

Figure 4.3: SENSITIVITY OF CONES TO FULL-FIELD FLICKER

Amplitude (a) and phase (b) of cone responses to temporal frequency from OPL circuit model. For the parameters values chosen, the cone's response peaks at 10cps, and levels off below 0.33cps.

the voltage of that node, and is the analog of the membrane potential of the cone. I will also plot frequency responses on a logarithmic scale, in dB³; spatial frequency in is units of cpd (cycles per degree), and temporal frequency is in units of cps (cycles per second).

4.3 Responses to Flicker and Gratings

Full-field flicker and stationary sinusoidal gratings are used by physiologists and psychophysicists to characterize the temporal and spatial responses of the visual system. In the same vein, I present analytical expressions that describe the model's response to these classic stimuli. I describe the salient features of these responses, and relate them, quantitatively, to the model's parameters. I validate the model by comparing its responses to biological measurements.

³20dB is equivalent to a tenfold increase in amplitude.

4.3.1 Full-Field Flicker

If the spatial frequency is sufficiently low (i.e. $\rho \ll \sqrt{\epsilon_c/\ell_c}, \sqrt{\epsilon_h/\ell_h}$), we can drop the spatial-frequency terms and obtain

$$\begin{aligned}\tilde{H}_c(0, \omega) &= \frac{1}{g_{ch} \tau_c \tau_h (i\omega)^2 + (\epsilon_h \tau_c + \epsilon_c \tau_h) i\omega + \epsilon_c \epsilon_h + 1}, \\ \tilde{H}_h(0, \omega) &= \frac{1}{g_{ch} \tau_c \tau_h (i\omega)^2 + (\epsilon_h \tau_c + \epsilon_c \tau_h) i\omega + \epsilon_c \epsilon_h + 1}.\end{aligned}$$

These expressions give the sensitivity of cones and horizontal cells to full-field flicker; the magnitude and phase of $\tilde{H}_c(0, \omega)$ are plotted in Figure 4.3.

The cones have a bandpass temporal-frequency response with a distinct peak at $\hat{\omega} \simeq 1/\sqrt{\tau_c \tau_h}$; the response rolls off at 20dB per decade beyond this point. Surprisingly, the peak frequency of the cone is determined not by the cone's individual time constant alone, but rather by the geometric mean of the time constants of the cone and the horizontal cell. The gain reaches a maximum value of $\sqrt{(\tau_h/\tau_c)Q_t}$ at the peak, where

$$Q_t \simeq \left(\epsilon_c \sqrt{\frac{\tau_h}{\tau_c}} + \epsilon_h \sqrt{\frac{\tau_c}{\tau_h}} \right)^{-1}.$$

These expressions for the peak frequencies and the peak gain are based on the approximation $\epsilon_c \epsilon_h \ll 1$. The cone's response levels off for frequencies below ϵ_h/τ_h . Decreasing ϵ_h , which increases the voltage gain from the cone to the horizontal cell, moves this point to lower frequencies, attenuating the low spatial frequencies further.

The phase is initially 0, rises gradually as the temporal frequency increases, reaches a maximum, and then decreases rapidly around the peak temporal frequency. The phase passes through 0 at the peak frequency $\hat{\omega}$, becomes negative, and approaches -90° .

Compare the flicker response of the model to the flicker-sensitivity curves of humans and cats, obtained from psychophysical and physiological measurements, shown in Figure 3.4a and Figure 3.6b,c, respectively. The model shows the same bandpass characteristic observed for high intensities and diffuse patterns in the human measurements and in the X-cell and Y-cell cat measurements.

The main differences between the model and the biology is that the model reproduces neither the steepness of the high-frequency cutoff, nor the rapid phase changes that occur there—cat X cells go out to 600° [90]. Kelly fitted this steep cutoff with a model of the long thin process that carries signals from the outer segment of the cone to the cone terminal; he used a continuous, passive, lossy cable model that produced an exponential rolloff [108]. Chen and Freeman fitted the sharp cutoff in Frishman’s X-cell data with a model of the phototransduction process that consisted of a cascade of eight first-order reactions [106]. These high-order cascades produced a steep cutoff and introduced a large delay, which resulted in rapid phase changes. In fact, Frishman and her colleagues showed that their phase measurements were approximated quite well with a pure delay of 24ms for the X cell and 20ms for the Y cell.

The simple model that I analyzed includes neither the phototransduction process nor the cable properties of the cone. Such a simple model cannot be expected to reproduce the biological responses exactly. My goal is to capture the bandpass character of the biological responses, and the model meets that goal. When the contribution of the phototransduction cascade to the cutoff and the phase is removed, the residual gain and phase shift look much like those of this simple model (see plots of contributions of individual stages in Chen and Freeman’s more elaborate model [106]).

The model reproduces the rise in sensitivity before the peak, as do the models proposed by Kelly and by Chen and colleagues. These models produce this behavior by placing lowpass filters in negative-feedback loops around lowpass feedforward stages (the feedback filters must have a frequency cutoff that is lower than that of the feedforward filters), just like this model does. Alternatively, the rise in sensitivity before the peak may be obtained by introducing a parallel pathway with a lower-frequency cutoff and subtracting that pathway’s output from the main feedforward pathway. These two architectures are called **feedback** and **feedforward**, respectively.

Kelly used two to four feedback loops to fit the data in the region where the gain is increasing; the number of loops needed increased with intensity [108]. On the other hand, Chen and Freeman tried to fit both the feedforward and the feedback models to the cat X-cell data, and found that the latter provided a much closer fit [106].

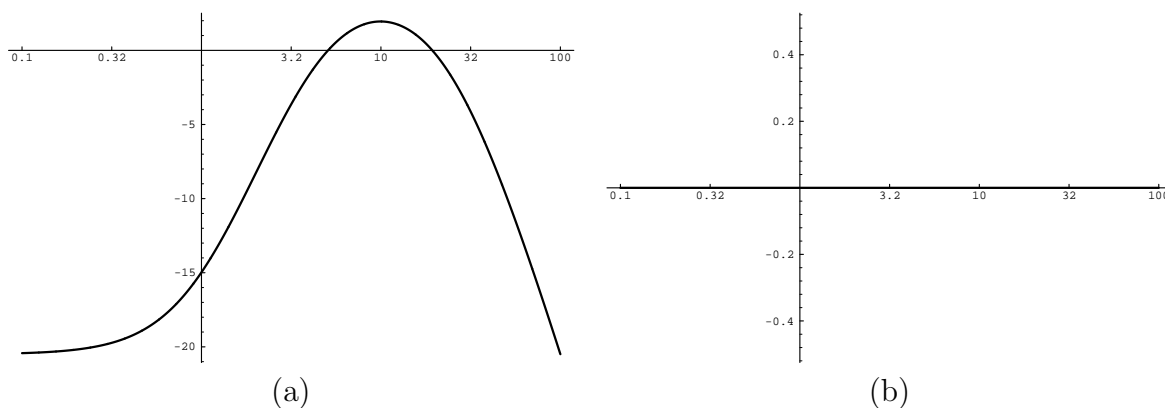


Figure 4.4: SENSITIVITY OF CONES TO STATIONARY GRATINGS

Amplitude (a) and phase (b) responses of cones to spatial frequency from OPL circuit model. For the parameter values chosen, the cone's response peaks at 10cpd, and levels off below 1.6cpd.

