

# Distribution of independent components of binocular natural images.

David W Hunter<sup>1</sup> Paul B Hibbard<sup>1,2</sup>

1. School of Psychology and Neuroscience, University of St Andrews

2. Department of Psychology, University of Essex

## Abstract

An influential theory of the function of early processing in the visual cortex is that it forms an efficient coding of ecologically valid stimuli. In particular, correlations and differences between visual signals from the two eyes are believed to be of great importance in solving both depth from disparity and binocular fusion. Techniques such as Independent Components Analysis have been developed to learn efficient codings from natural images; these codings have been found to resemble receptive fields of simple-cells in V1. However the extent to which this approach provides an explanation of the functionality of the visual cortex is still an open question. When binocular ICA components were compared with physiological measurements we found a broad range of similarities together with a number of key differences. In common with physiological measurements we found components with a broad range of both phase and position disparity tuning. However we have also found a larger population of binocularly anti-correlated components than has been found physiologically. We found components focused narrowly on detecting disparities proportional to half-integer multiples of wavelength rather than the range of disparities found physiologically. We present the results as a detailed analysis of phase and position disparities in Gabor-like components generated by Independent Components Analysis trained on binocular natural images and compare these results to physiology. We find strong similarities between components learned from natural images that indicate that ecologically valid stimuli are important in understanding cortical function, but with significant differences that suggest that our current models are incomplete.

Keywords: Binocular vision; Binocular disparity; Natural image statistics; Independent component analysis.

## Introduction

A perception of depth is necessary for humans and other animals to understand, interact with and navigate around our environment. One of our principal sources of depth information is binocular vision, such that the small differences between the images formed in our two eyes are used to infer the three-dimensional structure of the environment. Although the responses of some neurons in the LGN are affected by the images presented to both eyes (Tong, Guido, Tumosa, Spear and Heidenreich, 1992) it is widely held that the first stages of disparity processing occur in the primary visual cortex (V1) (Cumming and DeAngelis, 2001; Parker, 2007; Roe, Parker, Born and DeAngelis, 2007). Here, cells are found that are tuned to binocular disparity. The disparity sensitivity of these neurons is well characterised by the binocular energy model (Fleet, Wagner and Heeger, 1996; Ohzawa, Deangelis, and Freeman, 1997; Prince, Pointon, Cumming and Parker, 2002; Read and Cumming, 2003). Under this model, the responses of linear Gabor filters are summed between the

two eyes, then squared to produce an energy response. Disparity tuning is introduced by summing across filters with different shapes and/or locations of receptive fields for the two eyes.

Physiological studies of the visual cortex have found widespread evidence of cells that respond to binocular stimuli, see (Neri, 2005; Parker, 2007; Roe et al., 2007) and (Howard, 2002; Howard and Rogers, 2002) for detailed reviews. Hubel and Wiesel (1962) studied V1 in cats and found cells that had Gabor-like receptive fields in both eyes. Similar cells have been found in macaque monkeys (Poggio and Fischer, 1977; Prince, Cumming and Parker, 2002; Prince, Pointon, et al., 2002), sheep (Clarke, Donaldson and Whitteridge, 1976) and the visual Wulst of the barn owl (Pettigrew and Konishi, 1976). Numerous studies have characterised these cells in increasing detail, both in terms of the responses of individual cells, and their distributions.

Differences in the location and/or phase of the receptive fields of simple cells between the two eyes are common, and are generally thought to form the basis for disparity estimation. Position-disparity tuned cells have the same shape of receptive field, but with a shift in location, in the two eyes. Evidence for position-disparity tuned cells has for example been found in V1 of the cat (Anzai, Ohzawa and Freeman, 1999; Nikara, Bishop and Pettigrew, 1968; Pettigrew, 1972). Conversely, phase-disparity tuned cells have receptive fields with an identical location, but a difference in their shape, between the two eyes. Specifically the wave-function of the Gabor-like receptive field is shifted in phase in one eye compared to the other. Such cells have also been found in V1 of the cat by Ohzawa and colleagues (DeAngelis, Ohzawa and Freeman, 1991, 1995; Ohzawa, DeAngelis, and Freeman, 1990). Subsequently researchers found evidence that disparity tuned V1 cells in both cats and macaques exhibit a mixture of phase and position disparity sensitivity (Anzai et al., 1999; Prince, Cumming, et al., 2002; Prince, Pointon, et al., 2002; Tsao, Conway and Livingstone, 2003). It is generally believed that the outputs of these simple cells feed into V1 complex cells according to some variant on the energy model (Fleet et al., 1996; Ohzawa et al., 1997; Prince, Pointon, et al., 2002; Read and Cumming, 2003). Together, these cells are assumed to form the basis for the estimation of differences in the locations of corresponding features across the two eyes, and thus the perception of depth from binocular disparity.

Since binocular cells in area V1 are tuned for orientation and spatial frequency, it is also possible that their tuning for these properties might differ between the two eyes. Indeed, differences in the preferred orientation between eyes have been found for neurons in the macaque cortex, and areas 17 and 21a of the cat (Blakemore, Fiorentini and Maffei, 1972; Bridge and Cumming, 2001; Nelson, Kato, and Bishop, 1977; Wieniawa-Narkiewicz, Wimbome, Michalski and Henry, 1992). These interocular differences in the orientation tuning of receptive fields are potentially valuable in the encoding of surface orientation. Orientation disparities, defined as differences in the orientation of corresponding features in the two eyes' images, are created when surfaces are slanted away from frontoparallel. These orientation disparities could therefore be used to determine the orientation of surfaces (Greenwald and Knill, 2009). Although Bridge and colleagues (Bridge and Cumming, 2001; Bridge, Cumming and Parker, 2001) argued that the type of response to orientation found in binocular V1 cells is not well-suited to the analysis of orientation disparities, psychophysical evidence suggests that they could contribute directly to the perception of depth (Heydt, Adorjani, Hännny and Baumgartner, 1978; Ninio, 1985). Differences in the size of the corresponding features in the two eyes' images could play a similar role (Tyler and Sutter, 1979).

While electrophysiological studies have demonstrated a wide variety of disparity tuning in cortical neurons, they do not directly provide any understanding of why disparity is encoded in this way. Significant understanding of the nature of the computations performed by the visual system can be gained from the analysis of typical natural images. The vast majority of natural image inputs to the visual system are redundant; a recent estimate by Field and Chandler (2012) put this redundancy at 64% in local regions of monocular natural images. This redundancy includes both noise in the imaging system and dependencies within the data. In particular, natural images exhibit significant

spatial redundancy; the intensities of neighbouring sample locations are not independently distributed. Analysis of these relationships is an important factor in understanding local image structure, and how this can be efficiently encoded.

Barlow (1961) proposed that neurons perform an energy-efficient coding of the visual input by removing this redundancy from the signal. Numerous techniques have been applied to samples from natural images to examine ways in which information can be represented efficiently. For example, Olshausen and Field created a sparse linear decomposition of image patches by applying a Cauchy prior to favour coefficients (responses) with low values (Olshausen and Field, 1996). This resulted in a set of 'edge-like' basis functions that were spatially localised, oriented and bandpass. These Gabor-like functions show many similarities in their overall structure to the receptive field structure of V1 neurons. Similar results have been found by minimising mutual information between filter outputs (Bell and Sejnowski, 1997), and maximising the kurtosis of the population response (Hyvärinen, Hurri and Hoyer, 2009). A detailed analysis of the similarities of these learned filters with monocular V1 simple cell responses was carried out by van Hateren and van der Schaaf (1998). Again, clear similarities were noted between the distributions of frequency tuning, orientation and receptive field size and measurements of these properties in V1 simple cells in the macaque. These results have been taken as evidence that the visual cortex achieves an efficient coding of visual input using components that are independent, exhibit a sparse response to inputs and capture non-Gaussian spatial relations in the data. Ringach (2002) compared the RF sizes of ICA components and components generated using sparse coding (Olshausen 2002) to those measured in cats and monkeys. He found that in general both sparse and ICA components had larger, more narrowly tuned, features than those observed in physiological measures. Although the image statistics did not successfully describe the RF of simple cells in V1 he did not conclude that the basic principles were flawed as a way understanding the function of cortical cells.

Methods for the efficient or optimal encoding of information have also been proposed in order to understand the responses of binocular neurons. A range of approaches to this problem have been taken, in each case with a different optimisation goal in mind. Li and Atick (1994a) proposed that the brain encodes information in a way that reduces the redundancy present in natural binocular images. Clearly, the images formed in our left and right eyes are very similar. This similarity is a significant source of redundancy in the visual information that we receive. Li and Atick proposed that binocular information is encoded in a way that de-correlates the two eyes' images. This is achieved by creating two channels, one that additively combines the left and right eyes' views, and one that subtracts them (Li and Atick, 1994a). This two-channel structure is supported by psychophysical evidence (Chen and Li, 1998; May, Zhaoping and Hibbard, 2012). Li and Atick (1994a) combined binocular decorrelation with whitening of the image as a way of encoding binocular information. The resulting binocular filters exhibited disparity tuning that showed a number of similarities to that found in cortical cells.

Burge and Geisler (2014) derived binocular filters that were optimised for estimating disparity in natural images. This was achieved using the accuracy maximisation analysis proposed by Geisler, Najemnik and Ing (2009). Optimal binocular filters were learned for planar samples, with known disparities, created from natural images. Again, the binocular receptive fields of the learned filters were similar in shape to those found in the visual cortex.

Independent Component Analysis has also been applied to binocular images. Here, the aim is to maximise the independence of the learned components, and this is attempted through the maximisation of the kurtosis of responses. A sparse coding of binocular image patches was created by Hoyer and Hyvärinen using Independent Component Analysis (ICA) (Hoyer and Hyvärinen, 2000). This analysis was performed on patches taken from locations around a simulated convergence point in the left and right views. As a result, their analysis focused on features that were close to alignment.

The components learned closely resembled Gabor functions for each eye's view, with a similar orientation and frequency for each eye. As such, they were very similar in form to the receptive fields

of binocular neurons in V1 (Anzai et al., 1999). A disparity tuning function for each component was calculated. Following Poggio and Fischer (1977) these were then classified as tuned-excitatory, tuned-inhibitory, near or far cells. Tuned excitatory cells are those that show a clear response peak at zero disparity. Conversely, tuned inhibitory cells show a clear trough in their response at zero disparity. Near and far cells show a peak in response for near or far disparities, respectively. Hoyer and Hyvärinen found that the majority of their cells were tuned excitatory, near or far cells, with a phase disparity close to 0 (Hoyer and Hyvärinen, 2000). This feature of the disparity tuning is again remarkably similar to that found in cortical cells. For example, Prince, Cumming et al. (2002) showed that the distribution of phase-disparity tuning for V1 cells showed a very clear peak at 0, with a fall-off in the number of cells tuned to larger phase differences. Okajima (2004) performed a similar study using DoG filtered Gaussian noise and natural images using synthetic horizontal displacement. They generated Gabor-like components by minimising mutual information. Like Hoyer and Hyvärinen they found components with similar frequency and orientation. Their analysis identified components exhibiting phase disparity, position disparity and both.

Taken as a whole, these results show that ICA applied to natural binocular images generates components with a number of close similarities to binocular neurons. These results suggest that the responses of binocular neurons to natural images will in turn be sparse and independent. The purpose of the current study is to extend this approach, in order to provide the first detailed qualitative comparison between the components learned by ICA, and the responses of binocular neurons.

Our fundamental approach was similar to that of Hoyer and Hyvärinen (2000), in that we performed Independent Component Analysis on patches cut from corresponding locations in the left and right images of binocular pairs. We adapted and expanded on their approach in a number of important ways.

Firstly, Hoyer and Hyvärinen (2000) assessed the disparity tuning of their learned components, but did not attempt to quantitatively model their receptive fields. While the components analysed were described as Gabor like, they were not fitted with Gabor or other functions. In their study, disparity was quantified by measuring the responses of the component to stimuli that consisted of the components themselves, presented with a range of positional disparities. The shapes and relative locations of the components for each eye's input were not assessed. We modelled the components as Gabor filters, in order to allow us to make direct comparisons with physiological data.

The second difference is that this then allows for a more fine-grained assessment of the components' disparity tuning. In particular, modelling binocular receptive fields allows us to directly assess the position- and phase-disparity tuning of each component. This goes beyond the categorisation of tuning functions into tuned-excitatory, tuned-inhibitory, near and far cells. It should be noted that, in the alert rhesus monkey, tuned excitatory cells are also found that show clear tuning to either a crossed or uncrossed disparity (Poggio, 1991; Poggio and Fischer, 1977; Poggio, Gonzalez and Krause, 1988; Poggio, Motter, Squatrito and Trotter, 1985; Poggio and Talbot, 1981), and that cells are better viewed as forming a continuum of tuning characteristics, rather than falling into these discrete categories (DeAngelis et al., 1991; Freeman and Ohzawa, 1990; LeVay and Voigt, 1988; Ohzawa, DeAngelis and Freeman, 1996). Our analysis allows a detailed assessment of the relationship between the tuning to positional and phase disparities, as has been performed for physiological data (Prince, Cumming, et al., 2002; Prince, Pointon, et al., 2002).

The third difference is that, as well as assessing tuning for horizontal disparity, we can also assess the tuning for vertical disparity, and for disparities in orientation and spatial frequency.

Finally, our study also differed in the way that we sampled from binocular images. Hoyer and Hyvärinen took their samples from an area around simulated fixation points, which were chosen to make the samples from the two eyes relatively similar. This is an important consideration with binocular images, as their statistical properties are spatially non-stationary (Hibbard, 2007, 2008; Liu,

Bovik and Cormack, 2008). Since we tend to fixate the same point with each eye, the disparity in the centre of the image is expected to be close to zero. As we move away from this point, the range of expected disparities will increase. The dependence of disparity range on eccentricity is reflected in the tuning of the visual system to disparity, as measured in physiological (DeAngelis and Uka, 2003; Durand, Zhu, Celebrini and Trotter, 2002; Prince, Pointon, et al., 2002) and psychophysical (Hampton and Kertesz, 1983; Qin, Takamatsu and Nakashima, 2006) studies. The consequence for ICA is that, if we sample from the same image location in each eye, the similarity between the two samples will decrease as we move from the centre of the image to more peripheral locations. This will mean that the left and right halves of the components will be expected to be more similar for central locations than for other locations. In the current study we broadened the sample range to 20° (square) using binocular images taken with calibrated cameras (Hibbard, 2008). Our samples are thus more representative of binocular images in general, rather than the special case of samples relatively close to fixation.

The overall aims of the current study were to perform a detailed analysis of the results of ICA applied to natural binocular images; to provide a comparison between the components learned by ICA, and the responses of binocular neurons; and to determine the extent to which this approach can provide an explanation of the function of binocular simple cells in the visual cortex.

## Methods

Following the methods of Hoyer and Hyvärinen, we processed the images in four stages. Patch pairs were cut from an image set and normalised, followed by a whitening stage and finally the computation of the independent components. In the next section we describe this method and its reasoning. This method is also similar to that used on synthetic patches (Okajima, 2004).

### *The data set*

The methods for capturing and processing the binocular images are described in Hibbard (2008). These methods will have significant effects on the statistics of binocular images. For example, the convergence of the cameras determines how the disparity statistics will vary as a function of the image location. The details of the image capture process (Hibbard, 2008) are therefore repeated here. Images were captured using two Nikon Coolpix 4500 digital cameras, harnessed in a purpose-built mount that allowed the inter-camera separation, and the orientation of each camera about a vertical axis, to be manipulated. This is a simplification of the situation for human binocular vision, in which there are potentially three degrees of freedom for each eye (rotations about horizontal and vertical axes, and the line of sight). The analyses presented here focus on situations in which vergence is approximately symmetrical and elevation is close to zero. In this case, the expected cyclovergence, which is not possible in the camera setup used, is negligible (e.g. Porrill, Ivins and Frisby, 1999). In all cases, an inter-camera separation of 65 mm (representative of the typical human interocular separation) was used. The cameras were oriented so that the same point in the scene projected to the centre of each camera's image, so as to mimic the typical human fixation strategy.

Two classes of scene were investigated. In the first, images were collections of natural objects (fruit, vegetables, stones, shells, plants) arranged in "still-life" collections. These were displayed in a Verivide light cabinet, with D65 illumination, and were viewed from a distance of less than 1m in all cases. The second collection was of outdoor scenes, taken in the quad of St Mary's college in St Andrews (to include trees, flowers, lawns) or on the beach (to include the beach, rocks). Since the cameras were fixated on a target object in each image pair, and a range of distances was sampled, the images contain a range of convergence distances, from approximately 50 cm to tens of metres.

Images were captured at a resolution of 1600 x 1200 pixels. They were then calibrated to take account of the characteristics of the cameras. Firstly, images were calibrated using a camera

calibration toolkit that is available online<sup>1</sup>. This allowed us to correct for lens distortions, calculate the effective focal lengths of the cameras, and to transform the images into a ‘pinhole-camera’ model. That is, the spatial location of each pixel in the image is described in terms of the visual direction through the centre of the lens that will project onto that pixel. The final resolution of the images was 1 pixel per arc minute of visual angle. The images were also calibrated to take account of the colour characteristics of the cameras, by capturing colour patches from a Macbeth Colorchecker DC chart, and using these to map RGB camera values to CIE LAB values (Hong, Luo and Rhodes, 2001). Subsequent analyses were performed on the luminance information only.

