PERGAMON

Vision Research 41 (2001) 955–964

Vision Research

www.elsevier.com/locate/visres

Occlusions contribute to scaling in natural images

Rosario M. Balboa *, Christopher W. Tyler [1], Norberto M. Grzywacz [2]

Smith–Kettlewell Eye Research Institute, 2318 Fillmore Street, San Francisco, CA 94115-1813, USA

Received 3 May 2000; received in revised form 16 October 2000

Abstract

Spatial power spectra from natural images fall approximately as the square of spatial frequency, a property also called scale invariance (scaling). Various theories for visual receptive fields consider scale invariance key. Two hypotheses have been advanced in the literature for why natural images obey scale invariance. The first is that these images have luminance edges, whose spectra fall as frequency squared. The second is that scale invariance follows from natural images being essentially a collage of independent, constant-intensity regions, whose sizes follow a power-law distribution. Recently, an argument by example was made against the first hypothesis. Here we refute that argument and show that the first hypothesis is consistent with the scaling under a wide variety of distributions of sizes. There are two reasons for this: first, for every frequency, the log–log slope of the rotationally averaged power spectrum of an image is the weighted mean of the log–log slopes from the independent regions of the image formed by objects occluding one another. Second, the log–log slopes of the spectrum envelope for a constant-intensity region are 0 and -3 for frequencies corresponding to periods much larger and much smaller than the region's size, respectively. Therefore, it is not surprising that natural images have log–log slopes between -1.5 and -3, with a mean near -2.

Keywords: Natural images; Scaling; Power spectra

1. Introduction

Barlow (1961a,b, 1989) postulated that the function of receptive fields early in the visual system is to eliminate spatial redundancies in the image. Several theoreticians (Srinivasan, Laughlin, & Dubs, 1982; Atick & Redlich, 1990, 1992; Field, 1994; McCarthy & Owen, 1996) have proposed specific receptive-field models that implement Barlow's redundancy ideas. One of these models (Srinivasan et al., 1982) was based on the autocorrelation function, which is a statistical technique to quantify image redundancies. The other models used the spatial power spectrum of the image, which is the Fourier transform of the autocorrelation. It is thus important to measure the power spectra of the images that form the visual input. Several investigators have made careful measurements of such spectra, concluding that these spectra fall approximately with the square of spatial frequency (Carlson, 1978; Burton & Moorhead, 1987; Field, 1987; van Hateren, 1992; Tolhurst, Tadmor, & Chao, 1992; Field, 1993; Ruderman & Bialek, 1994; McCarthy & Owen, 1996; Balboa, 1997), a property also called scale invariance (scaling).

One hypothesis for why power spectra from natural images fall with frequency squared is that these spectra are strongly influenced by the spectral slope of edges in the images (Carlson, 1978). An alternative hypothesis postulates that natural images have objects that obey self-similar distributions of sizes (Mandelbrot, 1977; Field, 1987, 1994). For instance, Field (1994) showed that images formed by a sum of self-similar functions are scale-invariant in their contrast (their spectra fall with frequency squared). Ruderman (1997) generalized Field's arguments to include more realistic situations. Through some mathematical approximations, he argued that only images whose autocorrelation functions fall with a power-law produce power-law power spectra. He then proceeded to argue through some new measurements that only power-law distributions of sizes

* Corresponding authvor. Tel.: +1-415-3452134; fax: +1-415-3458455.

E-mail addresses: rbalboa@ski.org (R.M. Balboa), cwt@ski.org (C.W. Tyler), nmg@ski.org (N.M. Grzywacz).

[1] Tel.: +1-415-3452105.

[2] Tel.: +1-415-3452174.

0042-6989/01/$ - see front matter
PII: S0042-6989(00)00302-3

would give rise to power-law autocorrelation functions and power spectra. When taken to their logical conclusion, these arguments can be used to generate an example that would refute the luminance-edge hypothesis for scaling. Ruderman did so by building an image composed of luminance edges, whose autocorrelation function does not obey a power law. Unfortunately, the power spectrum of this image was not computed, because if it was done, then one would realize that it does fall with a power-law. Using his method to build images with a particular autocorrelation function (Ruderman, 1997), we generate the image on the left side of Fig. 1. This artificial image does not have a power-law autocorrelation function (right upper panel of Fig. 1), but this image has a power-law power spectrum with a log–log slope[3] of -2.25. (A similar result, but with a slope of -2.8 was derived by directly scanning Ruderman's image.) Therefore, non power-law autocorrelation functions and thus, non power-law distributions of sizes can produce power-law power spectra. Although Ruderman demonstrated through approximations that power-law autocorrelation functions yield power-law power spectra, Fig. 1 shows that the former is not a sufficient condition for the latter. Balboa (1997) provided a further counterexample by calculating without approximations the autocorrelation functions corresponding to spectra of the form $B/(C^2+f^2)^{p/2}$, where B, C, and p are positive constants and f is spatial frequency. Although the spectra fall as power laws (f^{-p}), the autocorrelation functions do not. For instance, for $p=3$, the autocorrelation function is $A_3\exp(-r/\lambda)$, where A_3 and λ are positive constants.