Only the feedback configuration could produce the sharp resonant peak evident in the full-field flicker responses; the gain at the peak is five or six times higher than a feedforward scheme predicts. However, physiological measurements in monkeys [93], and psychophysical measurements in humans [108], do not show such high resonances and are fitted by feedforward models well. The feedback model was also the only one of eight configurations studied by Chen and Freeman that satisfied the restrictions placed on the relative strengths and the relative delays between the center and the surround by the experimental measurements [106].

In summary, a simple linear two-layer feedback model:

- Accounts for the bandpass responses to temporal frequency observed for high intensities and diffuse patterns in human psychophysics and cat physiology.
- But it does not reproduce the steepness of the high-frequency cutoff, nor does it reproduce the large phase accumulation.
- This shortcoming is, most likely, because the model does not include the cable properties of the photoreceptor neurites, nor does it include the cascade of chemical reactions involved in phototransduction.

4.3.2 Stationary Gratings

If the temporal frequency is sufficiently low (i.e. $\omega \ll \epsilon_c/\tau_c, \epsilon_h/\tau_h$), we can drop the temporal frequency terms and obtain

$$\begin{aligned}\tilde{H}_c(\rho, 0) &= \frac{1}{g_{ch}} \frac{\ell_h^2 \rho^2 + \epsilon_h}{\ell_c^2 \ell_h^2 \rho^4 + (\epsilon_c \ell_h^2 + \epsilon_h \ell_c^2) \rho^2 + \epsilon_c \epsilon_h + 1}, \\ \tilde{H}_h(\rho, 0) &= \frac{1}{g_{ch}} \frac{1}{\ell_c^2 \ell_h^2 \rho^4 + (\epsilon_c \ell_h^2 + \epsilon_h \ell_c^2) \rho^2 + \epsilon_c \epsilon_h + 1}.\end{aligned}$$

These expressions give the sensitivities to stationary gratings; the magnitude and phase of $\tilde{H}_c(\rho, 0)$ are plotted in Figure 4.4.

The spatial-frequency responses parallel the temporal ones; the cones also have a bandpass spatial response. However, the amplitude of the spatial responses rises twice as fast as does the temporal response, on a log-log plot, and rolls off twice as fast as well. Another difference is that the phase of the spatial response never deviates from 0.

The cone's response peaks at $\hat{\rho} \simeq \sqrt{(1 - \epsilon_h \ell_c / \ell_h)} \sqrt{(\ell_c \ell_h)}$, attaining a maximum gain of $(\ell_h / \ell_c) Q_x$, where

$$Q_x \simeq \left(2 + \epsilon_c \frac{\ell_h}{\ell_c} - \epsilon_h \frac{\ell_c}{\ell_h} \right)^{-1}.$$

Again, I made the approximation $\epsilon_c \epsilon_h \ll 1$. In close analogy to the temporal behavior, the peak frequency is determined by the geometric mean of the space constants of the decoupled syncytia. The cone response levels off for frequencies below $\sqrt{\epsilon_h / \ell_h}$.

Compare the grating response of the model to the grating-sensitivity curves of humans and cats, obtained from psychophysical and physiological measurements, shown in Figure 3.4b and Figure 3.6a, respectively. The OPL model shows the same bandpass characteristic observed for high intensities and low temporal modulation in the human measurements and the cat X-cell measurements.

Again, the main difference between the biological responses and the model's are the model's inability to produce steep rolloff. Kelly and other researchers used an

exponentially weighted function of the form $\rho^2 e^{-\rho}$ to fit the steep rolloff found in psychophysical measurements [109, 110, 111]. This choice is consistent with Rodieck's difference-of-Gaussians (DOG) model for the receptive field center and surround, which also results in an exponential rolloff with frequency. The DOG model fits Frishman's cat ganglion-cell measurements perfectly. More detailed models, based on retinal anatomy, have shown that the gaussian-like spatial profile of the receptive-field center arises from spatial summation by the bipolar-cell dendrites [112]. Smith showed that the gaussian-like spatial profile of the receptive-field surround arises from the presence of two types of horizontal cells [112].

The simple model that I analyzed does not include bipolar convergence, and it has only one type of horizontal cell. Nevertheless, the model captures the bandpass character of the biological responses, and reproduces the increase in sensitivity before the peak, much like the other models.

For the spatial behavior, Kelly found that a feedforward model with a single stage in the parallel inhibitory pathway could account for increasing gain; the contribution from the inhibitory pathway falls off as intensity decreases [109]. At extremely low intensity levels, the response becomes lowpass and can be fitted with the exponential cut-off function over the entire frequency range [110, 111]. Thus, the trends for spatial and temporal frequencies are identical: The sensitivity to higher frequencies, and the peak contrast gain, both increase with intensity. On the other hand, Chen and Freeman used a single feedback stage to fit the spatial responses of cat X cells measured at high intensity, just as this model does.

In summary, a simple linear two-layer feedback model:

- Accounts for the bandpass responses to spatial frequency observed for high intensities and slow temporal modulation in human psychophysics and cat physiology.
- But it does not reproduce the steepness of the high-frequency cutoff,
- This shortcoming is, most likely, because the model does not include spatial summation in the bipolar cell dendrites, nor does it include a second horizontal

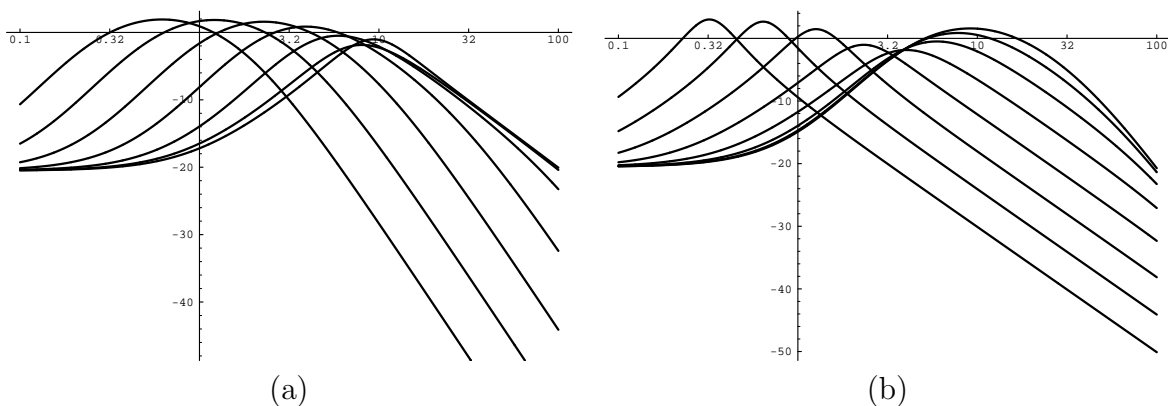


Figure 4.5: SENSITIVITY OF CONES TO MOVING GRATINGS

Response to gratings moving at various speeds, plotted versus (a) temporal frequency or (b) spatial frequency. The low-speed temporal-frequency response in (a) looks like the stationary-grating frequency response, and the high-speed spatial-frequency response in (b) looks like the flicker-frequency response. At speeds below $\hat{\omega}/\hat{\rho} = 1$ dps, the peak in the temporal-frequency response tracks the speed; at speeds below 1dps, the peak in the spatial-frequency response tracks the speed. As the curves shift, their shape remain the same.

cell network.

4.3.3 Moving Gratings

A grating with spatial frequency ρ produces the temporal frequency $\omega = v\rho$ when it moves with velocity v in the direction of its orientation. Therefore, it is easy to predict the response to such a stimulus. We simply substitute $v\rho$ for ω , and evaluate $\tilde{H}(\rho, v\rho)$; or, we substitute ω/v for ρ , and evaluate $\tilde{H}(\omega/v, \omega)$. The resulting expression tells us how spatial filtering and temporal filtering, respectively, depend on speed. In fact, we can draw salient conclusions without doing any algebra.

