INSTITUTE OF PHYSICS PUBLISHING NETWORK: COMPUTATION IN NEURAL SYSTEMS

Network: Comput. Neural Syst. **12** (2001) 255–270 www.iop.org/Journals/ne PII: S0954-898X(01)23518-0

Characterizing the sparseness of neural codes

B Willmore[1] **and D J Tolhurst**

Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, UK

E-mail: bw200@cam.ac.uk

Received 4 December 2000

Abstract

It is often suggested that efficient neural codes for natural visual information should be 'sparse'. However, the term 'sparse' has been used in two different ways—firstly to describe codes in which few neurons are active at any time ('population sparseness'), and secondly to describe codes in which each neuron's lifetime response distribution has high kurtosis ('lifetime sparseness'). Although these ideas are related, they are not identical, and the most common measure of lifetime sparseness—the kurtosis of the lifetime response distributions of the neurons—provides no information about population sparseness.

We have measured the population sparseness and lifetime kurtosis of several biologically inspired coding schemes. We used three measures of population sparseness (population kurtosis, Treves–Rolls sparseness and 'activity sparseness'), and found them to be in close agreement with one another. However, we also measured the lifetime kurtosis of the cells in each code. We found that lifetime kurtosis is uncorrelated with population sparseness for the codes we used.

Lifetime kurtosis is not, therefore, a useful measure of the population sparseness of a code. Moreover, the Gabor-like codes, which are often assumed to have high population sparseness (since they have high lifetime kurtosis), actually turned out to have rather low population sparseness. Surprisingly, principal components filters produced the codes with the highest population sparseness.

1. Introduction

Neural codes for visual information have probably evolved to maximize their metabolic or informational efficiency (Barlow 1989). It has been suggested that the most efficient neural codes perform a *sparse* encoding of natural input. This idea has taken two forms, which we will refer to as 'population sparseness' and 'lifetime sparseness'.

[1] Corresponding author.

1.1. Population sparseness

We use the term population sparseness to refer to a property discussed by Field (1987, 1994) in his discussion of *sparse-dispersed*[2] encoding of natural input. In a sparse-dispersed code, there is a large population of coding neurons available, but only a small subset of this population is active in response to any single stimulus—i.e. the code has high population sparseness. Also, *different* small subsets of the coding population will be activated by different stimuli—i.e. the coding is 'dispersed' across the population. Field suggests that sparse-dispersed coding is efficient because relatively few cells are producing action potentials at any time, and so relatively little energy is consumed (but see the discussion).

Sparse-dispersed coding is usually contrasted with *compact* coding, in which there is a relatively small population of coding neurons, but a large proportion of these cells is activated by each stimulus. Compact codes are efficient in the sense that they require small numbers of coding neurons. However, it has been suggested that compact codes require more action potentials than population-sparse codes (Field 1994). Principal components analysis (PCA) can be used to produce compact codes for visual information (Hancock *et al* 1992).

It is clear from the description above that population sparseness is a property of a *population* of coding neurons, not of individual cells. An individual neuron cannot be 'sparse' in this sense; it can only be a member of a sparse-coding population. To measure the population sparseness of a code directly, it would therefore be necessary to measure the responses of the entire coding population and to find the proportion of neurons which were active in response to each stimulus. The overall population sparseness would then be the average proportion of cells which were active at any time. For real neurons, this is clearly an impossible task, since it is not possible to record from large numbers of individual neurons simultaneously. However, for computational models of neural coding, where all responses can be calculated, it is simple enough to measure directly population sparseness. In this paper, we suggest measures for doing this.

1.2. Lifetime sparseness—kurtosis of a neuron's lifetime response distribution

It has also been suggested that cells in efficient neural codes should have high *lifetime sparseness*, or *lifetime kurtosis* (e.g. Field 1987, 1994). This means that each neuron should respond only rarely but, when a neuron does respond, it should produce responses whose magnitudes are relatively large. This idea is consistent with recent work on independent components analysis (ICA) (Comon 1994), which has demonstrated (Bell and Sejnowski 1997, van Hateren and van der Schaaf 1998, van Hateren and Ruderman 1998) that simple learning rules can produce codes in which individual coding units have high lifetime kurtosis (and therefore low entropy). This may be desirable because it means that action potentials (which occur rarely) each carry a relatively large amount of information.

Two lines of evidence have suggested that the receptive fields of simple cells in primary visual cortex (V1) may have been selected in evolution to have high lifetime kurtosis. First, ICA and related methods have been used to learn codes with high lifetime kurtosis for the spatial information in monochrome photographs of natural scenes (Fyfe and Baddeley 1995, Harpur and Prager 1996, Olshausen and Field 1996, 1997, Bell and Sejnowski 1997, van Hateren and van der Schaaf 1998, van Hateren and Ruderman 1998). These codes contain coding units which closely resemble the orientation-specific and spatial-frequency-specific receptive fields of simple cells in V1 (Hubel and Wiesel 1962, Movshon *et al* 1978, Jones and Palmer 1987,

[2] Field uses the term *sparse-distributed* to refer to this type of coding. However, 'distributed' has been used previously (Hinton *et al* 1986) to mean 'many neurons are active in response to each stimulus'. We therefore prefer the term 'sparse-dispersed'. This is discussed in Willmore *et al* (2000).

DeAngelis *et al* 1993). Second, direct measurement of the responses of individual simple cells in cat, monkey and ferret has shown that they each respond relatively rarely to natural visual input (Baddeley *et al* 1997, Tolhurst *et al* 1999, Vinje and Gallant 2000), suggesting they may have high lifetime kurtosis.

1.3. Sparseness of the responses of a population of neurons

The lifetime response kurtosis of the cells in a code might seem to be related to the population sparseness of coding; both concepts of sparseness imply that neural activation by natural stimuli should be a relatively uncommon event. However, we demonstrate here that, although lifetime kurtosis is an interesting property of neural codes, it is not the same as population sparseness. Population sparseness is a property of the instantaneous responses of a population, whilst lifetime kurtosis is a property of the lifetime response of an individual cell. We show that a code may have high lifetime kurtosis without having high population sparseness. Moreover, the codes with the highest population sparseness are not those with the highest lifetime kurtosis.