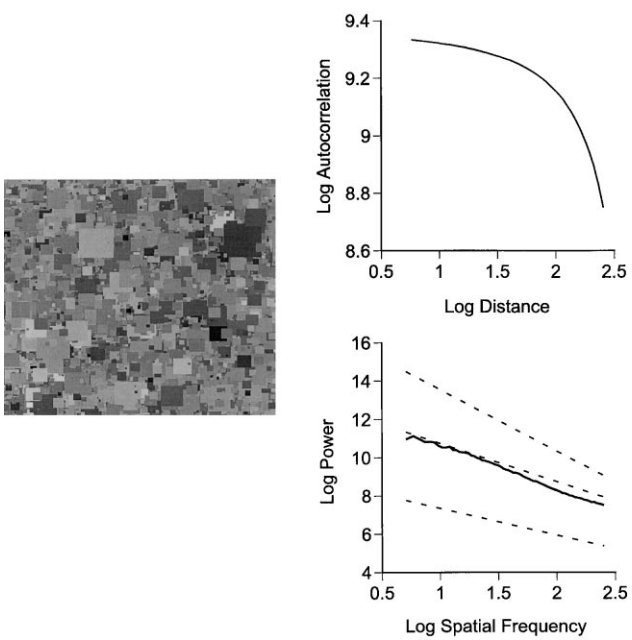

Fig. 1. Artificial image, and its autocorrelation function (upper-right panel) and orientation-averaged power spectrum (lower-right panel). The image was built by the method that was used by Ruderman to generate his Fig. 9. The image consists of square patches occluding each other. Their sizes and intensity levels followed exponential and Gaussian distributions, respectively. The image size was 512×512 pixels; the length constant of the exponential distribution was 10 pixels and the mean and S.D. of the Gaussian distribution were 128 and 35, respectively. Autocorrelation distances were in pixels and power-spectrum frequencies in cycles/image. The image autocorrelation function does not fall as a power law (Plotting this function in semi-log axes revealed that it had the shape of an exponential function over a wide range of frequences, as predicted by Ruderman (1997)) (a straight line in a log–log plot). And yet, the power spectrum does so, falling linearly in log–log axes with a slope of -2.25. The external dashed lines give the range of measured power-spectrum log–log slopes from natural images (Field, 1987). The middle dashed line has a slope of -2. Therefore, the image in this figure not only falls as a power law, but its log–log slope matches well the experimental value for natural images. Frequency units in the power spectrum are in cycles/image.

Because Ruderman's example yields a power-law power spectrum, this example does not rule out the luminance-edge hypothesis for scale invariance. This paper now shows that luminance edges impart much structure to an image's spectrum. This structure is such that the spectrum will be nearly scale invariant for a large variety of distribution of sizes. Such relative independence on sizes has strong implications for theories based on redundancy removal in the visual processing. The majority of the theories described above use the -2 slope in the log–log power spectrum as an indicator that the distribution of sizes in natural images are scale invariant. Ruderman's work (Ruderman, 1997) attempts to provide even more evidence in favor of a scale-invariant distribution. However, his evidence is indirect and no one has measured directly the natural distribution of sizes. By refuting Ruderman's example, we show that the distribution of sizes in natural images with a self-similar power spectrum is not constrained to be scale invariant. Consequently, the power spectrum becomes a weak statistical indicator of the structure in the visual world.

Some of this work has appeared before in abstract form (Balboa, De Juan, & Grzywacz, 1997).

2. Intuition

This section develops the intuition behind our ideas on natural-image power spectra, leaving the enunciation of more precise mathematical results to the next section.

A good starting point to understand the contribution of luminance edges (occlusions) to spectra is the behavior in one dimension. It is well known that one-dimensional (1-D) edges yield power spectra whose envelopes fall as frequency squared. Does the same result apply to two-di-

[3] The spatial frequencies over which we calculated slopes ranged from about 5 to 250 cycles/image. This range matched or was slightly wider than the range used by Field (1987).

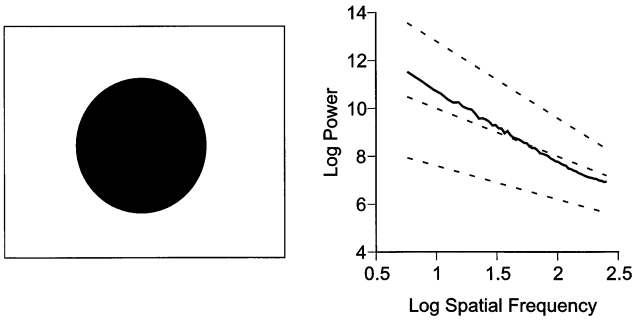

Fig. 2. An image consisting of a black disk on a white background has a power-spectrum slope inside the range of natural images (dashed lines as in Fig. 1). The power spectrum of the disk is nearly linear in log–log axes, with a slope of -2.8, which is close to the theoretic value of -3 for individual constant-intensity regions (Section 4). Frequency units in the power spectrum are in cycles/image.

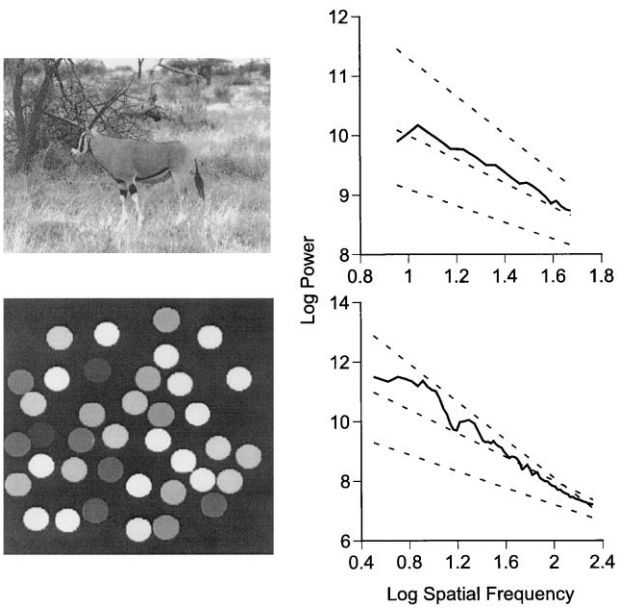

Fig. 3. Two images and their orientation-averaged power spectra (solid lines; spatial frequency in cycles/image). The left-upper panel is a natural image of an oryx, while the left-lower panel is an artificial image with randomly-placed disks of identical radii and arbitrary gray levels (from a homogeneous distribution). These images show slopes near -2 and within the range of natural images (dashed lines as in Fig. 1). For the artificial image, this is so despite having a single object size. In the text, we argue that the slope for the natural image (-1.9) is shallower than that for the artificial image (-2.8), since the former has a wider range of object sizes than the latter. That this artificial image has non-occluding objects does not mean that the arguments in this paper work only for such images. On the contrary, the proofs in Appendix A consider occlusions explicitly.

mensional (2-D) objects with sharp edges? Fig. 2 (left panel) shows an image consisting of a black disk on a white background. The rotationally averaged power spectrum of this disk plotted on a log–log axis is shown in Fig. 2 (right panel). This plot is nearly linear, with a slope of -2.8, which means that the spectrum is falling approximately as $f^{-2.8}$. Consequently, different from 1-D objects, individual 2-D objects with sharp edges do not have spectra that fall as frequency squared. The log–log slopes of these spectra envelopes at high frequencies are actually close to -3 regardless of the shape of the object. Despite being more negative than the mean of -2.2, these slopes are inside the range of slopes measured for natural images (Carlson, 1978; Burton & Moorhead, 1987; Field, 1987; van Hateren, 1992; Tolhurst et al., 1992; Field, 1993; Ruderman & Bialek, 1994; McCarthy & Owen, 1996; Balboa, 1997) (dashed lines).