For slow speeds, the temporal frequencies, $v\rho$, produced by the motion are low. Hence, the temporal terms drop out, and the response is identical to that for stationary gratings, $\tilde{H}_c(\rho, 0)$, and does not depend on speed. However, if we plot the response versus temporal frequency (i.e., $\tilde{H}_c(\omega/v, 0)$), we find that the response has the same shape as the grating-sensitivity curve, but shifts to higher temporal frequen-

cies as the speed increases. In particular, at each speed, v , the peak response occurs at the temporal frequency $v\hat{\rho}$.

On the other hand, for fast speeds, the spatial frequencies, ω/v , produced by the motion are small. Hence, the spatial terms drop out, and the response has the same shape as the flicker curve, $\tilde{H}_c(0, \omega)$, and does not depend on speed. However, if we plot the response versus spatial frequency (i.e. $\tilde{H}_c(0, \rho v)$), we find that the response has the same shape as the flicker sensitivity curve, but shifts to lower spatial frequencies as the speed increases. In particular, at each speed, v , the peak response occurs at the spatial frequency $\hat{\omega}/v$.

This behavior holds for a whole class of the spatiotemporal filters, since my argument does not depend on the detailed form of the transfer function $\tilde{H}_c(\rho, \omega)$. The argument works whenever the spatial- and temporal-frequency terms become negligible at low frequencies. Consequently, for all spatiotemporal filters that satisfy this requirement, we can state the following general results:

- As speed decreases, the shape of the temporal frequency sensitivity curve asymptotically approaches that for spatial frequency, but it shifts to proportionately lower temporal frequencies.
- As speed increases, the shape of the spatial frequency sensitivity curve asymptotically approaches that for temporal frequency, but it shifts to proportionately lower spatial frequencies.

We confirm these conclusions by computing and plotting the spatial- and temporal-frequency responses for gratings moving at various speeds, as shown in Figure 4.5.

Compare these moving-grating responses to the psychophysical measurements from humans shown in Figure 4.6a. The model reproduces the dependence of the peak frequency on speed. Kelly fitted the horizontal displacement of the peak with speed, over the range from 0.15dps to 32dps, using the expression $\hat{\rho} = 7.3/(v + 2)$. This quantitative relation reveals that the peak spatial frequency is indeed inversely proportional to speed for high speeds, as predicted by the model. This OPL circuit model does not account for the dependence of the peak height on speed; additional

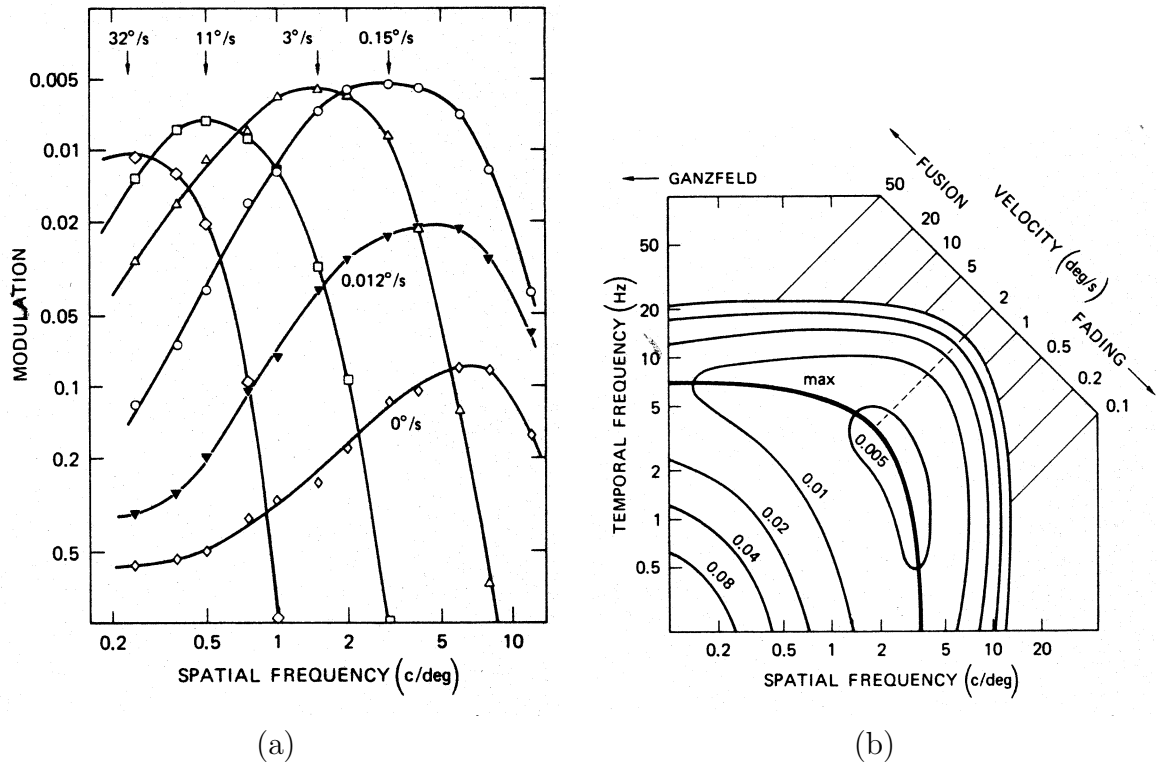


Figure 4.6: HUMAN SENSITIVITY TO MOTION

(a) Response to sinusoidal gratings, moving at six different speeds, versus spatial frequency. The relative amplitude of the modulation of the baseline intensity (i.e., threshold contrast) is plotted; the baseline intensity was 300td. The subject's eye movements were tracked and compensated for in these experiments. The response was always bandpass, and its peak remained at about 5 cpd for speeds below 2 dps. At higher speeds, the peak position shifted to lower spatial frequencies. The peak amplitude also changed, rising rapidly initially, and then falling slowly above 2 dps. Reproduced from [89]. (b) Contour plot of spatiotemporal-contrast-threshold surface (same as in Figure 3.5b). Sensitivity doubles from one contour to the next. The heavy line represents the maximum sensitivity at each velocity; a velocity axis is included on the upper right. The surface is roughly symmetric about the line $v = 2$ dps. Reproduced from [89].

spatial and temporal filtering in the inner retina could account for the dependence of the height of the peak on speed.

The model also predicts that the peak temporal frequency is proportional to speed for low speeds. This behavior is also evident in the psychophysical data (Figure 4.6), although Kelly did not plot his data versus temporal frequency. The curves for 0dps, 0.012dps, and 0.15dps peak at about 6cpd, 4cpd, and 3cpd, respectively. We obtain the temporal frequencies produced by these moving gratings by multiplying spatial frequency by speed, which gives 0cps, 0.048cps, and 0.45cps, respectively. Hence, the peak temporal frequency is roughly proportional to speed.

In summary, a simple linear two-layer feedback model:

- Accounts for the dependence of the peak spatial and temporal frequencies on speed observed for high and low speeds, respectively, in human psychophysics.
- But it does not reproduce the dependence of the peak height on speed.
- This shortcoming is, most likely, because the model does not include highpass temporal and spatial filtering in the inner retina.

This description of the locus of the peak position completes my discussion of the model's sensitivity to stimuli used in classic psychophysical and physiological experiments. To understand exactly how these responses arise, and to extend the description to arbitrary dynamic patterns, we must turn to the complete three-dimensional spatiotemporal transfer function.