2. Methods

We investigated several linear codes, all of which have been used previously for encoding visual information. We investigated the population and lifetime sparseness of these codes when encoding a large set of natural image fragments.

2.1. Coding schemes

We examined nine different codes in total. Examples of filters from each code are shown in figure 1. Three are possible models of simple cell receptive fields: Gabor filters (Daugman 1980, Marcelja 1980, Field and Tolhurst 1986, Jones and Palmer 1987), independent component filters (Bell and Sejnowski 1997) and Olshausen–Field bases (Olshausen and Field 1996, 1997). Two are non-orthogonal sets: Gaussian filters and difference-of-Gaussian filters. The remaining four are complete orthogonal codes: principal components filters, sinusoids, two-dimensional Walsh functions and single-pixel filters. Each set consisted of 256 16×16-pixel linear filters, and each filter was standardized to have a mean of zero and standard deviation of 1.0, giving each filter an equal *a priori* expectation of response.

Principal components filters were obtained by finding the eigenvalues of the covariance matrix of one of the sets of 10 000 natural image fragments described below. This was done separately on raw, log-transformed and pseudo-whitened image sets (see section 2.2).

The independent components filters were obtained by performing ICA on sets of raw and log-transformed images, using the FastICA package (Hyvärinen 1999) under Matlab. The Olshausen–Field bases were obtained from pseudo-whitened images using Matlab software kindly provided by Bruno Olshausen. Parameter λ/σ was set to 0.14, and ε was reduced through training, starting with 1000 iterations at $\varepsilon = 500$, then 1000 iterations at $\varepsilon = 250$ and finally 1000 iterations at $\varepsilon = 100$.

The Gabor filters were self-similar, even-symmetric filters with peak frequencies f_N, $f_N/2$, $f_N/4$ and $f_N/8$, where f_N is the Nyquist frequency (i.e. 8 cycles/fragment). Four orientations were used at each sampling point (0°, 45°, 90° and 135°). Each filter had spatial frequency bandwidth (ratio of half-maximum values) of 1 octave and full angular bandwidth at half-height of 41°. This set was designed, following Navarro *et al* (1996), to maximize the completeness of the code whilst maintaining near-biological bandwidths.

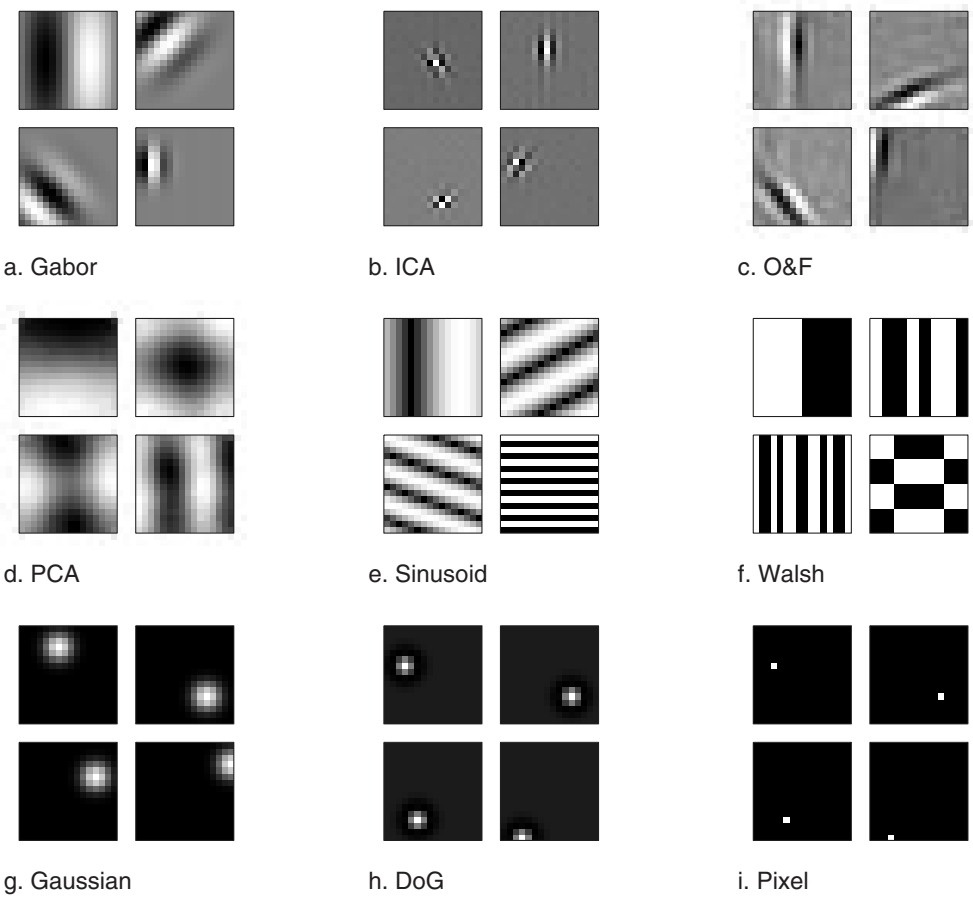

Figure 1. Examples of each of the filter types used in our analysis. PCA was performed separately on raw, log-transformed and pseudo-whitened images; only the PCs (nos 2, 4, 5 and 7) of raw images are shown. ICA was performed separately on raw and log-transformed images; only the ICs of log-transformed images are shown.

The Gaussian filters had standard deviation of 1.5 pixels, whilst the difference-of-Gaussians filters were produced from two Gaussians with standard deviations of 0.5 (centre) and 2 pixels (surround), and equal volumes.

2.2. Natural image fragments

We obtained a set of 64 monochrome natural scenes (256 × 256 pixels each), whose subjects included vegetation, landscapes, buildings, animals and people. As described in Tolhurst *et al* (1992), the photographs were calibrated against photographs of Munsell grey-paper charts, to enable conversion of each pixel value into a luminance value. This procedure corrects for the luminance non-linearities of the film and photographic processes, and resulted in more than 1000 grey levels being represented in each image. Sample images are shown in figure 2.

