Vision Res. Vol. 33, No. 2, pp. 257–267, 1993
Printed in Great Britain.

0042-6989/93 $5.00 + 0.00

Spatiotemporal Contrast Sensitivity of Early Vision

J. H. VAN HATEREN*

Received 16 March 1992; in revised form 1 August 1992

Based on the spatial and temporal statistics of natural images, a theory is developed that specifies spatiotemporal filters that maximize the flow of information through noisy channels of limited dynamic range. Sensitivities resulting from these spatiotemporal filters are very similar to the human spatiotemporal contrast sensitivity, including the dependence on ambient light intensity. The theory predicts several psychophysical laws: Ferry–Porter's law, the de Vries–Rose law, Weber's law, Bloch's law, Ricco's law, and Piper's law.

Natural images Human vision Weber's law Bloch's law Ricco's law

INTRODUCTION

Barlow (1961) proposed that a major task of early vision is the reduction of redundancy present in natural images. By removing redundancy through lateral inhibition in the spatial domain and self-inhibition in the temporal domain, the incoming information is conditioned to fit more efficiently into the channels transporting it to higher brain centres. Similar ideas were formulated and explored by, for example, Laughlin (1981, 1983, 1987) and Srinivasan, Laughlin and Dubs (1982).

Indeed, work on the statistics of television images (Kretzmer, 1952) and more recent work by Field (1987) and Huang and Turcotte (1990) show that natural images have remarkably constrained statistics. Moreover, the temporal structure as perceived by an animal's visual system depends mostly on the movements of the animal itself, and will therefore possess characteristic statistics as well. These spatial and temporal characteristics are such that much of the spatial and temporal information in natural images is predictable, and thus redundant.

Although reducing redundancy is a good strategy for images with good signal-to-noise ratios, it can be counterproductive if the signal-to-noise ratio becomes small (such as at low ambient light intensities, due to photon noise). Then it can even be a better strategy to increase redundancy, by spatial pooling and temporal smearing, in order to obtain more reliable signals. A general strategy that works for arbitrary signal-to-noise ratios is the principle of maximizing the amount of information transferred to the brain (see e.g. Snyder, Laughlin & Stavenga, 1977, for an application to the theory of sampling and eye design).

In this article I apply this principle of maximizing information to spatiotemporal processing in the human visual system. Information on the spatiotemporal structure of natural images is combined with known properties of the eye's optical apparatus and of the temporal properties of cones. On the assumption that the visual system samples its surroundings through an array of noisy channels of limited dynamic range, the theory results in the construction of a spatiotemporal filter that maximizes the flow of information through each channel. Interestingly, spatiotemporal filters thus constructed are similar to the human spatiotemporal contrast sensitivity measured psychophysically, including the dependence on ambient light intensity. Moreover, the theory predicts several psychophysical laws: Ferry–Porter's law (the critical flicker frequency depends linearly on the logarithm of the background light intensity), the de Vries–Rose law (sensitivity proportional to the square root of the background light intensity), Weber's law (contrast constancy), Bloch's law (threshold contrast inversely proportional to stimulus duration for short durations), Ricco's law (threshold contrast inversely proportional to stimulus area for small areas), and Piper's law (threshold contrast proportional to the square root of stimulus area for larger areas).

THEORY AND RESULTS

Below I will first discuss the spatiotemporal structure of natural images, then outline and explain a general theory of early vision aimed at maximizing information transfer, and subsequently apply this to the human visual system.

Spatiotemporal structure of natural images

Recently, Field (1987) showed that the spatial power spectra of several natural images depend on the spatial

*Department of Biophysics, University of Groningen, Nijenborgh 4, NL-9747 AG Groningen, The Netherlands.

frequency, f_s, as $1/f_s^2$. Similar results were obtained by Burton and Moorhead (1987), and by Huang and Turcotte (1990) on satellite images of the surface of the earth. To investigate this further, I computed the power spectra of 117 images of widely varying natural scenes (see van Hateren, 1992, for details), and found that virtually all are fitted well by a straight line when the spectra are drawn in a double logarithmic plot. The mean and standard deviation of the slopes are -2.13 ± 0.36, thus confirming Field's results. For the calculations below I will assume a $1/f_s^2$-behaviour.

Interestingly, the power spectrum of a step function behaves as $1/f_s^2$, and it is probably the abundance of edges in natural images (correlated to object boundaries) that produces their $1/f_s^2$-behaviour. Thus, although the power spectra discussed here are a global property of images, similar considerations apply to more local parts of images, as long as these contain edges.

Laughlin (1983) found that the average contrast of natural scenes is about 40%, and I will use this value below. (My set of images did not allow an accurate estimate of the average contrast of natural scenes; see the Appendix for a definition of contrast.)

Most of the temporal variation encountered by the eye of an organism will be produced by its own movements, be it locomotion, head movements, or eye movements. Although movements of other agents (e.g. predators or prey) can be biologically very important, I will assume that early vision is best tuned to the most common of movements, namely those caused by the organism itself. Thus we need a description of the resulting distribution of velocities as perceived by the eye. This distribution, which I will call the velocity model, is difficult to determine exactly, because it not only depends on the organism's movements, but also on how it moves through a three-dimensional world full of objects. It can be shown, however, that if an organism is moving in a straight line through a world filled uniformly with objects, the velocity model behaves as $1/v^2$ for large velocities v (see the Appendix). As a simple function that follows this behaviour, and also behaves well for low speeds, I will use

$$a_v(v) = \frac{c}{(|v| + \sigma_v)^2}, \qquad (1)$$

as an approximation of the velocity model, with c a calibration constant. It gives the probability distribution a_v of velocities v of components in the image as perceived by the eye, with σ_v a parameter determining the width of the distribution. I assume here that this function is also a reasonable approximation of the distribution of the velocities due to rotations rather than translations of the visual system.

The spatial structure of natural images together with the velocity model determine a power density giving the average power expected for a given spatial frequency moving at a given velocity. For theoretical reasons and because psychophysical measurements are usually made in the space-time domain rather than in the space-velocity domain, we will transform the power density to the space-time domain by a change of variables (see Appendix). The result is shown in Fig. 1, which gives the power spectrum of the image stream (the image as a function of time) expected on average by the organism when it is moving through a natural environment. Only one of the two spatial dimensions (f_x and f_y) is shown in the figure. Note that most of the power is in low spatial and temporal frequencies, with progressively less power in both higher spatial and higher temporal frequencies.

A theory of early vision

The theory is developed for a single information channel in the visual system. The visual system is assumed to consist of any array of similar channels, each looking at a different position in visual space. Although it is the entire array that determines the ultimate limits of visual performance, we will see that the single channel performance as predicted by the theory is very similar to what is observed experimentally for the entire system.

Figure 2 shows a scheme of the theory elaborated below. Natural image streams contain power over a virtually unlimited range of spatial and temporal frequencies. A visual system limits this amount of information basically by low-pass filtering. This is partly because of physical limitations (eye size limits spatial resolving power because of diffraction, metabolic cost puts bounds on the temporal bandwidth of photoreceptors), and presumably partly because of limitations of brain size and complexity: the brain must be able to cope with the amount of information allowed access by the photoreceptors. Thus we assume that the image-stream is first low-pass filtered, in space and time, by a prefilter (Fig. 2). The resulting prefiltered image is degraded by noise: photon shot noise due to the absorption of light by the visual pigment, transducer noise produced by the phototransduction process, and any other noise source involved in the processes of prefiltering. The resulting noisy image is eventually transferred to a channel of limited dynamic range (i.e. it only supports a certain range of response amplitudes) which adds some noise of its own (channel noise) to the image.

