and right retinal images of the pendulum are identical. The neutral density filter in front of the right eye is assumed to introduce a time delay of 4 pixels in the temporal responses of the model cells' right receptive fields. The computed equivalent disparity d at each spatiotemporal location of the pendulum is shown in Fig. 2b. It can be seen from the figure that when the pendulum is moving to the right (left) the computed equivalent disparity is negative (positive), indicating that the pendulum appears closer to (further away from) the observer, in agreement with the perception. The projection of the 3D plot onto the d–x plane forms a closed path similar to the ellipse in depth in Fig. 1 (notice that d and x are plotted with different scales in Fig. 2).

The details of our simulations are as follows. Since our theoretical results (equations 7, 9 and 10) demonstrate that the exact forms of receptive field profiles are not important so long as they satisfy some general properties, we used spatiotemporal Gabor filters for receptive field profiles in our simulations for convenience. For each pendulum position, the equivalent disparity was computed with 24 model binocular complex cells with their receptive fields centered at that position and with their $\Delta\phi$ parameter evenly distributed in $[-\pi, \pi]$. The maximum response of the cell population was located through a parabolic interpolation and the interpolated $\Delta\phi$ parameter was divided by ω_x^o to obtain the equivalent disparity (Qian, 1994). The total preferred frequency of all cells, defined as $\sqrt{(\omega_x^o)^2 + (\omega_t^o)^2}$, was fixed at $\pi\sqrt{2}/16$ radian/pixel, and the preferred temporal to spatial frequency ratio was set to the instantaneous velocity of the pendulum. The Gaussian widths σ_x and σ_t of all cells' receptive fields were equal to 16 pixels. The simulation results were not very sensitive to the parameters of the model cells; the only essential requirement is that the preferred spatial frequency (ω_x^o) should be small enough such that the expected equivalent disparity falls in the range of $[-\pi/\omega_x^o, \pi/\omega_x^o]$ (Qian, 1994; Zhu and Qian, 1996). For example, we obtained nearly identical results when the total preferred frequency was scaled up and the receptive field size scaled down by a factor of 4. All simulations were performed on a Sun SPARCstation 10.

The generalized Pulfrich effect to arbitrary spatiotemporal patterns

The result in equation 10 can be generalized to an arbitrary spatiotemporal stimulus which may or may not contain any coherent motion. Again, assume that a neutral density filter introduces a temporal delay of Δt in the response of the right receptive field of binocular cells. The complex cell response, constructed from a quadrature pair of simple cells well tuned to spatiotemporal frequencies ($\omega_x^o,\ \omega_y^o,\ \omega_t^o$) and with a phase difference $\Delta\phi$ between the left and right receptive fields, to the stimulus is approximately (see the Appendix):

$$r_c \approx c^2\ |\tilde{I}(\omega_x^o,\ \omega_y^o,\ \omega_t^o)|^2\ \cos^2(\frac{\Delta\phi}{2} - \frac{\omega_t^o \Delta t}{2})\ [\iiint_0^\infty |\tilde{f}_l(\omega_x,\ \omega_y,\ \omega_t)| d\omega_x d\omega_y d\omega_t]^2. \tag{12}$$

This expression is identical to equation 9 except that here the integration in the last term is carried over both spatial and temporal frequencies and the motion constraint equation 8 is not required since we do not assume any coherent motion in the stimulus. The equivalent disparity for this cell, which is determined by the $\Delta\phi$ dependent cosine term in the above expression, is therefore also given by equation 10. Thus, for any stimulus that can significantly excite cells tuned to frequencies $(\omega_x^o, \omega_y^o, \omega_t^o)$, an interocular time delay is equivalent to a binocular disparity given by equation 10 from the cells' point of view.

The above result can explain the observation that the Pulfrich effect is still present when viewing flickering dynamic random noise patterns on a monitor screen instead of an oscillating pendulum (Tyler, 1974; Falk, 1980). There are two aspects in this phenomenon that need to be explained: when a time delay is introduced by a neutral density filter placed in front of the right eye, (1) the original flat noise pattern appears to have depths both in front of and behind the monitor screen, and (2) the front surface appears to move to the right and the back surface appear to move to the left even though the original noise pattern does not have any clear motion in either direction. The first aspect can be explained by the fact that the noise pattern has a broad spatial and temporal frequency spectrum. It can thus drive a wide range of cells, including those tuned to either positive or negative temporal frequencies. Consequently, the pattern appears to have depths both behind and in front of the screen according to equation 10. In addition, since the cells with positive and negative temporal frequency preferences, which are responsible for the perception of the back and front surfaces, are tuned to the left and right directions of motion respectively, the back surface should therefore appear to move to the left and the front surface to the right. This explains the second aspect of the phenomenon.

An example of our computer simulations with the dynamic noise patterns is shown in Fig. 3. The spatiotemporal representation of the noise pattern at a fixed y position is given in Fig. 3a. The two eyes see the same pattern and a time delay of 4 pixels is assumed for the right receptive fields of the model cells. Because there is no coherent motion trajectory in the dynamic noise pattern, we cannot use the same format as in Fig. 2b to display the simulation results. Instead, we consider a given spatiotemporal location and compute the equivalent disparities at this location using several different families of complex cells. Cells in the same family have identical spatiotemporal frequency tuning (and therefore preferred horizontal velocity) but with their phase parameter differences uniformly distributed in $[-\pi, \pi]$. Different cell families are tuned to different horizontal velocities. An equivalent disparity is computed from each cell family[6] and the results from 11 different families are shown in Fig. 3b. In this figure, the preferred horizontal velocity of each cell family is represented by an arrow, and the corresponding equivalent disparity reported by the family is indicated by the vertical position of the arrow. It is clear from the figure that cell families tuned to different preferred horizontal velocities report different equivalent disparities as predicted by

[6]The simulation procedure is the same as that for the pendulum except that a spatial pooling step is added when computing complex cell responses (Zhu and Qian, 1996). This pooling step does not make any difference for simple input stimuli such as the pendulum while it greatly improves the reliability of disparity tuning to stimuli such as the noise pattern, whose Fourier phase is not a smooth function of the frequencies (see the Appendix of Zhu and Qian (1996)). The inclusion of the pooling step in computing complex cell responses is well justified by the physiological observation that the receptive field sizes of complex cells are somewhat larger than that of simple cells at the same eccentricity (Zhu and Qian, 1996)

equation 10.

In our simulation of the oscillating pendulum considered in the previous subsection, we assumed that at a given instant, the perceived disparity is given by the cell family whose preferred velocity matches that of the pendulum. This is a reasonable assumption because the cells in this family are maximally activated. On the other hand, the dynamic noise pattern considered here has a very broad frequency spectrum and consequently, cells tuned to different spatiotemporal frequencies (velocities) are about equally activated. One therefore cannot easily determine the equivalent disparity reported by which cell family dominates the perception, and the different disparities reported by different cell families must be simultanously present in our perception. This is consistent with our informal observation that the Pulfrich effect with the dynamic noise stimulus is not as clear as that with a pendulum and that the noise appears to revolve in a volume rather than on a thin surface. However, a bias toward a particular disparity may be generated by the distribution of the numbers of cells in the cortex tuned to different velocities.

It is also important to note that even without the temporal delay, the original dynamic noise pattern has a broad spatial and temporal frequency spectrum and therefore should activate cells tuned to all directions of motion. The pattern, however, does not appear to move in any direction because there is a suppression stage in the motion pathway at which motion energies from different directions locally inhibit each other (Qian and Andersen, 1994; Snowden et al., 1991). The introduction of a time delay causes motion signals for the left and right directions to appear in different disparity channels (as defined by the $\Delta\phi$ parameter). Since the inhibition between opposite directions of motion is disparity specific (Qian et al., 1994a; Qian and Andersen, 1994; Bradley et al., 1995), the left and right motion signals at the front and back surfaces no longer cancel each other and net motion on each surface is therefore perceived.

The Pulfrich effect with stroboscopic stimuli

Our model can explain another interesting variation of the Pulfrich effect reported by Burr and Ross (1979) (see also Morgan (1975), and Ross and Hogben (1975)). In their experiments, a spot of light is shown stroboscopically on a sequence of horizontal locations at regular time intervals (τ). The two eyes see the same sequence of the light spot undergoing apparent motion except the left eye's version is delayed with respect to the right eye by a small amount (δt). Since the delay δt is smaller than the time interval τ, the two eyes never see any spot of light at the same time. There is therefore no spatial disparity, defined in the usual sense, present in the stimulus at any time. However, the Pulfrich depth is perceived as if the light spot were moving continuously instead of stroboscopically. It has been suggested that the missing intermediate positions of the light spot are first reconstructed by the brain and then the stereo mechanism works on the reconstructed version of the display (Poggio and Poggio, 1984; Burr, 1979). The observed effect can be explained naturally and almost trivially by our model without introduction of any additional assumptions. Our model does not assume an explicit spatial disparity in the stimulus at any given time but relies on responses of cells with spatiotemporal receptive fields. Since the temporal response functions of the primary visual cortical cells have a width of about 100–200 msec, much larger than the time delay δt (less than 2 msec) used in the experiments, there is a substantial overlap between the temporal responses of the left and right receptive fields and equation 12 remains valid for the stroboscopic stimuli.

