Proc. R. Soc. Lond. B 239, 129–161 (1990)
Printed in Great Britain

A model for the estimate of local image velocity by cells in the visual cortex

BY NORBERTO M. GRZYWACZ[1] AND A. L. YUILLE[2]

[1] *Center for Biological Information Processing, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, E25–201, Cambridge, Massachusetts* 02139, *U.S.A.*

[2] *Harvard University Division of Applied Sciences, G12e Pierce Hall, Cambridge, Massachusetts* 02138, *U.S.A.*

(*Communicated by H. B. Barlow, F.R.S. – Received* 8 *September* 1989)

Some computational theories of motion perception assume that the first stage en route to this perception is the local estimate of image velocity. However, this assumption is not supported by data from the primary visual cortex. Its motion sensitive cells are not selective to velocity, but rather are directionally selective and tuned to spatio-temporal frequencies. Accordingly, physiologically based theories start with filters selective to oriented spatio-temporal frequencies. This paper shows that computational and physiological theories do not necessarily conflict, because such filters may, as a population, compute velocity locally. To prove this point, we show how to combine the outputs of a class of frequency tuned filters to detect local image velocity. Furthermore, we show that the combination of filters may simulate 'Pattern' cells in the middle temporal area (MT), whereas each filter simulates primary visual cortex cells. These simulations include three properties of the primary cortex. First, the spatio-temporal frequency tuning curves of the individual filters display approximate space-time separability. Secondly, their direction-of-motion tuning curves depend on the distribution of orientations of the components of the Fourier decomposition and speed of the stimulus. Thirdly, the filters show facilitation and suppression for responses to apparent motions in the preferred and null directions, respectively. It is suggested that the MT's role is not to solve the aperture problem, but to estimate velocities from primary cortex information. The spatial integration that accounts for motion coherence may be postponed to a later cortical stage.

1. INTRODUCTION

The assumption that the visual system estimates velocity locally is central to some computational theories for visual motion perception (Hildreth 1984; Yuille & Grzywacz 1988*a*, *b*). These theories, and others, combine these estimates spatially (Hildreth 1984; Yuille & Grzywacz 1988*a*, *b*; Bulthoff *et al.* 1989) and temporally (Grzywacz *et al.* 1989) to explain coherent motion percepts and to solve the aperture problem (Marr & Ullman 1981; Adelson & Movshon 1982; Hildreth 1984). (This problem is the impossibility to measure locally, velocity components other than that parallel to the luminance gradient.) The velocity-estimate assumption is supported by the high precision with which humans estimate motion direction

(Levinson & Sekuler 1976) and speed (McKee 1981; McKee & Nakayama 1984; McKee *et al.* 1986).

However, motion-sensitive cells in the primary visual cortex do not detect velocities, but rather are directionally selective and tuned to spatio-temporal frequencies (for reviews on cortical motion analysis see Nakayama (1985); Maunsell & Newsome (1987); Andersen & Siegel (1989)). One can roughly decompose these cells' spatio-temporal tuning curves into the product of separate spatial and temporal frequency responses (Ikeda & Wright 1975; Tolhurst & Movshon 1975; Holub & Morton-Gibson 1981) (although this decomposition does not hold in the cat's Area 18) (Bisti *et al.* 1985; Galli *et al.* 1988). Further evidence against velocity selectivity and for frequency tuning (Movshon *et al.* 1980) is the dependency of single cells' directional tuning on stimulus shape (Hammond 1979, 1981) and speed (Hammond & Reck 1981).

Accordingly, physiologically motivated theories use directionally selective frequency tuned filters (Poggio & Reichardt 1976; Adelson & Bergen 1985; Watson & Ahumada 1985) of which, the spatio-temporally oriented are the most relevant for this paper (motion energy filters, Adelson & Bergen (1985)). The main computational motivation underlying the use of spatio-temporally oriented filters is that image motion is characterized by orientation in space-time (Fahle & Poggio 1981; Adelson & Bergen 1985). There is evidence that cells in the primary visual cortex detect such orientation (Emerson *et al.* 1987*a*, *b*; McLean *et al.* 1987). Furthermore, it has been shown that spatio-temporally oriented models correctly predict the facilitation and suppression for responses to movement in the preferred and null directions, respectively (Emerson *et al.* 1987*a*). (The preferred direction is the one for which a stimulus elicits the maximal response from a cell. On the other hand, the null direction is the one yielding the weakest response.)

An additional step necessary for the success of these 'physiological' theories is to compute image velocities. Heeger (1987) presented an elegant model that computes velocities through the spatio-temporal integration of the outputs of Gabor motion energy filters (Gabor 1946; Daugman 1985). Unfortunately, there is no computational rationale to integrate these outputs as in his model, and the model assumes that the image's power spectrum is flat; this is often incorrect. Another model that uses directionally selective frequency-tuned filters to compute the component of velocity in the direction orthogonal to oriented structure in the image has been proposed (Fleet & Jepson 1989). However, this model does not compute the full velocity vector.

We introduce a method to estimate local velocity from the outputs of motion-energy filters that is correct for any pure translation. These local velocity estimates allow one to use computationally motivated schemes for the spatial (Hildreth 1984; Yuille & Grzywacz 1988*a*,*b*; Bulthoff *et al.* 1989) and temporal (Grzywacz *et al.* 1989) integration of velocities. Also, this locality may be critical for the system's ability to detect motion boundaries in a natural way (§6). This paper proves theorems on the velocity estimates and shows that the method is generally consistent with cortical physiology. In particular, it is shown that each filter shares three properties with cells in the primary visual cortex: approximate spatio-temporal separability; directional tuning dependency on stimulus shape and

speed; facilitation and suppression for preferred- and null-direction motions, respectively. Moreover, one can wire up the filters' outputs to new cells such that they are velocity selective and consistent with MT 'Pattern' cells (Movshon *et al.* 1985; Rodman & Albright 1989). This suggests that the Pattern cells may correspond to the velocity selective cells found in MT (Newsome *et al.* 1983; see §6). The circuitries that we describe use the wide bandwidth of the primary cortex temporal tuning curves to their advantage; (wide, that is compared with the relatively well-tuned spatial tuning curves). Also, these circuitries profit from the inverse relation between the optimal spatial frequencies and receptive field size.

The intention here is not to fit cortical data in precise detail, but to propose a general theory of early visual computations that accounts qualitatively for several critical experiments. In particular, the new model makes two assumptions that are 'dubious' in physiological details: Gabor models for receptive fields (Daugman 1985; Heeger 1987) and the square of the output of the filters as cells' responses (Adelson & Bergen 1985; Heeger 1987). To achieve generality, the effects of relaxing these assumptions are discussed.

The organization of the sections of this paper is as follows: §2 introduces our method, proves that it estimates local image velocities for general translations correctly, and suggests neural circuitries to implement the new model. Section 3 compares the new method with Heeger's (1987). Sections 4 and 5 compare the behaviour of the method with that of the primary visual cortex and MT, respectively. Finally, §6 discusses the possible implications of the results of this paper with particular emphasis on computation and physiology.

The material in this paper has previously been presented as an abstract (Yuille & Grzywacz 1989*a*).

2. The model

The model has two stages. The first measures motion energies (the output of motion-energy filters) and the second estimates velocity from these energies. This section starts with the description of the first stage (§2.1). Then, in §2.2, we analyse the distribution of motion energies over the filters for image translations. Finally, strategies and neural implementations for the estimate of velocity from this distribution are presented (§2.3).

2.1. *Description of the model*

The starting point for the model (figure 1) is the observation that image motion is characterized by orientation in space–time (Fahle & Poggio 1981; Adelson & Bergen 1985).

Adelson & Bergen (1985) suggested that one can detect this orientation with spatio-temporally oriented filters. Such a filter is

$$F(\boldsymbol{x}, t : \Omega, \boldsymbol{n}, \Omega_t, \sigma, \sigma_t) = \frac{1}{(2\pi)^{\frac{3}{2}}\sigma^2(\sigma_t)} \exp\left(-\frac{|\boldsymbol{x}|^2}{2\sigma^2}\right) \times \exp(-i\Omega \boldsymbol{n}\cdot\boldsymbol{x}) \exp\left(-\frac{t^2}{2\sigma_t^2}\right) \exp(-i\Omega_t t), \quad (2.1.1)$$

where $\boldsymbol{x}$ and t are a spatial location in the image and time, respectively, $\sigma > 0$,

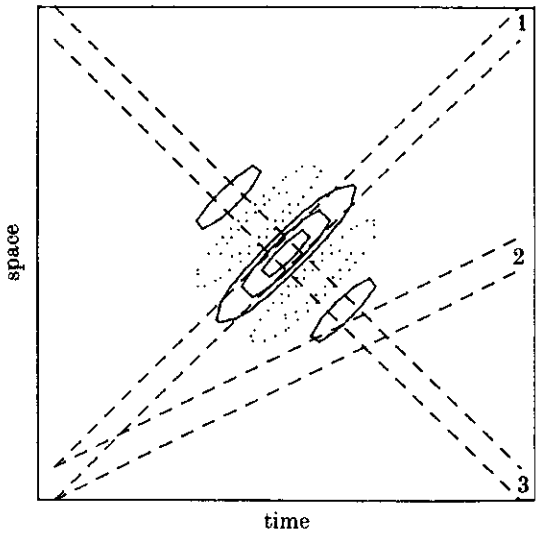

FIGURE 1. Image translation as space–time orientation and its detection by spatio-temporally oriented filters. The figure illustrates the contour plot of a cosine Gabor filter (equation (2.1.2)) and three velocities of a slit of light. Velocity 1 would elicit the maximal response from the filter, as it crosses its central positive portion. On the other hand, velocity 2 is too slow and fails to elicit a strong response from the filter. Finally, velocity 3 has the right speed but the wrong direction, thus crossing negative portions of the filter; (——), positive; (·····), negative.

$\sigma_t > 0$, Ω, and Ω_t are scalar parameters, and $\boldsymbol{n} = (\cos\theta, \sin\theta)$ is a unit vector. For convenience, we will sometimes combine the spatial magnitude Ω and direction $\boldsymbol{n}$ into the vector $\boldsymbol{\Omega} = (\Omega_x, \Omega_y) = \Omega\boldsymbol{n}$. Later in this section, we discuss the physiological meanings of the parameters and variables. This filter is similar to the ones used by Heeger (1987), as its real part is the cosine-phase Gabor filter (Gabor 1946; Daugman 1985):

$$Gab_{\cos}(\boldsymbol{x}, t) = \frac{1}{(2\pi)^{\frac{3}{2}}(\sigma)^2(\sigma_t)} \exp\left(-\frac{|\boldsymbol{x}|^2}{2\sigma^2}\right) \exp\left(-\frac{t^2}{2\sigma_t^2}\right) \cos(\Omega\boldsymbol{n}\cdot\boldsymbol{x} + \Omega_t t), \quad (2.1.2)$$

and its imaginary part is minus the sine-phase Gabor filter:

$$Gab_{\sin}(\boldsymbol{x}, t) = \frac{1}{(2\pi)^{\frac{3}{2}}(\sigma)^2(\sigma_t)} \exp\left(-\frac{|\boldsymbol{x}|^2}{2\sigma^2}\right) \exp\left(-\frac{t^2}{2\sigma_t^2}\right) \sin(\Omega\boldsymbol{n}\cdot\boldsymbol{x} + \Omega_t t). \quad (2.1.3)$$

This filter is oriented in space-time (figure 1).

From equation (2.1.1), we model the responses of directionally selective cells in the primary visual cortex to an image, $I(\boldsymbol{x}, t)$, as the nonlinear filter,

$$N(\boldsymbol{x}, t : \Omega, \boldsymbol{n}, \Omega_t, \sigma, \sigma_t) = |F(\boldsymbol{x}, t : \Omega, \boldsymbol{n}, \Omega_t, \sigma, \sigma_t) * I(\boldsymbol{x}, t)|^2, \quad (2.1.4)$$

where $*$ represents convolution. This definition is similar to the one proposed by Adelson & Bergen (1985), who call it motion energy.

To understand to what aspects of the signal such filters are tuned, it is useful to look at the Fourier transform of equation (2.1.1):

$$\overline{F(\omega, \omega_t : \Omega, \boldsymbol{n}, \Omega_t, \sigma, \sigma_t)} = \frac{1}{(2\pi)^{\frac{3}{2}}} \exp\left(-\frac{(\boldsymbol{n}\cdot\omega - \Omega)^2 \sigma^2}{2}\right) \times \exp\left(-\frac{(\boldsymbol{n}^*\cdot\omega)^2 \sigma^2}{2}\right) \exp\left(-\frac{(\omega_t - \Omega_t)^2 \sigma_t^2}{2}\right), \quad (2.1.5)$$

where the overline stands for Fourier transform, $\boldsymbol{n}^* = (-\sin\theta, \cos\theta)$ is a unit vector orthogonal to $\boldsymbol{n}$, and ω and ω_t are a location in the spatial-frequency domain and temporal frequency, respectively. This equation suggests that the non-linear filter defined in equation (2.1.4) is tuned to sinusoidal gratings that have spatial frequency Ω, travel in the direction $\boldsymbol{n}$, and have temporal frequency Ω_t, with σ and σ_t determining the sharpness of the tunings. Furthermore, these tunings are separable in $\boldsymbol{n}$, Ω, and Ω_t.

Equation (2.1.5) suggests a physiological interpretation for the model as follows. The variable N is the response at time t of a primary visual cortex cell, whose centre of receptive field lies at position $\boldsymbol{x}$ in the image. Alternatively, one may interpret N as the sum of the responses of two cells with the same preferred direction, but whose spatio-temporal profile is 90° out of phase, that is, a quadrature pair (Pollen & Ronner 1981). The sizes of the receptive field and its temporal window are σ and σ_t, respectively. The direction, spatial frequency and temporal frequency of the sinusoidal stimulus eliciting the maximal response are $\boldsymbol{n}$, Ω, and Ω_t, respectively. Following these interpretations, the analyses below do not assume that σ and σ_t are constant for all cells. Cells with large optimal spatial frequency have small receptive field size and vice versa (Hochstein & Shapley 1976; Maffei & Fiorentini 1977; Andrews & Pollen 1979). For generality, and from data appearance (Bisti *et al.* 1985), we also allowed the cells' optimal temporal frequency and temporal window to interdepend. At some stages, the analyses use, with computational advantage, the assumption that the bandwidth of the temporal frequency tuning curves is relatively wide compared with the spatial bandwidth. The assumption states that for all velocities, $\boldsymbol{v}$, to which the cells respond the following relation holds: $(|\boldsymbol{v}|\,\sigma_t)^2 \ll \sigma^2$. Informally, it was verified by literature inspection that typically $3 \leqslant (\sigma/(|\boldsymbol{v}|\,\sigma_t))^2 \leqslant 60$.

