
Local contrast in natural images: normalisation and coding efficiency

Nuala Brady

Department of Psychology, University College Dublin, Belfield, Dublin 4, Ireland;
e-mail: nuala.brady@ucd.ie

David J Field

Department of Psychology, Cornell University, Ithaca, NY 14853, USA
Received 1 October 1999, in revised form 27 September 2000

Abstract. The visual system employs a gain control mechanism in the cortical coding of contrast whereby the response of each cell is normalised by the integrated activity of neighbouring cells. While restricted in space, the normalisation pool is broadly tuned for spatial frequency and orientation, so that a cell's response is adapted by stimuli which fall outside its 'classical' receptive field. Various functions have been attributed to divisive gain control: in this paper we consider whether this output nonlinearity serves to increase the information carrying capacity of the neural code. 46 natural scenes were analysed with the use of oriented, frequency-tuned filters whose bandwidths were chosen to match those of mammalian striate cortical cells. The images were logarithmically transformed so that the filters responded to a luminance ratio or contrast. In the first study, the response of each filter was calibrated relative to its response to a grating stimulus, and local image contrast was expressed in terms of the familiar Michelson metric. We found that the distribution of contrasts in natural images is highly kurtotic, peaking at low values and having a long exponential tail. There is considerable variability in local contrast, both within and between images. In the second study we compared the distribution of response activity before and after implementing contrast normalisation, and noted two major changes. Response variability, both within and between scenes, is reduced by normalisation, and the entropy of the response distribution is increased after normalisation, indicating a more efficient transfer of information.

1 Introduction

The idea that our perceptual systems code information efficiently is an old one (eg Attneave 1954; Barlow 1961). However, it is only in recent years, as our understanding of the structure of natural signals has grown, that such ideas have been formally tested. There are now a large number of studies which support the notion that our visual system presents an efficient means for coding the statistical structure found in the natural environment (eg Field 1987, 1994; Burton and Moorhead 1987; Atick and Redlich 1992; Van Hateren 1993). This research suggests that much of the known physiology of the early visual system allows for an efficient representation of natural scenes. Much of the research has concentrated on the relatively linear aspects of visual coding. For example, simple cells in V1 are usually treated as linear filters with parameters of orientation, spatial frequency, and position, and why the visual system might have this particular set of parameters has been the subject of many of the recent studies. However, it is well known that nonlinearities exist at the very earliest stages of visual processing, and it is important to determine whether an information-processing account can also help to explain these aspects of visual coding.

A code or representation is inefficient when its outputs are redundant or predictable in some way. In measuring efficiency, a distinction is made between first-order redundancy in which the symbols of a code (eg different neurons in a network or different response states of a single neuron) occur with unequal frequency, and higher-order redundancy in which the predictability comes from interdependences among symbols (Field 1987; see Atick 1992 for a review). This same distinction is crucial in understanding the redundancy of natural signals. In figure 1 we plot the distribution of luminance in

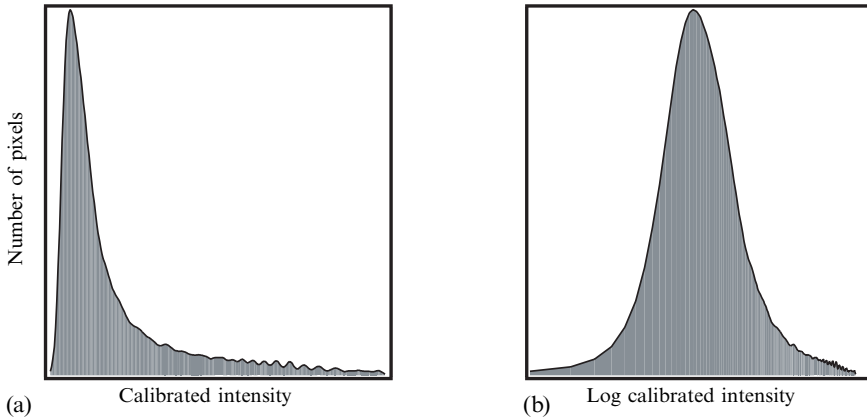


Figure 1. The distribution of luminance in natural scenes is (a) skewed toward the low end of the luminance range, and (b) is approximately normal on a log luminance axis. The luminance distribution is estimated from a set of 46 calibrated 8-bit images, where the pixel intensity is linearly related to the luminance in the original scene.

natural scenes as estimated from a set of 46 calibrated images. The distribution is skewed toward the low end of the luminance range and is approximately normal on a log luminance axis. A visual system whose neurons respond linearly to luminance will inherit this first-order redundancy; low responses will be very common and high responses relatively infrequent. A logarithmic-like transformation of luminance ensures a more equal distribution of responses (Richards 1981).

Early attempts to describe the higher-order structure of natural images focus mainly on the second-order statistics, specifically on the correlation between pairs of luminance values at different points in an image (Burton and Moorhead 1987; Field 1987; Tolhurst et al 1992). Neighbouring points in natural scenes are highly correlated, with correlation falling exponentially as the distance between the points increases. In the frequency domain, this structure is described by the image power spectrum. Natural scenes have a spectrum in which power falls as the inverse of frequency squared so that contrast energy is approximately constant in octave bands. A 'pixel' code, in which all sensors are narrowly localised in space, will inherit the redundancy described by the autocorrelation function. On the other hand, the sensors of a Fourier code which are narrowly localised in spatial frequency will be uncorrelated in the presence of natural scenes, since they are aligned with the principal components of the image (Field 1994). In this case, the second-order redundancy of the image is converted into first-order redundancy of the sensor outputs.

In visual cortex, cells are localised in both space and spatial frequency, and the trade-off in resolution is given by a cell's bandwidth. The frequency bandwidths of both simple and complex cells increase with the optimal spatial frequency of the cell and show approximate octave constancy (De Valois et al 1982). This produces a system in which a neuron's gain increases in proportion to its spatial frequency, and a code in which neurons tuned to different frequencies show roughly equal response activity in the presence of natural scenes (Field 1987; Brady and Field 1995).

In addition to considering the relative response activity of different neurons, we can also consider the histogram of response activity as measured across a population of neurons. When natural images are filtered by means of functions that resemble cortical receptive fields (eg Gabor functions) the distribution of filter outputs is highly kurtotic, peaking at zero or low response and having long exponential tails (Field 1994; Ruderman 1994). This distribution has low entropy and high redundancy relative to a Gaussian, which is the distribution having the highest entropy for a fixed variance (Rieke et al 1997).