It is also interesting to note that Burr and Ross (1979) reported that with their experimental paradigm, the Pulfrich depth effect is clearly observed only when the time interval (τ) of the apparent motion is smaller than 200 msec. This can be explained by the fact that the significant portion the temporal response profiles of V1 cells is typically less than 200 msec (Hamilton, Albrecht and Geisler, 1989; DeAngelis et al., 1993). When τ is larger than 200 msec these cells are no longer sensitive to the apparent motion of the stimulus although the observer as a whole may still see the motion using some higher level long range motion mechanisms. Consequently, similar to the case with the noise patterns discussed in the previous subsection, cells tuned to different velocities report different equivalent disparities and no particular disparity dominates the perception. The Pulfrich effect should thus be much weaker when τ is larger than 200 msec or the effect may not even be observable because in the paradigm used by Burr and Ross (1979) there is only a single dot present intermittently instead of many dots in the noise pattern.

We have performed computer simulations with the stroboscopic stimuli. An example is shown in Fig. 4. Fig. 4a is the spatiotemporal representation of the stroboscopic dot patterns presented to the left and right eyes. Each dot lasts for 1 time pixel, the time interval (τ) of the apparent motion is 50 pixels, and the time delay between the two eyes' views is 4 pixels. Note that here the interocular time delay is generated electronically in the patterns presented to the two eyes (Burr and Ross, 1979) instead of by a neutral density filter. As can be seen from the figure, at any instant of time only one of the two eyes sees a dot. The computed equivalent disparity is shown in Fig. 4b. The result is rather similar to the case of continuous motion in Fig. 2. The simulation procedure is same as that used in Fig. 2. Again, the results are not very sensitive to the cell parameters used. However, here one should use model cells with large enough spatiotemporal receptive fields so that they are sensitive to the apparent motion in the stimulus.

Additivity of time delay and real disparity

There is yet another aspect of the Pulfrich effect that can be explained by our model. It has been found that the perceived depth caused by temporal delay combines additively with actual disparity in the experimental paradigm of Burr and Ross (see also Julez and White (1969) for similar results with a different paradigm). It can be shown that when there is both a real disparity D and a temporal delay Δt present, the cosine term in equation 12 will become:

$$\cos^2\left(\frac{\Delta\phi}{2} - \frac{\omega_x^o D}{2} - \frac{\omega_t^o \Delta t}{2}\right), \tag{13}$$

and the equivalent disparity is thus given by:

$$d \approx D + \frac{\omega_t^o}{\omega_x^o}\Delta t. \tag{14}$$

Therefore, the effects of real disparity D and of the interocular time delay Δt enhance or cancel each other additively depending on their signs.

An intuitive explanation of the Pulfrich effects

The central idea in our above explanations of the various Pulfrich-like phenomena is the equivalence between an interocular time delay and a binocular disparity from the visual cortical cells' point of view. The details of our formal mathematical demonstration of this equivalence is given in the Appendix. Here we provide an intuitive explanation. Fig. 5 shows schematically the left and right receptive field profiles of three simple cells. The left receptive fields of all three cells are exactly the same while their right receptive fields differ. The right receptive field of the cell in Fig. 5a is identical to its left receptive field (notice the reference crosses are centered on the grey areas of both receptive fields). Therefore, a complex cell constructed from a quadrature pair of such simple cells should prefer zero disparity. In contrast, the right receptive field of the cell in Fig. 5b is phase shifted with respect to the left receptive field, and this generates a horizontal displacement of the right receptive field (notice the different relative position of the grey area with respect to the cross). The corresponding complex cell should therefore prefer a non-zero disparity. Finally, the cell in Fig. 5c is the same cell shown in Fig. 5b except that its right receptive field has now been delayed in time (i.e., shifted upwards) due to a neutral density filter placed in front of the right eye. The important thing to notice is that, due to the space-time slant, this time delay also creates an apparent horizontal shift of the excitatory and inhibitory regions of the right receptive field at a given time which cancels the effect of the phase shift in Fig. 5b (notice now the cross is again centered on the grey area). Since the disparity tuning of a cell is determined by the horizontal relationship between the left and right receptive fields, the corresponding complex cell in Fig. 5c should be tuned to zero disparity just like the cell in Fig. 5a. We therefore conclude that a complex cell originally tuned to a non-zero disparity may prefer zero disparity when an appropriate interocular time delay is introduced. When such a cell is activated it does not "know" whether (1) the stimulus has a non-zero disparity or (2) the stimulus has zero disparity *and* there is an interocular time delay.

To determine the how much horizontal shift is generated by a given temporal delay, we first draw auxiliary lines through the center of the excitatory and inhibitory subregions of a given receptive field profile (see Fig. 6a). The horizontal and vertical distances between two adjacent lines (indicated by the two thin short lines) are approximately equal to the preferred spatial period λ_x $(= 2\pi/\omega_x^o)$ and temporal period λ_t $(= 2\pi/\omega_t^o)$ of the cell. Now suppose a time delay of Δt is introduced such that the new receptive field profile is marked by the dashed lines as in Fig. 6b. It is obvious that the horizontal shift d generated by the time delay is given by

$$d = \frac{\lambda_x}{\lambda_t}\Delta t = \frac{\omega_t^o}{\omega_x^o}\Delta t.$$

This is exactly what we derived in equation 10.

Positional shift vs. phase-parameter difference

The binocular cell model proposed by Freeman et al. (Ohzawa et al., 1990; Freeman and Ohzawa, 1990; DeAngelis et al., 1991) assumes that the left and right receptive field profiles of a given cell have the same envelopes (on the corresponding left and right retinal locations) but different phase parameters for the excitatory/inhibitory modulations within the envelopes. An alternative is that there may be an overall shift (for both the envelopes and modulations) between the two profiles (Bishop, Henry and Smith, 1971; Maske, Yamane and Bishop, 1984; Wagner and Frost, 1993). The third and most general model assumes that the two profiles differ by both an overall positional shift and a phase parameter difference for the modulations (DeAngelis, Ohzawa and Freeman, 1995; Zhu and Qian, 1996). Although there are subtle differences between them (Zhu and Qian, 1996), we have shown previously (Zhu and Qian, 1996) that our stereo vision model (Qian, 1994) works equally well under all three possibilities. In this subsection we show that the main conclusions in this paper are not affected by the different choices of receptive field models either.

It is sufficient to consider the most general case where the left and right receptive field profiles of a simple cell differ by both an overall horizontal positional shift Δx and a phase parameter difference $\Delta\phi$. It can be shown (see the Appendix) that equation 7 (the response of a complex cell constructed from a quadrature pair of simple cells to a stimulus with both motion and disparity) should now be written as:

$$r_c \approx c^2\ |\tilde{I}(\omega_x^o, \omega_y^o)|^2\ \cos^2\left(\frac{\Delta\phi + \omega_x^o \Delta x}{2} - \frac{\omega_x^o D}{2}\right)\ \left[\iint_0^\infty |\tilde{f}_l(\omega_x,\ \omega_y,\ \omega_t')| d\omega_x d\omega_y\right]^2. \qquad (15)$$

Just like equation 7, the cell is tuned to both disparity and motion. The disparity tuning of the cell is now determined by both Δx and $\Delta\phi$, and the preferred disparity is given by $D_{pref} = \Delta x + \Delta\phi/\omega_x^o$. The motion selectivity of the cell is still determined by its spatiotemporal frequency tuning. Thus our previous conclusion of using a population of complex cells to recover stimulus motion and disparity simulationsly remains valid.

It can also be shown that with the hybrid receptive field model, equation 9 for the Pulfrich effect becomes:

$$r_c \approx c^2\ |\tilde{I}(\omega_x^o,\ \omega_y^o)|^2\ \cos^2\left(\frac{\Delta\phi + \omega_x^o \Delta x}{2} - \frac{\omega_t^o \Delta t}{2}\right)\ \left[\iint_0^\infty |\tilde{f}_l(\omega_x,\ \omega_y,\ \omega_t')| d\omega_x d\omega_y\right]^2. \qquad (16)$$

Again, by comparing equations 15 and 16 we find that an interocular time delay Δt is equivalent to a binocular disparity as indicated by equation 10. Similar arguments apply to the generalized Pulfrich effects with the noise patterns and the stroboscopic stimuli. Our conclusions on the Pulfrich effects thus remain the same.

Temporal stretching vs. temporal delay

In our above explanations of the Pulfrich-like phenomena, we have assumed that the effect of a neutral density filter placed in front of one eye is to introduce a time delay in neuronal responses of the cells' receptive fields in that eye. There is considerable experimental evidence supporting this assumption (Mansfield and Daugman, 1978; Lennie, 1981; Cynader et al., 1978; Carney et al., 1989). However, a recent study by (Kaufman and Palmer, 1990) suggests that this assumption may be an oversimplification. Specifically, these investigators found that attenuating the luminance of the input stimulus causes a temporal "stretching", not a pure delay, of the spatiotemporal receptive fields of simple cells. Thus, although the peak response is delayed, the effect of the filter cannot be simply characterized by shifting the cells' temporal response profiles. We show here that the Pulfrich effects can also be explained by the temporal stretching.