The images were resized using a bicubic interpolation function, as implemented by MATLAB’s `imresize` command, such that each pixel corresponded to 4 arc min of visual angle; the effects of this rescaling were subsequently assessed in detail, as discussed below. From the left images a set of 100,000 25 by 25 pixel image patches were cut from random locations. Another set of image patches were cut from identical locations in the corresponding right images. That is, the samples from the left and right eyes came from the same position in the image, with the same pixel coordinates, rather than from locations necessarily corresponding to the same physical structure in the scene. The input samples to the ICA analysis were created by concatenating the samples from the two images. Hoyer and Hyvärinen sampled from a 300x300 pixel region around a simulated vergence point, arguing that this matched the converged and focused configuration of typical viewing. As the statistics of binocular disparity (Hibbard, 2007, 2008; Liu, Bovik and Cormack, 2008), and the disparity tuning of cortical neurons (DeAngelis and Uka, 2003; Durand et al., 2002; Prince, Pointon, et al., 2002) vary with the position in the image relative to the fixation point, it is important to also consider points away from fixation. Therefore we sampled uniformly from the whole image. As a control condition, to test the extent to which the binocular properties of the components were determined by the binocular redundancy in the images, ICA was also computed on samples drawn separately, from unrelated positions, in the two images.

### **Normalisation and Gain**

Rather than transmit absolute intensities across the optic nerves, retinal ganglion cells encode and transmit local changes in intensity (Laughlin, 1981; Srinivasan, Laughlin and Dubs, 1982). We modelled this by subtracting from each patch its mean intensity (Hyvärinen, 1999; Okajima, 2004; Olshausen and Field, 1996).

We call the  $i^{\text{th}}$  combined, vectorised image patch  $x_i = \{x_{l,i}, x_{r,i}\}$ , where  $x_{l,i}$  denotes the  $i^{\text{th}}$  left patch and  $x_{r,i}$  the  $i^{\text{th}}$  right patch. The luminance-centred left image patch is:

$$\bar{x}_{l,i} = x_{l,i} - \langle x_{l,i} \rangle \quad (1)$$

where  $\langle x \rangle$  denotes the mean of  $x$ . The same method was applied to the right view patches.

By centring the patch we removed the effects of local illumination on a scale roughly the size of the patch. In order to normalise the contrast and to remove illumination differences between the two views we normalised both left and right patches separately, by dividing each vector by its norm ( $\| \cdot \|$ ):

---

<sup>1</sup> [http://www.vision.caltech.edu/bouguetj/calib\\_doc/](http://www.vision.caltech.edu/bouguetj/calib_doc/).

$$\widehat{\mathbf{x}}_{l,i} = \frac{\overline{\mathbf{x}}_{l,i}}{\|\overline{\mathbf{x}}_{l,i}\|}, \widehat{\mathbf{x}}_{r,i} = \frac{\overline{\mathbf{x}}_{r,i}}{\|\overline{\mathbf{x}}_{r,i}\|} \quad (2)$$

This represents a linear approximation of the logarithmic functions believed to occur in early vision as measured by Tong et al. (1992) and developed from psychophysical experimentation by Ding and Sperling (2006). Ding and Sperling's model normalises input in a non-linear fashion using ratios of the left/right intensity at particular phases and frequencies. As we are interested in the phase and frequency distributions of binocular images in the current study, we have preferred whole patch normalisation in order to avoid altering those distributions at this stage. Equation (2) has the effect of making each patch vector into a unit vector, thus equalising patch contrast, across both views and between patch locations. Using this equation, contrast ratios are constrained to be 1 whereas the method of Ding and Sperling allows for all possibilities between one-eye dominance to a unitary left/right ratio. This approximation allows us to study the special case of equal left/right intensity. Adding an extra dimension such as intensity ratio would have increased both the complexity of the analysis and the number of components needed to generate accurate distributions.

Most models of binocular neurons assume a linear summation of inputs followed by a non-linear post-processing step (DeAngelis et al., 1995; Fleet et al., 1996; Hibbard, 2008; Hyvärinen, 1999; Okajima, 2004) consistent with physiological measurements (Anzai et al., 1999; Prince, Cumming, et al., 2002; Prince, Pointon, et al., 2002). The left and right components of the patch were then simply concatenated to form the binocular sample:

$$\mathbf{x}_i = [\widehat{\mathbf{x}}_{l,i}, \widehat{\mathbf{x}}_{r,i}] \quad (3)$$

The assumptions of the FastICA algorithm (Hyvärinen, 1999) require each sample to be of unit length, so we further normalised the concatenated vectors. As these vectors were already of unit length, this amounted to a division by  $\sqrt{2}$ , making no further adjustments to the relative strengths of the left or right signals.

### **Whitening**

Whitening with Principal Components Analysis is an important pre-processing stage in ICA. It removes linear correlations from the data, and in the case of natural-image statistics acts as a low-pass filter (Hyvärinen, 1999). The patch itself acts as a windowing function and a high-pass filter; combined with whitening this acts as a band-pass filter. The normal power spectrum of an image follows a  $\frac{1}{f^\alpha}$  curve (where  $f$  is the frequency and  $\alpha$  is a constant, normally  $\alpha \approx 2$ ), whitening acts to flatten this curve, normalising the responses at each frequency. As the signal strength is modulated by  $\frac{1}{|f^\alpha|}$  and noise strength is uniform in the frequency domain (assuming Gaussian white noise), noise dominates the signal at high frequencies (Atick & Redlich, 1992). We performed a low-pass filtering to remove the higher noise-dominated frequencies, by truncating the PCA model. A similar role in noise reduction at high frequencies has been proposed for the retina (Atick & Redlich, 1992). We found that 200 eigenvectors generated by PCA on the image patches explained a mean of 86.7% of the variance in the image patches. ICA is performed in the  $\mathfrak{R}^{200}$  space generated by the eigenvectors. Components were converted back to the image space by applying the inverse of the whitening matrix.

### **Independent Component Analysis**

Independent Component Analysis attempts to find a linear decomposition of the data such that each component is maximally independent. We used the FastICA method of Hyvärinen as it is rapid and generally converges. The algorithm uses gradient descent in an attempt to find the components with the minimum mutual information, using kurtosis as a measure of independence. It should be noted that a general limitation of this approach is that it is not guaranteed to minimise dependencies. This

will only be achieved if the samples are a linear superposition in independent, heavy-tailed sources. We used a hyperbolic tan as the derivative of the non-linearity and initialised the weights from a random Gaussian distribution. The number of components generated by the algorithm is restricted by the number of eigenvectors generated by the whitening stage. We used 200 eigenvectors from the whitening stage, the same number of independent components were then generated by the FastICA algorithm. The components were converted from the PCA generated eigenvector space back into image space by applying the inverse of the whitening matrix. In total 100,000 binocular image patches were used to train the ICA model cut from uniformly random locations from 139 left/right image pairs. In order to calculate accurate distributions of the components we repeated the component generation 200 times using different patch sets (from the same images) each time, thus producing 40000 components from which reasonably accurate histogram distributions could be calculated.

### ***The meaning of the ICA components***

It should be noted that ICA suffers from some limitations. Firstly, as PCA is frequency bandwidth limited, the resulting components are also frequency bandwidth limited. This limits the range of frequencies that can be observed and results in components that are more narrowly tuned in frequency than might otherwise be observed (Ringach, Sapiro, & Shapley, 1997). As a consequence of the oriented edge-like structure of the components we also expect them to be narrowly tuned in orientation. While the frequency content of the components will be restricted, there is no reason to believe that the phase will be substantially altered as PCA is not limited in phase. Secondly unlike PCA, the components are not ordered. Thirdly, there is no test for statistical significance. Fourthly, the algorithm assumes both noiseless input, and no output noise. Finally, the algorithm is not particularly robust to initial starting conditions (Hyvärinen, 2011). A consequence of these issues is that it is not trivial to determine the explanatory power of a given component, in the same way that we can for a PCA component. Given the sparse nature of the algorithm, it is conceivable that a component will explain only a small amount of data peculiar to a particular sample set, rather than being a useful descriptor of the general population. This is especially true if noise is present in the input (Hyvärinen, Sarela, & Vigário, 1999). An indirect method to determine the explanatory power is to generate numerous ICA models and check for recurring similar components (Himberg, Hyvärinen and Esposito, 2004). As our analysis looks at overall trends in the calculated components rather than analysing individual components in detail, we did not attempt to validate individual components. Instead we calculated 200 separate ICA models using different image patches with 200 components in each model, thus generating 40,000 components in total. Highly similar and thus significant components will form clusters in this analysis. Non-significant components will appear as outliers. Not all the potential independent components will be found, since ICA is restricted to an arbitrary number of components by the Principal Components Analysis pre-processing stage, in our case 200.

The distributions of components can be argued to show the relative abundance of a particular independent feature in the binocular image data. High prevalence components will form clusters, and can thus be thought of as more prevalent in the data, than low prevalence components. It should be noted that, while repetition of the ICA computation will allow the significance of highly abundant components to be qualitatively determined, it will not readily lead to the discovery of less abundant components, as the highly abundant components will simply be recalculated in each iteration. It should also be further noted that the components generated by recalculation are not independent or orthogonal, as the independence constraint is only applied between components generated by a single FastICA computation. However, general trends in the data should be captured.

### **Fitting Gabor Functions.**

The methods of (Okajima, 2004; Prince, Cumming, et al., 2002; Prince, Pointon, et al., 2002) were followed to fit Gabor functions to each of the components. We fitted Gabor functions separately to each of the left and right view parts of each component by minimising the  $L^2$ -norm between the function and the component. The 2D Gabor function is defined as:

$$g(x, y; \theta, f, \phi, \sigma_w, \sigma_h, \psi) = w(x, y, \sigma_w, \sigma_h, \psi) c(x, y, f, \phi, \theta) = \exp\left(-\frac{\hat{x}^2}{2\sigma_w^2} - \frac{\hat{y}^2}{2\sigma_h^2}\right) \cos(2\pi\hat{x}f + \phi) \quad (4)$$

$$\hat{x} = x \cos \psi - y \sin \psi$$

$$\hat{y} = x \sin \psi + y \cos \psi$$

$$\hat{x} = x \cos \theta - y \sin \theta$$

The Gabor function consists of two components: a wave-generating function ( $c(x, y, f, \phi, \theta)$ ) which is constrained in image space by a windowing function ( $w(x, y, \sigma_w, \sigma_h, \psi)$ ). The wave-generating function describes a cosine pattern with frequency  $f$  and phase  $\phi$ , this pattern is rotated about the origin by an angle  $\theta$ . The windowing function constrains the image space span of the wave-generating function to a Gaussian window of width  $\sigma_w$  and height  $\sigma_h$ ; this window is rotated independently of the wave function by  $\psi$ . Previous authors have fixed  $\theta = \psi$ , such that the windowing function rotates with the wave-generating function (Prince, Cumming, & Parker, 2002). However, we have removed this constraint to allow the Gabor fitting function to describe a greater range and variety of Gabor-like components. All the Gabor functions are centred at zero and generated over a two dimensional image  $x = [-n/2 : n/2]$  and  $y = [-n/2 : n/2]$  where  $n$  is the size of the component patch.

In order to match the Gabor function to our component patches we must add horizontal and vertical displacement terms  $h$  and  $v$ . Our equation becomes:

$$G(x, y; h, v, \theta, f, \phi, \sigma_w, \sigma_h, \psi, s) = g(x - h, y - v; \theta, f, \phi, \sigma_w, \sigma_h, \psi) s \quad (5)$$

Where  $s$  is a scaling parameter that models the amplitude of the Gabor function. The parameters of the model were fitted to the data using the Nelder-Mead simplex method (Nelder and Mead, 1965) initialised with a Genetic Algorithm, using MATLAB's implementation.

### **Gabor symmetry**

As a result of numerous symmetries linking the parameters of the 2D Gabor function (equation 4) they are only independent within a particular range.

Rotation of the wave function of the Gabor ( $\theta$ ) by  $\pi$  radians is equivalent to reflecting the phase ( $\phi$ ) about 0 i.e.  $G(\dots; \dots, \theta, \dots, \phi, \dots) = G(\dots; \dots, \theta - \pi, \dots, -\phi, \dots)$ . The other parameters have been omitted here for clarity.

### **Interaction of Phase and Position shifts**

Changes in the spatial location of stimuli can be encoded by Gabor functions by two methods, a phase shift and a position shift. The phase shift is encoded by varying the phase parameter  $\phi$ , the position shift by varying  $v$  and  $h$  in the direction parallel to the wave generating function. Phase shifts can be converted to position shifts and visa-versa by

$$|v \cos \theta - h \sin \theta| = \frac{\phi}{2\pi f} \quad (6)$$

in the range  $-\pi < \phi < \pi$ .  $|\cdot|$  is the magnitude of the vector and the cos and sin terms rotate the shifts into the orientation of the wave generating function. Equation 6 is derived from the well-known Fourier

shift theory. While equation 6 maps phase and position in the wave-function ( $c$ ) phase and position shifted Gabor functions differ in terms of the windowing function ( $w$ ). For example, an even phase Gabor phase shifted by  $\frac{\pi}{2}$  radians will become odd, an even phase Gabor function shifted in position by an amount equivalent to  $\frac{\pi}{2}$  radians (by equation 6) will still be even phase.

## Results

Figure 1 shows the results of applying the FastICA algorithm to 100,000 binocular image patches taken from 139 image pairs. 200 components are generated in each batch, all 200 are shown. Across the 200 runs, the PCA whitening process explained on average 86.7% of variance in the image patches. Each patch shows the concatenated left/right parts of the components. Pairs of Gabor-like components are clearly visible, exhibiting a wide range of orientations, frequencies and locations.

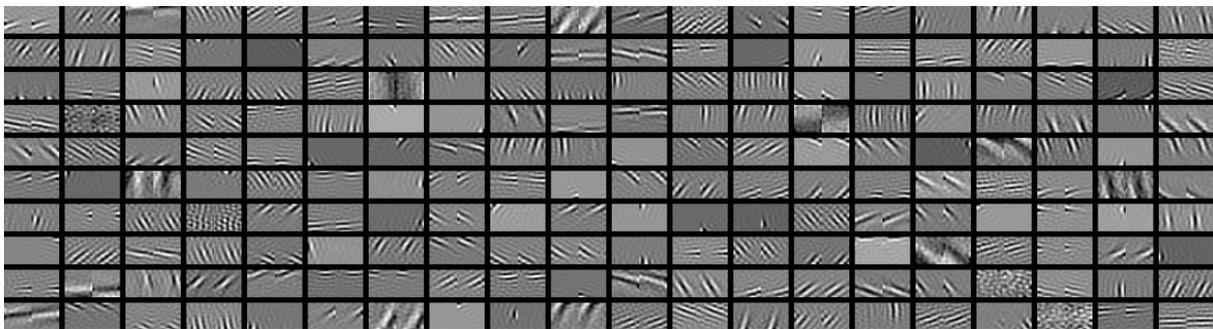


Figure 1. Example of components generated by ICA. 200 components generated by a single batch are displayed. The left half of each component corresponds to the left view, the right half to the right view. The Gabor-like components are clearly visible in most components.

## Accuracy of Fitting

The process of generating the ICA components contains many elements that depend on random processes. The patches are sampled from uniformly distributed random locations and the ICA algorithm iterates from a normally distributed random initial state. The accuracy of the fitting of the Gabor functions can be assessed in terms of both their trueness (lack of bias) and their precision. As the distributions we are interested in are not *a priori* known we cannot directly assess their trueness, we can however estimate their precision using a bootstrapping technique. This will illustrate the range of distributions produced by this method and allows us to estimate confidence intervals for the distributions. Unlike the overall process, the trueness of the Gabor-fitting sub-process can be evaluated by comparison with synthetic image patches showing Gabor functions with known ground-truth values. If these are sufficiently accurate, we can have some confidence in the accuracy of the values of Gabor functions fitted to components generated by the ICA. We will first describe the process of assessing the trueness of the Gabor-fitting function, then the bootstrapping we used to generate CI for the overall process.

In order to assess the trueness we generated 400 Gabor functions by sampling their parameters from a uniform distribution. The range of the distributions was determined either by the constraints and symmetries of the Gabor function (window and wave-function orientation), the patch sizes and resolution (horizontal and vertical position, intensity scaling and frequency maxima), or from the range of parameters produced by fitting Gabor components on the ICA components (window size and frequency minima). The Gabor functions generated by this method were rendered in 400 25 by 25 pixel square image patches and supplied as input to the fitting process described above in the same manner as a set of ICA components. The parameters of the fitted components were compared to the parameters of the Gabor functions that generated the image patch. Results of this comparison can be

seen in table 1. The distributions of the errors are highly non-normal with most of the parameter errors close to zero (see Median Absolute Deviation column). However, a minority exhibit large outliers that drive larger mean squared errors. In most cases both the MAD and the mean squared errors are less than the unit of measurement (e.g. pixels). The only exception is the window sizes, which have mean squared error of 6.4 and 6.25 pixels<sup>2</sup>. Even here, half of all errors are below 0.20 and 0.23 pixels. For the measures we are principally interested in, i.e. phase, frequency, orientation and position, the error values are extremely low and less than the sampling rate of the image (i.e. less than one pixel), although the much larger MSE values indicate the presence of large outliers. We conclude that the fitting method produces a generally accurate reflection of the true value of the underlying Gabor functions.

Parameter	Unit	Range [min,max]	Median Absolute Deviation	Mean Squared Error	
Horizontal centre $h$	Gabor	Pixels	[0,25*]	0.0014	0.1052
Vertical centre $v$	Gabor	Pixels	[0,25*]	0.0013	0.1512
Window width ( $\sigma_w$ )		Pixels	[4,12.5]†	0.2014	6.4004
Window height ( $\sigma_h$ )		Pixels	[4,12.5] †	0.2271	6.2531
Window orientation ( $\psi$ )		Radians	[0, $\pi$ ]	<0.0001	0.2658
Wave function frequency ( $f$ )		Cycles per pixel	[0.1†,0.5‡]	<0.0001	0.0006
Wave function phase ( $\phi$ )		Radians	$[-\pi, \pi]$	0.0012	0.1083
Wave function orientation ( $\theta$ )		Radians	[0, $\pi$ ]	<0.0001	0.2658
Scaling Factor ( $s$ )		Intensity	[0.1 1.0]†	<0.0001	0.2221

Table 1 Estimates of the accuracy of fitting of Gabor functions. Each parameter of the randomly generated Gabor function is sampled from a uniform distribution of the ranges shown. Unless marked the ranges chosen are constraints of the Gabor functions. \* 25 is the size of the image patch in pixels, † at least 95% of fitted Gabors generated from the ICA model are between these values. ‡ - the Nyquist limit, wavelengths shorter than this cannot be detected.

