Modeling V1 Disparity Tuning to Time-Varying Stimuli

YUZHI CHEN,[1] YUNJIU WANG,[2] AND NING QIAN[1]
[1]Center for Neurobiology and Behavior and Department of Physiology and Cellular Biophysics, Columbia University,
New York, New York 10032; and [2]Laboratory of Visual Information Processing, Institute of Biophysics, Chinese Academy
of Sciences, Beijing 100101, China

Chen, Yuzhi, Yunjiu Wang, and Ning Qian. Modeling V1 disparity
tuning to time-varying stimuli. *J Neurophysiol* 86: 143–155, 2001.
Most models of disparity selectivity consider only the spatial proper-
ties of binocular cells. However, the temporal response is an integral
component of real neurons' activities, and time-varying stimuli are
often used in the experiments of disparity tuning. To understand the
temporal dimension of V1 disparity representation, we incorporate a
specific temporal response function into the disparity energy model
and demonstrate that the binocular interaction of complex cells is
separable into a Gabor disparity function and a positive time function.
We then investigate how the model simple and complex cells respond
to widely used time-varying stimuli, including motion-in-depth pat-
terns, drifting gratings, moving bars, moving random-dot stereograms,
and dynamic random-dot stereograms. It is found that both model
simple and complex cells show more reliable disparity tuning to
time-varying stimuli than to static stimuli, but similarities in the
disparity tuning between simple and complex cells depend on the
stimulus. Specifically, the disparity tuning curves of the two cell types
are similar to each other for either drifting sinusoidal gratings or
moving bars. In contrast, when the stimuli are dynamic random-dot
stereograms, the disparity tuning of simple cells is highly variable,
whereas the tuning of complex cells remains reliable. Moreover, cells
with similar motion preferences in the two eyes cannot be truly tuned
to motion in depth regardless of the stimulus types. These simulation
results are consistent with a large body of extant physiological data,
and provide some specific, testable predictions.

INTRODUCTION

Numerous physiological studies have documented disparity-
tuned cells in V1 (Barlow et al. 1967; Freeman and Ohzawa
1990; Poggio and Poggio 1984). To understand the mechanism
of tuning, many researchers have also investigated how the
disparity responses of a cell may be explained by the under-
lying binocular receptive field (RF) structure. Since disparity is
a spatially defined property, nearly all stereo models are solely
based on spatial considerations while leaving out the temporal
dimension as irrelevant. Specifically, most models (Fleet et al.
1996; Nomura et al. 1990; Ohzawa et al. 1990; Qian 1994;
Sanger 1988; Zhu and Qian 1996) only consider how the
spatial RFs of binocular cells may respond to *static* stimuli and
generate the physiologically observed disparity tuning curves,
such as the tuned, near, and far types found in V1 (Poggio and
Fischer 1977; Poggio et al. 1988). However, the spatial and
temporal response properties always come together for real
neurons. More importantly, physiological studies of disparity

tuning often use time-varying stimuli such as motion-in-depth
patterns, drifting gratings, moving bars, moving random-dot
stereograms, or dynamic random-dot stereograms in addition
to static images. To fully understand these data, the temporal
response properties of cortical cells must be considered.

There is also a functional reason to include time into stereo
modeling: consistent with the physiological finding that many
visual cortical cells are tuned to both disparity and motion
(Bradley et al. 1995; Maunsell and Van Essen 1983; Ohzawa et
al. 1996), there is increasing psychophysical evidence indicat-
ing that motion and stereo interact with each other in generat-
ing our perception (Anstis and Hassis 1974; Nawrot and Blake
1989; Qian et al. 1994a; Regan and Beverley 1973). We have
already proposed a model for motion-stereo integration based
on the general properties of binocular, spatiotemporal RFs of
visual cortical cells (Qian 1994; Qian and Andersen 1997; Qian
et al. 1994b). However, we did not explicitly model the dis-
parity tuning curves of cortical cells to specific time-varying
stimuli. In this paper, we first present a simple function that
conveniently describes the temporal response profiles of real
V1 cells and incorporate this function into the disparity energy
model (Ohzawa et al. 1990; Qian 1994). We then apply the
model to investigate V1 disparity responses to a variety of
time-varying stimuli used in physiological experiments. Some
of the results were reported previously in abstract form (Chen
et al. 2000).

METHODS

It is well established that the spatial RFs of V1 simple cells can be
accurately fit by Gabor functions (Daugman 1985; Jones and Palmer
1987; Marcělja 1980; Ohzawa et al. 1990). Since we are concerned
with disparity tuning instead of orientation tuning in this paper, we
only consider vertically oriented binocular cells whose left and right
RFs are given by (DeAngelis et al. 1991; Ohzawa et al. 1990, 1996)

$$g_l(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \cos(\omega_x^o x + \phi_l) \qquad (1)$$

$$g_r(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \cos(\omega_x^o x + \phi_r) \qquad (2)$$

where ω_x^o is the preferred horizontal spatial frequency, σ_x and σ_y
determine the RF dimensions along the horizontal and vertical axes,
respectively, and ϕ_l and ϕ_r are the phase parameters for the left and
right RFs, respectively. For oriented stimuli (e.g., bars and gratings),
we assume that the stimulus orientations are aligned with the cells'

preferred orientation. For moving stimuli, we assume that the direction of motion is perpendicular to the orientation of the RFs.

Unlike the spatial RFs, the temporal response of cortical cells is not Gabor-like (DeAngelis et al. 1993a, 1999; Ohzawa et al. 1996). We examined the temporal profiles of real V1 cells and found that they can be conveniently described by an envelope of the gamma probability density function, multiplied by a sinusoidal modulation

$$h(t) = \begin{cases} \dfrac{1}{\Gamma(\alpha)\tau^\alpha} t^{\alpha-1} \exp\left(-\dfrac{t}{\tau}\right) \cos(\omega_t^o t + \phi_t) & t \geq 0 \\ 0 & t < 0 \end{cases} \quad (3)$$

Here τ is the time constant for the envelope, α determines the degree of skewness, and $\Gamma(\alpha)$ is the standard gamma function for normalization; for simplicity, we let $\alpha = 2$ in this paper, and $\Gamma(2) = 1$. The sinusoidal term with frequency ω_t^o generates alternating on and off responses. Since for many real cells the first half cycle of the temporal response is shorter by various amounts than the second half cycle, the parameter ϕ_t is introduced to reduce the length of the first half cycle. (Due to the rapid decay of the exponential, the durations of the 3rd and later half-cycles are not important.) The ϕ_t parameter also determines whether the initial response is on or off. Although previously proposed functions can fit the real temporal responses just as well (Adelson and Bergen 1985; DeAngelis et al. 1999; Watson and Ahumada 1985), we prefer *Eq. 3* because all parameters have simple, intuitive meanings. *Equation 3* is plotted for two different sets of parameters in Fig. 1A. The two curves are representative of the real temporal responses from V1 (DeAngelis et al. 1993a; Ohzawa et al. 1996).

The frequency tuning of *Eq. 3* is determined by its Fourier transform, which can be calculated analytically as

$$\mathcal{H}(\omega_t) = \frac{1}{2\tau^2}\left\{ \frac{\exp(i\phi_t)}{[1/\tau + i(\omega_t - \omega_t^o)]^2} + \frac{\exp(-i\phi_t)}{[1/\tau + i(\omega_t + \omega_t^o)]^2} \right\} \quad i = \sqrt{-1} \quad (4)$$

for $\alpha = 2$. Note that because of ϕ_t, ω_t^o may not be close to the preferred temporal frequency of the function. The amplitude spectra for the temporal responses in Fig. 1A are plotted in B, showing band-pass and low-pass characteristics, respectively. These two types of frequency tuning behavior correspond to transient and sustained responses, respectively (Hawken et al. 1996)

The temporal function $h(t)$ can then be combined with the spatial function $g(x, y)$ to model three-dimensional spatiotemporal RFs of simple cells (Adelson and Bergen 1985; Watson and Ahumada 1985). For binocular simple cells, this can be done for the left and right RFs separately

$$f_l(x, y, t) = g_l(x, y)h(t) + \eta \bar{g}_l(x, y)\bar{h}(t) \quad (5)$$

$$f_r(x, y, t) = g_r(x, y)h(t) + \eta \bar{g}_r(x, y)\bar{h}(t) \quad (6)$$

where $\bar{g}$ and $\bar{h}$ functions are obtained from the corresponding g and h functions by replacing all the cosine terms by the sine terms. The constant weighting factor η, between 0 and 1, is introduced to model various degrees of directional sensitivity (Adelson and Bergen 1985; Watson and Ahumada 1985).

The response of simple cells to a stereo image pair $I_l(x, y, t)$ and $I_r(x, y, t)$ can be approximated by linear spatiotemporal filtering

(DeAngelis et al. 1993b; Jones and Palmer 1987; Ohzawa et al. 1990), followed by half-squaring (Anzai et al. 1999a,b; Heeger 1992)

$$r_s(t) = \Theta\left[\int\int\int_{-\infty}^{+\infty} \{f_l(x, y, t - t')I_l(x, y, t') + f_r(x, y, t - t')I_r(x, y, t')\}dxdydt' \right] \quad (7)$$

where the half squaring operation is defined as

$$\Theta[X] = \begin{cases} X^2 & X \geq 0 \\ 0 & X < 0 \end{cases} \quad (8)$$

For some simulations, we also included a threshold to be subtracted from the integral in *Eq. 7* before half-squaring. These will be mentioned specifically in RESULTS. The threshold tends to make tuning curves sharper by removing small responses.

Under the assumption that the RF size is much larger than the horizontal disparity D of the stimulus, it can be shown that the simple cell response is approximately (see APPENDIX)

$$r_s(t) \approx \Theta\left[2B(t)\cos\left(\theta(t) + \frac{\phi_+ - \omega_x^o D}{2}\right)\cos\left(\frac{\phi_- + \omega_x^o D}{2}\right) \right] \quad (9)$$

where

$$\phi_+ \equiv \phi_l + \phi_r, \quad \phi_- \equiv \phi_l - \phi_r \quad (10)$$

and $B(t)$ and $\theta(t)$ (defined in APPENDIX) are independent of ϕ_l, ϕ_r and D. *Equation 9* is a generalization to our previous results obtained with spatial RFs only (Qian 1994; Qian and Zhu 1997). It indicates that in addition to stimulus disparity, simple cells are also sensitive to $\theta(t)$, which depends on the spatiotemporal details (or Fourier phase) of the stimulus.