What does first-order entropy tell us about coding efficiency? For any given cell, we can calculate the maximum information it can convey from the distribution of its responses using the statistical measure of entropy. Tolhurst (1989) estimated the amount of information transmitted about image contrast by V1 neurons in the cat, using response histograms generated by grating stimuli which varied in contrast. Owing to high response variability or 'noise', these cells are reported to convey less than 1 bit of information about contrast in 500 ms. This estimate of information transmission depends crucially on assumptions about 'noise', eg the information-carrying capacity of the code may be increased by averaging responses across a number of neurons.

In this paper, we ask whether filter output nonlinearities, specifically contrast normalisation, may serve to increase coding efficiency. The answer to this question depends on the form of the output distribution considered to be most efficient, as different distributions maximise entropy subject to different constraints (Rieke et al 1997). When the maximum firing rate or response range is constrained, a uniform distribution is optimal (Laughlin 1981), whereas the exponential distribution maximises information transmission given a constraint on the average firing rate (Baddeley et al 1997; Rieke et al 1997). The question is particularly interesting in light of the fact that response distributions with high kurtosis have become synonymous with what has been called 'sparse-distributed' coding (Field 1994; Olshausen and Field 1996). In a sparse-distributed representation each cell has roughly equal probability of firing when this is measured across the class of images being represented, but the probability of a particular cell being active for any particular image is very low. The computational advantages of this type of representation, in which the cell responses are highly independent, have been discussed (Field 1987; Barlow 1989; Field 1994), and it has been shown that neural networks which aim to maximise sparseness when trained with natural images produce localised-bandpass receptive fields similar to those in striate cortex (Olshausen and Field 1996).

Before considering how output nonlinearities may transform the response distribution of the cortical code, we provide a brief review of models of cortical-cell responses and contrast normalisation.

Cortical-cell responses are typically described with a two-stage model (Heeger 1991). In the initial stage, the response of a cortical simple cell depends on a linear sum of local image intensity, or on a nonlinear combination of linear subunits in the case of complex cells (Hubel and Wiesel 1962). But when tested physiologically, simple cells show departures from strict linearity, including a threshold or contrast below which they do not respond, and response compression and saturation at high contrasts (Movshon et al 1978; Dean 1981; Albrecht and Hamilton 1982). In addition, the excitatory response of a simple cell to its preferred stimulus may be suppressed or inhibited by the presence of a second stimulus, and this suppression has been found to be broadly tuned for spatial frequency and orientation, and to be independent of direction of motion (Bonds 1989). It has been proposed that this mutual inhibition between cells serves to normalise a cell's response with respect to the stimulus contrast within a given neighbourhood (Bonds 1991). In recent nonlinear models of cortical-cell responses the initial linear stage is followed by a normalisation stage in which the cell's response is divided by the pooled activity of a number of neighbouring cells (Heeger 1991; Carandini and Heeger 1994). The normalisation model explains a number of nonlinear aspects of cell behaviour, including response saturation, cross-orientation inhibition (Heeger 1991), and threshold effects (Tolhurst and Heeger 1997).

While various functions have been attributed to contrast normalisation, most accounts stress how divisive gain control allows a system with limited dynamic range to code visual information independently of contrast. If information about local image features (eg their size, orientation, direction of motion) is carried by the distribution

of response activity across some population of cells, it is important that the ratio of cell responses is invariant of contrast levels. The potentially damaging consequences of response saturation for such a scheme are avoided if saturation is set by the level of image contrast rather than by the neural response level (Ohzawa et al 1985; Albrecht et al 1984). A different view of contrast normalisation is presented by Bonds (1991), who reports that, when cortical inhibition is removed experimentally, the peak firing rate of cortical neurons substantially increases while the relation between contrast and response is retained. Contrast normalisation may not be necessary for maintaining a contrast-invariant representation of image features, and Bonds (1991) suggests instead that it is a means of restricting the flow of information to higher levels of processing.

A nonlinear relation between stimulus contrast and response is also characteristic of mammalian lateral geniculate nucleus and retinal ganglion cells. As there is no evidence for inhibitory contrast gain control at this stage of visual processing (Ohzawa et al 1985), it is likely that compression and saturation are hardwired. Laughlin (1981) suggests that such contrast nonlinearities serve to increase the efficiency of information transmission, and his study of contrast coding by neurons in fly retina is one of the first to demonstrate the strengths of considering neural coding from the perspective of information theory. Laughlin estimated the distribution of contrast in the environment and showed that the contrast-response function of the fly's retinal neurons matched the integral of this distribution rather well. The technique of transforming a variable such as grey-scale or colour by the integral of its frequency distribution is known as 'histogram equalisation' in image processing, where it is used to enhance the quality of degraded images. If the fly neurons responded linearly to contrast, their different response states would occur with different frequency, reflecting the frequency distribution of contrast in the environment. But with a nonlinear response to contrast which is matched to the contrast distribution, different response states occur with approximately equal probability. In information-theory terms, first-order redundancy is removed. Below, we consider whether contrast normalisation in mammalian cortex may achieve a similar end, by dynamically matching cell responses to prevailing contrasts in natural scenes. In our first study we estimate the magnitude of contrast in natural images. The methods we use have been reported previously (Brady and Field 1993) and are similar to those used by others (Tolhurst 1996; Tolhurst et al 1997).

2 Methods

2.1 Images

A total of 46 natural images were analysed in these studies; these included images of people, trees, plants, and woodland. All scenes were photographed in and around Ithaca, NY, and in Southern California with a Miranda MS-3 camera on Ilford XPI black and white slide film. The images were digitised from slide negatives at a spatial resolution of 1600×1024 pixels with a 12-bit Barneyscan slide scanner, and one or two 512×512 pixels images were taken from each scan. The images were calibrated to correct for the nonlinear transformation of luminance by the photographic film. The luminance calibration involved photographing a strip of 31 Munsell grey-scale swatches (Munsell Color, Baltimore, MD), whose reflectance ranged from 3.1% to 90% in approximately equal logarithmic steps. The chart was photographed outdoors at three different camera exposures, providing an input intensity range of around 7500 : 1. The photographs of the Munsell swatches were processed and scanned in the same way as the natural scenes, and a fifth-order polynomial function relating the digitised swatch values to the corresponding input values was used to linearise the scene data. Scenes with particularly large luminance ranges, which involved extrapolating beyond the input range, were discarded. Further details of the collection and calibration processes

are described elsewhere (Field 1994; Field and Brady 1997). After calibration, the 512×512 pixels digitised images had 8 bits of grey-level with the 256 levels linearly spaced.