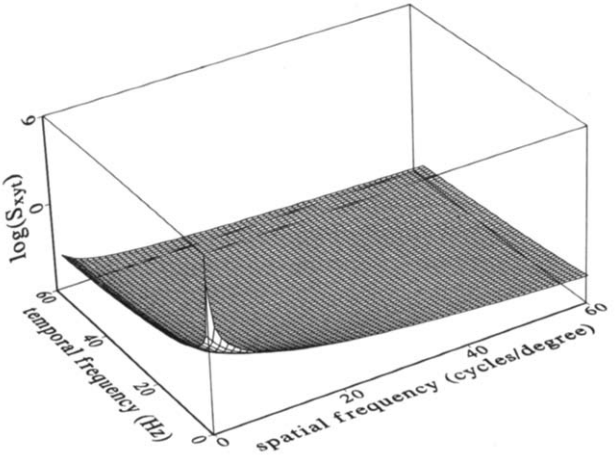

FIGURE 1. Spatiotemporal power spectrum of natural images observed by a visual system with a velocity model as described in the text.

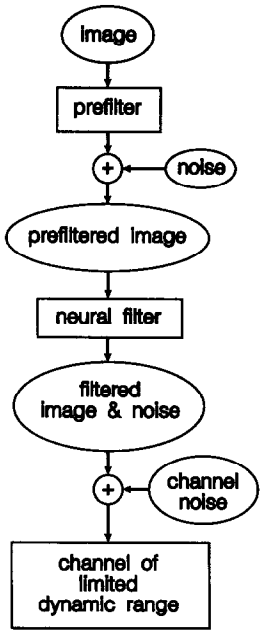

FIGURE 2. Scheme of the theory. An image is low-pass filtered by a prefilter, and noise is added to the result. A neural filter subsequently delivers a filtered image and filtered noise to a channel of limited dynamic range, which also adds noise to the result. The basic assumption of the theory is that the amount of information transferred by the channel is maximized by a suitable choice of the neural filter.

Before the prefiltered image is transferred to this channel, however, it is first transformed, in space and time, by a neural filter. The basic assumption of the theory is that this filter is tuned to maximize the total amount of information that is transferred by the channel. In the Appendix, I show how this neural filter can be computed.

Note that the fact that the channel is noisy and that it has a limited dynamic range is not only realistic, but also essential for the concept of maximizing information to work. No noise or an unlimited dynamic range would put no limits on the amount of information that could be transferred through the channel, and would leave the neural filter undetermined.

Before presenting results on the full spatiotemporal case, I will first discuss a one-dimensional example (in the time domain) to clarify the theory.

An example in the time domain

Figure 3(a) shows the square root of the temporal power spectrum of a natural image stream. It was calculated by integrating the power spectrum of Fig. 1 over the two spatial dimensions, and taking the square root. This spectrum is subsequently filtered by a temporal prefilter [Fig. 3(b)]. For this filter I chose a multistage low-pass filter (e.g. Watson, 1986; see also below), adjusted such that it corresponds to an impulse–response with a temporal full width at half-maximum of 40 msec. This value was recently measured in light adapted macaque cones (Schnapf, Nunn, Meister & Baylor, 1990). The resulting amplitude spectrum is shown in Fig. 3(c) (solid line).This also shows the noise amplitude spectrum (dashed line) added to the prefiltered image. The noise amplitude is chosen such that the average signal-to-noise ratio ($\overline{SNR}$, see the Appendix for a definition) equals 100. The power density spectra of neither the noise at the prefilter nor the channel noise are known in detail. As a simple first-order approximation we will assume that they are flat in the region of frequency-space where the analysis is performed. This assumption is not essential for the formulation of the theory, however, and it could be loosened in future elaborations.

Image and noise are subsequently filtered by a neural filter [Fig. 3(d)], determined such that it maximizes information flow (see below). Figure 3(e) shows the result of this filtering: filtered image (solid line), filtered noise (dashed line), and additive channel noise (dots). The amplitude of the channel noise was chosen such that it occupies one-tenth of the available dynamic range of the channel. This means that the root-mean-square (r.m.s.) value of channel noise is one-tenth of the r.m.s. value of the total amplitude fluctuations in the channel, due to both signal (image) and noise. The exact amount of channel noise is not very critical for the calculations below [for the spatiotemporal filter of Fig. 4(a), varying the channel noise by a factor of 10 changes the position of the peak by $<10\%$, and changes the sensitivity most at the lowest spatial and temporal frequencies, by less than a factor of 2]. The total r.m.s. value in the channel can be determined by adding, integrating, and taking the square root of the power spectra corresponding to the three amplitude spectra of Fig. 3(e) (see the Appendix for details). One limitation imposed upon the neural filter of Fig. 3(d) is that is must have a gain such that the dynamic range of the channel is fully utilized, but not exceeded.

From Fig. 3(e) we can determine the signal-to-noise ratio as a function of frequency (SNR; note the different use of SNR and $\overline{SNR}$: SNR is a function of frequency, while $\overline{SNR}$, an average of the SNR over frequency, is not). From the SNR the information density as a function of frequency follows directly [Fig. 3(g); see Goldman, 1953; and also the Appendix]. The total amount of information flowing per second through the channel is given by the integral of information density over all temporal frequencies. In fact, this information rate was maximized by a suitable choice of the neural filter: the neural filter of Fig. 3(d) is such that it produces the maximum possible rate of information [integral over the trace in Fig. 3(g)] while still keeping channel amplitude fluctuations [as determined from Fig. 3(e)] within the limits of the channel's dynamic range. We see in Fig. 3(g) that low temporal frequencies contain more information than higher ones, but this bias is much less than one would expect on the basis of Fig. 3(c). The reason is that, with $\overline{SNR} = 100$, the neural filter [Fig. 3(d)] is reducing low temporal frequencies, while favouring higher frequencies.

It can be shown (van Hateren, 1993) that the peak of the neural filter is always at a frequency where the SNR of the prefiltered image equals 1 [cf. Fig. 3(c, d)]. Lower frequencies are reduced because they are so strongly present in natural images that they threaten to