We model complex cell responses using the well-known quadrature pair method for disparity energy computation (Adelson and Bergen 1985; Emerson et al. 1992; Ohzawa et al. 1990; Pollen 1981; Qian 1994; Watson and Ahumada 1985). The complex cells derive both their spatial and temporal properties from the constituent simple cells. Because of the half-wave rectification contained in the half-squaring operation for each complex cell, we need to sum the responses of four simple cells (Ohzawa et al. 1990), all with identical ϕ_- but with their $\phi_+/2$ differing in steps of $\pi/2$. (This is exactly equivalent to summing the squared responses of two simple cells without the half squaring.) The resulting complex cell response is approximately

$$r_q(t) \approx \left[2B(t)\cos\left(\frac{\phi_- + \omega_x^o D}{2}\right) \right]^2 \quad (11)$$

which has more reliable disparity tuning because it is no longer a function of $\theta(t)$. The preferred disparity of the cell is thus

$$D_{pref} \approx -\frac{\phi_-}{\omega_x^o} \quad (12)$$

which is same as for the static case (Qian 1994).

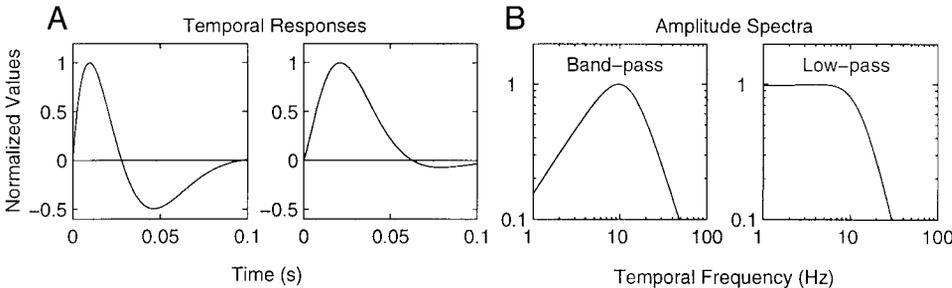Previously, we pointed out that for both physiological and compu-

FIG. 1. *A*: temporal responses of *Eq. 3* plotted for two sets of parameters. The positive and negative values represent on and off responses, respectively. For both curves, $\omega_t^o/2\pi = 7.2$ Hz and $\tau = 0.016$ s, but $\phi_t = 0.1\pi$, and -0.4π, respectively. *B*: the corresponding Fourier amplitude spectra on a log-log scale showing the band-pass and low-pass behavior, respectively. These temporal response profiles and amplitude spectra closely resemble those of real V1 cells.

tational reasons, a spatial pooling step should be added after the quadrature-pair construction to better simulate complex cell responses (Qian and Zhu 1997; Zhu and Qian 1996). We add this step for modeling complex cell responses to the random-dot type of stimuli, as such pooling significantly improves the reliability of disparity tuning (Fleet et al. 1996; Qian and Zhu 1997; Zhu and Qian 1996). The pooling step is omitted for bar and grating stimuli because it does not make any difference for those stimuli. The weighting function for the spatial pooling is a normalized, circularly symmetric two-dimensional Gaussian with a σ equal to σ_x in *Eqs. 1* and *2*.

RESULTS

Binocular interaction RFs of complex cells

Equations 5 and *6* can be used to model simple cells' binocular, spatiotemporal RFs (results not shown), which are first-order kernels of the white noise analysis (Adelson and Bergen 1985; Anzai et al. 1999a; DeAngelis et al. 1999; Ohzawa et al. 1996). One cannot obtain similar first-order RFs for complex cells because complex cells do not have separated on and off subregions. However, as Ohzawa, DeAngelis, and Freeman (1997) have shown, real complex cells have well-defined binocular interaction RFs, which are the impulse response functions obtained by flashing a line at the preferred orientation at time t to locations x_l and x_r in the two eyes, respectively. It is a first-order temporal and second-order spatial kernel. Previously, Ohzawa et al. (1997) have modeled the second-order spatial kernel. Here we add the time variable and compare our simulations with the experimental data.

It can be shown that the binocular interaction RF defined by Ohzawa et al. (1997) for a complex cell can be written as (see APPENDIX)

$$F_c(D, t) = S(D)H(t) \tag{13}$$

where

$$S(D) = \frac{4}{\sqrt{\pi}\,\sigma_x} \exp\left(-\frac{D^2}{4\sigma_x^2}\right) \cos(\omega_x^o D + \phi_-) \tag{14}$$

$$H(t) = h^2(t) + \eta^2 \bar{h}^2(t) \tag{15}$$

Remarkably, *Eq. 13* is separable in disparity and time regardless of whether the underlying simple cells for the complex cell are spatiotemporally separable or not (i.e., $\eta = 0$ or not). This is true so long as the simple cells are described by *Eqs. 5* and

6 and therefore have the matched degrees of spatiotemporal orientation in the two eyes (Ohzawa et al. 1996). Also note that $S(D)$ is a Gabor function of disparity D (Zhu and Qian 1996) and that unlike the temporal response $h(t)$ for the constituent simple cells, the temporal response $H(t)$ of the complex cell's binocular interaction RF is always positive, indicating that the Gabor disparity tuning of complex cells do not vary over time. These features are consistent with experimental data (Ohzawa et al. 1997).

Equation 13 is plotted in Fig. 2 for four model complex cells. The time-integrated tuning curves are also shown at the bottom of each panel, indicating that these cells are tuned-excitatory (TE), tuned-inhibitory (TI), near (NE), and far (FA) types, respectively, according to Poggio's classification. The disparity-time separability in *Eq. 13* is clearly exhibited in the figure for both the nondirectional cell ($\eta = 0$, Fig. 2A) and the strongly directional cell ($\eta = 1$, Fig. 2B).

Another feature in Fig. 2 is that the $D - T$ profiles of nondirectional or weakly directional complex cells (Fig. 2, *A* and *C*) have two peaks along the time axis, while strongly directional complex cells (Fig. 2, *B* and *D*) are unimodal over time. This originates from *Eq. 15*. When the directional factor $\eta = 0$, the complex cell temporal response function becomes

$$H(t) = \left[\frac{t}{\tau^2} \exp\left(-\frac{t}{\tau}\right) \cos(\omega_t^o t + \phi_t)\right]^2 \tag{16}$$

which can show multiple peaks in time because of the cosine term. On the other hand, when the direction factor $\eta = 1$, we have

$$H(t) = \left[\frac{t}{\tau^2} \exp\left(-\frac{t}{\tau}\right)\right]^2 \tag{17}$$

which can only have one peak. This relationship between directionality and the peak number along the time dimension in $D - T$ plots is a testable prediction.

Motion in depth

When an object is moving toward or away from an observer, the binocular disparity of the object changes over time, and the motion speeds or directions in the two eyes are different. The fact that the disparity tuning of complex cells does not vary with time (Fig. 2) implies that these cells are not tuned to
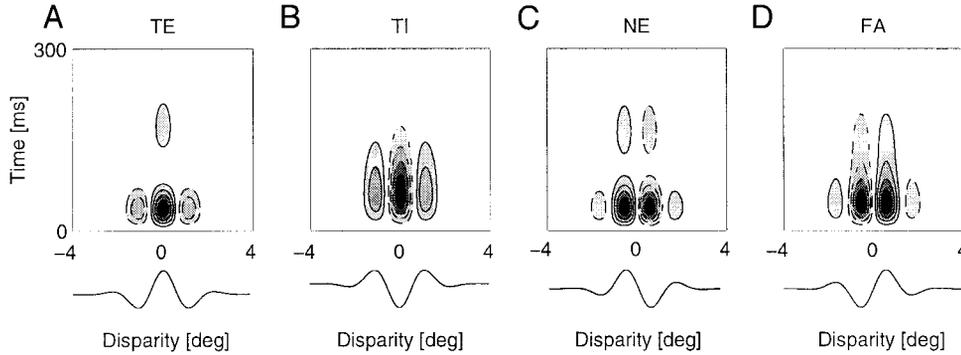

FIG. 2. Binocular interaction RFs (or $D - T$ profiles) of 4 model complex cells plotted according to *Eq. 13*. The solid and dashed contours represent the positive and negative values, respectively. *Below* each panel is the disparity tuning curve generated by integrating the $D - T$ profile along the time axis. These complex cells are constructed from simple cell RFs all with $\omega_x^o/2\pi = 0.4$ cycles/deg, $\omega_t^o/2\pi = 2$ Hz, $\sigma_x = 0.8°$, $\sigma_y = 1.2°$, $\tau = 60$ ms, and $\phi_t = 0.1\pi$. The ϕ_- and η parameters are A: 0, 0; B: $-\pi$, 1; C: $-\pi/2$, 0.3; D: $\pi/2$, 0.6, respectively. Therefore *A* is a tuned-excitatory (TE) and nondirectional complex cell; *B* is a tuned-inhibitory (TI) and strongly directional complex cell; *C* and *D* are near (NE) and far (FA) complex cells, respectively, with intermediate degrees of directional selectivity.

motion in depth (Ohzawa et al. 1997; Qian 1994; Qian and Andersen 1997). Consistent with this, most V1 cells have the same motion preference for the two eyes, and give the strongest response to the frontoparallel motion at the preferred disparity (Ohzawa et al. 1996, 1997; Poggio and Talbot 1981). In addition, Maunsell and Van Essen (1983) reported that no MT (V5) cells were found to be truly tuned for motion in depth when the motion trajectories of the stimuli were properly positioned (see following text).

We have simulated motion-in-depth tuning curves under a variety of conditions (Figs. 3–5). The format of each plot in each figure is identical to that used by Maunsell and Van Essen (1983). Twelve motion trajectories, represented "around the clock," were considered for each tuning curve. The 0 and 180° paths represent the rightward and leftward motions, respectively, in a frontoparallel plane; the 90 and 270° represent motions straight away from and toward the observer, respectively. The remaining eight trajectories represent intermediate, oblique paths in depth. Maunsell and Van Essen (1983) pointed out that to properly assess the motion-in-depth tuning, the mid-points of all trajectories should meet at a point with the preferred disparity of the cell. In this case, the 0 and 180° trajectories are on the cell's preferred disparity plane if it exists.

The 12 trajectories for the moving stimuli are specified by the horizontal speeds for the two eyes (Maunsell and Van Essen 1983). Starting from the 0° path and going counterclockwise, the 12 speed pairs for the left and right eyes used in our simulations are (1.8, 1.8), (0.6, 1.8), (−0.6, 1.8), (−1.8, 1.8),

(−1.8, 0.6), (−1.8, −0.6), (−1.8, −1.8), (−0.6, −1.8), (0.6, −1.8), (1.8, −1.8), (−1.8, −0.6), and (1.8, 0.6), in deg/s.

MOVING BARS. Figure 3 shows the results for a directional simple cell (*A*) and the corresponding complex cell (*B*) in response to a moving bar stimulus. The two rows are for the cases with and without a threshold term in *Eq. 7*, respectively. Since both the left and right RFs of the model cells prefer leftward motion, it is not surprising that the tuning curves are peaked in the left, frontoparallel direction, indicating that these cells are not tuned to motion in depth. We have also performed simulations with nondirectional model cells (results not shown). In this case, the tuning curves usually had two peaks pointing at 0 and 180° directions, and for simple cells, there were additional, smaller peaks at 90 and 270° directions, again indicating the absence of motion-in-depth tuning. These results are consistent with the physiological data for the majority of visual cortical cells (Maunsell and Van Essen 1983; Poggio and Talbot 1981). The inclusion of a threshold term (*2nd row*) makes the tuning curves sharper because it suppresses small responses from the nonpreferred paths. This could explain some sharp tuning curves found experimentally (Maunsell and Van Essen 1983; Poggio and Talbot 1981).