Let the left and the right receptive field profiles of a binocular simple cell be denoted by $f_l(x,\ y,\ t)$ and $f_r(x,\ y,\ t)$. Assume that the effect of the neutral density filter placed in front of the right eye is to stretch the right receptive field with respect to the $t = 0$ point by a factor of $k > 1$ along the time axis. It can then be shown (see the Appendix) that the complex cell response to a moving stimulus with disparity D is given by:

$$r_c \approx c^2\ |\tilde{I}(\omega_x^o,\ \omega_y^o)|^2\ [(1-r)^2+4r\cos^2(\frac{\Delta\phi}{2}-\frac{\omega_x^o D}{2}-\frac{\Delta\alpha}{2})]\ [\iint_0^\infty |\tilde{f}_l(\omega_x,\ \omega_y,\ \omega_t')|d\omega_x d\omega_y]^2. \quad (17)$$

where

$$r \equiv \frac{1}{k}\left|\frac{\tilde{f}_l(\omega_x^o,\ \omega_y^o,\ \omega_t^o/k)}{\tilde{f}_l(\omega_x^o,\ \omega_y^o,\ \omega_t^o)}\right| \quad (18)$$

and

$$\Delta\alpha \equiv \arg[\tilde{f}_l(\omega_x^o,\ \omega_y^o,\ \omega_t^o)] - \arg[\tilde{f}_l(\omega_x^o,\ \omega_y^o,\ \omega_t^o/k)]. \quad (19)$$

The arg function represents the phase angle of a complex quantity. As before, the $\Delta\phi$ dependent cosine term determines the disparity tuning of the cell. Even when there is no real disparity in the stimulus ($D = 0$), the temporal stretching ($k > 1$) produces an equivalent disparity of

$$d \approx \frac{\Delta\alpha}{\omega_x^o}. \quad (20)$$

This relation provides the theoretical basis of the Pulfrich effect under the assumption of temporal stretching. Equation 20 also holds for the generalized Pulfrich effects to the random noise patterns and the stroboscopic stimuli. Obviously, when there is no temporal stretching ($k = 1$), we have $r = 1$ and $\Delta\alpha = 0$, and the equivalent disparity is zero.

It can be shown that for the Gabor filters, equation 20 can be reduced to a form similar to equation 10:

$$d \approx \frac{\omega_t^o/k}{\omega_x^o}\overline{\Delta t}. \quad (21)$$

where $\overline{\Delta t}$ is the difference between the Gaussian center locations of the left and the (stretched) right receptive fields along the time axis. For the Gabor filters with their Gaussian envelopes centered at $t = 0$, stretching with respect to $t = 0$ will not change the center location, and therefore these filters will not generate the Pulfrich effects. However, these filters are non-causal and they never exist in the real brain.

We have also performed computer simulations similar to that shown in Fig. 2 but with the temporal delay replaced by the temporal stretching. An example is shown in Fig. 7 where the neutral density filter placed in front of the right eye is assumed to introduce a stretching factor of $k = 1.1$. The stretching is relative to the $t = 0$ point which is set at 2.5σ to the left of the Gaussian center. This particular value of k was chosen because it generates a shift of 4 time pixels between the Gaussian centers of the left and right receptive fields and therefore its effect is likely to match that of the 4-pixel temporal delay in Fig. 2. All the other parameters in the two simulations are identical. We conclude based on the similarity of the two figures that the temporal stretching can explain the Pulfrich effects just as well as the temporal delay. When all the other parameters are fixed, larger values of k generate larger equivalent disparities. For large k, however, the curve in Fig. 7b will become somewhat less smooth than the corresponding curve in Fig. 2b (results not shown) because the stretching of the right receptive field causes a mismatch of the preferred spatial frequencies of the left and right receptive fields, which in turn makes the model complex cells somewhat less independent of the stimulus Fourier phases.

Discussion

In this paper, we have developed an integrated model of motion and stereo vision using physiological properties of real binocular cells. Specifically, we have shown that under the general assumption that the left and right receptive fields of a binocular simple cell are well tuned to the same spatiotemporal frequencies, and that the main difference between the two receptive fields is a phase difference and/or a positional shift, the model complex cell constructed from a quadrature pair of such simple cells are tuned to both motion and binocular disparity. We have derived an explicit expression for the complex cell responses as a function of the cell parameters (see equation 7). The expression shows that the cell's preferred spatiotemporal frequencies determine its motion selectivity while the phase difference (and/or positional shift) and the preferred horizontal spatial frequency determine its disparity tuning. Therefore, by using a population of cells with their preferred frequencies and phase differences (and/or positional shift) covering a wide range, one could estimate the stimulus velocity and disparity simultaneously.

To our knowledge, our model is among the first integrated models of motion and stereopsis based solely on physiological mechanisms. On the other hand, there have been many psychophysical observations on motion-stereo interaction. It is therefore interesting to apply our model to explain these observations. We have previously employed a special version of the model to explain the disparity facilitation of transparent motion perception in paired dot patterns (Qian et al., 1994b; Qian et al., 1994a). In this paper we applied the model to explain a family of the Pulfrich-like phenomena. The depth illusion in these phenomena are all created by an interocular time delay produced either electronically or through a neutral density filter. The visual patterns used, however, are quite different in different experiments. It has been suggested previously that different neural mechanisms might be responsible for these phenomena. Our analysis demonstrates that they can all be explained in a unified way by our motion-stereo model. We also considered the possibility that the effect of the neutral density filter may be a temporal stretching instead of a pure delay and showed that the Pulfrich effects can be explained just as well.

There is a fundamental difference between our explanation and the standard explanation of the Pulfrich effect. The standard explanation asserts that the motion of the pendulum converts an interocular time delay into a real binocular disparity in the *stimulus*. According to this view, the Pulfrich effect is a stereo problem in disguise, and any purely stereo vision algorithm can explain the illusion. No temporal aspects need to be included in the algorithm. Indeed, if there were only stereo mechanisms but no motion mechanisms in the brain, or if motion and stereo were processed in completely separate neural pathways, the Pulfrich illusion would still be predicted by the standard explanation. Our explanation, on the other hand, does not assume any physical disparity in the *stimulus*, but instead makes the equivalence between an interocular time delay and a binocular disparity at the level of *neuronal responses*. Because of this, it is necessary that our model includes the temporal aspect of neuronal responses. The model relies on the fact that, based on the known spatiotemporal properties of real binocular cells in the brain, these cells cannot distinguish an interocular time delay from a binocular disparity. The two explanations are equivalent for the classical Pulfrich pendulum effect. However, the standard explanation fails to explain the generalized Pulfrich effects to dynamic noise patterns and stroboscopic stimuli while our model can explain these variations almost trivially. For the dynamic noise patterns the standard explanation does not work because there is simply no coherent motion to convert a time delay into a real disparity in the stimuli. One might argue that random correspondence in the noise pattern may provide the required motion signal. This argument is non-physiological, however, since a typical cell will contain in its receptive fields many noisy dots and cannot be said to detect a particular random correspondence while ignoring many others (see below). Our model explains this phenomenon naturally without any additional assumptions because the model is built on units with spatiotemporal frequency tuning. Dynamic noise patterns have a broad spatiotemporal spectrum and can excite these units, and therefore the effect should still be present. For the stroboscopic stimuli, the standard explanation fails because at any given time, only one of the two eyes sees a stimulus and therefore there is absolutely no disparity present in the stimulus at any time. A purely stereo vision algorithm would predict no depth in this case. Again, our model explains this phenomenon naturally without any additional assumptions because the temporal response properties of the units automatically "fill" in the time gaps in the stimuli.

We would like to emphasize the generality of our results as our derivations (see the Appendix) do not rely on any specific functional forms of the cell's receptive field profiles. Instead, we only made some rather general assumptions about cells' properties. We discuss two of these assumptions here in more detail. The first is the quadrature pair method for constructing complex cells from simple cells. This method was first used in motion energy models (Adelson and Bergen, 1985; Watson and Ahumada, 1985; van Santen and Sperling, 1985). It was later adopted to model disparity sensitive complex cells by Ohzawa et al. (1990). The mathematical justification of using the quadrature pair construction as a method of getting phase-independent disparity tuning was given by Qian (1994). Although there is no direct evidence supporting this construction, Freeman and coworkers (Ohzawa et al., 1990; Freeman and Ohzawa, 1990) found that this method models the responses of binocular complex cells quite well. Therefore, even if the brain does not literally use the method for constructing complex cells, it is valid as a phenomenological description of complex cell responses. We would like to point out that just as in the case of stereo vision (Qian, 1994), the quadrature pair method is not an indispensable part of our motion-stereo model either. To go from the simple cell response equation 40 to the complex cell response equation 44 in the Appendix, one can simply sum up the squared responses of many simple cells with their receptive field Fourier phases (θ_f) uniformly covering the entire range of 2π. One can even replace some of these simple cells with a set of properly aligned LGN center-surround cells so that the resulting complex cell is constructed from a mixture of simple and LGN cells.

The second assumption that warrants further discussion is that the frequency tuning of simple cells are much sharper than the Fourier spectra of the retinal images. This assumption is used when we go from equation 39 to equation 40. This is usually a good assumption because the natural environment is rich in complex textures and sharp boundaries. However, in the rare case when the visual system is looking at a sine wave grating this assumption is clearly violated. In general, if the retinal image has a Fourier spectrum much sharper than the frequency tuning of the cells, the equations we derived (equations 3, 7, 9, 10, 12, 13, 14 15, 16, 17 and 20) still maintain their forms but ω_x^o, ω_y^o and ω_t^o in these equations should now represent the dominant spatiotemporal frequencies of the image instead of the preferred frequencies of the cells. The preferred disparity and velocity of a given cell will thus be different for different stimulus frequencies. Consequently, if one uses a single family of cells at a fixed frequency scale to estimate stimulus disparity and velocity, the results will not be accurate unless the dominant stimulus frequencies match the preferred frequencies of the cells. This, however, does not pose a serious problem for the real visual system, except for the stimulus with very high or low frequencies (see the next paragraph), because the brain contains cells tuned to a wide range of frequencies and the cells with the highest responses are those whose preferred frequencies do match those of the stimuli.