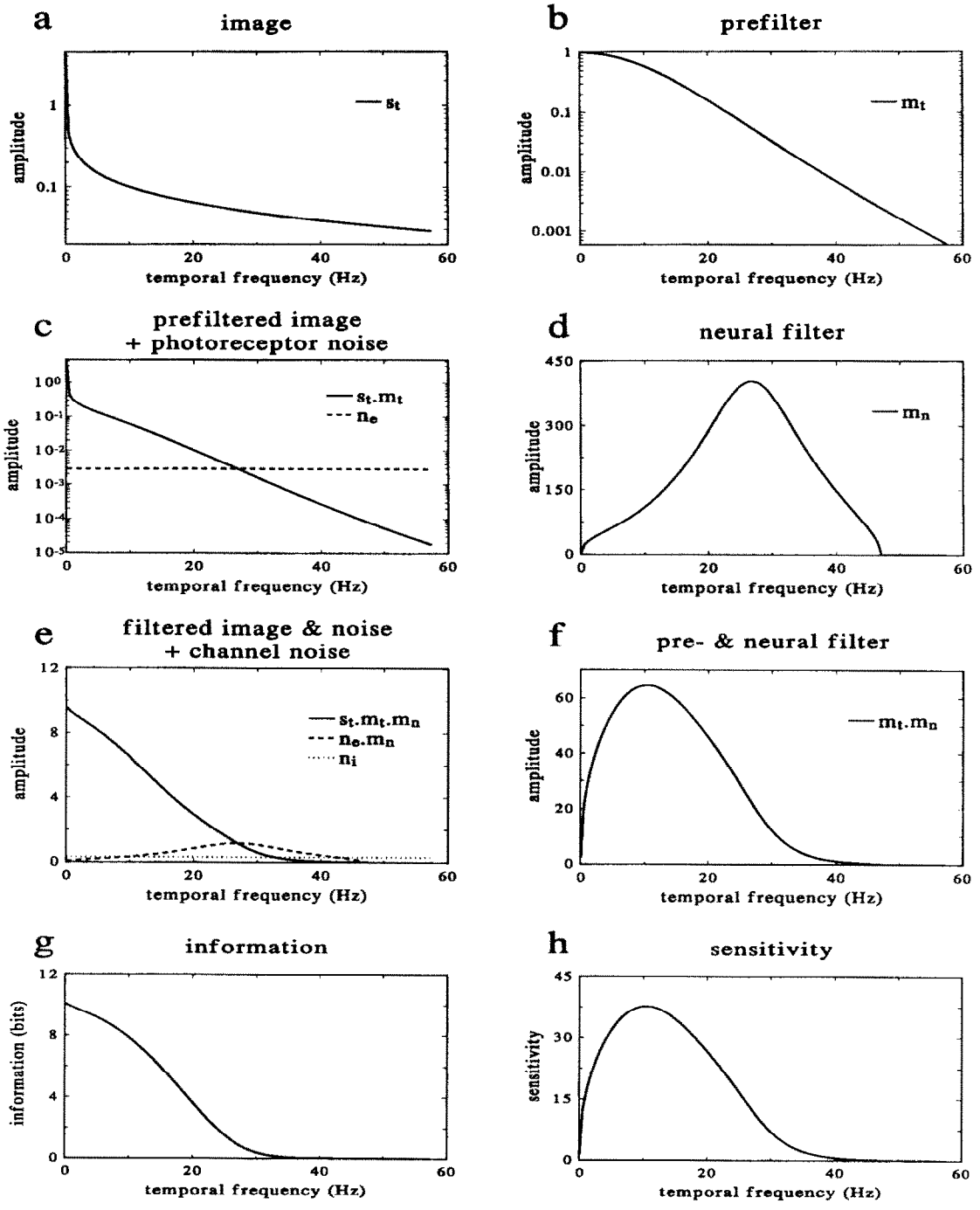

FIGURE 3. An example in the time domain. (a) Amplitude spectrum s_t of the temporal variations of a natural image stream. (b) Temporal prefilter, m_t. (c) Solid line, prefiltered image, $s_t m_t$; dashed line, noise n_e added to the prefiltered image. (d) Neural filter m_n maximizing the information rate in the channel. (e) Solid line, image after pre- and neural filtering, $s_t m_t m_n$; dashed line, noise after neural filtering, $n_e m_n$; dotted line, additive channel noise n_i. (f) Combination of prefilter and neural filter, $m_t m_n$. (g) Information derived from the SNR following from (e). (h) Sensitivity (see text for explanation).

occupy much of the dynamic range of the channel at the expense of other frequencies. In general, it is better to have many different frequencies of moderate SNR than to have some of very high and some of very low SNR {this is because information is proportional to the log $[1 + (\mathrm{SNR})^2]$, which increases most strongly when $\mathrm{SNR} = 1$}. The highest frequencies are also reduced by the neural filter, because their original SNRs are already so much smaller than 1 that their contribution to the information is negligible. Thus it is better to reduce these frequencies in order to prevent the associated noise to occupy too much of the channel's dynamic range. The position of the peak of the neural filter depends on the level of noise in Fig. 3(c). For higher $\overline{\mathrm{SNR}}$ it shifts to the right, and for lower $\overline{\mathrm{SNR}}$ to the left.

Figure 3(f, h) finally show filter characteristics that could be observed directly. Figure 3(f) shows the combination of prefilter and neural filter, which is the transfer function of the total system. Figure 3(h) gives the sensitivity of the system as a function of frequency. The sensitivity is defined here as the signal-to-noise ratio in the channel resulting from presenting a single temporal

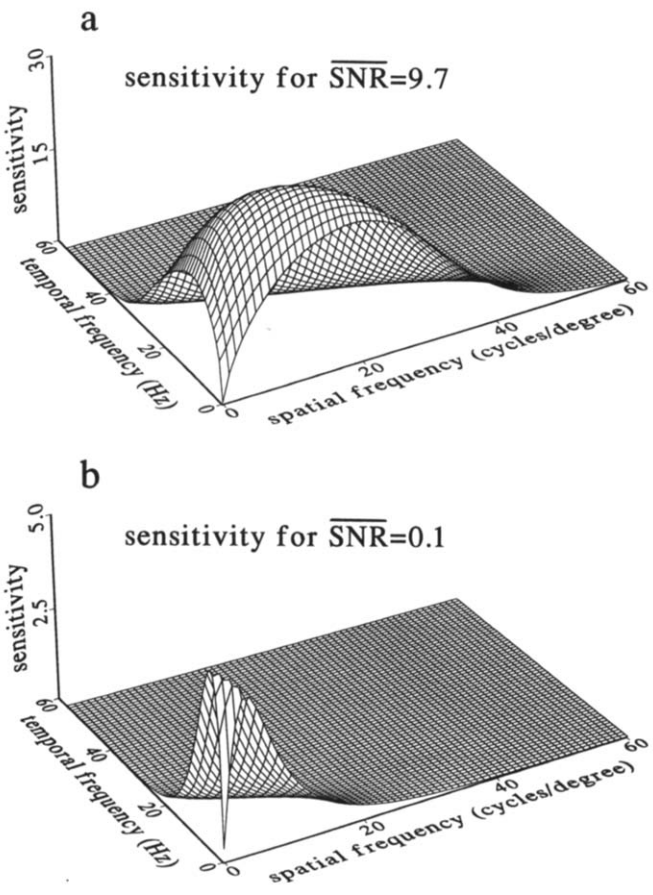

FIGURE 4. Spatiotemporal contrast sensitivities for two $\overline{SNR}$s.

frequency of 100% modulation (I will use $\widetilde{SNR}$ to denote this signal-to-noise ratio). This $\widetilde{SNR}$ is obtained by taking the square root of the ratio of signal power and noise power, i.e. the signal power in the channel resulting from the stimulus, and the total noise power in the channel (integrated over all temporal frequencies). Note that this is a rather conservative estimate of the signal-to-noise ratio: it assumes that a decision on the presence or absence of the stimulus would be based only upon the average power in the channel. At a sensitivity of 1 ($\widetilde{SNR} = 1$) there would be twice the amount of power in the channel than there would have been without the presence of the stimulus. Obviously, with more sophisticated analysis on the channel's signal (i.e. looking at signal and noise in narrow frequency bands, averaging over time, or pooling channels over space, etc.) the sensitivity would be higher (see e.g. Watson, 1992).

Spatiotemporal examples

The transfer function of the temporal prefilter, already mentioned in the previous section, is given by (Watson, 1986)

$$m_t(f_t) = \frac{1}{[(2\pi f_t \tau)^2 + 1]^{n/2}}, \qquad (2)$$

with $n = 9$ and $\tau = 5.65$ msec. This value of τ yields a temporal impulse–response with a full width at half-maximum of 40 msec (Schnapf *et al.*, 1990).

For the spatial prefilter I chose an approximation to a transfer function as determined by Campbell and Gubisch (1966) for the optics of the human eye. I found that the data points of their Fig. 9, pupil diameter 2 mm, are reasonably well described by the transfer function

$$m_s(f_s) = \sqrt{1 - f_s/f_{s,max}}\, e^{-f_s/(f_{s,max}\, a)}, \qquad (3)$$

where f_s is the spatial frequency, a a constant ($a = 0.28$), and $f_{s,max}$ the cutoff frequency of a lens of diameter $D = 2$ mm and a wavelength $\lambda = 570$ nm ($f_{s,max} = D/\lambda = 3.5 \times 10^3$ c/rad = 61.2 c/deg). I used equation (3) subsequently for both directions of spatial frequencies (f_x and f_y).

Finally, I treated the constant σ_v needed in the velocity model [equation (1)] as a free parameter in the theory. I adjusted σ_v such that spatiotemporal sensitivities resulted close to those measured psychophysically in the human visual system. The data I present below are for $\sigma_v = 0.63$ deg/sec. The results are not very sensitive to the exact value of σ_v: varying σ_v by a factor of 2 shifts the position of the sensitivity peak of Fig. 4(a) (see below) by about 25% (see also the Discussion).