Next, we asked whether the spectra of multi-object images may be thought of as some (hopefully simple) combination of the spectra of each the objects in the image. In Fig. 3, we provide two extreme examples of multi-object images and their spectra. In the upper panels, the natural image of an oryx has a large number of objects of different sizes, yielding a log–log power-spectrum slope of -1.9. The lower panels show an artificial image of 39 randomly placed disks of equal size and the power spectrum of this image. The log–log slope of the spectrum is -2.8, which is not surprising, since the image is composed of disks, the same objects as in Fig. 2. Nevertheless, this image raises two important issues, especially if compared to the oryx image: First, power spectra from individual sharp-edge objects have a structure such that even if there is only one object size, the overall spectrum of the image will be within the experimental range for natural images (dashed lines). Second, the existence of multiple object sizes, such as in the oryx image, may indeed bring the spectrum closer to full scale invariance.

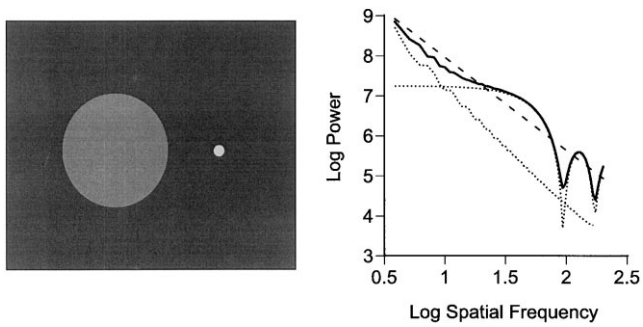

Fig. 4. Left: An image with two disks of different sizes and different intensities. Right: Orientation-averaged power spectrum of the image (solid line; spatial frequency in cycles/image), its linear fit (dashed line), and the individual power spectra of each disk (dotted lines). The power spectrum of the small disk is flat until reaching the spatial frequency whose corresponding period matches the size of the disk. After that point, the spectrum starts to fall with oscillations at high frequencies. The power spectrum of the large disk falls linearly in log–log axes with a slope of -2.8. The overall spectrum in nearly a weighted mean of the individual spectra, giving its linear fit an intermediate slope (-2.3). Again, this slope matches well those observed for natural images (Carlson, 1978; Burton & Moorhead, 1987; Field, 1987; van Hateren, 1992; Tolhurst et al., 1992; Field, 1993; Ruderman & Bialek, 1994; McCarthy & Owen, 1996; Balboa, 1997).

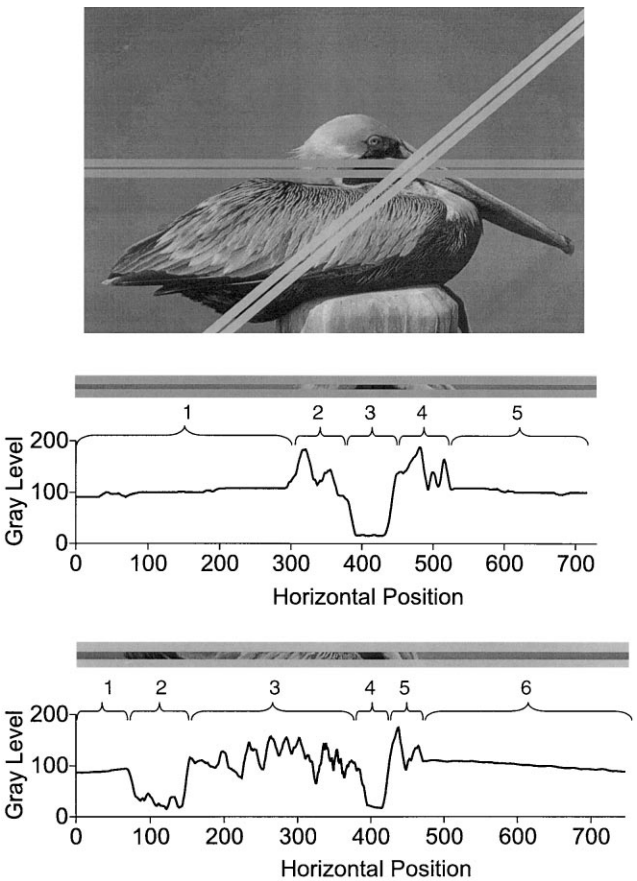

Fig. 5. Top: A natural image with two one-dimensional cuts superimposed (shown thick for clarity). Bottom: Gray levels for each cut (the horizontal one on the top), their graphs (positions in pixels), and numbered braces indicating a stepwise approximation for the graphs. For the horizontal cut, the Braces 1, 2, 3, 4, and 5 correspond to sky, head, face, beak, and sky, respectively. For the diagonal cut, the Braces 1, 2, 3, 4, 5, and 6 correspond to sky, ventral plumage, lateral plumage, face, beak, and sky, respectively. A stepwise approximation is possible regardless of the orientation of the cuts.

2.1. Spectral slopes for images of discrete objects

In Fig. 4, we provide an intuition for how single-object spectra and the distribution of sizes interact to give rise to the power spectrum. The image on the left panel has two disks with different radii and intensities. The two dotted lines are the individual power spectra for the two disks. For the large disk, the dotted line is practically straight and represents a power law with a slope of -2.8, as in Fig. 2. In contrast, the power spectrum for the small disk has a plateau at low spatial frequencies and then falls with oscillations. The envelope of the fall has, as expected, a slope of approximately -3. In turn, the plateau goes from zero frequency up to a frequency whose corresponding spatial period is roughly of the size of the disk. Such plateaus exist for images where the objects do not fill the image frame, regardless of their shape. The solid line is the power spectrum of the full image and the dashed line is the best linear fit to this spectrum. This fit has a slope of -2.3. This example shows that particular cases of a few discrete objects may approximate a slope of -2 even without an extended distribution of object sizes.

It is not difficult to understand how having two sizes made the slope closer to -2, even though each disk tends to produce a slope of -3 at high frequencies. At each frequency, the log–log slope of the image spectrum is approximately the mean of the slopes from both disks, with the weights being proportional to their spectral energies at that frequency. Hence, at low frequencies, the spectrum of the image is dominated by the power spectrum of the large disk. However, the slope is shallower than -3, since the slope for the small disk is 0 and this slope is contributing to the final slope. At intermediate and high frequencies, the spectrum of the image is dominated by the power spectrum of the small disk. Because the disk is so small, the envelope of its spectrum only reaches a slope of -3 at very high frequencies. In Fig. 4, the slope is -2.3 partly because, in this example, the contrast of the small disk is higher than that of the large disk. We are not claiming that in general the contrasts of natural objects increase as their sizes decrease. Indeed, we have successfully modeled the distribution of natural contrasts assuming that the brightnesses of objects are independent of their sizes (Balboa & Grzywacz, 2000b). If in Fig. 4, the contrast of the small disk were closer to that of the large disk, then the slope would not be -2.3, although it still would be shallower than -3.