Because equation (2.1.5) is separable in $\boldsymbol{n}$, Ω, and Ω_t, it strongly suggests that for moving images, the filter (equation (2.1.4)) is not tuned to any particular velocity. Figure 2, based on calculations presented later in this paper, confirms this suggestion.

One filter cannot estimate the velocity, $\boldsymbol{v}$, but, as we will show in the next section, the set of filters responding most vigorously can.

2.2. *Velocity estimate*

This section shows that the largest responses of the motion-energy filters as a function of their optimal spatial frequency, optimal temporal frequency and optimal direction of motion can determine velocity uniquely. To do so, three

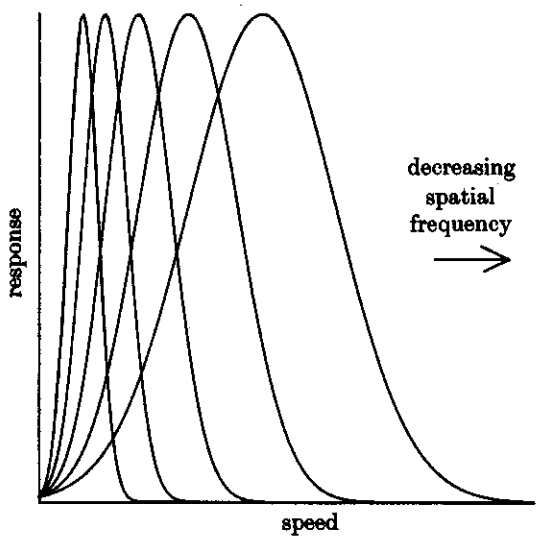

FIGURE 2. Motion-energy filters are not speed selective. The plots are the speed tuning curves for the responses of a motion-energy filter to translating sinusoidal gratings (equation (4.2.2)). These curves' maxima occur at increasingly higher speeds as the gratings' spatial frequency decreases. If the filters were speed selective, all the maxima would occur at the same speed.

theorems and two corollaries are proven. Before each of them, we briefly describe their physiological meaning.

The starting point of these mathematical results is the observation that the spatio-temporal power spectrum of a translating image lies on the plane $\boldsymbol{\omega}\cdot\boldsymbol{v}+\omega_t=0$ in the frequency domain (Watson & Ahumada 1985; Heeger 1987; Daugman 1988). This suggests using the combination of the outputs of cells tuned to specific spatio-temporal frequencies to detect this plane. Our results show how to combine these cells' responses in a computationally sensible way.

The following theorem says that 'if one defines primary visual cortex cells by their optimal temporal frequency and two optimal spatial frequencies (the horizontal and vertical components of the optimal spatial-frequency vector), then in this three-dimensional space, for translations, the maximal responses lie on a known plane.' This result is not a trivial consequence of the knowledge that the spatio-temporal power spectrum of a translating image lies on a plane (Watson & Ahumada 1985; Heeger 1987; Daugman 1988). The plane that the theorem refers to is a plane in the space of the cells' parameters. Actually, one can show that filters other than Gabor filters do not have the same property (this is related to the scale–space theorems; Yuille & Poggio (1986)). The theorem is strictly correct only when the receptive field sizes and temporal windows are constant for all cells. However, in Theorem 3 and its corollary, we show that this constancy requirement can be relaxed under physiological conditions. Figure 3 illustrates the conclusion reached with Theorem 1.

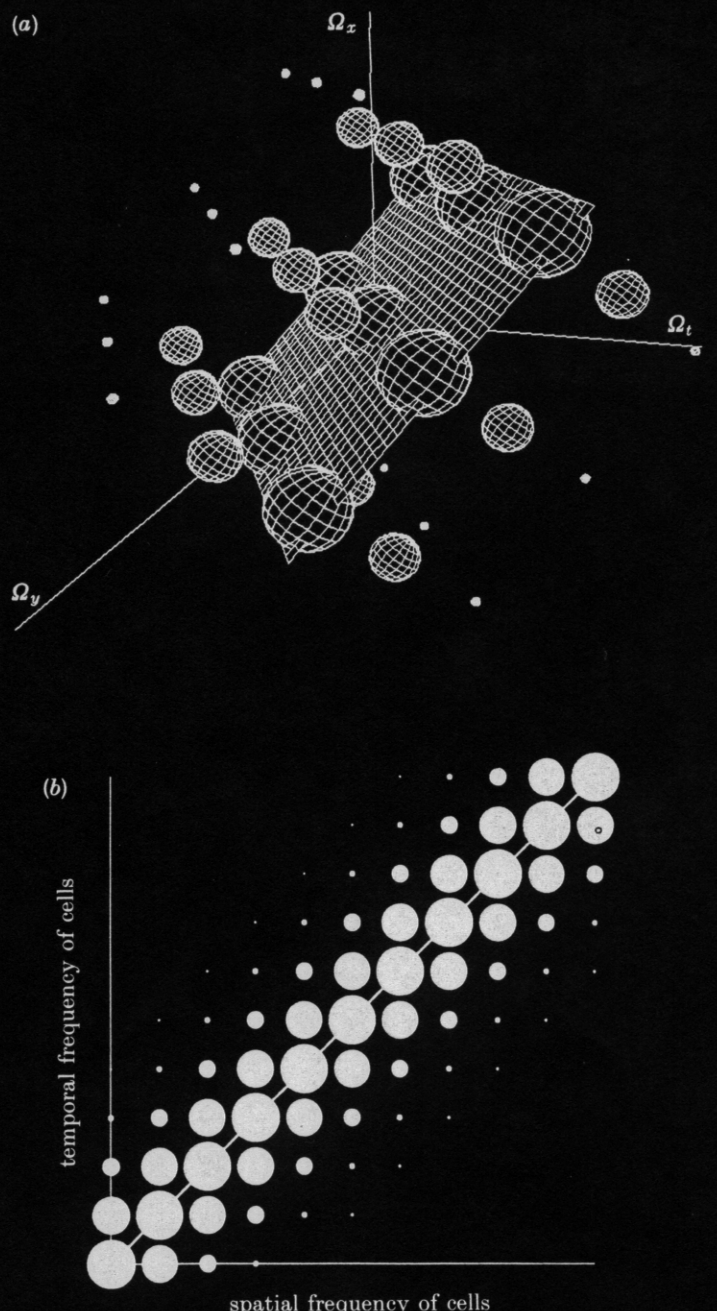

FIGURE 3. Most of the distribution of motion energies lie near the plane $\Omega \boldsymbol{n} \cdot \boldsymbol{v} + \Omega_t = 0$ in the space of the cells' optimal frequencies ($\Omega \boldsymbol{n} = (\Omega_x, \Omega_t)$ and Ω_t). (*a*) This figure shows this distribution for a translating dot (equation (4.2.6)) and indicates the plane where the motion energies (sum of responses of quadrature pairs of directionally-selective frequency tuned cells) are maximal. The motion energies rapidly decrease as the distance of the filters' optimal frequencies from the plane increases; diameter = cell response. (*b*) This figure shows a two-dimensional cross section of a distribution like the one in (*a*).

THEOREM 1. *If σ and σ_t are constants, then the local maxima of $N(\boldsymbol{x}, t: \boldsymbol{\Omega}, \boldsymbol{n}, \Omega_t, \sigma, \sigma_t)$ as a function of $(\boldsymbol{\Omega}, \boldsymbol{n}, \Omega_t)$ lie on the plane $\boldsymbol{\Omega n} \cdot \boldsymbol{v} + \Omega_t = 0$ for all images that move with a constant velocity, $\boldsymbol{v}$*

This result follows from a corollary of a stronger result: Theorem 2. Theorem 2 will provide the response distribution in the three-dimensional space defined by the cells' optimal spatial and temporal frequencies.

THEOREM 2. *The response $N(\boldsymbol{x}, t: \boldsymbol{\Omega}, \Omega_t, \sigma, \sigma_t)$ is weakly separable as follows: a function p exists such that $N(\boldsymbol{x}, t: \boldsymbol{\Omega}, \Omega_t, \sigma, \sigma_t) = p(\boldsymbol{x}, t: \sigma^2 \boldsymbol{\Omega} - \sigma_t^2 \Omega_t \boldsymbol{v}, \sigma, \sigma_t) \exp(-(\sigma_t^2 \sigma^2)(\Omega_t + (\boldsymbol{\Omega} \cdot \boldsymbol{v}))^2 / (2(\sigma^2 + \sigma_t^2 v^2)))$. Hence the only dependence of N on the spatial characteristics of the stimuli occurs within the function p.*

Proof. See Appendix 1.

The following corollary shows that if the receptive-field sizes and temporal windows are constant, then the responses follow a known Gaussian distribution centred on the plane of theorem 1. Thus the claim in Theorem 1 follows from this corollary. Figure 4 illustrates that this Gaussian distribution has a constant orientation relative to the plane, simplifying the plane's search.

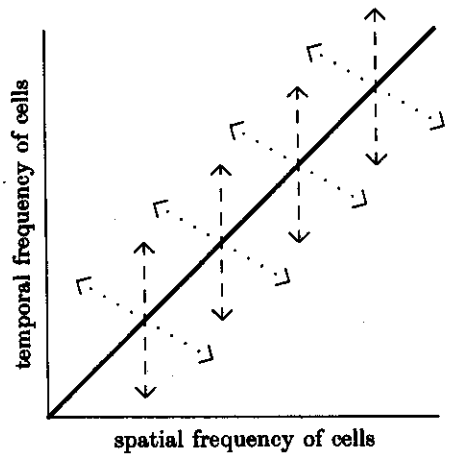

FIGURE 4. Two special cases where the distribution of motion energies in the space of the filters' optimal frequencies is simple. One such case occurs (corollary 1) when the receptive field sizes, σ, and temporal windows, σ_t, are constant. In this case, the motion-energy distribution follows a Gaussian distribution centred on the plane of theorem 1 (the solid line shows the plane's cross section) and with constant orientation relative to the plane. This orientation is not necessarily perpendicular to the plane. Another special case occurs (corollary 2) when σ_t is small in comparison to σ, that is, when $(\sigma_t |\boldsymbol{v}|)^2 \ll \sigma^2$, where $\boldsymbol{v}$ is the plane's velocity. In this case, even if σ_t and σ depend on the filters' optimal frequencies, the motion-energy distribution has a constant orientation: parallel to the temporal frequency axis. Moreover, this distribution has its maximum on the plane. The assumption of wide temporal tuning is approximately correct under physiological conditions; (– –), wide temporal tuning; (···), constant tunings.

COROLLARY 1. *If σ and σ_t are constants, then the variation of $N(\boldsymbol{x}, t: \boldsymbol{\Omega}, \Omega_t, \sigma, \sigma_t)$ in the $(\boldsymbol{\Omega}, \Omega_t)$ space in the direction $((\boldsymbol{v}\sigma_t)/\sigma^2, 1/\sigma_t)$ is a Gaussian function centred on the plane $\boldsymbol{\Omega n} \cdot \boldsymbol{v} + \Omega_t = 0$, and dependent only on $\boldsymbol{v}$, σ, and σ_t.*

Proof. The arguments $(\sigma^2\boldsymbol{\Omega}-\sigma_t^2\Omega_t\boldsymbol{v})$ of the function p do not vary in this direction. The only variation is therefore due to the Gaussian function $\exp(-(\sigma_t^2\sigma^2)(\Omega_t+(\boldsymbol{\Omega}\cdot\boldsymbol{v}))^2/(2(\sigma^2+\sigma_t^2v^2)))$. This Gaussian is centred on the plane $\boldsymbol{\Omega n}\cdot\boldsymbol{v}+\Omega_t=0$.

Theorem 3 asserts: 'If the temporal tuning curves' bandwidths are wide relative to the spatial ones, then irrespective of the receptive field sizes and temporal windows being constant, the response distribution in the optimal-frequency space is of a simple form.' This result is important, because the receptive field sizes and temporal windows may depend on the cells' optimal frequencies (§2.1). We denote these dependencies by $\sigma(\boldsymbol{\Omega})=K(|\boldsymbol{\Omega}|)/|\boldsymbol{\Omega}|$ and $\sigma_t(\Omega_t)=K_t(\Omega_t)/|\Omega_t|$, where K and K_t are functions that are mildly dependent, or perhaps independent, of σ and σ_t, respectively. More precisely, Theorem 3 assumes that for all velocities, $\boldsymbol{v}$, to which the cells respond, $(|\boldsymbol{v}|\,\sigma_t)^2 \ll \sigma^2$. An informal literature study seems to justify this assumption (§2.1).

THEOREM 3. *Given the approximation* $|\boldsymbol{v}|^2 \ll (\sigma/\sigma_t)^2$, *and remembering that* $\sigma_t=\sigma_t(\Omega_t)$ *and* $\sigma=\sigma(\boldsymbol{\Omega})$, *we find that the response* $N(\boldsymbol{x},t:\boldsymbol{\Omega},\Omega_t,\sigma,\sigma_t)$ *is weakly separable in the sense that there exists a function* r, *independent of* Ω_t *and* σ_t, *such that* $N(\boldsymbol{x},t:\boldsymbol{\Omega},\Omega_t,\sigma,\sigma_t)\approx r(\boldsymbol{x},t:\boldsymbol{\Omega},\sigma)\exp(-\sigma_t^2(\Omega_t+(\boldsymbol{\Omega}\cdot\boldsymbol{v}))^2/2)$.

Proof. See Appendix 2.