We pseudorandomly extracted ten sets of 10 000 16 × 16-pixel fragments from these natural scenes. It is common to preprocess images using either a log transform (van Hateren and van der Schaaf 1998) or a pseudo-whitening filter (which attempts to flatten the approximately $1/f$

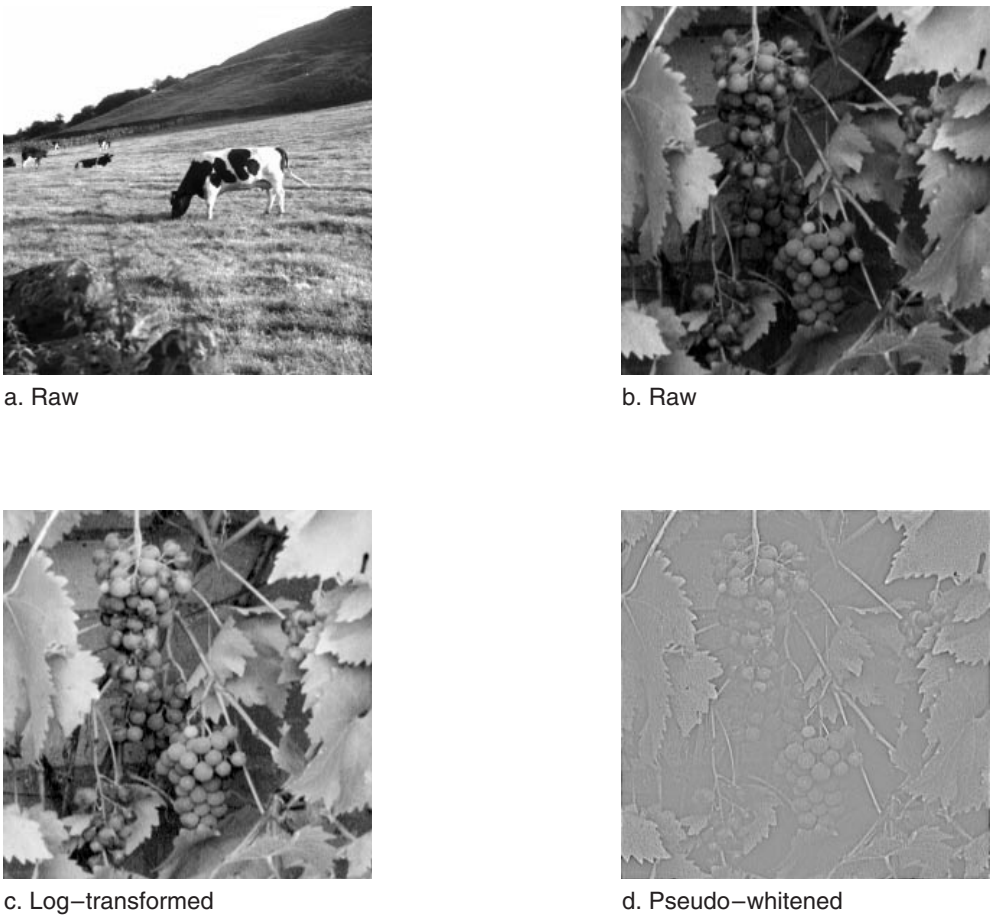

Figure 2. Sample images: (*a*), (*b*) raw (calibrated but otherwise unprocessed); (*c*) log-transformed; (*d*) pseudo-whitened. The images have been gamma-corrected for printing (square-root transform) and have been plotted so that each fills the grey-level range.

amplitude spectrum of natural images by multiplying each Fourier coefficient by f) described in Olshausen and Field (1997) and Willmore *et al* (2000). These forms of preprocessing may reflect some of the properties of retinal processing. We therefore used the images in three conditions—unprocessed (or 'raw'), log-transformed and pseudo-whitened.

Three sets of principal components filters were produced, one from images in each preprocessing condition. These three filter sets were tested on images which had been preprocessed in the corresponding way. The other filter sets, which were produced by analysing natural images (independent component filters and Olshausen–Field bases), were tested only on images which had been preprocessed in the same way as the images that had originally been analysed—i.e. raw and log transformed for independent components filters, and pseudo-whitened for Olshausen–Field bases. All the remaining filter sets were tested on images from all three preprocessing conditions.

2.3. Calculation of the responses of model cells

The response of each model cell to each image fragment was calculated by taking the scalar product of the receptive field with the pixel values in the fragment. This process was repeated for every model cell, for every image fragment, producing a set of $256 \times 10\,000$ model responses for each image set. We took this set of responses and calculated its mean population sparseness and mean lifetime kurtosis. We repeated the calculations on all ten sets of 10 000 image fragments, in order to obtain estimates of the standard error of measurements made on these responses. We found that the standard errors were extremely small, and these will not be discussed further.

2.4. Measurement of lifetime kurtosis

The lifetime kurtosis of a model neuron is commonly used to measure how rarely that neuron responds to visual input (e.g. Bell and Sejnowski 1997, Olshausen and Field 1997). Kurtosis is the fourth moment of a distribution of responses, and measures the 'peakedness' of the distribution. It can only be used if the distribution is unimodal, and if it has approximately symmetrical positive and negative tails. Since real simple cells rarely have symmetrical response distributions, kurtosis cannot be used usefully to measure their responses (but see Vinje and Gallant 2000). However, kurtosis is appropriate for measuring the response distributions of linear filters, since these are usually unimodal and approximately symmetrical.

If M stimuli are presented to a neuron, and the neuron produces responses $r_1 \ldots r_M$, then the lifetime kurtosis, K_L, is given by

$$K_L = \left\{ \frac{1}{M} \sum_{i=1}^{M} \left[\frac{r_i - \bar{r}}{\sigma_r} \right]^4 \right\} - 3 \tag{1}$$

where $\bar{r}$ and σ_r are the mean and standard deviation of the responses. A distribution will have high positive kurtosis if it contains many responses which are small (compared to the standard deviation σ_r), and only a few responses which are very large. This corresponds to a strongly peaked response distribution. A distribution will have large negative kurtosis if all responses are present equally often in the distribution. The Gaussian distribution is used as a comparison, and therefore has zero kurtosis.

To obtain an overall lifetime kurtosis value for each code, we took the mean of the lifetime kurtosis values of all the 256 cells in the code.