2.2 Code

The images were analysed with log Gabor filters whose parameters of orientation, spatial frequency, and phase were chosen to match those of mammalian striate cortical cells. While a two-dimensional Gabor function provides a good descriptor of the spatial weighting function of cat cortical cells (Jones and Palmer 1987), other functions have been proposed as more suited to capturing the low-frequency bias seen in the cells' amplitude response to spatial frequency (Hawken and Parker 1987). One such function is the log Gabor (Field 1987), whose frequency response is Gaussian on log frequency axis, and is described as:

$$\exp \left\{ - \ln \left(\frac{f}{f_0} \right) / 2 \left[\ln \left(\frac{\sigma}{f_0} \right) \right]^2 \right\},$$

where f_0 is the optimal spatial frequency and σ is a measure of frequency bandwidth. Spatial-frequency bandwidths were chosen to increase in direct proportion to optimal spatial frequency, giving a 1.58 octave bandwidth which is within the range reported for primate simple and complex cells (De Valois et al 1982). The orientation half-bandwidth of the filters was set to 25° for a fixed aspect ratio of approximately 0.9, which is comparable to measurements on real cells (De Valois et al 1982). We restricted the analyses to five optimal frequencies (8, 16, 32, 64, and 128 cycles/image), and four central orientations (0 to $3\pi/4$ radians, in steps of $\pi/4$). Edge effects were avoided by looking only at the inner 384×384 pixels of the filtered images. For each spatial position, frequency, and orientation, we calculated the response of both even and odd symmetric filters. All images were logarithmically transformed prior to filtering, so that the filters responded to the logarithm of a luminance ratio: $\log(A) - \log(B) = \log(A/B)$. The advantage of this approach is that equal changes in contrast produce equal changes in filter response independent of the local intensity in the image. We note that transformations other than a logarithmic one have been proposed to model light adaptation (Naka and Rushton 1966), and that in another study contrast in complex scenes has been measured by the method of dividing the response of a bandpass filter by that of a low-pass filter of the same scale (Peli 1991).

In study 1, the response of each filter was calibrated relative to its response to a logarithmically transformed sinusoidal grating whose spatial frequency and orientation were matched to those of the filter. The end result is a receptive field that responds linearly to contrast. It is important to recognise that we are not arguing that cells respond linearly to contrast. Rather, we investigate whether a nonlinear response to contrast like that found in the visual system increases the information-carrying capacity of the neurons. The calibration allows us to express image contrast in terms of the commonly used Michelson metric. Local contrast ranged from 0 to 100% and was quantised in steps of 1%. The distribution is based upon the responses of 271 319 040 sensors (384×384 pixels, 46 images, 5 spatial frequencies, 4 orientations, and 2 phases). In study 2, we measured the distribution of response activity without calibration, using 26 images, 3 spatial frequencies, 4 orientations, and 2 phases. This distribution was also collected after implementing contrast normalisation in the code where the response of each filter was normalised by the integrated response of neighbouring filters plus a constant. Following Heeger (1991) we choose the normalisation pool for a given filter to include all orientations and neighbouring spatial frequencies, eg a filter tuned to 16 cycles/image with a central orientation of $\pi/4$ is normalised by the combined outputs of twelve filters tuned to 8, 16, and 32 cycles/image at all 4 orientations as shown in figure 2.

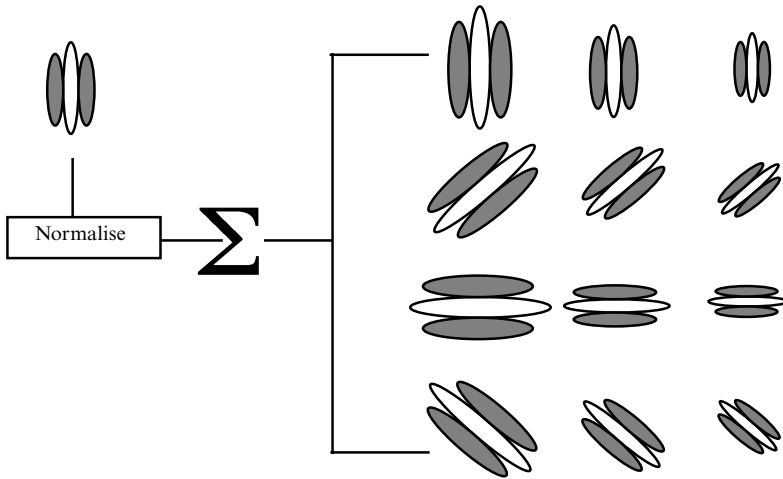


Figure 2. The contrast normalisation pool includes filters tuned to neighbouring spatial frequencies and all orientations. See text for further detail.

The constant used in the normalisation pool is known as the ‘semisaturation constant’. When the filter responses are low and the semisaturation constant is relatively high, normalisation is expansive; correspondingly, at high response levels, normalisation is compressive. In choosing the semisaturation constant we were guided by the original distribution of contrasts in natural scenes, which is heavily biased toward low contrasts. We ran the analyses using two different values of the semisaturation constant, corresponding to the response to gratings of 2% and 10% contrast.

3 Study 1: The distribution of local contrast in natural images

The purpose of this study was to estimate the distribution of contrasts in natural scenes as ‘seen’ by mammalian striate cortical neurons, and to estimate the variability in local contrast which such cells encounter. As our knowledge of how cortical cells respond to contrast is largely based upon studies with sinusoidal grating patterns, we measured image contrast in terms of the Michelson metric. As discussed in section 2, the images were logarithmically transformed and then analysed by means of sensors whose spatial filtering properties matched those of striate cortical cells. The response of each sensor was calibrated relative to its response to a sinusoidal grating of preferred spatial frequency and orientation, giving an estimate of Michelson contrast at each point in the images.

3.1 Results and discussion

The histogram of image contrasts is plotted in figure 3a. The form of the distribution is familiar from previous studies (Field 1994; Ruderman 1994): it peaks at low contrasts and falls exponentially with increasing contrast. Of particular interest are the third and fourth moments. The skew, or third moment, of a distribution characterises the asymmetry of a distribution about its mean, where zero skew indicates perfect symmetry. The distribution of contrasts in figure 3a is close to zero, showing that the distributions of positive and negative contrasts are very similar. The kurtosis, or fourth moment, is a measure of the peakedness of a distribution and, like skew, is only meaningful when compared to the (two-sided) Gaussian or normal distribution which has a kurtosis of zero. The distribution of local contrasts for this set of images has a positive kurtosis of 2.20.