Figure 4 shows spatiotemporal sensitivities for two different $\overline{SNR}$s. Although only one spatial frequency axis is shown, calculations were actually performed in two spatial dimensions (and one temporal dimension). At high $\overline{SNR}$ the sensitivity is band-pass, whereas at low $\overline{SNR}$ it is low-pass for most spatial and temporal frequencies. Also note in Fig. 4(a) that the filter is spatially band-pass for low temporal frequencies, spatially low-pass for high temporal frequencies, temporally band-pass for low spatial frequencies, and temporally low-pass for high spatial frequencies. This is precisely what has been found in numerous psychophysical investigations of human spatial, temporal, and spatiotemporal contrast sensitivities (e.g. de Lange, 1958; Kelly, 1961, 1977, 1979; van Nes, Koenderink, Nas & Bouman, 1967; Koenderink, Bouman, Bueno de Mesquita & Slappendel, 1978; Koenderink & van Doorn, 1979).

Spatiotemporal contrast sensitivity as a function of light intensity

In the previous sections results were presented for the various $\overline{SNR}$s (the average signal-to-noise ratio). Often, this $\overline{SNR}$ is not manipulated directly in psychophysical experiments, but it is varied indirectly by using different background intensities. Thus, in order to compare the theoretical results with psychophysical results we need a relation between the ambient light intensity and the resulting $\overline{SNR}$ in the photoreceptors. This appears not yet to have been measured, and therefore we will rely on an educated guess. If I is the light intensity (in arbitrary, dimensionless units), and $\overline{SNR}_{max}$ the maximum $\overline{SNR}$ that the photoreceptors can accomplish (due to inherent noise limitations), the resulting $\overline{SNR}$ is assumed to be given by

$$\overline{SNR} = \frac{\sqrt{I}\,\overline{SNR}_{max}}{\sqrt{I + \overline{SNR}_{max}}}. \qquad (4)$$

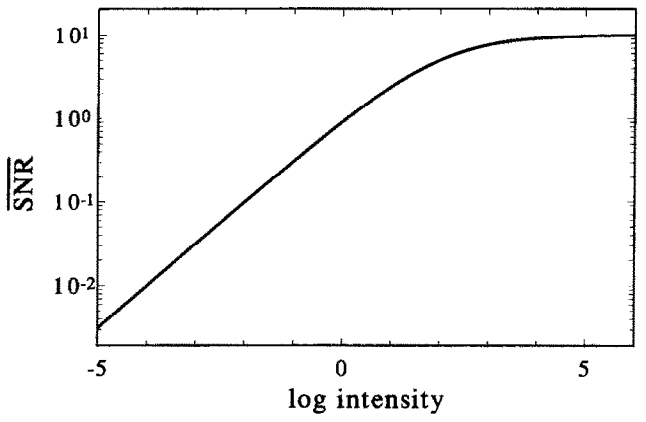

FIGURE 5. Resulting $\overline{SNR}$ as a function of intensity according to equation (4).

This function is depicted in Fig. 5. For small I equation (4) approximates $\sqrt{I}$, thus the $\overline{SNR}$ increases as the square root of the intensity, in accord with the quantum fluctuation theory (de Vries, 1943; for a review see Bouman, van de Grind & Zuidema, 1985). For high I, equation (4) goes asymptotically to $\overline{SNR}_{max}$, which has been set to 10 in Fig. 5. This value for $\overline{SNR}_{max}$ has also been used for all calculations below. Although it is only a guess, it seems likely that it is of the right order of magnitude. Recently, I found that the $\overline{SNR}$ in blowfly photoreceptors saturates for high light intensities at about 20 (van Hateren, 1992; see also Howard & Snyder, 1983; Howard, Blakeslee & Laughling, 1987). Because blowfly photoreceptors are electrically coupled and probably of lower input resistance than human cones, they are likely to have slightly higher $\overline{SNR}$s.

The intensity axis of Fig. 5 is in arbitrary, dimensionless units. However, assuming photon noise to be the dominating source of noise at the lowest intensities, assuming a 30% quantum efficiency of the eye (proportion of photons absorbed in cones to photons available at the cornea), and using a conversion formula due to Boynton (cited in Pokorny & Smith, 1986, pp. 8–14), I estimate that $I = 10$ in Fig. 5 (and subsequent figures) corresponds to a retinal illuminance of 5 td. This gives only an indication of the order of magnitude, however, because of the uncertain value of the quantum efficiency of the eye.

Using equation (4) we can now compute how the spatiotemporal sensitivity varies with background light intensity (i.e. adaptional level). Figure 6 shows sensitivities for combinations of four different light intensities, and two spatial and two temporal frequencies. Sensitivities are low-pass for low light intensities, and for high temporal and spatial frequencies, but band-pass for

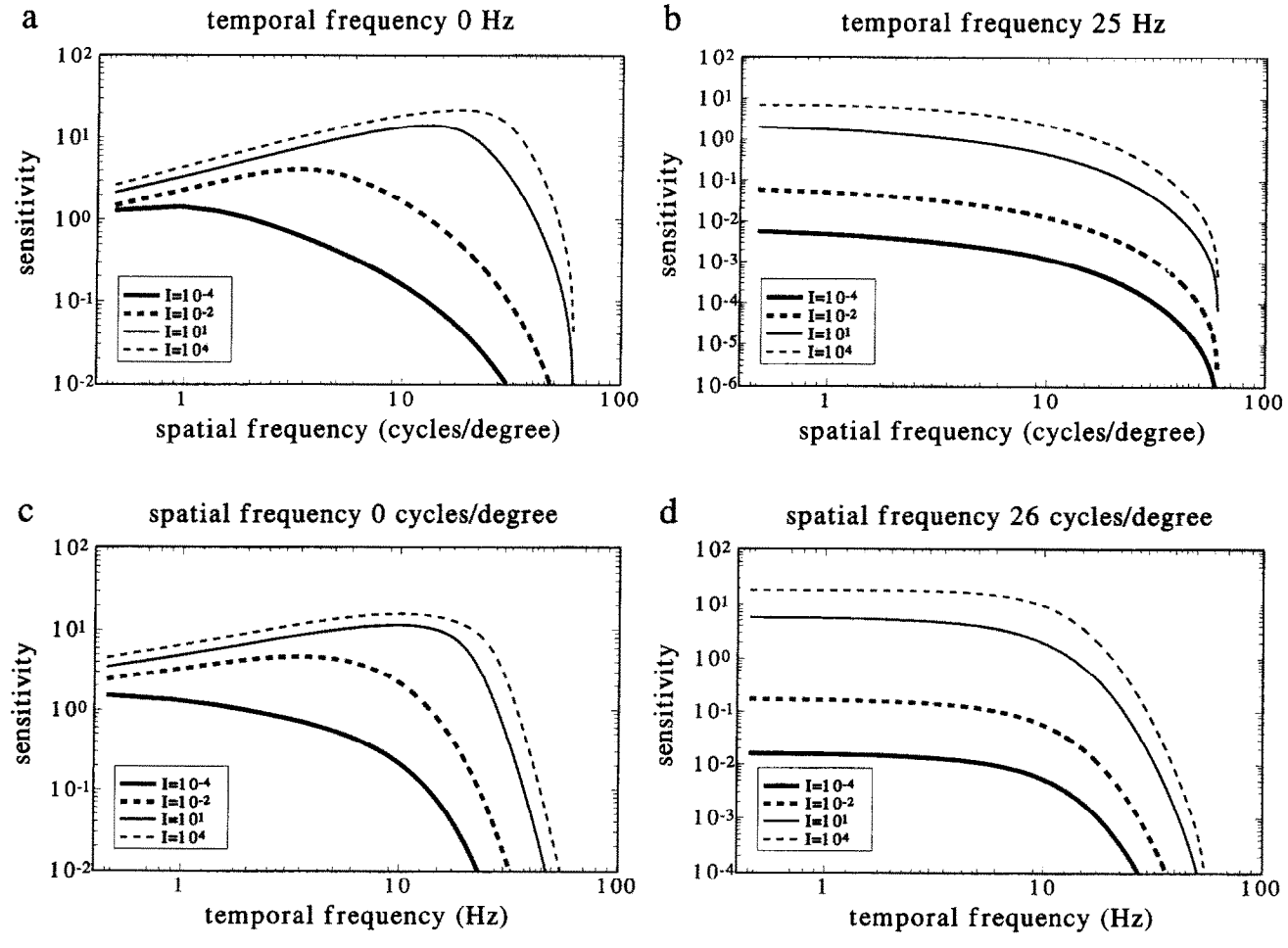

FIGURE 6. Contrast sensitivities at various spatial and temporal frequencies, and various intensities I. The intensity I is in arbitrary units, where $I = 10$ approximately corresponds to a retinal illuminance of 5 td.