Corollary 2 shows that in the three-dimensional space of optimal frequencies, the response distributions as function of temporal frequency have maxima on the plane defined in Theorem 1. This means that the overall distribution has a maximal ridge on the plane. Figure 4 illustrates that under the approximation $(|\boldsymbol{v}|\,\sigma_t)^2 \ll \sigma^2$,

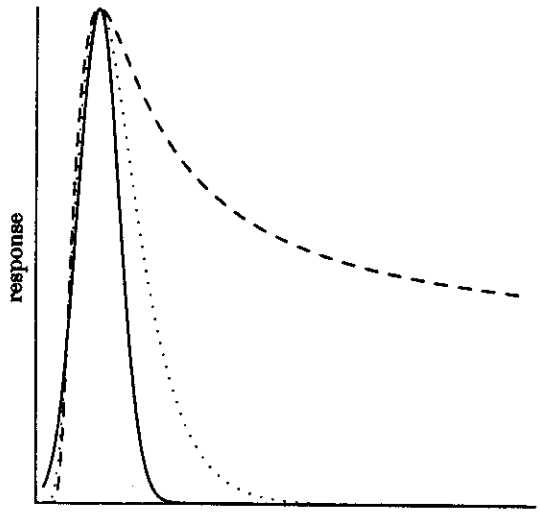

FIGURE 5. The motion-energy distribution when the cells' temporal window, σ_t, depends on their optimal temporal frequency, Ω_t (equation (2.3.1)). This distribution, which is peaked on the translation plane, is only Gaussian when σ_t is constant. Otherwise, for monotonically falling dependencies of σ_t on Ω_t, a positive skewness appears; (——), σ_t = const.; (···), const./Ω_t^2; (—), const./Ω_t.

the distribution of motion energies is oriented parallel to the temporal frequency axis. Figure 5 illustrates the distributions along such a line for different dependencies of σ_t on Ω_t.

COROLLARY 2. *With the same assumptions and approximations as Theorem* 3, *along a one-dimensional line parallel to* Ω_t *axis, the maximum of* $N(\boldsymbol{x}, t: \boldsymbol{\Omega}, \Omega_t, \sigma, \sigma_t)$ *lies on the plane* $\boldsymbol{\Omega} \cdot \boldsymbol{v} + \Omega_t = 0$.

Proof. Consider the set of lines parallel to the Ω_t axis. The only variation of $N(\boldsymbol{x}, t: \boldsymbol{\Omega}, \Omega_t, \sigma, \sigma_t)$ is due to the $\exp(-\sigma_t^2(\Omega_t + (\boldsymbol{\Omega} \cdot \boldsymbol{v}))^2/2)$ term, which is unimodal with maximum at $\boldsymbol{\Omega} \cdot \boldsymbol{v} + \Omega_t = 0$.

2.3. *Strategies and neural implementations for velocity estimate*

We now describe three related methods for finding the velocity of the stimulus by using the mathematical results of the previous section, and discuss possible neural implementations.

The previous section suggests that although primary visual cortex cells are not velocity selective, their population responses may be so. We think of these cells as forming a three-dimensional space defined by their optimal temporal frequency and two optimal spatial frequencies (the $\boldsymbol{\Omega}$ components). If these parameters span sufficiently large ranges, then the cells' strongest responses lie close to the plane $\boldsymbol{\Omega n} \cdot \boldsymbol{v} + \Omega_t = 0$ for an image translating with velocity $\boldsymbol{v}$. In cat, these ranges are five to six octaves large (Holub & Morton-Gibson 1981).

The problem is how to estimate velocity from the combination of the outputs of motion-energy cells (quadrature pairs of directionally selective frequency tuned cells), whose centres of receptive field lie in a single spatial location. This locality may be critical for the system's ability to detect motion boundaries in a natural way. We discuss some computational, psychophysical and implementational aspects of this problem.

Computations

As there are only a finite, though probably very large (Nauta & Feirtag 1986), number of cells, efficient sampling is important (Jasinschi 1988). One cannot, for example, compute derivatives of responses with respect to the filter parameters to determine the velocity from the strongest responses by using Theorem 1. We require, however, that the theory yields the correct velocity as the sampling becomes arbitrarily dense. The principal issues, given the need for sampling, are how to weigh the responses of cells depending on their magnitudes and distance from the origin in the three-dimensional space defined by the cells' optimal frequencies. The magnitude of a cell's response depends on the spatial characteristics of the image as well as on the velocity. For example, one may get a cell far away from the plane responding more strongly than a cell in the plane. From the separability result in Theorem 3, we know that this difference is due to these two cells having different spatial frequencies. Thus it would be unfair to compare them directly. One should prefer to compare cells with similar spatial frequencies. The further away a cell is from the origin, the better the cell is for estimating $\boldsymbol{v}$ (figure 6), so it is desirable that the sampling size in this space grows linearly with distance

from the origin. Data supporting this scaling comes from the decrease of receptive field size when the cells' optimal spatial frequency increases (Hochstein & Shapley 1976; Maffei & Fiorentini 1977; Andrews & Pollen 1979).

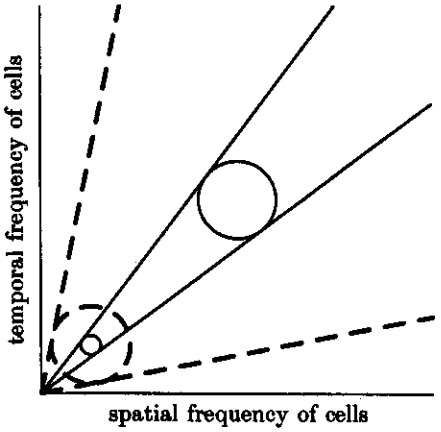

FIGURE 6. It is desirable that the sampling size in the space defined by the cells' optimal frequencies grows linearly with distance from the origin. In this figure, the diameters of the circles represent the frequency bandwidth of motion-energy cells. The angle between the lines starting at the origin and tangential to the circles represents the velocity uncertainty of the motion-energy cells. This is because the velocity plane (Theorem 1) crosses the origin. If all the cells had the same bandwidth, cells near the origin would have large uncertainty (angle between dashed lines). On the other hand, if the bandwidth scaled with distance from the origin, then all cells would have the same uncertainty (angle between solid lines).

Psychophysics

The phenomenon of transparency shows that humans can perceive several velocities at the same point; the theory must be able to deal with this effect. In §6, we describe how the theory may cope with the transparency and the perhaps related problem of motion boundaries. The theory must also deal with the aperture problem at large; i.e. if the image motion is consistent with an infinite set of possible velocities, then the smallest velocity is perceived.

Implementation

Neuronally plausible elements must be the basis of the theory's implementation. To some degree, this means that one should prefer to use computational elements that have been identified by neurobiologists. But most important, neuronal plausibility strongly suggests the *Principle of sloppy workmanship* (Huggins & Licklider 1951; Ratliff 1965). This principle states that neural networks in the nervous system perform very well without relying on the precision of their anatomical details and neural responses (von Neumann 1958). Thus the computational success of such a network should not heavily depend on the network's mathematical details. We argue that this principle suggests a simplicity of connectivity. A 'sloppy-workmanship' developmental process would prefer to wire up the brains' computations with the minimal number of connections to minimize errors.

There are many possible strategies for computing the velocities, given these results. In this paper, we discuss three closely related examples.

The ridge strategy

This strategy uses corollary 2 as a starting point and proposes excitatory connections from each motion-energy cell to the velocity selective cells most consistent with it (figure 7). These connections should have a weak preference for velocities with small components perpendicular to the preferred direction, so as to give a unique answer for the aperture problem in the large. It is a straightforward method that might be favoured by a neural system, because of its robustness.

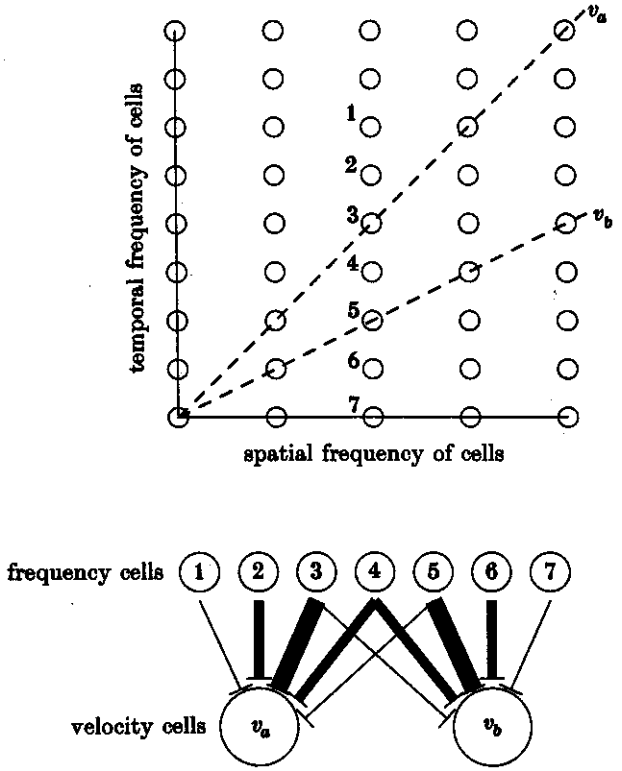

FIGURE 7. The Ridge strategy. In the top part of the figure, the centre of the open circles represent some sampling locations in the space of the cells' optimal frequencies. The diameters of the circles are not bandwidth here. The cross sections of two velocity planes (corresponding to velocities v_a and v_b) are shown and seven motion-energy cells (directionally selective frequency tuned cells) are labelled. In the bottom part of the figure, we show how each of these seven cells make excitatory connections to cells tuned to the velocities v_a and v_b. The number of lines in each connection represents the connection's strength. Motion-energy cells, whose parameters are close to a velocity plane make strong connections to the corresponding velocity cells. Otherwise, if the motion-energy cells are far from the plane, then the connections are weak. To calculate the input's strengths, we used equation (2.3.1). A part of the model not shown here is a winner-take-all mechanism to choose the strongest velocity cell. Also, we do not illustrate the implementation of the solution to the aperture problem in the large (equation (2.3.1)).

Suppose we have a set of M motion-energy cells $(\boldsymbol{\Omega}^\mu, \Omega_t^\mu, \sigma^\mu, \sigma_t^\mu)$ with $\mu = 1, \ldots, M$. A possible implementation is to define the response, $R(\boldsymbol{x}, t : \boldsymbol{v})$, at time t of a velocity selective cell tuned to velocity $\boldsymbol{v}$, and whose receptive field is centred at position $\boldsymbol{x}$, by:

$$R(\boldsymbol{x}, t : \boldsymbol{v}) = A \sum_\mu N(\boldsymbol{x}, t : \boldsymbol{\Omega}^\mu, \Omega_t^\mu, \sigma^\mu, \sigma_t^\mu)\, \mathrm{e}^{-(\sigma_t^\mu)^2 (\Omega_t^\mu + (\boldsymbol{\Omega}^\mu \cdot \boldsymbol{v}))^2/2}\, \mathrm{e}^{-(\boldsymbol{v} \cdot \boldsymbol{\Omega}^{\mu *}/k)^2}, \quad (2.3.1)$$

where $\boldsymbol{\Omega}^*$ is orthogonal to $\boldsymbol{\Omega}$, and A and k are constant parameters.

This equation suggests that the strength of the connection between cell $(\boldsymbol{\Omega}^\mu, \Omega_t^\mu, \sigma^\mu, \sigma_t^\mu)$ and the velocity selective cell tuned to the velocity $\boldsymbol{v}$ should be $\exp(-(\sigma_t^\mu)^2 (\Omega_t^\mu + (\boldsymbol{\Omega}^\mu \cdot \boldsymbol{v}))^2/2) \exp(-(\boldsymbol{v} \cdot \boldsymbol{\Omega}^{\mu *}/k)^2)$.

This method is similar to correlation and template matching methods in computer vision. If we fix $\boldsymbol{\Omega}$ and let Ω_t vary, then from Theorem 3, we know that the form of the variation of the filtered response is $\exp(-\sigma_t^2 (\Omega_t + (\boldsymbol{\Omega} \cdot \boldsymbol{v}))^2/2)$; this defines our template. The largest value of the correlation of this template with $N(\boldsymbol{x}, t : \boldsymbol{\Omega}^\mu, \Omega_t^\mu)$, as we vary the value of $\boldsymbol{v}$ while fixing $\boldsymbol{\Omega}$, gives an estimate for the velocity. To combine the results as $\boldsymbol{\Omega}$ varies, we simply add the magnitude of the responses for each $\boldsymbol{\Omega}$. The factor $\exp(-(\boldsymbol{v} \cdot \boldsymbol{\Omega}^{\mu *})/k)^2$ is designed to prevent the aperture problem in the large (if the image motion is consistent with an infinite set of possible velocities, then the smallest velocity is perceived). The parameter k should be sufficiently large to maintain the validity of the results of §2.3.

Several velocity selective cells will be excited and the one with the largest response corresponds to the velocity estimate. A winner-take-all mechanism (Feldman & Ballard 1982; Koch & Ullman 1985; Yuille & Grzywacz 1989*b*) may then select the maximally responding cell. Such a mechanism is consistent with the inhibition of Pattern cells when stimulated with a motion not parallel to their preferred direction (J. A. Movshon, personal communication). A supralinear dependence (with positive second derivative) of the activation of the velocity selective cells on their input might account for the facilitatory way that directionally selective units appear to interact (J. A. Movshon, personal communication; Ferrera & Wilson 1987).

This method is local, parallel, and instantaneous, gives the right answer for arbitrarily dense sampling, and degrades well as the sampling becomes more sparse.

The estimation strategy

This strategy attempts to estimate the image's spatial characteristics and compute the velocity simultaneously by minimizing a goodness-of-fit criterion. Theorem 3 underlies this strategy, by providing the form of the variation of the motion-energy cells' response distribution in the Ω_t direction. If one places several cells on the same line in this direction, then one can estimate the distribution on this line (figure 8) obtaining measurements for velocity computations. Without such an alignment the problem would be ill-posed, because with only a finite number of cells, there would be insufficient information to estimate the signal and the velocity.

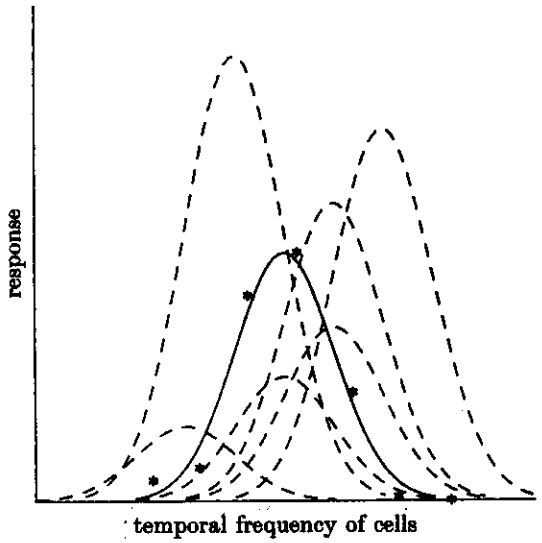

FIGURE 8. The estimation strategy. The figure shows the motion energies of a moving dot sampled by seven motion-energy cells aligned parallel to the temporal frequency axis. The estimation strategy computes the image's spatial characteristics and velocity by finding the amplitude and centre of the motion-energy distribution. The procedure finds the best fit of the expected distribution (theorem 3) to the data; (*), motion energies; (——) correct estimate; (– –), incorrect estimates.