## Log Gabor functions

Although most studies of V1 (Prince, Cumming et al., 2002) have fitted Gabor functions to response data, it has been suggested that Log Gabor functions are a more accurate fit to observed data (Field 1987). In the context of binocular stereopsis, it has also been shown that an energy model based on Log-Gabor functions can be developed, and would lead to more accurate estimation of binocular disparity (Faria, Batista and Araújo, 2013). Unlike the standard Gabor function, Log-Gabor functions are defined in Fourier frequency space rather than the image space. Log-Gabor functions were defined in the polar Fourier domain as:

$$LG(f, \theta | \theta_0, \theta_\sigma, f_0, f_\sigma) = e^{-\frac{(\theta - \theta_0)^2}{2\theta_\sigma^2}} e^{\frac{-(\ln \frac{f}{f_0})^2}{2(\ln \frac{f_\sigma}{f_0})^2}} \quad (7)$$

where  $f$  is the radius (frequency) and  $\theta$  is the angle of the polar coordinates.  $f_0$  is the principal frequency and  $f_\sigma$  the bandwidth of the frequency component (Fischer et al. 2007).  $\theta_0$  is the principal orientation,  $\theta_\sigma$  is the orientation bandwidth. Log-Gabor functions have some significant advantages over standard Gabor functions. Unlike Gabor functions, which have a positive DC component, Log-Gabor functions have zero DC. Log-Gabor functions also have a long tail in frequency space which more closely matches observations in primates (Hawken and Parker, 1987). However for this study

Log-Gabor functions have two significant disadvantages. Firstly most studies which have carried out physiological measurements have fitted Gabor functions to the data making Log-Gabor functions less directly comparable to these data, and to the standard binocular energy model. Secondly Log-Gabor functions do not possess a windowing function with a clearly defined centre as a standard Gabor function does, rendering the analysis of position disparity more complex.

Parameters for equation 7 were determined for ICA components by fitting a Log-Gabor function in a similar manner to fitting a Gabor component. First, the ICA component was converted to Fourier space then the mean squared error between the absolute value of the ICA component in Fourier space and the Log-Gabor function was minimised using Matlab's `fminsearch` function, initialised by a Genetic Algorithm (as described in the Gabor fitting section above). We did not evaluate the accuracy of the Log-Gabor fitting in as much detail for the Gabor fitting, however we found the algorithm to be extremely consistent. 400 test Gabor functions (see Gabor fitting section above) were fitted 10 times with Log-Gabor functions using the `fminsearch/GA` and random initialisation. The MAD of the deviations divided by the median of the deviations of the fitting error was used to determine the consistency of the fit. A high value would indicate an inconsistent algorithm. The values were divided by the median of the fitting error to prevent large fitting errors dominating the statistic. Over the 400 test Gabor functions, the mean of the MAD was 0.005. This indicates that the fitting algorithm is consistent and we can be confident that the error values it produces when fitted to ICA components are reliable.

In order to compare the accuracy of Gabor functions to Log-Gabor functions as a description of the learned components of the ICA model, we first fitted both Gabor functions and Log-Gabor functions to 8000 components generated by the ICA model (4000 left and 4000 right view components). The accuracy of the fit was determined by measuring the mean squared difference between the fitted Gabor (or Log-Gabor) function and the original component. As Log-Gabor components are only defined in Fourier space the ICA components and Gabor components were transformed to Fourier space and the mean squared difference calculated on the absolute values of the Fourier components. This provided a measure of the accuracy of fit of the envelope only, phase is lost in this comparison. For 8000 components (chosen randomly from the full set to reduce computation times) 7897 were successfully fitted with both Gabor and Log Gabor functions.

The fitting error between the learned components and the Gabor functions and the fitting errors between the learned components and the Log Gabor functions were highly correlated (Spearman's  $\rho$  was 0.99986). Differences between the two measures were standardised by dividing by the mean of both Log-Gabor and Gabor errors, thus differences are specified in terms of overall fitting error. The median difference between the Gabor and Log-Gabor error measurements was 0.003. This is less than the estimated level of consistency in the fitting (0.005, see above). 43.7% of fitted components exhibited standardised differences in error of less than the estimated level of consistency. For 37.3% of fitted components the Gabor function was slightly more accurate than the Log-Gabor function (median standardised error 0.015) and for 19.0% the Log-Gabor functions were slightly more accurate than the Gabor functions (median standardised error 0.054). We concluded that Log-Gabor functions were equally capable of describing the ICA components compared to Gabor functions, however they are unable to describe the position of the RF, which is important in our analysis, without an additional fitting stage in the spatial domain.

## Bootstrapping

The accuracy of the fitting is only one source of error in the process. The computation of the ICA components depends on both the locations of the patches chosen (uniform random distribution across the image) and the initial starting point of the FastICA algorithm (random normal distribution with mean zero) (Hyvärinen, 1999).

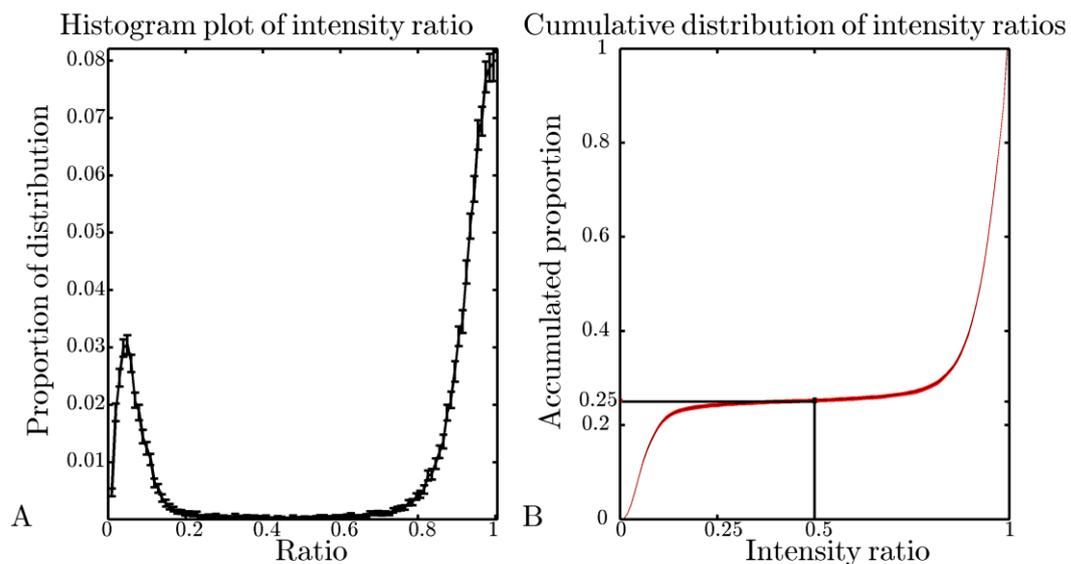
We used a simple bootstrapping method to generate new sample sets from the fitted Gabor functions. To generate a bootstrapped sample-set whole Gabor functions were sampled uniformly at random from the 40000 fitted Gabor functions with replacement. 200 sample-sets of 40000 Gabor functions were generated in this fashion. In order to calculate a distribution from the data, e.g. the distribution of phase disparities, a separate histogram was calculated for each of the bootstrapped sample-sets. Identical bins were used for each histogram. 95% confidence intervals (CI) were computed separately for each bin by sorting the data and taking the values of the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentile respectively. These confidence intervals can be seen as the vertical bars on the histograms. For singular values, bootstrapped confidence intervals are calculated in a similar manner, the statistical value is calculated separately for each of the bootstrapped sample-sets and 2.5% and 97.5% confidence intervals calculated from the 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles. Results of the bootstrapping analysis will be shown where appropriate for each of the distributions analysed, generally this is restricted to histograms rather than scatter plots as the CI of individual points cannot be computed by this method.

## Properties of the ICA components

A detailed analysis was carried out on the parameters of the fitted Gabor functions on each of the 40,000 ICA components.

### Degree of binocularity

Although the samples are left/right normalised to account for local illumination differences, there is no guarantee that the components generated from the ICA algorithm will contain binocular features. Monocular features, i.e. components with weak or non-existent signals from one or other view, will be generated when features in one view occur independently from features in the other view. A measure analogous to ocular dominance can be calculated from the ratio of intensity ( $s$  in equation 5) between left and right component pairs. The larger of the two values was chosen as the denominator. The resulting ratio is directionless with a ratio of 1 being a binocular component equally weighted in each eye, and a ratio of 0 being a fully monocular component with no input from the contralateral eye.



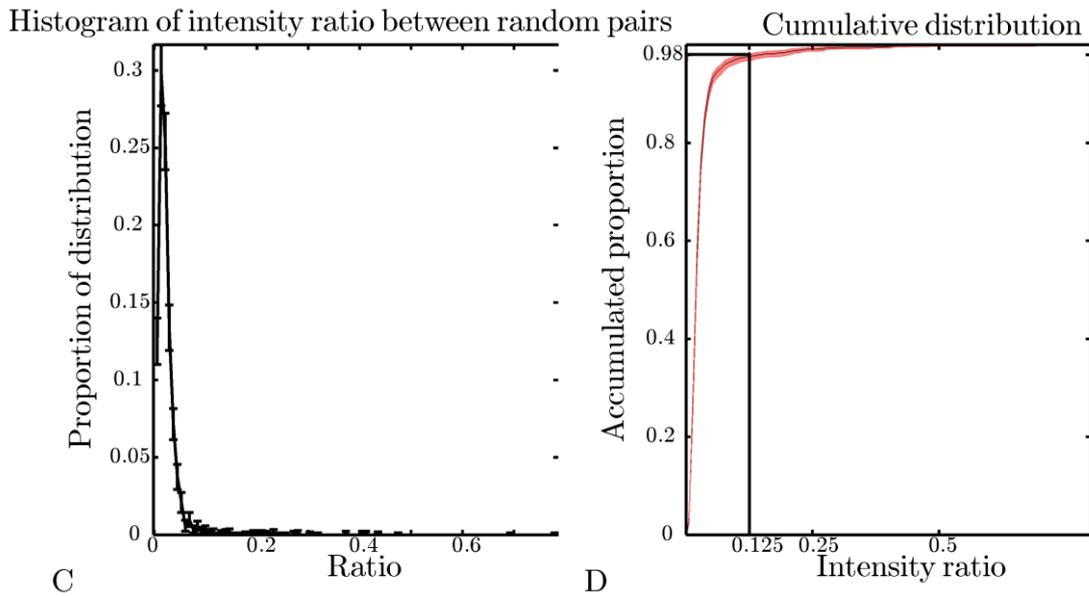


Figure 2. Ratio of intensities between left and right pairs of components. Subplot A shows a bootstrapped histogram of intensities ratio from 39998 component pairs. 95% Confidence Intervals are shown as black error bars. Subplot B shows the cumulative distribution of these ratios. The median of the cumulative distribution for each histogram bin is shown as a black line, the range covered by the 95% Confidence Intervals is shown in red. The distribution is bimodal with the bulk of intensities ranged towards left/right equality of intensity (i.e. 1). From subplot B we can see that 75% of component pairs have ratios greater than 0.5. By way of contrast subplots C and D show the intensity ratios of ICA components trained on randomised patch pairs. C shows the distribution of random patch ratios as a histogram, 98% of the components have a ratio of less than 0.125 (see figure D).

From figure 2A and B we can see that a bimodal distribution of luminance ratios emerges. The larger of the two groups peaks at 1, and indicates a set of binocularly tuned components. 75% of components have a binocular ratio greater than 0.5. This is a conservative estimate as 72% of components have a binocular ratio of greater than 0.8. The smaller group that we designate 'monocular' contains 25% of components. Members of the monocular group have ratios of less than 0.5. Again, this is a conservative estimate as 24% have binocular ratios of less than 0.2. A clear majority of components produced by the ICA analysis show binocular tuned features, with little difference in intensity between left and right components. To test that the binocularity evident in the majority of components truly reflects the interocular redundancies in the test images, luminance ratios were also computed for 4000 ICA components trained on unrelated samples, drawn independently from the left and right images. Figures 2C and D shows the luminance ratios for components trained on randomised patch pairs. Most (98%) components have a luminance ratio of less than 0.125, the proportion of components with a ratio of at least 0.5 was not measurable with any estimate of accuracy (2 out of 4000). The presence of binocular components in randomised patch pairs was negligible, therefore we are able to conclude that binocular components are generated by a relationship between the interocular signals.

### Orientation and frequency analysis of individual components

Figure 3 shows the basic results of fitting Gabor functions to the generated components. A 2D histogram of the locations of the centres of the Gabor functions fitted to both left and right eyes can be seen in Figure 3A. The area of the heat-map corresponds to the dimensions (in arc min) of the image patches, the cells are coloured according to the count of Gabor functions with centres in each cell. From the figure a clear boundary effect can be seen with a large proportion of cells having a zero or close to a zero position in either the x or y-axis; this is an effect of the constraints on the fitting function. Away from the boundary area the full range of available possible positions is represented, with negligible clustering.

The distribution of window sizes can be seen in Figure 3B as a 2D histogram (heat-map) of the window width ( $\sigma_w$ ) against window height ( $\sigma_h$ ) in terms of cycles in the wave-function. As the windows are rotated by  $\psi$ ,  $\sigma_w$  and  $\sigma_h$  do not conform to the x and y-axes, the rotation is also independent of the rotation of the wave-function ( $\theta$ ). The windows are biased towards oval shapes, few show circular shapes (shown on the graph as the dashed black line), however these ovals are not generally particularly elongated. Measuring the window size in terms of cycles also provides a useful indication of the bandwidth of frequency and orientation tuning (Ringach, 2002); a low value for the window size results in a broadband frequency tuned component, a high value results in a narrowband frequency tuned component. Similar logic follows for orientation tuning. The results show a strong tendency towards narrowband tuning, with values for the standard deviation of window size generally greater than 1 in one of the principal directions (either  $\sigma_w, \sigma_h$ ) and generally around 0.5 in the other. As noted by Ringach (2002) this is a substantial deviation from physiology as most cells observed in the V1 are of the macaque visual cortex have window sizes of less than 1 and are therefore much more broadly tuned in frequency and orientation than the components learned using ICA. The median frequency bandwidth of the components was 0.675 octaves (0.673 0.677 95% CI). The median frequency bandwidth for cells in the visual cortex of the macaque is higher than this, around  $\sim 1.4$  octaves (DeValois, Albrecht & Thorell, 1982). It is worth noting however that as the image patches were pre-processed using PCA, which is also a bandwidth limiting process, the narrowband tuning of the learned components is likely to be a result of band-pass filtering in the preprocessing stage.

A bootstrapped rose histogram of the wave-function orientations ( $\theta$ ) is shown in Figure 3C. The black lines show the median of the distributions with the 95% confidence intervals shown as red bars. The orientations cover the range of possible values ( $0^\circ$  to  $180^\circ$ ), with a strong bias towards  $90^\circ$  and  $0^\circ$  ( $180^\circ$  is equivalent to  $0^\circ$ ). Although the distribution of edges in natural images is biased towards  $0^\circ$  and  $90^\circ$  (Hansen and Essock, 2004), it has been observed that ICA tends to produce results in which the orientation and frequency are aligned with the sampling grid (van der Schaaf and van Hateren, 1996). Consequently we are not able to determine the extent to which these results are due to the prevalence of horizontal and vertical features in the binocular natural images or due to biases in the ICA algorithm.

The distribution of the phases of the fitted Gabor functions ( $\phi$ ) is shown in Figure 3D. Again, the medians of the distributions are shown as black lines and the 95% CI in red. The distributions show a generally uniform distribution of phases.

The distribution of frequencies, in cycles per arc min, is plotted as a histogram in Figure 3E. The error bars show the 95% confidence intervals. The range of frequencies is constrained by the minimum wavelength detectable from the sampling lattice, to be less than the Nyquist limit (0.5 cycles/pixel, 0.125 cycles/arcmin) and greater than or equal to 0 cycles/pixel (Shannon, 1949). A strong peak can be seen at 0.085 cycles/arc min (5.1 cycles/degree). The distribution of frequencies is strongly influenced both by the range of frequencies in the training image sets and the implementation of the ICA algorithm (van Hateren and van der Schaaf, 1998). Although the Nyquist limit is 0.125 cycles/arcmin, this requires that the wave-function be perfectly aligned with the sampling lattice, thus the limit is lower in practice. The size of the windowing function is also a factor as there are more ways to pack smaller Gabor functions into the space of the sample image while maintaining the independence of the samples.

