Computing relative motion with complex cells

BABETTE K. DELLEN, JOHN W. CLARK, AND RALF WESSEL
Department of Physics, Washington University in Saint Louis, One Brookings Drive, St. Louis

(Received May 6, 2004; Accepted January 20, 2005)

Abstract
Contextual influences shape our perception of local visual stimuli. Relative-motion stimuli represent an important contextual influence, yet the mechanism subserving relative-motion computation remains largely unknown. In the present work, we investigated the responses of an established model for simple and complex cells to relative-motion stimuli. A straightforward mathematical analysis showed that relative-motion computation is inherent in the nonlinear transformation of the complex-cell model. Tuning to relative velocity is achieved by applying a temporal filter to the complex-cell response. The mathematical inference is supported by simulations that quantitatively reproduce measured complex-cell responses in both cat and monkey to a variety of relative-motion stimuli. Importantly, the posited mechanism for cortical computation of relative motion does not require an intermediate neural representation of local velocities and does not require lateral or feedback interactions within a network.

Keywords: Contextual influences, Energy model, Relative motion, Simple and complex cells, Visual cortex

Introduction
Our percepts are shaped by the spatiotemporal context within which local features are presented (Albright & Stoner, 2002). Understanding the mechanism of global stimulus integration remains a central challenge in computational systems neuroscience. An especially important problem of global stimulus integration is the computation of relative motion, which is required for solving many complex tasks of the visual system (Nakayama, 1985), such as depth perception by motion parallax and motion-induced figure-ground segregation. Neural correlates of relative-motion computation are abundant in the visual cortex (Orban et al., 1987; Lamme, 1995; Li et al., 1999; Cao & Schiller, 2003) and subcortical structures (Frost & Nakayama, 1983; Bender & Davidson, 1986; Sun et al., 2002), but little is known about the mechanism of this computation.

In the primary visual cortex, two main types of neurons, simple and complex cells, are typically distinguished on the basis of their response properties to visual stimuli (Hubel & Wiesel, 1962; Movshon et al., 1978a,b; Skottet et al., 1991). Simple cells are found to sum the contributions from different locations within the visual field linearly, while complex cells employ nonlinear transformations that have been successfully modelled as nonlinear functions of simple-cell transformations (Hubel & Wiesel, 1962; Adelson & Bergen, 1985; Emerson et al., 1992; Chance et al., 1999). Many V1 neuron response properties, such as spatial and temporal frequency selectivity, velocity tuning, spatial-phase independence, and frequency doubling, are reproduced with these models (Dayan & Abbott, 2002). In the work being reported, we studied the response properties of an established model for simple and complex cells to relative-motion cues. The results of the mathematical investigation and the simulation reveal that this descriptive model of complex cells is particularly suited for the computation of relative motion.

Materials and methods
The quantitative model of the spatial simple-cell response specified by eqn. (2) involves two parameters, namely (1) the parameter $k$, which determines the two-dimensional vector $k = ke_\gamma$, and (2) the standard deviation $\sigma$ in the normalized Gaussian function $W(x) = 1/(\pi\sigma^2)\exp[-(x^2+y^2)/\sigma^2]$ assumed for the window function $W(x)$ of the position $x$ in the visual plane. The parameter $k$ is to be interpreted as the preferred spatial frequency of the model, while $\sigma$ determines the size of the receptive field of the model cell. We have chosen $k = 0.5$ cycles/deg and $\sigma = 1.6/\sqrt{2}$ deg. These choices imply $k\sigma = 3.55$, which lies well within the range of experimental estimates for the product $k\sigma$ (Movshon et al., 1978c; Ringach, 2002). The resulting Gabor filter is plotted in Fig. 1A. The same parameters apply to all figures except Figs. 5 and 6.

For temporal filtering, we use the function

$$h_{\text{temp}}(t) = a \exp[-\alpha(t - \tau_{\text{temp}})] \times \{[\alpha(t - \tau_{\text{temp}})]^a_{\text{temp}}/a_{\text{temp}}! - [\alpha(t - \tau_{\text{temp}})]^b_{\text{temp}}/b!\},$$  (1)
where the parameters $a$, $a$ and $b$ differ for simple ($s$) and complex ($c$) cells. The function adopted in eqn. (1) is basically identical to the filter that was proposed by Adelson and Bergen (1985), except for a newly introduced latency parameter $\tau$.

We have chosen the temporal filter of the simple cells to be lowpass with $\tau_s = 0.03$ s, $\alpha_s = 6000/45$ s$^{-1}$, $a_s = 3$, and $b_s = 5$. For the complex cell, we employ a bandpass filter with $\tau_c = 0.035$ s, $\alpha_c = 4000/15$ s$^{-1}$, $a_c = 5$, and $b_c = 7$. Both of these filters, which are plotted in Fig. 1B, resemble real temporal responses from V1 (Adelson & Bergen, 1985; DeAngelis et al., 1993). Again, the same parameters choices apply to all figures except Figs. 5 and 6. The complex-cell model used for the simulations is depicted schematically in Fig. 1C.