Although most cortical cells are like those shown in Fig. 3, preferring frontoparallel motion with fixed disparity, there is evidence that some cells in areas V1 and V2 are tuned to motion toward or away from the observer (Cynader and Regan 1978; Poggio and Talbot 1981). However, cells preferring frontoparallel motion may appear to be tuned to motion in depth if the mid-points of the stimulus trajectories meet at a point outside the preferred disparity plane (Maunsell and Van Essen 1983). Under this condition, the 0 and 180° trajectories are not in the cell's preferred disparity plane and thus may not evoke the strongest responses. By contrast, the cell may be most excited by the oblique depth-path that happens to have the best overlap with the preferred disparity plane. The tuning curves under this "off-preferred-plane" situation for the same simple and complex cells in Fig. 3 are shown in the *top row* of Fig. 4. Here, the mid-points of all paths meet at a point with a disparity of −0.04° while the cells' preferred disparity is 0.04°. As predicted by Maunsell and Van Essen (1983), now the cells appear to prefer motion along oblique paths in depths. Thus some cells may appear tuned to motion in depth simply because of the improper choice of the test paths in an experiment. However, this possibility does not rule out the existence of cortical cells that are truly tuned to motion in depth. These cells should have different preferred directions or speeds in the two eyes (Cynader and Regan 1978; Poggio and Talbot 1981) and can thus show motion-in-depth tuning even when the stimulus paths are properly chosen. Our simulation-results for a simple and a complex cell preferring opposite directions of motion in the two eyes are shown in the *bottom row* of Fig. 4. The cells are tuned to motion straight away from the observer. Unlike the cells in the *top row,* these true motion-in-depth cells have a single prominent peak in their tuning curves.

RANDOM-DOT STEREOGRAMS. We have also simulated motion-in-depth tuning curves of the same simple and complex cells in Fig. 3 (with threshold) to coherently moving random-dot stereograms (MRDSs), and dynamic random-dot stereograms (DRDSs), and examined the effect of spatial pooling (see METHODS) for the complex cell responses. The dots of a MRDS are all on the same disparity plane at a given time and the
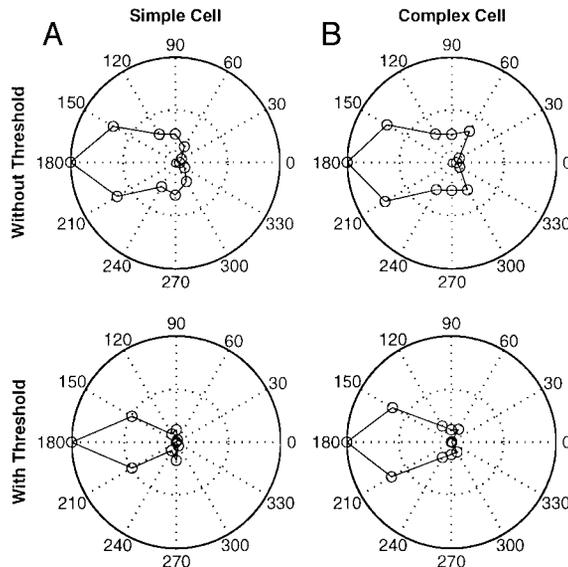

FIG. 3. Motion-in-depth tuning curves of a model simple cell (*A*) and a model complex cell (*B*) to a bar, moving along 12 paths whose mid-points coincide at a point on the cells' preferred disparity plane. The *two rows* in *A* and *B* are for the cases with and without threshold, respectively. The threshold is equal to 20% of the maximum response of the linear filtering in *Eq. 7*. The RF parameters of the simple cell (*A*) are $\omega_x^o/2\pi = 4$ cycles/deg, $\omega_t^o = 6$ Hz, $\sigma_x = 0.1°$, $\sigma_y = 0.2°$, $\tau = 20$ ms, $\phi_l = 0°$, $\phi_r = 60°$, $\phi_t = 0.1\pi$, and $\eta = 0.6$. The complex cell (*B*) receives inputs from the simple cell and 3 other simple cells according to the quadrature method. The bar size and duration are $0.1 \times 1°$ and 0.33 s, respectively. The integrated responses over the 0.33 s period are plotted. The cells have a preferred disparity of 0.04° and a preferred speed of 1.8°/s. (Note that ω_t^o/ω_x^o is not close to the preferred speed because ω_t^o is not close to the preferred temporal frequency of the cell.) The RFs are computed in a three-dimensional region of $0.5° \times 1° \times 0.1$ s. The spatial and temporal sampling steps used in the simulations are 0.01° and 5 ms, respectively.
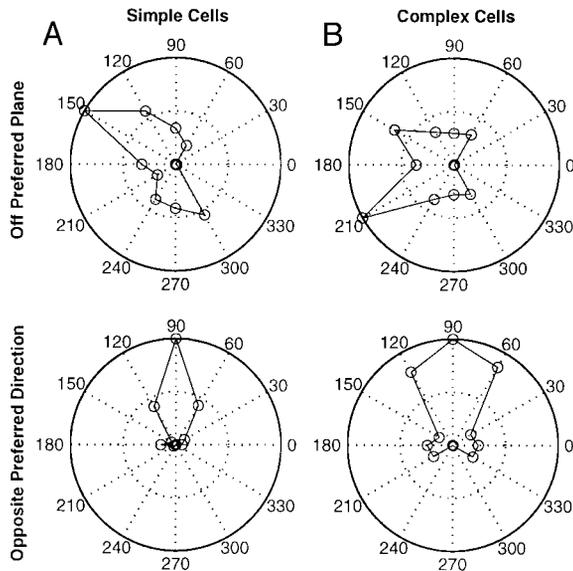
FIG. 4. Two ways of having tuning peaks away from frontoparallel planes. *Top*: the simple (*A*) and complex (*B*) cells are identical to those in Fig. 3 (with threshold). Although they actually prefer frontoparallel motion, they appear tuned to motion in depth here because the mid-points of the stimulus paths meet at a point with disparity −0.04° instead of the cells' preferred disparity 0.04°. *Bottom*: on the other hand, these cells are truly tuned to motion in depth because the directional preferences of left and right RFs are opposite. The parameters are identical to those for Fig. 3 except that $\omega_t^o/2\pi$ for the right RF has been changed from 6 to −6 Hz to generate opposite directional preference.

whole plane moves along each of the 12 motion paths mentioned in the preceding text. Each MRDS is large enough so that it covers the cells' RFs at all times without the edge effect. A DRDS is identical to the corresponding MRDS in terms of disparity change over time, but the dot positions are randomly replotted for each frame. To investigate the reliability of the tuning curves, we simulated two tuning curves for each case, with two sets of independently generated MRDSs or DRDSs. The results are shown in Fig. 5. It can be seen that the tuning for MRDSs is very similar to that for moving bars (Fig. 3), except that the curves are narrower because there are more weak responses for MRDSs than for moving bars that are suppressed by the threshold. The curves for DRDSs, on the other hand, are quite different. First, because DRDSs, by

definition, can only have disparity changes over time, but no directions of motion, the tuning curves are symmetrical with respect to the 90–270° axis. This is independent of the direction selectivity of the cell. Second, the two curves from the two independent simulations are very different from each other for the simple cell but are quite similar to each other for the complex cell with spatial pooling. This indicates that complex cells have more reliable tuning to DRDSs than do simple cells. Finally, the tuning curves for DRDSs are not as narrow as those for moving bars or MRDSs. For the simple cell, the main peak location is often located outside the preferred disparity plane. These specific features of motion-in-depth tuning to MRDSs and DRDSs can be tested experimentally, and have implications for some relevant psychophysical observations (see DISCUSSION).

Similar to Fig. 4 for the bar stimuli, MRDSs and DRDSs can also give false motion-in-depth tuning if the motion paths are not properly chosen, and real motion-in-depth tuning can only be obtained with cells preferring opposite directions in the two eyes.

Disparity tuning curves

DRIFTING SINUSOIDAL GRATINGS AND BARS. Unlike the motion-in-depth stimuli discussed in the preceding text, all stimuli in this and subsequent subsections have a constant disparity over time. Ohzawa and Freeman (1986a,b) used binocular drifting sinusoidal gratings to test the disparity tuning of V1 cells in the cat. Figure 6 shows the response time courses and disparity tuning curves of a model simple and complex cell stimulated by drifting sinusoidal gratings of various interocular phase differences. The parameters are chosen to simulate the data shown in Fig. 3 of Ohzawa and Freeman (1986b) for the simple cell, and Fig. 1 of Ohzawa and Freeman (1986a) for the complex cell. Since that particular simple cell had shorter active half-cycles than the silent half-cycles, we include a threshold equal to 20% of the maximum value of the linear-filtering-result in *Eq. 7*. The spatial and temporal frequencies of gratings match the preferred frequencies of the cells, as in the actual experiments. Ohzawa and Freeman (1986b) used the first harmonic amplitude of the simple cell response for plotting the tuning curve. We simply use the time-integrated total
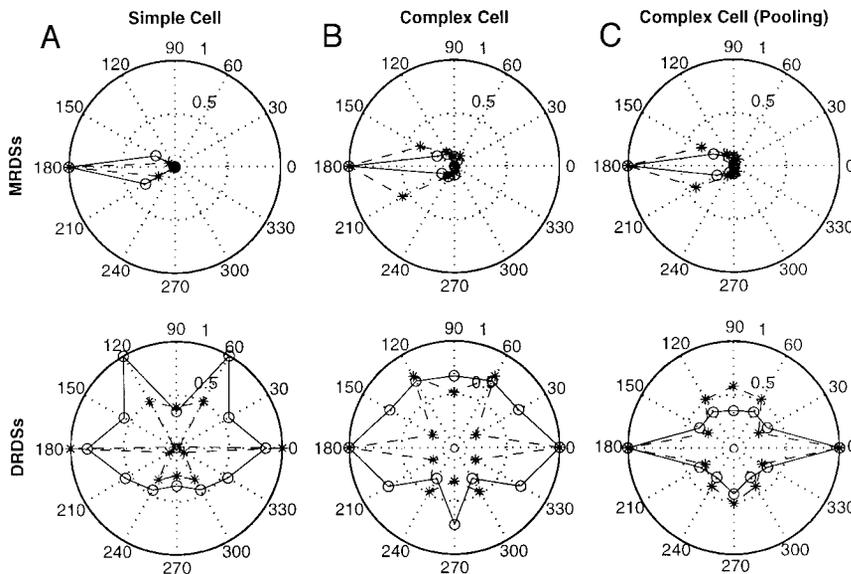