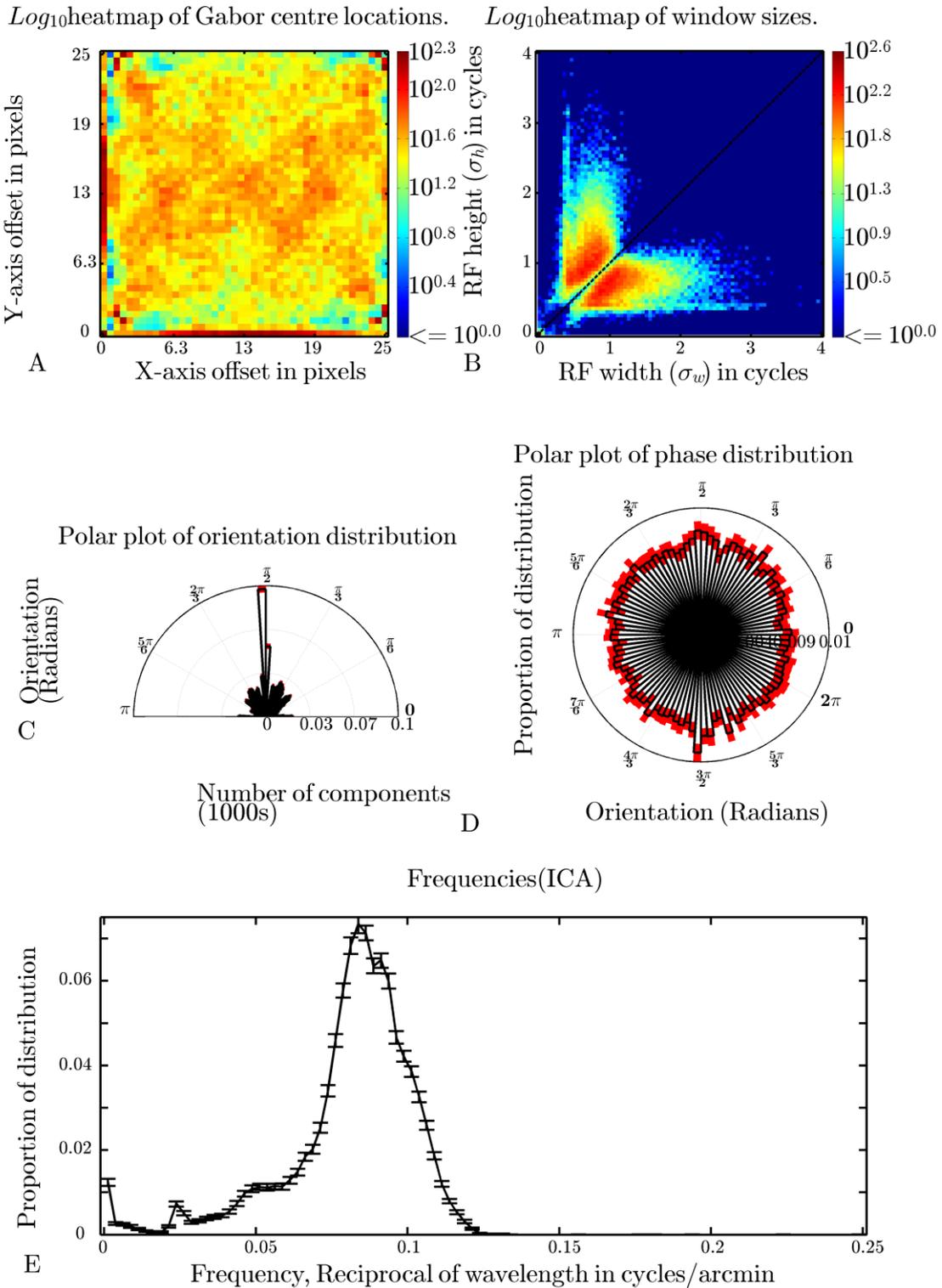


Figure 3. Results of fitting Gabor functions to the components generated from successive ICA of image patches. Subplot A shows a (log<sub>10</sub>) heat-map of locations of the Gabor functions, as measured from the centre point of the windowing function. The boundary effect can be seen as a prevalence of high cell counts (red) along the

$x = 0$  and  $y = 0$  lines. Subplot B shows the distribution of two radii of the windowing function. There is a general tendency towards slightly elliptical functions. A bootstrapped rose histogram of the wave-function orientation can be seen in sub-plot C. The values on the radius axis are histogram counts in 1000s. A significant bias towards  $\frac{\pi}{2}$  radians and 0 radians can be seen. Sub-plot D shows a bootstrapped rose histogram of the phases of the fitted Gabor functions. The values on the radius axis are histogram counts in 1000s. The plot shows a generally even distribution of phases. Subplot E shows the bootstrapped histogram of frequencies of both left and right Gabor functions combined. The frequencies are  $1/\text{wavelength}$  in arcmin. The frequencies of the filters are highly clustered, this is most likely a result of the windowing effect of constraining the data to the size of the image patches. This issue is addressed in detail later in the subsection on 'Scale'.

## Disparity analysis

### Frequency and Orientation disparity analysis

Figure 4A shows a scatterplot of left view orientation against right view orientation. The difference in orientation is shown in Figures 4B&C. There is a clear peak around zero, showing that most interocular differences in receptive field orientation are small. Components are present across the whole range of orientation differences but clustered around zero. The left and right orientations are extremely highly correlated [Pearson's  $r^2=0.99$  (0.993, 0.993 95%CI),  $p<0.001$  ( $<0.001$ ,  $<0.001$  95% CI)]. The spread (median of absolute deviation) of orientation disparities was 0.0196 radians (0.02025, 0.02024 95% CI), the standard deviation was 0.086 radians (0.08122, 0.09395 95% CI).

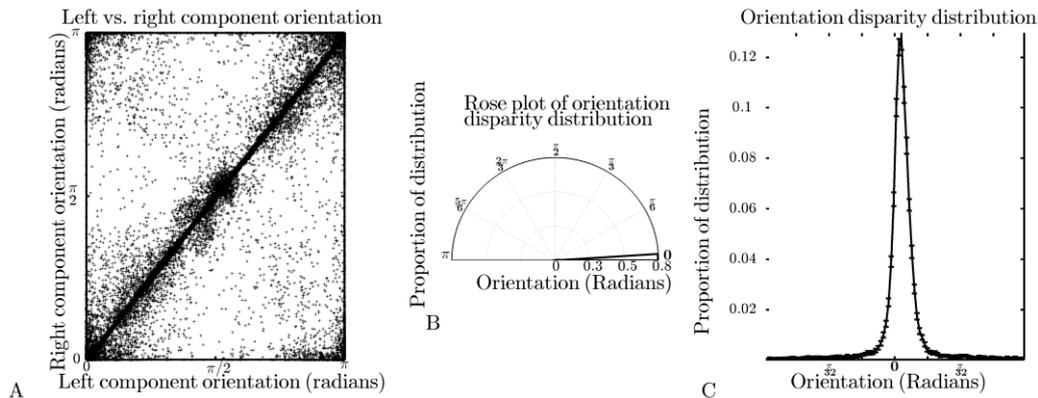


Figure 4. Comparisons of frequency and orientation between left/right pairs of Gabor functions fitted to ICA components. Subplot A shows the relationship between the orientations of the left and right parts of the components. Subplot B shows a bootstrapped rose histogram of the absolute angle differences between left and right fitted Gabors. Subplot C shows a scaled-up subset of the angle differences between  $-\frac{\pi}{16}$  and  $\frac{\pi}{16}$ . The main black line shows the median bootstrapped distribution with the error bars showing the 95% confidence intervals. Most fits produce Gabor functions with similar orientations.

Figure 5 shows the relationship between left and right frequencies for the fitted functions. Again there is a clear linear correlation between left/right frequencies [Pearson's  $r^2=0.98$  (0.982, 0.984 95% CI),  $p = <0.001$  ( $<0.001$ ,  $<0.001$  95% CI)]. As before, a minority of components do not fit the linear profile and appear as outliers in the plot. The vast majority of components are tuned to the same frequency in each view (see figure 5b).

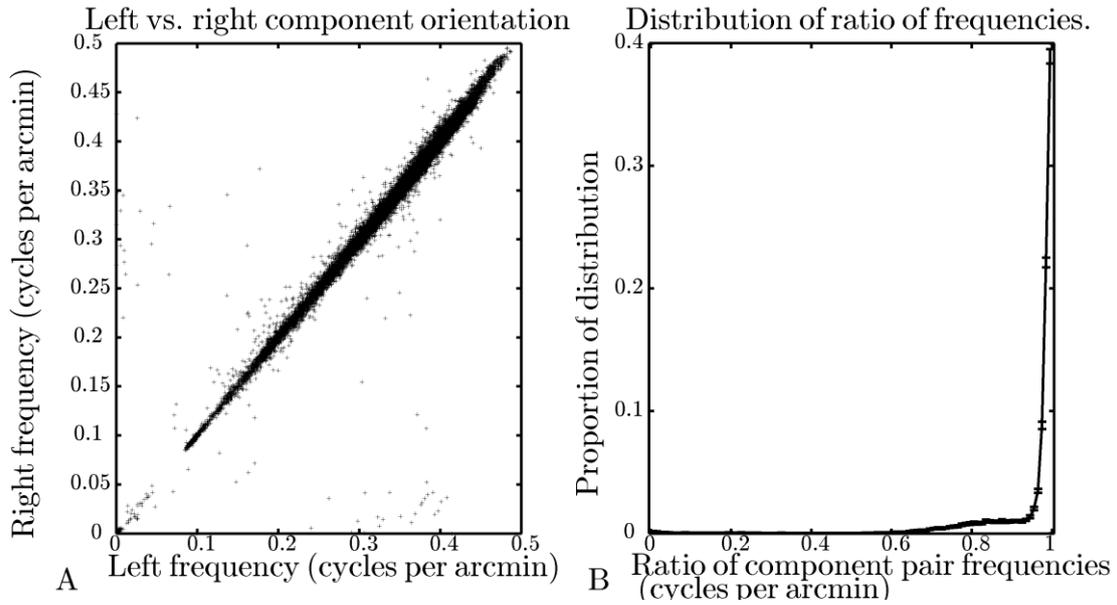


Figure 5. Distribution of tuning frequencies of binocular Gabor pairs. Subplot A shows the distribution of pairs as a scatter plot. A clear linear relationship is visible between left and right frequencies. Subplot B shows the bootstrapped histogram of left/right frequency ratios. For consistency ratios greater than 1 have been inverted to ensure a smallest/largest ratio. Most components have a frequency difference close to zero. The main black line shows the median bootstrapped distribution with the error bars showing the 95% confidence intervals.

### **Phase disparity analysis**

Given twin left/right Gabor responses, two forms of disparity can be calculated: phase disparity and position disparity. A position disparity is a shift in the location of an otherwise identical receptive field between the two views. In contrast, a phase disparity is a change in the shape of the filter, in the form of a shift in the Gabor phase component. This shift is orthogonal to the direction of the component's orientation by definition. Using the phase information from the Gabor functions fitted to the ICA components we calculated the phase difference as the shortest distance around a circle from the two angular phase positions. The distributions of the observed phase differences can be seen in Figure 6, as a polar histogram in Figure 6A and a standard bar histogram constrained to the range  $[0, \pi]$  in Figure 6B. The plots show a strongly bimodal distribution of phase disparity, with peaks at 0 and  $\pi$  radians and troughs at  $\frac{\pi}{2}$ . The distribution is also asymmetric with a bias toward  $\pi$  phase components, indicating a bias in the ICA results towards anti-phase components.

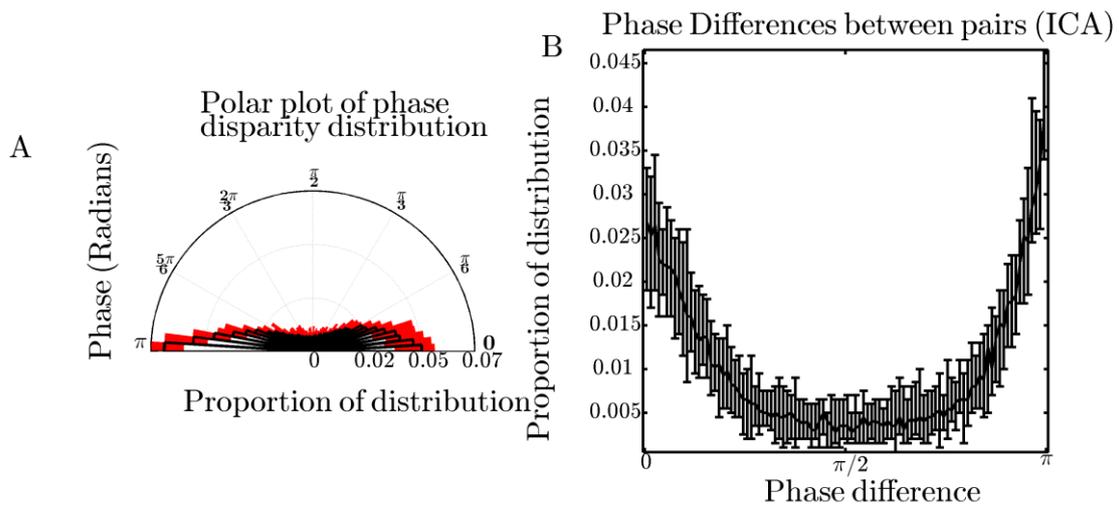


Figure 6. Distribution of absolute phase disparity in the components. Plotted as a polar histogram of the angular distance between left and right phases (A), the black boxes show the median of the bootstrapped distribution for each angular cell, the red boxes show the 95% CI for each cell. Subplot B shows bootstrapped histogram plot of the same results with 95% CI shown as black bars. A bimodal distribution can be clearly seen in the two plots, with the difference between peak and trough that is clearly larger than the estimated error in the distribution. The distribution of binocular phase differences are clearly asymmetric (about  $\frac{\pi}{2}$ ) with a significant difference between the 0 and  $\pi$  phase components.

### **Position disparity analysis**

The Gabor functions consist of two parts, a sinusoid and a windowing function. For each ICA component the centre of the windowing function is found for the left and right parts of the component separately. The displacement disparity between left and right Gabor functions is measured as the distance between the centres of the windowing functions. This can be measured in the horizontal and vertical directions, or in the directions parallel and orthogonal to the orientation of the filter. We consider both.

Simple vertical and horizontal disparities can be calculated by subtracting the left view position coordinates from the right view position coordinate ( $h_r - h_l$ , and  $v_r - v_l$ ). Figure 7 shows the marginal distributions of the horizontal and vertical disparities respectively. For horizontal disparities, negative values indicate components tuned to detect 'near' disparities, positive values indicate components tuned to detect 'far' disparities. For vertical disparities, positive values indicate that the receptive field is shifted upward in the right eye compared to the left eye, and negative values the reverse. As the distributions had very long tails, the plots only show data between the 1.25<sup>th</sup> and 98.75<sup>th</sup> percentiles. The distributions were calculated using 100 bin histograms and confidence intervals calculated using 200 bootstraps.

The distributions of both horizontal and vertical disparities are highly peaked and roughly symmetric about 0, indicating an even mix of near and far tuned features, and appear to obey a double-sided power law. As can be seen in Figure 7A&D the majority of displacements are less than 1 pixel. These small displacements are likely to represent true disparities, rather than chance fluctuations, since the medians of the absolute position disparity distributions (0.39 pixels [0.390,0.402 95% CI] for horizontal disparity and 0.20 [0.198,0.194 95% CI] for vertical disparity) are much greater than the MAD estimated when the accuracy of the fitting was analysed (0.0028 and 0.0026 respectively). The 95% CIs show the range of the bootstrapped distributions. As the cameras, mimicking our eyes, are separated horizontally, we would expect a wider range of displacements on the horizontal axis compared to the vertical axis. The standard deviation of the horizontal displacements was 0.94 pixels

(0.924,0.948 95% CI) and for the vertical displacements 0.42 pixels (0.413,0.422 95% CI). These non-overlapping values match expectations that the distribution of horizontal disparities is broader than that of vertical disparities. Both horizontal and vertical position disparity distributions had similar excess kurtosis. The bootstrapped kurtosis is 3.52 (3.384, 3.731 95% CI) for the horizontal distribution and 3.70 (3.536,3.891 95% CI) for the vertical distribution. An excess kurtosis of 3 indicates that the distribution of both horizontal and vertical position disparity follows a double sided Laplacian. The first and last half percentiles of each distribution were rejected as outliers in this analysis.

Figures 7A&D show the distributions of position disparity in pixels. These are replotted in Figures 7B&E to show disparities as a ratio of the wavelengths of the individual components. The corresponding cumulative distributions are plotted in Figures 7C&F. From these, we can see that 89.1% (88.64-89.42% 95%CI) of horizontal and 99.1% (98.90-99.17% 95%CI) of vertical position disparities are less than half a cycle.

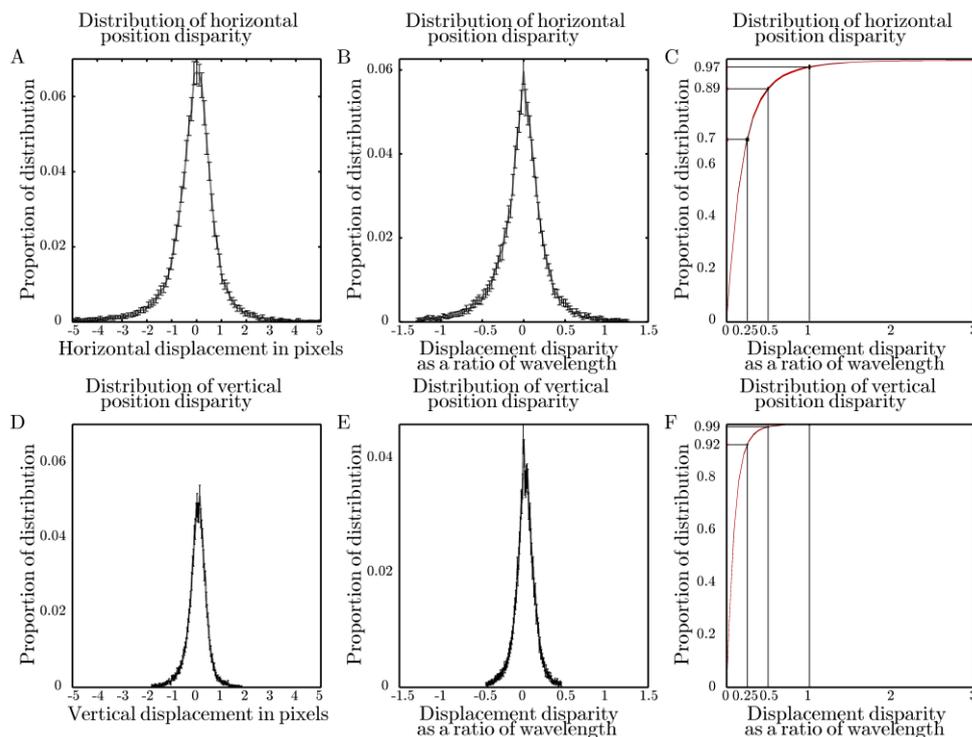


Figure 7. (A,D) Marginal distributions of the horizontal and vertical disparities between left and right view fitted Gabor functions, computed as bootstrapped histograms with 100 bins. The distributions are limited to 98.25% double sided quintiles. The distributions are clearly peaked at 0, broadly symmetric and highly kurtotic. B&E show the displacements as a function of the frequencies of the fitted functions. C&F show the cumulative distributions of the horizontal and vertical displacements respectively. The median of the computed distributions is shown as a black line, the 95% CI intervals are shown in red. The proportions of the distributions with disparities of less than  $\frac{1}{4}$ ,  $\frac{1}{2}$ , 1 cycles are marked along with the CI of proportions shown as red lines on the vertical axis.