Thus, mixing objects with different sizes and intensities typically produces spectra with mean slopes shallower than -3 and thus closer to -2. This conclusion is essentially independent of the distribution of sizes. The overarching idea is that at each frequency, the slope is a mixture of -3 slopes from relatively large objects, 0 slopes from small objects, and intermediate slopes from medium-sized objects. It follows that the final slope should seldom be steeper than -3 (Section 3 will explain why this may happen occasionally).

The only conditions necessary for this conclusion are that images are a collage of independent, constant-intensity patches. Ruderman (1997) provided strong evidence that natural images meet this condition. In other words, to a first approximation, the natural world produces stepwise images. The reason for such images is that nature comprises discrete objects of relatively constant physical properties and thus, of relatively constant albedo. Even regions that are not objects per se, such as the sky, have relatively constant physical properties over space, such as light scattering. Fig. 5 illustrates the general concept of natural images with stepwise regions. From the pelican image[4] at the top,

[4] For clarity, Fig. 5 has only 32 intensity levels and thus, only changes above 3% are meaningful. However, in the discussion that follows, the intensity steps that we are talking about are far deeper than 3%.

two 1-D sections (surrounded by gray strips in the figure for clarity) are extracted. The graphs below show that, independent of the orientation of the section, the image can be approximated by a few step functions. We label five such steps with numbered braces in the middle graph, which corresponds to the horizontal section. Brace 1 corresponds to the sky on the left, and Braces 2 and 3 to the pelican's head. The sky leads to an intermediate, relatively constant intensity level (≈ 90), whereas the pelican's head changes from higher intensities to very low intensities (Brace 3), approximated by three steps. In continuation, Braces 4 and 5 corresponds to the pelican's peak and the sky, respectively. Not surprisingly, the 'sky' steps (Braces 1 and 5) have about the same intensities, while the 'pelican' step (Braces 2, 3, and 4) has the highest and lowest intensities.

3. Results

In Appendix A, we formalize mathematically the intuitions highlighted in the preceding section[5]. Here, we just state these mathematical results and give pointers to the sections in Appendix A where these results appear.

In the first section in Appendix A (Section 5.1), we analyze the power spectrum of 1-D images consisting of regions of uniform intensity. For frequencies so low that the corresponding periods are longer than the longest region in the image, the power spectrum fluctuates, but its envelope remains roughly constant with frequency. As the spatial frequency increases, the power spectrum continues to fluctuate, but its envelope tends to fall in inverse proportion to the square of the frequency. Hence, spectra of 1-D collage of independent, constant-intensity objects occluding one another show the same type of scaling as natural images. This scaling occurs without making assumptions on the sizes of objects.

This analysis is extended to the 2-D case in Section 5.2. There, we consider 2-D images, whose cuts have similar profiles as the ones used in the 1-D analysis. As for the 1-D case, the 2-D power spectrum tends to be constant at low frequencies until the frequency becomes sufficiently high. Again, the corresponding spatial period of the critical frequency is approximately the size of the largest constant-intensity region in the image. After this frequency, the spectrum tends to fall with frequency. The exact mathematical form of this fall is analyzed in the two following sections. The first section (Section 5.3) shows that the spectral envelope for an individual constant-intensity region tends to be flat at low frequencies (spatial periods larger than the size of the region). This envelope then tends to fall as frequency cubed at high frequencies. The second section (Section 5.4) analyzes the log–log frequency dependence of the power spectra for images composed of several independent, constant-intensity regions. The result is that the slopes across an intermediate range of frequencies are to a good approximation a weighted mean of the slopes from the individual regions comprising the image. The slopes weights are proportional to the spectral energy of these regions at that frequency. Therefore, spectra of 2-D collage of independent objects occluding one another show log–log slopes between 0 and -3 (the most shallow and steep slopes from individual regions). The slope should tend to -3 at very high frequencies, that is, frequencies whose corresponding period is smaller than the smallest region in the image. However, in natural images there is often no evidence for a scale of the 'smallest object.' Regions can be any size, since they can correspond to objects at any distance from the camera or from the eye.

Consequently, it should not be surprising that the log–log slopes of 2-D natural-image power spectra range from -1.4 to -3.2 (Field, 1987). Little assumption has to be made about the sizes of objects to obtain such slopes.

Finally, Appendix A considers the effects on these conclusions of violating the assumption that images are stepwise. For this purpose, Section 5.5 considers violations of the constant-intensity assumption caused by biased illuminants producing non-homogeneous illumination. Because the 1-D and 2-D analyses yield similar results, we restrict the gradient analysis there to the 1-D case. This analysis shows that such gradients cause the power spectrum to fall faster than without the gradients. Hence, such gradients could contribute to the variability in slopes observed across natural images.

4. Discussion

Our two main results are as follows: First, we found flaws in Ruderman's (Ruderman, 1997) arguments against the luminance-edge hypothesis for the squared fall of the log–log power spectrum slope of natural images (Fig. 1). Contrary to these arguments, non power-law autocorrelation and thus, non power-law distributions of sizes can produce power-law power spectra. Second, we showed that images comprising independent, constant-intensity regions forming sharp luminance edges typically give rise to power spectra that fall with log–log slopes between -1.5 and -3 (Fig. 4 and Section 3). This result is consistent with measurements from natural images (Carlson, 1978;

[5] Quantal noise (Rushton, 1961; Fuortes & Yeandle, 1964; Baylor, Lamb, & Yau, 1979) will not be considered here, since its flat power spectrum can be shown only to add a constant to the power spectrum.

Burton & Moorhead, 1987; Field, 1987; van Hateren, 1992; Tolhurst et al., 1992; Field, 1993; Ruderman & Bialek, 1994; McCarthy & Owen, 1996; Balboa, 1997) and allows approximating them as stepwise functions (Fig. 5). This approximation is somewhat similar to that by Ruderman (1997), who views natural images as a collage of regions corresponding to independent objects. However, our explanation for the approximately square fall of the power spectrum diverges from his. Ruderman proposes that such a fall is caused by objects having a power-law distribution of sizes. In Fig. 1, we show that assuming such a distribution is not necessary. A wide variety of size distributions gives rise to log–log slopes within the experimental range. Therefore, the distribution of object sizes in natural images ought to be analyzed directly to determine what distributions are typical in nature.