FIG. 5. Motion-in-depth tuning curves of a model simple cell (*A*) and a model complex cell without (*B*) and with (*C*) spatial pooling to moving and dynamic random-dot stereograms (MRDSs and DRDSs, respectively), with paths centered on a point at the cells' preferred disparity plane. The cell parameters are identical to those in Fig. 4 except that a spatial pooling step was added in *C*. The pooling function is a normalized, symmetric 2-dimensional Gaussian with a σ of 0.1°. Two curves shown in each panel (○ and *) are obtained with 2 independently generated sets of stimuli. The dot size is 0.02 × 0.02° and dot density is 10%. The overall size, refresh rate, and duration of each stimuli are 0.5 × 1°, 50 Hz, and 0.5 s, respectively.
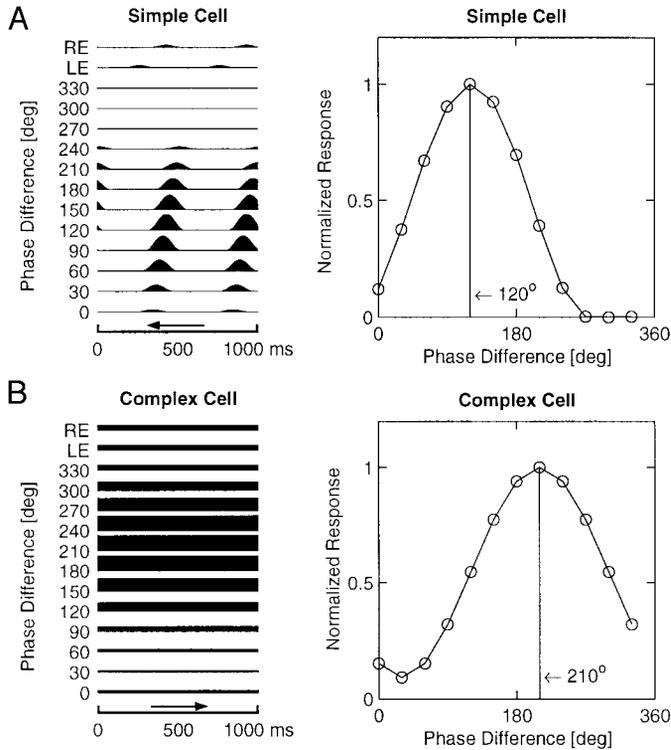
FIG. 6. Response time courses and disparity tuning curves of a model simple cell (*A*) and a model complex cell (*B*) stimulated by drifting sinusoidal gratings. *Left*: the response time courses as the interocular phase difference of the grating varied from 0 to 330° in 30° steps. The initial 0.3 s of transient responses has been excluded to show the steady-state behavior. The left and right monocular responses (LE and RE) of the cells are also shown. *Right*: the disparity tuning curves created by integrating the responses over a 1-s period. The vertical lines indicate the predicted preferred disparities according to *Eq. 12*. The simple cell (*A*) has spatiotemporally inseparable binocular RFs, with $\omega_x^o/2\pi$ = 0.3 cyc/deg, $\omega_t^o/2\pi$ = 2 Hz, σ_x = 1°, σ_y = 1.6°, τ = 60 ms, ϕ_l = 0°, ϕ_r = −120°, ϕ_t = 0.1π, and η = 0.6. The RFs are computed in a 3-dimensional region of 5° × 8° × 0.3 s. The threshold value is equal to 20% of the maximum linear filtering response of the simple cell. The RF parameters of the complex cell (*B*) are $\omega_x^o/2\pi$ = 0.4 cycles/deg, $\omega_t^o/2\pi$ = −2 Hz, σ_x = 0.8°, σ_y = 1.2°, τ = 60 ms, ϕ_- = 210°, ϕ_t = −0.1π, and η = 0.6. The RFs are computed over a region of 4° × 6° × 0.3 s. The spatial and temporal frequencies of the gratings match the preferred spatial and temporal frequencies of the cells. The initial phase of the right image is fixed at 60° for both cells and that of the left image is varied from 60° to 390° in steps of 30°. The spatial and temporal sampling intervals for the simulations are 0.1° and 10 ms, respectively.

response because it is proportional to the first harmonic in the context of our model. Figure 6 shows that the responses of both the simple and complex cells depend on the interocular phase difference (proportional to disparity) of the gratings. The simple cell's responses are modulated sinusoidally in time followed by rectification, while the complex cell responses are sustained. These features agree with the experimental data (Ohzawa and Freeman 1986a,b).

Another feature in Fig. 6*A* is that the temporal responses of the simple cell are tilted to the right as the interocular phase difference increases. This is also consistent with the physiological results in Fig. 3 of Ohzawa and Freeman (1986b). It can be shown that this tilt stems from the specific way of introducing binocular disparity. In both the experiments (Ohzawa and Freeman 1986a,b) and our simulations, the disparity is generated by keeping the grating phase of one eye's image fixed while varying the phase in the other eye. If the disparity is symmetrically divided between the two eyes, then the tilt disappears (results not shown). The reason is that the asym-

metric disparity generates a small positional change that leads to a temporal delay in the simple cell's response.

The model cells used in the preceding simulations are ocularly balanced. However, similar results can be obtained when one eye is more dominant than the other. There are two ways to introduce ocular dominance into the model. The first method is to introduce a weighting factor in front of one of the two RF profiles in *Eq. 7*. Mathematically, this is equivalent to presenting a stereogram with different contrast scales (but of the same contrast sign) to the two eyes. As we have shown previously (Qian 1994; Qian and Mikaelian 2000), the tuning curves will maintain the same shape under this condition although the pedestal will be higher and the amplitude will be smaller. The second method for introducing ocular dominance is to assume that one eye has a higher response threshold than the other. We find through simulations that again similar tuning curves can be obtained unless one of the thresholds is so high that the corresponding eye does not respond (results not shown).

We have also simulated response time courses and disparity tuning curves of simple and complex cells to moving bars (results not shown). Like the grating case, the tuning curves for both simple and complex cells peak at locations predicted by *Eq. 12*, and the vertical alignment of the response time courses depends on whether the disparities are introduced symmetrically in the two eyes or not. For directional cells, the disparity tuning curves for the preferred and anti-preferred directions have the same peak locations although the responses amplitudes differ markedly. These features are consistent with the experimental data in Fig. 4 of Poggio and Fischer (1977). For each bar sweep, the complex cells give longer responses than the corresponding simple cells because the former do not have the discrete on and off RF subregions.

RANDOM-DOT STEREOGRAMS. Poggio et al. (1985, 1988) also applied DRDSs to measure disparity tuning curves. In their experiments, each stereogram maintained a constant disparity during a trial, but the actual dot locations were randomly re-plotted from frame to frame. They found that simple cells do not show reliable disparity tuning to DRDSs but that complex cells do.

To investigate how reliably our model simple and complex cells were disparity-tuned to DRDSs, we computed, for each cell type, 1,000 disparity tuning curves from 1,000 independent sets of DRDSs, all generated from the same parameters. All DRDSs had a refresh rate of 100 Hz as in Poggio et al.'s experiments. Figure 7 shows the results. We also considered the effect of adding a spatial pooling stage to the complex cell responses (Fig. 7*C*, see METHODS). For clarity, only 30 randomly picked curves for each cell are shown in the *top panels*. The distribution histograms of the preferred disparities (*bottom panels*) are compiled from all 1,000 curves. It is clear from the figure that the peak location of the tuning curves is much more variable for the simple cells than for the complex cells and that spatial pooling helps to further improve the reliability of the complex cell responses. Specifically, 40, 77, and 99% of the tuning curves peak within 0.02° of the predicted preferred disparity for the simple cell, the complex cell without pooling, and the complex cell with pooling, respectively. Additional simulations show that for complex cells, the standard deviation of the peak locations is inversely proportional to the σ of the two-dimensional Gaussian used for the spatial pooling. Since the number of cells (*N*) pooled is proportional to σ^2, the
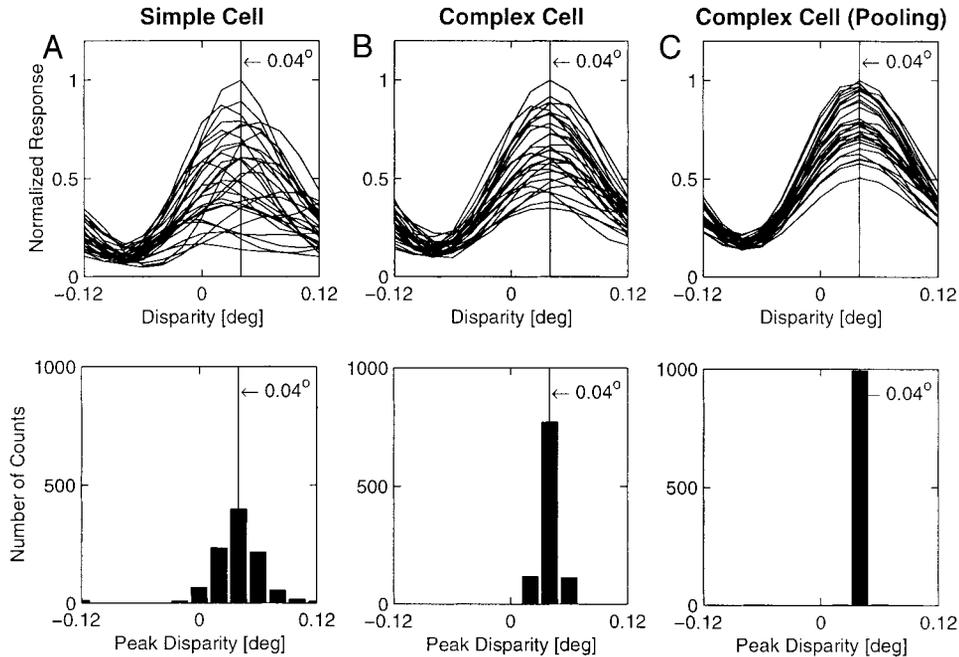
FIG. 7. Disparity tuning curves of a model simple cell (*A*), and a model complex cell without (*B*) and with (*C*) spatial pooling, in response to DRDSs. *Top*: 30 disparity tuning curves obtained from 30 independent DRDSs. Each point on a curve was obtained by integrating the response over a period of 500 ms. The curves in a panel are normalized by the strongest response. *Bottom*: the distribution histograms of the peak locations, each compiled from 1,000 disparity tuning curves. The bin size of the histograms is 0.02°. The vertical lines indicate the predicted preferred disparities according to *Eq. 12*. The RF parameters of the simple cell (*A*) are $\omega_x^o/2\pi = 4$ cycles/deg, $\omega_t^o/2\pi = 6$ Hz, $\sigma_x = 0.1°$, $\sigma_y = 0.2°$, $\tau_y = 0.2°$, $\tau = 20$ ms, $\phi_l = \phi_r = 60°$, $\phi_t = 0.1\pi$, and $\eta = 0.6$. The RFs are computed in a 3-dimensional region of $0.5° \times 1° \times 0.1$ s. The complex cell (*B*) receives inputs from the simple cell and 3 other simple cells according to the quadrature method. *C*: the spatial pooling procedure (see METHODS) is added to the complex cell in *B*. The pooling function is a normalized, symmetric 2-dimensional Gaussian with a σ of 0.1°. The dot size is $0.02 \times 0.02°$ and dot density was 10%. The overall size, refresh rate, and duration of the stimuli are $1° \times 1.2°$, 100 Hz, and 0.5 s, respectively. The spatial and temporal sampling steps for these simulations are 0.01° and 5 ms, respectively.

variability of the peak locations follows the inverse $\sqrt{N}$ law, as expected. However, the improvement from the simple cell to the complex cell (without pooling) is about twice that expected from the inverse $\sqrt{N}$ law because the four simple cells in the quadrature method are specifically picked to reduce variability.