As real cortical cells signal image contrast by their firing rate and typically have very low maintained rates, positive and negative contrasts are signalled by separate

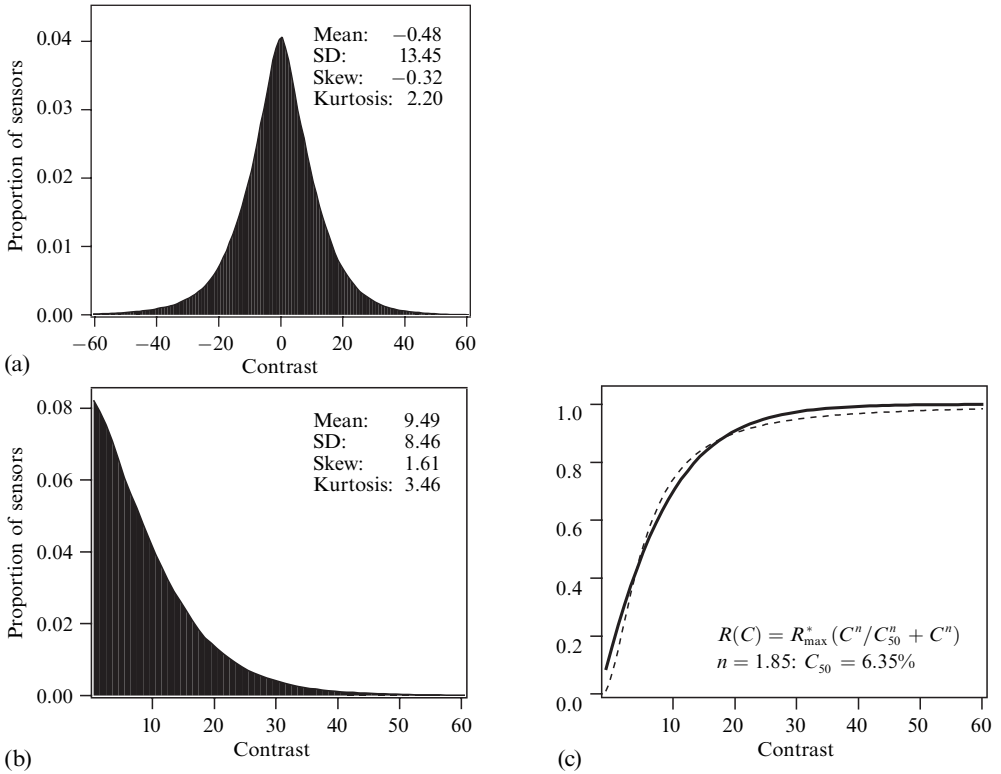


Figure 3. The distribution of image contrast in natural scenes: (a) both positive and negative, and (b) positive alone. In this study, sensor responses were pooled across 46 images, 5 spatial frequencies, and 4 orientations. The contrast bin width was 1%. (c) The integral of the positive-contrast histogram shown by the solid line defines the optimal contrast-response function. A hyperbolic function shown by the dotted line with $R_{\max} = 1.0$, $C_{50} = 6.35\%$, and $n = 1.85$ provides a good fit to the data. SD = standard deviation.

channels, ie groups of cells which are phase-shifted by 180° relative to each other. In figure 3b we plot the distribution of positive contrasts and report its moments. Given the symmetry of the two-sided histogram in figure 3a, the values for the distribution of negative contrasts (not reported) are very similar. The distribution has a mean contrast of 9.5% and a standard deviation of 8.5%, confirming previous reports that contrasts are low in the natural environment (Laughlin 1981).

Ruderman (1994) suggests that it is useful to think of the distribution of contrasts in natural images as a composite of different distributions corresponding to regions of relatively high or low texture. An analysis of the individual scenes in the set reveals much variability among the contrast distributions. Figure 4 shows three images from the set alongside their local (positive) contrast distributions. The image of the dark trees on a white, snowy background has regions of very high contrast, but the overall distribution (with a low mean of 3.67% and high kurtosis of 31.62) reflects the large number of blank or relatively featureless regions. By comparison, the more uniformly textured image of the Brussels sprouts has a higher mean contrast (11.11%) and lower kurtosis (2.60). Figure 5a shows a scatter plot of mean local contrast and kurtosis for all 46 images. As expected, mean image contrast decreases with increasing kurtosis (a linear regression of mean contrast on kurtosis returns a slope of -0.19 , Pearson's $r = -0.74$). Skew and kurtosis show a strong positive relation (slope = 0.09 , Pearson's $r = 0.97$) as do standard deviation and mean contrast (slope = 0.81 , Pearson's $r = 0.93$), as shown in figures 5b and 5c. Given that standard deviation increases with mean image

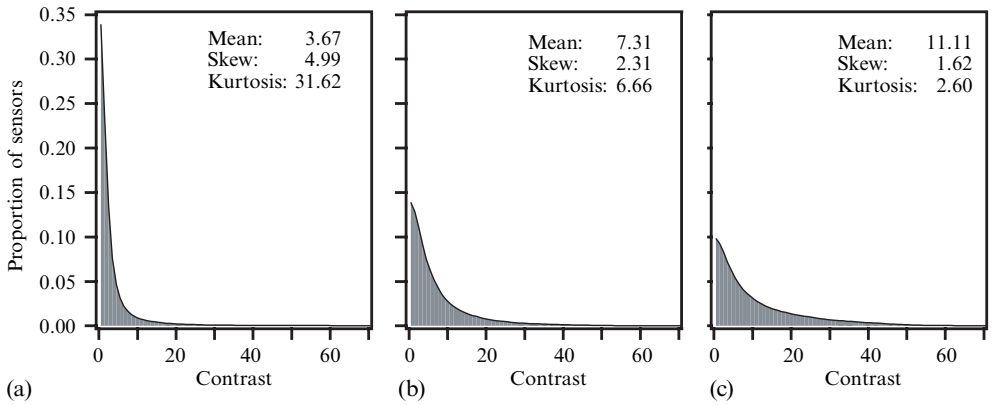


Figure 4. 3 images from the set of 46 with their contrast distributions plotted beneath. Image (a) on the left, with lots of featureless regions, has a high kurtosis (31.62), while the uniformly textured image (c) on the right has low kurtosis (2.60).