The joint distributions of horizontal and vertical disparities are plotted in figure 8. Horizontal and vertical disparities are uncorrelated (Pearson's  $r=0.028$ ,  $p<0.001$ ,  $n=37028$ ) and the mutual information is low (0.0846, calculated using a 2D histogram with 1098 bins using base 2), indicating that the distributions are independent.

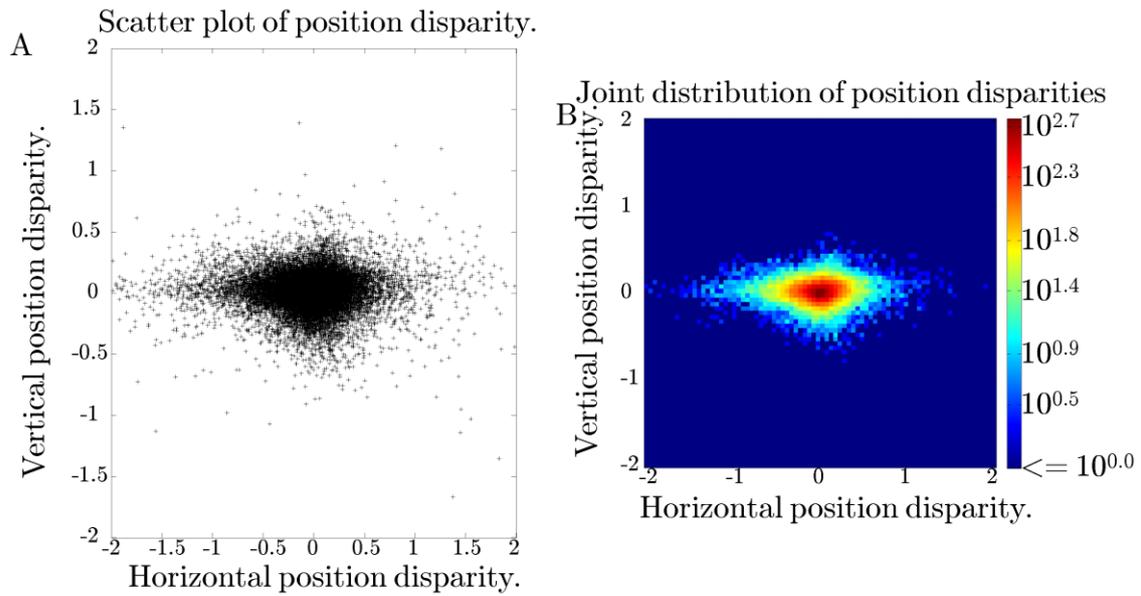


Figure 8. The joint distributions of the horizontal and vertical position disparities as a ratio of wavelength. Subplot A shows a scatter plot of all the computed locations (27661 in total), B shows a 2D heat-map of the distribution with each cell colour coded to show the natural log of the cell count.

An alternative way to describe the distribution of two-dimensional disparities is in terms of the magnitudes of disparity in directions parallel and orthogonal to the orientation tuning of the filters. Analysing disparity in this way is of interest since previous work has assessed the extent to which the direction of disparity to which neurons are tuned is related to their orientation tuning (Cumming, 2002; Gonzalez, Justo, Bermudez and Perez, 2003; Read and Cumming, 2004b). The distributions of disparities are plotted in this way in figure 9. As before we cut the long tailed distributions at the 98.25% percentile and generated confidence intervals using 200 bootstraps. Again, the distributions are clustered about zero, with an exponentially decreasing proportion of components as disparity increases. As can be seen in figure 9B, the vast majority of components, (98.0% [97.75%-98.57% 95%CI] in the orthogonal direction, 90.0% [89.38%-90.01% 95%CI] in the parallel direction) have disparities of less than half of the wavelength. The excess kurtosis is 3.04 (2.915, 3.156 95% CI) for components in the direction orthogonal to the orientation tuning, and 1.52 (1.434, 1.576 95% CI) in the parallel direction.

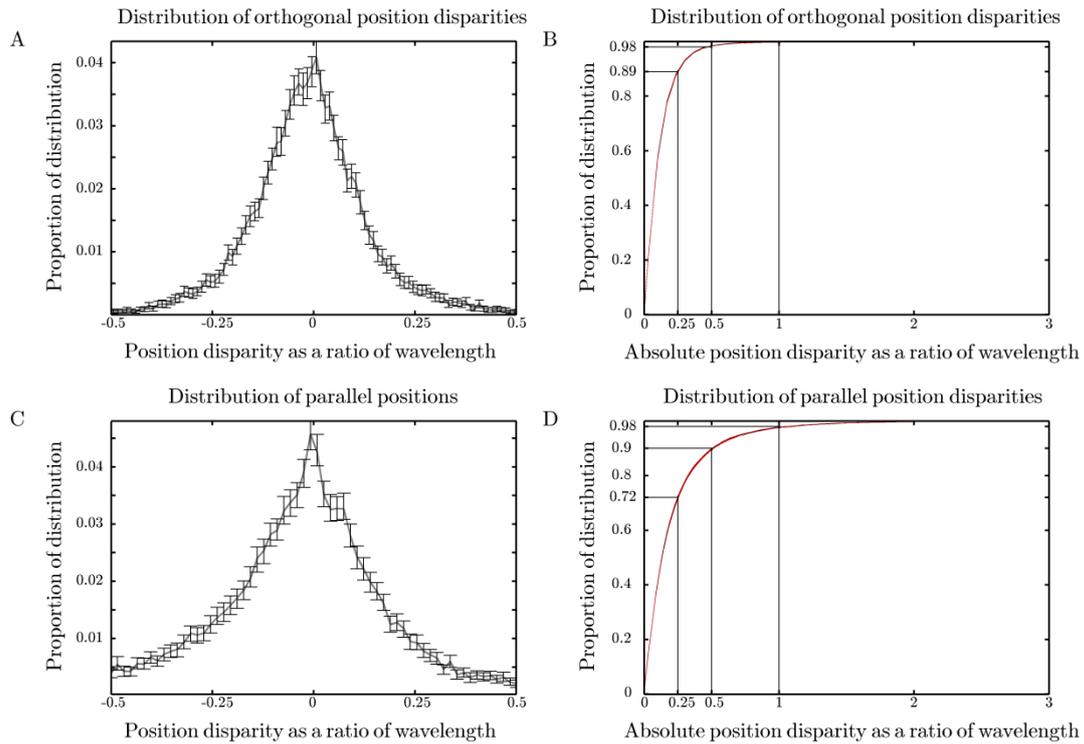


Figure 9. Distributions of position disparities as a ratio of wavelength. Subplots A&C show bootstrapped histograms of (A) position disparities oriented orthogonal to the Gabor orientation (C) and parallel to the Gabor orientation. Subplots B&D show the bootstrapped cumulative distributions for subplots A&C respectively. All distances are ratios of the component wavelength. Positive disparities indicate components tuned to detect 'far' type disparities, negative disparities indicate components tuned to detect 'near' type disparities.

Next, we directly compared the direction of positional disparity with the orientation tuning of the components. The direction was calculated as the inverse tangent ( $\text{atan2}$ ) of the vector between left and right Gabor centres. This angle thus determines the direction of positional displacement in an image-based (horizontal/vertical) rather than component based (parallel/orthogonal) coordinate system. Figure 10A shows a rose plot of the bootstrapped distribution of disparity directions independently of the components' orientation. The distribution is clearly biased towards horizontal (Hodges-Ajne test  $p < 0.001$  (Berens 2009)). Figure 10B shows a heatmap of orientation of position disparity against orientation of the components. From the heatmap no clear association is visible between position disparity orientation and orientation of the filters, no correlation was found (using directional statistics) between position disparity orientation and orientation of the filters ( $\rho = 0.0863$ ,  $c = -0.0111$  (Jammalamadaka and Sengupta 2001 as implemented by Berens 2009)).

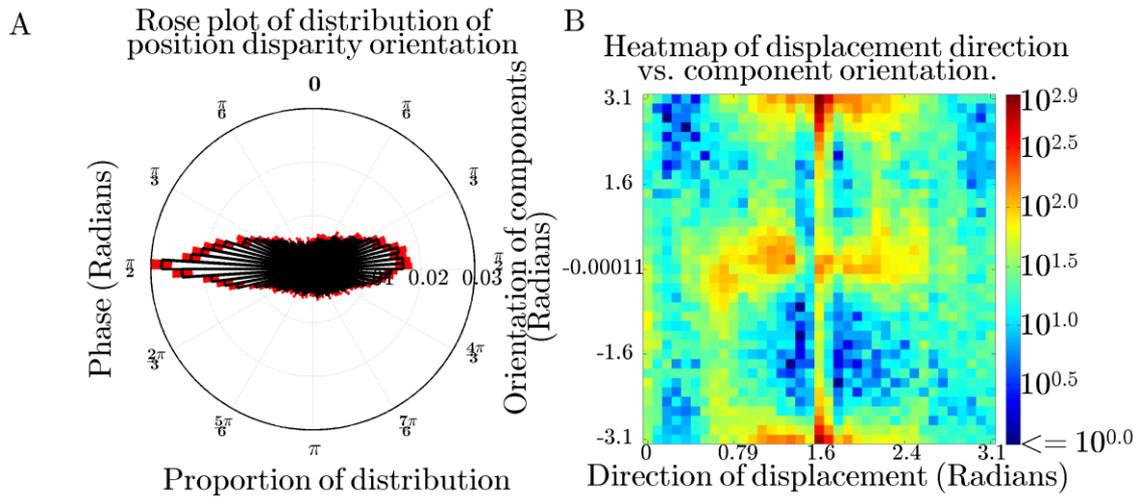


Figure 10. Distribution of direction of disparity. Subplot A shows a rose plot of the distribution of disparities between left and right fitted components. The black lines show the median counts in each bin, the red bars show the range of the 95% CI of the bootstrapped distributions. An angle of 0 radians indicates a vertically oriented Gabor function with positive angles indicating counter-clockwise rotation. Similarly, a displacement angle of 0 radians indicates a vertically oriented displacement. A clear and consistent bias towards horizontal rather than vertical position disparities is also visible, with the distributions showing a smooth transition in between the horizontal and vertical directions. Subplot B shows a Log10 heatmap showing the joint distribution of the orientation of the components ( $\theta$ ) against the direction of position disparity.

***The relationship between position and phase disparity tuning***

Figure 11 shows a scatterplot of phase disparity against position disparity. As the position disparities are long tailed the plot shows only disparities of less than 1 wavelength. This captures 97.7% of the components. The figure shows that the components span the full range of joint position and phase disparities. In the figure we can see the bimodal structure of the phase distribution as clusters around 0 and  $\pm\pi$ .

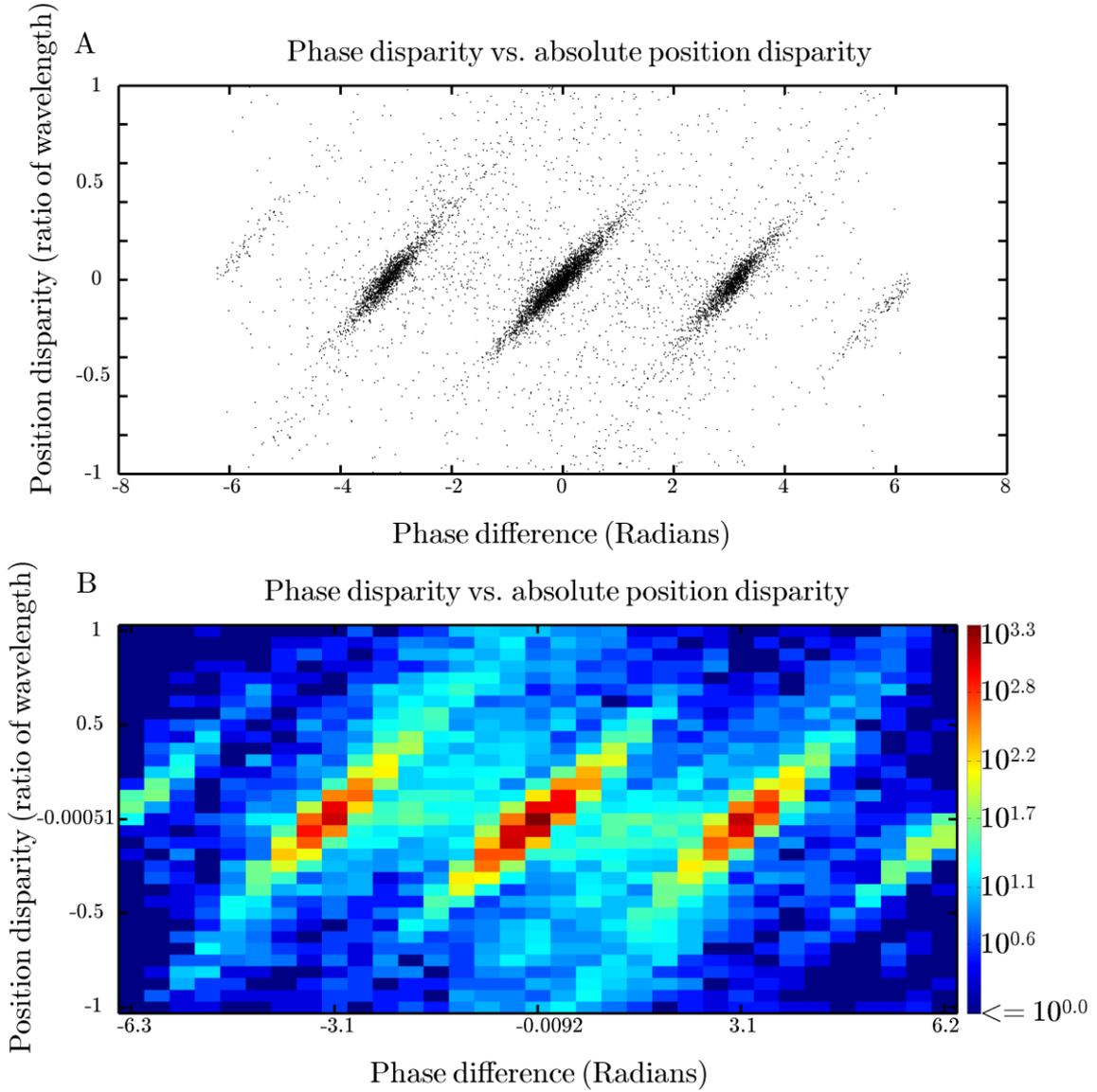


Figure 11. Phase displacement in radians against position disparity as a fraction of wavelength. The lines that appear suggest a link between phase and position disparity. The central cluster shows correlated binocular components, the left and right clusters show anti-correlated components.

Phase and position disparity are related by the Fourier shift theory, thus given a phase disparity, a similar position disparity can be calculated according to equation 7. Part of the interaction between phase and frequency can be calculated by:

$$d_c = (\hat{x}_r - \hat{x}_l) - \frac{(\phi_r - \phi_l)}{2\pi f} \quad (8)$$

Where  $\hat{x}_r$  and  $\hat{x}_l$  are the right and left component positions of the Gabor windows projected onto the wave-function direction (see equation 4).  $\phi_r$  and  $\phi_l$  are the phase of the left and right component Gabor functions.  $d_c$  is effectively the difference in the underlying wave-function of both components.

When measured in terms of wavelength ( $d_{cf}$ ) an integer value (e.g. 0,1 etc.) indicates that the peaks and troughs of the wave-function of the left and right components align, such that the peaks and troughs fall in exactly the same locations in the receptive fields. A  $d_{cf}$  of half integer values indicates that the wave-function is anti-correlated and the peaks in one eye align with the troughs in the other and visa-versa. Note in both cases the windowing function is free to move so the components can have a different configuration of side-bands as the windowing function covers/uncovers different parts of the wave-function.

A plot of the distribution  $d_c$  for the fitted Gabor functions can be seen in figure 12. Values of  $d_c$  are shown in terms of the both pixels and the wavelength of the individual fitted Gabor functions, and are strongly clustered around half integer multiples of the wavelength. This fits well with the strongly correlated and anti-correlated phase results above as correlated components would be separated by integer multiples of the wavelength, and anti-correlated results would be separated by integer + half wavelengths. By calculating the proportion of components contained within each half wavelength band we found that a substantial proportion of components (35.6%, 34.86% Lower, 36.16% Upper 95% CI) are tuned to zero disparity. A larger proportion are tuned to anti-correlated components, 46.65% (45.38% Lower, 47.59% Upper 95%CI) are in the combined -0.5 wavelengths with +0.5 wavelengths categories.