2.5. Measurement of population sparseness

We used three different measures of population sparseness: population kurtosis, a modified Treves–Rolls measure and 'activity sparseness'. We found the population sparseness of each code's responses to each image fragment using each of these three measures. We then obtained a single overall value of each population sparseness measure for each code by taking the mean population sparseness of that code in response to each of the 10 000 images that was presented.

2.5.1. Population kurtosis. Like lifetime kurtosis, population sparseness is a measure of the infrequency of strong neural responses—the difference is the set of responses over which this is measured. It is therefore possible to measure population sparseness using *population kurtosis*. This is the kurtosis of the responses of the entire population of neurons to a single

stimulus. If there are N neurons, then the population kurtosis for a single image is

$$K_{\mathrm{P}} = \left\{ \frac{1}{N} \sum_{j=1}^{N} \left[\frac{r_j - \bar{r}}{\sigma_r} \right]^4 \right\} - 3. \tag{2}$$

Equations (1) and (2) are identical, except that lifetime kurtosis is measured for the distribution of responses of a single neuron to many (M) stimuli, whereas population kurtosis is measured for the distribution of responses of the entire coding population of N neurons to a single stimulus.

2.5.2. Treves–Rolls population sparseness. Treves and Rolls (1991) propose an alternative measure of population sparseness. Like kurtosis, it measures the shape of a response distribution. Unlike kurtosis, however, it is appropriate only for responses (such as those of real neurons) which range from 0 to $+\infty$. The Treves–Rolls population sparseness, S_{T}, of a distribution is given by

$$S_{\mathrm{T}} = \frac{\left[\sum_{j=1}^{N} r_j/N\right]^2}{\sum_{j=1}^{N} \left[r_j^2/N\right]}. \tag{3}$$

We have made two modifications to the Treves–Rolls measure: we add a modulo sign, so that the measure can be used for responses (like those of linear filters) which range between $-\infty$ and $+\infty$; also we subtract the sparseness value from one, to give a measure, S'_{T}, that increases as population sparseness increases:

$$S'_{\mathrm{T}} = 1 - \frac{\left[\sum_{j=1}^{N} |r_j|/N\right]^2}{\sum_{j=1}^{N} \left[r_j^2/N\right]}. \tag{4}$$

This measure could also be used to characterize the lifetime response distributions of single neurons. If the measure were applied to the M different responses that a single neuron produced in response to M stimulus presentations (cf equation (1)), then it would produce a measure of the proportion of strong responses, which would be closely related to the lifetime kurtosis of that neuron.

2.5.3. Activity sparseness. We also use a measure of population sparseness which is a direct implementation of the standard definition of population sparseness: 'how few cells are active at any one time'. First, we take the distribution of responses of the coding population to a single image. Then we set a threshold value for the responses to each image, equal to the standard deviation of the responses in the distribution. Any neural responses whose magnitudes are larger than this threshold value are considered to be 'on'. Responses whose magnitudes are smaller than the threshold are 'off'. Then the *activity sparseness* of the code is simply the number of cells which are 'off' in response to a particular stimulus.

We have chosen to set the threshold equal to the standard deviation of the population response to each picture. Clearly, many other threshold criteria are possible. We chose this criterion for two reasons. Firstly, we are effectively standardizing by dividing by the standard deviation (since the response distributions of our zero-dc filters have means which are close to zero), and this is analogous to the divisions performed in the equations for kurtosis and the Treves–Rolls measure. Secondly, a standardization of this sort may be related to the energy normalization believed to occur in primary visual cortex (Heeger 1992).

Table 1. The first three columns show the mean kurtosis of the lifetime responses of coding units in the nine different models, studied with up to three image-preprocessing conditions ('Raw' = calibrated but otherwise unprocessed images; 'Log' = log-transformed images; 'Whi' = pseudo-whitened images). The remaining three columns show the mean kurtosis of the population response distributions of each model.

	Lifetime response kurtosis			Population response kurtosis		
	Raw	Log	Whi	Raw	Log	Whi
Gabor	18.5	9.03	18.47	21.66	21.81	5.37
ICA	18.74	9.58		6.42	5.32	
O&F			17.21			2.17
PCA	8.24	4.28	8.13	32.64	34.82	3.07
Sinusoid	10.33	5.15	12.05	27.12	27.73	4.62
Walsh	10.69	4.99	10.91	27.75	28.29	4.01
Gaussian	7.37	4.02	8.93	0.21	0.12	0.52
DoG	9.67	4.44	11.2	1.70	1.29	1.74
Pixel	6.76	3.26	11.13	1.66	1.04	2.68

3. Results

We took nine coding schemes for visual information (Gabor filters, independent components filters, Olshausen–Field bases, principal components filters, Gaussian filters, difference-of-Gaussian filters, sinusoids, single pixels and Walsh functions), and calculated their responses to fragments of natural images, which were either 'raw' (calibrated but otherwise unprocessed) or had been log-transformed or had been pseudo-whitened. This makes a total of 24 data sets (not all models were studied with all three image preprocessing protocols). We then evaluated the distributions of these responses by measuring their lifetime response kurtosis, and three measures of population sparseness—population kurtosis, modified Treves–Rolls sparseness and activity sparseness.

First, we ensured that the three measures of population sparseness produced comparable results. Figure 3 gives scatterplots which show how the Treves–Rolls (figure 3(*a*)) and activity sparseness (figure 3(*b*)) of our models' responses varied with the population kurtosis. The correlations are good (Pearson's $r = 0.96$ and 0.97 respectively; 22 dof). This suggests that these measures are indeed quantifying the same property of the responses. It also confirms that the population kurtosis and Treves–Rolls measures *are* consistent with the standard definition of population sparseness, quantified as *activity sparseness*: 'how few neurons are active at any one time'. The relationships in figure 4 are not straight lines, because kurtosis is calculated with a higher exponent than the other two measures. A plot of lifetime kurtosis against a lifetime Treves–Rolls measure (not shown) also had a high correlation ($r = 0.90$).

Having ascertained that the relative population sparseness values are not seriously affected by the particular sparseness measure we choose, we use *population kurtosis* as our one measure of population sparseness for the rest of this paper.