The model responses were calculated by integrating the response of the model complex cell over a time interval given by the travel time of the object through the center of the window function. To obtain the tuning curves, we first computed the normalized firing rate and the corresponding standard error for each stimulus configuration by averaging over 50 trials with refreshed random-dot patterns. The tuning curves of Figs. 4A and 4C were then scaled and shifted to permit a comparison with the experimental data, which were plotted in the form presented in the original papers.

In Fig. 5, we simulated a population of complex cells that consisted of nine subpopulations, each containing 121 identical model complex cells. The spatial frequencies across the different subpopulations were ranging from 0.2 cycles/deg to 1 cycle/deg, in steps of 0.1 cycles/deg. The receptive-field size was varied so as to keep $k \sigma = 3.55$ constant. The preferred temporal frequency of the temporal filter was changed along with the preferred spatial frequency of the model cell to obtain the desired speed tuning. We computed the mean firing rate of every model complex cell by averaging over 30 trials with refreshed random-dot patterns and integrated the response over a time interval of 1.5 s. The resulting values were normalized independently for each subpopulation. The normalized responses of the model complex cells were summed and the integral value was displayed at the position of the receptive-field center of the respective model neurons (Fig. 5B). For Fig. 5C,
we scaled the simulation data to fit the experimental data, which were plotted in the form presented in the original paper.

In Fig. 6, the product $k\sigma$ ranges from 0.4 to 6.8. The spatial frequency is kept constant at 0.5 cycles/deg, while the size of the receptive field is varied. The modulation ratio is defined as the ratio of the mean complex-cell response for incoherent motion to that for coherent motion of the stimulus. The mean complex-cell responses were computed by averaging over 60 trials with refreshed random-dot patterns.

During the computations, the response of the complex cell was updated at time steps of $\Delta t = 0.0016$ s. All simulations were computed with a spatial resolution of 0.05 deg.

Results

Derivation of model complex-cell responses to two moving spots

The response of nondirectional simple cells to a dynamic intensity distribution $I(x, t)$ can be derived by applying a spatial filter and a temporal filter in succession. The spatial filter is often described by a Gabor function, and the spatial simple-cell response becomes

$$ s_p(t) = \int_{-\infty}^{\infty} I(x, t) W(x) \cos(k \cdot x - \phi) \, dx. $$

(2)

Here $W(x)$ is the Gaussian window function, $k = |k|$ is the preferred spatial frequency of the filter, and $k/k$ defines the orientation of the spatial filter in space. The parameter $\phi$ is the spatial phase of the filter. The vectors $x$ and $k$ are two-dimensional.

Subsequent temporal filtering with a filter $h_s(t)$ yields the spatiotemporally filtered simple-cell response:

$$ s_p(t) = \int_{-\infty}^{\infty} h_s(t-t') s_p(t') \, dt'. $$

(3)

Additional half-wave rectification can be applied to prevent negative responses.

One of the basic properties of this linear, nondirectional spatiotemporal filter is its sensitivity to visual motion. We demonstrate this for a stimulus that consists of a single spot of light that moves with a velocity $v_a$. The intensity distribution of the moving spot may be written as a delta function $I(x, t) = \delta(x - v_a t)$ and filtering with a Gabor filter yields

$$ s_p(t) = \int_{-\infty}^{\infty} \delta(x - v_a t) W(x) \cos(k \cdot x - \phi) \, dx $$

$$ = W(v_a t) \cos(k \cdot v_a t - \phi). $$

(4)

(5)

When the spot moves through the window function, $s_p(t)$ oscillates with a frequency $k \cdot v_a / 2\pi$; hence, the movement of the stimulus is represented by the temporal oscillation of the Gabor-filtered intensity distribution. The factor $W(v_a t)$ is approximately constant in time so long as the stimulus stays roughly within the window. Consequently, the subsequent application of a temporal filter $h_s(t)$, tuned to a preferred temporal frequency, produces tuning of the response amplitude to a preferred stimulus velocity. Velocity tuning of real simple-cell responses can be reproduced on the basis of such a descriptive model (Adelson & Bergen, 1985).

Many neurons in the primary visual cortex are also sensitive to relative motion, that is, the relative velocity of two independently moving parts of the visual scene (Orban et al., 1987; Li et al., 1999; Cao & Schiller, 2003). Therefore we investigate the model’s response to relative-motion stimuli as well.

We consider a stimulus that consists of two spots of lights, moving with velocities $v_a$ and $v_b$ and starting at positions $x_a$ and $x_b$, where the subscripts will later stand for object and background, respectively. We describe the stimulus by two delta functions $I(x, t) = \delta(x - v_a t + x_a) + \delta(x - v_b t + x_b)$. The spatially filtered model simple-cell response becomes

$$ s_p(t) = \int_{-\infty}^{\infty} [\delta(x - v_a t + x_a) + \delta(x - v_b t + x_b)] $$

$$ \times W(x) \cos(k \cdot x - \phi) \, dx $$

$$ = W(v_a t - x_a) \cos(k \cdot v_a t - \phi - k \cdot x_a) $$

$$ + W(v_b t - x_b) \cos(k \cdot v_b t - \phi - k \cdot x_b). $$

(6)

(7)