More precisely, from Theorem 3 we know that:

$$N(\boldsymbol{x}, t : \boldsymbol{\Omega}, \Omega_t, \sigma, \sigma_t) \approx r(\boldsymbol{x}, t : \boldsymbol{\Omega})\, \mathrm{e}^{-\sigma_t^2(\Omega_t^2 + (\boldsymbol{\Omega}\cdot\boldsymbol{v}))/2}. \tag{2.3.2}$$

The function $r(\boldsymbol{\Omega})$ is unknown and depends on the form of the image. However it will be constant as we vary Ω_t keeping $\boldsymbol{\Omega}$ constant.

To estimate velocity and the image's spatial characteristics, one can minimize a goodness-of-fit criterion $E(\boldsymbol{v}, r(\boldsymbol{\Omega}))$, both with respect to $\boldsymbol{v}$ and $r(\boldsymbol{\Omega})$, given a set of measurements $N(\boldsymbol{x}, t : \boldsymbol{\Omega}^\mu, \Omega_t^\mu, \sigma^\mu, \sigma_t^\mu)$ for $\mu = 1, \ldots M$. We choose the standard least-squares fit criterion:

$$E(\boldsymbol{v}, r(\boldsymbol{\Omega}), \sigma, \sigma_t) = \sum_\mu (N(\boldsymbol{x}, t : \boldsymbol{\Omega}^\mu, \Omega_t^\mu, \sigma, \sigma_t) - r(\boldsymbol{x}, t : \boldsymbol{\Omega}^\mu)\, \mathrm{e}^{-\sigma_t^2(\Omega_t^\mu + (\boldsymbol{\Omega}^\mu\cdot\boldsymbol{v}))^2/2})^2. \tag{2.3.3}$$

Now, suppose we have M motion-energy cells $(\boldsymbol{\Omega}^\mu, \Omega_t^\mu)$, which are arranged to lie on L lines $(\nu = 1, \ldots, L)$ in the Ω_t direction. How many cells and how many lines does one need to estimate velocity? Along a line, the motion-energy distribution has only two parameters that one can estimate: the amplitude and the optimal temporal frequency. It is possible to show that to find these parameters, one needs measurements from two cells (two equations with two variables). Of those two parameters, the one depending on velocity is the optimal temporal frequency, which equals $-\boldsymbol{\Omega}\cdot\boldsymbol{v}$. Thus the optimal temporal frequency provides a single equation for two variables: the velocity components. It follows that one needs at least two lines to estimate local velocity. In summary, the theoretical minimum necessary to estimate local velocity is to have two lines with two cells each. In

practice, however, it is best to have many lines and more than two measurements per line. Denote the values of $r(\boldsymbol{\Omega})$ on the lines as r^ν for $\nu = 1, \ldots L$. Then the goodness-of-fit criterion becomes:

$$E(\boldsymbol{v}, r^\nu) = \sum_\mu (N(\boldsymbol{x}, t : \boldsymbol{\Omega}^\mu, \Omega_t^\mu, \sigma, \sigma_t) - r^\nu(\mu)\, \mathrm{e}^{-\sigma_t^2(\Omega_t^\mu + (\boldsymbol{\Omega}^\mu \cdot \boldsymbol{v}))^2/2})^2. \tag{2.3.4}$$

One of the ways to find the velocity $\boldsymbol{v}$ that minimizes this equation is as follows. Because the goodness-of-fit criterion, $E(\boldsymbol{v}, r^\nu)$, is quadratic in r^ν, we can obtain by differentiation a system of L linear equations and L variables, whose solution gives the best r^ν as a function of $\boldsymbol{v}$. By substituting back for r^ν one obtains a cost function, $\bar{E}(\boldsymbol{v})$. This function may be fed to velocity selective cells, that is, a cell selective to velocity $\boldsymbol{v}$ would receive input $\bar{E}(\boldsymbol{v})$. Among these cells, the one with the smallest response corresponds to the velocity estimate.

This method is local, can be implemented in parallel, is instantaneous, gives the correct answer even for a finite number of cells (in principle), and should degrade well with noise. Unfortunately it may not be biologically plausible.

The extra information strategy

This strategy uses the outputs of purely spatial frequency tuned cells to calculate the spatial characteristics of the image. This information can then be used to modify the estimation strategy by giving estimates for the form of $r(\boldsymbol{\Omega})$. We do not discuss this method in detail here.

3. Comparison with Heeger's method

This section compares our model to the method presented by Heeger (1987).

As in our model, his method starts with the calculation of motion energies (equation (2.1.4)). He also points out that motion energies are not velocity selective mechanisms, but rather are tuned to particular spatio-temporal frequencies.

To extract velocities from motion energies, Heeger first convolves the motion energies with a three-dimensional Gaussian window. Thus he estimates the average velocity within this window.

Next, the implementation compares the distribution of this convolution over the filters with the distribution expected for a particular stimulus: a moving flat-power-spectrum texture. In the case that the stimulus indeed has a flat power spectrum, the correct velocity can be found by matching the predicted and the measured energies. He also deals with a case where this condition is not met; when the contrasts are different for different spatial orientations.

Finally, Heeger suggests a parallel network that can compare the motion energies of the image and of the flat-power-spectrum texture.

In general, Heeger's method gives accurate velocity estimates for translating textured patterns, some sine-grating plaid patterns, and natural textures, and appears to simulate psychophysical data on the coherence of sine-grating plaids (Adelson & Movshon 1982).

However, we see three main problems with Heeger's model.

The first problem is the flat-power-spectrum assumption, which will lead to

incorrect velocity estimates for several images. His correction for different contrasts at different orientations is a limited attempt to solve this problem. Our method, however, does not suffer from this assumption and estimates velocity correctly for a greater variety of stimuli (§2.2).

A second problem with Heeger's method has to do with his spatio-temporal integration of motion energies. The range of his spatial integration is not limited to the cells' receptive field size, but also spans across cells whose centres of receptive field lie in different spatial locations. A manifestation of this problem is the smoothing that occurs when the integration windows straddle motion boundaries (Heeger 1987). As we have shown, this type of integration is unnecessary to compute velocities. We argue that there is no computational rationale to integrate motion energies over space as his method does. In contrast, our method performs local velocity estimates (for example, equations (2.3.1) or (2.3.3)), and thus allows for integration methods that have an explicit computational rationale (Hildreth 1984; Yuille & Grzywacz 1988*a*, *b*; Bulthoff *et al.* 1989; Grzywacz *et al.* 1989).

Finally, our weakest objection to Heeger's method is the suggestion that his parallel implementation, to compare the motion energies of the image and of the flat-power-spectrum texture, is a model for MT cells. The method uses mathematical operations that may not be easy to implement biologically (Ratliff 1965; Grzywacz & Koch 1987; Grzywacz & Poggio 1989, cf. §2.3).

4. Comparison with primary visual cortex

This section shows that the behaviour of the new model may account for some of the data obtained in the primary visual cortex. In particular, three characteristics of the primary visual cortex are discussed: space-time separability (§4.1), directional tuning (§4.2) and responses to apparent motions (§4.3). The part of the model identified with the primary visual cortex are the outputs of the motion-energy filters (equation (2.1.4)). Section 6 discusses how the combination of these outputs (§2.3) may account for the behaviour of MT.

4.1. *Space–time separability*

This section shows that the model accounts for the rough space-time separability of spatio-temporal tuning curves of primary visual cortex cells (Ikeda & Wright 1975; Tolhurst & Movshon 1975; Holub & Morton-Gibson 1981). We show that the model is not strictly separable, but that it is approximately so. This approximation follows from the separability of the Fourier Transform of the Gabor filters (equation (2.1.5)). However, this approximation is not related to the separability discussed by Poggio & Reichardt (1973, 1976). They showed that models consisting of linear filters separable in the space-time domain and followed by second-order nonlinearities have separable spatio-temporal frequency tuning curves in the average. (Examples of second-order nonlinearities include multiplication and squaring.)

To explore whether the model displays this separability, one must calculate equation (2.1.4) for moving sine gratings, whose luminance gradients are parallel

to the filter's preferred direction, $\boldsymbol{n}$. Let us neglect the background luminance and write the equation for these gratings as:

$$I(\boldsymbol{x}-\boldsymbol{v}t) = I_1 \sin(\lambda \boldsymbol{n}\cdot(\boldsymbol{x}-\boldsymbol{v}t)). \qquad (4.1.1)$$

The experimental spatio-temporal separability is claimed for time-averaged responses. Accordingly, we calculate the average response (see Appendix 3):

$$\langle N(\boldsymbol{x}, t : \Omega, \boldsymbol{n}, \Omega_t, \sigma, \sigma_t)\rangle = I_1^2 e^{-(\lambda-\Omega)^2\sigma^2} e^{-(\lambda \boldsymbol{n}\cdot\boldsymbol{v}+\Omega_t)^2\sigma_t^2} + I_1^2 e^{-(\lambda+\Omega)^2\sigma^2} e^{-(\lambda \boldsymbol{n}\cdot\boldsymbol{v}-\Omega_t)^2\sigma_t^2}. \qquad (4.1.2)$$

As a function of the spatial frequency, λ, and temporal frequency, $\alpha = \lambda \boldsymbol{n}\cdot\boldsymbol{v}$, this equation consists of two Gaussians centred at $(\lambda, \alpha) = \pm(\Omega, -\Omega_t)$ (figure 9).

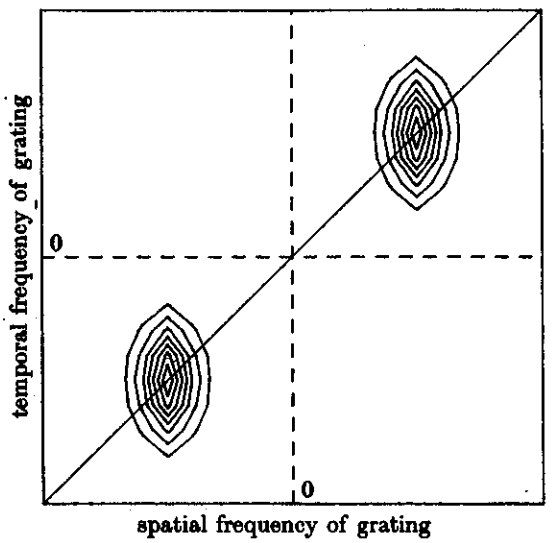

FIGURE 9. Approximate space–time separability of the spatio-temporal tuning curves. This is a contour plot of the spatio-temporal tuning curve for a motion-energy cell (equation (4.1.2)). The solid line represents a cross section of the velocity plane. Because sinusoidal motions with all their parameters changing sign remains the same, and as the plots are confined within the quadrants, the cell's tuning curve is approximately separable. This means that for two fixed temporal frequencies, the response dependencies on spatial frequency are roughly the same up to normalization. A similar conclusion holds if one fixes the spatial frequency and varies the temporal frequency.

Equation (4.1.2) shows that the separability found experimentally also occurs in the model. In the experiments, the tuning curves die well before the zero spatial and temporal frequencies. From the model's perspective, this means that $\Omega\sigma \gg 1$ and $\Omega_t\sigma_t \gg 1$, because, in this case, the contribution of each Gaussian to the other Gaussian's quadrant is negligible. Then, because sinusoidal motions with all the parameters changing signs remain the same, the tuning curve represented by equation (4.1.2) is $I_1^2 \exp(-(\lambda-\Omega)^2)\exp(-(\alpha+\Omega_t)\sigma_t^2)$. This tuning curve is separable in space and time.

4.2. *Directional tuning*

The directional tuning of directionally selective complex cells tends to be unimodal, when stimulated with bars or low speed dot textures but bimodal for high speed dot textures (Hammond 1979, 1981; Hammond & Reck 1981). Whenever this tuning is bimodal, the two preferred directions are distributed symmetrically about the bar's preferred direction.

This section shows that the model accounts for these observations. Our results confirm the arguments of Movshon *et al.* (1980) that these phenomena are consistent with spatio-temporal frequency tuned cells. Their idea comes from the observation that dots have Fourier components in all directions. Furthermore, the speed of a given component for a moving dot decreases with this component's angle with the direction of motion. Thus, if a dot moves fast, to elicit maximal response from a given spatio-temporal filter, the dot should not move parallel with the filter's best direction because in that case, the optimal Fourier component for the filter can move at the optimal speed. However, if a dot moves slowly, then it should move in the best direction of the filter, so that the optimal Fourier component has the best possible, though not optimal, speed. A potential problem with these arguments is that they do not take into account the contributions of non-optimal Fourier components. Our results show that this problem can be neglected.

For the sake of simplicity, we use sine gratings instead of bars and single dots instead of dot textures.

Consider the equation for the sine grating, whose luminance gradient unit vector, $\boldsymbol{\xi}$, may not be parallel to $\boldsymbol{n}$, and whose velocity is in the direction $\boldsymbol{\xi}$, that is, $\boldsymbol{v} = |\boldsymbol{v}|\boldsymbol{\xi}$:

$$I(\boldsymbol{x}-\boldsymbol{v}t) = I_1 \sin(\lambda\boldsymbol{\xi}\cdot(\boldsymbol{x}-\boldsymbol{v}t)). \tag{4.2.1}$$

We can calculate the time-averaged value of N as in equation (4.1.2):

$$\langle N(\boldsymbol{x}, t:\boldsymbol{\Omega}, \boldsymbol{n}, \Omega_t, \sigma, \sigma_t)\rangle$$
$$= I_1^2 e^{-(\lambda-\boldsymbol{\Omega}\cdot\boldsymbol{\xi})^2\sigma^2} e^{-(\lambda|\boldsymbol{v}|+\Omega_t)^2\sigma_t^2} e^{-(\lambda\boldsymbol{\Omega}\cdot\boldsymbol{\xi}^*)^2\sigma^2} + I_1^2 e^{-(\lambda+\boldsymbol{\Omega}\cdot\boldsymbol{\xi})^2\sigma^2} e^{-(\lambda|\boldsymbol{v}|-\Omega_t)^2\sigma_t^2} e^{-(\lambda\boldsymbol{\Omega}\cdot\boldsymbol{\xi}^*)^2\sigma^2}. \tag{4.2.2}$$

Let ϕ correspond to the orientation of the sine grating ($\boldsymbol{\xi} = (\cos\phi, \sin\phi)$), then by differentiating equation (4.2.2), one obtains:

$$\frac{\partial\langle N(\boldsymbol{x}, t:\boldsymbol{\Omega}, \boldsymbol{n}, \Omega_t, \sigma, \sigma_t)\rangle}{\partial\phi} \approx I_1^2\,\boldsymbol{\Omega}\sigma^2\lambda\boldsymbol{n}\cdot\boldsymbol{\xi}^* N(\boldsymbol{x}, t:\boldsymbol{\Omega}, \boldsymbol{n}, \Omega_t), \tag{4.2.3}$$

where $\boldsymbol{\xi}^* = (-\sin\phi, \cos\phi)$. This approximation assumes that we are close to one of the Gaussians, that is, in the separable region of §4.1. Thus the optimal directions of motion approximately satisfy the following equation:

$$\boldsymbol{n}\cdot\boldsymbol{\xi}^* = 0. \tag{4.2.4}$$

It follows that the preferred direction to the sine grating is approximately the direction of the filter's orientation, $\boldsymbol{n}$ (figure 10).