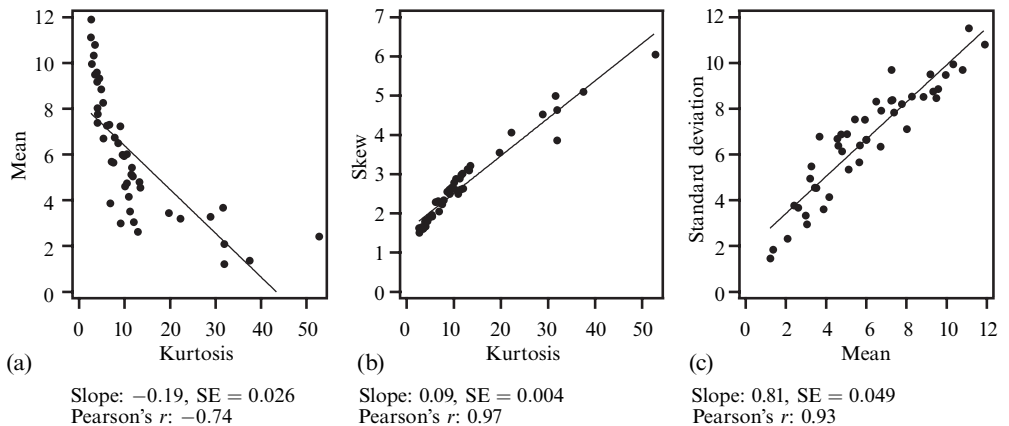


Figure 5. The relation between various moments of the contrast distributions as estimated from 46 images. (a) Mean contrast varies inversely with kurtosis; (b) skew and kurtosis are positively correlated as are (c) the standard deviation and the mean.

contrast, we use the coefficient of variation (which expresses the standard deviation as a fraction of the mean) as a measure of the variability in local contrast within scenes; for the current set it ranges from a low of 0.89 to a high of 1.85.

The range of contrasts within and between scenes is of particular interest in considering neural coding. In figure 6a, the mean image contrast is plotted for each of the 46 images in order of increasing mean, with error bars to indicate ± 1 coefficient

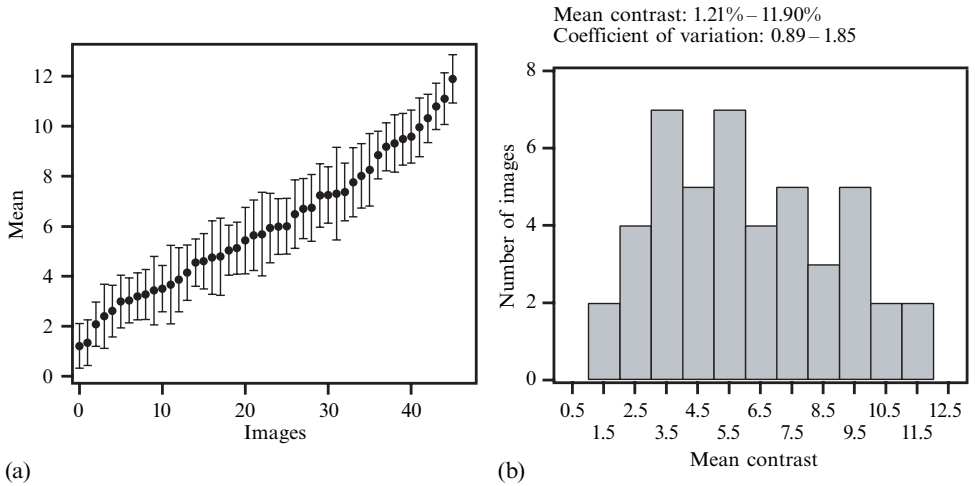


Figure 6. Variability of local contrast within and between images. (a) The mean contrast is plotted for each of 46 images in order of increasing mean and ranges from a low of 1.21% to a high of 11.9%. The error bars show ± 1 coefficient of variation which is roughly similar across images. (b) The distribution of mean contrasts for the set of forty-six images is roughly normal.

of variation. The mean varies from a low of 1.21% to a high of 11.90% with the coefficient of variation within images being roughly similar across the range of mean contrasts. The lower and upper quartiles of the distribution of mean contrasts in figure 6b are 3.67% and 8.02%, respectively; and the lower and upper deciles are 2.99% and 9.59%, respectively. Mean contrast therefore varies by a factor of 2.2 to 3.2. Tolhurst et al (1997) report a comparable estimate in a published abstract, although it is difficult to make a comparison with that study without the full details.

If individual cells convey 1 bit of information or less (Tolhurst 1989), the visual system must have a strategy for handling the natural range of contrasts. One possibility is that there are contrast ‘channels’, analogous to spatial-frequency or orientation channels, which together cover the full range of contrasts. Support for this idea comes from the fact that cells in mammalian striate cortex vary in the placement of their response range, with thresholds measured under the same experimental conditions ranging from less than 0.5% to as high as 7% in cat (Dean 1981), and the semisaturation constant in cat and in primate cells ranging from 1% to greater than 40% (Albrecht and Hamilton 1982). Contrast adaptation is another likely candidate. Ohzawa et al (1985) show that a large percentage of cat striate cortical cells shift their operating range horizontally on the contrast axis when exposed to new average contrast levels. This contrast gain control mechanism appears to be unique to cortex and is not found in retinal ganglion or LGN neurons. Bonds (1991) has shown that contrast gain control is relatively rapid and can be activated by quite low contrasts such as those in the natural environment. In his study, cells show a clear contrast hysteresis, producing a higher response to a given test contrast if it has been preceded by contrasts lower than the test than if preceded by contrasts higher than the test. Contrast gain dropped by an average of 0.36 log units when this protocol was used, with gain shifts occurring as rapidly as within 3 s.

Finally, we ask whether the contrast-response functions of cortical neurons are matched to the distribution of contrasts in the natural environment. Following Laughlin (1981), we plot the integral of the positive contrast distribution in figure 3c. It defines the ‘optimal’ contrast-response function of cortical cells, assuming a goal of equalising the different response states of a neuron. The shape of this function is qualitatively similar to the contrast-response function of cortical cells, having an initial linear region

which spans approximately 1 log unit of contrast and a region of compression and saturation. Quantitatively, too, there are similarities. Albrecht and Hamilton (1982) use a hyperbolic function to describe the contrast-response function of cortical cells in both cat and primate cortex. This function, like the Naka–Rushton equation which is used to describe the intensity-response function of retinal neurons (Naka and Rushton 1966), describes a saturating system and is characterised by three parameters: R_{\max} is the maximum response, C_{50} is the semisaturation constant or contrast at which the response reaches half its maximum, and n is an exponent which describes the rate of change:

$$R(C) = R_{\max} \frac{C^n}{C_{50}^n + C^n}.$$

Cells vary considerably in the positioning of their dynamic range and the current estimate of 6.35% for the semisaturation constant is within the range reported for real cells. Similarly, there is variability in the exponent with a value between 2.0 and 3.0 providing a good fit for the majority of cells (Albrecht and Hamilton 1982), and recent models of contrast normalisation suggest that this exponent should be 2.0 (Carandini and Heeger 1994). In figure 3c the exponent is 1.85. But while the optimal contrast-response function, defined here as the integral of the contrast distribution for natural scenes, is similar in form to that of cortical cells, we must take caution in concluding that this is the most efficient use of a limited dynamic range. As discussed above, producing a uniform distribution of response activity may not be the goal of neural coding, as different distributions maximise entropy subject to different constraints.