If the two spots move with the same absolute velocity $|v_a| = |v_b|$ in parallel $v_a \parallel v_b$ and start at symmetric distances $x_a = -x_b$ from the center of the window function, eqn. (7) simplifies to

$$ s_p(t) = W(v_a t - x_a) $$

$$ \times [\cos(k \cdot v_a t - \phi - k \cdot x_a) + \cos(k \cdot v_b t - \phi + k \cdot x_a)] $$

$$ = 2W(v_a t - x_a) \cos[k \cdot (v_a - v_b) t/2 - k \cdot x_a] $$

$$ \times \cos[k \cdot (v_a + v_b) t/2 - \phi], $$

(8)

(9)

where the trigonometric identity $\cos x + \cos y = 2 \cos[(x - y)/2] \cos[(x + y)/2]$ was used in the last step. Hence, the spatial response of the model simple cell $s_p(t)$ oscillates with a frequency $k \cdot v_a / 2\pi$ not only when the two spots move in the same direction, but also when they move in opposite directions. The temporally filtered spatial response of the model simple cell will therefore be identical for both cases, giving

$$ s_p(t) = \alpha(|v_a|) 2W(v_a t - x_a) \cos[k \cdot (v_a - v_b) t/2 - k \cdot x_a] $$

$$ \times \cos[k \cdot (v_a + v_b) t/2 - \phi], $$

(10)

where $\alpha(|v_a|)$ is a scaling factor due to the temporal filtering that depends only on the absolute velocities of the spots (for simplicity, we omit irrelevant changes to the functional form of $s_p(t)$ caused by the application of the temporal filter). To distinguish between relative and coherent motion with such a model, we have to separate the two oscillatory factors from each other. One possibility is to sum two squared simple-cell responses $s_p(t)$ with phases 0 and $\pi/2$, yielding

$$ c(t) = s_p^2(t) + s_p^2(t). $$

(11)

The function $c(t)$ is identical to the description of complex-cell responses in the energy model (Adelson & Bergen, 1985). This is intrinsically very interesting, since the removal of the coherent-motion term has led us directly to a computation that has already been identified in the primary visual cortex.
For the example of the two moving spots, the model complex-cell response becomes

\[
c(t) = 4a^2(\mathbf{v}_a) W^2(\mathbf{v}_a t - \mathbf{x}_a) \cos^2[\mathbf{k} \cdot (\mathbf{v}_a - \mathbf{v}_b) t - 2 \mathbf{k} \cdot \mathbf{x}_a]
\]

(12)

\[
= 2a^2(\mathbf{v}_a) W^2(\mathbf{v}_a t - \mathbf{x}_a)[1 + \cos[\mathbf{k} \cdot (\mathbf{v}_a - \mathbf{v}_b) t - 2 \mathbf{k} \cdot \mathbf{x}_a]],
\]

(13)

where the trigonometric identities \(\cos^2 x + \sin^2 y = 1\) and \(2 \cos^2 x = \cos(2x) + 1\) were used. The model complex-cell response oscillates when the two spots move in opposite directions, but not when they move in the same direction. Tuning of the response amplitude to relative motion can be achieved by applying a temporal filter \(h_c(t)\) to the response \(c(t)\), yielding

\[
\tilde{c}(t) = \int_{-\infty}^{\infty} h_c(t - t') c(t') \, dt'.
\]

(14)

Additional half-wave rectification can be invoked to prevent negative responses. Application of a temporal filter to the model complex-cell response is a natural extension of the energy model. Frequency doubling is still observed in the extended complex-cell model, since the frequency of the oscillation is not affected by the linear temporal filter, but only its amplitude. Since the temporal filter does not have a spatial dimension, the model complex-cell response also remains invariant to the spatial phase of the stimulus.

*Extension to arbitrary moving stimuli*

The special case discussed above can be generalized from two moving spots to arbitrary moving stimuli. Using the trigonometric identity \(\cos(x - y) = \cos x \cos y + \sin x \sin y\), the spatial simple-cell response is readily expressed as

\[
s_b(t) = \int_{-\infty}^{\infty} I(x, t) W(x) \cos(k \cdot x - \phi) \, dx
\]

(15)

\[
= \int_{-\infty}^{\infty} I(x, t) W(x) \cos(k \cdot x) \cos \phi \, dx

+ \int_{-\infty}^{\infty} I(x, t) W(x) \sin(k \cdot x) \sin \phi \, dx
\]

(16)

\[
= \text{Re} \mathcal{F}_k[I(x, t) W(x)] \cos \phi + \text{Im} \mathcal{F}_k[I(x, t) W(x)] \sin \phi
\]

(17)

with

\[
\mathcal{F}_k[I(x, t) W(x)] = \int_{-\infty}^{\infty} I(x, t) W(x) \exp(i k \cdot x) \, dx.
\]

(18)

Here, \(\text{Re} \mathcal{F}_k\) denotes the real part of the spatial Fourier transform, and \(\text{Im} \mathcal{F}_k\) its imaginary part. With the identities

\[
\mathcal{F}_k[I(x, t) W(x)] = |\mathcal{F}_k[I(x, t) W(x)]| \exp(i \Phi)
\]

(19)

and

\[
\Phi = \arctan \frac{\text{Im} \mathcal{F}_k[I(x, t) W(x)]}{\text{Re} \mathcal{F}_k[I(x, t) W(x)]}
\]

(20)

we write

\[
s_b(t) = |\mathcal{F}_k[I(x, t) W(x)]| \cos \Phi \cos \phi + \sin \Phi \sin \phi
\]