4.4 Spatiotemporal Sensitivity

Plots of the magnitude and phase of the cone's spatiotemporal-frequency transfer function, $\tilde{H}_c(\rho, \omega)$, are shown in Figure 4.7. The function is more or less symmetric about the 45° degree axis, because interchanging ρ^2 and $i\omega$ in Equations 4.3 and 4.4 produces homomorphic equations. Consequently, everything that we say about spatial frequency with respect to temporal frequency is still true when the two

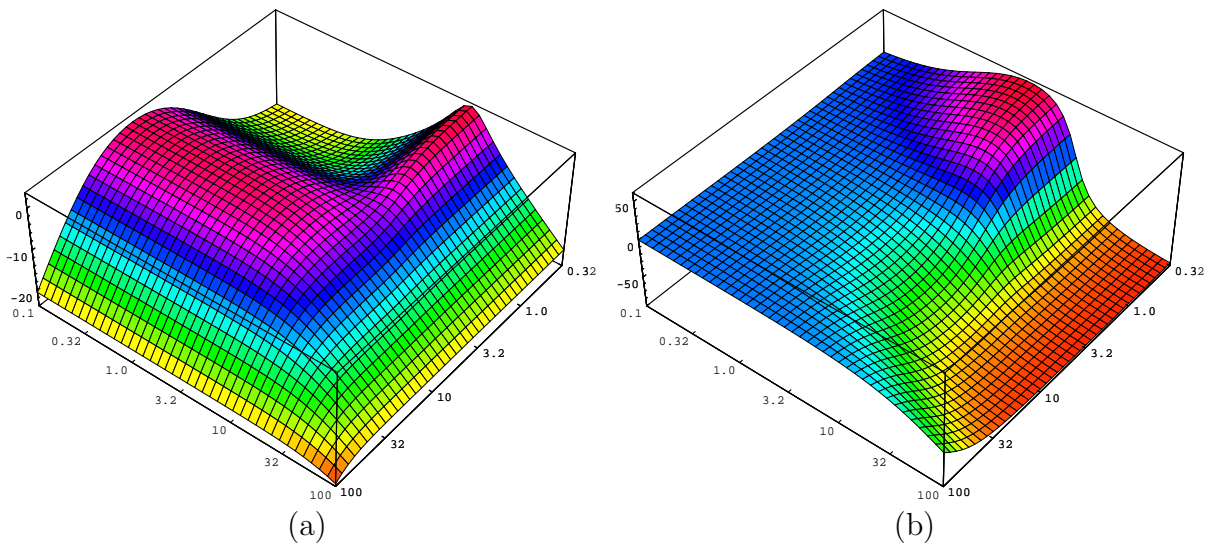


Figure 4.7: SPATIOTEMPORAL SENSITIVITY OF CONES

Three-dimensional plots showing (a) magnitude and (b) phase of cone responses from OPL circuit model versus spatial and temporal frequency. At higher spatial frequencies, the low-frequency temporal sensitivity increases, and vice versa. The phase is positive for low temporal frequencies, negative for high temporal frequencies and tends to 0 at high spatial frequencies.

are interchanged. Bear in mind this duality as we discuss the salient features of the spatiotemporal frequency responses.

I begin the discussion by taking the cross-sections of these surfaces that correspond to the flicker and grating curves presented in Figure 4.3a,b and Figure 4.4a,b. The flicker and grating curves are defined by the intersections of the spatiotemporal surface with the $\rho = 0$ and $\omega = 0$ planes, respectively. As expected, the amplitude response (Figure 4.7a) is bandpass at these planes.

As we move the spatial-frequency plane away from the $\rho = 0$ plane, to higher spatial frequencies, we observe increasingly strong responses to low temporal frequencies. Similarly, as we move the temporal-frequency plane away from the $\omega = 0$ plane, to higher temporal frequencies, we observe increasingly strong responses to low spatial frequencies. Thus, the filters for temporal frequency and for spatial frequency are letting through more low-frequency energy—they are becoming less bandpasslike and more lowpasslike. When the planes reach the peak temporal frequency, $\hat{\omega}$, and the peak spatial frequency, $\hat{\rho}$, the transition is complete, and the filters become purely lowpass, without any peak whatsoever. Both filters remain lowpass as the planes move beyond the peak frequencies.

The model's spatiotemporal frequency sensitivity mirrors that for humans, shown as a family of curves in Figure 3.5a, as a three-dimensional plot in Figure 3.5b, and as a contour plot in Figure 4.6b; and mirrors that of cats, shown as a family of curves in Figure 3.6b. In particular, this simple linear two-layer feedback model shares four salient features with human psychophysics and cat physiology:

1. Spatial filtering is bandpass at low temporal frequencies, and is tuned to a particular spatial frequency $\hat{\rho}$.
2. Temporal filtering is bandpass at low spatial frequencies, and is tuned to a particular temporal frequency $\hat{\omega}$.
3. Spatial filtering becomes lowpass at high temporal frequencies, and tuning disappears completely when the temporal frequency exceeds $\hat{\omega}$.

4. Temporal filtering becomes lowpass at high spatial frequencies, and tuning disappears completely when the spatial frequency exceeds $\hat{\rho}$

So far, we have discussed how the model's transfer function modulates the amplitude of the input spatiotemporal sinusoid. Let us now consider the phase shift introduced by the transfer function.

As usual, the flicker and grating curves are defined by the intersections of the spatiotemporal surface with the $\rho = 0$ and $\omega = 0$ planes, respectively (see Figure 4.7). At the $\rho = 0$ plane, the response leads for frequencies below the peak temporal frequency, $\hat{\omega}$; it lags for frequencies above $\hat{\omega}$; and the phase decreases rapidly around $\hat{\omega}$, passing through 0 at $\hat{\omega}$. The behavior at the $\omega = 0$ plane also is as expected, with a phase shift of zero.

As we move the spatial-frequency plane away from the $\rho = 0$ plane, to higher spatial frequencies, the phase lead decreases across the entire $\omega < \hat{\omega}$ region, going towards 0—and even changes to a small phase lag over a small subregion near $(\hat{\rho}, \hat{\omega})$. The phase lag also decreases across the entire $\omega > \hat{\omega}$ region, going towards 0. Both transitions appear progressively, occurring at different spatial frequencies for different temporal frequencies—the point at which the leads and lags disappear is roughly proportional to the temporal frequency.

Rapid changes in phase occur roughly along a horizontal line defined by $\omega = \hat{\omega}$, and along a diagonal line defined $\omega = (\hat{\omega}/\hat{\rho})\rho$. (The contour plot in Figure 4.8b shows this clearly.) Taken together, these lines divide the phase plot into three distinct regions. For frequencies below the diagonal line, the phase is close to 0. For frequencies above the diagonal line and below the horizontal line, there is a large phase lead. And for frequencies above both lines, there is a large phase lag.

In summary, the model's transfer function, with respect to phase shift, has three salient features:

1. There is no phase shift when the spatial frequency is above $(\omega/\hat{\omega})\hat{\rho}$ or the temporal frequency is below $(\rho/\hat{\rho})\hat{\omega}$.
2. A large phase lead occurs when the temporal frequency is below $\hat{\omega}$ and the

spatial frequency is below $(\omega/\hat{\omega})\hat{\rho}$.

3. A large phase lag occurs when the temporal frequency is above $\hat{\omega}$ and the spatial frequency is below $(\omega/\hat{\omega})\hat{\rho}$.