4 Study 2: Response distributions before and after normalisation

In this study, 26 images were analysed by using simulated simple cells and the response distributions compared in two conditions. In the first condition, responses were collected across filters tuned to 3 spatial frequencies (16, 32, and 64 cycles/image), 4 orientations (0 to $3\pi/4$ radians in steps of $\pi/4$ radians), and 2 phases (0° and 180°). In the second condition, the response distribution was collected across the same range of frequency, orientation, and phase, but with the response of each filter now normalised by the integrated responses of neighbouring filters plus a semisaturation constant as described in section 2.

4.1 Results and discussion

The distribution of response activity (scaled to a range of 0 to 100 units) before normalisation is shown by the solid line in figure 7a. Less than 0.001% of sensor responses exceeded 100 units after scaling. Not surprisingly, the distribution is very similar in shape to the distribution of local contrasts in figure 3b. It shows a sharp peak, with $\sim 26\%$ of sensor responses falling in the lowest bin (0 to 1 unit). We calculated the moments of the distribution over the first 54 response bins after first imposing a threshold which excludes this first bin. In terms of the local contrast measurement, excluding the first bin corresponds to a threshold of around 1% contrast. After thresholding, the mean response is 7.08 and the standard deviation is 7.50. The distribution has a high positive kurtosis of 6.83 and a skew of 2.39.⁽¹⁾ The corresponding distribution after normalisation is shown by the dashed line in figure 7a. The distribution is less peaked and has a less extensive tail over the high-contrast region. The mean response level is 7.57 and the standard deviation is 5.41. Both kurtosis (2.37) and skew

⁽¹⁾Note that the high kurtosis in part reflects the fact that this is a one-sided distribution. While the kurtosis of a two-sided distribution may be meaningfully compared to that of a two-sided Gaussian which is defined to have a kurtosis of zero, this is not the case here. We restrict ourselves to comparing the moments of the one-sided distributions before and after contrast normalisation and we use a measure of entropy when considering how coding efficiency changes.

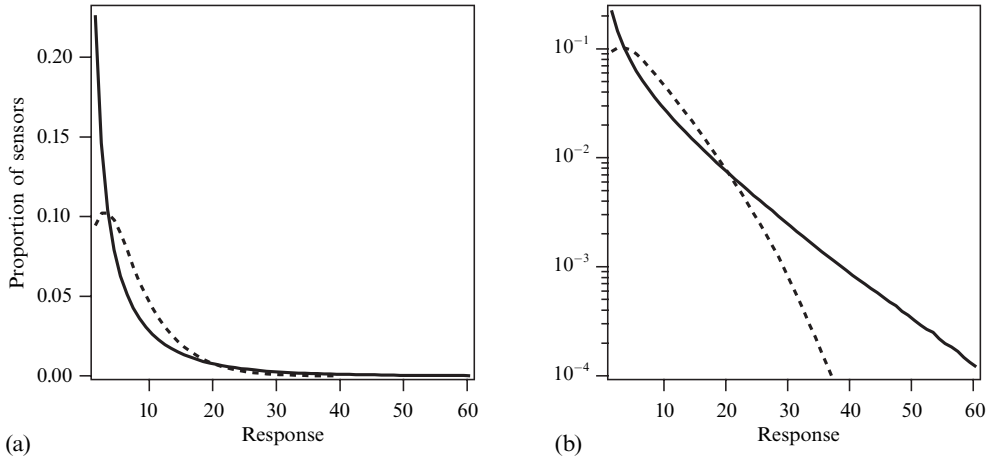


Figure 7. The distribution of response activity before (solid line) and after (dashed line) contrast normalisation shown on (a) linear and (b) log–linear plots. The distributions are based on 26 images, 3 spatial frequencies, 4 orientations, and 2 phases.

(1.40) are considerably reduced. This distribution was generated by using a semisaturation constant corresponding to a simple-cell response to a grating of 10% contrast. Lower semisaturation values give similar results; a 2% semisaturation constant leads to a distribution with a kurtosis value of 1.91 and a skew of 1.26 for this set of images.

Both distributions are replotted in figure 7b on log–linear axes, on which an exponential fall-off appears as a straight line. We note that the response distribution, particularly before normalisation, closely approximates an exponential distribution. The exponential distribution is of particular interest in considering the efficiency of neuronal coding, in that it is the distribution which maximises entropy, subject to the constraint of a fixed mean firing rate (Rieke et al 1997). There is some evidence that real neurons code information efficiently according to this criterion: Baddeley et al (1997) measured the mean firing rate of cells in areas V1 of cat and area IT of macaque in response to video sequences of natural scenes, and report that the distributions are well described by an exponential.

In order to compare the information-carrying capacity of the code before and after normalisation, we measured the entropy, S , of the two distributions using the Shannon metric, $S = -\sum_i p_i \log_2 p_i$, where p_i is the proportion of sensor responses falling in the i th bin). The distributions in figure 7 clearly have different variance. However, since we are interested in the shape of the distribution, we first correct for unequal variance. We find that the entropy for the distributions before and after the nonlinear normalisation is 3.55 and 4.12: a difference of 0.57 bits. If entropy is measured without correction for unequal variance, there is little difference between the distributions (~ 0.13 bits).

The effects of contrast normalisation are also obvious when we compare the statistics of the individual image distributions before and after normalisation. In figure 8, we compare mean response level, standard deviation, skew, and kurtosis before and after normalisation for each of the 26 images. The distributions were generated with a semisaturation constant of 10%. (A semisaturation constant of 2% yields very similar results.) In all plots, the images are ordered according to increasing kurtosis of the original distributions. Both skew and kurtosis are considerably reduced after normalisation. More importantly, response variability between scenes is much reduced after normalisation, as seen by the more restricted range of both the mean response level and the standard deviation. Variability within images is also reduced as evidenced by the smaller error bars in the top-left plot for the normalised images.