Based on the above discussion, we can also determine how the predicted disparity by the model deviates from the actual values for sinusoidal stimuli with very high or low spatial frequencies. We consider the model with either the phase-parameter based or the position-shift based receptive field profiles (Zhu and Qian, 1996). If the phase-parameter based receptive field description is used, the model predicts that the disparities of those gratings with very high spatial frequencies will be underestimated while those with very low frequencies will be overestimated. The deviation will be more significant for the gratings with spatial frequencies further away from the main tuning range of the visual cortical cells. On the other hand, the position-shift based algorithm should always give the actual disparity value of the stimuli (within one spatial period of the gratings) because their preferred disparity is given by the shift parameter Δx independent of the stimulus frequencies. This result provides an opportunity for distinguishing the two types of receptive field descriptions via visual psychophysical experiments.

Two additional testable predictions can be made based on our theoretical results. First, we predict that the response of a binocular cell to an interocular time delay can be approximately matched by a binocular disparity according to equation 10. To test this prediction, one can first measure a cell's tuning curves to binocular disparity and to interocular time delay, then measure the preferred spatial frequency (ω_x^o) and temporal frequency (ω_t^o) of the same cell, and finally examine if the two tuning curves are related to each other by the scaling factor ω_t^o/ω_x^o along the horizontal axis. The second prediction is also based on equation 10. The equation predicts that cells with different preferred spatial to temporal frequency ratios will, by themselves, "report" different apparent Pulfrich depths for a given temporal delay. If we assume that the perceived depth corresponds to the disparities reported by the most responsive cells in a population (or by the population average of all cells weighted by their responses), then the perceived Pulfrich depth should vary according to equation 10 as we selectively excite different populations of cells by using stimuli with different spatial and temporal frequency contents. This prediction is particularly interesting when stimuli without coherent motion are used. Note that both predictions cannot be readily made by the standard explanation of the Pulfrich effect because it says nothing about the neurons in the brain.

Both motion detection and stereo vision have been formulated as solving a correspondence problem in the past. Algorithms based on this view often relies on explicit matching of fine image features in successive frames (for motion) or in the left and right images (for stereopsis). This explicit matching procedure, however, is unlikely to be physiological because the receptive field sizes of typical cells in the visual cortex are larger than the fine image features such as a dot or a zero-crossing in a random dot stereogram. Indeed, even the cells in monkey foveal striate cortex have a receptive field size of about 0.1 degree (Dow, Snyder, Vautin and Bauer, 1981). A cell simply integrates contributions of all image features in its receptive fields. It is hard to imagine that a cell could selectively mark out a certain feature among many other similar ones within its receptive field and try to match it with another feature in the next time frame or in the other retina. Our motion-stereo model does not suffer from this problem as it is based on the spatiotemporal receptive field properties of real cells, and like other energy based models (Adelson and Bergen, 1985; Watson and Ahumada, 1985; Heeger, 1987; Qian, 1994), it does not assume any explicit feature extraction or matching and the correspondence problem is solved in an implicit way through correlation-like operations (Qian and Zhu, 1995).

In conclusion, we have derived a unified model of motion and stereo vision using physiological mechanisms and have provided a comprehensive and quantitative explanation of a family of Pulfrich-like phenomena. We also made specific predictions for further experimental tests of the model. We are currently exploring applications of the model to other phenomena of motion-stereo interaction. Our work demonstrates how computational modeling can help bridge the gap between physiology and perception. It also suggests that it may be more fruitful to construct computational theories of vision based on neurophysiology than to treat theories as abstract concepts independent of physiological implementations (Marr, 1982).

Appendix

In this appendix we derive the complex cell response expressions under various conditions discussed in the text.

Derivation of equations 7 (motion-stereo integration)

Since a complex cell is constructed from a pair of simple cells, we first derive simple cell responses. For a binocular simple cell with left and right spatiotemporal receptive fields $f_l(x,\ y,\ t)$ and $f_r(x,\ y,\ t)$, its response to a stimulus with left and right retinal images $I_l(x,\ y,\ t)$ and $I_r(x,\ y,\ t)$ is given by (Freeman and Ohzawa, 1990; Ohzawa et al., 1990; Qian, 1994):

$$r_s(t) = \iiint_{-\infty}^{+\infty} dx dy dt' \; [f_l(x,y,t'-t)I_l(x,y,t') + f_r(x,y,t'-t)I_r(x,y,t')]. \tag{22}$$

Although formally the integration is carried over the entire spatiotemporal space, the actual domain is limited by the extent of the receptive fields. Note that the convolution operation is applied to the temporal dimension but not to the two spatial dimensions because we only need to consider neurons at a given spatial location. Apply the Fourier power theorem and use tilde to denote the Fourier transform of a function, we have:

$$r_s(t) = \iiint_{-\infty}^{+\infty} d\omega_x d\omega_y d\omega_t \; [\tilde{f}_l(\omega_x,\omega_y,\omega_t)\tilde{I}_l^*(\omega_x,\omega_y,\omega_t) + \tilde{f}_r(\omega_x,\omega_y,\omega_t)\tilde{I}_r^*(\omega_x,\omega_y,\omega_t)]e^{-i\omega_t t}, \tag{23}$$

where ω_x, ω_y, and ω_t are the Fourier frequencies along the x, y and t dimensions respectively, and $*$ denotes complex conjugate. We have used the fact that the Fourier transforms of $f_l(x,y,t'-t)$ and $f_l(x,y,t')$ are related by:

$$\mathcal{F}(f_l(x,y,t'-t)) = e^{-i\omega_t t}\mathcal{F}(f_l(x,y,t')) \tag{24}$$

in equation 23.

Freeman and coworkers (DeAngelis et al., 1991; DeAngelis et al., 1995) proposed based their quantitative physiological studies that the left and right receptive fields of a binocular simple cell have corresponding retinal locations but different phase parameters for the (excitatory/inhibitory) modulations within the receptive fields, as represented by equations 1 and 2. It is easy to show that, in the Fourier domain, equations 1 and 2 differ by $e^{i\,\mathrm{sign}(\omega_x)\Delta\phi}$ for well-tuned receptive fields, where $\Delta\phi$ is the phase parameter difference defined in equation 4, and the sign function is equal to 1 when its argument is positive, and -1 otherwise.[7] We can therefore assume that in general the Fourier transforms of the left and right receptive fields are related by[8]

$$\tilde{f}_r(\omega_x,\omega_y,\omega_t) \approx \tilde{f}_l(\omega_x,\omega_y,\omega_t)e^{i\,\mathrm{sign}(\omega_x)\Delta\phi}. \tag{25}$$

[7]Note that under the alternative assumption of an overall horizontal positional shift (Δx) between the left and right receptive fields (Zhu and Qian, 1996; Wagner and Frost, 1993; DeAngelis et al., 1995), the two Fourier transforms will differ by $e^{i\omega_x\Delta x}$. The consequence of this assumption will be considered below.

[8]More generally, one can assume a spatiotemporal phase instead of associating the phase with the x dimension. The sign function will then depend on all three frequency variables. The essentially identical results can be derived.

We first derive equation 7. The left and right images of a stimulus patch with constant disparity D and velocity (V_x, V_y) can be written as[9]:

$$\begin{aligned} I_l(x,y,t) &= I(x - V_x t, y - V_y t), && (26)\\ I_r(x,y,t) &= I(x - V_x t + D, y - V_y t). && (27)\end{aligned}$$

Using the definition of Fourier transform, it is easy to show that

$$\begin{aligned} \tilde{I}_l(\omega_x,\omega_y,\omega_t) &= \delta(\omega_x V_x + \omega_y V_y + \omega_t)\tilde{I}(\omega_x,\omega_y) && (28)\\ \tilde{I}_r(\omega_x,\omega_y,\omega_t) &= \tilde{I}_l(\omega_x,\omega_y,\omega_t)e^{i\omega_x D} && (29)\end{aligned}$$

where $\delta()$ is the Dirac delta function. Substituting equations 25, 28 and 29 into equation 23 and using the delta function to carry out the integration over ω_t, we have:

$$r_s = \iint_{-\infty}^{\infty} d\omega_x d\omega_y\ \tilde{I}(\omega_x,\omega_y)\tilde{f}_l(\omega_x,\omega_y,\omega_t')[1 + e^{i\,\mathrm{sign}(\omega_x)\Delta\phi - i\omega_x D}]e^{-i\omega_t' t} \tag{30}$$

where

$$\omega_t' = -\omega_x V_x - \omega_y V_y \tag{31}$$

is the motion constraint (Watson and Ahumada, 1983).

