PERGAMON

Vision Research 40 (2000) 3427–3434

Vision Research

www.elsevier.com/locate/visres

A Bayesian model for the measurement of visual velocity

David Ascher, Norberto M. Grzywacz *

The Smith-Kettlewell Eye Research Institute, 2318 Fillmore St, San Francisco, CA 94115, USA

Received 8 July 1999; received in revised form 9 May 2000

Abstract

Several models have been proposed for how the brain measures velocity from the output of motion-energy units. These models make some unrealistic assumptions such as the use of Gabor-shaped temporal filters, which are non causal, or flat spatial spectra, which are invalidated by existing data. We present a Bayesian model of velocity perception, which makes more realistic assumptions and allows the estimation of local retinal velocity regardless of the specific mathematical form of the spatial and temporal filters used. The model is consistent with several aspects of speed perception, such as the dependence of perceived speed on contrast. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Model; Motion; Velocity; Bayesian; Optimal

1. Introduction

There exist several models of velocity perception based on the distributed output of motion-energy units tuned in the spatial-frequency (SF) and temporal-frequency (TF) domains (Heeger, 1987; Grzywacz & Yuille, 1990; Schrater & Simoncelli, 1998). However, all of these models make one or more assumptions that are not realistic. Some models assume that the spatial frequency spectrum is flat (Heeger, 1987), which has been shown not to be true in measured spectra of natural scenes (e.g. Field, 1987; van der Schaaf & van Hateren, 1996; Ruderman, 1997). Other models do not make that assumption, but assume non-causal, Gabor temporal filters, because of their forgiving mathematical properties (Grzywacz & Yuille, 1990). These limitations could raise doubts about the validity of such models. Here, we present an alternative model, which addresses both of these concerns. One important feature of our model is that it was developed within the Bayesian framework. This means that the performance of the model reflects optimal treatment of internal noise.

2. Theory

2.1. Intuition

Fig. 1 is a schematic of the distributed responses to a stimulus translating at a constant velocity of a set of motion-energy units tuned to different SF and TF. As shown by Grzywacz and Yuille (1990), with some assumptions about the shapes of the temporal and spatial filters which underly the tuning of the mechanisms, the speed can be computed exactly by estimating the best-fitting line. Our model investigates how this speed can be estimated when some of those assumptions are violated, and other, less restrictive, assumptions are made instead.

Our model preserves the assumption by Grzywacz and Yuille (1990) that the temporal bandwidth of the motion energy unit is much broader than its spatial bandwidth. This is a realistic assumption as demonstrated by published estimates on these bandwidths (Watson, 1986; Wilson, McFarlane, & Phillips, 1983). However, unlike some other models (Grzywacz & Yuille, 1990; Heeger, 1987), our model does not assume that the temporal filters of the underlying motion-energy units have Gabor-shaped profiles. This is because these filters are non-causal (the response at a given time depends on the stimulus later in time) and the shapes of the temporal filters which have been found to fit psy-

* Corresponding author. Tel.: +1-415-5611795; fax: +1-415-3458455.

E-mail address: nmg@ski.org (N.M. Grzywacz).

PII: S0042-6989(00)00176-0

chophysical data do not have a Gabor-like profile (Watson, 1986; Fredericksen & Hess, 1998; Ascher & Grzywacz, 2000). Nevertheless, the filters are still linear and their outputs are processed by a nonlinearity (such as a half-wave rectifying power law, Heeger, 1993; Simoncelli & Heeger, 1998), a requirement which is consistent with known physiology of complex cells in visual cortex (Emerson, Bergen, & Adelson, 1992). The model also requires that within the SF bandwidth of a given filter, the power spectrum of any given stimulus varies little. Because natural image spectra are not flat (Field, 1987; van der Schaaf & van Hateren, 1996; Ruderman, 1997), this requirement leads to a piecewise flat approximation of the spectrum. This approximation argues for a large number of narrowly tuned channels, which is consistent both with psychophysical