(21)

\[
= |\mathcal{F}_k[I(x, t) W(x)]| \cos(\phi - \Phi).
\]

(22)

Squaring and adding two model simple-cell responses shifted by \(\pi/2\), one obtains the model complex-cell response, namely

\[
c(t) = s_b^2(t) + s_b^2(t)
\]

(23)

\[
= |\mathcal{F}_k[I(x, t) W(x)]|^2
\]

(24)

\[
= |\text{Re} \mathcal{F}_k[I(x, t) W(x)]|^2 + |\text{Im} \mathcal{F}_k[I(x, t) W(x)]|^2.
\]

(25)

Here, we have neglected the temporal filtering performed by the model simple cells. This is justified as long as the speeds of the moving stimuli are close to the preferred speed of the model simple cells. Thus, the complex-cell response emerges as the square of the absolute value of the Fourier transform of the window-filtered intensity distribution.

For a coherently moving intensity distribution \(I_b(x, t) = I_0(x - v_a t)\), we make the substitution \(x \rightarrow x + v_a t\) and find

\[
\text{Re} \mathcal{F}_k[I_b(x, t) W(x)]
\]

(26)

\[
= \int_{-\infty}^{\infty} I_b(x - v_a t) W(x) \cos(k \cdot x) \, dx
\]

\[
= \cos(k \cdot v_a t) \int_{-\infty}^{\infty} I_b(x) W(x + v_a t) \cos(k \cdot x) \, dx

- \sin(k \cdot v_a t) \int_{-\infty}^{\infty} I_b(x) W(x + v_a t) \sin(k \cdot x) \, dx
\]

(27)

\[
= \cos(k \cdot v_a t) \cos \eta - \sin(k \cdot v_a t) \sin \eta
\]

(28)

\[
= A_o \cos(k \cdot v_a t + \eta)
\]

(29)

and

\[
\text{Im} \mathcal{F}_k[I_b(x, t) W(x)] = A_o \sin(k v_a t + \eta),
\]

(30)

where

\[
A_o = |\mathcal{F}_k[I_b(x) W(x + v_a t)]|
\]

(31)

and

\[
\eta = \arctan \frac{\text{Im} \mathcal{F}_k[I_b(x) W(x + v_a t)]}{\text{Re} \mathcal{F}_k[I_b(x) W(x + v_a t)]}.
\]

(32)

In the same way, we find exactly the same expression for the other intensity distribution \(I_a(x, t) = I_0(x - v_b t)\), except that the subscript \(o\) is replaced everywhere by \(b\).

Thus, for a stimulus consisting of two moving intensity distributions

\[
I(x, t) = I_o(x - v_a t) + I_b(x - v_b t),
\]

(32)
Computing relative motion

where the trigonometric identities \( \cos^2 x + \sin^2 y = 1 \) and \( \cos(x - y) = \cos x \cos y + \sin x \sin y \) were used. The integrals \( A_o \) and \( A_b \) are approximately constant in time so long as the image stays roughly within the window \( W(x) \).

Eqn. (35) contains a modulatory term that oscillates at the frequency \( \{ k \cdot (v_o - v_b) + \eta \} / 2\pi \), thereby reporting the presence of a velocity difference in the image. Hence, subsequent temporal filtering of \( c(t) \) via [Eqn. (14)] will produce tuning to relative motion.

Directionally-selective complex cells are treated in the Appendix.

Simulations

To illustrate the key features of the model and to compare responses with experimental data, we simulated model complex-cell responses to a variety of relative-motion cues. The complex-cell model is sketched in Figs. 1A–1C. We chose a representative example of a model complex cell consisting of two model simple cells, phase shifted by \( \pi/2 \), with spatial and temporal parameters drawn from the experimentally determined range (Materials and methods and Figs. 1A–1B). The two independently moving parts of the visual stimulus moved with velocities \( v_o \) and \( v_b \). The directions of motion of the object and the background were chosen to be parallel to the direction of \( k \); the model complex cell was not directionally selective. In the simulations, the background spanned \( 10 \times 10 \text{ deg}^2 \) of the visual field. For comparison, the approximate width of the Gaussian window function of the model complex cell is shown as a gray circle in the left panels of Fig. 4.

The first stimulus was a single spot of light moving through the area of integration of the model complex cell. The model simple cells oscillated in response to the movement of the luminous spot (Fig. 2A). But, since the model simple cells are phase shifted by \( \pi/2 \), the temporal responses are phase shifted by the same amount. Thus, when adding the squared simple-cell responses, the oscillations cancelled and the total response \( c(t) \) reflected merely the shape of the window function (Fig. 2B, dashed line). Subsequent temporal filtering led to a reduction of the amplitude (Fig. 2B, solid line). However, when stimulating the model cells with two spots that moved in opposite directions, both cells oscillated in phase (Fig. 2C), so the sum of their squares oscillated as well (Fig. 2D, dashed line). The temporal filter of the model complex cell magnified the amplitude of this oscillation, thus enhancing the cell’s response to relative motion (Fig. 2D, solid line).