Since $I(x,y)$ and $f_l(x,y,t)$ are real functions, their Fourier transforms satisfy the following properties:

$$\tilde{I}(-\omega_x,-\omega_y) = \tilde{I}^*(\omega_x,\omega_y) \tag{32}$$

and

$$\tilde{f}_l(-\omega_x,-\omega_y,-\omega_t) = \tilde{f}_l^*(\omega_x,\omega_y,\omega_t). \tag{33}$$

Change the integration variables ω_x and ω_y to $-\omega_x$ and $-\omega_y$ in equation 30 and apply the above identities, we have

$$r_s = \iint_{-\infty}^{\infty} d\omega_x d\omega_y\ \tilde{I}^*(\omega_x,\omega_y)\tilde{f}_l^*(\omega_x,\omega_y,\omega_t')[1 + e^{-i\,\mathrm{sign}(\omega_x)\Delta\phi + i\omega_x D}]e^{i\omega_t' t} \tag{34}$$

Since the integrands of equations 30 and 34 are conjugate to each other, we add the two equations to obtain:

$$r_s = \iint_{-\infty}^{\infty} d\omega_x d\omega_y\ Re\{\tilde{I}(\omega_x,\omega_y)\tilde{f}_l(\omega_x,\omega_y,\omega_t')[1 + e^{i(\mathrm{sign}(\omega_x)\Delta\phi - \omega_x D)}]e^{-i\omega_t' t}\} \tag{35}$$

where Re denotes the real part of a complex quantity. The terms in the integrand are in general complex, and each of them can be written as an amplitude multiplied by a complex phase term:

$$\begin{aligned} \tilde{I}(\omega_x,\omega_y) &= |\tilde{I}(\omega_x,\omega_y)|e^{i\theta_I(\omega_x,\omega_y)} && (36)\\ \tilde{f}_l(\omega_x,\omega_y,\omega_t) &= |\tilde{f}_l(\omega_x,\omega_y,\omega_t)|e^{i\theta_f(\omega_x,\omega_y,\omega_t)} && (37)\\ 1 + e^{i(\mathrm{sign}(\omega_x)\Delta\phi - \omega_x D)} &= 2|\cos(\frac{\mathrm{sign}(\omega_x)\Delta\phi}{2} - \frac{\omega_x D}{2})|e^{i\theta(\omega_x)} && (38)\end{aligned}$$

[9]The disparities and velocities of real world stimuli are, of course, not constant. However, this is a good approximation within the spatiotemporal windows of visual cortical cells.

Equation 35 can then be written as:

$$r_s = 2 \iint\limits_{-\infty}^{\infty} d\omega_x d\omega_y \; |\tilde{I}(\omega_x, \omega_y)| \; |\tilde{f}_l(\omega_x, \omega_y, \omega_t')| \; |\cos(\frac{\text{sign}(\omega_x)\Delta\phi}{2} - \frac{\omega_x D}{2})| \; \cos(\theta_I + \theta_f + \theta - \omega_t' t). \tag{39}$$

We did not explicitly write out the ω dependence of the θ's in the above equation for clarity. Most primary visual cortical cells are well tuned to spatial frequencies. Assume that the cell in equation 39 is tuned to the frequencies (ω_x^o, ω_y^o) and that its tuning is significantly sharper than that of the other terms in the equation, we can then approximate $\tilde{f}_l(\omega_x, \omega_y, \omega_t)$ by two delta functions, one peaked at (ω_x^o, ω_y^o) and the other at $(-\omega_x^o, -\omega_y^o)$, and simplify equation 39 into:

$$r_s \approx 4|\tilde{I}(\omega_x^o, \omega_y^o)| \; |\cos(\frac{\Delta\phi}{2} - \frac{\omega_x^o D}{2})| \; \cos(\theta_I + \theta_f + \theta - \bar{\omega}_t t) \iint\limits_{0}^{\infty} d\omega_x d\omega_y \; |\tilde{f}_l(\omega_x, \omega_y, \omega_t')|, \tag{40}$$

where

$$\bar{\omega}_t = -\omega_x^o V_x - \omega_y^o V_y. \tag{41}$$

Here we have let $\text{sign}(\omega_x^o) = 1$ since, without loss of generality, we can assume $\omega_x^o > 0$. We also used the fact that all three θ's satisfy the relation $\theta(-\omega_x^o, -\omega_y^o) = -\theta(\omega_x^o, \omega_y^o)$. Equation 40 is the expression for the simple cell response.

We now compute complex cell responses using the quadrature pair construction. It is easy to show that the response of the simple cell that forms a quadrature pair with the simple cell in equation 40 is given by:

$$r_s^q \approx 4|\tilde{I}(\omega_x^o, \omega_y^o)| \; |\cos(\frac{\Delta\phi}{2} - \frac{\omega_x^o D}{2})| \; \sin(\theta_I + \theta_f + \theta - \bar{\omega}_t t) \iint\limits_{0}^{\infty} d\omega_x d\omega_y \; |\tilde{f}_l(\omega_x, \omega_y, \omega_t')|. \tag{42}$$

This is because the θ_f's of the two simple cells differ by $\pi/2$ while all the other parameters are the same. The response of a complex cell constructed from this quadrature pair is then given by:

$$r_c = (r_s)^2 + (r_s^q)^2 \tag{43}$$

$$\approx 16|\tilde{I}(\omega_x^o, \omega_y^o)|^2 \; \cos^2(\frac{\Delta\phi}{2} - \frac{\omega_x^o D}{2})[\iint\limits_{0}^{\infty} d\omega_x d\omega_y \; |\tilde{f}_l(\omega_x, \omega_y, \omega_t')|]^2. \tag{44}$$

This completes the derivation of equation 7.

Derivation of equation 9 (Pulfrich's pendulum)

To derive equation 9, $f_r(x, y, t)$ should now be replaced by $f_r(x, y, t + \Delta t)$ in equation 22. Or equivalently, its Fourier transform $\tilde{f}_r(\omega_x, \omega_y, \omega_t)$ should be replaced by $\tilde{f}_r(\omega_x, \omega_y, \omega_t)e^{i\omega_t \Delta t}$. Also, disparity D in equation 27 should be set to zero. Here we assume that the cells are well tuned to spatiotemporal frequencies $(\omega_x^o, \; \omega_y^o, \; \omega_t^o)$. For the cells to have good responses $\bar{\omega}_t$ should be equal to ω_t^o. All the other steps of derivation are the same as above.

Derivation of equation 12 (the generalized Pulfrich effects)

To derive equation 12, $f_r(x,y,t)$ should be replaced by $f_r(x,y,t+\Delta t)$ in equation 22 and equations 26 and 27 should be replaced by

$$I_l(x,y,t) = I_r(x,y,t) = I(x,y,t), \tag{45}$$

because here we only assume a general spatiotemporal pattern which may or may not contain any coherent motion. When we get to the stage of equation 40, we need to approximate $\tilde{f}_l(\omega_x,\omega_y,\omega_t)$ by two delta functions peaked at $(\omega_x^o,\omega_y^o,\omega_t^o)$ and $(-\omega_x^o,-\omega_y^o,-\omega_t^o)$. All the other steps of derivation are the same as those for deriving equation 7.

Derivation of equations 15 and 16 (receptive fields with both positional shift and phase difference)

When there is both a horizontal positional shift Δx and a phase parameter difference $\Delta\phi$ between the left and right receptive fields of a simple cell, equation 25 should be replaced by

$$\tilde{f}_r(\omega_x,\omega_y,\omega_t) \approx \tilde{f}_l(\omega_x,\omega_y,\omega_t) e^{i\,\mathrm{sign}(\omega_x)\Delta\phi + i\omega_x\Delta x}. \tag{46}$$

All the other steps for deriving equations 15 and 16 are same as those for deriving equations 7 and 9 above.

Derivation of equation 17 (the Pulfrich effects explained by a temporal stretching)

When the right receptive field $f_r(x,y,t)$ is temporally stretched by a factor of k with respect to the $t=0$ point, its mathematical description becomes $f_r(x,y,kt)$. Equation 25 should therefore be modified as:

$$\mathcal{F}(f_r(x,y,kt)) = \frac{1}{k}\tilde{f}_r(\omega_x,\omega_y,\frac{\omega_t}{k}) \approx \frac{1}{k}\tilde{f}_l(\omega_x,\omega_y,\frac{\omega_t}{k}) e^{i\,\mathrm{sign}(\omega_x)\Delta\phi}. \tag{47}$$

Using the procedures similar to that for deriving equation 35 above, we found that the simple cell's response to a stimulus with motion and disparity is given by:

$$r_s = \iint_{-\infty}^{\infty} d\omega_x d\omega_y \; Re\{\tilde{I}(\omega_x,\omega_y)\tilde{f}_l(\omega_x,\omega_y,\omega_t') \left[1 + \frac{\tilde{f}_l(\omega_x,\omega_y,\omega_t'/k)}{k\tilde{f}_l(\omega_x,\omega_y,\omega_t')}\right] e^{i(\mathrm{sign}(\omega_x)\Delta\phi - \omega_x D)} e^{-i\omega_t' t}\} \tag{48}$$

Let

$$f_l(\omega_x,\omega_y,\omega_t') \equiv |f_l(\omega_x,\omega_y,\omega_t')| e^{i\alpha(\omega_x,\omega_y,\omega_t')} \tag{49}$$

and define

$$r(\omega_x,\omega_y,\omega_t',k) \equiv \frac{1}{k}\frac{\tilde{f}_l(\omega_x,\omega_y,\omega_t'/k)}{\tilde{f}_l(\omega_x,\omega_y,\omega_t')}, \tag{50}$$

and

$$\Delta\alpha(\omega_x,\omega_y,\omega_t',k) \equiv \alpha(\omega_x,\omega_y,\omega_t') - \alpha(\omega_x,\omega_y,\omega_t'/k), \tag{51}$$

we have:

$$r_s = \iint_{-\infty}^{\infty} d\omega_x d\omega_y \; Re\{\tilde{I}(\omega_x,\omega_y)\tilde{f}_l(\omega_x,\omega_y,\omega_t')[1 + r e^{i(\mathrm{sign}(\omega_x)\Delta\phi - \omega_x D - \Delta\alpha)}] e^{-i\omega_t' t}\} \tag{52}$$

Applying the similar procedures that led us from equation 35 to equation 44, we obtain equation 17.