Consider now a dot travelling with constant speed:

$$I(\boldsymbol{x}-\boldsymbol{v}t) = I_1\,\delta(\boldsymbol{x}-\boldsymbol{v}t). \tag{4.2.5}$$

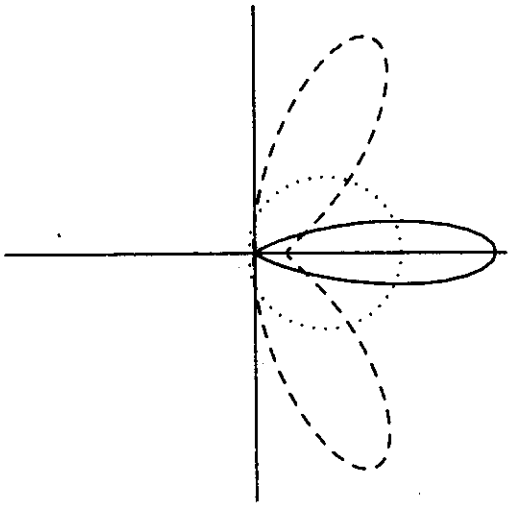

FIGURE 10. Dependence of single cells' directional tuning on stimulus shape. These polar plots are directional-tuning curves of the responses of a motion-energy cell to translating sinusoidal gratings (equation (4.2.2)) and dots (equation (4.2.6)). The plots are such that the cell response and stimulus direction correspond to the distance from the origin and angle, respectively. The optimal direction for the grating and the dots moving slowly is identical. However, when the dots move fast their directional tuning curve becomes bimodal with its lobes being symmetrical about the grating's best direction; (——), grating; (--), dot: high speed; (···) dot: low speed.

For simplicity, we give the response of the cells for which the dot passes through the centre of their receptive field, and for the time t at which the dot is at this centre:

$$N(\Omega, \boldsymbol{n}, \Omega_t, \sigma, \sigma_t) = \frac{\sigma^2\sigma_t^2}{\sigma^2+\sigma_t^2 v^2} \mathrm{e}^{-(\sigma\sigma_t)^2(\boldsymbol{v}\cdot\boldsymbol{n}\Omega+\Omega_t)^2/(\sigma^2+\sigma_t^2 v^2)}. \tag{4.2.6}$$

One can show that for other trajectories and times this section's conclusions will be the same. By differentiating this equation with respect to the direction of motion, one finds that the extrema of $N(\Omega, \boldsymbol{n}, \Omega_t)$ occur when:

$$(\boldsymbol{v}\cdot\boldsymbol{n}\Omega+\Omega_t)\,\boldsymbol{v}^*\cdot\boldsymbol{n}\Omega = 0, \tag{4.2.7}$$

where $\boldsymbol{v} = |\boldsymbol{v}|(\cos\phi, \sin\phi)$ and $\boldsymbol{v}^* = |\boldsymbol{v}|(-\sin\phi, \cos\phi)$. By substituting back into equation (4.2.6), one can determine that these extrema are maxima.

Thus the dot's and sine's preferred directions are similar when the dot moves slowly, but the dot's tuning is bimodal for high speeds (figure 10). If $|\boldsymbol{v}| \leqslant |\Omega_t/\Omega|$, then only $\boldsymbol{v}^*\cdot\boldsymbol{n} = 0$ satisfies equation (4.2.7), implying that $\boldsymbol{v}$ is in the direction $\boldsymbol{n}$. However, if $|\boldsymbol{v}| \geqslant |\Omega_t/\Omega|$, there exist two preferred directions symmetrical about $\boldsymbol{n}$ satisfying $\boldsymbol{v}\cdot\boldsymbol{n}\Omega+\Omega_t = 0$, with $\boldsymbol{v}^*\cdot\boldsymbol{n} = 0$ now determining a minimum.

4.3. *Apparent motion*

We define facilitation and suppression as positive and negative signs of the subtraction from an apparent motion response of the sum of the responses to the individual apparent motion slits, respectively. In other words, facilitation occurs

when one gets more response from the apparent motion than expected from the responses to the slits alone. And conversely, if suppression occurs one gets less response than expected.

Data from directionally selective complex cells in the cat's primary visual cortex show facilitation and suppression for the preferred and null directions, respectively (Emerson *et al.* 1987*b*). The suppression appears only if the distance and delay between the apparent motion slits is sufficiently long.

It was argued (Emerson *et al.* 1987*a*) that energy models, but not correlation models, for the measurement of visual motion (Hassenstein & Reichardt 1956; van Santen & Sperling 1984) conform with these data.

We confirm that energy models account for these data and suggest a new property: suppression may occur for all velocities not satisfying the filter's condition for translations, $\Omega \boldsymbol{n} \cdot \boldsymbol{v} + \Omega_t = 0$, even if the motion is in the preferred direction. The reason for the model to produce facilitation is the squaring operation (equation (2.1.4)). Under the correct velocities for facilitation, the positive portions of the response to one slit tend to occur when the response to the other slit is positive or small. A similar tendency exists for the negative portions of the responses. These tendencies lead to facilitation, because the square of the sum of two numbers of the same sign is larger than the sum of their squares. Similarly, suppression occurs if the interfering portions of the responses have opposite sign. The difficulty with these arguments, resolved by our calculations, is that the situation is often not so clear. Positive responses to a slit can occur during positive and negative responses to the other slit.

Consider a dot flashed in the image:

$$I(\boldsymbol{x}, t) = I_1 \, \delta(\boldsymbol{x} - \boldsymbol{x}_1) \, \delta(t - t_1). \tag{4.3.1}$$

One can compute this dot's N:

$$N(\boldsymbol{x}, t : \Omega, \boldsymbol{n}, \Omega_t, \sigma, \sigma_t) = I_1^2 \, \mathrm{e}^{-(\boldsymbol{x}-\boldsymbol{x}_1)^2/\sigma^2} \, \mathrm{e}^{-(t-t_1)^2/\sigma_t^2}. \tag{4.3.2}$$

Now, for a two-dot apparent motion we have:

$$I(\boldsymbol{x}, t) = I_1(\delta(\boldsymbol{x} - \boldsymbol{x}_1) \, \delta(t - t_1) + \delta(\boldsymbol{x} - \boldsymbol{x}_2) \, \delta(t - t_2)). \tag{4.3.3}$$

Also, in this case we can compute N:

$$\begin{aligned} N(\boldsymbol{x}, t : \Omega, \boldsymbol{n}, \Omega_t, \sigma, \sigma_t) = I_1^2 (&\mathrm{e}^{-(\boldsymbol{x}-\boldsymbol{x}_1)^2/\sigma^2} \, \mathrm{e}^{-(t-t_1)^2/\sigma_t^2} + \mathrm{e}^{-(\boldsymbol{x}-\boldsymbol{x}_2)^2/\sigma^2} \, \mathrm{e}^{-(t-t_2)^2/\sigma_t^2} \\ &+ 2 \, \mathrm{e}^{-(\boldsymbol{x}-\boldsymbol{x}_1)^2/(2\sigma^2)} \, \mathrm{e}^{-(t-t_1)^2/(2\sigma_t^2)} \, \mathrm{e}^{-(\boldsymbol{x}-\boldsymbol{x}_2)^2/(2\sigma^2)} \, \mathrm{e}^{-(t-t_2)^2/(2\sigma_t^2)} \\ &\times \cos(\Omega \boldsymbol{n} \cdot (\boldsymbol{x}_1 - \boldsymbol{x}_2) + \Omega_t (t_1 - t_2))). \end{aligned} \tag{4.3.4}$$

The first two terms on the right-hand side of equation (4.3.4) are the contributions from each dot, and the third term, the residual, is the quantity of interest:

$$\begin{aligned} Res(\boldsymbol{x}, t : \Omega, \boldsymbol{n}, \Omega_t) = 2 \, &\mathrm{e}^{-(\boldsymbol{x}-\boldsymbol{x}_1)^2/(2\sigma^2)} \, \mathrm{e}^{-(t-t_1)^2/(2\sigma_t^2)} \, \mathrm{e}^{-(\boldsymbol{x}-\boldsymbol{x}_2)^2/(2\sigma^2)} \, \mathrm{e}^{-(t-t_2)^2/(2\sigma_t^2)} \\ &\times \cos((\Omega \boldsymbol{n} \cdot \boldsymbol{v} + \Omega_t)(t_1 - t_2)), \end{aligned} \tag{4.3.5}$$

where $\boldsymbol{v} = (x_2 - x_1)/(t_2 - t_1)$. To know whether the interaction is facilitatory or suppressive we must consider the sign of *Res*:

$$Sgn(Res) = Sgn\,(\cos\,((\Omega \boldsymbol{n}\cdot\boldsymbol{v} + \Omega_t)\,(t_2 - t_1))), \tag{4.3.6}$$

where $Sgn(y) = 1$ if $y \geqslant 0$ and $Sgn(y) = -1$ if $y < 0$.

There is always a facilitation when the apparent motion fulfills $\Omega \boldsymbol{n}\cdot\boldsymbol{v} + \Omega_t = 0$ (figure 11). Direct substitution of this condition into equation (4.3.6) yields $Sgn(Res) = 1$.

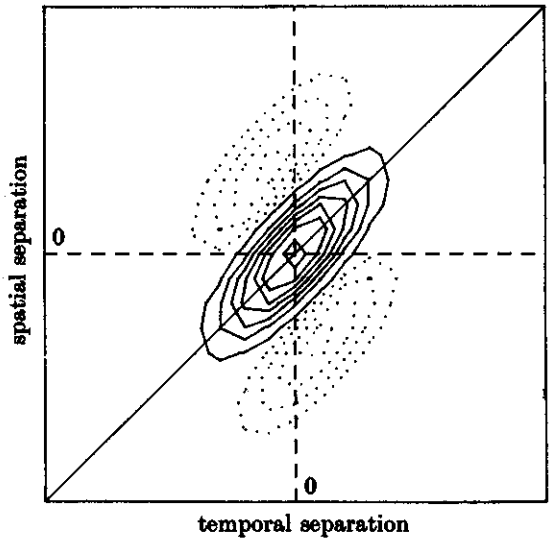

FIGURE 11. Facilitation and suppression in apparent motions. The figure displays a contour plot of the subtraction from an apparent motion response of the sum of the responses to the individual apparent motion slits (equation (4.3.5)). Positivity and negativity are given by the solid and dotted contours, respectively. The diagonal line represents a cross section of the translation plane, for which there is always facilitation (positivity). However, for every other velocity, including motions in the preferred direction, it may be possible to detect a suppression (negativity). The null direction is the easier in which to find suppression; for this direction, facilitation may occur if the delay between the slits is small.

However, for every other velocity it may be possible to detect a suppression. If $\boldsymbol{v}$ is such that $\Omega \boldsymbol{n}\cdot\boldsymbol{v} + \Omega_t \neq 0$, then one can choose $t_2 - t_1$ sufficiently large so that $Sgn(Res) = -1$. This condition may be true even for motions in the preferred direction (figure 11). These preferred-direction suppressions may not always be observable, as the Gaussians in equation (4.3.5) may reduce significantly the absolute value of *Res*.

Suppression is more likely to be observed for the null direction than for other motion directions (figure 11), as the null direction leads to the largest value of $|\Omega \boldsymbol{n}\cdot\boldsymbol{v} + \Omega_t|$. Nevertheless, even for the null direction, if $t_2 - t_1$ is small, one observes facilitation.

In summary, our model seems to account for the cortical cells' facilitation and suppression phenomenology, although the model suggests that suppression may be observable under wider stimulus conditions than those studied so far.

5. Comparison with MT

We now discuss how the new model may account for the detection of velocity in MT. The results of this section use the complete model, including the outputs of the motion-energy filters (equation (2.1.4)) and the combination of these outputs (§2.3). This is different to that in §4, which only required the motion-energy filters.

Movshon *et al.* (1985) distinguished between two classes of directionally selective cells in monkey MT: Component cells and Pattern cells (see also Rodman & Albright (1989)).

The Component cells respond to the motion direction of single-oriented contours but not of complex patterns. For these patterns, these cells only respond when an oriented portion of the patterns moves perpendicularly to the cell's preferred orientation. All primary visual cortex cells that are directionally selective and about 40% of MT cells appear to belong to this class.

The Pattern cells respond both to the direction of motion of single oriented contours and of complex patterns. These cells appear to represent about 25% of MT cells. The other 35% of MT cells cannot be clearly classified as either Component or Pattern cells. Movshon *et al.* (1985) argue that this ambiguity is mostly due to the statistical insensitivity of their methods. However, it is possible that other classes of cell exist.

To distinguish between these cell types, researchers (Movshon *et al.* 1985; Rodman & Albright 1989) measured the difference in response to moving sinusoidal grating and sinusoidal plaid stimuli. (The latter is the sum of a pair of crossed sinusoidal gratings.) The Component cells' directional tuning was bimodal for the plaids. The optimal directions roughly occurred when the plaid gratings were perpendicular to the cells' preferred direction as determined by single gratings. On the other hand, the optimal direction of the Pattern cells for the plaid were approximately coincident with that of the single gratings.

We now show that the motion energy filters (equation (2.1.4)) behave like Component cells, whereas the filters' combination by one of the methods of §2.3 behave like Pattern cells.