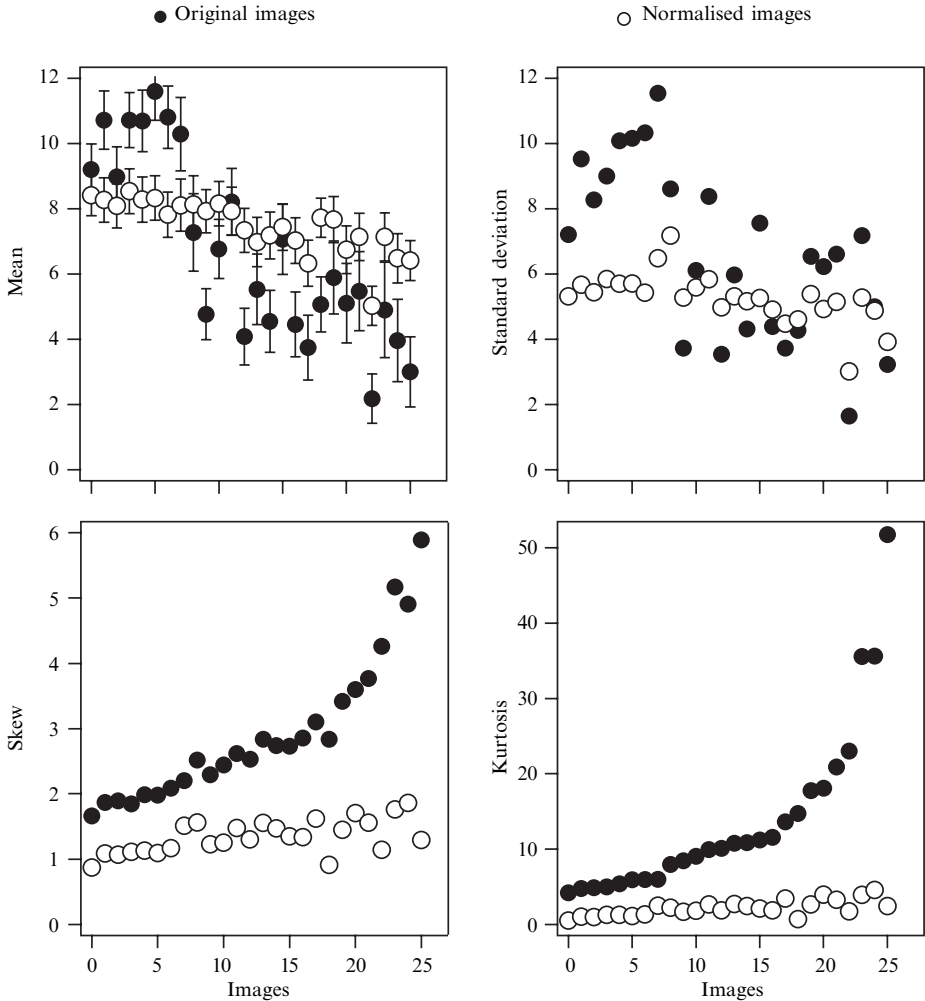


Figure 8. A comparison of mean response level, standard deviation, skew, and kurtosis before (solid symbols) and after (open symbols) normalisation for each of the 26 images. In all plots the images are ordered according to increasing kurtosis of the original distributions. The error bars in the top left plot indicate ± 1 coefficient of variation.

5 General discussion

In study 1 we estimated the distribution of local contrast in natural scenes. Our measure of image contrast is similar to that of Peli (1991) in that it is based on the output of a cortical code. In our study, the filter responses to natural scenes are calibrated with respect to their response to a sinusoidal grating, so that the contrast at each point in an image is given a Michelson value. This allows us to relate what we know about the neuronal coding of contrast to what we know about contrast in natural scenes. Both the shape of the distribution and the range of contrast are of interest, and we first consider the issue of variability.

Contrast is quite low in natural scenes; we estimate a mean local contrast of $\sim 9.5\%$ (standard deviation $\sim 8.5\%$) from a set of 46 calibrated images. Very high contrasts, such as those associated with sharply defined boundaries, are particularly rare. The variation of contrast within scenes is high, with standard deviation varying in direct proportion to mean contrast. Variation in contrast between scenes is also quite high. Mean local contrast ranged from a low of 1.21% to a high of 11.90% in

our set of 46 images. The distribution of mean contrasts is approximately normal and the mean contrast varies by a factor of 2.2 or 3.2, where these estimates are based on the upper and lower quartiles and deciles of the distribution, respectively.

The dynamic range of cortical cells is quite low. Albrecht and Hamilton (1982) estimate the range of cat and primate cortical cells to cover somewhat less than 1 log unit of contrast and Tolhurst (1989) estimates that individual neurons in cat cortex convey less than 1 bit of information (two levels of contrast) by their instantaneous firing rate. What strategies does the visual system use to handle the range of naturally occurring contrasts? As suggested previously, response averaging, either in time or across neurons, may boost the information-carrying capacity of the cortical code (Tolhurst 1989). Alternatively, different ranges of contrast may be coded by different groups of cells, analogous to how orientation and frequency are coded by relatively independent 'channels'. There is evidence that cortical cells differ in the positioning of their dynamic range as shown by the variability in estimates of the semisaturation constant, which is the contrast at which cells reach half their maximum response (Albrecht and Hamilton 1982). The distribution of semisaturation constants for primate cells (Albrecht and Hamilton 1982, figure 13) is not dissimilar in shape to the distribution of mean contrasts in natural images (our figure 6), although it peaks at higher contrasts.

The process of contrast gain control (Ohzawa et al 1985) or contrast normalisation (Bonds 1989, 1991) has also been proposed as a means of overcoming a limited dynamic range. This study shows that contrast normalisation clearly reduces the variability in local image contrast, or rather the variability in the neural response to contrast. As such, it allows for a more efficient transfer of information. Although the process leads to a loss of information about absolute contrast across a scene, the ratio of responses of cells tuned to different frequencies or orientations within a given region of a scene remains roughly constant, as their gain is controlled by a similar normalisation pool. As has been noted by a number of researchers, this allows information about the size, the orientation, and other aspects of image features, which is carried by the distribution of activity across different cells, to be signalled independently of contrast (Bonds 1991; Heeger 1991). In this way contrast gain control is similar to light adaptation which makes the visual response to image contrast invariant to overall light levels.

We now consider the form or shape of the response distribution. As noted in section 1, a response distribution with high kurtosis in which a large percentage of neurons exhibit zero or very low response has become associated with sparse coding. Does contrast normalisation, which clearly reduces kurtosis, make the code less sparse? The answer is no, and the issue of a neural threshold is crucial here. In our study we have implemented contrast normalisation after imposing a threshold which corresponds to a simple-cell response to a sinusoidal grating of $\sim 1\%$ contrast. The choice is arbitrary, and we note that physiological studies (eg Dean 1981; Albrecht and Hamilton 1982) show that cortical cells vary in their thresholds. Dean (1981) reports a range of 0.5% to 7% grating contrast for cat cells. But without imposing a threshold, our implementation of contrast normalisation would lead to a very 'busy' image in which extremely low responses (all well below the semisaturation constant) are boosted upward. By imposing a threshold, a large percentage of cells ($\sim 26\%$ for our set of images with a threshold of $\sim 1\%$ contrast) remain silent. The kurtosis of the response distribution will remain high.