Acknowledgments

We would like to thank Dr. Yudong Zhu for his help with Fig. 2b, and the two anonymous reviewers for their insightful comments. NQ is supported by NIH grant MH54125 and a research grant from the McDonnell-Pew Program in Cognitive Neuroscience. RAA is supported by NIH grant EY07492 and the Sloan Center for Theoretical Neurobiology at Caltech.

References

Adelson, E. H. and Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion, *J. Opt. Soc. Am. A* **2**(2): 284–299.

Adelson, E. H. and Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns, *Nature* **300**(5892): 523–525.

Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque, *J. Neurophysiol.* **52**: 1106–1130.

Anstis, S. M. and Hassis, J. P. (1974). Movement aftereffects contingent on binocular disparity, *Perception* **3**: 153–168.

Bishop, P. O., Henry, G. H. and Smith, C. J. (1971). Binocular interaction fields of single units in the cat striate cortex, *J. Phsiol.* **216**: 39–68.

Bradley, D. C., Qian, N. and Andersen, R. A. (1995). Intrgration of motion and stereopsis in cortical area MT of the macaque, *Nature* **373**: 609–611.

Burr, D. C. (1979). Accuity for apparent Vernier offset, *Vision Res.* **19**: 835–837.

Burr, D. C. and Ross, J. (1979). How does binocualr delay give information about depth?, *Vision Res.* **19**: 523–532.

Carney, T., Paradiso, M. A. and Freeman, R. D. (1989). A physiological correlate of the pulfrich effect in cortical neurons of the cat, *Vision Res.* **29**: 155–165.

Cynader, M. S., Gardner, J. C. and Douglas, R. M. (1978). Neural mechanisms underlying stereoscopic depth perception in cat visual cortex, *in* S. J. Cool and E. L. Smith III (eds), *Frontiers in Visual Science*, Springer, Berlin, pp. 373–386.

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, *J. Opt. Soc. Am. A* **2**: 1160–1169.

DeAngelis, G. C., Ohzawa, I. and 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. and Freeman, R. D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. general characteristics and postnatal development, *J. Neurophysiol.* **69**: 1091–1117.

DeAngelis, G. C., Ohzawa, I. and Freeman, R. D. (1995). Neuronal mechanisms underlying stereopsis: how do simple cells in the visual cortex encode binocular disparity?, *Perception* **24**: 3–31.

Dow, B. M., Snyder, A. Z., Vautin, R. G. and Bauer, R. (1981). Magnification factor and receptive field size in foveal striate cortex of the monkey, *Exp. Brain Res.* **44**: 213–228.

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

Falk, D. S. (1980). Dynamic visual noise and the stereophenomenon: Interocular time delays, depth and coherent velocities, *Percept. Psychophys.* **28**: 19–27.

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

Gardner, J. C., Douglas, R. M. and Cynader, M. S. (1985). A time-based stereoscopic depth mechanism in the visual cortex, *Brain Res.* **328**: 154–157.

Grzywacz, N. M. and Yuille, A. L. (1990). A model for the estimate of local image velocity by cells in the visual cortex, *Proc. R. Soc. Lond. A* **239**: 129–161.

Hamilton, D. B., Albrecht, D. G. and Geisler, W. S. (1989). Visual cortical receptive fields in monkey and cat: spatial and temporal phase transfer function, *Vision Res.* **29**: 1285–1308.

Heeger, D. J. (1987). Model for the extraction of image flow, *J. Opt. Soc. Am. A* **4**(8): 1455–1471.

Hildreth, E. C. (1984). Computations underlying the measurement of visual motion, *Artificial Intelligence* **23**(3): 309–355.

Hubel, D. H. and Wiesel, T. (1968). Receptive fields and functional architecture of the monkey striate cortex., *J. Phsiol.* **195**: 215–243.

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

Julez, B. and White, B. (1969). Short term visual memory and the Pulfrich phenomenon, *Science* **222**: 639–641.

Kaufman, D. A. and Palmer, L. A. (1990). The luminance dependence of spatiotemporal response of cat striate cortical units, *Invest. Opthalmol. and Vis. Sci. Suppl. (ARVO)* **31**: 398.

Lennie, P. (1981). The physiological basis of variation in visual latency, *Vision Res.* **21**: 815–824.

Mansfield, R. J. W. and Daugman, J. D. (1978). Retinal mechanisms of visual latency, *Vision Res.* **18**: 1247–1260.

Marcelja, S. (1980). Mathematical description of the responses of simple cortical cells, *J. Opt. Soc. Am. A* **70**: 1297–1300.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, W. H. Freeman and Company, San Fransisco.

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

Marr, D. and Poggio, T. (1979). A computational theory of human stereo vision, *Proc. R. Soc. Lond. B* **204**: 301–328.

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

Maunsell, J. H. R. and Van Essen, D. C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey II. Binocular interactions and sensitivity to binocular disparity, *J. Neurophysiol.* **49**: 1148–1167.

Morgan, M. J. (1975). Stereo illusion based on visual persistence, *Nature* **256**: 639–640.

Morgan, M. J. and Thompson, P. (1975). Apparent motion and the pulfrich effect, *Perception* **4**: 3–18.

Nawrot, M. and Blake, R. (1989). Neural integration of information specifying structure from stereopsis and motion, *Science* **244**: 716–718.

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

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

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

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

Prazdny, K. (1985). Detection of binocular disparities, *Biol. Cybern.* **52**: 93–99.

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

Qian, N. and Andersen, R. A. (1994). Transparent motion perception as detection of unbalanced motion signals II: Physiology, *J. Neurosci.* **14**: 7367–7380.

Qian, N. and Sejnowski, T. J. (1989). Learning to solve random-dot stereograms of dense and transparent surfaces with recurrent backpropagation, *Proceedings of the 1988 Connectionist Models Summer School*, pp. 435–443.

Qian, N. and Zhu, Y. (1995). Physiological computation of binocular disparity, *Soc. Neurosc. Abs.* **21**: 1507.

Qian, N., Andersen, R. A. and Adelson, E. H. (1994a). Transparent motion perception as detection of unbalanced motion signals I: Psychophysics, *J. Neurosci.* **14**: 7357–7366.

Qian, N., Andersen, R. A. and Adelson, E. H. (1994b). Transparent motion perception as detection of unbalanced motion signals III: Modeling, *J. Neurosci.* **14**: 7381–7392.

Regan, D. and Beverley, K. I. (1973). Disparity detectors in human depth perception: evidence for directional selectivity, *Nature* **181**: 877–879.

Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system, *in* W. A. Rosenblith (ed.), *Sensory Communication*, Wiley, New York.

Reid, R. C., Soodak, R. E. and Shapley, R. M. (1987). Linear mechanisms of directional selectivity in simple cells of cat striate cortex, *Proc. Nat. Acad. Sci. USA* **84**: 8740–8744.

Ross, J. (1974). Stereopsis by binocular delay, *Nature* **248**: 363–364.

Ross, J. and Hogben, J. H. (1975). The Pulfrich effect and the short term memory of stereopsis, *Vision Res.* **15**: 1289–1290.

Sanger, T. D. (1988). Stereo disparity computation using gabor filters, *Biol. Cybern.* **59**: 405–418.

Snowden, R. J., Treue, S., Erickson, R. E. and Andersen, R. A. (1991). The response of area MT and V1 neurons to transparent motion, *J. Neurosci.* **11**: 2768–2785.

Tyler, C. W. (1974). Stereopsis in dynamic visual noise, *Nature* **250**: 781–782.

van Santen, J. P. H. and Sperling, G. (1985). Elaborated Reichardt detectors, *J. Opt. Soc. Am. A* **2**: 300–321.

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

Watson, A. B. and Ahumada, A. J. (1983). A look at motion in the frequency domain, *in* J. K. Tsotsos (ed.), *Motion: representation and perception*, North-Holland, pp. 1–10.

Watson, A. B. and Ahumada, A. J. (1985). Model of human visual-motion sensing, *J. Opt. Soc. Am. A* **2**: 322–342.

Yeshurun, Y. and Schwartz, E. L. (1989). Cepstral filtering on a columnar image architecture – A fast algorithm for binocular stereo segmentation, *IEEE Pat. Anal. Mach. Intell.* **11**: 759–767.

Zhu, Y. and Qian, N. (1996). Binocular receptive fields, disparity tuning, and characteristic disparity, *Neural Comp.* (in press).

Figure Captions

1. A schematic drawing of the classical Pulfrich effect (top view). A pendulum is oscillating back and forth in the frontoparallel plane indicated by the solid line. When a neutral density filter is placed in front of the right eye, the pendulum appears to move in an elliptical path in depth as indicated by the dashed line. The direction of rotation in depth, marked by the arrows in the figure, will reverse if the neutral density filter is placed in front of the left eye.