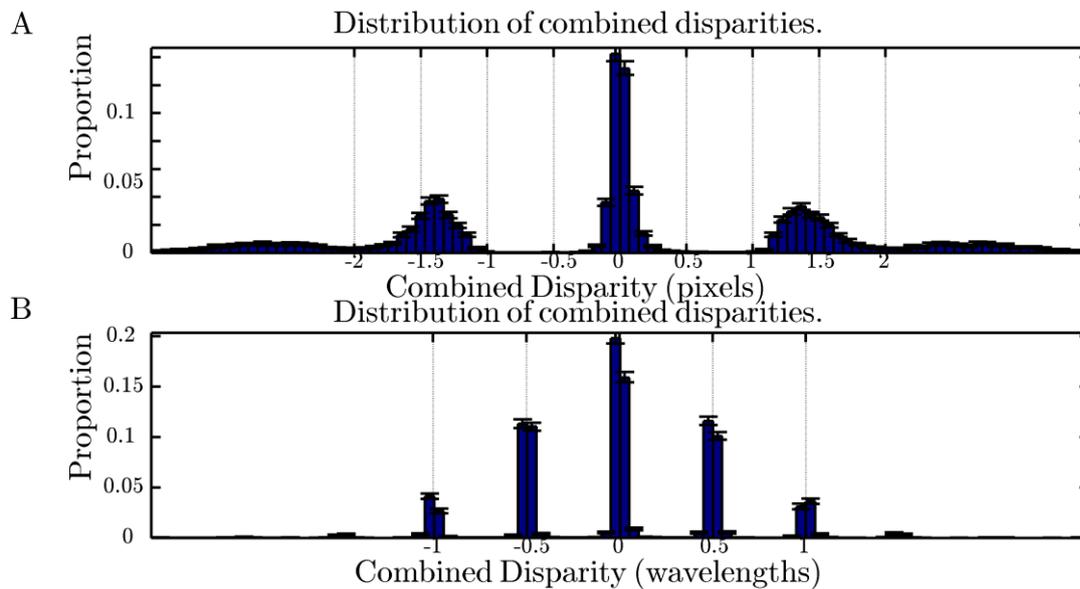


Figure 12. Distribution of combined disparity (disparity remaining when phase and position disparity are accounted for) disparity,  $d_c$  from equation 7) for valid ICA components, calculated using 100 uniformly spaced bins. 95% CI are shown as black error bars. The top plot shows the combined disparity measured in pixels, the bottom chart shows the combined disparity in terms of the wavelength of the individual filters.

The discrete nature of the combined disparity cannot be entirely explained by clustering of the position disparity components around 0 disparity (see figure 9) as the combined disparities are peaked at more locations than 0 and the position disparity function are much broader than the combined disparity clusters. Nor can it be explained entirely by clustering of phase disparity, although the 0 and  $\pi$  radian phase disparities could account for the 0 and  $\frac{1}{2}$  combined disparity peaks, again the distribution is too narrow. Instead the effect is produced by the interaction of phase and disparity.

### Scale

As shown in Figure 3E the ICA components capture only a narrow range of frequencies. To widen the range of frequencies captured in their analysis, van Hateren and van der Schaaf (1998) varied the size of the patches sampled. In order to capture the coarsest scales here would require image

patches too large to feasibly compute using ICA. Thus, rather than vary the size of image patch we kept this constant at 25 by 25 pixels and rescaled the images. 10 scales were chosen, each an octave apart, such that one pixel in the patches covered an area from 10 by 10 arc minutes at the coarsest scale to 1 by 1 arc min at the finest. Components were computed using ICA and Gabor functions were fitted using the method described above. Distributions were calculated using 200 bin histograms with confidence intervals calculated by bootstrapping using 2.5<sup>th</sup> and 97.5<sup>th</sup> percentiles to mark the 95% confidence intervals.

The exact distribution of frequencies depends on both the ICA process and the frequency content of the images. To disambiguate the two we calculated the distributions of ICA components relative to two frames: the sampling grid, i.e. relative to pixels in the image patches, and the visual field of the images measured in arc minutes. If the distributions are constrained by the sampling grid the distributions of the components will be identical across scales when calculated in the sample-grid frame, but differ in the visual field frame (arc min). By contrast, if the sample grid has no effect on the ICA component distributions, the distributions will be identical across scales in the visual field frame (arc min) but differ in the sample-grid frame. It should be noted that the effect we are discussing is a 'windowing' effect, i.e., the components must exist within the image set in order to be detected. The sampling grid simply constrains our view of the data-set.

### ***Degree of binocularity across scale***

As discussed earlier, not all components will contain binocularly tuned features. The proportion of monocularly tuned components will reflect the degree of independence between the left and right views. As the size of features detected is likely to vary across scale while the actual disparities remain constant, it is likely that the proportion of binocular components will change also. In particular we would expect to find fine scale features exhibiting a greater degree of independence between views as the disparities become greater than the wavelength.

We assessed the degree in binocularity using the same ratio of intensity as before. Figure 13B shows the proportion of monocular components generated for each of the 10 scales. Here a monocular component is defined as having a ratio of less than 1/19, i.e. more than 95% of the energy in the component is in the dominant eye. Coarse scales, those of 6 arc minutes per pixel or more, show almost no monocular components, while the vast majority of components at the fine scale are monocular. While the distribution of actual disparities in the images is not known, it seems likely that the tendency for monocular components at the finest scales is due to the disparities in the scene being larger than the features detected. Thus coarse scale feature detectors are better tuned to detect the disparities found in the image set.

As we are primarily interested in binocular disparity, we excluded monocular components from further analysis. Due to the small number of binocular components available from the finest two scales we restricted further analysis to scales coarser than 3 arc minutes per pixel.

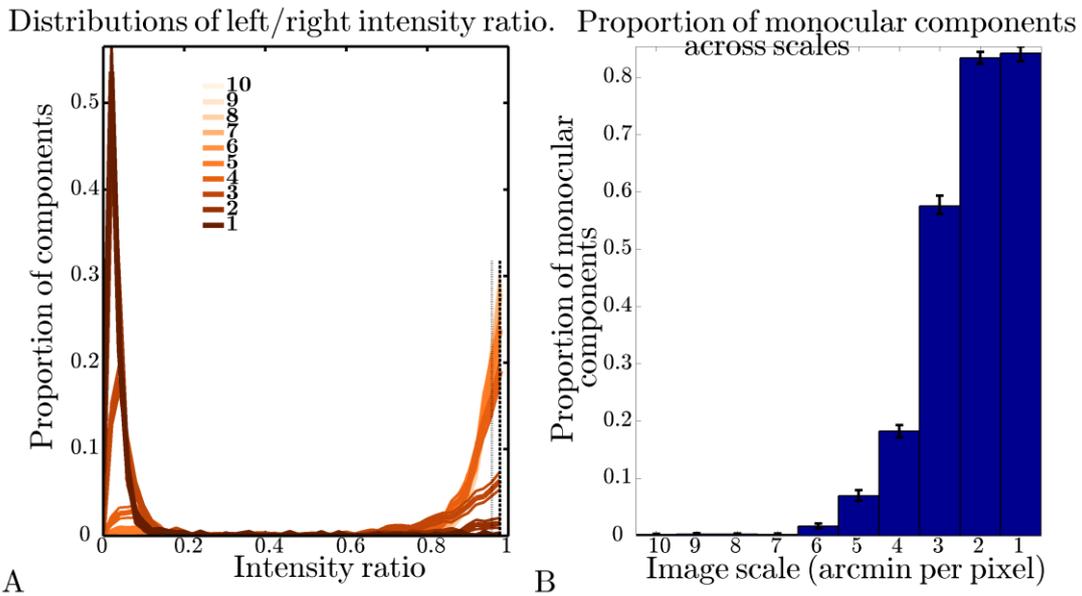


Figure 13. The effect of image scale of the proportion of monocular features. Images at the coarsest scale, 10' per pixel, show the smallest proportion of monocular tuned components, images at the finest scale show a high proportion of monocular tuned components. Subplot A shows the bootstrapped histograms of the intensity ratios at each scale. The median of the bootstrapped distribution is shown as a thick line, the 95% CI are shown as thin-lines. Fine scales show strong peaks at ratios close to zero (monocular) and coarse scales show a bias towards a ratio of 1 (binocular). Subplot B shows the proportion of monocular components at each scale (in arcmin per pixel), the blue bars show the median proportion of monocular components, the 95% CI are shown by the black error bars.

**Frequencies across image scale**

Figure 14 shows the distribution of frequencies for components learned across a range of scales. The distributions are plotted in two modes; in the first the frequencies are measured in cycles per pixel, this shows the effects of the sampling grids and patch size on the frequencies. If the detected frequencies of the components depend principally on the size of the patch, we would expect the frequency envelopes to be identical or similar at all image scales. The second mode shows the frequencies rescaled to cycles per arc minute of visual angle and allows comparison with actual image features. For all but the finest scale the distributions, measured in cycles per pixel, are almost identical (Figure 14A). The largest deviation is found in the 3 arc minute scale. Figure 14B shows the frequency distributions rescaled to show their true wavelengths in cycles per arcmin. The frequencies of the components cover a broad range from 0.02 cycles per arcmin to 0.16 cycles per arcmin (1.2-9.6 cycles/degree). The full distribution is shown as a red line. The distributions are highly bandwidth limited due to patch and sampling grid size. The number of components generated at each scale is fixed and only affected by the proportion of monocular components. It is clear that the resulting overall distributions of frequency tuning owe more to the ICA sampling method than to the distribution of frequencies in the image.

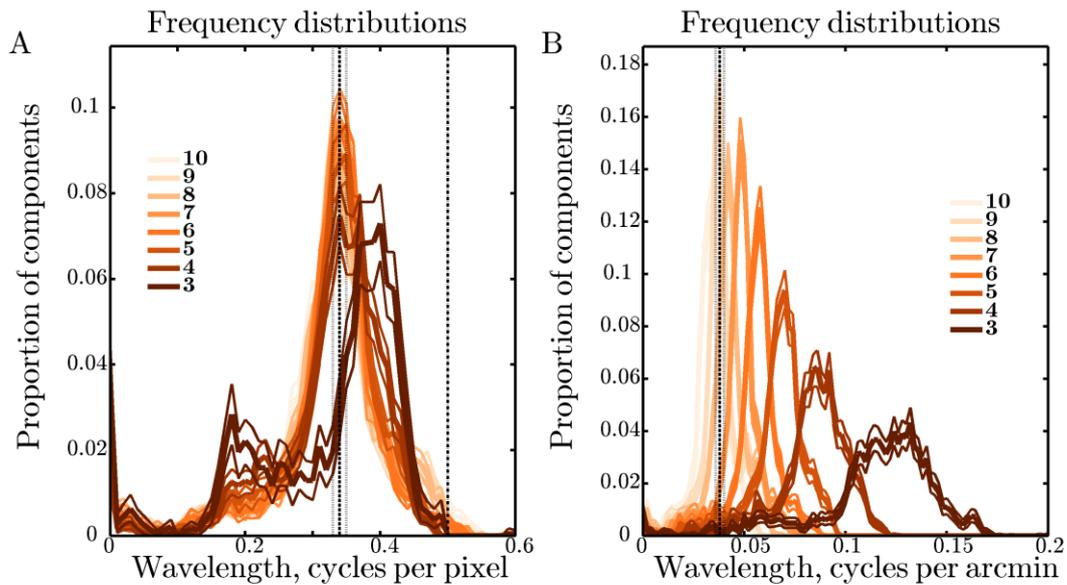


Figure 14. Results of ICA analysis at varying scales. The scales are measured in arc-minutes per pixel. All plots show bootstrapped histograms with the median shown as a thick line and the 95% CI shown as thin lines. The edges of two of the histogram bins are shown as dashed and dotted lines. Subplot A shows the distribution of frequencies in cycles per *pixel* for each of the 10 scales. The Nyquist limit is shown as a dot-dash line. Most of the coarse scaled frequencies show highly similar distributions with some bias towards higher frequencies at the finest scales. Subplot B shows the same distributions as cycles per *arc-minute*. At coarser scales, the tuning shifts to lower frequencies.

### **Position disparities across scale**

The original image set contains a wide but unknown range of disparities that may not be adequately captured at the scale chosen in the detailed analysis presented above. Widening the scale will capture a wider range of disparities, however as with frequency the question of whether the distributions are affected by the ICA method remains. Again, we can test this by comparing the disparity distributions across scale and comparing them relative to the ICA sampling grid and to the original image dimensions. Figure 15 shows the distributions of position disparity across scales, both relative to the sampling grid (pixels) and relative to the visual field (arc min). The distributions are shown both in absolute terms (Figure 15A&B) and in horizontal (C&D) and vertical directions (E&F). Measured relative to the sampling grid the position disparity distributions show a trend towards more kurtotic (peaked) distributions at coarse scales compared to fine scales. The trend is most marked in the horizontal and vertical directions (Figure 15C&E). When measured in terms of the actual visual angle this trend is reversed, with fine scales showing a more kurtotic distribution than coarse scales (Figure 15B&F), except in the horizontal direction where no effect is visible. If the distributions are heavily biased by the ICA algorithm or sampling grid we would expect highly similar distributions across scales.

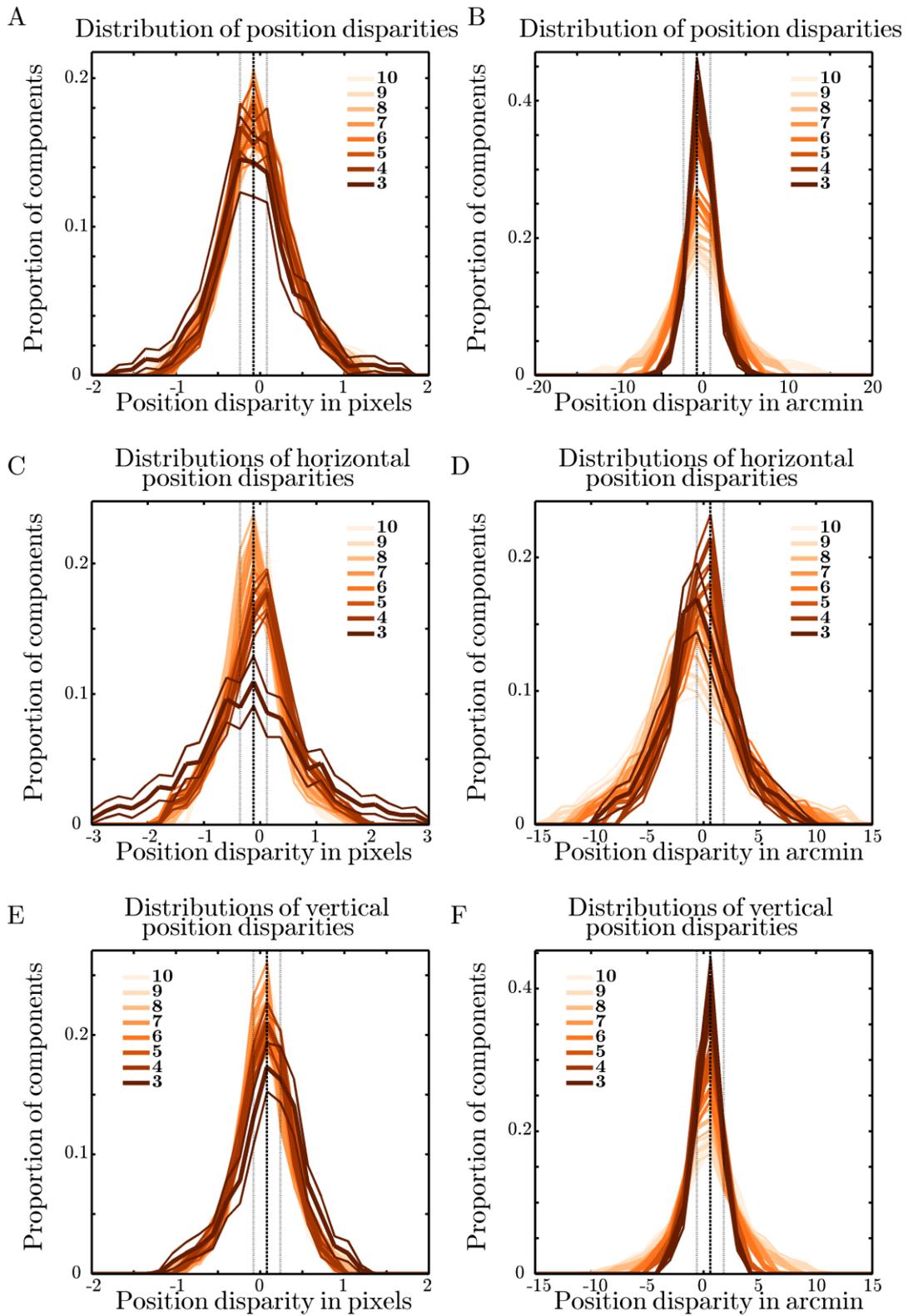


Figure 15. Comparison of disparity distributions across scales. Scales between 3 and 10 arc-minutes per pixel are shown as bootstrapped distributions of varying shades. As before the thick lines denote the median of the distribution, thin lines the 95% CI and the dotted and dashed lines the edges of histogram bins. Subplots A&B show the distribution of absolute position disparities across the scales measures in pixels(A) and arc-minutes(B). Fine scales show a strong bias towards small disparities while coarse scales show a wider coverage. Subplots C&D show the horizontal position disparities in pixels(C) and arc-minutes(D), and subplots E&F show vertical position disparities in pixels(E) and arc-minutes(F). As with the absolute displacements (A&B) fine scales shows a bias towards small disparities while coarse scales show a wider distribution, however in the horizontal case it is weaker. Unlike the frequency distributions, which are strongly tied to the size of the patches, the position disparities vary across scales.

The large variance in the distributions of position disparity in both the horizontal and vertical direction, when plotted in pixel units, suggests that the sampling grid alone is not driving the distribution. Similarly, if driven purely by the sampling grid the range of position disparities, measured in arc minutes, would double when the sampling rate halved. While the range of position disparities in arc minutes is wider at coarse scales, the width does not double with a halving of the scale.

### **Phase disparities across scales.**

The distributions of phase disparity across scales can be seen in figure 16. The distributions are highly similar across scales. This indicates that the results hold over a range of frequencies between ~2 cycles per degree and ~10 cycles per degree.