Our simulation result, that disparity tuning curves to DRDSs are more reliable in complex cells than in simple cells, is in qualitative agreement with the experimental data of Poggio and coworkers (Poggio et al. 1985, 1988). Quantitatively, however, there may be some discrepancies. Although they did not publish any simple cell tuning curves to DRDSs, Poggio et al. (1985, 1988) reported that nearly all neurons responding to DRDSs are complex cells and that simple cells are not tuned to these stimuli. In contrast, the simulated tuning curves in Fig. 7*A* are not completely random but show a tendency to peak around the preferred disparity of the corresponding complex cell (marked by the vertical line in the figure). A close examination reveals that the disparity tuning trend of the model simple cell results from the fact that a small number of frames in each DRDS generate relatively reliable tuning because they happen to contain dot distributions that excite the cell strongly.

A closely related problem in Fig. 7*A* is that the response amplitudes of the simple cell to different sets of DRDSs fluctuated over a very large range (because some DRDSs happen to contain more frames that strongly excite the cell than other DRDSs). However, experimental data show that although some V1 cells occasionally give a strong response to one random-dot pattern and a weak response to another pattern, most cells have comparable responses to different random dot

stimuli (Qian and Andersen 1995; Skottun et al. 1988; Snowden et al. 1992).

The preceding two problems can be resolved by introducing the following contrast response function to replace the half-squaring operation in *Eq. 8*

$$R[X] = \begin{cases} R_{max}X^n/(X^n + X_{50}^n) & X \geq 0 \\ 0 & X < 0 \end{cases} \quad (18)$$

where R is the simple cell response, X is the result of linearly filtering a stereo stimulus through the binocular spatiotemporal RFs of the simple cell, and R_{max}, X_{50}, and n denote, respectively, the maximum response, the X at which the response reaches half its maximum value, and the exponent that determines the steepness of the function (Albrecht and Hamilton 1982; Sclar et al. 1990). It has been shown that this type of contrast response can be implemented by a normalization procedure following the half-squaring operation (Heeger 1992). Like the discharge of real simple cells, *Eq. 18* saturates at high stimulus contrast. When $n = 2$, the equation reduces to half-squaring at low stimulus contrast. Since this function compresses the response range, it should effectively increase the contributions to tuning curves from those frames in a DRDS that evoke relatively weak responses, and consequently reduce the tuning reliability of the model simple cells because weak responses usually generate poor tuning curves. The simulation results confirm this expectation (Fig. 8). The simple cell's disparity tuning to DRDSs became much more variable while the tuning of the complex cell remained reliable, especially with spatial pooling. These results are more consistent with
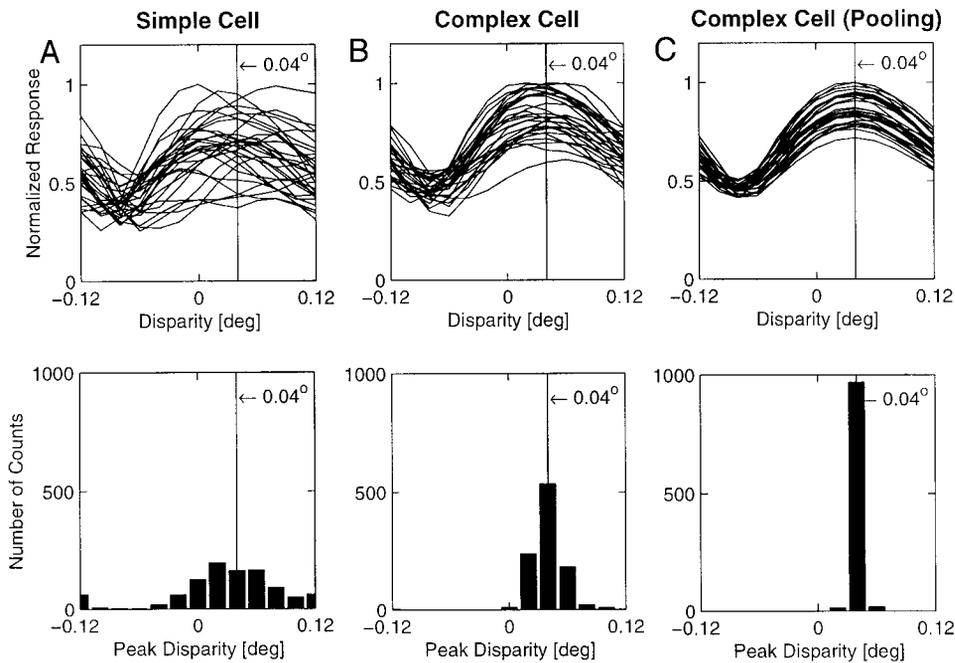
FIG. 8. Disparity tuning curves to DRDSs with contrast saturation. The simulations are identical to those in Fig. 7 except that *Eq. 8* is replaced by *Eq. 18.* The parameters of the contrast response function are $R_{max} = 1$, $X_{50} = 10$, and $n = 2$.

Poggio's experimental reports (Poggio et al. 1985, 1988) than are those in Fig. 7, although we cannot make a quantitative comparison due to the lack of published experimental data.

We next simulated the responses of the cells used for Fig. 8 to coherently MRDSs. The results are shown in Fig. 9. Obviously, the simple cell's disparity tuning to MRDSs is much more reliable than to DRDSs. The reason is that $\theta(t)$ in *Eq. 9* varies randomly over time for DRDSs, while it changes smoothly for MRDSs. Since the temporal averaging of a continuous $\theta(t)$ is much closer to a constant than is the averaging of some random values, coherently moving stereograms should always generate more reliable disparity tuning curves than the random frames unless a very large number of frames (>200) is used (in which case both types of tuning curves become reliable). This is a specific prediction that can be tested physiologically. Poggio et al. measured disparity tuning of some V1

cells to MRDSs (Poggio et al. 1985, 1988). Unfortunately, they did not systematically compare the cells' responses to DRDSs and MRDSs but instead appeared to group the two types of stereograms together as the "cyclopean stimuli."

Finally, for the purpose of comparison, we also simulated the disparity tuning of the cells in Fig. 8 to static random-dot stereograms (SRDSs). The results are shown in Fig. 10. Consistent with our previous simulations with spatial RFs only (Qian 1994; Qian and Zhu 1997; Zhu and Qian 1996), the simple cell showed completely random disparity tuning curves when different sets of SRDSs were used, while the complex cell maintained reasonable tuning reliability when the spatial pooling is applied. Moreover, for all cell types, disparity tuning to SRDSs is not as reliable as that to DRDSs, which in turn is not as reliable as the tuning to MRDSs. This is easy to understand because for static patterns there is only a single value for
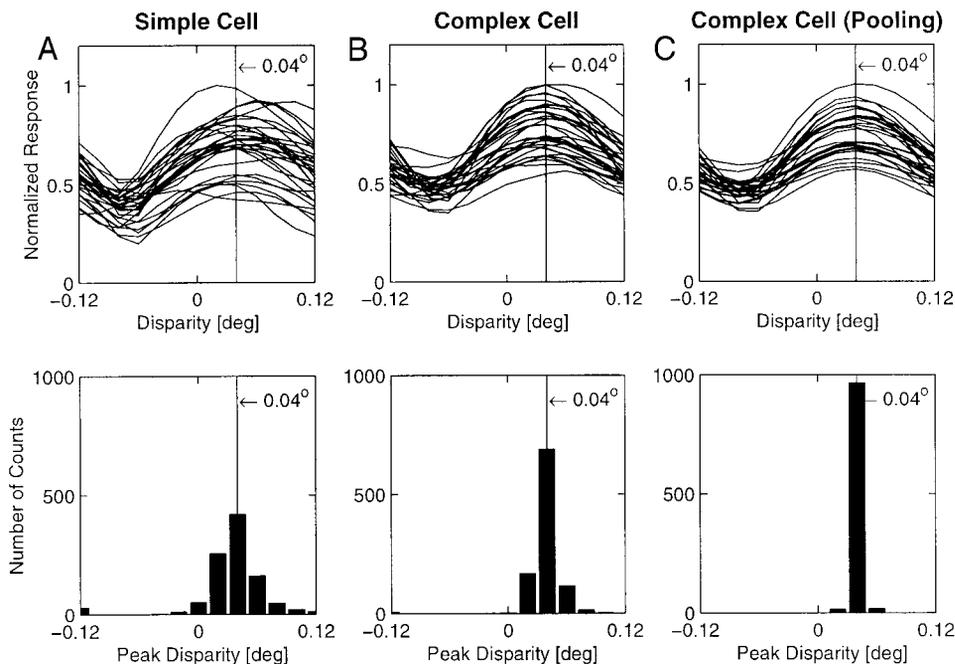

FIG. 9. Disparity tuning curves to MRDSs with contrast saturation. The simulations are identical to those in Fig. 8 except that MRDSs are used. All MRDSs move leftward at a speed of 2°/s.
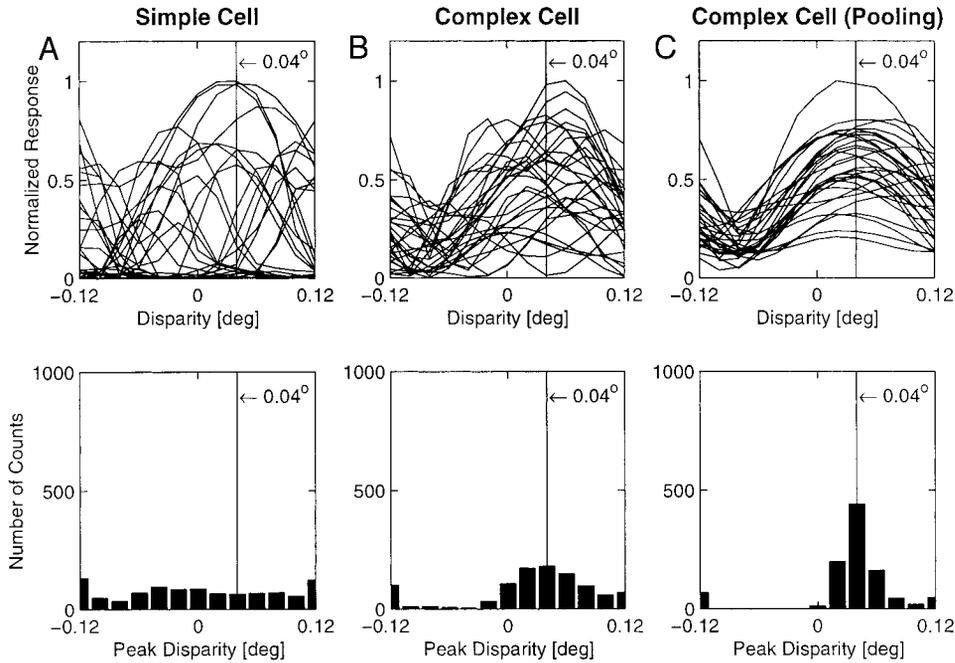
FIG. 10. Disparity tuning curves to SRDSs with contrast saturation. The simulations are identical to those in Fig. 8 except that SRDSs are used.