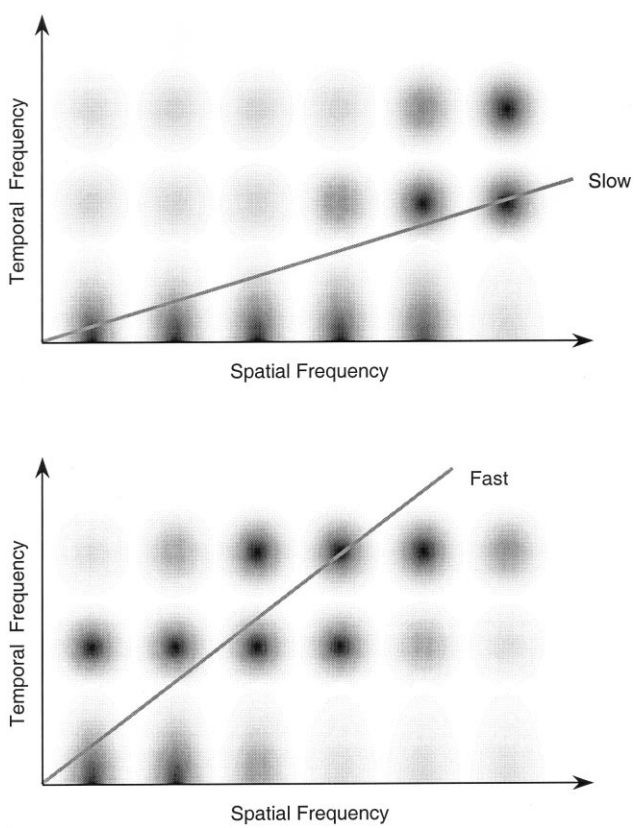

Fig. 1. This figure illustrates the responses of 18 spatial- and temporal-frequency tuned channels. A column corresponds to the set of filters with a shared SF tuning, while a row corresponds to filters with shared TF tuning. For clarity of illustration, the channels do not overlap in the schematic — evidence suggests that their sensitivity profiles overlap considerably. Such an arrangement of channels is the core of a general model of speed perception (Heeger, 1987; Grzywacz & Yuille, 1990; Simoncelli & Heeger, 1998). In the top panel, the distribution of activities for a slowly moving stimulus are depicted by the varying levels of gray, along with the best-fitting line, whose slope corresponds to the most likely speed of the stimulus. In the bottom panel, the distribution for the same stimulus moving faster is depicted, along with its best-fitting line.

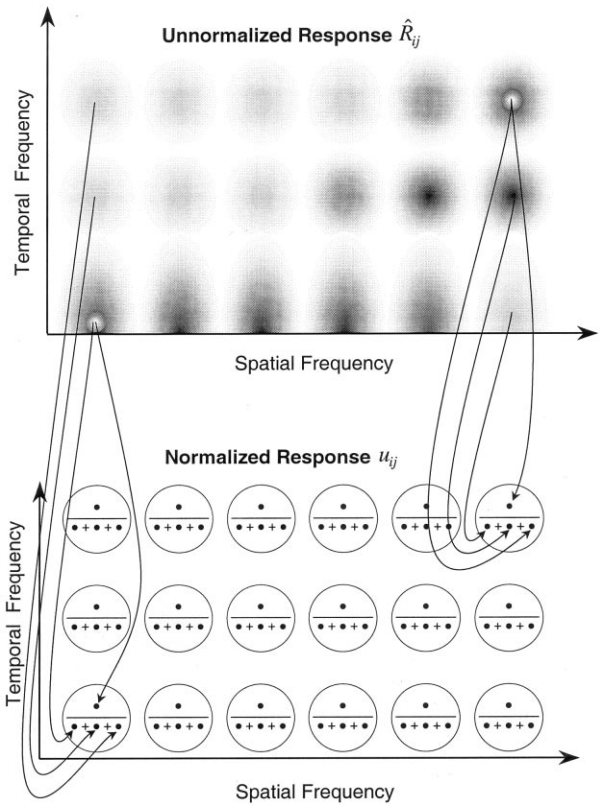

Fig. 2. This figure illustrates the normalization process described in Eq. (4), using the same stimulus as for the top panel of Fig. 1. The normalization of two of the filters is depicted. The circles on the bottom panel indicate a ratio of a filter's response by the sum of the responses of all the filters with same SF tuning. Only two normalizations are fully illustrated, to avoid clutter.

(Wilson et al., 1983; Watson, 1986) and physiological data (Holub & Morton-Gibson, 1981). Thus, in each spatial-frequency band, the response is independent of the spatial-frequency content of the scene. It follows that the spatial-frequency content can be factored out of the velocity-estimation process by a process analogous to the contrast normalization used in some other models (Heeger, 1992; Carandini, Heeger, & Movshon, 1996). Finally, the motion of the stimulus is assumed to be at least as long as the integration period of the temporal filter and the stimulus size is assumed to be much larger than the size of the receptive field. This last assumption allows us, as in Grzywacz and Yuille (1990), to limit the analysis to locally translatory motions.

Given these assumptions, we can compute velocity as sketched in Fig. 1, independently of the spatial structure of the stimulus. To eliminate spatial-structure dependence, the model takes advantage of the piecewise-flat spatial spectrum to normalize the responses of motion-energy unit outputs across TF bands and within a SF channel, as derived below and illustrated in Fig. 2.

If the filters were noise-free, then these idealized normalized responses would allow the exact computation of the stimulus velocity. Because there is noise in the motion sensing process, we derive an optimal estimate of speed, which uses knowledge of the noise statistics. We use the Bayesian framework in the derivation below as illustrated in Fig. 3.