4.1. Why do power spectra not always fall as the square of spatial frequency?

The fall of the power spectrum has a log–log slope of around -2, but the exact value depends on the composition of each individual image (Carlson, 1978; Burton & Moorhead, 1987; Field, 1987; van Hateren, 1992; Tolhurst et al., 1992; Field, 1993; Ruderman & Bialek, 1994; McCarthy & Owen, 1996; Balboa, 1997). Section 3 provides two reasons for the slope not being exactly -2: First, the exact slope depends on the shapes, border jaggedness (Field & Brady, 1997), sizes, and intensities of the independent regions in the image (see also Fig. 4). Second, intensity gradients caused by focal illumination or shadows can steepen the slopes. Here, we comment on other reasons for variations of slope besides gradients and geometric properties of regions. One can find (Carlson, 1978; Burton & Moorhead, 1987; Field, 1987; van Hateren, 1992; Tolhurst et al., 1992; Field, 1993; Ruderman & Bialek, 1994; McCarthy & Owen, 1996; Balboa, 1997) images with slopes more negative than -2 in any habitat, but this is particularly true underwater (Balboa, 1997). This suggests that typical slopes from different habitats might be slightly different. Why are underwater slopes steeper? One reason is blur, which causes loss of high spatial frequencies and thus steeper slopes (Field & Brady, 1997; Tolhurst & Tadmor, 1997). A cause of underwater blur is image movement registered by insufficiently fast photography. Another related source of blur is low illumination due to strong light absorption in water (Mobley, 1995), forcing longer photographic exposures and thus making images more prone to movement blurring. Finally, light scattering, which is particularly strong in water (Mobley, 1995), would cause further blur and make the slopes steeper. But not all slopes are steeper than -2; there are also images with slopes less negative than that. A possible reason for such shallow slopes is that the power spectrum is flat at low frequencies up to a frequency whose corresponding spatial period is approximately the size of the largest object in the image (Section 3). Therefore, if the slope is extracted with some portion of this flat zone included, one can have slopes less negative than -2.

Besides not having power spectra that fall as the square of spatial frequency, natural images often produce spectra with significant curvature. Tolhurst et al. (1992) made this point explicitly, but one can observe this curvature in many other data sets in the literature. In our model, the slope is allowed to change with frequency (Fig. 3 and Section 3), thus being consistent with this curvature. The changes of slope occur because the spectra of individual objects change from 0 to -3 as frequency rises.

4.2. Possible implications for the visual system

From the preceding section, we can argue that the retinal circuits of underwater animals might be different from those of land and flying animals to compensate for the loss of high spatial frequencies and for the low amounts of light (Mobley, 1995) in underwater natural images. Models in the literature would predict such a difference (Balboa & Grzywacz, 2000a). For related reasons, other natural habitats might lead to different retinal circuits. Interesting examples include foggy habitats in which pigeons and sea-birds must sometimes fly for hours or days (Sinclair, 1985), turbid lakes and rivers, and Arctic and Antarctic environments. And there are even different underwater habitats depending on factors such as the organic particles in the water (Mobley, 1995).

Does the design of retinal circuitry take into account that power spectra fall roughly as the square of spatial frequency? If this happened, we would argue that it is because the retina 'assumes' that objects in the world have uniform luminance and occlude each other. We would also argue that the retina may assume $1/f^2$ spectra almost independently of the objects sizes in the image. However, it is unclear whether the assumption of $1/f^2$ spectra plays a major role in retinal design. We recently analyzed theories that use $1/f^2$ power spectra to describe early retinal receptive fields. It was found that none of these Information-Theory-like theories could account for the biological behavior of horizontal cells with light intensity (Balboa & Grzywacz, 2000c). To explain this behavior, we proposed an alternative hypothesis based on the statistics of contrasts that belong to occluding objects (Balboa & Grzywacz, 2000a). Hence, although the power spectrum may not be the directly relevant information from natural images at the very early stages of visual computation, one of the underpinnings of the power spectrum, the occlusions, may be so. And the power spectrum could still be

a relevant statistic for later stages of visual processing (Atick & Redlich, 1990; Dan, Atick, & Reid, 1997).

Acknowledgements

We thank Dr David Field for critical comments, Daniel Grzywacz for identifying the animal in Fig. 3, and the two anonymous referees of the journal for thoughtful suggestions. This work was supported by National Eye Institute Grants EY08921 and EY11170, and the William A. Kettlewell Chair to N.M. Grzywacz and by a National Eye Institute Core Grant to Smith-Kettlewell.

Appendix A

A.1. One-dimensional analysis

Let a 1-D cut of the image consist of M regions of uniform intensity, such that the intensity profile of the jth region is

$$I_j(x) = \begin{cases} 0 & x < x_j,\ x > x_{j+1} \\ \hat{I}_j & x_j \le x \le x_{j+1} \end{cases} \tag{1}$$

where the constants $\hat{I}_j > 0$ and x_j are obtained by applying some pre-specified, arbitrary distributions of intensity and sizes of regions, respectively. Consequently, the image 1-D cut is

$$I(x) = \sum_{j=1}^{M} I_j(x) \tag{2}$$

The Fourier transform of I is the sum of the Fourier transforms of I_j, that is,

$$\tilde{I}(\omega) = \frac{-i}{\omega} \sum_{j=1}^{M} \hat{I}_j(\exp(i\omega x_{j+1}) - \exp(i\omega x_j)) \tag{3}$$

where $\omega = 2\pi f$, with f being the spatial frequency. Therefore, the power spectrum of I is

$$|\tilde{I}(\omega)|^2 = \frac{1}{\omega^2} \left| \sum_{j=1}^{M} \hat{I}_j(\exp(i\omega x_{j+1}) - \exp(i\omega x_j)) \right|^2 \tag{4}$$

Consequently, as the spatial frequency increases, the power spectrum tends to fluctuate (the parenthetical term on the right-hand side), but its envelope falls inversely proportionally to the square of the frequency (the $1/\omega^2$ term). The only exception occurs at very low spatial frequencies: when $\omega \ll 1/(x_{j+1} - x_j)$ for all j, one can approximate the term inside the parenthesis in Eq. (4) with the first term of the Taylor series, yielding

$$|\tilde{I}(\omega)|^2 \approx \left| \sum_{j=1}^{M} \hat{I}_j(x_{j+1} - x_j) \exp(i\omega x_j) \right|^2 \tag{5}$$

As a result, for frequencies so low that the corresponding periods are longer than the longest step in the 1-D cut, the envelope of the power spectrum tends to remain constant with frequency.