higher light intensities at low spatial and temporal frequencies. Also note that sensitivities increase with increasing light intensity. The ordinate (sensitivity) shows the $\widetilde{\text{SNR}}$ for a single channel (as before, the $\widetilde{\text{SNR}}$ is the signal-to-noise ratio at one particular frequency). On the basis of a single channel it would be difficult to perceive a spatiotemporal stimulus if the $\widetilde{\text{SNR}}$ would be around or below 1. However, in psychophysical experiments there are many channels available that could be used by the visual system for discrimination of a particular stimulus. By pooling these channels the combined result would yield a better $\widetilde{\text{SNR}}$ than that shown in Fig. 6. As the amount of this pooling (possibly also extending over time) is not known exactly, we have no absolute calibration of the sensitivity axis of Fig. 6 against human contrast sensitivity. Nevertheless, most calculations compare favourably with psychophysical results if a $\widetilde{\text{SNR}} \approx 0.1$ is assumed as the threshold, which implies a pooling of about 100 statistically independent units. Note that a full correspondence between the performance of a single channel and psychophysical performance may be further complicated by processes such as probability summation. Recently, Bijl (1991) presented a model designed to predict psychophysical performance from the behaviour of retinal X-cells.

We can compare the shape, if not the absolute amplitude, of the curves with psychophysical results (e.g. Kelly, 1979; Koenderink *et al.*, 1978). The general conclusion is that, although there are certainly differences in details, the general behaviour of these curves as a function of spatial frequency, temporal frequency, and light intensity is very well predicted by the theory.

Kelly (1979, his Fig. 6) noted that the shape of the spatiotemporal sensitivity surface was very similar along lines of fixed velocity [i.e. skew lines, originating in the origin, in e.g. Fig. 4(a) of this article], apart from a shift along the spatial frequency axis. This result is also produced by the theory presented here, as depicted in Fig. 7.

Various psychophysical laws

The theory predicts several classical psychophysical laws. Figure 8(a) shows the Ferry–Porter law (the critical flicker frequency depends linearly on the logarithm of the background light intensity). As the critical flicker frequency (CFF) depends on the number of channels pooled when making psychophysical judgements, I show two cases, one with the CFF at $\widetilde{\text{SNR}} = 1$ (one channel utilized), and one with the CFF at $\widetilde{\text{SNR}} = 0.1$ (e.g. effectively 100 statistically independent channels utilized), the latter case being probably more realistic than the former. For $\widetilde{\text{SNR}} = 0.1$ there is a qualitative agreement with the results of Kelly (1961, his Fig. 8), although the theoretical CFF at the highest intensities is about 40% lower than actually measured (note, however, that Kelly's wide-field stimulus also stimulated the parafoveal retina, whereas the present calculations deal with foveal properties only; see also the Discussion).

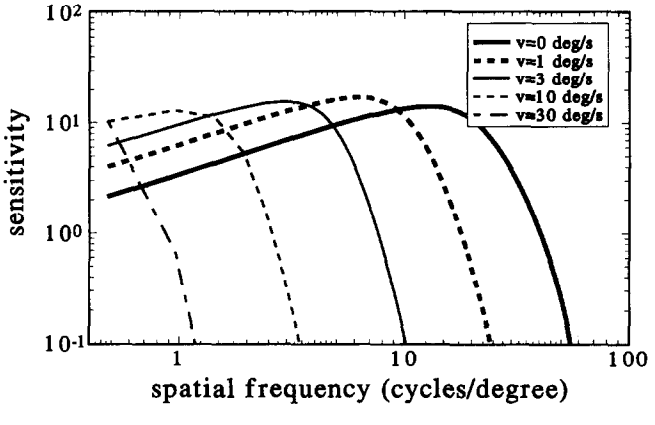

FIGURE 7. Contrast sensitivities at fixed velocities v (covarying spatial and temporal frequencies) for an intensity $I = 10$.

Figure 8(b) shows the de Vries–Rose law for low intensities (sensitivity proportional to the square root of the background light intensity) and Weber's law for high light intensities. Weber's law is here indicated by the constant sensitivity for higher intensities, with sensitivity defined as before, namely the signal-to-noise ratio in the channel resulting from a stimulus of 100% modulation. If the signal-to-noise ratio in the channel is constant, thresholds are a constant proportion of the signal amplitude, which is a formulation of Weber's law. The thin line on the left in Fig. 8(b) shows a slope of $\frac{1}{2}$ (corresponding to the de Vries–Rose law). The curves shown are for $f_s = 0$ c/deg, and three different temporal frequencies. Figure 8(b) shows that the transition to the Weber regime is occurring at lower light intensities for low temporal frequencies than for high temporal frequencies, as is also observed experimentally (e.g. Kelly, 1961). I obtained analogous results as in Fig. 8(b) for a temporal frequency $f_t = 0$ Hz, and various spatial frequencies.

Figure 8(c) presents results (for two different intensities and two stimulus sizes) showing Bloch's law (threshold contrast inversely proportional to stimulus duration for short durations, with threshold contrast being proportional to 1/sensitivity), depicted by the slope of -1 of the leftmost thin line. For longer durations, the curves have roughly a slope of $-\frac{1}{2}$ (compare with the rightmost thin line), i.e. the inverse square law. For a large stimulus size and a high light intensity, the curves get slopes closer to 0. The results in Fig. 8(c) are very similar to those obtained by Barlow (1958, Fig. 2). Sensitivity is here defined, as before, as the $\widetilde{\text{SNR}}$ obtained from taking the square root of the ratio of the total power due to the stimulus, and the total noise power.

Finally, Fig. 8(d) illustrates Ricco's law (threshold contrast inversely proportional to stimulus area for small areas; i.e. a slope of -1 in a plot with logarithmic axes) and Piper's law (threshold contrast proportional to the square root of stimulus area, i.e. a slope of $-\frac{1}{2}$). Again, these results are quite similar to Barlow's (1958, Fig. 3), apart from a spatial scaling factor due to the fact that Barlow measured parafoveally.

Note that the results of Fig. 8(c, d) follow directly from the shape of the spatiotemporal contrast sensitivity