2. (a) The horizontal position of the pendulum as a function of time for one full cycle of oscillation. The pendulum first swings to the right (positive x direction), it then reverses direction and moves to the left, and finally it moves to the right again. The maximum speed is 1 space pixel per time pixel. (b) The computed equivalent disparity as a function of horizontal position and time (see part a). The data points from the simulation are shown as small solid circles. Lines are drawn from the data points to the x–t plane in order to indicate the spatiotemporal location of each data point. A time delay of 4 pixels is assumed for the right receptive fields of all the model cells. The pendulum has negative equivalent disparity (and therefore is seen as closer to the observer) when it is moving to the right and have positive equivalent disparity (further away from the observer) when it is moving to the left. The projection of the 3D plot onto the d–x plane forms a closed path similar to the ellipse in Fig. 1.

3. (a) The spatiotemporal representation of a dynamic noise pattern. Each dot has a size of 1 spatial pixel and stays for 1 time pixel before its polarity is randomly reassigned with 0.5 probability. (b) The computed equivalent disparities with 11 families of complex cells. A time delay of 4 pixels is assumed for the right receptive fields of all the model cells. The preferred horizontal velocity of each cell family is indicated by an arrow and the corresponding equivalent disparity reported by that family is represented by the vertical distance from the zero disparity point. The longest arrows in the figure represent a speed of 1 space pixel per time pixel.

4. (a) The spatiotemporal representation of a stroboscopic pendulum. The two sets of dots are the left and right eyes' views respectively. The time delay between the two sets of dots (4 time pixels) and the duration of each dot (1 time pixel) are exaggerated in the drawing for the purpose of illustration. (b) The computed equivalent disparity at each dot location, presented in the same format as in Fig. 2. The result is very similar to the continuous case shown in Fig. 2.

5. Schematic drawings of three simple cells' left and right spatiotemporal receptive field profiles, illustrating the approximate equivalence of an interocular temporal delay to a binocular disparity. The grey and white lobes represent excitatory and inhibitory subfields respectively. The rectangular frames and the crosses inside are drawn for facilitating comparisons between different profiles. (a) The left and right receptive profiles are exactly identical. (b) The left profile is identical to that in (a) while the right profile has been phase shifted (notice the relative position of the grey area to the cross). The phase shift generates a horizontal offset between the left and right

receptive field modulations. (c) The left profile is identical to that in (b) while the right profile has been delayed (shifted upwards) in time. The time delay also generates an apparent horizontal offset between the left and right receptive field modulations, which cancels the effect of the phase shift in (b).

6. A geometric explanation of equation 10. (a) Lines are drawn through the central ridges of the excitatory and inhibitory regions of a receptive field profile. The horizontal and vertical distances between these lines (indicated by the short thin lines in the figure) are approximately equal to the preferred spatial and temporal periods of the cell. (b) If the receptive field profile is now delayed by Δt in time (i.e., shifted upwards) as indicated by the dotted lines, an apparent horizontal shift of d is also introduced.

7. (a) The spatiotemporal representation of an oscillating pendulum same as in Fig. 2a. (b) The computed equivalent disparity, presented in the same format as in Fig. 2, when a temporal stretch factor $k = 1.1$ is introduced for the right receptive fields of all the model cells. The results are very similar to those generated by a temporal delay of 4 pixels in Fig. 2b.