2.2. Mathematical derivation

The model consists of a set of filters F_{ij}, where $i \in 1 \ldots n_r$ and $j \in 1 \ldots n_t$ with n_r and n_t being the number of spatial- and temporal-frequency tunings optima, respectively. The response R_{ij} of each filter F_{ij} to a stimulus S at position $\mathbf{r}$ and time t is

$$R_{ij}(\mathbf{r}, t) = \left|S(\mathbf{r}, t) * F_{ij}(\mathbf{r}, t)\right|^m \tag{1}$$

where $*$ stands for convolution and m is an exponent to be determined (see Section 4). The convolution can be expressed using Fourier's convolution theorem as

$$S * F_{ij}(\mathbf{r}, t) = \frac{1}{(2\pi)^2} \int \mathrm{d}\omega_{\mathbf{r}}\, \mathrm{d}\omega_t\, \tilde{S} \cdot \tilde{F}_{ij}(\omega_{\mathbf{r}}, \omega_t)\, \mathrm{e}^{-i(\omega_{\mathbf{r}} \cdot \mathbf{r} + \omega_t t)}$$

where $\omega_{\mathbf{r}}$ and ω_t, $\tilde{S}$ and $\tilde{F}$ are spatial frequency (times 2π), temporal frequency (times 2π), the Fourier transform of the stimulus, and the Fourier transform of the filter, respectively. If we assume that the receptive field is small compared to the stimulus size, then we can limit our analysis to translatory motions, which allows us to factor $\tilde{S}$ into a component dependent only on the spatial structure of the stimulus, $g(\omega_{\mathbf{r}})$, and a velocity dependent component, that is, $\tilde{S} = g(\omega_{\mathbf{r}})\delta(\omega_{\mathbf{r}} \cdot \mathbf{v} + \omega_t)$, where δ is the Dirac delta function (Grzywacz & Yuille, 1990). Thus,

$$S * F_{ij}(\mathbf{r}, t) = \frac{1}{(2\pi)^2} \int \mathrm{d}\omega_{\mathbf{r}}\, g(\omega_{\mathbf{r}}) \tilde{F}_{ij}(\omega_{\mathbf{r}}, -\omega_{\mathbf{r}} \cdot \mathbf{v})\, \mathrm{e}^{-i(\omega_{\mathbf{r}} \cdot \mathbf{r} - \omega_{\mathbf{r}} \cdot \mathbf{v} t)} \tag{2}$$

As we assume that the SF spectrum of the stimulus varies little within the bandwidth of the SF channels, we can take the spatial component out of the integral by computing the value of g at the channel's center frequency, $\mathbf{\Omega}_{\mathbf{r}}$, and obtain

$$S * F_{ij}(\mathbf{r}, t) = \frac{1}{(2\pi)^2} g(\mathbf{\Omega}_{\mathbf{r}}) \int \mathrm{d}\omega_{\mathbf{r}}\, \tilde{F}_{ij}(\omega_{\mathbf{r}}, -\omega_{\mathbf{r}} \cdot \mathbf{v}) \times \mathrm{e}^{-i\omega_{\mathbf{r}} \cdot (\mathbf{r} - \mathbf{v}t)}$$

By using the definition of the Fourier transform for $\tilde{F}$ and putting this equation back into Eq. (1), we get

$$\hat{R}_{ij} = \left|g(\mathbf{\Omega}_{\mathbf{r}})\right| \left| \int \mathrm{d}t' F_{ij}\, \mathbf{r} + \mathbf{v}(t' - t), t' \right|^m$$

where $\hat{R}_{ij}$ is the ideal response if all the assumptions hold and in the absence of noise. If we assume that the stimulus duration is long compared to the filter integration constant and that the stimulus is large compared to the receptive field, then we obtain a steady state solution

$$\hat{R}_{ij} = \left|g(\mathbf{\Omega}_{\mathbf{r}})\right|^m \left| \int \mathrm{d}t' F_{ij}(t'\mathbf{v}, t') \right|^m \tag{3}$$

which is independent of position $\mathbf{r}$ or time t.

To eliminate from the model's estimation of velocity the dependence on the spatial structure, $g(\mathbf{\Omega}_{\mathbf{r}})$, we can normalize the response of all of the filters which fall within a spatial frequency band. The *ideal*, normalized response of filter F_{ij} to a given stimulus is thus