We use plaids that are the superposition of two orthogonal sine wave gratings travelling with the same velocity $\boldsymbol{v}$:

$$I(\boldsymbol{x}-\boldsymbol{v}t) = I_1 \sin(\lambda\boldsymbol{\xi}\cdot(\boldsymbol{x}-\boldsymbol{v}t)) + I_1 \sin(\lambda\boldsymbol{\xi}^*\cdot(\boldsymbol{x}-\boldsymbol{v}t)), \tag{5.1}$$

where λ is the gratings' spatial frequency, and $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^*$ are the gratings' directions. The time-averaged response is given by:

$$\begin{aligned} N(\boldsymbol{x},t:\boldsymbol{\Omega},\Omega_t,\sigma,\sigma_t) = {} & (I_1^2/4)\,(\mathrm{e}^{-\sigma^2(\lambda-\boldsymbol{\Omega}\cdot\boldsymbol{\xi})^2}\mathrm{e}^{-\sigma^2(\boldsymbol{\Omega}\cdot\boldsymbol{\xi}^*)^2}\mathrm{e}^{-\sigma_t^2(\lambda(\boldsymbol{v}\cdot\boldsymbol{\xi})+\Omega_t)^2} \\ & + \mathrm{e}^{-\sigma^2(\lambda+\boldsymbol{\Omega}\cdot\boldsymbol{\xi})^2}\mathrm{e}^{-\sigma^2(\boldsymbol{\Omega}\cdot\boldsymbol{\xi}^*)^2}\mathrm{e}^{-\sigma_t^2(\lambda(\boldsymbol{v}\cdot\boldsymbol{\xi})-\Omega_t)^2}) \\ & + (I_1^2/4)\,(\mathrm{e}^{-\sigma^2(\lambda-\boldsymbol{\Omega}\cdot\boldsymbol{\xi}^*)^2}\mathrm{e}^{-\sigma^2(\boldsymbol{\Omega}\cdot\boldsymbol{\xi})^2}\mathrm{e}^{-\sigma_t^2(\lambda(\boldsymbol{v}\cdot\boldsymbol{\xi}^*)+\Omega_t)^2} \\ & + \mathrm{e}^{-\sigma^2(\lambda+\boldsymbol{\Omega}\cdot\boldsymbol{\xi}^*)^2}\mathrm{e}^{-\sigma^2(\boldsymbol{\Omega}\cdot\boldsymbol{\xi})^2}\mathrm{e}^{-\sigma_t^2(\lambda(\boldsymbol{v}\cdot\boldsymbol{\xi}^*)-\Omega_t)^2}). \end{aligned} \tag{5.2}$$

This is a linear combination of Gaussians centred on $(\boldsymbol{\Omega},\Omega_t) = (\lambda\boldsymbol{\xi}, -\lambda\boldsymbol{v}\cdot\boldsymbol{\xi}) = (-\lambda\boldsymbol{\xi}, \lambda\boldsymbol{v}\cdot\boldsymbol{\xi}) = (\lambda\boldsymbol{\xi}^*, -\lambda\boldsymbol{v}\cdot\boldsymbol{\xi}^*) = (-\lambda\boldsymbol{\xi}^*, \lambda\boldsymbol{v}\cdot\boldsymbol{\xi}^*)$. Because these Gaussians are

centred on the plane $\boldsymbol{\Omega} \cdot \boldsymbol{v} + \Omega_t = 0$, both the Ridge and the Estimation strategies of §2.3 will yield the true velocity.

Figure 12 plots the directional tuning curve to the plaid (equation (5.1)) of a Component and a Pattern cell assuming the Ridge strategy of §2.3. The optimal directions in these tuning curves agree with those recorded in MT (Movshon *et al.* 1985; Rodman & Albright 1989). The directional tuning is sharper for Component than for Pattern cells coinciding with the sharper tuning in the primary visual cortex when compared with MT. This raises the possibility that primary visual cortex cells feed directly into Pattern cells without using the MT Component cells as intermediate (like Scheme a of fig. 9 of Rodman & Albright (1989)). Also, the tuning difference in figure 12 might be reduced by a mutual inhibitory network implementing a partial winner-take-all mechanism (Yuille & Grzywacz 1989*b*).

Finally, the model also appears to be consistent with the finding that type II cells in Albright's classification of MT cells (Albright 1984) correspond to the Pattern cells (Rodman & Albright 1989). The property characterizing type II cells is that their preferred direction for moving spots and preferred orientation for stationary slits are parallel. Our model accounts for the correspondence between type II and Pattern cells as follows. Because the slits are stationary, they would activate Pattern cells in MT through motion-energy cells tuned to low temporal frequencies. The only such cells consistent with a given velocity vector are those tuned to gratings parallel to it. This is because, these are the only gratings not expected to move. Thus the best stationary slits are those whose Fourier

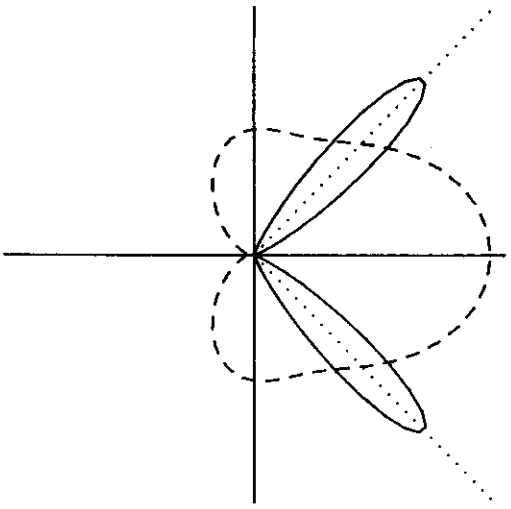

FIGURE 12. The directional tuning curve of the responses to the plaid of a Component (——) and a Pattern (– –) cell. These polar plots have the same convention as in figure 10. The Component cells respond (equation (4.4.2)) mainly when one of the gratings composing the plaid moves in parallel to the cell's preferred direction (to the right). This occurs when the plaid moves parallel to the dotted lines. The Pattern cells' responses were simulated by using the Ridge strategy without a winner-take-all mechanism (§2.3). The directional tuning is sharper for Component than for Pattern cells, coinciding with the sharper tuning in the primary visual cortex when compared to MT. This tuning difference might be reduced by a mutual inhibitory network implementing a partial winner-take-all mechanism (Yuille & Grzywacz 1989).

components are parallel to the Pattern cell's preferred direction. This explanation is essentially the same one that was provided by Rodman & Albright (1989).

6. Discussion

We presented a model, qualitatively consistent with physiology, that postulates two stages for cortical velocity estimate: the first measuring motion energies (Adelson & Bergen 1985) from the moving stimulus and the second estimating velocity from these energies. The first stage might correspond to the primary visual cortex and the second to MT (Movshon *et al.* 1985), but other alternatives are possible (Rodman & Albright 1989). The intuition for how the second stage combines the first-stage outputs is similar to that suggested byMcKee *et al.* (1986). The model would yield arbitrarily correct velocity if the sampling of spatio-temporal frequencies by the first stage was arbitrarily dense. Evidence for high-density measurements for spatial frequencies has been provided (Silverman *et al.* 1989).

The main ideas behind this model are general and may have applications to other early visual computations. These ideas are of a fast filtering of information followed by the combination of the filters' outputs to compute the relevant visual parameters. It has been suggested that texture discrimination (Clark *et al.* 1987; Bergen & Adelson 1988; Daugman 1988), feature detection (Morrone & Burr 1988) and stereopsis (Sanger 1988; Yeshurun & Schwartz 1987) may follow such a strategy.

The model's second stage computes velocity *locally* from the motion-energy distribution across first-stage cells, and thus may explain (see also Adelson (1987); Fleet & Jepson (1989)) the perhaps related phenomena of motion transparency (Adelson & Movshon 1982) and discontinuities (Anstis 1970). The computation uses only motion-energy cells, whose centres of receptive field lie in a single spatial location. If two different motion fields are adjacent, then a bimodal distribution of motion energies is generated, implying two velocities (figure 13*a*). Bimodal distributions would also occur in motion transparency (figure 13*b*). To detect two velocities from these bimodal distributions, the Ridge strategy (§2.3) might use a *local* winner-take-all mechanism. However, it is possible that the brain uses a *global* winner-take-all strategy. In this case, with no winner in MT, another visual pathway might compute velocities directly from the primary visual cortex. An alternative is that with a global winner-take-all mechanism, the winners would switch transiently between themselves as the image changes leading to the perception of transparency. Schemes that integrate velocity signals (Hildreth 1984; Yuille & Grzywacz 1988*a*, *b*; Bulthoff *et al.* 1989; Grzywacz *et al.* 1989) may have to do so by segregating velocities that differ locally.

Why do we say that the model works locally despite the theorems' assumption that the velocity is constant over the whole image? The reason is that in practice this assumption can be greatly relaxed. Because the cells have essentially a limited spatio-temporal range, determined by σ and σ_t, the velocity only needs to be constant over this range. Thus the model provides good velocity etimate almost everywhere for classes of motion, such as rotation or expansion, that can be locally approximated as translation.

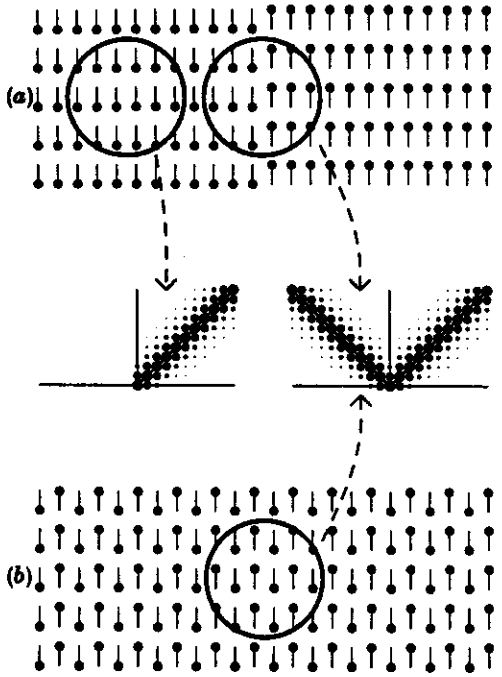

FIGURE 13. Transparency and motion discontinuties. (*a*) Two adjacent fields of dots move in opposite directions, thus forming a boundary of motion discontinuity. (*b*) Two superimposed fields of dots move in opposite directions, thus forming a region of motion transparency. The diameters of the circles represent the receptive field sizes of motion-energy cells. If such cells are near the discontinuity or are seeing motion transparency, then the motion-energy distribution for these cells lie around two planes (cf. figure 3*b*). Otherwise, if there is only one motion field on these cells' receptive fields, then the motion energies lie around a single plane. To detect two velocities from the above bimodal distributions, the Ridge strategy (§2.3) might use a local winner-take-all mechanism. Schemes that integrate velocity signals may have to do so by segregating velocities that differ locally.

We will now discuss two assumptions of the model's first stage that are probably incorrect in physiological details: Gabor filtering (equation (2.1.1)) and filter's output squared as the cells' responses (equation (2.1.4)). This incorrectness follows from the modelling of cell responses by simple mathematics instead of realistic biophysics (Grzywacz & Poggio 1989). However, our main idea, the combination of motion-energy filters, seems to be conveniently modelled by this paper's methods.

The Gabor function is, strictly speaking, the only filter for which we can guarantee that the extrema of responses in the cells' optimal-frequency space lie on a ridge (unpublished calculations). This is due to the fact that the Gaussian is the only separable rotationally invariant function. If, however, the filters are similar, but not exactly, like Gabors, then we expect the results of §2.2 to be true most of the time. This expectation is confirmed by the velocity computation in real images with filters that were built by a self-organizing developmental model, and that only approximately resemble Gabor functions (Yuille & Cohen 1989).

To what extent are Gabor functions good models of the properties of primary

visual cortex cells? Some researchers suggest that these functions are good models of the receptive-field structure in cat (Jones & Palmer 1987). Also, physiologically plausible models that resemble spatio-temporal Gabor filters have been proposed (Adelson & Bergen 1985). However, there are at least three problems with Gabor models. First, they predict that the cells' spatial tuning curves follow a Gaussian function (equation (4.1.2)). However, the data show that these tuning curves follow a log-normal function (Gaussian when plotted on a logarithmic scale (Gaddum 1945; Holub & Morton-Gibson 1981)), typically with more than 85% of the area under the curves being for frequencies higher than optimal. Such log-normal behaviour might be significant, as it would naturally provide the scaling properties necessary for good velocity estimates (figure 6). If tuning curves are log-normal, the frequency range of a cell is proportional to its optimal frequency. An alternative to Gabor models, which would have the same scaling property, are Wavelet models (Grossmann & Morlet 1984; Mallat 1988). Secondly, the Gabor-filter temporal part is non-causal, that is, because it is not zero for negative times, it uses future information in its computations. A filter, whose temporal tuning curve is log-normal, as in cortical cells (Holub & Morton-Gibson 1981), might correct this problem (typically more than 99% of this tuning curve occurs for frequencies higher than optimal). On the other hand, the non-causality may also imply that velocity computation is not 'on-line'; a delay that is consistent with the temporal integration necessary to estimate velocity accurately (Nakayama & Tyler 1981; Regan & Beverly 1984; McKee & Welch 1985). Thirdly, the use of a Gabor filter implies an initially linear mechanism, which is not always supported by experiments (Emerson & Citron 1988; Reid & Shapley 1988). One neural mechanism that destroys the linearity of the cells is the ON–OFF rectification. Another neural nonlinearity that may have a role in velocity computations is shunting inhibition, which seems to account for *retinal* directional selectivity (Torre & Poggio 1978; Marchiafava 1979; Amthor & Grzywacz 1990).

It is improbable that the cells' responses correspond to the square of the filters' output (equation (2.1.4)), as squaring operations may not be part of the neural 'vocabulary' (Grzywacz & Koch 1987; Grzywacz & Poggio 1989). However, in our model, this operation underlies the facilitation in preferred-direction apparent motions (figure 11). Thus, in physiologically realistic models, the squaring operation must be substituted by other supralinear operations (relations with a positive second derivative).

Our final argument is that biological motion perception may use at least three stages; two for velocity estimation and at least one more to deal with coherent motion (Yuille & Grzywacz 1988*a*, *b*). The first two stages are similar to those proposed by Adelson & Movshon (1982) and Movshon *et al.* (1985), although we suggest a different role for these stages than they do. Direct psychophysical evidence for two stages in the computation of velocity has recently been given (Welch 1989). Interestingly, in theory, velocity estimation requires only one stage (Verri *et al.* 1989).