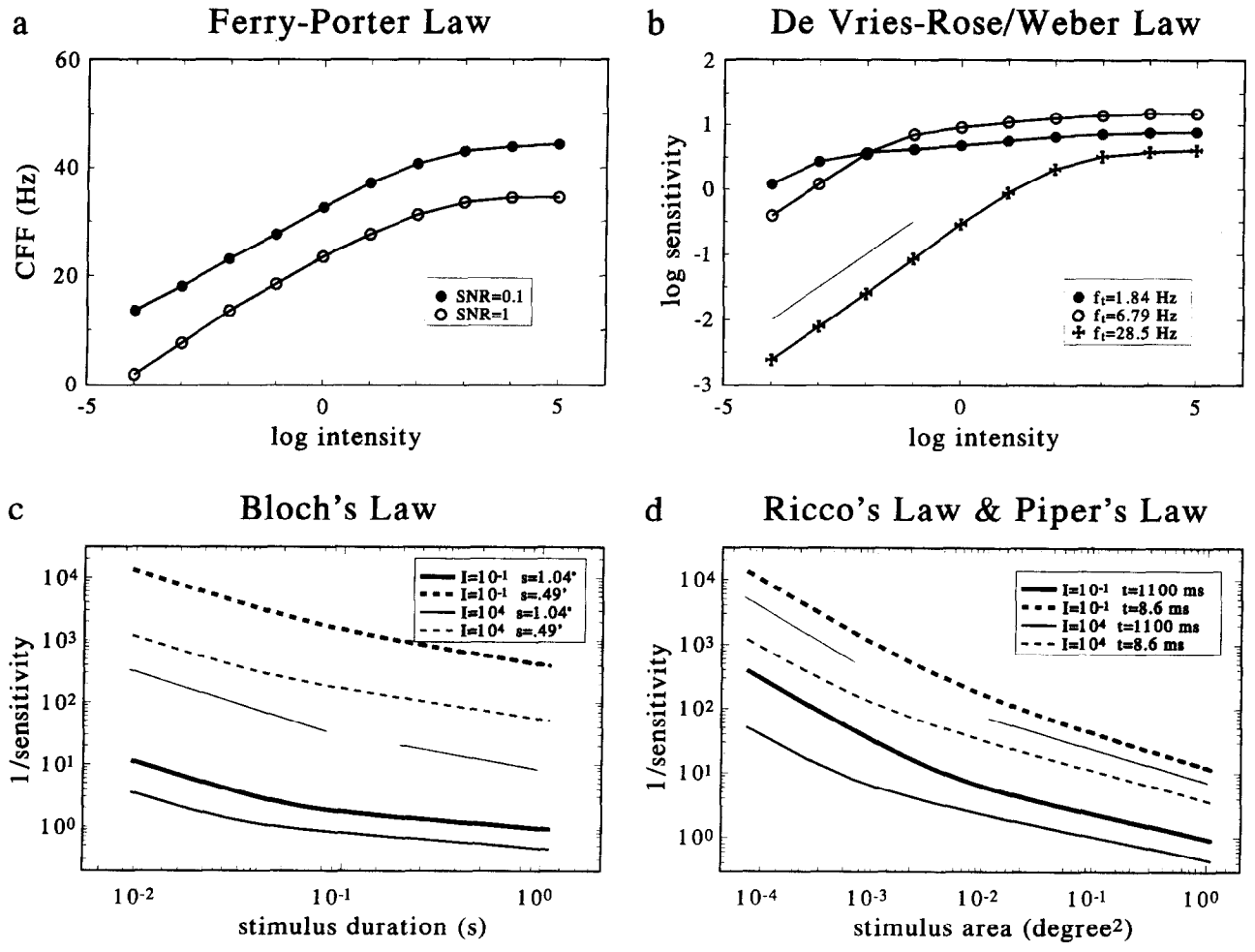

FIGURE 8. Several psychophysical laws as predicted by the theory. (a) Ferry–Porter law; CFF, critical flicker frequency; the figure shows the CFF for $\widetilde{SNR} = 1$ and $\widetilde{SNR} = 0.1$. (b) de Vries–Rose/Weber law. The spatial frequency is 0 c/deg. (c) Bloch's law for two intensities I and two stimulus sizes (the stimulus is a square of width s). (d) Ricco's law and Piper's law for two intensities I and two stimulus presentation times t.

[e.g. the surface of Fig. 4(a)], without further assumptions.

DISCUSSION

In this article I presented spatiotemporal contrast sensitivities of early vision, constructed on the premise that a main task of early vision is to maximize the information available to the brain, given noisy channels of limited dynamic range. Based on the spatiotemporal structure of natural image streams, and utilizing known properties of the human visual system for the prefilter, the theory produces results that correspond remarkably well with psychophysical measurements:

- spatiotemporal sensitivities are mostly low-pass in space and time for low intensities, and they are band-pass for all but the highest spatial and temporal frequencies for high intensities;
- sensitivities extend to much higher spatial and temporal frequencies for higher light intensities than for lower ones;
- spatiotemporal sensitivities are shaped similarly when plotted for constant velocity;
- the theory predicts several psychophysical laws (Ferry–Porter, de Vries–Rose, Weber, Bloch, Ricco, Piper), including many of the details displayed when varying spatial or temporal frequency, stimulus size or duration, and light intensity.

I want to emphasize that these results were not obtained by a complicated model with many free parameters, but instead by a basically very simple theory (Fig. 2), almost completely based on first principles. The only important free parameter is σ_v, the parameter determining the velocity model, though there is some experimental justification for the value of this parameter (see below). The main determinant of the theoretical results, however, is the spatial structure of the natural world, and, of course, the assumption that maximizing information flow is the basic strategy of early vision.

The role of v_c and σ_v

The spatial and temporal prefilters used can be characterized by two constants, $\sigma_s = 14.7$ c/deg and $\sigma_t = 9.7$ Hz, which correspond to the 1/e-values of the spatial and temporal transfer functions. I will call their ratio the

characteristic velocity v_c of the visual system, with $v_c = \sigma_t/\sigma_s = 0.66$ deg/sec. The characteristic velocity is a measure of the velocity required to cross a typical photoreceptor receptive field in a typical photoreceptor integration time (see Glantz, 1991, for similar considerations on invertebrate vision). It is illuminating to compare σ_v [the parameter determining the width of the velocity distribution, see equation (1)] to v_c. The results presented in this article are based on $\sigma_v/v_c \approx 1$. For this ratio the spatiotemporal filter changes, as a function of light intensity, similarly for spatial and temporal frequencies. However, if $\sigma_v/v_c \gg 1$, it is mainly the spatial properties that change as a function of light intensity, with virtually fixed temporal properties. If $\sigma_v/v_c \ll 1$, on the other hand, spatial properties are hardly changing (width remains the same, although some lateral inhibition develops), whereas temporal properties are changing very much. This case may be expected for human extrafoveal vision (σ_t about constant or even increasing somewhat, and σ_s declining appreciably, thus v_c increasing; v_c increases probably much more than σ_v, which will also increase because the viewing point will on average be more misaligned with the heading point than for foveal vision). A similar case with $\sigma_v/v_c = 0.1$ was recently investigated in detail in the fly visual system (van Hateren, 1992).

The value of σ_v was adjusted to yield good correspondence with psychophysical results. The value used, $\sigma_v = 0.63$ deg/sec, is of the same order of magnitude as that of the velocity distributions reported by Steinman and Collewijn (1980) on residual retinal image motion during head movements.

Limitations and extensions of the theory

The contrast sensitivities calculated seem to extend to slightly higher spatial frequencies and slightly lower temporal frequencies than measured contrast sensitivities, although it is difficult to be certain because there is quite some variation in psychophysical results among different authors. Both of these discrepancies could easily be resolved by assuming slightly different parameters of the prefilters. The band-width of the temporal prefilter is based on measurements of Schnapf *et al.* (1990) on isolated macaque cones, and it is quite possible that human cones in the intact retina show somewhat faster impulse–responses when completely light-adapted. It is also possible that other retinal elements are the real limiting factors and should be considered as determining the temporal prefilter. Similar arguments apply to the spatial prefilter, which is likely to be an overestimation of the spatial-frequency–response, as it only takes into account the lens optics. In particular, the cone aperture (waveguide effects), optical coupling of cones, intraretinal light scattering, electrical coupling, involvement of horizontal cells, etc. are all factors potentially influencing the spatial prefilter.

In fact, the components of the theory should not be taken too literally: they do not represent specific components in the nervous system. In particular, all adaptive properties of the theory schematized in Fig. 2 are projected into the neural filter rather than partly into the prefilter. This alone already makes the theoretical prefilter a much simpler device than a real photoreceptor: the adaptive properties of the photoreceptor are, for simplicity, incorporated into the (adaptive) frequency–response of the neural filter. Preferably, the scheme of Fig. 2 should be considered as an abstraction of early vision, not as a specific model of a particular visual system.

Although the theory predicts the amplitude of the neural transfer function, it does not predict its phase. Spatial phase is not a problem if we assume circularly symmetrical receptive fields, but temporal phase is a harder nut to crack. One possibility yielding impulse–responses minimally spread in time is a minimum phase filter combined with a pure time delay (see e.g. Roufs, 1972). This kind of filter is apparently realized in second order neurons of the blowfly visual system (van Hateren, 1992), but there are certainly other reasonable possibilities.