Similar effects were observed for more complex relative-motion stimuli. The next stimulus (Fig. 3A, inset) consisted of a stationary circular window that contained a drifting sine-wave grating moving with a velocity \( v_o \). The center of the circular window coincided with the center of the window function of the model complex cell and was surrounded by a background sine-wave grating that moved with a velocity \( v_b \). The width of the window function (\( \approx 5.5 \text{ deg} \)) exceeded the size of the circular window (\( \approx 3 \text{ deg} \)), thus ensuring that a part of the relative-motion cue was stimulating the model complex cell. The model complex-cell responses were calculated as a function of the background velocity \( v_b \) for three different object velocities, namely \( v_o = 2 \text{ deg/s} \) (square), \( v_o = 4 \text{ deg/s} \) (diamond), and \( v_o = 8 \text{ deg/s} \) (triangle) (Fig. 3A). We observed tuning to relative motion rather than absolute motion of the backgrounds. In all three cases, the tuning curves had a minimum at \( (v_b - v_o)/v_o = 0 \), that is, zero relative motion of the object and background grating. The model’s relative-motion sensitivity did not depend on the phase of the gratings (data not shown); rather, it originates from the oscillation of the relative-motion term that is inherent in the simple-cell model and is isolated by means of a complex-cell transformation.

Next, we explored how the background motion affects the speed tuning of the model complex cell to the center stimulus (Figs. 3B–3D). First, the tuning curve is calculated as a function of the object velocity of the center grating for a stimulus consisting of the center grating only (Fig. 3B). Since the model complex cell was not selective for direction, it responded to negative as well as to positive directions of movement in a symmetrical fashion, and the tuning curve had two peaks, whose locations define the “optimal speed”. The tuning profile is a result of both the lowpass properties of the model simple cells and the bandpass property of the temporal filter of the model complex cell. When the background grating was added to the stimulus, the tuning profile changed in a manner dependent on the velocity of the background. For a background moving slower than the optimal speed of the center grating (Fig. 3C), the model complex cell was responding to the relative motion between center and background; hence the right peak of the tuning curve shifted to the high object-velocity side. For a background moving faster than the optimal speed of the center grating, the right part of the tuning curve shifted even further to the high object-velocity side, yet only the left peak survived temporal filtering (Fig. 3D). In comparison with Fig. 3B, the tuning curve is apparently shifted to the low-velocity side. The model simulation thus reproduced relative-motion effects that have been observed in cortical area 18 of cat (Li et al., 1999). There, targeted cortical neurons exhibit shifts in speed selectivity to higher or lower object velocity if the background velocity is respectively lower or higher than the optimal speed of the center grating.

Next we simulated the response of the model complex cell to an opaque random-dot bar (object) that was moving in front of a moving random-dot background (Fig. 4A). When the background and the bar were moving in the same direction (to the right) and with the same velocity, the random-dot bar was indistinguishable from the background, and the model complex cell did not respond upon movement of the bar through the Gaussian window function. Clusterings of random dots provided absolute-motion cues that caused the complex-cell response to fluctuate with a low frequency. In contrast, when the background was moving in the opposite direction (to the left) but with the same speed as the bar, the model complex cell responded vigorously upon motion of the bar through the Gaussian window function. We calculated the model complex-cell responses as a function of the background velocity \( v_b \) for three different object velocities, namely \( v_o = 2 \text{ deg/s} \) (square), \( v_o = 4 \text{ deg/s} \) (diamond), and \( v_o = 8 \text{ deg/s} \) (triangle). The third panel of Fig. 4A clearly demonstrates tuning to relative motion, that is, to \( v_b - v_o \), that is similar to the responses obtained for the grating stimulus investigated previously (Fig. 3A). For high object and background velocities, the tuning to relative motion was weakened, and the previously V-shaped tuning curve became distorted. We compared the simulation results with experi-
imental data (red circle) from neurons in monkey primary visual cortex in response to the same relative-motion stimulus (Cao & Schiller, 2003). Both model and experiment showed V-shaped relative speed tuning and velocity cut-off at high background and/or object velocities.

Similar behavior was found for a transparent random-dot bar (object) (Fig. 4B) and an opaque short solid bar (object) (Fig. 4C), both moving in front of the random-dot background. Under the conditions of transparency, dots of the background were visible between the gaps of the dots of the random-dot bar. In both cases, the model complex cell responded with a pronounced oscillation when presented with the relative-motion stimulus. However, since both the transparent random-dot bar and the short solid bar provided additional absolute-motion cues, we observed an elevation in the firing rate also in the coherent case, even though the response was generally weaker and showed less modulation than in the incoherent case. The speed tuning curves for the transparent random bar and the short solid bar showed pronounced tuning to relative motion. The tuning curve for the short solid bar was compared with experimental data from complex cells in cat primary visual cortex which had been classified as anti-phase cells (Orban et al., 1987). The model reproduces the basic shape of the measured complex-cell tuning curve (red line).