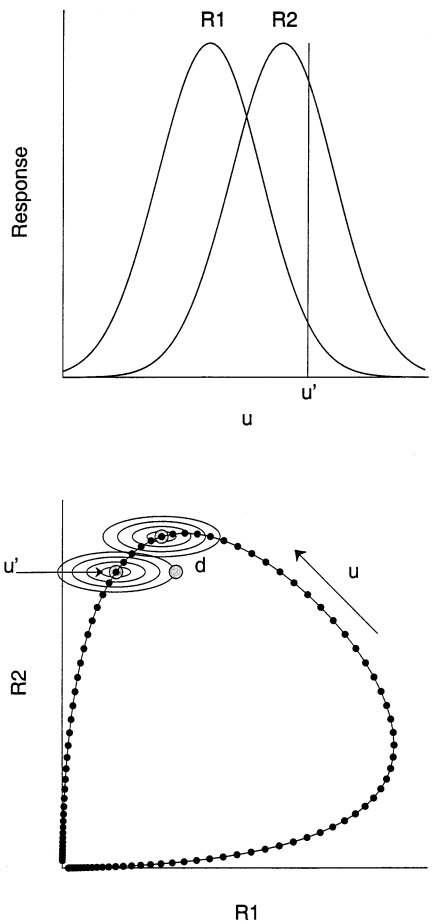

Fig. 3. This figure illustrates the process of estimating the stimulus (in our case, velocity), which is most likely to account for a given noisy measurement. The top panel depicts two detectors with overlapping sensitivity curves. Most stimulus values (u) lead to partial activation of the two filters $R1$ and $R2$. This is depicted in the bottom panel, which is a phase plot; each point corresponds to a given value of u, with the abscissa and ordinates specified by the values of $R1$ and $R2$, respectively. The task of the model is to find the u that is most likely given a measurement d. That u will not be necessarily closest to d in a linear metric, but that which requires adding the least amount of noise to yield the measurement d. In this schematic, the u' for the left most gray dot is more likely than the u for the topmost gray dot because, while the latter is absolutely closer to **d**, the latter is farther from **d** in terms of standard deviations of the noise (as illustrated by the smaller axis of the noise ellipses along the $R2$ dimension).

$$u_{ij} = \frac{\hat{R}_{ij}}{\sum_{j=1}^{n_t} \hat{R}_{ij}} = \frac{\left| \int \mathrm{d}t F_{ij}(t'\mathbf{v}, t') \right|^m}{\sum_{j=1}^{n_t} \left| \int \mathrm{d}t F_{ij}(t'\mathbf{v}, t') \right|^m} \quad (4)$$

In contrast, a good model for the *actual* response of the filter to a real stimulus S is R_{ij}, and its corresponding measurable, noisy, normalized output is

$$d_{ij} = \frac{(R_{ij} + \varepsilon_{r_{ij}})(1 + \varepsilon_{c_{ij}})}{\sum_{j}^{n_t} (R_{ij} + \varepsilon_{r_{ij}})} \quad (5)$$

where $\epsilon_{r_{ij}}$ is an additive noise due to subcortical processes and $\epsilon_{c_{ij}}$ is a multiplicative noise at the level of the cortical-normalization process.[1] The normalization is assumed to occur over a large number of cells (even though they may come from a small number of channels), and the noise $\epsilon_{r_{ij}}$ is assumed to be zero-mean and independent across cells. Thus, the sum of $\epsilon_{r_{ij}}$ in the denominator becomes negligible compared to the sum over R_{ij} and is neglected in the following derivations. A final assumption is that $\epsilon_{r_{ij}} \ll R_{ij}$ and $\epsilon_{c_{ij}} \ll 1$. Hence, one can expand Eq. (5) and neglect the term containing $\varepsilon_{r_{ij}} \varepsilon_{c_{ij}}$, to obtain

$$d_{ij} \approx u_{ij}(1 + \varepsilon_{c_{ij}} + \varepsilon'_{r_{ij}}) \quad (6)$$

where $\varepsilon'_{r_{ij}} = \varepsilon_{r_{ij}}/R_{ij}$.

The goal of the model is to estimate the velocity $\mathbf{v}$ from the measurements $\mathbf{d} = \{d_{ij}\}$. Thus, we wish to find the $\mathbf{v}$ that maximizes the conditional probability $P(\mathbf{v}|\mathbf{d})$. Using Bayes' theorem, we get

$$P(\mathbf{v}|\mathbf{d}) = \frac{P(\mathbf{d}|\mathbf{v})P(\mathbf{v})}{P(\mathbf{d})} \quad (7)$$

Because $\mathbf{d}$ is given, $P(\mathbf{d})$ is just a normalizing constant. In turn, $P(\mathbf{v})$ is a prior distribution and would ideally be measured from natural images (a first attempt was made by Dong & Atick, 1995). As argued by many, the distribution $P(\mathbf{v})$ is not flat, being biased towards slow speeds (Ullman & Yuille, 1989; Weiss & Adelson, 1998). This bias leads the model to disambiguate speeds in favor of slower speeds. Finally, we must specify $P(\mathbf{d}|\mathbf{v})$. If one assumes that the filters have independent sources of noise, then we can estimate $\mathrm{P}(\mathbf{d}|\mathbf{v})$ as

$$P(\mathbf{d}|\mathbf{v}) = \prod_{i,j} P(d_{ij}|\mathbf{v}) \quad (8)$$

The probability of a normalized response d_{ij} given a speed $\mathbf{v}$ is the probability of that response given the ideal response u_{ij}, since that response is deterministically related to $\mathbf{v}$. We model this probability through a zero-mean Gaussian noise with variance $u_{ij}^2\sigma_c^2$:

$$P(d_{ij}|\mathbf{v}) = \frac{1}{\sqrt{2\pi}u_{ij}\sigma_c} \mathrm{e}^{-((d_{ij}-u_{ij})^2/(2u_{ij}^2\sigma_c^2))} \quad (9)$$

This model assumes that the brain uses as a prior that moving stimuli have sufficiently high contrast so that $\epsilon'_{r_{ij}} \ll \epsilon_{c_{ij}}$. As we will see in the next section, this assumption leads to interesting effects when one uses low-contrast stimuli in the laboratory.