Unfortunately, the detailed phase characteristics are not available for the biological systems, so we cannot make a comparison.

A theme that unifies the amplitude and phase characteristics is the dependence of spatial filtering on temporal frequency, and vice versa. In this model, spatial filtering cannot be separated from temporal filtering: $\tilde{H}_c(\rho, \omega) \neq H_x(\rho)H_t(\omega)$. That is, we cannot realize the filtering performed by the model by cascading a spatial filter, $H_x(\rho)$, with a temporal filter, $H_t(\omega)$. This **spatiotemporal inseparability** arises because the same elements in the circuit are used to perform both spatial filtering and temporal filtering. A purely spatial filter cannot have any time dependencies in its wires; a purely temporal filter cannot have any crosstalk with its neighbors. The OPL model adheres to neither of these edicts: Signals are spread out in time as they are spread out in space—and vice versa—as they are passed from place to place by the internode conductances and as they are passed from time to time by the node capacitances.

4.5 Responses to Moving Images

To understand how the OPL model responds to motion, we will find it most instructive to display the three-dimensional spatiotemporal-frequency transfer function as a contour plot, and to superimpose the input spectrum onto this plot. The transfer function is replotted in this fashion in Figure 4.8; observe the similarity between this plot and the contour plot of human spatiotemporal contrast sensitivity in Figure 4.6b.

The speed $\hat{v} \equiv \hat{\omega}/\hat{\rho}$, given by the ratio between the peak temporal frequency and the peak spatial frequency, plays a decisive role in the model. I call this speed the **pivotal speed**, because it demarcates the border between the low-speed region, where motion produces higher temporal frequencies, and the high-speed region, where

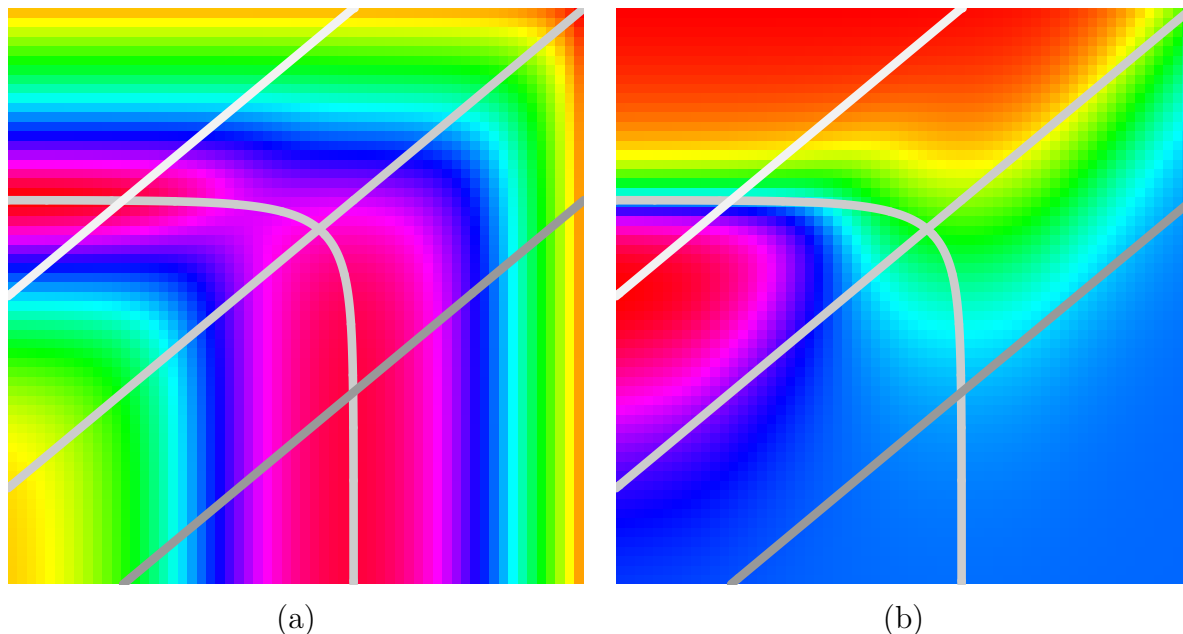


Figure 4.8: MOTION AND SPATIOTEMPORAL SENSITIVITY OF CONES

Color-coded contour plots showing amplitude (a) and phase (b) of the cone responses from the OPL circuit model. A cyclic color encoding was used, starting with red at the bottom end of the scale, and coming back to red at the top. (a) The amplitude plot looks like a mountain range with a sharp bend in it; the L-shaped ridge that runs along the top is shown. (b) The phase plot looks like a plain in the area below the $\omega = \rho \hat{v}$ line, where \hat{v} is the ratio between the peak temporal frequency, $\hat{\omega}$, and the peak spatial frequency, $\hat{\rho}$. The area above this line is shared by a mountain and a valley, with the $\omega = \hat{\omega}$ line demarcating the border between them. Each diagonal bar in these plots is the support of the input image's spectrum, for uniform translation at a different speeds. The support is defined by the line $\omega = \rho v$, where v is the speed. On $\log(\omega)$ - $\log(\rho)$ scales, this line is always at an angle of 45° ; it shifts to higher temporal frequencies as speed increases. The three speeds shown are $0.1\hat{v}$, \hat{v} , and $10\hat{v}$. The intersection of these diagonal lines with the spatiotemporal surface gives the sensitivity of the model to moving gratings, and produces the curves plotted in Figure 4.5a,b, when projected onto the temporal-frequency axis or onto the spatial-frequency axis, respectively.

motion produces lower spatial frequencies. These two distinct behaviors arise because the line $\omega = \rho\hat{v}$ bisects the L-shaped ridge of the amplitude plot into two arms, one running horizontally and the other running vertically; the ridges takes the corner right at the pivotal speed line, $\omega = \rho\hat{v}$.

At speeds below \hat{v} , the spectrum intersects the vertical arm, and the locus of the peak remains at the same spatial frequency $\hat{\rho}$, but it moves to higher temporal frequencies $\hat{\rho}v$ with increasing speed, v . At speeds above \hat{v} , the spectrum intersects the horizontal arm, and the locus of the peak remains at the same temporal frequency $\hat{\omega}$, but it moves to lower spatial frequencies $\hat{\omega}/v$ with increasing speed, v . This picture explains the responses obtained for moving gratings plotted in Figure 4.5a,b.

We can derive the locus of the peak by differentiating $|\tilde{H}_c(\rho, \rho v)|$ with respect to ρ and setting the derivative to 0 to find the maximum. However, a simple approximate expression describes the locus of the peak, $(\hat{\rho}(v), \hat{\omega}(v))$, quite well.

$$\frac{\hat{\rho}(v)}{\hat{\rho}(0)} + \frac{\hat{\omega}(v)}{\hat{\omega}(\infty)} = 1,$$

where $\hat{\rho}(0) = \hat{\rho}$ is the peak spatial frequency for the stationary-grating sensitivity curve, and $\hat{\omega}(\infty) = \hat{\omega}$ is the peak temporal frequency for the full-field-flicker sensitivity curve. This expression accounts for the peak position in both the high-speed and the low-speed regions. To locate the peak on the temporal-frequency axis, we replace $\hat{\rho}(v)$ with $\hat{\omega}(v)/v$; to locate the peak on the spatial-frequency axis, we replace $\hat{\omega}(v)$ with $\hat{\rho}(v)v$. Making these substitutions gives

$$\begin{aligned}\hat{\omega}(v) &= \frac{v}{\hat{v} + v}\hat{\omega}(\infty), \\ \hat{\rho}(v) &= \frac{\hat{v}}{\hat{v} + v}\hat{\rho}(0),\end{aligned}$$

where $\hat{v} \equiv \hat{\omega}(\infty)/\hat{\rho}(0)$ is the speed at which the behavior crosses over from the low-speed regime to the high-speed regime.