We suggest that the primary visual cortex and MT represent two stages needed to estimate velocity, and argue, contrary to Movshon *et al.* (1985), that MT is not concerned with the aperture problem. To do so, we first define what we mean by

the aperture problem. It is the impossibility to measure locally, velocity components other than that normal to the luminance gradient (Marr & Ullman 1981). Movshon *et al.* postulate that the role of the primary visual cortex is to analyse motions of one-dimensional patterns. These researchers also advance the idea that MT solves the resulting aperture problem. In particular, they propose a model assuming that primary visual cortex cells compute the velocity *normal* to one-dimensional patterns. This proposition is consistent with the finding that primary visual cortex cells are 'velocity selective' for moving bars (Movshon 1975; Orban *et al.* 1986; Baker 1988). As a curiosity, our model predicts this selectivity (unpublished calculations). However, the receptive-field size of primary visual cortex cells is typically larger than regions of the visual world where significant curvature and texture exist. Thus these cells have to deal with two-dimensional patterns. Under these conditions, the assumptions of Movshon *et al.* lead to incorrect estimates of velocity. We suggest that the primary visual cortex does not assume a one-dimensional visual world. Thus we argue that the role of MT is to compute local velocity without having to deal with the aperture problem. It is tempting to identify the speed-selective cells found in MT (Newsome *et al.* 1983) with its pattern cells (Movshon *et al.* 1985; Rodman & Albright 1989). A problem with this identification is that studies with sinusoidal gratings suggest that it applies only to a small percentage of MT cells (J. A. Movshon, personal communication). However, one must be careful in interpreting results from sinusoidal-grating experiments, as they may, for two reasons, represent poor stimuli for the Pattern cells. The first reason is that sinusoidal gratings are one-dimensional and the Pattern cells may not be designed to solve the aperture problem in the whole (§2.3). The second reason is that sinusoidal gratings may stimulate effectively very few of the motion-energy cells, leading to signal-to-noise problems. The facilitatory interaction between component cells (J. A. Movshon, personal communication; Ferrera & Wilson 1987) emphasizes that a good stimulus should probably comprise several Fourier components.

At least one more motion-processing stage is needed to deal with another problem, which is related to the aperture problem. This new problem occurs when objects larger than the receptive field size of MT cells move. In this case, the cells may compute different velocities for different portions of these objects. Nevertheless, it is often important to assign a global motion to the objects (Hildreth 1984; Yuille & Grzywacz 1988*a*, *b*). The solution for this coherence problem requires a motion-processing stage, which might occur in later cortical areas (Tanaka *et al.* 1986; Saito *et al.* 1986), which perform spatial integration over large receptive fields.

Some of the motivation for this work was provided by discussions with Dave Heeger and John Daugman. We thank Lyle Borg-Graham, Dave Heeger, Ellen Hildreth, Tommy Poggio, Jeff Schall, Tai Sing and Jim Smith for critical reading of the manuscript. N.M.G. was supported by the grant BNS-8809528 from the National Science Foundation and the Sloan Foundation; Tomaso Poggio, Ellen Hildreth and Peter Schiller by a grant from the Office of Naval Research, Cognitive and Neural Systems Division; Tomaso Poggio, Ellen Hildreth and Edward

Adelson by grant IRI-8719394 from the National Science Foundation. A. L. Y. was supported by the Brown-Harvard-MIT Center for Intelligent Control Systems with the United States Army Research Office grant DAAL03-86-K-0171.

References

Adelson, E. H. 1987 Transparency in motion perception. *Invest. Ophthal. vis. Sci.* **28**, 232.

Adelson, E. H. & Bergen, J. 1985 Spatiotemporal energy models for the perception of motion. *J. opt. Soc. Am.* A **2**, 284–299.

Adelson, E. H. & Movshon, J. A. 1982 Phenomenal coherence of moving visual patterns. *Nature, Lond.* **300**, 523–525.

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

Amthor, F. R. & Grzywacz, N. M. 1990 The non-linearity of the inhibition underlying retinal directional selectivity. (In the press.)

Andersen, R. A. & Siegel, R. M. 1989 Motion processing in primate cortex. In *Signal and sense: local and global order in perceptual maps* (ed. G. M. Edelman, W. E. Gall & W. M. Cowan). New York: John Wiley. (In the press.)

Andrews, B. W. & Pollen, D. A. 1979 Relationship between spatial frequency selectivity and receptive field profile of simple cells. *J. Physiol., Lond.* **287**, 163–176.

Anstis, S. 1970 Phi movement as a subtraction process. *Vision Res.* **10**, 1411–1430.

Baker, C. L. jr. 1988 Spatial and temporal determinants of directionally selective velocity preference in cat striate cortex neurons. *J. Neurophysiol.* **59**, 1557–1574.

Bergen, J. R. & Adelson, E. H. 1988 Early vision and texture perception. *Nature, Lond.* **333**, 363–364.

Bisti, S., Carmignoto, L., Galli, L. & Maffei, L. 1985 Spatial-frequency characteristics of neurones of area 18 in the cat: dependence on the velocity of the visual stimulus. *J. Physiol., Lond.* **359**, 259–268.

Bulthoff, H., Little, J. & Poggio, T. 1989 A parallel algorithm for real-time computation of optical flow. *Nature, Lond.* **337**, 549–553.

Clark, M., Bovik, A. C. & Geisler, W. S. 1987 Texture segmentation using a class of narrowband filters. In *Proceedings of the international conference on acoustic systems and signal processing*. Washington, D.C.: IEEE Press.

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

Daugman, J. G. 1988 Pattern and motion vision without Laplacian zero-crossings. *J. opt. Soc. Am.* A **5**, 1142–1148.

Emerson, R. C., Bergen, J. R. & Adelson, E. H. 1987*a* Movement models and directionally selective neurons in the cat's visual cortex. *Neurosci. Abstr.* **13**, 1623.

Emerson, R. C. & Citron, M. C. 1988 How linear and nonlinear mechanisms contribute to directional selectivity in simple cells of cat striate cortex. *Invest. Ophthalmol. vis. Sci.* **29**, 23.

Emerson, R. C., Citron, M. C., Vaughn, W. J. & Klein, S. A. 1987*b* Nonlinear directionally sensitive subunits in complex cells of cat striate cortex. *J. Neurophysiol.* **58**, 33–65.

Fahle, M. & Poggio, T. 1981 Visual hyperacuity: spatio-temporal interpolation in human vision. *Proc. R. Soc. Lond.* B **213**, 451–477.

Feldman, J. A. & Ballard, D. H. 1982 Connectionist models and their properties. *Cog. Sci.* **6**, 205–254.

Ferrera, V. P. & Wilson, H. R. 1987 Direction specific masking and the analysis of motions in two dimensions. *Vis. Res.* **27**, 1783–1796.

Fleet, D. J. & Jepson, A. D. 1989 Computation of normal velocity from local phase information. *Tech. Rep. Res. Biol. Comp. Vis.* no. RBCV-TR-89-27. Department of Computer Science, University of Toronto, Toronto, Canada.

Gabor, D. 1946 Theory of communication. *J. Inst. Electr. Eng.* **93**, 429–457.

Gaddum, J. H. 1945 Lognormal distributions. *Nature, Lond.* **156**, 463–466.

Galli, L., Chalupa, L., Maffei, L. & Bisti, S. 1988 The organization of receptive fields in area 18 neurones of the cat varies with the spatio-temporal characteristics of the visual stimulus. *Expl Brain Res.* **71**, 1–7.

Grossmann, A. & Morlet, J. 1984 Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM Jl appl. Math.* **15**, 723–736.

Grzywacz, N. M. & Koch, C. 1987 Functional properties of models for direction selectivity in the retina. *Synapse* **1**, 417–434.

Grzywacz, N. M. & Poggio, T. 1989 Computation of motion by real neurons. In *An introduction to neural and electronic networks* (ed. S. F. Zornetzer, J. L. Davis & C. Lau). Orlando, Florida: Academic Press. (In the press.)

Grzywacz, N. M., Smith, J. A. & Yuille, A. L. 1989 A common theoretical framework for visual motion's spatial and temporal coherence. In *Proceedings of an IEEE workshop on visual motion, Irvine, California*, pp. 148–155. Washington, DC: IEEE Computer Society Press.

Hammond, P. 1979 Stimulus-dependence of ocular dominance and directional tuning of complex cells in area 17 of the feline visual cortex. *Expl Brain Res.* **35**, 583–589.

Hammond, P. 1981 Simultaneous determination of directional tuning of complex cells in cat striate cortex for bar and for texture motion. *Expl Brain Res.* **41**, 364–369.

Hammond, P. & Reck, J. 1981 Influence of velocity on directional tuning of complex cells in cat striate cortex for texture motion. *Neurosci. Lett.* **19**, 309–314.

Hassenstein, B. & Reichardt, W. E. 1956 Systemtheoritsche analyse der zeit-, reihenfolgen- und vorzeichenauswertung bei der bewegungsperzeption des russelkafers *chlorophanus*. *Z. Naturforsch* **11b**, 513–524.

Heeger, D. 1987 A model for the extraction of image flow. *J. opt. Soc. Am.* A **4**, 1455–1471.

Hildreth, E. C. 1984 *The measurement of visual motion*. Cambridge, Massachusetts: MIT Press.

Hochstein, S. & Shapley, R. M. 1976 Quantitative analysis of retinal ganglion cell classifications *J. Physiol., Lond.* **262**, 237–264.

Holub, R. A. & Morton-Gibson, M. 1981 Response of visual cortical neurons of the cat to moving sinusoidal gratings: Response-contrast functions and spatiotemporal integration. *J. Neurophysiol.* **46**, 1244–1259.

Huggins, W. H. & Licklider, J. C. R. 1951 Place mechanisms of auditory frequency analysis. *J. acoust. Soc. Am.* **23**, 290–299.

Ikeda, H. & Wright, M. J. 1975 Spatial and temporal properties of 'sustained' and 'transient' neurones in area 17 of the cat's visual cortex. *Expl Brain Res.* **22**, 363–383.

Jasinschi, R. S. 1988 Space-time sampling with motion uncertainty: constraints on space–time filtering. In *Proceedings of the 2nd international conference on computer vision, Tampa, Florida*, pp. 428–434. Washington, DC: IEEE Computer Society Press.

Jones, J. P. & Palmer, L. A. 1987 An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiol.* **58**, 1233–1258.

Koch, C. & Ullman, S. 1985 Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiol.* **4**, 219–227.

Levinson, E. & Sekuler, R. 1976 Adaptation alters perceived direction of motion. *Vision Res.* **16**, 779–781.

McKee, S. P. 1981 A local mechanism for differential velocity detection. *Vision Res.* **21**, 491–500.

McKee, S. P. & Nakayama, K. 1984 The detection of motion in the peripheral visual field. *Vision Res.* **24**, 25–32.

McKee, S. P., Silverman, G. H. & Nakayama, K. 1986 Precise velocity discrimination despite random variations in the temporal frequency and contrast. *Vision Res.* **26**, 609–619.

McKee, S. P. & Welch, L. 1985 Sequential recruitment in the discrimination of velocity *J. opt. Soc. Am.* A **2**, 243–251.

McLean, J., Raab, S. & Palmer, L. 1987 Spatiotemporally oriented simple receptive fields: local linear motion detectors. *Neurosci. Abstr.* **13**, 1623.

Maffei, L. & Fiorentini, A. 1977 Spatial frequency rows in the striate visual cortex. *Vision Res.* **17**, 257–264.

Mallat, S. G. 1988 Review of multifrequency channel decompositions of images and wavelet models. *Robot. Res. Tech. Rep.* no. 412. Computer Science Division, New York University, New York.

Marchiafava, P. L. 1979 The responses of retinal ganglion cells to stationary and moving visual stimuli. *Vision Res.* **19**, 1203–1211.

Marr, D. & Ullman, S. 1981 Directional selectivity and its use in early visual processing. *Proc. R. Soc. Lond.* B **211**, 151–180.

Maunsell, J. H, R,. & Newsome, W. T. 1987 Visual processing in monkey extrastriate cortex. In *Annual review of Neuroscience* (ed. W. M. Cowan, E. M. Shooter, C. F. Stevens & R. F. Thompson), vol. 10, pp. 363–401. Palo Alto, California: Annual Reviews Inc.

Morrone, M. C. & Burr, D. C. 1988 Feature detection in human vision: a phase-dependent energy model. *Proc. R. Soc. Lond.* B **235**, 221–245.

Movshon, J. A. 1975 The velocity tuning of single units in cat striate cortex. *J. Physiol., Lond.* **249**, 445–468.

Movshon, J. A., Adelson, E. H., Gizzi, M. S. & Newsome, W. T. 1985 The analysis of moving visual patterns. In *Pattern recognition mechanisms* (ed. C. Chagas, R. Gattas & C. G. Gross), pp. 117–151. Rome: Vatican Press.

Movshon, J. A., Davis, E. T. & Adelson, E. H. 1980 Directional movement selectivity in cortical complex cells. *Neurosci. Abstr.* **6**, 230.

Nakayama, K. 1985 Biological motion processing: a review. *Vision Res.* **25**, 625–660.

Nakayama, K. & Tyler, C. W. 1981 Psychophysical isolation of movement sensitivity by removal of familiar position cues. *Vision. Res.* **21**, 427–433.

Nauta, W. J. H. & Feirtag, M. 1986 *Fundamental neuroanatomy*. New York, New York: W. H. Freeman.

von Neumann, J. 1958 *The computer and the brain*. New Haven, Connecticut: Yale University Press.

Newsome, W. T., Gizzi, M. S. & Movshon, J. A. 1983 Spatial and temporal properties of neurons in macaque MT. *Invest. Ophthalmol. vis. Sci.* **24**, 106,

Orban, G. A., Kennedy, H. & Bullier, J. 1986 Velocity sensitivity and direction selectivity of neurons in areas V1 and V2 of the monkey: influence of eccentricity. *J. Neurophysiol.* **56**, 462–480.

Poggio, T. & Reichardt, W. E. 1973 Considerations on models of movement detection. *Kybernetics* **13**, 223–227.

Poggio, T. & Reichardt, W. E. 1976 Visual control of orientation behaviour in the fly: part II: towards the underlying neural interactions. *Q. Rev. Biophys.* **9**, 377–438.

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

Ratliff, F. 1965 *Mach bands: quantitative studies on neural networks in the retina*. San Francisco: Holden-Day.

Regan, D. & Beverley, K. I. 1984 Figure ground segregation by motion contrast and by luminance contrast. *J. opt. Soc. Am.* A **1**, 433–442.