How do we calculate from these equations the most likely speed given the actual normalized responses of the filters (d_{ij})? By substituting Eq. (4) for u_{ij} in Eq. (9), one gets an expression of $P(d_{ij}|\mathbf{v})$, which is explicit on $\mathbf{v}$. Inserting this expression in Eq. (8), gives the explicit dependence on $\mathbf{v}$ of $P(\mathbf{d}|\mathbf{v})$. This dependence multiplied by $\mathrm{P}(\mathbf{v})$ provides a formula for $P(\mathbf{v}|\mathbf{d})$ explicitly on $\mathbf{v}$ (Eq. (7)). This formula is all we need to estimate the most likely velocity given the filter data. We find the $\mathbf{v}$ that maximize this probability $P(\mathbf{v}|\mathbf{d})$. In other words, we use the maximum likelihood of this probability distribution to identify the velocity $\mathbf{v}$ that is most likely to have yielded the observed responses $\mathbf{d}$.

3. Simulations

Some interesting psychophysical and physiological results on velocity perception can be accounted for by this model. One well-known psychophysical result is the effect of contrast on perceived speed. As shown by Thompson (1982), at temporal frequencies below 8 Hz, a decrease in contrast reduces perceived speed, while at higher temporal frequencies, reductions in contrast increase perceived speed. To test whether the model can account for this contrast effect, we implemented our model with qualitatively realistic filters. The impulse responses of the filters were Alpha functions, as used by Watson and Ahumada (1985) (but the argument in the next paragraph shows that this choice is not crucial). Using Watson's notation, the parameters for our low-pass filter were $\tau = 4.12$ ms and $n = 9$. The bandpass filter used $\tau = 3.88$ ms, $\kappa = 1.33$, $n_1 = 9$, $n_2 = 10$, and $\xi = 0.8$. The specific values of these parameters are somewhat meaningless as the profiles chosen are only rough approximations of the channel profiles. The important features of these profiles is that they overlap over the relevant range of temporal frequencies, and that one filter is more sensitive at the low-TF range, while the other is more sensitive at the high-TF range. Prior distributions were computed assuming $\sigma_c = 0.05$. Test filters were then calculated for a variety of contrasts with an additive noise of $1.5\,\mathrm{e}^{-5}$, and a probability distribution was computed from these responses for each velocity and test contrast. The velocity match was chosen to be the velocity that corresponded to the peak

[1] Physiological studies have consistently shown that cortical noise is approximately proportional to response, that is, multiplicative (Snowden, Treue, Erickson, & Andersen, 1991; Tolhurst, Movshon, & Dean, 1983).

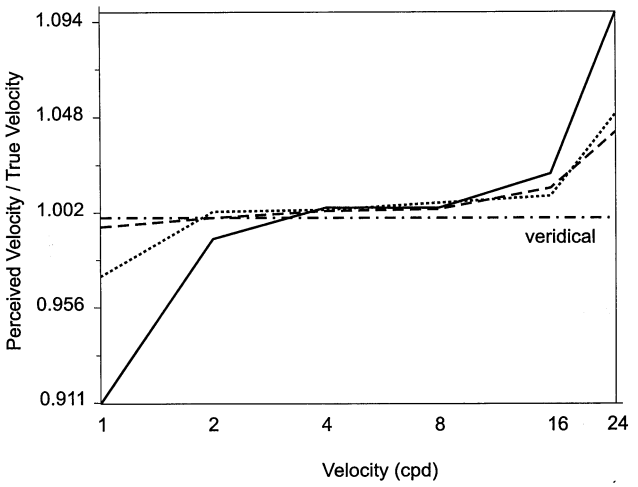

Fig. 4. Simulation results. This plot shows the computed ratio of perceived speed by veridical speed as stimulus speed varies for a variety of stimulus contrasts. The thick black line corresponds to a contrast of 4%, the dotted line to a contrast of 8% and the dashed line to a contrast of 40%. Whereas at high contrasts, perceived speed is veridical, at low contrast, perceived speed is underestimated and overestimated for low and high speeds, respectively.

of that probability distribution. Fig. 4 shows the results of a simulation of our model. The model shows the same qualitative behavior as human observers (see for example Figs. 1 and 2 in Thompson, 1982). Specifically, at low contrast, the model underestimates perceived velocity at low velocities, and overestimates it at high velocities. Judgments become more veridical as contrast increases.

An intuition for how the misperception of velocity at low contrasts occurs can be drawn by analyzing the schematic of the filter-response distributions shown in Fig. 5. The top panel shows the probability distribution for the low-pass filter's normalized response (labeled d_0) as a function of velocity. As can be seen, high normalized responses for the low-pass filter are associated with low velocities, and low normalized responses are associated with high velocities. The converse relationships are shown in the middle panel for the high-pass filter. The process of determining perceived speed from these filters given a specific stimulus speed is shown in the lower panel. For the speed shown (chosen low for purposes of illustration), the distributions of $P(\mathbf{d}|\mathbf{v})$ for the two filters can be sliced out of the upper panels. The multiplication of those two probability distributions yields a joint distribution, the peak of which determines the chosen, perceived velocity. At low velocities, the $P(\mathbf{d}|\mathbf{v})$ distribution is *broader* for the lowpass filter than for the highpass filter (this is because the multiplicative noise has more impact when the response is larger — see lower panel). The narrower highpass distribution has the most impact on the location of the peak in the compound probability distribution. In other words, the perceived velocity at low velocities is determined mostly by the highpass filter responses (and conversely at high velocities).