We then looked at the lifetime kurtosis values for each code, and compared them with the population sparseness (as measured by population kurtosis). The raw values are shown in table 1, and a scatterplot comparing population sparseness and lifetime kurtosis is shown in figure 4. The Olshausen–Field and independent component filters are selected to have high lifetime kurtosis, and this is seen in our results. However, it is immediately clear from figure 4 that there is no significant relationship between population sparseness and lifetime kurtosis (Pearson's $r = -0.37$; dof $= 22$, $P > 0.05$); in fact, if there is a correlation, it is negative. Thus it is clear that, although lifetime kurtosis is often used as an indication of population

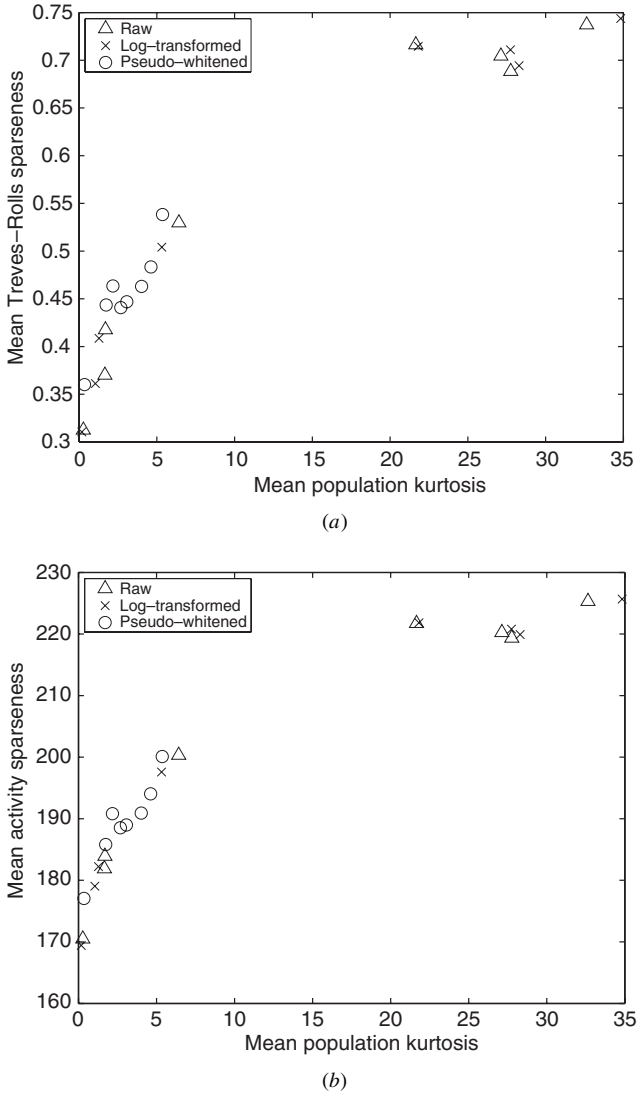

Figure 3. Scatterplots showing the relationships between the three sparseness measures. (*a*) Mean Treves–Rolls sparseness plotted against mean kurtosis of population responses; (*b*) mean activity sparseness plotted against mean kurtosis of population responses. Triangles indicate raw images were used to test the filter sets, crosses indicate log-transformed images and circles indicate whitened images.

sparseness, *lifetime kurtosis cannot be used as a measure of population sparseness*: the two kurtosis values measure properties of a coding scheme which are different in principle and in practice. For example, although the Olshausen–Field and independent component filters have high lifetime kurtosis, they have surprisingly low population kurtosis.

The Olshausen–Field algorithm includes a post-processing 'sparsification' stage, which increases the population kurtosis (Olshausen and Field 1997). We performed this stage on the responses of the Olshausen–Field filters, and found it had a limited effect on population sparseness (increasing the population kurtosis from 2.17 to 2.46), at the expense of lifetime

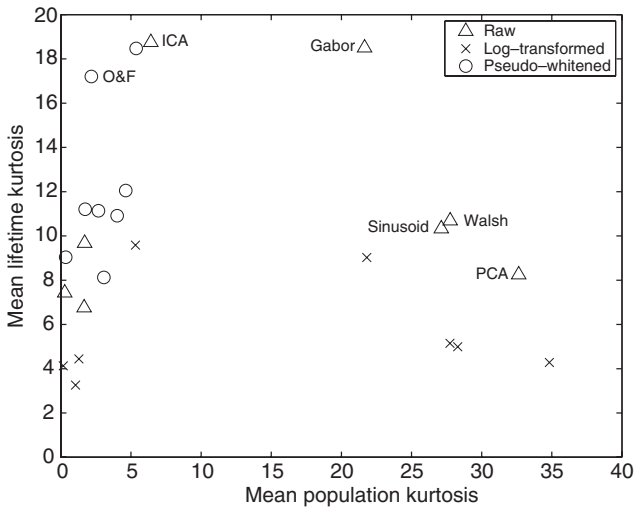

Figure 4. Scatterplot showing the relationship between the mean kurtosis of the neurons' lifetime response distribution and the mean kurtosis of the population response to each image. Triangles indicate raw images were used to test the filter sets, crosses indicate log-transformed images and circles indicate whitened images. Selected points are labelled; the remaining values can be found in table 1.

kurtosis (which decreased from 17.21 to 15.89). This post-processing could be applied equally well to any of the non-orthogonal coding models.

A striking feature of this graph is that the codes fall into two sharply defined categories. The first category contains five of the codes (Olshausen–Field bases, independent components, Gaussians, difference-of-Gaussians and single pixels) which have relatively low population kurtosis values (under 10), but have widely ranging lifetime kurtosis values (between 3 and 19). In this category, there is a positive correlation between population and lifetime kurtosis ($r = 0.71$, dof $= 14$, $P < 0.01$). The other category contains four codes (Gabor filters, principal components, sinusoids and Walsh functions), which may have fairly low lifetime kurtosis, but (for raw and log-transformed images) have substantially higher population kurtosis than all of the other codes. Indeed it is only these codes which have population kurtosis values that are 'interestingly' high—i.e. noticeably higher than the population kurtosis of the Gaussian code.