Reid, R. C. & Shapley, R. M. 1988 Complex temporal stimuli increase relative sensitivity of cat striate cortical neurons to high temporal frequencies. *Invest. Ophthalmol. vis. Sci.* **27**, 142.

Rodman, H. R. & Albright, T. D. 1989 Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). *Expl Brain Res.* **75**, 53–64.

Saito, H., Yukio, M., Tanaka, K., Hikosaka, D., Fukuda, Y. & Iwai, E. 1986 Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. *J. Neurosci.* **6**, 145–157.

Sanger, T. 1988 Stereo disparity computation using Gabor filters. *Biol. Cyber.* **59**, 405–418.

van Santen, J. P. H. & Sperling, G. 1984 A temporal covariance model of motion perception. *J. opt. Soc. Am.* A **1**, 451–473.

Silverman, M. S., Grosof, D. H., De Valois, R. L. & Elfar, S. D. 1989 Spatial-frequency organization in primate striate cortex. *Proc. natn. Acad. Sci. U.S.A.* **86**, 711–715.

Tanaka, K., Hikosaka, K., Saito, H., Yukie, M., Fukada, Y. & Iwai, E. 1986 Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey. *J. Neurosci.* **6**, 134–144.

Tolhurst, D. J. & Movshon, J. A. 1975 Spatial and temporal contrast sensitivity of striate cortical neurons. *Nature, Lond.* **257**, 674–675.

Torre, V. & Poggio, T. 1978 A synaptic mechanism possibly underlying directional selectivity to motion. *Proc. R. Soc. Lond.* B **202**, 409–416.

Verri, A., Girosi, F. & Torre, V. 1989 Mathematical properties of the 2D motion field: From singular points to motion parameters. In *Proceedings of an IEEE workshop on visual motion, Irvine, California*, pp. 190–200. Washington, DC: IEEE Computer Society Press.

Watson, A. B. & Ahumada, A. J. 1985 Model of human visual-motion sensing. *J. opt. Soc. Am.* A 2, 322–341.

Welch, L. 1989 The perception of moving plaids reveals two motion-processing stages. *Nature, Lond.* 337, 734–736.

Yeshurun, Y. & Schwartz, E. L. 1987 Cepstral filtering on a columnar image architecture: a fast algorithm for binocular stereo segmentation. *Robot. Res. Tech. Rep.* New York University, Courant Institute of Mathematical Sciences, technical report no. 286.

Yuille, A. L. & Cohen, D. S. 1989 The development and training of motion and velocity sensitive cells. Harvard Robotics Laboratory technical report no. 89–9.

Yuille, A. L. & Grzywacz, N. M. 1988*a* A computational theory for the perception of coherent visual motion. *Nature, Lond.* 333, 71–74.

Yuille, A. L. & Grzywacz, N. M. 1988*b* The motion coherence theory. In *Proceedings of the 2nd international conference on computer vision, Tampa, Florida*, pp. 344–353. Washington, DC: IEEE Computer Society Press.

Yuille, A. L. & Grzywacz, N. M. 1989*a* A model for the estimate of local image velocity by cells in the visual cortex. *Invest. Ophthalmol. vis. Sci.* 30, 425.

Yuille, A. L. & Grzywacz, N. M. 1989*b* A winner-take-all mechanism based on presynaptic inhibition feedback. *Neural Comp.* 1, 334–347.

Yuille, A. L. & Poggio, T. 1986 Scaling theorems for zero-crossings. *PAMI*-8 1, 15–25.

Appendix 1

In this appendix, we give a proof for Theorem 2. We first calculate $f(\boldsymbol{x}, t)$, the spatio-temporal convolution of the image with the complex filter in 2.1.1. The response of our nonlinear filter N is then given by the relation $N(\boldsymbol{x}, t : \boldsymbol{\Omega}, \boldsymbol{n}, \Omega_t, \sigma, \sigma_t) = |f(\boldsymbol{x}, t)|^2$. By using the convolution theorem we find:

$$f(\boldsymbol{x}, t : \boldsymbol{\Omega}, \boldsymbol{n}, \Omega_t, \sigma, \sigma_t) = \int \overline{F(\boldsymbol{\omega}, \omega_t : \boldsymbol{\Omega}, \boldsymbol{n}, \Omega_t, \sigma, \sigma_t)}\, \overline{I(\boldsymbol{\omega}, \omega_t)}\, \mathrm{e}^{-i\boldsymbol{\omega}\cdot\boldsymbol{x}}\, \mathrm{e}^{-i\omega_t t}\, \mathrm{d}\boldsymbol{\omega}\, \mathrm{d}\omega_t, \tag{A 1.1}$$

where $\overline{I(\boldsymbol{\omega}, \omega_t)}$ is the Fourier transform of the image $I(\boldsymbol{x}, t)$. If the image is moving with constant velocity $\boldsymbol{v}$, then:

$$\overline{I(\boldsymbol{\omega}, \omega_t)} = (2\pi)^{\frac{1}{2}} \delta(\boldsymbol{v}\cdot\boldsymbol{\omega} + \omega_t)\, g(\boldsymbol{\omega}), \tag{A 1.2}$$

where $g(\boldsymbol{\omega})$ is independent of ω_t and δ is the Dirac delta function. Substituting equations (2.1.5) and (A 1.2) into equation (A 1.1) (by using Cartesian coordinates $\boldsymbol{\Omega} = \Omega\boldsymbol{n} = (\Omega_x, \Omega_y)$) yields:

$$f(\boldsymbol{x}, t : \boldsymbol{\Omega}, \Omega_t, \sigma, \sigma_t) = \int \mathrm{e}^{-(\boldsymbol{\omega}-\boldsymbol{\Omega})^2\sigma^2/2}\, \mathrm{e}^{-(\boldsymbol{v}\cdot\boldsymbol{\omega}+\Omega_t)^2\sigma_t^2/2}\, g(\boldsymbol{\omega})\, \mathrm{e}^{-i\boldsymbol{\omega}\cdot(\boldsymbol{x}-\boldsymbol{v}t)}\, \mathrm{d}\boldsymbol{\omega}. \tag{A 1.3}$$

From equation (A 1.3) we see that N is the sum of the squares of two functions, f_1 and f_2 (the real and imaginary parts of f), of the form:

$$f_i(\boldsymbol{x}, t : \boldsymbol{\Omega}, \Omega_t, \sigma, \sigma_t) = \int \mathrm{e}^{-(\boldsymbol{\omega}-\boldsymbol{\Omega})^2\sigma^2/2}\, \mathrm{e}^{-(\boldsymbol{v}\cdot\boldsymbol{\omega}+\Omega_t)^2\sigma_t^2/2}\, g_i(\boldsymbol{x}, t : \boldsymbol{\omega}, \sigma, \sigma_t)\, \mathrm{d}\boldsymbol{\omega}, \tag{A 1.4}$$

where the g_i are the real and imaginary parts of $\mathrm{e}^{-i\boldsymbol{\omega}\cdot(\boldsymbol{x}-\boldsymbol{v}t)}$.

Hence the results will hold for N if they hold for f_1 and f_2. By using the summation convention ($a_i b_i = \sum_{i=1}^{N} a_i b_i$ and $a_i H_{ij} b_j = \sum_{i=1}^{N}\sum_{j=1}^{N} a_i H_{ij} b_j$) on

repeated indices i, j, k, l, p, but not on t, the argument of the exponent in the integrand of equation (A 1.4) can be written as:

$$(-1/2)\,((\boldsymbol{\Omega}-\boldsymbol{\omega})^2\sigma^2+(\Omega_t+\boldsymbol{v}\cdot\boldsymbol{\omega})^2\sigma_t^2)$$
$$=(-1/2)\,(\omega_i\omega_j(\delta_{ij}\sigma^2+v_iv_j\sigma_t^2)+2(\Omega_t\sigma_t^2v_i-\Omega_i\sigma^2)\,\omega_i+\sigma^2\Omega_i\Omega_i+\Omega_t^2\sigma_t^2). \tag{A 1.5}$$

We can complete the square and write it in the form:

$$(-1/2)\,(A_{ij}(\omega_i+A_{ki}^{-1}B_k)\,(\omega_i+A_{ij}^{-1}B_l)-B_kA_{kl}^{-1}B_l+\Phi), \tag{A 1.6}$$

where $A_{ij}=\delta_{ij}\sigma^2+v_iv_j\sigma_t^2$, $B_k=\Omega_t\sigma_t^2v_k-\sigma^2\Omega_k$ and $\Phi=\sigma^2\Omega_i\Omega_i+\sigma_t^2\Omega_t^2$.

The last two terms on the right hand side of equation (A 1.6) are independent of $\boldsymbol{\omega}$. Hence we can write:

$$f_i(\boldsymbol{x},t:\boldsymbol{\Omega},\Omega_t,\sigma,\sigma_t)=\mathrm{e}^{-(\Phi-B_kA_{kl}^{-1}B_l)/2}\int\mathrm{e}^{-A_{ij}(\omega_i+A_{ki}^{-1}B_k)(\omega_j+A_{lj}^{-1}B_l)/2}g_i(\boldsymbol{x},t:\boldsymbol{\omega})\,\mathrm{d}\boldsymbol{\omega}, \tag{A 1.7}$$

By calculating the inverse of A_{ij},

$$A_{ij}^{-1}=\frac{1}{\sigma^2}\delta_{ij}-[\sigma_t^2/(\sigma^2(\sigma^2+\sigma_t^2v_lv_l))]\,v_iv_j, \tag{A 1.8}$$

we find that

$$\Phi-B_iA_{ij}^{-1}B_j=[(\sigma_t^2\sigma^2)/(\sigma^2+\sigma_t^2v^2)]\,(\Omega_t+\Omega_iv_i)^2. \tag{A 1.9}$$

The integral term in equation (A 1.7) defines the function $p_i(\boldsymbol{x},t:\Omega_t\sigma_t^2\boldsymbol{v}-\sigma^2\boldsymbol{\Omega})$. It corresponds to convolving g_i with a function of $(\Omega_t\sigma_t^2\boldsymbol{v}-\sigma^2\boldsymbol{\Omega})$.

Defining $p=p_1^2+p_2^2$ proves the theorem.

Appendix 2

In this appendix, we give a proof for Theorem 3. These results can be derived from equation (A 1.6). We find by using equation (A 1.8):

$$A_{ij}(\Omega_i+A_{ki}^{-1}B_k)\,(\Omega_j+A_{lj}^{-1}B_l)=\sigma^2\left(\delta_{ij}+\frac{v_iv_j\sigma_t^2}{\sigma^2}\right)\left(\Omega_i-\omega_i+\frac{\sigma_t^2(\Omega_t+\boldsymbol{v}\cdot\boldsymbol{\Omega})\,v_i}{(\sigma^2+\sigma_t^2v^2)}\right)$$
$$\times\left(\Omega_j-\omega_j+\frac{\sigma_t^2(\Omega_t+\boldsymbol{v}\cdot\boldsymbol{\Omega})\,v_j}{(\sigma^2+\sigma_t^2v^2)}\right). \tag{A 2.10}$$

We now argue that we can approximate $A_{ij}(\Omega_i+A_{ki}^{-1}B_k)\,(\Omega_j+A_{lj}^{-1}B_l)$ by $\sigma^2(\Omega_i-\omega_i)\,(\Omega_i-\omega_i)$. From equation (A 1.7), we see that the response of the filter decays exponentially with $\Phi-B_iA_{ij}^{-1}B_j$. If we use the approximation $v^2\sigma_t^2\ll\sigma^2$ we can see that the ratio of the term $[\sigma_t^2(\Omega_t+\boldsymbol{v}\cdot\boldsymbol{\Omega})\,v_i/(\sigma^2+\sigma_t^2v^2)]$ to Ω_i is much smaller than $\sqrt{(\Phi-B_iA_{ij}^{-1}B_j)}$. Thus this term will only be important when the output of the filter is small, and hence we can set $\{\Omega_i-\omega_i+[\sigma_t^2(\Omega_t+\boldsymbol{v}\cdot\boldsymbol{\Omega})\,v_i/(\sigma^2+\sigma_t^2v^2)]\}\approx(\Omega_i-\omega_i)$. Again, by using the approximation $v^2\sigma_t^2\ll\sigma^2$ we can set $(\delta_{ij}+(v_iv_j\sigma_t^2/\sigma^2))\approx\delta_{ij}$. Thus we define $r(\boldsymbol{\Omega})=\exp\,(-(\sigma^2/2)\,(\omega_i-\Omega_i)\,(\omega_i-\Omega_i))$ and the result follows.

Appendix 3

Now we calculate the average response of motion-energy filters to moving sine wave gratings.

The equation for the gratings is:

$$I(x-vt) = I_1 \sin(\lambda n \cdot (x-vt)). \tag{A 3.1}$$

This has Fourier transform

$$\overline{I(\omega,\omega_t)} = (2\pi)^{\frac{1}{2}} I_1 \delta(\omega \cdot v + \omega_t)(1/2i)(\delta(\omega \cdot n - \lambda) - \delta(\omega \cdot n + \lambda))\delta(\omega \cdot n^*). \tag{A 3.2}$$

Substituting equation (A 3.2) into equation (A 1.1) and integrating with respect to (ω, ω_t), we obtain:

$$f(x,t:\Omega,n,\Omega_t,\sigma,\sigma_t) = I_1 e^{-(\lambda-\Omega)^2\sigma^2/2} e^{-(\lambda n\cdot v+\Omega_t)^2\sigma_t^2/2} e^{-i\lambda n\cdot(x-vt)} - I_1 e^{-(\lambda+\Omega)^2\sigma^2/2} e^{-(\lambda n\cdot v-\Omega_t)^2\sigma_t^2/2} e^{i\lambda n\cdot(x-vt)}. \tag{A 3.3}$$

Thus the response is:

$$N(x,t:\Omega,n,\Omega_t,\sigma,\sigma_t) = I_1^2 e^{-(\lambda-\Omega)^2\sigma^2} e^{-(\lambda n\cdot v+\Omega_t)^2\sigma_t^2} + I_1^2 e^{-(\lambda+\Omega)^2\sigma^2} e^{-(\lambda n\cdot v-\Omega_t)^2\sigma_t^2} - 2I_1^2 e^{-(\lambda^2+\Omega^2)\sigma^2} e^{-((\lambda n\cdot v)^2+\Omega_t^2)\sigma_t^2} \cos(2\lambda n\cdot(x-vt)). \tag{A 3.4}$$