To understand the net effect of additive noise on average perceived velocity, one then needs to understand how small shifts in the response **d** of the high-pass filter affect the perceived velocity. As can be seen in the middle panel, a small shift up in **d** shifts the perceived speed to a slightly higher velocity. A small shift *down* in **d**, on the other hand, shifts the perceived velocity to very low values, as the filter flattens out at low velocities. According to our model, this imbalance in the effect on perceived velocity on the highpass filter probability map is the source of the effect reported by Thompson — at low contrast, slow stimuli appear to move slower. The converse effect, also noted by Thompson, is explained by a similar analysis at high velocities, and is controlled by the shape of the lowpass filter.

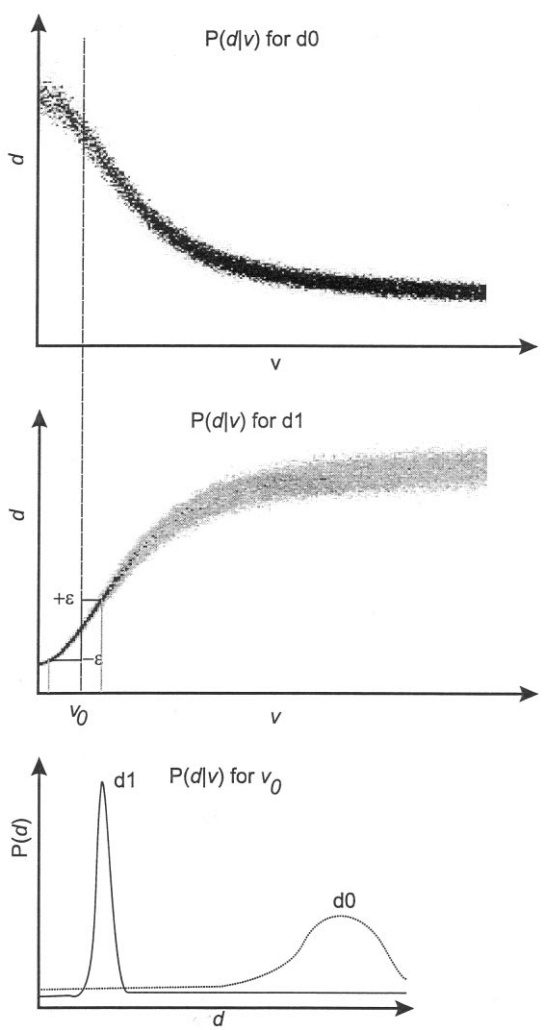

Fig. 5. Schematic of the density functions ($P(\mathbf{d}|\mathbf{v})$) for a lowpass (upper panel) and bandpass (middle panel) filters, and an explanation of how the bandpass filter determines perceived speed at low speeds and at low contrasts. See text for details.

At low contrast, the contribution of the additive noise $\epsilon_{r_{ij}}$ is not negligible. This is shown in Fig. 5 by the fairly broad band of responses. As a result, at low contrast, the noise of the system fails the prior assumptions (the assumption of no additive noise). On the other hand, at high contrast, the additive noise becomes relatively negligible compared to the absolute responses and thus, there is no violation of the prior assumptions by the noise, causing the estimated speed to become more veridical.

We also would like to point out another result that this model explains due to the use of multiplicative noise. Several studies have examined the contrast-dependence of neuronal noise and found it to scale with mean firing rate (e.g. Tolhurst, Movshon, & Dean, 1983; Snowden et al., 1991). As soon as the contrast level reaches a certain level, most of the noise in our model follows the same multiplicative relationship.

4. Discussion

How reasonable are the assumptions of the model? We already justified the assumptions of wide temporal and narrow spatial bandwidths. Here, we address four more: first, the assumption of linear detectors is consistent with physiological data (Ikeda & Wright, 1975; Tolhurst, Movshon, & Thompson, 1981; De Valois & De Valois, 1990) and appears in a wide range of models of motion (Adelson & Bergen, 1985; Watson & Ahumada, 1985). Second, the model uses the absolute value raised to a power as the nonlinearity on the output of the detectors. Other nonlinearities are allowable. In Appendix A, we show that any half-wave or full-wave rectifying power function can be used. However, any other nonlinearity would not allow the factorization of $g(\mathbf{\Omega}_\mathbf{r})$ and the TF-dependent term in Eq. (3) (see Appendix A). Hence, this provides a strong rationale for why the brain might use rectifying power functions as its filter nonlinearities, as in models of motion perception (Heeger, 1993; Simoncelli & Heeger, 1998). Third, the assumption of independence of filters is also very common, although work by Zohary, Shadlen, and Newsome (1994) has shown that due to anatomical connections, cells in MT may not have completely independent noise. A way to incorporate inter-channel noise dependence is to eliminate Eq. (8) and to use multidimensional Gaussian distributions with cross d_{ij} terms instead of Eq. (9). Such changes would lead to diagonal ellipses in Fig. 3. Fourth, the last assumptions, namely those regarding the small size of the receptive field and short integration time relative to the size and duration of the stimulus respectively, are important; they allow us to consider only translatory (Eq. (1)), steady state (Eq. (3)) motions. One can predict the errors that the model may make in cases where either or both of these assumptions are not met. These may correspond to human illusions or biases, such as the increase in apparent speed for short-duration stimuli reported by Giaschi and Anstis (1989), that radial motion appears to move faster than translatory motion (Bex & Makous, 1997), and that apparent speed of rotating displays varies with display scale (Werkhoven & Koenderink, 1993).