A.2. Two-dimensional analysis

Let us now define the two-dimensional (2-D) model of the image from the 1-D cuts and analyze its power spectrum. The definition is as follows: If $I(r)$ is the image model, then $I(\mathbf{a} + z\mathbf{b})$ has the same properties as our 1-D cut, where $\mathbf{a}$ is a constant vector, $\mathbf{b}$ is a constant unit vector ($|\mathbf{b}| = 1$), and z is a variable. In the analysis that follows, $\tilde{I}$ is calculated only in the horizontal direction of the frequency vector ($\boldsymbol{\omega} = (\omega_x, \omega_y)$). It will be apparent that the general tendencies (or envelope) of the Fourier transform of $I(r)$ is the same regardless of the orientation of $\boldsymbol{\omega}$. Hence, isotropic averaging of $\tilde{I}(\omega)$ is not necessary. From Eq. (3), the Fourier transform of $I(r)$ in the horizontal orientation ($\mathbf{b} = (1, 0)$) is

$$\begin{aligned} \tilde{I}(\omega_x) &= \int_{-\infty}^{+\infty} \mathrm{d}y \int_{-\infty}^{+\infty} \mathrm{d}x \exp(i\omega_x x)\, I(x,y) \\ &= \frac{-i}{\omega_x} \left[\int_{-\infty}^{+\infty} \mathrm{d}y \sum_{j=1}^{M(y)} \hat{I}_j(\exp(i\omega_x x_{j+1}(y)) \right. \\ &\quad \left. - \exp(i\omega_x x_j(y))) \right] \end{aligned} \tag{6}$$

where $M(y)$ is the number of steps in the section at the vertical position y, and $\hat{I}_j(y)$ and $x_j(y)$ are the amplitude and beginning of the jth step at that position, respectively. Thus, the power spectrum is

$$\begin{aligned} &|\tilde{I}(\omega_x)|^2 \\ &= \frac{1}{\omega_x^2} \left| \int_{-\infty}^{+\infty} \mathrm{d}y \sum_{j=1}^{M(y)} \hat{I}_j(\exp(i\omega_x x_{j+1}(y)) - \exp(i\omega_x x_j(y))) \right|^2 \end{aligned} \tag{7}$$

Because Eqs. (4) and (7) are similar, one can reach similar (though not identical) conclusions for 2-D stepwise images as those for the 1-D cuts. For 2-D stepwise images, the power spectrum tends to be constant at low frequencies until the frequency becomes sufficiently high that its corresponding spatial period is approximately of the size of the largest constant-intensity region in the image. After this frequency, the spectrum tends to fall with frequency as indicated by the $1/\omega_x^2$ term in Eq. (7).

However, the spectrum of the 2-D model does not fall as the square of the frequency as those of the 1-D cuts. The main reason is that for 2-D images, the $x_j(y)$ term in Eq. (7) may change slowly with y. This y-dependence coupled to the integration by y may remove some of the fluctuations of the absolute term of the

right-hand side of Eq. (7) and impart a ω_x-dependent trend. Another way to understand this trend is to rearrange Eq. (6) as follows:

$$\tilde{(}\omega_x) = \int_{-\infty}^{+\infty} \mathrm{d}x \, \exp(i\omega_x x) \int_{-\infty}^{+\infty} \mathrm{d}y I(x, y) \tag{8}$$

This equation is essentially the 1-D Fourier transform of the function obtained by averaging the image over the direction orthogonal to the oscillation of the Fourier component. This averaging would tend to smear the stepwise tendency of the image. This smearing would be equivalent to a low-pass filter, causing the spectrum to fall more steeply than with frequency squared.

A.3. Individual constant-intensity regions

Next, we prove that individual 2-D objects generally have spectra that fall as frequency cubed.

From Harmonic Analysis, one has the following theorem (Montgomery, 1994): Let C be a simple, closed, piecewise C^1 curve in the plane, let R denote its interior region, and let $\bar{\chi}_{\mathrm{R}}(\omega)$ denote the Fourier transform of the indicator function of R. Then, as $\omega \to \infty$

$$\int_{\omega' \geq \omega} \mathrm{d}\omega' \|\tilde{\chi}_R|(\omega')|^2 \approx \frac{L}{2\pi^2\omega} \tag{9}$$

In this theorem, piecewise C^1 means that except for a finite number of points, the curve is continuous and has continuous first-order partial derivatives, the indicator function is defined as 1 on R and 0 otherwise, and L is the length of the perimeter of C.

The rotationally averaged power spectrum of the indicator function at frequency ω,

$$\langle|\tilde{\chi}_R(\omega)|^2\rangle_\theta$$

is the integral of $|\tilde{\chi}_R(\omega^1)|^2$ in an infinitesimal ring between ω and $\omega + \delta\omega$ divided by the area of the ring. Therefore, from Eq. (9), as $\omega \to \infty$,

$$\langle|\tilde{\chi}_R(\omega)|^2\rangle_\theta \approx \frac{1}{4\pi^3\omega\delta\omega}\left(\frac{L}{2\pi^2\omega} - \frac{L}{2\pi^2(\omega+\delta\omega)}\right) \tag{10}$$

Using a Taylor expansion of Eq. (10), one obtains as $\omega \to \infty$

$$\langle|\tilde{\chi}_R(\omega)|^2\rangle_\theta \approx \frac{1}{4\pi^3\omega^3} \tag{11}$$

If one multiplies the indicator function by the intensity of the region, then

$$\langle|\tilde{\chi}_R(\omega)|^2\rangle_\theta = \frac{I^2L}{4\pi^3\omega^3} \tag{12}$$

In conclusion, the envelopes of the power spectra of single 2-D regions of constant-intensity tend to fall as the frequency cubed at high frequencies.