$\theta(t)$ in *Eq. 9,* and therefore temporal integration does not help to reduce the influence of the first cosine term in the equation.

DISCUSSION

The main goal of this paper is to understand how V1 cells respond to binocular disparity in time-varying stimuli. We introduced a specific function that conveniently describes temporal response profiles of real cortical cells including the transient (or band-pass) and the sustained (low-pass) types. We then incorporated this temporal function into the disparity energy model (Ohzawa et al. 1990; Qian 1994) and found that the binocular interaction RFs of V1 complex cells, with the typical disparity-time separability in the $D - T$ plot (Ohzawa et al. 1997), can be explained. The disparity part is a Gabor function and the time part is always positive. Finally, we investigated how the model simple and complex cells respond to various time-varying stimuli, including motion-in-depth patterns, drifting gratings, moving bars, MRDSs and DRDSs. We found that the simulated tuning curves agree with the extant experimental data quite well (Cynader and Regan 1978; Ohzawa and Freeman 1986a,b; Poggio and Fischer 1977; Poggio and Talbot 1981; Poggio et al. 1985). Our results indicate that both spatial pooling and temporal averaging can significantly improve the reliability of disparity tuning and that in general, complex cells are much better disparity detectors than simple cells (Ohzawa et al. 1990; Qian 1994), although the difference between the two cell types depends on the stimuli (see following text).

Tuning reliability

We pointed out previously that for static stereograms, simple cells do not have reliable disparity tuning since their responses are highly dependent on the Fourier phases of the stimuli (Qian 1994, 1997; Qian and Zhu 1997; Zhu and Qian 1996). For example, simple cells' tuning curves vary with the spatial phase of sinusoidal gratings and with the lateral position and contrast polarity of bars (Ohzawa et al. 1990). For coherently

moving stimuli considered in this paper, this Fourier-phase dependence is manifested as the temporal modulation of the response: as a stimulus such as a bar or a grating sweeps through the RFs of a cell and its Fourier phase changes continuously, and therefore the response changes accordingly in time. If the tuning curve of a simple cell is calculated by temporally integrating the responses over time, the phase dependence will be averaged out, and simple cells will then have reliable disparity tuning curves to moving stimuli. Indeed, we found that for moving bars and gratings, simple and complex cells show equally reliable disparity tuning curves. However, the situation is quite different for DRDSs. Here the simple cells' disparity tuning is still highly unreliable even with temporal integration of 50 different frames, and this lack of reliability is consistent with the experimental reports (Poggio et al. 1985, 1988). Intuitively, a DRDS only contains random samples of the possible Fourier phase values, while for coherently moving stimuli, the Fourier phase changes smoothly so that the full range of phase values can be quickly covered for every stimulus used in an experiment. Therefore temporal integration of simple cell responses is much more effective in improving disparity tuning for coherently moving stimuli than for DRDSs. In contrast to simple cells, complex cells have reliable disparity tuning to all of the stimulus types mentioned above, including DRDSs, and this is particularly true when the spatial pooling step is included for modeling complex cell responses. The simulated reliability of complex cell tuning is consistent with experimental data (Ohzawa et al. 1990; Poggio et al. 1985, 1988). The pooling reduces variability according the expected inverse $\sqrt{N}$ law, while the quadrature-pair construction for complex cells is about twice as effective as expected from the inverse $\sqrt{N}$ law.

One might conclude, based on the preceding discussion, that simple cells can reliably extract disparity for coherently moving stimuli but not for static patterns and DRDSs, whereas complex cells can do so for all stimulus types. This conclusion requires some qualification because for simple cells, the reliable tuning to coherently moving stimuli is only obtained after

integrating the responses over a certain period of time. The brain, however, may not have the luxury of waiting for the temporal integration to complete before responding to stimuli in the real world. In fact, disparity-triggered vergence eye movement has a latency of less than 60 ms in monkeys (Masson et al. 1997), only about 10 or 20 ms longer than the V1 response latency. Therefore the brain might have to extract disparity based on the responses over a time slice of only 10 or 20 ms. If this is the case, then simple cells may not be able to extract disparity reliably even for moving stimuli. Consider, for example, the simple cell response time courses to gratings (Fig. 6A). It is clear that tuning curves calculated from different brief time slices will have different peak locations. This problem does not exist for the complex cell in Fig. 6B because its responses are more sustained in time. We conclude that in general, complex cells are better suited than simple cells for disparity extraction.

Motion in depth

We have also shown that a cell with identical motion preference for its left and right RFs is not truly tuned to motion in depth. As Maunsell and Van Essen (1983) predicted, such a cell may give a false impression of motion-in-depth tuning if the stimulus paths are not properly aligned with the preferred disparity plane. True motion-in-depth tuning, however, can only be obtained for cells with different left and right motion preferences. Our simulations may help explain some relevant psychophysical findings. Westheimer (1990) reported that with line stimuli, the threshold for detecting disparity motion in depth is much higher than that for detecting the disparity difference of frontoparallel motions. This agrees with the fact that most visual cortical cells have the same motion preference in the two eyes (Maunsell and Van Essen 1983; Ohzawa et al. 1996, 1997; Poggio and Talbot 1981) and therefore are not tuned to motion in depth. Cumming and Parker (1994) found that stereomotion is primarily detected by means of the temporal change of binocular disparity, instead of the interocular velocity difference. Again, this is consistent with physiology because cells with identical motion preference in the two eyes cannot be sensitive to the interocular velocity difference. Finally, Harris and Watamaniuk (1995) concluded that the rate of pure disparity change is not a good cue for speed discrimination of DRDSs moving in depth. This could be due to the poor reliability and broad widths of the motion-in-depth tuning curves under this condition, as shown in Fig. 5.

Alternative methods

Although we used the phase-difference RF model and the quadrature pair construction proposed by Ohzawa et al. (1990) in all analyses and simulations presented here, similar results can be obtained for the position-shift RF model and for some other methods of constructing complex cell responses. As we demonstrated previously (Zhu and Qian 1996), there is little difference in disparity tuning between the phase-difference and the position-shift RF models for the broadband stimuli such as bars and random-dot patterns when the disparity range is smaller than the preferred spatial period (the inverse of preferred spatial frequency) of the RFs. For narrowband stimuli

like sinusoidal gratings, the main difference is a small horizontal shift of disparity tuning curves. But even this difference disappears when the grating frequency matches the cell's preferred spatial frequency, which is the case for the simulations reported here. We have also shown previously that the quadrature pair construction is exactly equivalent to a phase averaging procedure that integrates the responses of all simple cells with their ϕ_+ uniformly distributed in the entire 4π range (Qian and Mikaelian 2000). We can further demonstrate that squaring in the quadrature pair method is also not important because similar results can be obtained if the exponent of 2 in *Eq. 8* is replaced by a positive number n (Albrecht and Hamilton 1982; Sclar et al. 1990), and if the phase averaging procedure is used. In this case, *Eq. 11* for complex cell response simply becomes something very similar

$$r_q(t) \approx C(n) \left| 2B(t) \cos \left(\frac{\phi_- + \omega_x^o D}{2} \right) \right|^n \qquad (19)$$

where $C(n)$ is an unimportant function of n. Our computer simulations confirmed that indeed similar disparity tuning curves can be obtained (results not shown), the only difference being that larger n tends to generate sharper tuning curves. While the energy model is computationally more compact, the variations mentioned here may be more physiologically plausible.

Predictions

Several specific, testable predictions can also be made based on our analyses and simulations. First, strongly directional complex cells should only have a single peak along the time axis in the $D - T$ plot. Nondirectional cells should have more than one peak unless their temporal frequency bandwidths are so large (i.e., small τ and ω_t^o in *Eq. 16*) such that later peaks become too small to be observed. Second, cells with higher firing thresholds should have narrower disparity tuning curves. (The high threshold can be judged by the low spontaneous rate and the shorter active half-cycles than the silent half-cycles in response to drifting sinusoidal gratings.) Moreover, for cells with high response threshold, the motion-in-depth tuning curves for MRDSs should be much narrower than those for moving bars. Third, the observed tilt with increasing disparity in the time course of simple cells' responses to drifting stimuli should disappear if the stimulus disparity is introduced symmetrically into the two eyes. Fourth, cells' disparity tuning curves to MRDSs should be more reliable than those to DRDSs, which in turn should be more reliable than those to SRDSs. Here, the reliability is defined as how reproducible the tuning curves are when independent sets of random-dot patterns (all generated from the same sets of underlying parameters) are applied to the same cell. This predicted trend should be particularly strong for simple cells, but less pronounced for complex cells which have reasonably reliable disparity tuning to all stimulus types. Finally, for drifting bars and gratings, both simple and complex cells should have reliable disparity tuning when the time-averaged responses are used, while for static patterns and for dynamic random-dot stimuli, simple cells' disparity tuning should be much less reliable than that of complex cells. Experimental tests of these predictions will help determine the adequacy of the current understanding of V1 disparity selectivity.