Our model's use of normalization is not unusual. Several models have proposed that the output of V1 cells is normalized by the aggregate response of a set of related cells (Grzywacz & Yuille, 1990; Smith & Grzywacz, 1993; Heeger, 1991, 1993; Simoncelli & Heeger, 1998). Our model proposes that the purpose of the normalization is to allow the velocity-computation system to discount variations in the spatial structure of the stimulus, not just contrast as in other models. More generally, we postulate that the brain may use selective normalization to extract many types of stimulus invariance at various hierarchical levels of the visual pathway.

Acknowledgements

We would like to thank Dr Leslie Welch for her participation in the early development of this model. This work was supported by National Eye Institute grant EY06878-01 and a Rachel C. Atkinson Fellowship to D.A.; by Air Force Office of Sponsored Research grant FY96209810197, and National Eye Institute grants EY08921 and EY11170 to N.M.G., and by National Eye Institute Core grant EY06883 to the Smith-Kettlewell Eye Research Institute.

Appendix A

In the derivation of Eq. (3), we used as our nonlinearity on the output of the convolution the absolute value raised to a power. We now investigate the general form of the nonlinearities that can be used in the derivation above. For the derivation to hold, the nonlinearity cannot interfere with the factorization of the spatial component $g(\mathbf{\Omega}_\mathbf{r})$. In other words, we need to know what nonlinear functions f satisfy $f(xy) = f_1(x)f_2(y)$, where f_1 and f_2 are two unknown functions. By setting $y = 1$, one gets $f(x) = f_1(x)f_2(1)$ and thus if $f(x) \neq 0$, then $f(x)$ and $f_1(x)$ are proportional. Similarly, $f(x)$ and $f_2(x)$ are proportional. Hence,

$$\kappa f(xy) = f(x)f(y) \tag{10}$$

where κ is a constant. Let us explore the consequences of Eq. (10) for the nonlinearity f. When one sets $y = 0$, one obtains $f(0) = \kappa^{-1}f(0)f(x)$, and thus $f(0) = 0$ (since as we show below, $f(x)$ is not a constant). Given that $\kappa f(x) = f(1)f(x)$, as long as $f(x) \neq 0$, $f(1) = \kappa$. Moreover, $f(x) = \kappa^{-1}f(-1)f(-x)$, which leads to $f(x) =$

$\kappa^{-2}f(-1)^2f(x)$, and so if $f(x)\neq 0$, then $f(-1)=\pm\kappa$. Without loss of generality, we limit ourselves to $\kappa\geq 0$.

To derive a general form for f from Eq. (10), we consider the derivative of f

$$\dot{f}(x)=\lim_{\epsilon\to 0}\frac{f(x+\epsilon)-f(x)}{\epsilon}$$

Using Eq. (10), we get

$$\dot{f}=\lim_{\epsilon\to 0}\frac{\kappa^{-1}f(x)f(1+\epsilon/x)-f(x)}{\epsilon}$$

Factoring out $f(x)$ and defining $\epsilon'=\epsilon/x$, we obtain

$$\dot{f}=\lim_{\epsilon'\to 0}\frac{f(x)}{x}\frac{\kappa^{-1}f(1+\epsilon')-1}{\epsilon'}=\frac{f(x)}{x}\kappa^{-1}\dot{f}(1)$$

Since $\kappa^{-1}f(1)=1$ as shown above.
Defining $\alpha=\kappa^{-1}\dot{f}(1)$, one gets

$$x\dot{f}(x)=\alpha f(x) \tag{11}$$

As this is a first-order, homogeneous, ordinary differential equation, there is guaranteed to be a unique solution for each initial condition. When $x=0$, $\dot{f}(x)$ is undefined in Eq. (11), since $\dot{f}(0)=\alpha f(0)/0$ and $f(0)=0$. Thus there are two different ranges that must be addressed independently: $x>0$ and $x<0$. When $x>0$, the initial condition is uniquely specified by $f(1)=\kappa$. The solution is $f(x)=\kappa x^{\alpha}$, as one can verify by inserting it in Eq. (11) and checking the initial condition. When $x<0$,$f(x)=\kappa^{-1}f(-1)f(-x)=\kappa^{-1}f(-1)\kappa(-x)^{\alpha}=\pm\kappa(-x)^{\alpha}$.