Finally, it would probably be a mistake to consider the spatiotemporal filters presented here as the sole agents of early vision. Obviously, they can not transfer all the available information in the photoreceptor image, due to their limited dynamic range. Thus there seems to be ample opportunity for more specialized neurons, even early in the visual system, to fill the information gap.

REFERENCES

Barlow, H. B. (1958). Temporal and spatial summation in human vision at different background intensities. *Journal of Physiology, 141*, 337–350.

Barlow, H. B. (1961). Possible principles underlying the transformations of sensory messages. In Rosenblith, W. A. (Ed.), *Sensory communication* (pp. 217–234). Cambridge, Mass.: MIT Press.

Bijl, P. (1991). Aspects of visual contrast perception. PhD thesis, University of Utrecht, The Netherlands.

Bouman, M. A., van de Grind, W. A. & Zuidema, P. (1985). Quantum fluctuations in vision. In Wolf, E. (Ed.), *Progress in optics* (Vol. XXII, pp. 77–144). Amsterdam: North-Holland.

Bracewell, R. N. (1978). *The Fourier transform and its applications*. New York: McGraw–Hill.

Burton, G. J. & Moorhead, I. R. (1987). Color and spatial structure in natural scenes. *Applied Optics, 26*, 157–170.

Campbell, F. W. & Gubisch, R. W. (1966). Optical quality of the human eye. *Journal of Physiology, 186*, 558–578.

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*, 2379–2394.

Glantz, R. M. (1991). Motion detection and adaptation in crayfish photoreceptors. *Journal of General Physiology, 97*, 777–797.

Goldman, S. (1953). *Information theory*. New York: Dover.

van Hateren, J. H. (1992). Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation. *Journal of Comparative Physiology A, 171*, 157–170.

van Hateren, J. H. (1993). A theory of maximizing sensory information. *Biological Cybernetics*. In press.

Howard, J. & Snyder, A. W. (1983). Transduction as a limitation on compound eye function and design. *Proceedings of the Royal Society of London B, 217*, 287–307.

Howard, J., Blakeslee, B. & Laughlin, S. B. (1987). The intracellular pupil mechanism and photoreceptor signal: Noise ratios in the fly *Lucilia cuprina*. *Proceedings of the Royal Society of London B, 231*, 415–435.

Huang, J. & Turcotte, D. L. (1990). Fractal image analysis: Application to the topography of Oregon and synthetic images. *Journal of the Optical Society of America A, 7,* 1124–1130.

Kelly, D. H. (1961). Visual responses to time-dependent stimuli. I. Amplitude sensitivity measurements. *Journal of the Optical Society of America, 51,* 422–429.

Kelly, D. H. (1977). Visual contrast sensitivity. *Optica Acta, 24,* 107–129.

Kelly, D. H. (1979). Motion and vision. II. Stabilized spatio-temporal threshold surface. *Journal of the Optical Society of America, 69,* 1340–1349.

Koenderink, J. J. & van Doorn, A. J. (1979). Spatiotemporal contrast detection threshold surface is bimodal. *Optics Letters, 4,* 32–34.

Koenderink, J. J., Bouman, M. A., Bueno de Mesquita, A. E. & Slappendel, S. (1978). Perimetry of contrast detection thresholds of moving spatial sine wave patterns I–IV. *Journal of the Optical Society of America, 68,* 845–865.

Kretzmer, E. R. (1952). Statistics of television signals. *Bell System Technical Journal, 31,* 751–763.

de Lange, H. (1958). Research into the dynamic nature of the human fovea–cortex systems with intermittent and modulated light. I. Attenuation characteristics with white and colored light. *Journal of the Optical Society of America, 48,* 777–784.

Laughlin, S. B. (1981). A simple coding procedure enhances a neuron's information capacity. *Zeitschrift für Naturforschung, 36c,* 910–912.

Laughlin, S. B. (1983). Matching coding to scenes to enhance efficiency. In Braddick, O. J. & Sleigh, A. C. (Eds), *Physical and biological processing of images* (pp. 42–52). Berlin: Springer.

Laughlin, S. B. (1987). Form and function in retinal processing. *Trends Neuroscience, 10,* 478–483.

van Nes, F. L., Koenderink, J. J., Nas, H. & Bouman, M. A. (1967). Spatiotemporal modulation transfer in the human eye. *Journal of the Optical Society of America, 57,* 1082–1088.

Pokorny, J. & Smith, V. C. (1986). Colorimetry and color discrimination. In Boff, K. R., Kaufman, L. & Thomas, J. P. (Eds), *Handbook of perception and human performance* (Chap. 8). New York: Wiley.

Robson, J. G. (1966). Spatial and temporal contrast-sensitivity functions of the visual system. *Journal of the Optical Society of America, 56,* 1141–1142.

Roufs, J. A. J. (1972). Dynamic properties of vision—II. Theoretical relationships between flicker and flash thresholds. *Vision Research, 12,* 279–292.

Schnapf, J. L., Nunn, B. J., Meister, M. & Baylor, D. A. (1990). Visual transduction in cones of the monkey *Macaca fascicularis*. *Journal of Physiology, 427,* 681–713.

Snyder, A. W., Laughlin, S. B. & Stavenga, D. G. (1977). Information capacity of eyes. *Vision Research, 17,* 1163–1175.

Srinivasan, M. V., Laughlin, S. B. & Dubs, A. (1982). Predictive coding: A fresh view of inhibition in the retina. *Proceedings of the Royal Society of London B, 216,* 427–459.

Steinman, R. M. & Collewijn, H. (1980). Binocular retinal image motion during active head rotation. *Vision Research, 20,* 415–429.

de Vries, H. L. (1943). The quantum character of light and its bearing upon threshold of vision, the differential sensitivity and visual acuity of the eye. *Physica, 10,* 553–564.

Watson, A. B. (1986). Temporal sensitivity. In Boff, K. R., Kaufman, L. & Thomas, J. P. (Eds), *Handbook of perception and human performance* (Chap. 6). New York: Wiley.

Watson, A. B. (1992). Transfer of contrast sensitivity in linear visual networks. *Visual Neuroscience, 8,* 65–76.

van der Ziel, A. (1970). *Noise. Sources, characterization, measurement.* Englewood Cliffs, N.J.: Prentice–Hall.

Acknowledgements—I wish to thank Nomdo Jansonius and Doekele Stavenga for useful comments. This research was supported by the Netherlands Organization for Scientific Research (NWO).

APPENDIX

In the following we will assume a discrete spatiotemporal system, with N_x samples along the spatial x-axis spanning a width w_x, N_y samples along the spatial y-axis spanning a width w_y, and N_t samples along the temporal t-axis spanning a time w_t. This formulation leads to discrete mathematics, which can readily be implemented in a computer program. N_x samples spanning a width w_x lead to N_x spatial frequencies f_x, spaced at $\Delta f_x = 1/w_x$:

$$f_x = \left(\frac{-N_x}{2}+1\right)\frac{1}{w_x}, \ldots, \frac{-1}{w_x}, 0, \frac{1}{w_x}, \frac{2}{w_x}, \ldots, \frac{N_x}{2}\frac{1}{w_x} \tag{5}$$

(see e.g. Bracewell, 1978). Similarly, there are N_y spatial frequencies f_y, spaced at $\Delta f_y = 1/w_y$, and N_t temporal frequencies f_t, spaced at $\Delta f_t = 1/w_t$.