The four high-population-sparseness codes are different from the low-population-sparseness codes in one important respect—they contain filters which vary in their preferred spatial frequencies. In each of the low-population-sparseness codes (Olshausen–Field, ICs, Gaussians, DoGs and pixels), all filters have similar preferred spatial frequencies. The high-population-sparseness codes (Gabors, PCs, sinusoids and Walsh functions) all have some low-spatial-frequency filters as well as large numbers of high-spatial-frequency filters. Since natural images are known to have amplitude spectra which are approximately proportional to $1/f$ (Carlson 1978, Burton and Moorhead 1987, Field 1987, Tolhurst *et al* 1992), there is a large amount of variance in the low-frequency Fourier coefficients. Thus the low-frequency filters can be expected to have larger response magnitudes than the high-frequency filters. As a result, the few low-frequency filters often produce responses that are large compared with the responses of the many high-frequency filters, and the resulting representations often have high population sparseness.

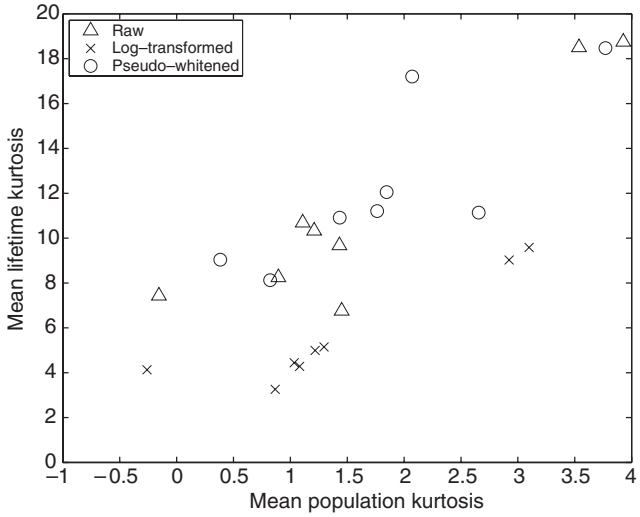

Figure 5. Scatterplot showing the relationship between the mean kurtosis of the neurons' lifetime response distribution and the mean kurtosis of the population response to each image, after each neuron's response distribution has been standardized to have a standard deviation of 1. Triangles indicate raw images were used to test the filter sets, crosses indicate log-transformed images and circles indicate whitened images.

This effect is particularly noticeable when the Gabor code is compared with the Olshausen–Field and ICA codes, since all three have similar 'receptive field' shapes. These three codes all have high lifetime kurtosis values, but only the Gabor model has a high population kurtosis. This is because the Olshausen–Field and ICA codes only contain filters whose preferred spatial frequencies are high (Olshausen and Field 1997, van Hateren and van der Schaaf 1998), and so are equally likely to produce large responses to natural images, whereas the Gabor code contains filters at a range of spatial frequencies, and so the low-frequency Gabors are more likely to respond strongly than the high-frequency Gabors.

This effect can be seen clearly by observing that the codes which produce population-sparse coding of raw (triangles) and log-transformed (crosses) images appear much less population sparse when encoding pseudo-whitened images (circles). The pseudo-whitened images have had their amplitude spectra approximately flattened, and so there is no longer a concentration of variance at low spatial frequencies. As a result, the responses of the low- and high-spatial-frequency neurons are more equally balanced, and the code appears to have lower population sparseness. This is particularly noticeable for the Olshausen–Field model, since it can only be used for pseudo-whitened images. However, whitening does not generally have a substantial effect on lifetime kurtosis.

Finally, it can be seen that log-transforming the images substantially reduces lifetime kurtosis. This is because the log transform is compressive, and reduces the pixel variance in the brighter picture fragments. The pixel variances in the darker and brighter fragments are now more similar, reducing one source of kurtosis in natural images (Baddeley 1996).

It can be seen from figure 4 that population sparseness does not only depend on how infrequently each cell responds (lifetime sparseness). Another important factor is how evenly variance is spread amongst the population of coding neurons—the codes which have many high-frequency (low-variance) units and a few low-frequency (high-variance) units have high population sparseness. The idea that coding variance should be spread evenly amongst coding

units is discussed by Field (1994), who uses the term 'distributed', and by Willmore *et al* (2000), who use the term 'dispersed'. Codes which are poorly dispersed (and so variance is concentrated in the responses of a few neurons) are likely to have very high population sparseness values, regardless of the lifetime sparseness of the coding units, whilst in well dispersed codes population sparseness is likely to be more closely related to lifetime sparseness. We have examined this effect of dispersal on population sparseness by standardizing each neuron's response distribution. The effect of this is that all neurons have the same response variance, and the effect of poor dispersal is removed. The resulting relationship between population and lifetime sparseness values for each code is plotted in figure 5. There are two effects. Firstly, there is a much clearer relationship between the two measures than for the unstandardized responses shown in figure 4. As expected, the two measures are now approximately proportional ($r = 0.76$, dof $= 14$, $P < 0.01$), although not nearly so closely related as the three population sparseness measures (figure 3). Secondly, all of the population sparseness values are very low (less than 4), indicating that the population response distributions are approximately Gaussian in shape.

4. Discussion

Lifetime kurtosis is often used as an indicator of population sparseness, yet we have found that the lifetime kurtosis and population sparseness (population kurtosis) of a number of biologically inspired coding schemes are poorly correlated. This seems counter-intuitive, since they are both intended to be measures of how infrequently strong neural responses are produced. However, this discrepancy is a result of the fact that lifetime and population kurtosis are measuring rather different properties of coding.

Kurtosis measures the shape (peakedness) and not the size (standard deviation) of the response distribution to which it is applied, and there is no reason why the shape of a cell's lifetime response distribution should be related to the shape of the instantaneous population response distribution. For example, consider a code which consists of a large number of *identical* Gabor filters. (Of course, this will be a very poor code, since it encodes only one feature of the input images, with very high redundancy.) Each neuron will have high response kurtosis (because Gabor filters form a high-lifetime-kurtosis code for natural images), but all neurons in the code will produce identical responses to each picture. As a result, the population kurtosis of the code will be extremely low, even though the lifetime kurtosis of the individual cells is high. Of course, a different Gabor code, with a variety of preferred orientations and spatial frequencies, could be constructed so that only one cell responded to any given image. This code would have similar mean lifetime kurtosis, but it would now also have high population sparseness.