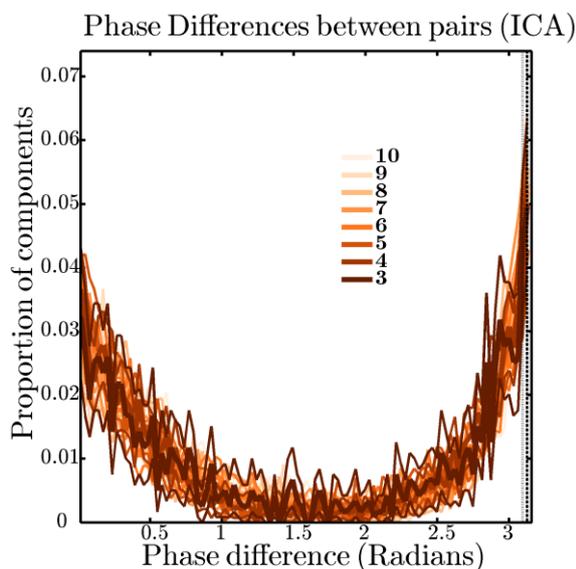


Figure 16. Bootstrapped distributions of phase differences across scales. The phase distributions follow the same bimodal pattern across the selected scales.

## **Relating the results to physiological findings.**

The ICA algorithm seeks to produce components that are maximally independent. In order to determine whether the resulting components provide an accurate model of binocular neurons, we directly compared the ICA components with physiological measurements. We describe the similarities in qualitative terms, paying close attention to similarities in the types of distributions and where applicable the ranges they cover.

### **Whitening**

It has been hypothesized that whitening of image information is performed by 'centre-surround' cells in the retina (Srinivasan, Laughlin and Dubs, 1982, Atick and Redlich, 1992). By analysing the spatial tuning characteristics of P-cells in the retina (using data published by Croner and Kaplan, 1995)

Graham, Chandler and Field (2006) found that the sensitivity of P cells is well-matched to the power spectra of natural scenes. However due to correlations that remained amongst neighbouring P-cells they came to the weaker conclusion that the P-cells performed an approximate response spectrum flattening. In our work we used PCA as it is a pre-requisite of Independent Components Analysis, however the spectral flattening performed in the retina provides some ecological validation for the PCA based whitening step.

However, it should also be noted that the PCA components generated in our analysis do not resemble the centre-surround responses found in retinal P-cells. The centre-surround responses of these cells are limited to half-cycle representations, while the PCA components have multiple cycles, the number of which depends on the wavelength. This results in eigenvectors (components used in the whitening) with a significantly larger RF than retinal ganglion cells. This may have implications for the RF size of ICA component that will be discussed later.

### **Orientation and spatial frequency disparity tuning**

Gabor functions tuned to cover the full range of orientations were found in both left and right view parts of the components (see figure 2c). However actual orientations were closely clustered on  $0$ ,  $\frac{\pi}{2}$  and  $\pi$ , corresponding to the sampling grid. ICA has a tendency to produce components aligned with the sampling grid as these have a lower energy state than un-aligned states (van der Schaaf and van Hateren, 1996). It is known that a particular tendency for horizontal and vertical orientations exists in photographic images (Hansen and Essock, 2004). While this may well to some extent reflect an anisotropic distribution of orientations in nature, it is also likely to result from the alignment of structures with the cardinal directions when composing photographs (van Hateren and van der Schaaf, 1998). It is not possible to attribute the anisotropy in our results to any corresponding anisotropy in the natural environment since it is likely to be driven to a large degree by the sampling grid in our photographs (van Hateren and van der Schaaf, 1998). The orientations of the left and right Gabor functions of the binocular component pairs were highly correlated ( $r^2=0.99$ ,  $p<0.001$ ). This is similar to results from physiology; Bridge and Cumming (2001) report an almost identical correlation of  $r^2=0.985$  and a spread (standard deviation) of orientation disparities (IDPO in their terminology) of  $9.22^\circ$ . We observed a spread (standard deviation) of  $3.55^\circ$ , around half that of Bridge and Cumming's result but of a similar order of magnitude. Due to the small angles involved, measurement noise could account for the discrepancy. This result supports the idea that a matching process, where features in one eye are matched with similarly orientated features in the other, is an efficient mechanism to code binocular scenes and therefore an effective strategy to compute binocular disparity.

The modal frequency of the components in the main analysis was approximately 4.8 cycles per degree and ranged between  $\sim 0$  and  $\sim 7$  cycles per degrees. This places the components within the range of frequency tuning of binocular simple cells in the macaque visual cortex (Prince, Pointon, Cumming, & Parker, 2002). This indicates that the range of frequencies selected by the band-pass filtering in the whitening stage is appropriate for comparison to physiology. The frequency distribution is skewed towards higher frequencies ( $\sim 4.8$  cycles per degree); this resembles the results from physiology (see (Prince, Pointon, et al., 2002) figure 4e).

Ringach (2002) noted a difference between receptive field sizes computed for ICA components and those measured physiologically in both cats and monkeys. Ringach noted that a substantially greater proportion of broad-band tuned cells compared to the tunings of ICA components (Ringach 2002). We found a similar effect in our data (see figure 3), in fact many of our ICA components are even more narrowly tuned than Ringach's. This indicates that a substantial number of V1 simple cells have a much smaller RF than measured here. A bias toward narrow-band frequency tuned components is a result of band-pass filtering in the PCA pre-processing stage, as discussed by Ringach et al (1997) band-pass filtering will remove high frequency information from the samples and subsequently from learned components, this can introduce ripple-like structures, which results in the components being biased towards narrow frequency band-width. As measurements of phase, frequency and orientation

are more reliable for narrow-band components than wide band components, we can be confident of the accuracy of our results.

### **Phase disparity**

The phase disparity tuning of our components non-uniformly spans the range of possible angular values, suggesting that the full range of phase disparity could be used to detect disparities in the visual scene. However, phase disparities of  $\pi/2$  are much less prevalent than disparities around 0 and  $\pi$ , implying that these disparities have less explanatory power.

The distribution of phase disparities was strongly bimodal (see figure 6), with peaks at 0 and  $\pi$  radians. The peak at 0 radians indicates the detection of correlated signals in each view. As the phase disparity is partially independent of the position disparity these correlated signals may be shifted in each view. The components around  $\pi$  radians are anti-correlated between the left and right eye. Their presence is consistent with Li and Atick's de-correlated channels theory of binocular vision (Li and Atick, 1994a). The 'plus' (correlated) and 'minus' (anti-correlated) channels that de-correlate single pixel sample inputs in their research are found in the interactions between multiple pixels in the ICA models as phase differences. As noted by Bell and Sejnowski (1997) and Ringach (2002), edge-like components produce sparse coding in the monocular case, locally decorrelating the images. The appearance of anti-correlated binocular sparse components is the logical extension of this to binocular image patches. The bias towards anti-correlated binocular components has been observed before in Fourier analysis by Li and Atick (1994a) and similar anti-correlated filters were also produced in Burge and Geisler's (2014) analysis of optimal filters for disparity estimation. They noted that a particular anti-correlated component could signal the presence of a stimulus at a particular disparity by *not* responding. This is related to the idea that such cells play an inhibitory role, of vetoing possible disparities when they do respond strongly (Read and Cumming, 2007). Recently an additional role for these anti-correlated cells in distinguishing object boundaries from texture edges has been suggested by Goutcher, Hunter and, Hibbard (2013).

Only half of this bimodal distribution has been found in physiological studies. Phase disparities in binocular cells of the macaque (Prince, Cumming, et al., 2002), cat (Anzai et al., 1999), and barn owl (Nieder and Wagner, 2000), as collated by Prince, Cumming, et al., (2002), showed a clear bias towards phase disparities of 0 radians, but with few anti-correlated neurons. Although such 'tuned inhibitory' neurons do exist (Poggio and Fischer, 1977; Poggio et al., 1985), they are not nearly as prevalent as would be predicted from the current analysis. We and previous authors have assumed that the visual cortex forms an efficient coding of the visual scene, but without any reference to the utility of the coding. In other words, neurons could code for stimuli that exist in the visual input but are not used in subsequent processing. This would be extremely energy inefficient. It is reasonable to assume that a pruning process could occur that selects for useful parts of the visual signal, although at present no biological process for such pruning is known. Analyses that are targeted at identifying the filters that are optimised for specific tasks, such as disparity estimation (Burge and Geisler, 2014) or scene parsing (Goutcher et al. 2013) can provide an additional level of understanding of the encoding of natural images.

### **Position disparity**

Distributions of position disparity tuning were highly peaked around 0. Components tuned to a disparity of less than half a wavelength dominate the distribution, with a clear majority tuned to a disparity of less than quarter of a wavelength. This result is qualitatively similar to neurophysiological measurements in V1 of macaques. Both Anzai et al. and Prince et al. found position disparities were mainly constrained to half the wavelength of the filter and clustered around zero (Anzai et al., 1999; Prince, Cumming, et al., 2002; Prince, Pointon, et al., 2002). Finding this result in ICA components indicates that the relatively small position disparities found in animal studies form an efficient coding

of binocular visual inputs, in line with geometrical considerations (Hibbard, 2007; Read and Cumming, 2004a). The distribution of tuning for vertical disparity was more strongly peaked at zero than that for horizontal disparity, again in line with geometrical predictions (Read and Cumming, 2004; Hibbard, 2007) and physiological findings (Cumming, 2002).

We do not have information about the range of disparities present in the images. This is however likely to be greater than the range of disparity tuning found in our components (between  $2^2$  and 8 arc minutes (see Figure 15)), since training patches were cut from both the verged and un-verged regions of the image. We can conclude that although disparities greater than the wavelength of the filter are most likely present in the scenes, filters tuned to these disparities do not form an efficient coding. An explanation of this apparent discrepancy can be found in the fact that filters do not exclusively respond to object boundaries, but also encode object texture. Unlike object boundary edges textures are frequently repetitive, thus a disparity detector tuned to a particular frequency could therefore make many good matches other than the correct disparity. As the ICA algorithm will find the most prevalent features regardless of actual disparity, the matches with the shortest distance between edges will dominate. While this results in an efficient encoding, this ambiguity needs to be resolved for the actual estimation of disparity. This highlights the important distinction between the initial encoding of binocular images in V1 (the stage that we seek to model here) and the subsequent estimation of disparity in higher cortical areas. It should also be noted that components tuned to larger disparities are present in ICA applied at coarser scales, consistent with the idea that larger disparities are detected by neurons with larger receptive fields (Allenmark and Read, 2011).

### **Orientation tuning and the direction of disparity**

Cumming (2002) found no correlation between the orientation tuning of neurons and the direction of disparity to which they were most sensitive. The oval shaped distribution of displacements shown in Figure 8A is similar to results found in monkeys (Cumming, 2002) where a similar bias towards detection of horizontal disparities was found. Like Cumming we found no association between the direction of position disparity and the orientation of the components. It should be noted however that the strong effect of grid-alignment found in the data might have masked any such association.

### **The relationship between phase and position disparity tuning**

Physiological studies have found evidence of mixtures of phase disparity and position disparity tuning. These tunings span a large proportion of the range of possible phase disparity and position disparity combinations and were found to be uncorrelated. Like our results, neural tuning clusters around 0 in both position and phase disparities (Anzai et al., 1999; Prince, Cumming, et al., 2002; Prince, Pointon, et al., 2002).

We found a strong linear relationship between phase disparity and position disparity that implies joint use of phase disparity and position disparity tuned components in scene disparity calculations. This relationship can be explained by the similarities between Gabor functions shifted in phase and shifted in position (the basis of the Fourier Shift theory). This result is intriguing. Rather than spanning the space of all possible phase and position disparities the components are clustered at particular combinations that have particular combined disparity, specifically multiples of half the wavelength. This result has not been observed in image statistics by other authors (Burge and Geisler (2014). Similarly it has not been observed in physiological studies. Cumming and De Angelis (2001) and Anzai et al. (1999) found a wide spread of phase and position disparities, Prince et al. did not convert their measurements into equivalent units and cannot be directly compared (Prince et al. 2002a). The phase and position distributions reported by these authors are marginal distributions combining results across a wide range of frequency tunings from broad-band to narrow-band. Our results are constrained in frequency tuning and are generally more narrow-band tuned than physiological measurements. It is possible that the broader range of phase and position disparities measured in

---

<sup>2</sup> The Nyquist limit.

primates is a consequence of broad frequency tuning functions, with narrow-band tuned cells exhibiting a similar distribution to one observed here.

Possible algorithms have been suggested to use these two measures of disparity. For example, Chen and Qian (2004) used phase disparity to estimate local shifts and position disparity to confirm the results. In contrast Read and Cumming (2007) used position disparity to calculate local shifts, and phase disparity to detect false positives. Read and Cumming claimed more accurate results for their method compared to Chen and Qian. These algorithms assume that phase and position disparity are not correlated.

## Scale

The range of disparities found across image scales (Figure 16) indicates a multi-scale approach to the detection of disparity (Allenmark and Read, 2011). At each scale the distribution of components spans a similar range of frequencies. A similar distribution was proposed mathematically by Li and Atick to perform two functions on binocular images: noise reduction, especially at higher frequencies, and whitening of the input signals (Li and Atick, 1994b). The range of disparities detected by components at the different scales shares some of their features with Li and Atick's model. At each scale the band-pass operation avoids signals close to the Shannon-Nyquist limit where high-frequency noise dominates the signal (Shannon, 1949). The range of frequencies detected also depends on the scale, with coarse scales detecting a wider range of frequencies and disparities than higher frequencies.

## Discussion

We used Independent Component Analysis to produce a sparse linear coding of binocular image patches. This produced Gabor-like features which we analysed in terms of phase disparity and position disparity.

As with other authors we have found a range of phase and position disparity coding components (Anzai et al., 1999; Okajima, 2004). Our analysis has produced many more, and more detailed measurements than previous studies. As a result, we observed new relationships in the data. We found a linear relationship between phase and position disparity that produced components with highly clustered overall disparity profiles. This is different from the physiological measurements of (Anzai et al., 1999; Prince, Cumming, et al., 2002) which show no correlation between phase and position disparity tunings. In terms of signal processing our results show a clear link between narrow-band phase and position disparities in each view.

Physiological measurements have found many fewer anti-correlated tuned neurons in V1 than predicted by our results (Prince, Cumming, et al., 2002). However, one important distinction is that such physiological measures have been taken from cells with receptive fields in the centre of the visual field, whereas the analysis presented here considers samples drawn from the whole image. While beyond the scope of the present paper, it is possible that this discrepancy reflects the difference in spatial sampling.

Overall, there are a number of ways in which our ICA results produce components with properties that are similar to binocular cortical neurons. Our components are well fitted by Gabor functions with similar orientation and spatial frequency tuning in each eye (Bridge and Cumming, 2001). These components showed position and phase disparity tuning, with most components having a combination of both (Prince, Cumming and Parker, 2002). The distribution of horizontal and vertical position disparity were both strongly peaked around zero, with a greater spread for horizontal disparity than for vertical disparity (Cumming, 2002; Read & Cumming, 2004b; Hibbard, 2007). There was also a local peak in the distribution of phase disparities around zero (Prince, Cumming and Parker, 2002).

There were however also a number of ways in which our results differed from physiological findings. Most notably, the largest peak in the distribution of phase disparities was found at  $\pi$ . Also, when the preferred disparity of each component was calculated, by taking account of both its position and phase tuning, peaks in the distribution at half-wavelength intervals were evident. These unexpected results represent aspects of the components learned that do not directly reflect attributed of cortical neurons (Ringach, 2002).

The ICA algorithm detected features based on their prevalence in the supplied image set. Matches between these components and the images do not necessarily mean that the actual disparities of objects in the image match that of the component. It is likely that a proportion of these matches will be 'false' i.e. the matched disparities are not the same as the actual, physical disparities. The algorithm has a strong bias towards narrowband features with large receptive fields, while the visual cortices of cats and monkeys have a larger proportion of broad-band features with relatively smaller receptive fields (Ringach 2002). In our analysis the components had a median frequency bandwidth of  $\sim 0.5$  octaves, this differs from the median of 1.4 octaves reported by De Valois et al.(1982) . Thus we have been comparing narrow-band tuned ICA component to the marginal statistics of features measure from V1 simple cells of a wide range of narrow to broad-band tunings. The similarities we have found have not been restricted to narrow-band features in V1 suggesting these features are not dependent of the band-width of the Gabor function. However the differences that we have observed may be due in part to the size of the receptive fields and only hold for narrow-band signals.

Taken on their own the components calculated via ICA cannot calculate disparity. ICA is simply an efficient coding method, and simple correlations between the scene and components will produce many false matches. Algorithms based on the standard energy model combine the outputs of two or more components, with a non-linear term, as the initial stage of disparity detection. The original model of Fleet et al. combined pairs of components with 0 position and  $\frac{\pi}{2}$  phase shift (Fleet et al., 1996), while also pooling information across orientation, scale and spatial position. Numerous possible combinations of components have been suggested, including multi-scale phase based models (Chen and Qian, 2004), a gated model where phase maxima close to zero are combined with position extrema (Read and Cumming, 2007), combining positive and negative energy model units (Haefner and Cumming, 2005) and filters based on learning the appropriate combinations from natural image data (Burge and Geisler 2014). Whatever combination of binocularly-encoded information is required for the estimation of disparity, this is likely to occur in visual areas beyond V1.

Our results, in line with previous studies, reveal some clear similarities between the components learned and the receptive fields of cortical neurons, but also a number of differences. This highlights the limitations of trying to explain these receptive field properties as the result of the encoding principles employed. One important limitation is that the result might reflect properties of the learning methods employed, such as image sampling and preprocessing, that do not reflect the relevant encoding principles, such as independence (Bell and Sejnowski, 1997; van Hateren and van der Schaaf, 1998; Ringach, 2002). Another important consideration is that the properties of cortical neurons are likely to be determined not by considerations of how to encode information so as to generate a full representation of the image (Ringach, 2002), but by how this information will subsequently be used, for example in the estimation of disparity (Burge & Geisler, 2014).