To illustrate the functional implications of relative-motion sensitivity, we considered a population of model complex cells with parameter values within the experimentally determined range (Materials and methods). The population consisted of nine subpopulations, each containing 121 model complex cells. The spatial frequencies across the different subpopulations ranged from 0.2 cycles/deg to 1 cycle/deg. The model complex cells were distributed evenly over the two-dimensional plane of the visual input. The visual scene consisted of a stationary square-shaped window of size $4 \times 4$ deg$^2$ that contained a moving random-dot pattern (Fig. 5A). The square-shaped window was surrounded by a random-dot background. The dots within the square were moving to the right with $v_o = 4$ deg/s, while the dots of the background were moving to the left with $v_o = -4$ deg/s. The responses of the model complex cells were integrated over a time interval of 1.5 s, normalized (Materials and methods), and the normalized firing rates of the different complex-cell types were summed at each
spatial location, resulting in a response map encoding motion discontinuities (Fig. 5B). The motion boundaries of the random-dot square can be inferred from the model complex-cell population response. The response of model complex cells along a horizontal cross section of the population for both the subpopulation with $k = 0.5\,\text{cycles/deg}$ (black cross) and for the average over different types (blue asterisk) was compared with “motion pop-out” responses produced by neurons in monkey primary visual cortex (Lamme, 1995) (red circle, Fig. 5C). The model population response qualitatively reproduces the cortical population response to motion discontinuities. We note that the model population response to the figure center was enhanced when neurons with larger receptive-field sizes were added to the population (data not shown).

**Dependence on receptive-field parameters**

How does the model complex cell’s relative-motion sensitivity depend on the bandwidth (Dayan & Abbott, 2002) of the model simple cells? Up to now, we have not discussed the influence of the window function on the temporal behavior of the simple-cell and complex-cell models. Looking again at eqn. (13) or eqn. (35), we realize that temporal modulations are also induced by the motion of image points through the window function, which will limit the ability of the model to sense relative motion. We consider a Gaussian window function and define the width of the Gaussian envelope to be $2\sqrt{2\sigma}$, following a previously introduced convention (DeAngelis et al., 1993). Hence, the two moving intensity distributions require the times $2\sqrt{2\sigma}/v_o$ and $2\sqrt{2\sigma}/v_b$ to cross the width of the Gaussian window, which results in temporal modulations with periods $\sqrt{2\sigma}/v_o$ and $\sqrt{2\sigma}/v_b$, respectively. Dividing these periods by the period $2\pi/k(v_o - v_b)$ of the oscillations induced by relative motion, gives the dimensionless ratios $\sqrt{2\sigma}/k(v_o - v_b)/2\pi v_o$ and $\sqrt{2\sigma}/k(v_o - v_b)/2\pi v_b$ (for simplicity it is assumed that $v_o$ and $v_b$ are parallel to $k$). It is seen that the ratio of periods so constructed is directly proportional to the parameter $k\sigma$, which determines the bandwidth of the spatial filter. Accordingly, we expect that the sensitivity of the model complex cells to relative motion will increase with increasing $k\sigma$.

We determined the model complex cell’s sensitivity to relative motion for the opaque solid short-bar stimulus of Fig. 4C as a function of the product $k\sigma$ (Fig. 6). The modulation ratio is defined as the ratio of the mean complex-cell response for incoherent motion to that for coherent motion of the target stimulus. For $k\sigma > 3$, we observed pronounced sensitivity to relative motion. In cat visual cortex, experimental values for $k\sigma$ range between 1.7 to 6.9 (Movshon et al., 1978c). Similar values have been established in monkey visual cortex (Ringach, 2002). We expect that cortical complex cells with large $k\sigma$ are more likely to give a pronounced response to relative-motion cues than complex cells with small $k\sigma$. The product $k\sigma$ is a measure for the spatial-frequency selectivity of the spatial filter (Dayan & Abbott, 2002), and is related to the bandwidth $b$ through

$$b = \log_2 \left[ \frac{|k|\sigma + \sqrt{2\ln 2}}{|k|\sigma - \sqrt{2\ln 2}} \right].$$

(36)
Fig. 4. Model complex-cell responses to relative-motion random-dot stimuli. A: Second-order motion stimulus. An opaque random-dot bar (object) of size $7.5 \times 0.5 \text{deg}^2$ moves to the right in front of a moving random-dot background ($10 \times 10 \text{deg}^2$) (left panel). The gray circle indicates the width of the window function, which is defined as the width of the Gaussian envelope at 5% of the peak amplitude. When the random-dot bar moves coherently with the background through the receptive-field center, both moving to the right at $4 \text{deg/s}$ ($v_b = v_o$), the model complex cell shows little response (center top panel). Fluctuations of the model complex-cell response are caused by the absolute motion of the random dots moving in and out of the window function. In contrast, when the random-dot bar moves to the right against the background, which moves to the left at the same speed ($v_b = -v_o$), the model complex-cell response oscillates strongly, encoding the relative velocity between background and bar (center bottom panel). To obtain representative tuning curves, firing rates have been averaged over 50 trials, and the standard error of the simulation is plotted with the data points. The speed tuning curves, plotted as a function of the background velocity for three different bar velocities, $v_o = 2 \text{deg/s}$ (square), $v_o = 4 \text{deg/s}$ (diamond), and $v_o = 8 \text{deg/s}$ (triangle), clearly show tuning to relative motion (right panel). The three tuning curves have been scaled and shifted by the same amount to allow a comparison with experimental data from monkey primary visual cortex obtained for the same stimulus (Cao & Schiller, 2003) (red circle). Both experimental and model curves exhibit (1) V-shaped tuning to relative motion within the dynamical range of the cell and (2) a velocity cut-off at high background and/or object velocities. B: Transparent motion stimulus. A transparent random-dot bar that moves in front of the moving random-dot background (left panel) provides an absolute-motion cue that causes an elevation in the model complex-cell firing rate for coherent motion (center top panel). For a relative-motion stimulus ($v_b = -v_o$), the model complex-cell response oscillates and is stronger compared to the coherent case (center bottom panel). The speed tuning curves (right panel) show pronounced tuning to relative motion within the sensitive range of the cell. C: Opaque solid-bar stimulus. For a short solid bar (object) of size $0.3 \times 0.6 \text{deg}^2$ moving in front of a moving random-dot background (left panel), the model complex cell again responds to the relative-motion stimulus (center bottom panel) with robust oscillations, and the tuning curves (right panel) exhibit tuning to relative motion. In its basic shape, the model curve resembles the experimental curve from cat primary visual cortex (Orban et al., 1987) (red line) (right panel).
Influence of the object size on relative-motion sensitivity