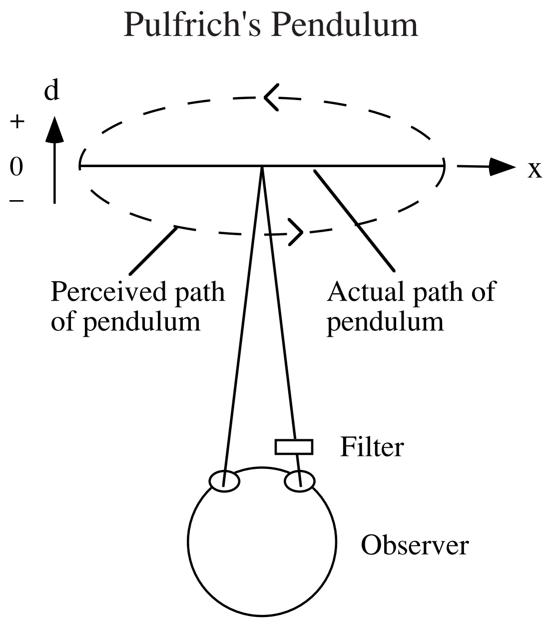

Fig. 1 Qian & Andersen

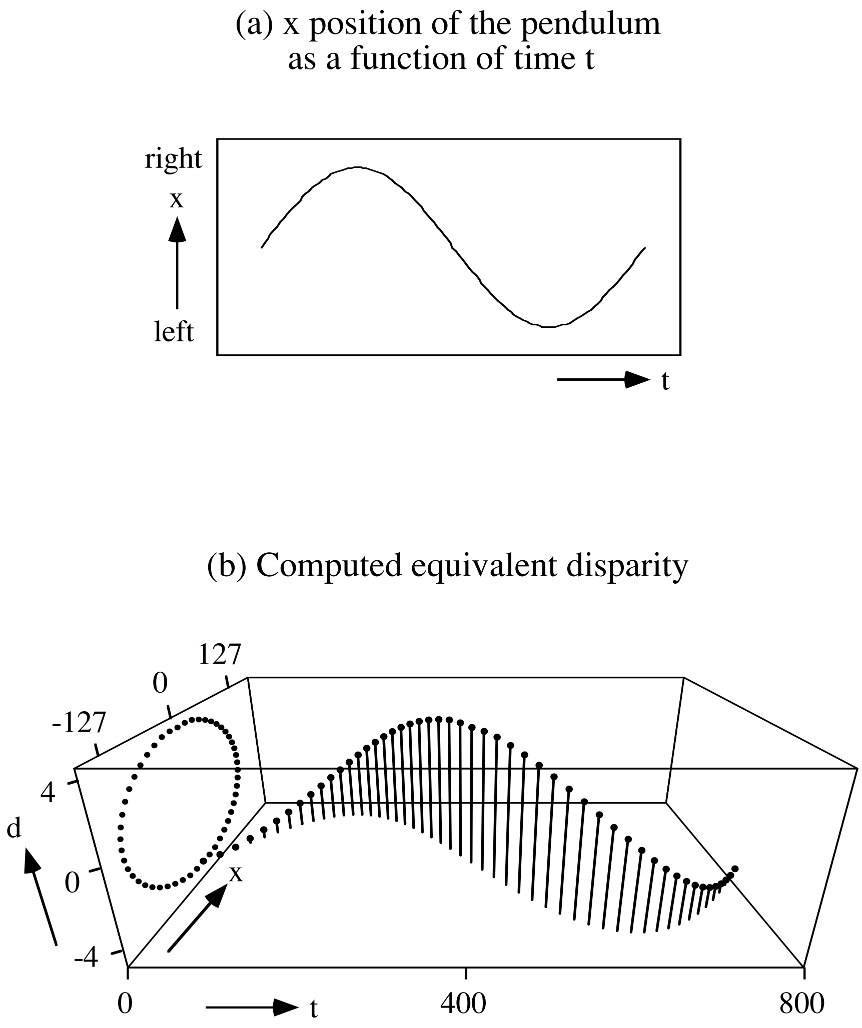

Fig. 2 Qian & Andersen

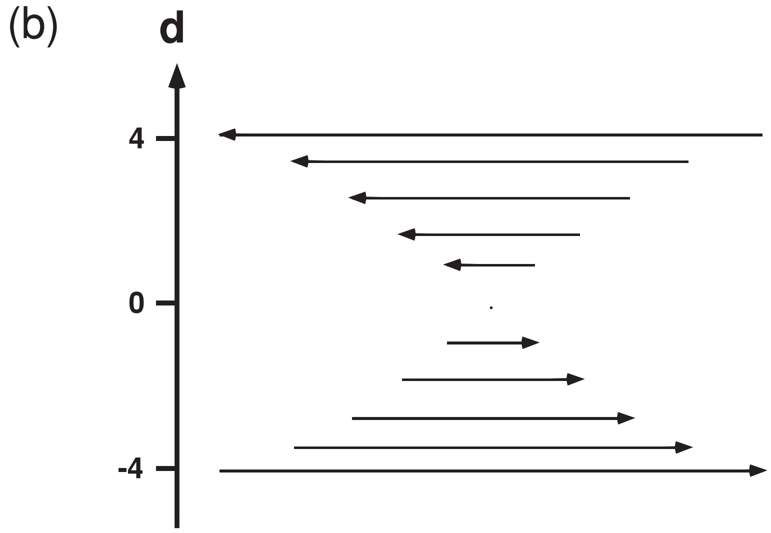

Fig. 3 Qian & Andersen

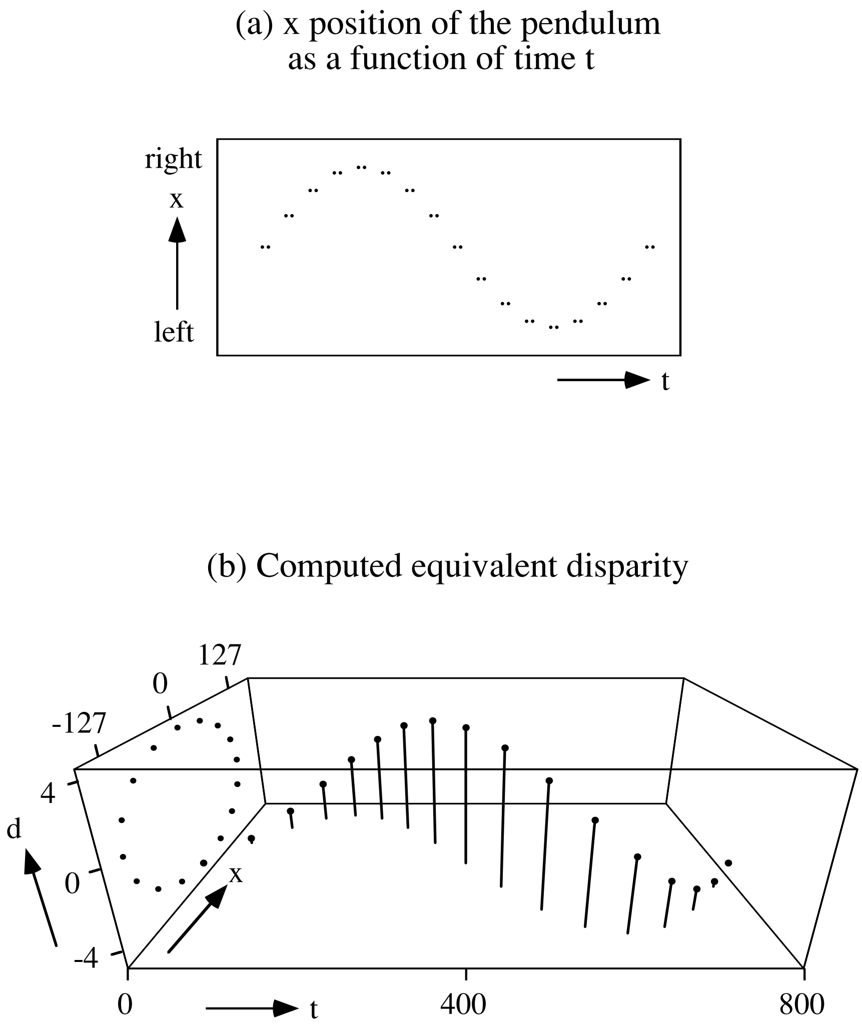

Fig. 4 Qian & Andersen

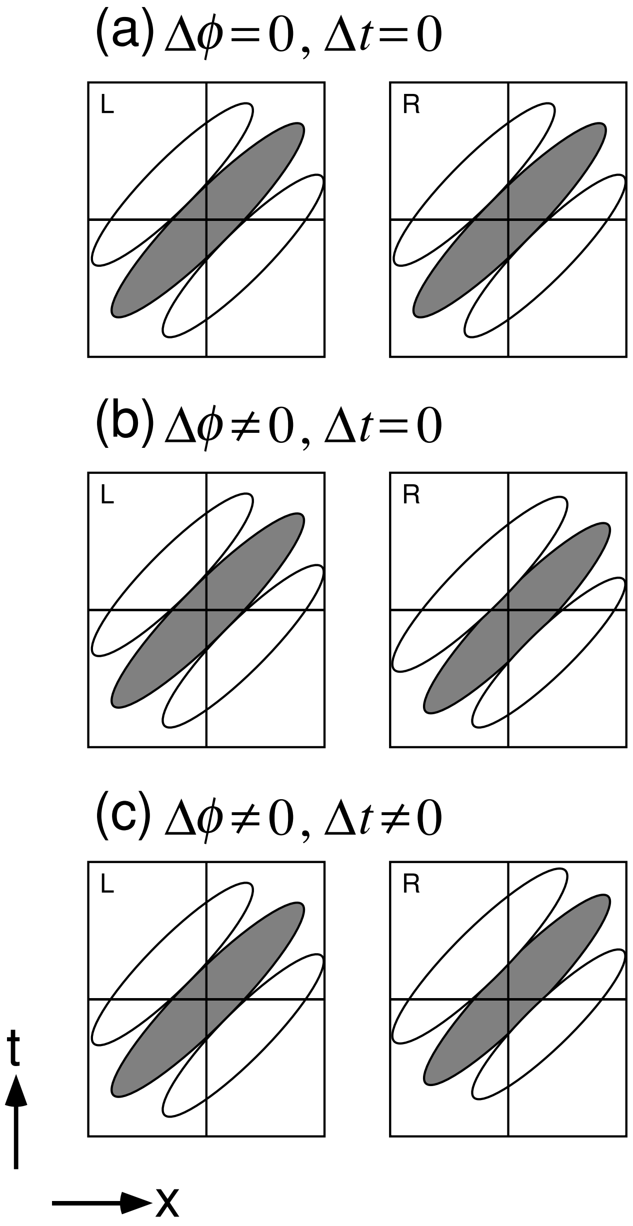

Fig. 5 Qian & Andersen

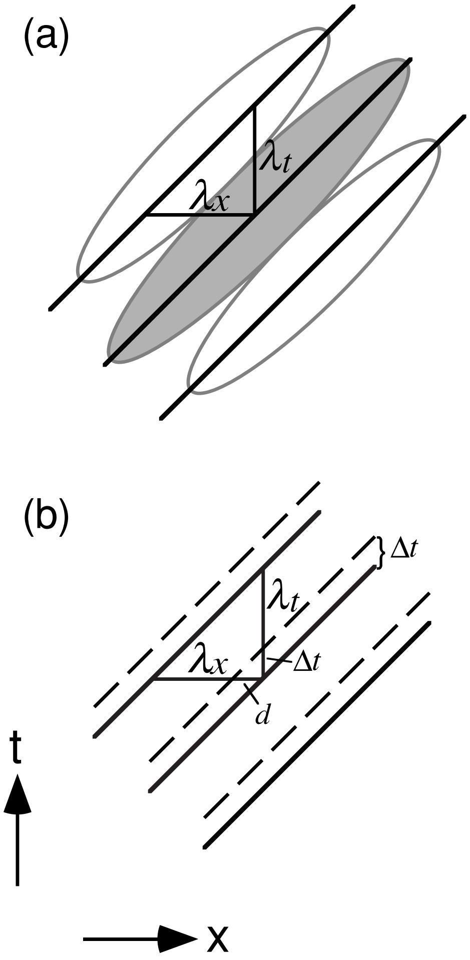

Fig. 6 Qian & Andersen

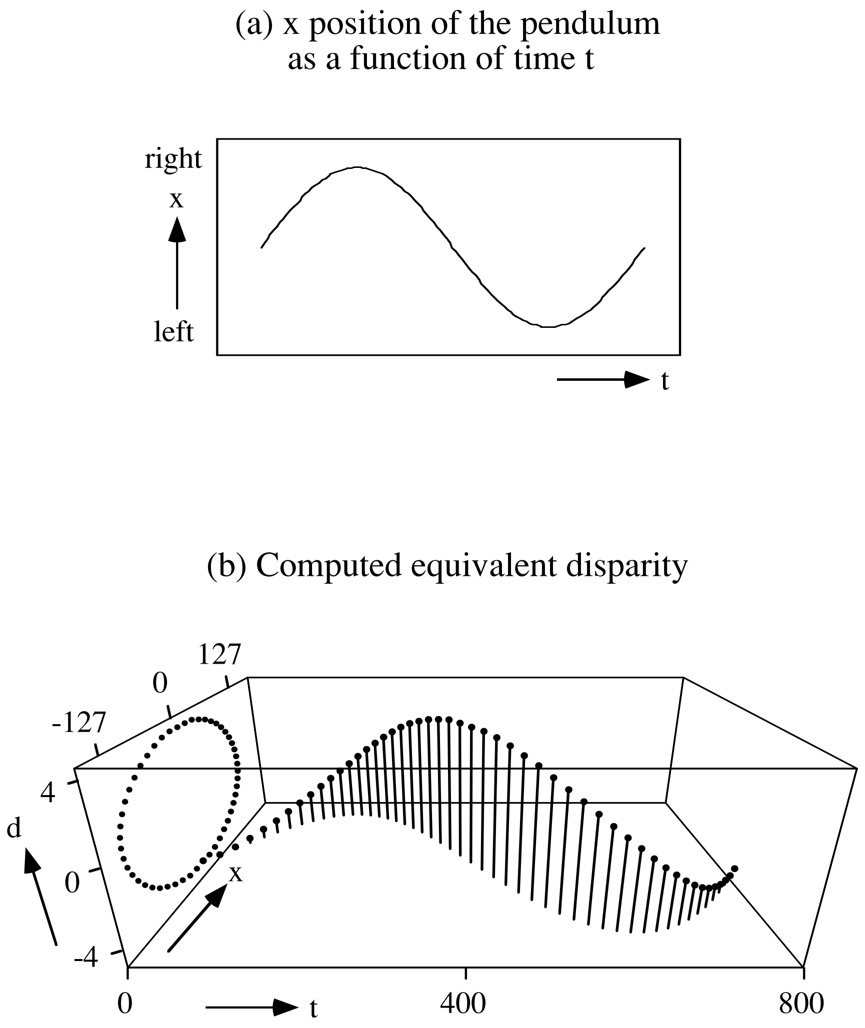

Fig. 7 Qian & Andersen