Problems with the disparity energy model

The disparity energy model has been highly successful in explaining a wide range of physiological and perceptual observations as demonstrated by this and numerous previous publications (Anzai et al. 1999b; Fleet et al. 1996; Mikaelian and Qian 2000; Ohzawa et al. 1990, 1997; Qian 1994; Qian and Andersen 1997; Qian and Zhu 1997; Qian et al. 1994b; Zhu and Qian 1996). This is quite remarkable given that the model is a relatively high-level abstraction that does not include detailed morphology, connectivity, and membrane biophysics of the visual cells. However, there are also some experimental findings that are inconsistent with the model. Ohzawa et al. (1997) noted that the spatial elongation of the binocular interaction RF of real complex cells is significantly larger than that predicted by a single quadrature-pair mechanism. This problem may be alleviated by adding a spatial pooling procedure for computing complex cell responses (Fleet et al. 1996; Qian and Zhu 1997; Zhu and Qian 1996), which also accounts for the larger RFs of complex cells compared with simple cells at the same eccentricity (Hubel and Wiesel 1962; Schiller et al. 1976). Another problem noted by Ohzawa et al. (1997) is that for real complex cells, the disparity frequency (obtained from the disparity tuning curves to broadband stimuli) is usually lower than the preferred spatial frequency (especially for high-frequency cells), while the energy model predicts equality of the two frequencies (Ohzawa et al. 1990; Qian 1994; Zhu and Qian 1996). However, the discrepancy may be, at least partially, due to something unrelated to the model: the disparity frequency was measured with the white-noise method while the preferred spatial frequency was measured with drifting sinusoidal gratings (Ohzawa et al. 1997). Since the spatial frequency measured with noise stimuli is lower than that measured with drifting gratings (Gaska et al. 1994), perhaps the white-noise method also underestimates the disparity frequency (Ohzawa et al. 1997). Indeed due to the time-consuming nature of the white-noise method, one might tend to chose a lower spatial sampling density for the noise stimuli than for the grating stimuli. We found through simulations that an insufficient spatial sampling density (which would be more likely to happen for cells with high spatial frequencies) can indeed lead to an underestimation of the measured disparity frequency (results not shown).

The energy model also predicts that when stimuli presented to the two eyes have opposite signs of contrast, the disparity tuning curve of a complex cell should be inverted in shape, with the same amplitude as the same-contrast-sign case (Ohzawa et al. 1990; Qian 1994; Qian and Mikaelian 2000). In reality, while many complex cells do show the predicted tuning curve inversion, the amplitude of tuning is typically reduced (Cumming and Parker 1997; Ohzawa et al. 1997). It has been suggested that an introduction of monocular thresholds at the simple cell stage may explain the reduced amplitude (Read et al. 2000). Finally, there are cells that appear monocular when the two eyes are tested separately but show a large binocular interaction (either disparity- or nondisparity-selective) when the two eyes are stimulated together (Ohzawa and Freeman 1986a,b; Poggio and Fischer 1977). This is presumably due to some subthreshold events and may be partially explained by adding a binocular threshold in *Eq. 7* after the summation of the monocular contributions. A full account, however, may require a highly nonlinear summation mechanism for combining the two monocular inputs. How to modify the energy model to resolve these and other problems without completely sacrificing its simplicity will be a challenge to future research.

APPENDIX

Derivation of Eq. 9

We derive the simple cell responses *Eq. 9* under the general assumption that the size of the RFs is much larger than the image disparity. First, rewrite g_1 and $\bar{g}_1$ in *Eq. 5* as

$$g_1(x, y) = \cos(\phi_1)g_{\cos}(x, y) - \sin(\phi_1)g_{\sin}(x, y) \tag{A1}$$

$$\bar{g}_1(x, y) = \cos(\phi_1)g_{\sin}(x, y) + \sin(\phi_1)g_{\cos}(x, y) \tag{A2}$$

where

$$g_{\cos}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \cos(\omega_x^o x) \tag{A3}$$

$$g_{\sin}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \sin(\omega_x^o x) \tag{A4}$$

Then *Eq. 5* becomes

$$f_1(x, y, t) = \cos(\phi_1)[g_{\cos}(x, y)h(t) + \eta g_{\sin}(x, y)\bar{h}(t)] - \sin(\phi_1)$$
$$\times [g_{\sin}(x, y)h(t) - \eta g_{\cos}(x, y)\bar{h}(t)] \tag{A5}$$

A binocular stimulus with a horizontal disparity D can be written as

$$I_1(x, y, t) = I(x, y, t), \quad I_r(x, y, t) = I(x + D, y, t) \tag{A6}$$

The linear filtering of the left image by the left RF in *Eq. 7* becomes

$$r_s^l(t) = \cos(\phi_1) \int\!\!\int\!\!\int_{-\infty}^{+\infty} [g_{\cos}(x, y)h(t - t') + \eta g_{\sin}(x, y)\bar{h}(t - t')]I(x, y, t')\mathrm{d}x\mathrm{d}y\mathrm{d}t'$$

$$- \sin(\phi_1) \int\!\!\int\!\!\int_{-\infty}^{+\infty} [g_{\sin}(x, y)h(t - t') - \eta g_{\cos}(x, y)\bar{h}(t - t')]I(x, y, t')\mathrm{d}x\mathrm{d}y\mathrm{d}t'$$

$$= B(t) \cos(\theta(t) + \phi_1) \tag{A7}$$

where

$$B(t) = \sqrt{B_1^2(t) + B_2^2(t)}, \quad \theta(t) = \arctan\left(\frac{B_2(t)}{B_1(t)}\right) \tag{A8}$$

$$B_1(t) = \int\!\!\int\!\!\int_{-\infty}^{+\infty} [g_{\cos}(x, y)h(t - t') + \eta g_{\sin}(x, y)\bar{h}(t - t')]I(x, y, t')\mathrm{d}x\mathrm{d}y\mathrm{d}t' \tag{A9}$$

$$B_2(t) = \int\!\!\int\!\!\int_{-\infty}^{+\infty} [g_{\sin}(x, y)h(t - t') - \eta g_{\cos}(x, y)\bar{h}(t - t')]I(x, y, t')\mathrm{d}x\mathrm{d}y\mathrm{d}t' \tag{A10}$$

The linear filtering of the right image by the right RF in *Eq. 7* is

$$r_s^r(t) = \int\!\!\int\!\!\int_{-\infty}^{+\infty} f_r(x - D, y, t - t')I(x, y, t')\mathrm{d}x\mathrm{d}y\mathrm{d}t' \tag{A11}$$

According to *Eq. 6,* we have

$$f_r(x - D, y, t) = g_r(x - D, y)h(t) + \eta \bar{g}_r(x - D, y)\bar{h}(t) \quad (A12)$$

where

$$g_r(x - D, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{(x - D)^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \cos(\omega_x^o x + \phi_r - \omega_x^o D) \quad (A13)$$

$$\bar{g}_r(x - D, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{(x - D)^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \sin(\omega_x^o x + \phi_r - \omega_x^o D) \quad (A14)$$

When the σ_x is much larger than the image disparity D, we can approximate

$$\exp\left(-\frac{(x - D)^2}{2\sigma_x^2}\right) \approx \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \quad (A15)$$

and a derivation similar to that for *Eq. A7* gives

$$r_s^r(t) \approx B(t) \cos(\theta(t) + \phi_r - \omega_x^o D) \quad (A16)$$

Finally, inserting *Eq. A7* and *Eq. A16* into *Eq. 7*, we obtain the simple cell response as

$$r_s(t) = \Theta[r_s^l(t) + r_s^r(t)]$$

$$\approx \Theta\left[2B(t) \cos\left(\theta(t) + \frac{\phi_+ - \omega_x^o D}{2}\right) \cos\left(\frac{\phi_- + \omega_x^o D}{2}\right)\right] \quad (A17)$$

which is *Eq. 9* in the text.

Derivation of Eq. 13

The binocular interaction RF for complex cells is the impulse response function obtained by flashing a line with preferred orientation (vertical in our case) at time t to locations x_l and x_r in the two eyes respectively. Because for vertical line stimuli, the Y dimension of *Eq. 7* simply integrates to a constant, we can ignore the Y dimension.

First, the response of the linear filtering of the dichoptically flashed line through the binocular simple cell RFs is given by

$$f_l(x_l, t) \pm f_r(x_r, t) = \frac{1}{\sqrt{2\pi}\sigma_x} \{[\cos(\omega_x^o x_l + \phi_l)ah(t) + \eta \sin(\omega_x^o x_l + \phi_l)a\bar{h}(t)]$$

$$\pm [\cos(\omega_x^o x_r + \phi_r)bh(t) + \eta \sin(\omega_x^o x_r + \phi_r)b\bar{h}(t)]\}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_x} \left\{\left[\cos\left(\omega_x^o x_l + \frac{\phi_+ + \phi_-}{2}\right)ah(t)\right.\right.$$

$$\left. + \eta \sin\left(\omega_x^o x_l + \frac{\phi_+ + \phi_-}{2}\right)a\bar{h}(t)\right]$$

$$\pm \left[\cos\left(\omega_x^o x_r + \frac{\phi_+ - \phi_-}{2}\right)bh(t)\right.$$

$$\left.\left. + \eta \sin\left(\omega_x^o x_r + \frac{\phi_+ - \phi_-}{2}\right)b\bar{h}(t)\right]\right\}$$

$$= C_1 \cos\left(\frac{\phi_+}{2}\right) - C_2 \sin\left(\frac{\phi_+}{2}\right) = C \cos\left(\gamma + \frac{\phi_+}{2}\right) \quad (A18)$$

where $\pm$ indicates whether the left and right eyes' lines have the same or opposite contrast signs

$$a = \exp(-x_l^2/2\sigma_x^2), \quad b = \exp(-x_r^2/2\sigma_x^2) \quad (A19)$$

$$C_1 = \frac{1}{\sqrt{2\pi}\sigma_x} \left[\left(\cos\left(\omega_x^o x_l + \frac{\phi_-}{2}\right)ah(t) + \eta \sin\left(\omega_x^o x_l + \frac{\phi_-}{2}\right)a\bar{h}(t)\right)\right.$$

$$\left. \pm \left(\cos\left(\omega_x^o x_r - \frac{\phi_-}{2}\right)bh(t) + \eta \sin\left(\omega_x^o x_r - \frac{\phi_-}{2}\right)b\bar{h}(t)\right)\right] \quad (A20)$$

$$C_2 = \frac{1}{\sqrt{2\pi}\sigma_x} \left[\left(-\sin\left(\omega_x^o x_l + \frac{\phi_-}{2}\right)ah(t) + \eta \cos\left(\omega_x^o x_l + \frac{\phi_-}{2}\right)a\bar{h}(t)\right)\right.$$

$$\left. \pm \left(-\sin\left(\omega_x^o x_r - \frac{\phi_-}{2}\right)bh(t) + \eta \cos\left(\omega_x^o x_r - \frac{\phi_-}{2}\right)b\bar{h}(t)\right)\right] \quad (A21)$$

$$C = \sqrt{C_1^2 + C_2^2} \quad (A22)$$

$$= \frac{1}{\sqrt{2\pi}\sigma_x} \sqrt{(a^2 + b^2 \pm 2ab \cos(\omega_x^o(x_l - x_r) + \phi_-))(h^2(t) + \eta^2\bar{h}^2(t))} \quad (A23)$$

$$= \sqrt{S_\pm(x_l, x_r)H(t)} \quad (A24)$$

$$\gamma = \arctan(C_2/C_1) \quad (A25)$$

with

$$S_\pm(x_l, x_r) = \frac{1}{2\pi\sigma_x^2}(a^2 + b^2 \pm 2ab \cos(\omega_x^o(x_l - x_r) + \phi_-)) \quad (A26)$$