With the aid of these contour plots, it is easy to see how the model's spatiotemporal inseparability shapes the model's response to moving gratings. The amplitude

of the response to a grating changes when that grating flickers or moves, because the gain of the spatial filter depends on temporal frequency. For temporal frequencies below the peak temporal frequency, $\hat{\omega}$, this dependence makes the response increase with flicker rate or with speed—except when the spatial frequency of the grating is equal to the peak spatial frequency, $\hat{\rho}$. For this exceptional situation, where the spatial filter is tuned to the spatial frequency of the grating, the response becomes speed invariant for temporal frequencies below $\hat{\omega}$ (i.e., speeds below \hat{v}). In general, however, the response of the OPL model is not speed invariant.

The phase of the grating response also changes drastically with increasing speed, and the change is nonmonotonic. The phase starts increasing after the speed exceeds the pivotal speed, \hat{v} , reaches a maximum, and then starts to decrease, reaching 0 when the speed is equal to $\hat{\omega}/\rho$, where ρ is the spatial frequency of the grating. However, when the spatial frequency of the grating is equal to the peak spatial frequency, $\rho = \hat{\rho}$, the phase does not change with speed—except for a small range around \hat{v} , where a small lag occurs.

4.5.1 Speed-Invariant Contrast Estimation

The model predicts that the amplitude and phase of the outer retina’s response depends on the speed of the moving grating. Speed-dependent responses give rise to the question: Can a speed-invariant estimate of contrast be obtained from the outer retina’s output signals?

If the input pattern has a broad spatial-frequency spectrum (e.g. an impulse, an edge, or a random-dot pattern), we can get a flicker- or speed-invariant estimate of contrast by measuring the energy at the spatial frequency to which the spatial filter is tuned. Since the response phase does not change much at this spatial frequency, we can get away with just measuring the peak amplitude. However, this strategy works in only the low-speed regime, where the temporal frequencies generated are below the cutoff point. In the high-speed region, we can obtain a speed invariant estimate of contrast by taking the dual approach.

Due to the dual relationship between spatial filtering and temporal filtering, the response is also speed-invariant when the temporal filter is tuned to the temporal frequency of the spatiotemporal sinusoid, and the spatial frequency is below $\hat{\rho}$ (i.e., speeds above \hat{v}). Thus, for a broadband signal, we can get a flicker- or speed-invariant estimate of contrast in the high-speed region, $v \gg \hat{v}$, by measuring the energy at the temporal frequency, $\hat{\omega}$. Again, since the response phase does not change at this temporal frequency, we can get away with just measuring the peak amplitude.

We should be aware that the circuit generates the energy we are measuring by amplifying energy at the spatial frequency $\rho = \hat{\omega}/v$. Therefore, we extract energy from lower spatial frequencies as speed increases, so the response is speed invariant only when the input energy is distributed uniformly across the spectrum.

In summary, we can obtain a speed-invariant estimate of contrast from the cone's output by proceeding as follows:

- For low speeds, $v \ll \hat{v}$, measure the energy at the spatial frequency, $\hat{\rho}$.
- For high speeds, $v \gg \hat{v}$, measure the energy at the temporal frequency, $\hat{\omega}$.

This strategy assumes that the input energy is distributed uniformly across spatial frequency. Since natural images, and edges, have a $1/\rho^2$ power spectrum, it will be smarter to tailor the algorithm to such a colored spectrum. Additional filtering in the inner retina could achieve this optimization.

To achieve speed-invariance in the transition region between the low-speed regime and the high-speed regime, we must match the peak flicker sensitivity, $\tilde{H}_c(0, \hat{\omega}) = \sqrt{(\tau_h/\tau_c)}Q_t$, to the peak grating sensitivity, $\tilde{H}_c(\hat{\rho}, 0) = (\ell_h/\ell_c)Q_x$. Equating these two expressions, and neglecting the term $\epsilon_h(\ell_c/\ell_h)^2$, we find that

$$\frac{\ell_c}{\ell_h} = \frac{\epsilon_h \tau_c}{2 \tau_h}. \quad (4.5)$$

As we want ϵ_h to be small, to attenuate low frequencies, we must make the space constants of the cone and horizontal cell syncytia disparate, and the time constants of the cone and horizontal cell similar, to satisfy this constraint.

4.5.2 Contrast-Invariant Speed Estimation

I now turn to the question of how to estimate the speed of the motion from the outer-retina's response. This computation is relatively straightforward when the input energy is distributed uniformly across frequency. In this case, the distribution of energy in the output is determined entirely by the intersection of the support of the input spectrum with the model's spatiotemporal-sensitivity surface. Hence, it is easy to see how the motion of such a broadband stimulus is encoded by the model.

The strongest spatial- and temporal-frequency components in the output are selected by the bandpass filtering performed by the model. For low speeds, $v \ll \hat{v}$, a particular spatial frequency, $\hat{\rho}$, is selected, and the energy shifts to higher temporal frequencies, $\omega = \hat{\rho}v$, with increasing speed. For high speeds, $v \gg \hat{v}$, a particular temporal frequency, $\hat{\omega}$, is selected, and the energy shifts to lower spatial frequencies, $\rho = \hat{\omega}/v$, with increasing speed.

In summary, the model predicts that we can obtain a contrast-invariant estimate of speed from the outer retina's response by proceeding as follows:

- For low speeds, $v \ll \hat{v}$, determine which temporal frequency, ω_{\max} , has the most energy, and compute $v = \omega_{\max}/\hat{\rho}$.
- For high speeds, $v \gg \hat{v}$, determine which spatial frequency, ρ_{\max} , has the most energy, and compute $v = \hat{\omega}/\rho_{\max}$.

The most economical way to implement this algorithm would be to use just two broadly-tuned bandpass filters, one tuned to the low end of the range and the other tuned to the high end, and interpolate between these two filters to determine the frequency of the input signal. This two channels may correspond to the magno and parvo pathways [113].

This algorithm for computing speed will not work for an image with a $1/\rho^2$ power spectrum because the bandpass filter is intentionally designed to whiten such a spectrum, and it equalizes the energy for all frequencies in its passband. To make the spatial or temporal frequency tuned in by the outer retina's spatiotemporal bandpass

stand out, we may use a highpass temporal filter or a highpass spatial filter to flatten such natural spectra. This strategy may be used by retina, since both of these highpass filtering operations occur in the inner retina.

4.5.3 Space–Time Effects

I bring my discussion about the outer-retina’s motion responses to a close by leaving the frequency domain to take a look at the response to a moving edge in space–time. The pertinent question is: How does the simple intuitive picture that we have painted in frequency coordinates translate into space–time coordinates?

The response of the model to a moving edge is shown in Figure 4.9, for five different speeds—quarter, half, once, twice, and four times the pivotal speed. I obtained analytical expressions for these responses by taking the inverse-Fourier transform of the model’s frequency-transfer functions. I evaluated and plotted these expressions for a particular choice of parameter values. No antialiasing was performed, so these responses contain energy at all frequencies, out to infinity.