A.4. Multi-region slopes

Is there a relationship between the spectra of individual constant-intensity regions and an image that is a collage of such regions? In this section, we show that if the regions are statistically independent, as proposed by Ruderman (1997), then such a relationship exists. At low frequencies, the envelope of the spectrum will have zero slope regardless of the shapes of the regions (Section 5.2). At intermediate and high frequencies, the log–log slope of the spectrum of the image is to a good approximation a weighted mean of the log–log slopes of the individual regions. Hence, at high frequencies, the envelope of the spectrum will theoretically fall as frequency cubed (Section 5.3), and at intermediate frequencies, the envelope of the spectrum will fall with a log–log slope between -3 and 0. (Those are the minimum and maximum log–log slopes of the envelopes for individual constant-intensity regions.) The range of intermediate frequencies will be wide, if the range of sizes of the regions of the image is itself wide.

If the image is a collage of constant-intensity regions, then its Fourier transform is

$$\tilde{F}(\omega_x, \omega_y) = \sum_j F_j(\omega_x, \omega_y) \exp(i\theta_j(\omega_x, \omega_y)) \tag{13}$$

where F_j and θ_j are the amplitude and phase of the transform of the jth region, respectively. The power spectrum of the image is $|\tilde{F}|^2$, which after expansion becomes

$$\begin{aligned}|\tilde{F}(\omega_x, \omega_y)|^2 = &\sum_j F_j^2(\omega_x, \omega_y) \\ &+ \sum_j \sum_{k \neq j} F_j(\omega_x, \omega_y) \exp(i\theta_j(\omega_x, \omega_y)) \\ &\times F_k(\omega_x, \omega_y) \exp(-i\theta_k(\omega_x, \omega_y))\end{aligned} \tag{14}$$

Next, we must take the mean of this equation over orientations. To do so, we define orientation in the frequency domain as ϕ, such that $\omega_x = \omega\cos(\phi)$ and $\omega_y = \omega\sin(\phi)$, with $\omega = \sqrt{\omega_x^2 + \omega_y^2}$. The mean spectrum over orientations is then

$$G(\omega) = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\phi \, |\tilde{F}(\omega\cos\phi, \omega\sin\phi)|^2 \tag{15}$$

which by using Eq. (14) becomes

$$\begin{aligned}G(\omega) = &\frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}\phi \sum_j F_j^2(\omega\cos\phi, \omega\sin\phi) \\ &+ \sum_j \sum_{k \neq j} F_j(\omega\cos\phi, \omega\sin\phi) \\ &\times \exp(i\theta_j(\omega\cos\phi, \omega\sin\phi)) \\ &\times F_k(\omega\cos\phi, \omega\sin\phi)\exp(-i\theta_k(\omega\cos\phi, \omega\sin\phi)\end{aligned} \tag{16}$$

Let us assume that the shapes of region j and of region k are statistically independent for all j and k. This assumption may not be true for particular pairs of regions in an image. However, in rich natural images, the vast majority of pairs of regions will have such independence. In this case, for the second term of this equation, the means of $F_j \exp(i\theta_j)$ and $F_k \exp(i\theta_k)$ can be evaluated separately, yielding

$$G(\omega) = \sum_j G_j(\omega) + \sum_j \sum_{k \neq j} H_j(\omega) \exp(i\alpha_j(\omega)) H_k(\omega) \exp(-i\alpha_k(\omega)) \quad (17)$$

where G_j and $H_j \exp(i\alpha_j)$ are the means over orientations of F_j^2 and $F_j \exp(i\theta_j)$, respectively, with H_j being a real positive function. Therefore,

$$G(\omega) = \sum_j G_j(\omega) + 2\sum_j \sum_{k>j} H_j(\omega) H_k(\omega) \cos(\alpha_j(\omega) - \alpha_k(\omega)) \quad (18)$$

The first thing that this equation shows is that if the image is a collage of constant-intensity regions, then its power spectrum is flat at low frequencies. This flatness follows from G_j and H_j being derived from the F_j in Eq. (16). Because F_j are individual-region spectra, their envelopes are flat at low frequencies.

How does the spectrum fall at intermediate and high frequencies? To answer this question, we first show that for these frequencies one can neglect the second term of Eq. (18). On one hand, the sum in the second term has more components. If the number of regions, M, is high, then there are M and $M^2/2$ components in the first and second terms, respectively. On the other hand, because the regions are independent, the α_j can attain any value, making the sign of the components in the second term essentially random. Consequently, the second term behaves as a random walk. It follows that the absolute value of this term is of the order

$$2(\sqrt{M^2/2})(2\bar{H}^2/\pi) = 2\sqrt{2}M\bar{H}^2/\pi$$

where $\bar{H}^2$ is the mean of H_j over all regions[6]. In comparison, the first term is of the order of $M\bar{G}$, where $\bar{G}$ is the mean of G_j over all regions. Hence, a comparison between the first and second terms of Eq. (18) essentially boils down to comparing the magnitudes of the G_j and the H_jH_k. At first sight, these terms appear to be similar, since they originate from F_j^2 and F_jF_k in Eq. (16), respectively. However, for the H_j but not the G_j, the mean over orientations in Eq. (16) undergoes sign changes due to variations in θ_j. These changes cause several of the H_k components to be of much lower magnitude than $\sqrt{G_j}$, making the second term of Eq. (18) much smaller than the first. Therefore, at intermediate and high frequencies, we have, to a good approximation,

$$G(\omega) = \sum_j G_j(\omega) \quad (19)$$

The log–log slope, $S(\omega)$, of the spectrum $G(\omega)$ is

$$S(\omega) = \frac{\mathrm{d}\log(G(\omega))}{\mathrm{d}\log(\omega)} = \frac{\omega \dfrac{\mathrm{d}G(\omega)}{\mathrm{d}\omega}}{G(\omega)} \quad (20)$$

By substituting Eq. (19) for $G(\omega)$ in Eq. (20), one gets

$$S(\omega) = \omega \sum_j \frac{\dfrac{\mathrm{d}G_j(\omega)}{\mathrm{d}\omega}}{\sum_k G_k(\omega)} \quad (21)$$

Let us denote by $S_j(\omega)$ the log–log slope of the spectrum of the individual region j. From Eq. (20)

$$S_j(\omega) = \frac{\omega \dfrac{\mathrm{d}G_j(\omega)}{\mathrm{d}\omega}}{G_j(\omega)} \quad (22)$$

By multiplying the numerator and denominator of Eq. (21) by G_j, and using the formula for S_j in Eq. (22), one gets

$$S(\omega) = \sum_j S_j(\omega) \frac{G_j(\omega)}{\sum_k G_k(\omega)} \quad (23)$$

In conclusion, at intermediate and high frequencies, the log–log slope of the spectrum is the weighted mean of the log–log slopes of the individual regions. Eq. (23) shows that the weights are the relative spectral energies.