We have investigated the extent to which the size of the object influences the response of the model complex cell to relative motion. Fig. 7A shows a plot of the modulation ratio of the model complex-cell response as a function of the length of the object. The stimulus consisted of an opaque random-dot bar that moved in front of a random-dot background, as already described in Fig. 4A. For simplicity, the object was square in shape. The sensitivity of the model complex cell to relative motion was found to increase with increasing size of the object. The model cell seemed to be very broadly tuned to different object sizes. However, for large sizes exceeding the width of the window function, we expect the cell’s relative-motion sensitivity to decline.

A different behavior was observed for the solid-bar stimulus of Fig. 4C. For the parameters considered, the model complex-cell response showed sharp tuning to small objects (Fig. 7B). However, tuning to object size depends on the parameters of the model complex cell.

Relation to other models

The energy-based complex-cell model (Adelson & Bergen, 1985; Emerson et al., 1992) chosen for this study is entirely descriptive, without any reference to hardware implementation. Hubel and Wiesel (1962) suggested a model for complex cells that does not involve a squaring nonlinearity. In this model, complex-cell responses are generated by summing over a population of simple cells of uniform orientation and frequency, but different spatial phases (Hubel & Wiesel, 1962; Dayan & Abbott, 2002). The response $c_{HW}(t)$ of the complex cell is produced by summing the rectified responses of several simple cells, according to

$$c_{HW}(t) = \int_{-\pi}^{\pi} [s_k(t)]_+ \, d\phi,$$

(37)

Fig. 5. Motion pop-out. A: The visual scene consists of a square-shaped window of $4 \times 4 \text{deg}^2$ that contains a moving random-dot pattern (object). The window itself is surrounded by a moving random-dot background. The dots within the window move to the right, while the background dots move to the left. Hence, the window is solely defined by relative-motion cues. B: The stimulus is presented to a population of 1089 nondirectional complex cells having spatial frequencies $|k|$ ranging from 0.2 cycles/deg to 1 cycle/deg. The receptive-field size, determined by the parameter $\sigma$, has been varied to keep the product $k \sigma$ determining the bandwidth of the simple cells constant at 3.55. The model complex-cell receptive fields are distributed evenly over a plane of size $10 \times 10 \text{deg}^2$ that covers the motion-defined figure. The firing rate of each cell is integrated over a time interval of 1.5 s and normalized with respect to the mean response of the respective subpopulation (Materials and methods). The plane is tiled into small squares, and the normalized firing rates of the different complex-cell types are summed at each spatial location. The resulting values are color-coded, and the squares at the corresponding spatial positions are filled with the appropriate color, resulting in a map of neuronal activity. The boundaries of the motion-defined square are readily discerned. C: Cross sections along the horizontal axis of the map for model complex cells with $k = 0.5$ cycles/deg (black cross) and for the whole population (blue asterisk) have been plotted, together with experimental values for neural responses from monkey primary visual cortex (red circle) in response to the same motion pop-out stimulus (Lamme, 1995). The simulation data has been scaled to cover the same range as the experimental data, which is presented as given in the original paper.
where the brackets $[]_+$ denote half-wave rectification. Using eqn. (22) and neglecting temporal filtering, $c_{HW}(t)$ becomes

$$c_{HW}(t) = \int_{-\infty}^{\infty} [\mathcal{F}[I(x,t)W(x)]] \cos(\phi - \Phi), d\phi$$

$$= [\mathcal{F}[I(x,t)W(x)]] \int_{-\infty}^{\infty} \cos(\phi - \Phi), d\phi. \quad (38)$$

The integration over $\phi$ reduces the complex-cell response $c_{HW}(t)$ to a constant term multiplied by $[\mathcal{F}[I(x,t)W(x)]]$. We conclude that the energy-based complex-cell model (Adelson & Bergen, 1985; Emerson et al., 1992) and the Hubel-Wiesel model (Hubel & Wiesel, 1962) carry out essentially the same mathematical operation. The equivalence of phase-averaging and energy methods has already been demonstrated for the computation of binocular disparity (Qian & Mikaelian, 2000). We expect that, within broad limits, the particular choice of complex-cell model does not alter the conclusions drawn from the present study. The energy model was chosen since, unlike the other models, it involves just two degrees of freedom, that is, two simple cells that are phase shifted by $\pi/2$.