Hence, the general form of the nonlinearity that one can use instead of $|\cdot|^m$ to derive Eq. (3) is

$$f(x)=\begin{cases}\kappa x^{\alpha} & x\geq 0\\ \pm\kappa(-x)^{\alpha} & x<0\end{cases} \tag{12}$$

Eq. (12) includes $|\cdot|^m$ as a particular case. If we abandon the requirement that $f(x)\neq 0$ and require that $x>0$ in the arguments following Eq. (10) (corresponding in the velocity model to a requirement that $g(\mathbf{\Omega}_r)>0$), then we can also accommodate a half-wave rectifying nonlinearity

$$f(x)=\begin{cases}\kappa x^{\alpha} & x\geq 0\\ 0 & x<0\end{cases}$$

References

Adelson, E. H., & Bergen, J. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America*, *2*(2), 284–299.

Ascher, D., & Grzywacz, N.M. (2000). A Bayesian model of temporal frequency masking. *Vision Research*, *40*(16), 2219–2232.

Bex, P. J., & Makous, W. (1997). Radial motion looks faster. *Vision Research*, *37*(23), 3399–3405.

Carandini, M., Heeger, D., & Movshon, J. (1996). Cerebral cortex. In: E. Jones & P. Ulinski, *Cortical models*, *vol. XII*. New York: Plenum Press.

De Valois, R. L., & De Valois, K. K. (1990). *Spatial vision*. Oxford: Oxford Science Publications.

Dong, D. W., & Atick, J. J. (1995). Statistics of natural time-varying images. *Network*: *Computation in Neural Systems*, *6*(3), 345–358.

Emerson, R. C., Bergen, J. R., & Adelson, E. H. (1992). Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research*, *32*(2), 203–218.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, *4*(12), 2379–2394.

Fredericksen, R., & Hess, R. (1998). Estimating multiple temporal mechanisms in human vision. *Vision Research*, *38*(7), 1023–1040.

Giaschi, D., & Anstis, S. (1989). The less you see it, the faster it moves: shortening the 'on-time' speeds up apparent motion. *Vision Research*, *29*(3), 335–347.

Grzywacz, N. M., & Yuille, A. L. (1990). A model for the estimate of local image velocity by cells in the visual cortex. *Proceedings of the Royal Society of London Series B*, *239*, 129–161.

Heeger, D. (1987). Model for the extraction of image flow. *Journal of the Optical Society of America*, *4*(8), 1455–1471.

Heeger, D. (1993). Modeling simple cell direction selectivity with normalized, half-squared, linear operators. *Journal of Neurophysiology*, *70*, 1885–1898.

Heeger, D. J. (1991). Nonlinear model of neural responses in cat visual cortex. In M. Landy & J. A. Movshon, Computational models of visual processing. (pp. 119–133). Cambridge, MA. MIT Press.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, *9*, 181–198.

Holub, R. A., & Morton-Gibson, M. (1981). Response of visual cortical neurons of the cat to moving sinusoidal gratings: response-contrast functions and spatiotemporal interactions. *Journal of Neurophysiology*, *46*(6), 1244–1259.

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

Ruderman, D. L. (1997). Origins of scaling in natural images. *Vision Research*, *37*, 3385–3389.

Schrater, P. R., & Simoncelli, E. P. (1998). Distributed velocity representation: evidence from motion adaptation. *Vision Research*, *38*(24), 3899–3912.

Simoncelli, E. P., & Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision Research*, *38*(5), 743–761.

Smith, J., & Grzywacz, N. M. (1993). A local model for transparent motions based on spatio-temporal filtering. In: F. Eeckman & J. Bower, *Computation and neural systems 1992*. Dordrecht: Kluwer Academic Press.

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

Thompson, P. (1982). Perceived rate of movement depends on contrast. *Vision Research*, *22*, 377–380.

Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research*, *23*, 775–785.

Tolhurst, D. J., Movshon, J. A., & Thompson, I. D. (1981). The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Experimental Brain Research*, *41*(3–4), 414–419.

Ullman, S., & Yuille, A. (1989). Rigidity and smoothness of motion. In: S. Ullman & W. Richards, *Image understanding*. Norwood, NJ: Ablex Publishing Corporation.

van der Schaaf, A., & van Hateren, J. H. (1996). Modelling the power spectra of natural images: statistics and information. *Vision Research*, *36*, 2759–2770.

Watson, A. B. (1986). Temporal sensitivity. In: K. R. Boff, L. Kaufman, & J. P. Thomas, Handbook of perception and human performance. (vol. I). New York: Wiley.

Watson, A. B., & Ahumada, A. J. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A*, *2*, 322–341.

Weiss, Y., & Adelson, E. (1998). *Slow and smooth: a Bayesian theory for the combination of local motion signals in human vision* (AI Memo No. 1624). Massachusetts Institute of Technology.

Werkhoven, P., & Koenderink, J. J. (1993). Visual size invariance does not apply to geometric angle and speed of rotation. *Perception*, *22*(2), 177–184.

Wilson, H. R., McFarlane, D. K., & Phillips, G. C. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research*, *23*, 873–882.

Zohary, E., Shadlen, M. N., & Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, *370*(6485), 140–141.