Kurtosis is one parameter which has been used to characterise the shape of the response distribution of simulated cortical cells when presented with natural scenes. The high kurtosis distinguishes the distribution from a Gaussian and can arise from various sources in the environment. It has been demonstrated that a neural network which attempts to maximise this high kurtosis or non-Gaussian output will result in

'neurons' selective to regions of space, spatial frequency (scale), and orientation (Olshausen and Field 1996). These results suggest that contrast is distributed sparsely across space, spatial frequency, and orientation. In contrast to Baddeley (1996), we think that all these aspects, including the local changes in contrast, are important (and 'interesting') to a visual system.

However, in terms of the information-carrying capacity of a single neuron, a kurtotic output is not the most efficient means of coding this sparse structure in images. Although the effect is not large, we have shown that nonlinear contrast gain control does increase the information-carrying capacity of the individual neurons. Although we cannot conclude that this is the reason for the gain control, such results suggest that it is likely to be one reason for this early nonlinearity.

References

- Albrecht D G, Farrar S B, Hamilton D B, 1984 "Spatial contrast adaptation characteristics of neurones recorded in the cat's visual cortex" *Journal of Physiology (London)* **347** 713–739
- Albrecht D G, Hamilton D B, 1982 "Striate cortex of monkey and cat: contrast response function" *Journal of Neurophysiology* **43** 217–237
- Atick J J, 1992 "Could information theory provide an ecological theory of sensory processing?" *Network* **3** 213–251
- Atick J J, Redlich A N, 1992 "What does the retina know about natural scenes?" *Neural Computation* **4** 196–210
- Attneave R, 1954 "Informational aspects of visual perception" *Psychological Review* **61** 183–193
- Baddeley R J, 1996 "Searching for filters with 'interesting' output distributions: an uninteresting direction to explore?" *Nature (London)* **381** 560–561
- Baddeley R, Abbott L F, Booth M C A, Sengpiel F, Freeman T, Wakeman E A, Rolls E T, 1997 "Responses of neurons in primary and inferior temporal visual cortices to natural scenes" *Proceedings of the Royal Society of London, Series B* **264** 1775–1783
- Barlow H B, 1961 "Possible principles underlying the transformations of sensory messages", in *Sensory Communication* Ed. W A Rosenblith (Cambridge, MA: MIT Press) pp 217–234
- Barlow H B, 1989 "Unsupervised learning" *Neural Computation* **1** 295–311
- Bonds A B, 1989 "Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex" *Visual Neuroscience* **2** 41–55
- Bonds A B, 1991 "Temporal dynamic of contrast gain in single cells of the cat striate cortex" *Visual Neuroscience* **6** 239–255
- Brady N, Field D J, 1993 "Local contrast in natural scenes" *Investigative Ophthalmology & Visual Science* **34**(4) 781
- Brady N, Field D J, 1995 "What's constant in contrast constancy? The effects of scaling on the perceived contrast of bandpass patterns" *Vision Research* **35** 739–756
- Burton G J, Moorhead I R, 1987 "Color and spatial structure in natural scenes" *Applied Optics* **26** 157–170
- Carandini M, Heeger D J, 1994 "Summation and division by neurons in primate visual cortex" *Science* **264** 1994
- Dean A F, 1981 "The relationship between the response amplitude and contrast for cat striate cortical neurones" *Journal of Physiology (London)* **318** 413–427
- De Valois R L, Albrecht D C, Thorell L G, 1982 "Spatial frequency selectivity of cells in macaque visual cortex" *Vision Research* **22** 545–559
- 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
- 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
- Hawken M J, Parker A J, 1987 "Spatial properties of neurones in the monkey striate cortex" *Proceedings of the Royal Society of London, Series B* **231** 251–288
- Heeger D J, 1991 "Nonlinear model of neural responses in cat visual cortex", in *Computational Models of Visual Processing* Eds M Landy, J A Movshon (Cambridge, MA: MIT Press) pp 119–133
- Hubel D H, Wiesel T N, 1962 "Receptive fields, binocular interaction and function architecture in the cat's visual cortex" *Journal of Physiology (London)* **160** 106–154
- Jones J P, Palmer L A, 1987 "The two-dimensional spatial structure of simple receptive fields in cat striate cortex" *Journal of Neurophysiology* **58** 1187–1211

-
- Laughlin S B, 1981 "A simple coding procedure enhances a neuron's information capacity" *Zeitschrift für Naturforschung C* **36** 910–912
- Movshon J A, Thompson I D, Tolhurst D J, 1978 "Spatial summation in the receptive field of simple cells in the cat's striate cortex" *Journal of Physiology (London)* **283** 53–77
- Naka K I, Rushton W A H, 1966 "S-potentials from luminosity units in the retina of fish (Cyprinidae)" *Journal of Physiology (London)* **185** 587–599
- Ohzawa I, Sclar G, Freeman R D, 1985 "Contrast gain control in the cat's visual system" *Journal of Neurophysiology* **54** 651–667
- Olshausen B A, Field D J, 1996 "Emergence of simple-cell receptive field properties by learning a sparse code for natural images" *Nature (London)* **381** 607–609
- Peli E, 1991 "Contrast in complex images" *Journal of the Optical Society of America A* **7** 2030–2040
- Richards W A, 1981 "A lightness scale from image intensity distributions", AI Memo No 648, Artificial Intelligence Laboratory, MIT, 545 Technology Square, Cambridge, MA
- Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W, 1997 *Spikes: Exploring the Neural Code* (Cambridge, MA: MIT Press)
- Ruderman D L, 1994 "The statistics of natural images" *Network: Computation in Neural Systems* **5** 517–548
- Tolhurst D J, 1989 "The amount of information transmitted about contrast by neurons in the cat's visual cortex" *Visual Neuroscience* **2** 409–413
- Tolhurst D J, 1996 "The limited contrast response of single neurones in cat striate cortex and the distribution of contrast in natural scenes" *Journal of Physiology (London)* **497** 64P
- Tolhurst D J, Heeger D J, 1997 "Comparison of contrast-normalization and threshold models of the responses of simple cells in cat striate cortex" *Visual Neuroscience* **14** 293–309
- Tolhurst D J, Laughlin S B, Lauritzen J S, 1997 "Modelling the variation in the contrast of natural scenes and the function of contrast normalization in the mammalian visual cortex" *Journal of Physiology (London)* **504** 123P–124P
- Tolhurst D J, Tadmor Y, Chao T, 1992 "The amplitude spectra of natural images" *Ophthalmic and Physiological Optics* **12** 229–232
- Van Hateren J H, 1993 "Spatiotemporal contrast sensitivity of early vision" *Vision Research* **33** 257–267