The spatial power density of natural images is given by

$$S_{xy}(f_x, f_y) = \begin{cases} \dfrac{1}{(1+c_s^2)\,\Delta f_x\,\Delta f_y} & \text{if } (f_x, f_y) = (0,0) \\ \dfrac{c_s^2}{c_1(1+c_s^2)(f_x^2+f_y^2)} & \text{otherwise,} \end{cases} \tag{6}$$

with c_1 a calibration constant, given by

$$c_1 = \sum_{\substack{f_x, f_y \\ (f_x, f_y) \neq (0,0)}} \frac{\Delta f_x\,\Delta f_y}{f_x^2+f_y^2}, \tag{7}$$

and the spatial contrast c_s of the image defined as

$$c_s = \left[\frac{\sum_{(f_x, f_y)\neq(0,0)} S_{xy}(f_x, f_y)\,\Delta f_x\,\Delta f_y}{S_{xy}(0,0)\,\Delta f_x\,\Delta f_y}\right]^{1/2}. \tag{8}$$

If the density of objects in the environment (i.e. in three dimensions) is α, and if we assume that objects in the world are homogeneously distributed, it follows that the probability $P(x)\mathrm{d}x$ of observing an object between a distance x and $x + \mathrm{d}x$ equals

$$P(x)\mathrm{d}x = \alpha \exp(-\alpha x)\mathrm{d}x, \tag{9}$$

which is similar to Beer's law for absorption of light in an absorbing medium. Assuming that the animal moves with a velocity v_t (perpendicularly to the direction of the object), the resulting velocity of the object in the image is

$$v = \frac{v_t}{x}. \tag{10}$$

With a change of variables (x to v) we find from equations (9) and (10) the probability $a_v(v)\mathrm{d}v$ of observing an object with an effective speed in the image between v and $v + \mathrm{d}v$:

$$a_v(v)\mathrm{d}v = \alpha v_t \frac{\exp(-\alpha v_t/v)}{v^2}\,\mathrm{d}v. \tag{11}$$

Thus for large v, $a_v(v)$ behaves as v^{-2}. This applies to all possible translational velocities v_t of the animal. As a reasonable guess for the total velocity distribution we then choose

$$a_v(v) = \frac{c_v}{(|v| + \sigma_v)^2}, \tag{12}$$

with σ_v a positive constant regulating the width of the distribution, and c_v a constant such that

$$\sum_v a_v(v)\,\Delta v = \sum_{f_t} a_v\left(\frac{\pi f_t}{2 f_r}\right)\frac{\pi}{2 f_r}\Delta f_t = 1. \tag{13}$$

The second part follows from a change of variables $(f_x, f_y, v) \rightarrow (f_x, f_y, f_t)$, with f_t the temporal frequency, and from [assuming no correlation between the direction of the velocity v and the spatial frequency vector (f_x, f_y)]

$$f_t = \frac{2}{\pi} v f_r, \tag{14}$$

with

$$f_r = \sqrt{f_x^2+f_y^2}, \tag{15}$$

i.e. f_r is the amplitude of the spatial frequency vector.

With the change of variables $(f_x, f_y, v) \rightarrow (f_x, f_y, f_t)$, and by approximating the contribution of frequencies lower than $1/w_r$ by substituting $f_r = 1/2w_r$ for the case $f_r = 0$ and $f_t \neq 0$, we finally get for the spatiotemporal power density

$$S_{xyt}(f_x, f_y, f_t) = \begin{cases} \dfrac{1}{(1+c_s^2)\,\Delta f_x\,\Delta f_y\,\Delta f_t} & \text{if } f_r = 0 \text{ and } f_t = 0 \\ \dfrac{4\pi c_s^2 w_r^3}{c_1(1+c_s^2)}\, a_v(\pi f_t w_r) & \text{if } f_r = 0 \text{ and } f_t \neq 0 \\ \dfrac{\pi c_s^2}{2c_1(1+c_s^2) f_r^3}\, a_v\left(\dfrac{\pi f_t}{2 f_r}\right) & \text{otherwise.} \end{cases} \tag{16}$$

For the spatial prefiltering we use a spatial modulation transfer function $m_s(f_x, f_y)$:

$$m_s(f_x, f_y) = \sqrt{1 - f_r/f_{r,\max}}\, e^{-f_r/(f_{r,\max}\, a)}, \tag{17}$$

with f_r defined in equation (15), $f_{r,\max}$ the cut-off frequency of the optics (here with $f_{r,\max} = 61.2$ c/deg), and a a constant (here with $a = 0.28$). For the temporal transfer function, $m_t(f_t)$ we use

$$m_t(f_t) = \frac{1}{[(2\pi f_t \tau)^2 + 1]^{n/2}}, \tag{18}$$

with $n = 9$ and $\tau = 5.65$ msec.

The power density of the signal in the photoreceptor after this filtering is

$$S_{rec}(f_x, f_y, f_t) = S_{xyt}(f_x, f_y, f_t) |m_s(f_x, f_y)|^2 |m_t(f_t)|^2. \tag{19}$$

Next, we assume that to this signal a noise power density N_e is added, assumed to be constant in the volume of space and time where the analysis is performed. We then define the average signal-to-noise ratio ($\overline{\text{SNR}}$) at the level of the photoreceptors as

$$\overline{\text{SNR}} = \left(\frac{\sum_{f_x, f_y, f_t} S_{rec}(f_x, f_y, f_t)\, \Delta f_x\, \Delta f_y\, \Delta f_t}{\sum_{f_x, f_y, f_t} N_e\, \Delta f_x\, \Delta f_y\, \Delta f_t} \right)^{1/2}. \tag{20}$$

Both signal and noise are subsequently filtered in space and time by a neural filter with a power transfer function $p_n(f_x, f_y, f_t)$. Finally, the result is delivered to a channel with a limited dynamic range, K, and a limited information capacity due to internal noise (power density N_i). For the calculations I chose $K/\sqrt{N_i} = 10$ and $K = 10$. The value of K (the r.m.s. value of the response in the channel) is not important for the results, however, because it only influences the scaling of the transfer functions.

Now we have for the signal power density S and the noise power density N in the channel

$$S(f_x, f_y, f_t) = S_{rec}(f_x, f_y, f_t) p_n(f_x, f_y, f_t) \tag{21}$$

$$N(f_x, f_y, f_t) = N_e p_n(f_x, f_y, f_t) + N_i. \tag{22}$$

The requirement that the total of signal and noise remain within the dynamic range of the channel leads to

$$\sum_{f_x, f_y, f_t} (S + N)\, \Delta f_x\, \Delta f_y\, \Delta f_t = K^2, \tag{23}$$

where we have used the fact that the mean square value of a signal equals the integral over its power spectrum (see e.g. van der Ziel, 1970). Finally, we require that the information transfer rate R is as large as possible, with R defined as (e.g. Goldman, 1953):

$$R = \sum_{f_x, f_y, f_t} \log_e\left(1 + \frac{S}{N}\right) \Delta f_x\, \Delta f_y\, \Delta f_t, \tag{24}$$

where we have chosen $\log_e$ rather than $\log_2$ for mathematical convenience. This means that if we want to express the information transfer rate in bits per second per steradian, we have to multiply the R of equation (24) by $\log_2 e$. Thus R has to be maximized while keeping the constraint in equation (23). Following Goldman (1953, p. 159), this problem can be solved using the method of Lagrange multipliers, which leads to

$$p_n = \frac{-N_i(2N_e + S_{rec}) + \left(N_i^2 S_{rec}^2 - 4N_e S_{rec} N_i \frac{1}{\lambda}\right)^{1/2}}{2N_e(N_e + S_{rec})}, \tag{25}$$

with λ a Lagrange multiplier (for a derivation see van Hateren, 1992b). Now we can find p_n by choosing λ such that equation (23) is satisfied. This has to be done numerically by varying λ, thus varying p_n [equation (25)] and thereby S [equation (21)], N [equation (22)], and finally the summation of equation (23).

The power transfer function $|m_{stn}|^2$ of the total spatiotemporal filter F transforming the original image is now given by

$$|m_{stn}(f_x, f_y, f_t)|^2 = |m_s(f_x, f_y)|^2 |m_t(f_t)|^2 p_n(f_x, f_y, f_t). \tag{26}$$

Through equation (26) we know the power transfer function of F, and thus the amplitude of its transfer function, $|m_{stn}|$.