As the speed increases, the response transforms from the spatial response to a static edge, which consists of an overshoot and an undershoot on either side of the edge, to the temporal response to a step input, which consists of an exponentially damped sinusoid that starts after the step occurs. By the time the speed changes by a factor of eight, a complete transformation has occurred, and the response changes from a perfectly symmetric spatial response to a completely assymmetric temporal response.

Due to causality, the temporal component of the response always trails the edge, occurring to the right of the edge in temporal coordinates, or to the left of the edge in spatial coordinates. Only the spatial component of the response can precede the edge; this effect occurs via long range transmission through the tightly-coupled horizontal-cell network. As the horizontal-cell network produces inhibition, the spatial signal gives rise to the overshoot to the right of the edge in space coordinates, or to the left of the edge in time coordinates. It takes time for signals to propagate through the

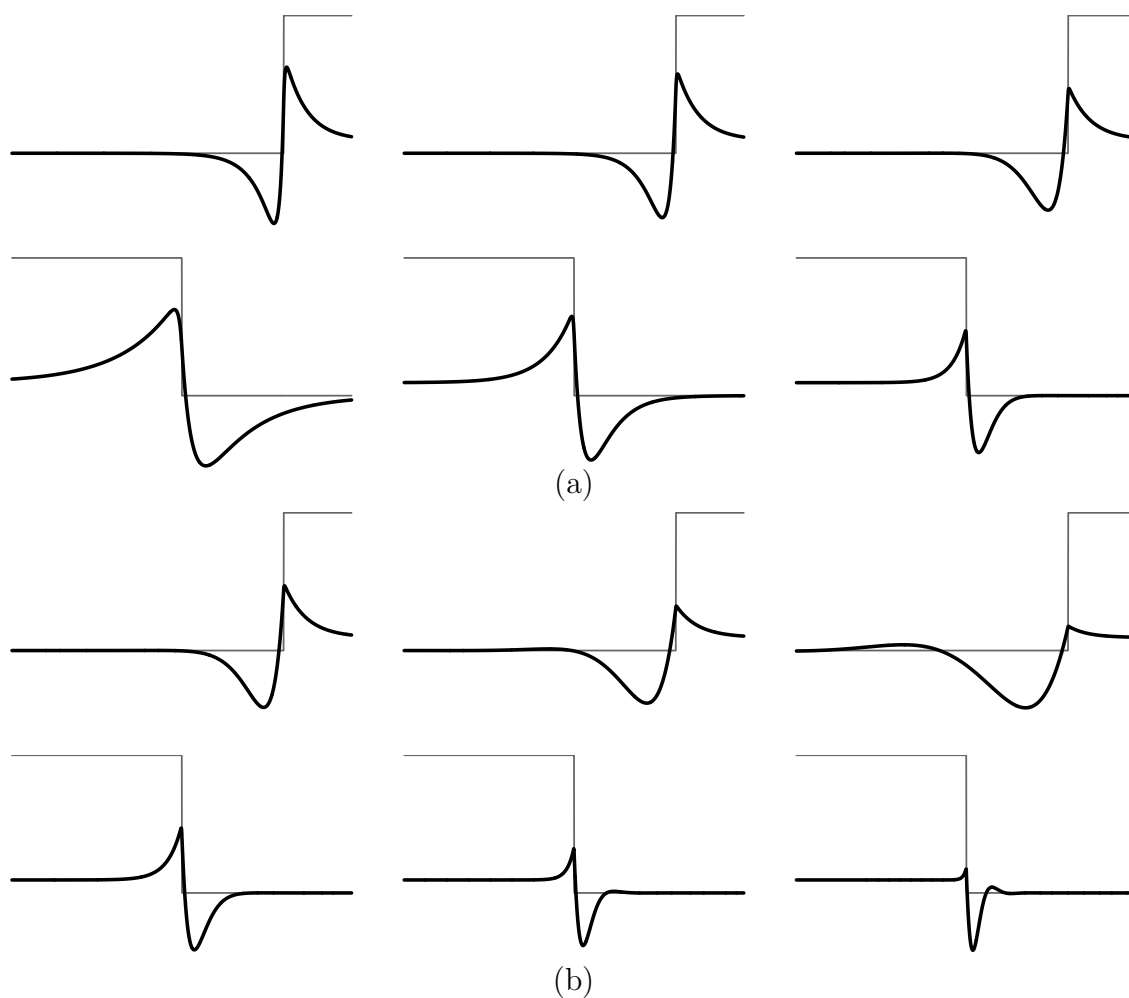


Figure 4.9: RESPONSE OF CONE TO MOVING EDGES

Each panel shows the responses of cones (bold line) to step edges (thin line) moving at three different speeds, obtained from the outer-retina circuit model: (a) Speeds equal to $0.25\hat{v}$, $0.5\hat{v}$, and \hat{v} . (b) Speeds equal to \hat{v} , $2\hat{v}$, and $4\hat{v}$. The input goes from 1 (white) to 0 (black) as the black region invades the white region. In each row, the edge's speed doubles from one graph to the next, going from left to right. In each panel, the top row shows responses plotted versus space, at a particular point in time, and the bottom row shows responses plotted versus time, at a particular location in space. In space coordinates, the edge moves to the right, and in time coordinates, the edge moves to the left.

network, and these inhibitory signals may be overtaken by excitatory signals from the short-range cone network if the edge moves fast enough.

Excitation overtakes inhibition when the speed exceeds the pivotal speed. At the pivotal speed, the time it takes for the edge to transverse the receptive field equals the time it takes for the cone–horizontal-cell feedback loop to settle. Therefore, for speeds below the pivotal speed, the system settles, and the response looks like the static edge response—there is no evidence of temporal behavior. Whereas, for speeds above the pivotal speed the system starts to respond after the edge has passed by, and the response looks like that to a full-field flash—there is no evidence of spatial behavior.

Consequently, two distinct behaviors are observed above and below the pivotal speed. Below the pivotal speed, the response is invariant with speed, when plotted versus position, whereas it is increasingly compressed when plotted versus time (See Figure 4.9a). Hence, the frequency responses look like the stationary grating response, and the energy shifts to higher temporal frequencies with increasing speed, as shown in Figure 4.5a. Above the pivotal speed, the response is invariant with speed, when plotted versus time, whereas it is increasingly drawn out, when plotted versus space (See Figure 4.9b). Hence, the frequency responses look like the full-field flicker response, and the energy shifts to lower spatial frequencies with increasing speed, as shown in Figure 4.5b.

4.6 Discussion

A simple physical model, consisting of two reciprocally-connected diffusive (signal-spreading) layers, captures the qualitative aspects of spatiotemporal filtering in the retina. The model reproduces the dependence of spatial filtering on temporal frequency and the dependence of temporal filtering on spatial frequency. In particular, spatial filtering is bandpass at low temporal frequencies, but becomes lowpass at high temporal frequencies. Conversely, temporal filtering is bandpass at low spatial frequencies, but becomes lowpass at high spatial frequencies.

Models of the retina similar to the one that I study here have been proposed and analyzed. However, none of the previous studies analyzed the effect of the model's spatiotemporal inseparability on motion. By studying a minimal model, and treating space as a continuum—using the continuous approximation—just like time, I was able to obtain closed-form analytic solutions, and to develop a clear intuitive picture of the spatiotemporal behavior of the retina.

I showed that the model's spatiotemporal inseparability has serious consequences for how information about contrast and speed is encoded by the retina. It also results in suboptimal filtering, as the model's spatiotemporal behavior deviates from the optimal filter for the ensemble of natural images.