A.5. Gradients of illumination

What effect does violating the assumption that images are stepwise have on our conclusions? The following example of violations through gradients of illumination shows that the power spectra have a steeper fall with these gradients than without them. Because the 1-D and 2-D analyses without gradients of illuminations yield similar results (compare Eqs. (4) and (7)), we restrict the analysis here to the 1-D case. Let us modify Eq. (1) to include the simplest model for these gradients, that is,

$$I_j(x) = \begin{cases} 0 & x < x_j,\ x > x_{j+1} \\ \hat{I}_j + \gamma x & x_j \le x \le x_{j+1} \end{cases} \quad (24)$$

[6] The first factor of -2 in the left hand side of this equality is the same as that before the double sum in Equation 18. The $\sqrt{M^2/2}$ factor is the square root of the number of steps. Finally, the $2/\pi$ is the mean of $|\cos|$ over all orientations.

where γ is a constant. Because Eq. (2) is still valid, one only needs to consider the Fourier transform of I_j, which is,

$$\tilde{I}_j(\omega) = \frac{-i}{\omega}\hat{I}_j(\exp(i\omega x_{j+1}) - \exp(i\omega x_j)) + \frac{\gamma}{i\omega}\left[(x_{j+1}\exp(i\omega x_{j+1}) - x_j\exp(i\omega x_j)) - \frac{1}{i\omega}(\exp(i\omega x_{j+1}) - \exp(i\omega x_j))\right] \quad (25)$$

The right-hand side of this equation is the sum of two terms. The first term is identical to the terms in Eq. (3). Therefore, this term will tend to make the power spectrum fall inversely with the square of the frequency (see Eq. (4)). However, when expanded, the second term of Eq. (25) contains a ω^{-2} sub-term. Because the power spectrum is $|\tilde{I}_j(\omega)|^2$, this sub-term leads the spectrum to be a polynomial of powers of ω^{-2}, including ω^{-4}. Consequently, the power spectrum will fall faster than just as ω^{-2}.

References

Atick, J. J., & Redlich, A. N. (1990). Towards a theory of early visual processing. *Neural Computation*, *2*, 308–320.

Atick, J. J., & Redlich, A. N. (1992). What does the retina know about natural scenes? *Neural Computation*, *4*, 196–210.

Balboa, R. M. (1997). *Teorías para la inhibición lateral en la retina basadas en autocorrelación, curtosis y aspereza local de las imágenes naturales*. Doctoral Dissertation. University of Alicante, Alicante, Spain.

Balboa, R. M., & Grzywacz, N. M. (2000a). The minimal local-asperity hypothesis of early retinal lateral inhibition. *Neural Computation*, *12*, 1485–1517.

Balboa, R. M., & Grzywacz, N. M. (2000b). Occlusions and their relationship with the distribution of contrast in natural images. *Vision Research*, *40*, 2661–2669.

Balboa, R. M., & Grzywacz, N. M. (2000c). The role of early retinal lateral inhibition: more than maximizing luminance information. *Visual Neuroscience*, *17*, 77–90.

Balboa, R. M., De Juan, J., & Grzywacz, N. M. (1997). Discrepancy between biology and the Predictive Coding Theory of lateral inhibition based on the statistics of natural scenes. *Investigative Ophthalmology and Visual Science*, *38*, S1017.

Barlow, H. B. (1961a). Possible principles underlying the transformation of sensory messages. In W. A. Rosenblith, *Sensory communication*. Cambridge, MA: MIT Press.

Barlow, H. B. (1961b). The coding of sensory messages. In W. H. Thorpe, & O. L. Zangwill, *Current problems in animal behaviour*. Cambridge: University Press.

Barlow, H. B. (1989). Unsupervised learning. *Neural Computation*, *1*, 295–311.

Baylor, D. A., Lamb, T. D., & Yau, K.-W. (1979). Responses of retinal rods to single photons. *Journal of Physiology*, *288*, 613–634.

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

Carlson, C. R. (1978). Thresholds for perceived image sharpness. *Photographic Science and Engineering*, *22*, 69–71.

Dan, Y., Atick, J. J., & Reid, R. C. (1997). Efficient coding of natural scenes in the lateral geniculate nucleus: Experimental test of a computational theory. *Journal of Neuroscience*, *16*, 3351–3362.

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*, *4*, 2379–2394.

Field, D. J. (1993). Scale-invariance and self-similar 'wavelet' transforms: An analysis of natural scenes and mammalian visual systems. In M. Farge, J. Hunt, & X. Vassilicos, *Wavelets, fractals and Fourier transforms: new developments and new applications* (pp. 151–191). Oxford: Oxford University Press.

Field, D. J. (1994). What is the goal of sensory coding? *Neural Computation*, *6*, 559–601.

Field, D. J., & Brady, N. (1997). Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Research*, *37*, 3367–3383.

Fuortes, M. G. F., & Yeandle, S. (1964). Probability of occurrence of discrete potentials waves in the eye of the *Limulus*. *Journal of Physiology*, *47*, 443–463.

Mandelbrot, B. B. (1977). *Fractals: Form, chance, and dimension*. San Francisco, CA: W.H. Freeman.

McCarthy, S. T., & Owen, W. G. (1996). Preferential representation of natural scenes in the salamander retina. *Investigative Ophthalmology and Visual Science*, *37*, S674 (Abstract).

Mobley, C. (1995). In M. Bass, *Handbook of optics*, vol. 1. Fundamentals, techniques and design. New York: McGraw-Hill.

Montgomery, H. L. (1994). Regularities of distribution. In *Ten lectures on the interface between Analytic Number Theory and Harmonic Analysis* (pp. 109–124). Providence: AMS.

Ruderman, D. L., & Bialek, W. (1994). Statistics of natural images: scaling in the woods. *Physics Review Letter*, *73*, 814–817.

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

Rushton, W. A. H. (1961). The intensity factor in vision. In W. D. McElroy, & H. B. Glass, *Light and life* (pp. 706–722). Baltimore, MD: Johns Hopkins Press.

Sinclair, S. (1985). *How animals see: other visions of our world*. London: Croom Helm.

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.

Tolhurst, D. J., Tadmor, Y., & Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic Physiology and Optics*, *12*, 229–232.

Tolhurst, D. J., & Tadmor, Y. (1997). Band-limited contrast in natural images explains the detectability of changes in the amplitude spectra. *Vision Research*, *37*, 3203–3215.

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.