and $H(t)$ defined in *Eq. 15*. According to the energy model, a complex cell sums up the half-squared outputs of four simple cells, all with identical RF parameters except their $\phi_+/2$ differing in steps of $\pi/2$. Thus impulse response function of the complex cell to the dichoptically flashed line is

$$F_c^\pm(x_l, x_r, t) = \Theta\left[C \cos\left(\gamma + \frac{\phi_+}{2}\right)\right] + \Theta\left[C \cos\left(\gamma + \frac{\phi_+ + \pi}{2}\right)\right]$$

$$+ \Theta\left[C \cos\left(\gamma + \frac{\phi_+ + 2\pi}{2}\right)\right] + \Theta\left[C \cos\left(\gamma + \frac{\phi_+ + 3\pi}{2}\right)\right]$$

$$= C^2 = S_\pm(x_l, x_r)H(t) \quad (A27)$$

An important feature of *Eq. A27* is that it is spatiotemporally separable into $S_\pm(x_l, x_r)$ and $H(t)$, regardless of whether the complex cell is directional or not. For comparison with the experiments, we transform the preceding expressions with the same procedures used by Ohzawa et al. (1997) for processing their experimental data to obtain the purely binocular interaction RF profiles: we first remove purely monocular responses (a^2 and b^2 in *Eq. A26*) by subtracting F^- from F^+, then change the variables x_l and x_r to

$$x_+ = x_l + x_r, \quad D = x_l - x_r \quad (A28)$$

to make disparity D explicit, and finally integrate $(F^+ - F^-)$ with respect to x_+. The final binocular interaction RF, also called $D - T$ profile (Ohzawa et al. 1997), for model complex cells is given by

$$F_c(D, t) = \int_{-\infty}^{\infty} [S_+(D, x_+)H(t) - S_-(D, x_+)H(t)]dx_+$$

$$= S(D)H(t) \quad (A29)$$

where $S(D)$ is defined in *Eq. 14*. This is *Eq. 13*. No approximation is used in this derivation.

REFERENCES

ADELSON EH AND BERGEN JR. Spatiotemporal energy models for the perception of motion. *J Opt Soc Am A* 2: 284–299, 1985.

ALBRECHT DG AND HAMILTON DB. Striate cortex of monkey and cat: contrast response function. *J Neurophysiol* 48: 217–237, 1982.

ANSTIS SM AND HASSIS JP. Movement aftereffects contingent on binocular disparity. *Perception* 3: 153–168, 1974.

ANZAI A, OHZAWA I, AND FREEMAN RD. Neural mechanisms for processing binocular information. I. Simple cells. *J Neurophysiol* 82: 891–908, 1999a.

ANZAI A, OHZAWA I, AND FREEMAN RD. Neural mechanisms for processing binocular information. II. Complex cells. *J Neurophysiol* 82: 909–924, 1999b.

BARLOW HB, BLACKMORE C, AND PETTIGREW JD. The neural mechanism of binocular depth discrimination. *J Physiol (Lond)* 193: 327–342, 1967.

BRADLEY DC, QIAN N, AND ANDERSEN RA. Integration of motion and stereopsis in cortical area MT of the macaque. *Nature* 373: 609–611, 1995.

CHEN Y, WANG Y, AND QIAN N. Simulations of disparity tuning curves of V1 simple and complex cells to dynamic, time-dependent stimuli (Abstract). *Invest Ophthalmol and Vis Sci Suppl* 41: 734, 2000.

CUMMING BG AND PARKER AJ. Binocular mechanisms for detecting motion-in-depth. *Vision Res* 34: 483–495, 1994.

CUMMING BG AND PARKER AJ. Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature* 389: 280–283, 1997.

CYNADER MS AND REGAN D. Neurones in cat parastriate cortex sensitive to the direction of motion in three-dimensional space. *J Physiol (Lond)* 274: 549–569, 1978.

DAUGMAN JG. 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, 1985.

DEANGELIS GC, GHOSE GM, OHZAWA I, AND FREEMAN RD. Functional micro-organization of primary visual cortex: receptive field analysis of nearby neurons. *J Neurosci* 19: 4046–4064, 1999.

DEANGELIS GC, OHZAWA I, AND FREEMAN RD. Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature* 352: 156–159, 1991.

DEANGELIS GC, OHZAWA I, AND FREEMAN RD. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *J Neurophysiol* 69: 1091–1117, 1993a.

DEANGELIS GC, OHZAWA I, AND FREEMAN RD. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. *J Neurophysiol* 69: 1118–1135, 1993b.

EMERSON RC, BERGEN JR, AND ADELSON EH. Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Res* 32: 203–218, 1992.

FLEET DJ, WAGNER H, AND HEEGER DJ. Encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision Res* 36: 1839–1858, 1996.

FREEMAN RD AND OHZAWA I. On the neurophysiological organization of binocular vision. *Vision Res* 30: 1661–1676, 1990.

GASKA JP, JACOBSON LD, CHEN HW, AND POLLEN DA. Space-time spectra of complex cell filters in the macaque monkey: a comparison of results obtained with pseudowhite noise and grating stimuli. *Vis Neurosci* 11: 805–821, 1994.

HARRIS JM AND WATAMANIUK SNJ. Speed discrimination of motion-in-depth using binocular cue. *Vision Res* 35: 885–896, 1995.

HAWKEN MJ, SHAPLEY RM, AND GROSOF DH. Temporal-frequency selectivity in monkey visual cortex. *Vis Neurosci* 13: 477–492, 1996.

HEEGER DJ. Normalization of cell responses in cat striate cortex. *Vis Neurosci* 9: 181–197, 1992.

HUBEL DH AND WIESEL T. Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *J Physiol (Lond)* 160: 106–154, 1962.

JONES JP AND PALMER LA. The two-dimensional spatial structure of simple receptive fields in the cat striate cortex. *J Neurophysiol* 58: 1187–1211, 1987.

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

MASSON GS, BUSETTINI C, AND MILES FA. Vergence eye movements in response to binocular disparity without depth perception. *Nature* 389: 283–286, 1997.

MAUNSELL JHR AND VAN ESSEN DC. 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, 1983.

MIKALIAN S AND QIAN N. A physiologically-based explanation of disparity attraction and repulsion. *Vision Res* 40: 2999–3016, 2000.

NAWROT M AND BLAKE R. Neural integration of information specifying structure from stereopsis and motion. *Science* 244: 716–718, 1989.

NOMURA M, MATSUMOTO G, AND FUJIWARA S. A binocular model for the simple cell. *Biol Cybern* 63: 237–242, 1990.

OHZAWA I, DEANGELIS GC, AND FREEMAN RD. Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science* 249: 1037–1041, 1990.

OHZAWA I, DEANGELIS GC, AND FREEMAN RD. Encoding of binocular disparity by simple cells in the cat's visual cortex. *J Neurophysiol* 75: 1779–1805, 1996.

OHZAWA I, DEANGELIS GC, AND FREEMAN RD. Encoding of binocular disparity by complex cells in the cat's visual cortex. *J Neurophysiol* 77: 2879–2909, 1997.

OHZAWA I AND FREEMAN RD. The binocular organization of complex cells in the cat's visual cortex. *J Neurophysiol* 56: 243–259, 1986a.

OHZAWA I AND FREEMAN RD. The binocular organization of simple cells in the cat's visual cortex. *J Neurophysiol* 56: 221–242, 1986b.

POGGIO GF AND FISCHER B. Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *J Neurophysiol* 40: 1392–1405, 1977.

POGGIO GF, GONZALEZ F, AND KRAUSE F. Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *J Neurosci* 8: 4531–4550, 1988.

POGGIO GF, MOTTER BC, SQUATRITO S, AND TROTTER Y. Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms. *Vision Res* 25: 397–406, 1985.

POGGIO GF AND POGGIO T. The analysis of stereopsis. *Annu Rev Neurosci* 7: 379–412, 1984.

POGGIO GF AND TALBOT WH. Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey. *J Physiol (Lond)* 315: 469–492, 1981.

POLLEN DA. Phase relationship between adjacent simple cells in the visual cortex. *Nature* 212: 1409–1411, 1981.

QIAN N. Computing stereo disparity and motion with known binocular cell properties. *Neural Comput* 6: 390–404, 1994.

QIAN N. Binocular disparity and the perception of depth. *Neuron* 18: 359–368, 1997.

QIAN N AND ANDERSEN RA. V1 responses to transparent and non-transparent motions. *Exp Brain Res* 103: 41–50, 1995.

QIAN N AND ANDERSEN RA. A physiological model for motion-stereo integration and a unified explanation of Pulfrich-like phenomena. *Vision Res* 37: 1683–1698, 1997.

QIAN N, ANDERSEN RA, AND ADELSON EH. Transparent motion perception as detection of unbalanced motion signals. I. Psychophysics. *J Neurosci* 14: 7357–7366, 1994a.

QIAN N, ANDERSEN RA, AND ADELSON EH. Transparent motion perception as detection of unbalanced motion signals. III. Modeling. *J Neurosci* 14: 7381–7392, 1994b.

QIAN N AND MIKAELIAN S. Relationship between phase and energy methods for disparity computation. *Neural Comput* 12: 279–292, 2000.

QIAN N AND ZHU Y. Physiological computation of binocular disparity. *Vision Res* 37: 1811–1827, 1997.

READ JCA, CUMMING BG, AND PARKER AJ. Local models can account for the reduced response of disparity-tuned V1 neurons to anti-correlated images. *Soc Neurosci Abstr* 26, 2000.

REGAN D AND BEVERLEY KI. Disparity detectors in human depth perception: evidence for directional selectivity. *Nature* 181: 877–879, 1973.

SANGER TD. Stereo disparity computation using gabor filters. *Biol Cybern* 59: 405–418, 1988.

SCHILLER PH, FINLAY BL, AND VOLMAN SF. Quantitative studies of single-cell properties in monkey striate cortex. I. Spatiotemporal organization of receptive fields. *J Neurophysiol* 39: 1288–1319, 1976.

SCLAR G, MAUNSELL JHR, AND LENNIE P. Coding of image contrast in central visual pathways of the macaque monkey. *Vision Res* 30: 1–10, 1990.

SKOTTUN BC, GROSOF DH, AND DEVALOIS RL. Responses of simple and complex cells to random dot patterns: a quantitative comparison. *J Neurophysiol* 59: 1719–1735, 1988.

SNOWDEN RJ, TREUE S, AND ANDERSEN RA. The response of neurons in areas V1 and MT of the alert rhesus monkey to moving random dot patterns. *Exp Brain Res* 88: 389–400, 1992.

WATSON AB AND AHUMADA AJ. Model of human visual-motion sensing. *J Opt Soc Am A* 2: 322–342, 1985.

WESTHEIMER G. Detection of disparity motion by the human observer. *Opt Vision Sci* 67: 627–630, 1990.

ZHU Y AND QIAN N. Binocular receptive fields, disparity tuning, and characteristic disparity. *Neural Comput* 8: 1611–1641, 1996.