**Discussion**

In the work reported here, we investigated the basic computations employed by a descriptive complex-cell model and showed that the inherent transformations are particularly suited for the computation of relative motion. We found that the spatiotemporal-filter description of nondirectional simple cells is inherently ambiguous with respect to coherent and relative motion. Relative motion further causes a characteristic oscillation of the nonlinear spatio-temporal filter that represents the complex cell. The frequency of this oscillation encodes the relative velocity of a stimulus. Hence, we suggested that the complex-cell model be extended by introduction of a temporal filter that acts on the complex-cell response to describe real cortical-cell tuning to relative velocity.

This result has important implications for our understanding of the classical and nonclassical receptive field. According to the model investigated here, relative-motion detection does not necessitate a nonclassical receptive field. The computation of relative motion is inherent in the spatiotemporal-filter description of complex cells. It is required, however, that the relative-motion cue lies within the area of integration of the complex cell. In the present work, we do not address additional contextual effects that might be
mediated by the center-surround organization of certain cortical cells.

Previous qualitative and quantitative models of relative-motion sensitivity have assumed the underlying mechanism to be an inhibitory interaction between directionally selective cells sensitive to absolute motion (Gruesser, 1971; Poggio et al., 1981; Reichhardt et al., 1989; Bradley et al., 1998). However, there is as yet no evidence that directionally selective mechanisms are involved in the computation of relative motion. Many visual neurons that are sensitive to relative motion are not selective for an absolute direction of motion, either of the target or of the surround (Frost & Nakayama, 1983; Orban et al., 1987; Davidson & Bender, 1991; Sun et al., 2002; Cao & Schiller, 2003). Moreover, currently prevalent models for directionally selective motion-sensitive cells employ inhibitory interactions between neurons that would silence neurons under conditions of transparency. Yet neurons in the primary visual cortex do respond to transparent motion (Snowden et al., 1991), a property that is reproduced by our model (Fig. 4B).

Recent reports on relative-motion sensitivity of neurons in the primary visual cortex show that V1 neurons predominantly process relative motion rather than absolute motion (Orban et al., 1987; Lamme, 1995; Li et al., 1999; Cao & Schiller, 2003). In such a system, self-induced background motion, caused by eye and body movements, is excluded from the perception from the very beginning, reducing the computational load of subsequent stages of visual processing.

Acknowledgments

We thank Charles Anderson for a seminal comment. The work was supported by grants from the Whitehall Foundation, the McDonnell Center for Higher Brain Function, and the NIH/NEI.

References


Cao, A. & Schiller, P.H. (2003). Neural responses to relative speed in the primary visual cortex of rhesus monkey. Visual Neuroscience 20, 77–84. [Data is taken from Fig. 3A. Note that there has been a misprint in the paper. The label on the x-axis should be log 2(Vs/Vf) (private correspondence with authors).]


Appendix

Directional selectivity and relative motion

We construct a directionally selective simple cell by applying first a linear, mathematically complex spatial filter to the intensity distribution of the visual space

$$S(t) = \int_{-\infty}^{\infty} I(x, t)W(x)e^{i[\mathbf{k} \cdot \mathbf{x}]}d\mathbf{x},$$

(39)

with the parameters defined as before. The capital $S$ distinguishes this construction from that of eqn. (2), which is based on a mathematically real spatial filter. Here, subsequent temporal filtering of $S(t)$ with a mathematically complex temporal filter produces tuning to the direction as well as the magnitude of the velocity. The square of the absolute value of this spatiotemporal response describes the directionally selective complex-cell response.

For a stimulus consisting of two moving intensity distributions

$$I(x, t) = I_{u}(x - v_{u}t) + I_{l}(x - v_{l}t),$$

(40)

the response $S(t)$ becomes

$$S(t) = \exp(\mathbf{k} \cdot \mathbf{v}_{u}t)F[k_{u}I_{u}(x)W(x + v_{u}t)]$$

$$+ \exp(\mathbf{k} \cdot \mathbf{v}_{l}t)F[k_{l}I_{l}(x)W(x + v_{l}t)].$$

(41)
Now, as stated before, subsequent temporal filtering with a complex filter will generate directional selectivity. The temporal filter has a preferred temporal frequency $k \cdot v$, corresponding to a preferred velocity $v$. We assume now that the object and the background move with velocity vectors close to the preferred velocity of the temporal filter, namely $v_o = v + \Delta v$ and $v_b = v - \Delta v$, where $|\Delta v| \ll |v|$. In this case, we can neglect the temporal filtering, and the directionally selective complex-cell response is found to be

$$C(t) = A_o^2 + A_b^2 + 2A_o A_b \cos(k \cdot \Delta v), \quad (42)$$

in analogy to eqn. (35). Subsequent temporal filtering with a preferred relative velocity $\Delta v$ will then generate tuning to relative velocity instead of absolute velocity. Based on this result, we predict that both directionally selective and directionally nonselective complex cells are suited to the computation of relative motion. However, sharp tuning to both direction and the magnitude of velocity does limit the range over which relative velocities can be detected. Directionally selective complex cells are expected to be less sensitive to opposing motion than nondirectionally selective complex cells.