4.1. Effect of variance variation on sparseness

Both population and lifetime sparseness are affected by the infrequency of neural responses. However, they are also heavily affected by two different types of inhomogeneity in the response set. It is this difference which means that the two sparseness measures rarely agree.

Baddeley (1996) has observed that there is a substantial effect of illumination variations on lifetime sparseness: if different stimuli are illuminated at different levels, then well-lit stimuli are likely to have high pixel variance, whilst poorly-lit stimuli will have low pixel variance. The result of this 'variance variation' is that the distribution of pixel values in the stimulus set is likely to have high kurtosis (since distributions of different variance have been added together), and, consequently, any filter used to analyse this stimulus set is likely to have its

lifetime response kurtosis exaggerated. Since the variance variation occurs *between* stimuli, it has no effect on the population kurtosis, meaning that the lifetime kurtosis is likely to be exaggerated relative to the population kurtosis.

There is an analogous effect which exaggerates population kurtosis. In a compact (or poorly dispersed) code, a large proportion of the coded variance is concentrated in the responses of a small number of coding units. As a result, these units will have high response variance, and so they will often produce responses which dwarf those of the rest of the coding population. The effect of this small number of large responses is that the set will have high population kurtosis values, even if the lifetime kurtosis of the individual units is low. PCA is an example of a compact code which has high population sparseness, but low lifetime sparseness.

This effect of dispersal on population sparseness can clearly be seen by comparing figure 4 with 5. In figure 4, the response variance of each coding unit is proportional to the stimulus variance it encodes, and so well-dispersed codes (such as ICA) tend to have lower kurtosis values than the poorly-dispersed (compact) codes (such as PCA). This strong effect of dispersal on population kurtosis obscures any relationship between lifetime and population sparseness. In figure 5, however, the response variances of the units have been standardized, making all of the codes artificially well dispersed. As a result, the differences due to dispersal are removed, and rough proportionality between the two measures can be seen.

4.2. Effect of variations in illumination on population sparseness

Illumination variations between images have a further effect of reducing population sparseness. Variations in illuminant level between different photographs, as well as shadows within photographs, mean that strongly illuminated image fragments will have much higher pixel variance than poorly illuminated ones. As a result, the strongly illuminated fragments will tend to excite many coding neurons, whilst the poorly illuminated fragments will excite few neurons. This introduces correlations into the responses of the coding neurons, and means that, although each neuron may have a high lifetime response kurtosis, the population kurtosis could remain low.

The models that we tested on raw or pseudo-whitened images will be susceptible to such local changes in pixel variance, whereas any model that makes a realistic attempt to mimic the visual system's coding for contrast rather than luminance will be less affected. It is significant, therefore, that we find lower lifetime kurtosis for the models studied with log-transformed images, since the log transformation is a better approximation to contrast coding than is, say, pseudo-whitening.

4.3. Effect of orthogonality

Principal components codes are orthogonal, whilst Gabor-like codes are not. As a result, each type of variability is encoded only once by a principal components code, whereas non-orthogonal codes may encode the same information repeatedly. Thus, it is likely that there will be substantial pairwise correlation between neural responses in non-orthogonal codes—i.e. two or more neurons will be active when only one is required. This effect has been observed in Gabor-like codes by Wegmann and Zetzsche (1990), and is likely to reduce the population sparseness of non-orthogonal codes when compared to orthogonal codes such as PCA (although the Gabor code we used was designed to be quasi-orthogonal, and therefore has high population sparseness). However, the Olshausen–Field (1997) 'sparsification' process slightly improves the population sparseness of their model; this process could equally be applied to the other non-orthogonal codes in our analysis.

4.4. Dispersal and completeness

We suggest that the real difference between sparse-dispersed and compact codes is not to be found in population sparseness—in any case, it is possible to add endless numbers of inactive units to any code in order to exaggerate its population sparseness. Instead, the salient difference between codes like that of Olshausen and Field and PCA codes lies in the *dispersal* of coding. In a PCA-based code, only a few cells respond to each stimulus, but it is always the same subset of cells (the first few components) that are active. In the Olshausen–Field and ICA codes, a larger number of cells may be active in response to each image but, crucially, the subset is different for each picture. The Olshausen–Field and ICA codes are highly dispersed, fitting with Field's (1994) concept of a sparse-dispersed code. This argument is presented in more detail in Willmore *et al* (2000). The question of a code's dispersal should also be considered alongside its completeness. If a code is incomplete, it will have 'spare' coding units which are likely to be active at the same time as other units; activity will be dispersed through the coding set, but in a redundant way. A proper characterization of a sparse code will require measures of completeness and dispersal as well as of lifetime and population kurtosis.

4.5. Is sparse-dispersed coding desirable?

Field (1994) suggests that population sparseness is desirable because, metabolically, large numbers of silent neurons (as found in a population-sparse code) are inexpensive, whereas small numbers of highly active neurons (as found in a compact code) are expensive. However, it has been argued by Levy and Baxter (1996) that both silent and spiking neurons are metabolically costly, and that a metabolically efficient code must balance these two costs. Laughlin and Attwell (2000) have produced estimates of the actual costs of dormancy and spiking in rat cerebral cortex neurons and they reach the surprising conclusion that maintenance of a neuron at rest absorbs a significant amount of energy. The generation of one action potential uses as much energy as maintaining a single neuron and its attendant glial cells at rest for only about 2 s.

Thus, sparse-dispersed coding may not be beneficial on energy grounds, since maintenance of many silent cells would require more energy than is used in the firing of the few active cells. However, it may have another advantage. The firing patterns of cortical neurons are notoriously variable, at least when activated by simple stimuli (Tolhurst *et al* 1983, Tolhurst 1989, Vogels *et al* 1989). In a compact code, like PCA, large amounts of input variance are encoded by small changes in the response level of a small number of cells. In the face of neural response variability, any compact code which was reliant on precise firing would be severely compromised. In contrast, in a sparse-dispersed code, cells can behave in an almost binary manner (Treves and Rolls 1991, Rolls and Tovee 1995) and still be useful and reliable encoders, since the exact details of firing rate are irrelevant. Thus sparse-dispersed coding may be a useful strategy for encoding visual information using highly noisy neurons.