In following subsections, I show how spatiotemporal inseparability goes hand in hand with local connectivity. As a consequence, nature must choose between a costly spatiotemporally separable optimal filter or a cheap spatiotemporally inseparable suboptimal filter, weighing coding efficiency against implementation efficiency. I also provide a summary of the procedures that I proposed to extract of information about contrast and speed from the outer retina's outputs.

4.6.1 Spatiotemporal Inseparability and Local Connectivity

The interdependence of spatial filtering and temporal filtering is a direct consequence of the locally connected character of the signal-spreading networks. Signals diffuse in space as they are relayed from node to node by the internode conductances. Signals also diffuse in time as they accumulate on the node capacitances. Consequently, the temporal scale on which signals are processed is intimately connected with the spatial scale at which they occur, and vice versa.

Simultaneous spatial and temporal diffusion places a constraint on the sum of the spatial frequency and the temporal frequency. The current spreading through the internode conductances is proportional to the second spatial derivative of the voltage, and the current charging the node capacitance is proportional to the first temporal derivative. Consequently, the sum of the rates at which the signal changes in space and

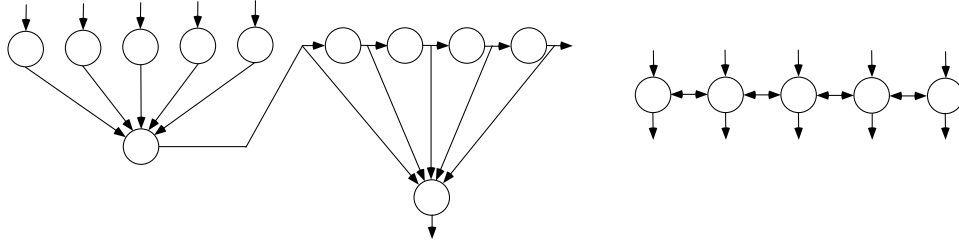


Figure 4.10: HARDWARE FOR SPATIOTEMPORAL FILTERS

Each node has the ability to form a weighted sum of its inputs and to provide a delayed version of the sum to its output; wires communicate signals instantaneously and do not attenuate or amplify them. Left: Separable spatiotemporal filter built from a spatial array and a tapped delay line. This configuration can realize any desired spatiotemporally separable filter. The number of nodes, and the number of wires, required per output is $O(n_r + n_t)$, where n_r is the order of the spatial filter and n_t is the order of the temporal filter. Right: Inseparable spatiotemporal filter built from a nearest-neighbor-connected spatial array. This configuration can realize only those inseparable filters whose spatiotemporal impulse response falls off smoothly. The number of nodes, and the number of wires, required per input is $O(1)$.

in time is constrained by the input current. This constraint translates into a constraint on the sum of the spatial frequency and the temporal frequency. Therefore, all the terms that appear in the transfer function of a locally-connected network involve sums of spatial frequency and temporal frequency, instead of products.

Spatiotemporal separability requires a multiplicative interaction between spatial frequency and temporal frequency, not a subtractive one. To obtain a multiplicative interaction, we must put a constraint on the product of spatial frequency and temporal frequency. Such a constraint produces frequency-sensitivity plots with contours running diagonally (for log-log coordinates), as shown in the frequency-response plot for the optimal filter (Figure 3.3b). In contrast, a sum constraint produces L-shaped contours (for log-log coordinates), as shown in the frequency-response plot of the model (Figures 4.7 and 4.8), and in the frequency response plot for humans (Figure 4.6b).

4.6.2 Efficient Coding Versus Efficient Implementation

A spatiotemporally inseparable filter cannot match the spectrum of the ensemble of natural images, which is more or less separable, over the entire range of spatial and temporal frequencies. It is possible to match the model's inseparable response to the separable response of the optimal filter, but in only certain restricted regions of the spatiotemporal frequency spectrum. We can achieve optimal spatial filtering at low temporal frequency and optimal temporal filtering at low spatial frequency. However, when matched to the optimal filter in these regions, the model does not filter out signals with poor SNR that occur at high spatial frequencies and high temporal frequencies. This mismatch between the model and the optimal filter predicts that the outer retina devotes more of its channel capacity to noise than is optimal.

Suboptimal outer-retina performance is a small price to pay for efficient implementation. The amount of hardware required to implement these two classes of filters—one separable and the other inseparable—is shown in Figure 4.10. By using nearest-neighbor connections, the inseparable network can share wires and nodes, and thus can use the hardware efficiently. A factor of $n = n_r + n_t$, where n_r (n_t) is the order of the spatial (temporal) filter, reduction in hardware translates to a similar reduction in pixel size and to a similar reduction in power consumption. These improvements in efficiency allow smaller and faster pixels to be used, increasing the spatiotemporal bandwidth of the retina.

Additional spatiotemporal filtering in the inner retina may compensate for suboptimal behavior of the outer retina, such that excessive noise in the latter's output is not passed on to the optic nerve and transmitted all the way to the brain. Possibly, this result is achieved by the presence of two channels, with one tuned to low temporal frequencies and high spatial frequencies (parvo pathway), and the other is tuned to high temporal frequencies and low spatial frequencies (magno pathway); neither one is tuned to the noisy signals that occur at high temporal frequencies and high spatial frequencies.

4.6.3 Encoding of Contrast and Speed of Moving Images

Understanding how the outer retina responds to motion led me to develop a natural set of procedures for obtaining a speed-invariant estimate of contrast and a contrast-invariant estimate of speed from the outer-retina's output signals.

In particular, there is a pivotal speed that demarcates the border between two distinct regimes. An edge moving at the pivotal speed sweeps across the receptive field in exactly the time it takes for the system to settle. Below the pivotal speed, the response is dominated by energy at the spatial frequency to which the spatial bandpass is tuned, and this energy moves to higher temporal frequencies as the speed increases. Above the pivotal speed, the response is dominated by energy at the temporal frequency to which the temporal bandpass is tuned, and this energy moves to lower spatial frequencies as the speed increases.

We can estimate contrast by measuring the amplitude of the response at the frequencies to which the bandpass spatial filter and the bandpass temporal filter are tuned, and taking the larger value. To see why this strategy works, we recall that the response is invariant when the sum of the temporal frequency and the spatial frequency is constant. So, by guaranteeing that the spatial frequency is high and the temporal frequency is low, we ensure that changing the temporal frequency has a negligible effect, making the response insensitive to speed. Making the temporal frequency high and the spatial frequency low also works, for the same reason.

We can estimate speed by finding the dominant spectral component, and taking the ratio between that component's temporal and spatial frequencies. In the low-speed regime, the dominant component occurs at the frequency where the spatial bandpass peaks. Hence, we already know the spatial frequency, and we need to determine only the temporal frequency. This strategy is analogous to using the spatial extent of the receptive field as a reference, and measuring the time it takes for the stimulus to cross the receptive field. In the high-speed regime, the strongest spectral component occurs at the temporal frequency where the temporal bandpass peaks. Hence, we already know the temporal frequency, and we need to determine only the spatial frequency.

This strategy is analogous to using the temporal extent of the receptive field as a reference, and measuring how far the stimulus travels during that time.

The advantage of this biomorphic motion algorithm is that it uses signals that occur at either the same location or at the same time—unlike other motion algorithms, which try to match up signals that occur at different locations at different times [114, 115, 116]. In general, this **correspondence problem** is difficult to solve, since there are many candidate matches and the correct one can be found only if the features within the field of view are sufficiently distinct to disambiguate. Note that the algorithm proposed here computes only speed—unlike these more general motion algorithms, which compute direction as well. We can use information about speed, however, to eliminate candidate matches, making the correspondence problem more tractable.