Acknowledgments

BW received a studentship from the BBSRC (UK). We are very grateful to Bruno Olshausen and to Hans van Hateren for their help with the generation of basis functions. Some of this work has been published in abstract form (Willmore and Tolhurst 2000a, b).

References

Baddeley R 1996 Searching for filters with 'interesting' output distributions: An uninteresting direction to explore? *Network: Comput. Neural Syst.* **7** 409–21

Baddeley R, Abbott L F, Booth M C A, Sengpiel F, Freeman T, Wakeman E A and Rolls E T 1997 Responses of neurons in primary and inferior temporal visual cortices to natural scenes *Proc. R. Soc.* B **264** 1775–83

Barlow H B 1989 Unsupervised learning *Neural Comput.* **1** 295–311

Bell A J and Sejnowski T J 1997 The 'independent components' of natural scenes are edge filters *Vis. Res.* **37** 3327–38

Burton G J and Moorhead I R 1987 Colour and spatial structure in natural scenes *Appl. Opt.* **26** 157–70

Carlson C R 1978 Thresholds for perceived image sharpness *Photogr. Sci. Eng.* **22** 69–71

Comon P 1994 Independent components analysis: a new concept? *Signal Process.* **36** 287–314

Daugman J G 1980 Two-dimensional spectral analysis of cortical receptive field profiles *Vis. Res.* **20** 847–56

DeAngelis G C, Ohzawa I and Freeman R D 1993 Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development *J. Neurophysiol.* **69** 1091–117

Field D J 1987 Relations between the statistics of natural images and the response properties of cortical cells *J. Opt. Soc. Am.* A **4** 2379–94

——1994 What is the goal of sensory coding? *Neural Comput.* **6** 559–601

Field D J and Tolhurst D J 1986 The structure and symmetry of simple-cell receptive-field profiles in the cat's visual cortex *Proc. R. Soc.* B **228** 379–400

Fyfe C and Baddeley R 1995 Finding compact and sparse-distributed representations of visual images *Network: Comput. Neural Syst.* **6** 333–44

Hancock P J B, Baddeley R and Smith L S 1992 The principal components of natural images *Network: Comput. Neural Syst.* **3** 61–70

Harpur G F and Prager R W 1996 Development of low entropy coding in a recurrent network *Network: Comput. Neural Syst.* **7** 277–284

Heeger D J 1992 Normalization of cell responses in cat striate cortex *Vis. Neurosci.* **9** 181–97

Hinton G E, McClelland J L and Rumelhart D E 1986 Distributed Representations *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* vol 1, ed D Rumelhart *et al* (Cambridge, MA: MIT Press) pp 77–109

Hubel D H and Wiesel T N 1962 Receptive fields, binocular interations, and functional architecture in the cat's visual cortex *J. Comp. Neurol.* **158** 267–94

Hyvärinen A 1999 Fast and Robust Fixed-Point Algorithms for Independent Component Analysis *IEEE Trans. Neural Networks* **10** 626–34

Jones J P and Palmer L A 1987 An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex *J. Neurophysiol.* **58** 1187–211

Laughlin S B and Attwell D 2000 An energy budget for glutamatergic signalling in grey matter of the rat cerebral cortex *J. Physiol.* P **525** 61P

Levy W B and Baxter R A 1996 Energy efficient neural codes *Neural Comput.* **8** 531–43

Marcelja S 1980 Mathematical description of the responses of simple cortical cells *J. Opt. Soc. Am.* A **70** 1297–300

Movshon J A, Thompson I D and Tolhurst D J 1978 Spatial summation in the receptive fields of simple cells in the cat's striate cortex *J. Physiol.* **283** 53–77

Navarro R, Tabernero A and Cristobal G 1996 Image representation with Gabor wavelets and its applications *Adv. Imaging Electron Phys.* **97** 1–84

Olshausen B A and Field D J 1996 Emergence of simple-cell receptive field properties by learning a sparse code for natural images *Nature* **381** 607–9

——1997 Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vis. Res.* **37** 3311–25

Rolls E T and Tovee M J 1995 Sparsenss of the neuronal representation of stimuli in the primate temporal visual cortex *J. Neurophysiol.* **73** 713–26

Tolhurst D J 1989 The amount of information transmitted about contrast by neurones in the cat's visual cortex *Vis. Neurosci.* **2** 409–13

Tolhurst D J, Movshon J A and Dean A F 1983 The statistical reliability of signals in single neurons in cat and monkey visual cortex *Vis. Res.* **23** 775–85

Tolhurst D J, Smyth D, Baker G E and Thompson I D 1999 Variations in the sparseness of neuronal responses to natural scenes as recorded in striate cortex of anaesthetised ferrets *J. Physiol.* P **515** 103P

Tolhurst D J, Tadmor Y and Tang Chao 1992 Amplitude spectra of natural images *Ophthalmic Physiol. Opt.* **12** 229–32

Treves A and Rolls E T 1991 What determines the capacity of autoassociative memories in the brain? *Network: Comput. Neural Syst.* **2** 371–97

van Hateren J H and Ruderman D L 1998 Independent component analysis of natural image sequences yields spatio-

temporal filters similar to simple cells in primary visual cortex *Proc. R. Soc.* B **265** 2315–20

van Hateren J H and van der Schaaf A 1998 Independent component filters of natural images compared with simple cells in primary visual cortex *Proc. R. Soc.* B **265** 359–66

Vinje W E and Gallant J L 2000 Sparse coding and decorrelation in primary visual cortex during natural vision *Science* **287** 1273–6

Vogels R, Spileers W and Orban G A 1989 The response variability of striate cortical neurons in the behaving monkey *Exp. Brain Res.* **77** 432–6

Wegmann B and Zetzsche C 1990 Statistical dependence between orientation filter outputs used in an human vision based image code *Proc. SPIE* **1360** 909–23

Willmore B and Tolhurst D J 2000a A comparison of natural-image based models of simple-cell coding *Eur. J. Neurosci.* **12** 196S

——2000b Measurement of sparseness and dispersal of simple-cell codes *J. Physiol.* P **526** 168P

Willmore B, Watters P A and Tolhurst D J 2000 Sparseness and kurtosis of computational models of simple-cell coding in primary visual cortex *Perception* **29** 1017–40