Another important limitation of our approach, in line with many other approaches to understanding efficient coding, is that it does not explicitly consider noise in either the input signal, nor in neural responses (Simoncelli & Olshausen, 2001). The levels of both forms of noise are important factors in determining an efficient encoding (for a detailed discussion see (Zhaoping, 2014)). Consideration of noise has contributed to our understanding of the efficient coding of information in the retina (Atick & Redlich, 1990, 1992) and in binocular vision (Li & Atick, 1994a), for example. The approach to learning independent components does however have some advantages from a consideration of the signal-to-noise ratio. Firstly, as discussed in the methods section, the PCA whitening stage, in which

later, higher-frequency component are truncated, bears some similarity to the explanation of retinal encoding proposed by Atick and Redlich (1992), in which noise is shown to dominate the signal at high frequencies (due to the  $\frac{1}{f^2}$  power spectrum typical of natural images) and truncation of the signal to lower frequencies increases the signal to noise ratio. Secondly, (Field, 1987, 1994) argued that in the context of uncorrelated signal noise, sparse coding may increase the signal to noise ratio as neurons will respond selectively to a subset of the signal space while uniform white noise is distributed across the entire space of possible signals. However as Hyvärinen Sarela & Vigário pointed out an ICA model trained on noisy input data will produce components tuned selectively to respond to a single (or close to) sample (Hyvärinen et al., 1999). In our work we have used bootstrapping in an attempt to assess the impact such outliers have on the distributions of components learned from ICA.

A final consideration is that, while the methods we used to calculate components have the aim of producing a sparse encoding, alternative metrics exist. Einhäuser, Kayser, König and Körding (2002) used a method that attempted to maximise 'temporal stability,' arguing that this was necessary for learning the features of 'complex' cells but not for learning 'simple' cell responses. It should be noted that although not explicitly referred to by the authors, this learning method was sparse in that only a subset of inputs was used for learning in any given iteration (Einhäuser et al., 2002). The idea of temporal stability was also explored by Hurri and Hyvärinen using a different method that was non-sparse (Hurri and Hyvärinen, 2003). Other statistical models exist that have demonstrable efficacy in describing natural images, in many cases outperforming ICA in terms of image data explained (log likelihood) (Zoran and Weiss, 2012). The class of models known as Mixture Models has received a significant amount of attention; variations on these models have been shown to outperform ICA. Both Gaussian Scale Models (Lyu and Simoncelli, 2007) and Gaussian Mixture Models (Xu and Jordan, 1996) have been found to be better descriptors of natural image patches in terms of their likelihood. Mixtures of elliptically contoured distributions have been shown psychophysically to produce more natural looking image patches (Gerhard, Wichmann, and Bethge, 2013). However, unlike factor models such as ICA, mixture models do not produce sets of individual components with which to compare to physiology, instead they produce linearly additive mixtures of distributions with no direct physiological analogue.

## Data-set

Binocular photographic image data and (Matlab) source code associated with this publication is available from GitHub at <https://github.com/DavidWilliamHunter/Bivis>

## Acknowledgements

This work was supported by EPSRC Strategic Partnership Funds 2012-13 with University of St Andrews, Scotland, UK and the BBSRC [ grant number BB/K018973/1]

## References

- Allenmark, F. & Read, J.C.A. (2011) Spatial stereoresolution for depth corrugations may be set in primary visual cortex, *PLoS Computational Biology*, 7(8): e1002142, doi:10.1371/journal.pcbi.1002142
- Atick, J. J., & Redlich, A. N. (1990). Towards a theory of early visual processing. *Neural Computation*, 2(3), 308-320.
- Atick, J.J & Redlich, A.N.. (1992). What does the retina know about natural scenes? *Neural Computation*, 4(2), 196-210.
- Anzai, A., Ohzawa, I. & Freeman, R. D. (1999). Neural mechanisms for encoding binocular disparity: receptive field position versus phase. *Journal of Neurophysiology*, 82, 874-890.

- Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. *Sensory communication*, 217-234.
- Bell, A. J. & Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vision Research*, 37(23), 3327-3338.
- Berens, P. (2009). "CircStat: a MATLAB toolbox for circular statistics." *Journal of Statistical Software* 31(10): 1-21.
- Blakemore, C., Fiorentini, A. & Maffei, L. (1972). A second neural mechanism of binocular depth discrimination. *The Journal of physiology*, 226(3), 725-749.
- Bridge, H. & Cumming, B. G. (2001). Responses of macaque V1 neurons to binocular orientation differences. *The Journal of Neuroscience*, 21(18), 7293-7302.
- Bridge, H., Cumming, B. G. & Parker, A. (2001). Psychophysical evidence against the use of orientation disparity in the perception of slant. *Journal of Vision*, 1(3), 172. doi: 10.1167/1.3.172
- Burge J, & Geisler W,S, (2014). Optimal disparity estimation in natural stereo-images. *Journal of Vision*, 14(2):1, 1-18
- Chen D. & Li Z. (1998) A psychophysical experiment to test the efficient stereo coding theory *Theoretical aspects of neural computation*, 225-235, Springer-verlag
- Chen, Y. & Qian, N. (2004). A coarse-to-fine disparity energy model with both phase-shift and position-shift receptive field mechanisms. [10.1162/089976604774201596]. *Neural Computation*, 16, 1545-1577.
- Clarke, P., Donaldson, I. & Whitteridge, D. (1976). Binocular visual mechanisms in cortical areas I and II of the sheep. *The Journal of physiology*, 256(3), 509-526.
- Croner, L.J. & Kaplan, E. (1995). Receptive fields of P and M ganglion cells across the primate retina. *Vision research*, 35(1), 7-24.
- Cumming, B. G. (2002). An unexpected specialization for horizontal disparity in primate primary visual cortex. *Nature*, 418(6898), 633-636.
- Cumming, B.G. & DeAngelis, G. C. (2001). The physiology of stereopsis. *Annual Review of Neuroscience*, 24, 203-238.
- DeAngelis, G. C., Ohzawa, I. & Freeman, R. D. (1991). Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature*, 352(6331), 156-159.
- DeAngelis, G. C., Ohzawa, I. & Freeman, R. D. (1995). Neuronal mechanisms underlying stereopsis: how do simple cells in the visual cortex encode binocular disparity? *Perception*, 24, 3-3.
- DeAngelis, G. C. & Uka, T. (2003). Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *Journal of Neurophysiology*, 89(2), 1094-1111.
- DeValois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. [10.1016/0042-6989(82)90113-4]. *Vision Res.*, 22, 545-559.
- Ding, J. & Sperling, G. (2006). A gain-control theory of binocular combination. *Proceedings of the National Academy of Sciences of the United States of America*, 103(4), 1141-1146.
- Durand, J.-B., Zhu, S., Celebrini, S. & Trotter, Y. (2002). Neurons in parafoveal areas V1 and V2 encode vertical and horizontal disparities. *Journal of Neurophysiology*, 88(5), 2874-2879.
- Einhäuser, W., Kayser, C., König, P. & Körding, K. P. (2002). Learning the invariance properties of complex cells from their responses to natural stimuli. *European Journal of Neuroscience*, 15(3), 475-486.
- Faria, F da C. e C., Batista, J, & Araújo, H. (2013). Stereoscopic Depth Perception Using a Model Based on the Primary Visual Cortex. *PLoS ONE*, 8(12), e80745. doi: 10.1371/journal.pone.0080745
- Field, D. J. (1994). What is the goal of sensory coding? *Neural Computation*, 6(4), 559-601.
- Field, D.J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *JOSA A*, 4(12), 2379-2394.
- Field, D. J. & Chandler, D. M. (2012). Method for estimating the relative contribution of phase and power spectra to the total information in natural-scene patches. *Journal of the Optical Society of America A*, 29(1), 55-67.
- Fischer, S., Sroubek, F., Perrinet, L., Redondo, R. & Cristóbal, G., Self invertible log-Gabor wavelets, *International Journal of Computer Vision*, 75, 231-246.
- Fleet, D. J., Wagner, H. & Heeger, D. J. (1996). Neural encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision Research*, 36(12), 1839-1857.
- Freeman, R. D. & Ohzawa, I. (1990). On the neurophysiological organisation of binocular vision. *Vision Research*, 30, 1661-1676.

- Geisler, W.S., Najemnik, J. & Ing, A.D. (2009) Optimal stimulus encoders for natural tasks, *Journal of Vision*, 9(13):17, 1-16.
- Gerhard, H. E., Wichmann, F. A. & Bethge, M. (2013). How sensitive is the human visual system to the local statistics of natural images? *PLoS Computational Biology*, 9(1), e1002873.
- Gonzalez, F., Justo, M. S., Bermudez, M. A. & Perez, R. (2003). Sensitivity to horizontal and vertical disparity and orientation preference in areas V1 and V2 of the monkey. *Neuroreport*, 14(6), 829-832.
- Graham, D.J., Chandler, D.M. & Field, D.J. (2006). Can the theory of “whitening” explain the center-surround properties of retinal ganglion cell receptive fields? *Vision Research*, 46(18), 2901-2913.
- Greenwald, H.S. & Knill, D. C. (2009). Orientation disparity: A cue for 3d orientation? *Neural Computation*, 21(9), 2581-2604.
- Goutcher, R, Hunter, D.W. & Hibbard, P.B. (2013). Tuned Inhibitory Responses in Binocular Natural Images. *i-Perception*, 4(7), 484-484.
- Haefner, R. & Cumming, B. G. (2005). Spatial nonlinearities in V1 disparity-selective neurons. *Society for Neuroscience Abstracts*, 583.9.
- Hampton, D.R. & Kertesz, A. E. (1983). The extent of Panum's area and the human cortical magnification factor. *Perception*, 12(2), 161-165.
- Hansen, B.C. & Essock, E.A. (2004). A horizontal bias in human visual processing of orientation and its correspondence to the structural components of natural scenes. *Journal of Vision*, 4(12).
- Hawken, M.J. & Parker, A.J. (1987). Spatial properties of neurons in the monkey striate cortex. *Proceedings of the Royal society of London. Series B. Biological sciences*, 231(1263), 251-288.
- Heydt, R., Adorjani, C., Hányy, P. & Baumgartner, G. (1978). Disparity sensitivity and receptive field incongruity of units in the cat striate cortex. *Experimental Brain Research*, 31(4), 523-545.
- Hibbard, P. B. (2007). A statistical model of binocular disparity. *Visual Cognition*, 15(2), 149-165.
- Hibbard, P. B. (2008). Binocular energy responses to natural images. *Vision Research*, 48(12), 1427-1439.
- Himberg, J., Hyvärinen, A. & Esposito, F. (2004). Validating the independent components of neuroimaging time series via clustering and visualization. *Neuroimage*, 22(3), 1214-1222.
- Hong, G., Luo, M.R., & Rhodes, P.A. (2001). A study of digital camera colorimetric characterisation based on polynomial modelling.
- Howard, I. P. (2002). *Seeing in depth, Vol. 1: Basic mechanisms*: University of Toronto Press.
- Howard, I. P. & Rogers, B. J. (2002). *Seeing in depth, volume 2: Depth perception. Toronto: I Porteous.*
- Hoyer, P. O. & Hyvärinen, A. (2000). Independent component analysis applied to feature extraction from colour and stereo images. *Network: computation in neural systems*, 11(3), 191-210.
- Hubel, D.H. & Wiesel, T.N. (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, *Journal of Physiology*, 160, 106-154.
- Hurri, J. & Hyvärinen, A. (2003). Simple-cell-like receptive fields maximize temporal coherence in natural video. *Neural Computation*, 15(3), 663-691.
- Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *Neural Networks, IEEE Transactions on Neural Networks*, 10(3), 626-634.
- Hyvärinen, A., Sarela, J., & Vigário, R. (1999). *Spikes and bumps: Artefacts generated by independent component analysis with insufficient sample size*. Paper presented at the First International Workshop on Independent Component Analysis and Signal Separation.
- Hyvärinen, A., Hurri, J. & Hoyer, P. O. (2009). *Natural Image Statistics: A probabilistic approach to early computational vision* (Vol. 39): Springer-Verlag New York Inc.
- Hyvärinen, A. (2011). Independent Component Analysis: Recent Advances, *Philosophical Transactions of the Royal Society of London A*, 371: 20110534.
- Jammalamadaka SR, Sengupta A (2001). *Topics in Circular Statistics*. World Scientific.
- Laughlin, S. (1981). A simple coding procedure enhances a neuron's information capacity. *Zeitschrift für Naturforschung*, 36, 910-912.
- LeVay, S. & Voigt, T. (1988). Ocular dominance and disparity coding in cat visual cortex. *Visual Neuroscience*, 1(4), 395-414.
- Li, Z., & Atick, J.J. (1994a). Efficient stereo coding in the multiscale representation\*. *Network: Computation in Neural Systems*, 5(2), 157-174.
- Li, Z., & Atick, J. J. (1994b). Toward a theory of the striate cortex. *Neural Computation*, 6(1), 127-146.
- Liu, Y., Bovik, A. C. & Cormack, L. K. (2008). Disparity statistics in natural scenes. *Journal of Vision*, 8(11):19, 1-14. doi:10.1167/8.11.19

- Lyu, S. & Simoncelli, E. P. (2007). Statistical modeling of images with fields of Gaussian scale mixtures. *Advances in Neural Information Processing Systems*, 19, 945.
- May, K. A., Zhaoping, L. & Hibbard, P. B. (2012). Perceived direction of motion determined by adaptation to static binocular images. *Current Biology*, 22, 28-32.
- Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7(4), 308-313.
- Nelson, J., Kato, H., & Bishop, P. (1977). Discrimination of orientation and position disparities by binocularly activated neurons in cat striate cortex. *Journal of Neurophysiology*, 40(2), 260-283.
- Neri, P. (2005). A stereoscopic look at visual cortex. *Journal of Neurophysiology*, 93(4), 1823-1826.
- Nieder, A. & Wagner, H. (2000). Horizontal-disparity tuning of neurons in the visual forebrain of the behaving barn owl. *Journal of Neurophysiology*, 83(5), 2967-2979.
- Nikara, T., Bishop, P. & Pettigrew, J. (1968). Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex. *Experimental Brain Research*, 6(4), 353-372.
- Ninio, J. (1985). Orientational versus horizontal disparity in the stereoscopic appreciation of slant. *Perception*, 14(3), 305-314.
- Ohzawa, I., DeAngelis, G.C. & Freeman, R.D. (1990). Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science*, 249(4972), 1037-1041.
- Ohzawa, I., DeAngelis, G. C. & Freeman, R.D. (1996). Reply. *Trends in Neurosciences*, 19(9), 386.
- Ohzawa, I., Deangelis, G. C. & Freeman, R.D. (1997). Encoding of binocular disparity by complex cells in the cat's visual cortex. *Journal of Neurophysiology*, 77(6), 2879-2909.
- Okajima, K. (2004). Binocular disparity encoding cells generated through an infomax based learning algorithm. *Neural Networks*, 17(7), 953-962. doi: 10.1016/j.neunet.2004.02.004
- Olshausen, B.A. (2002). Sparse codes and spikes. Probabilistic models of the brain: Perception and neural function, 257-272.
- Olshausen, B.A. & Field, D.J. (1996). Natural image statistics and efficient coding\*. *Network: computation in neural systems*, 7(2), 333-339.
- Parker, A.J. (2007). Binocular depth perception and the cerebral cortex. *Nature Reviews Neuroscience*, 8(5), 379-391.
- Pettigrew, J.D. (1972). The neurophysiology of binocular vision. *Scientific American*.
- Pettigrew, J.D. & Konishi, M. (1976). Neurons selective for orientation and binocular disparity in the visual Wulst of the barn owl (*Tyto alba*). *Science (New York, NY)*, 193(4254), 675.
- Porrill, J., Ivins, J.P. & Frisby, J.P. (1999). The variation of torsion with vergence and elevation. *Vision research*, 39(23), 3934-3950.
- Poggio, G.F. (1991). Physiological basis of stereoscopic vision. *Vision and visual dysfunction. Binocular vision*, 224-238.
- Poggio, G.F. & Fischer, B. (1977). Binocular interaction and depth sensitivity of striate and prestriate cortex of behaving rhesus monkey. *J. Neurophysiol.*, 40, 1392-1405.
- Poggio, G.F., Gonzalez, F. & Krause, F. (1988). Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *The Journal of Neuroscience*, 8(12), 4531-4550.
- Poggio, G.F., Motter, B. C., Squatrito, S. & Trotter, Y. (1985). Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms. [10.1016/0042-6989(85)90065-3]. *Vision Research*, 25, 397-406.
- Poggio, G.F. & Talbot, W.H. (1981). Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey. *The Journal of physiology*, 315(1), 469-492.
- Prince, S.J. P., Cumming, B.G. & Parker, A.J. (2002). Range and mechanism of encoding of horizontal disparity in macaque V1. *Journal of Neurophysiology*, 87, 209-221.
- Prince, S.J.P., Pointon, A., Cumming, B.G. & Parker, A.J. (2002). Quantitative analysis of the responses of V1 neurons to horizontal disparity in dynamic random-dot stereograms. *Journal of Neurophysiology*, 87(1), 191-208.
- Qin, D., Takamatsu, M. & Nakashima, Y. (2006). Changing Binocular Fusional Area with Retinal Shift in Binocular Vision. *Journal of Light & Visual Environment*, 30(1), 29-33.
- Read, J.C.A. & Cumming, B.G. (2003). Testing quantitative models of binocular disparity selectivity in primary visual cortex. *Journal of Neurophysiology*, 90, 2795-2817.
- Read, J.C.A. & Cumming, B.G. (2004a). Ocular dominance predicts neither strength nor class of disparity selectivity with random-dot stimuli in primate V1. *J. Neurophysiol.*, 91, 1271-1281.
- Read, J.C.A. & Cumming, B.G. (2004b). Understanding the cortical specialization for horizontal disparity. *Neural Computation*, 16(10), 1983-2020.

- Read, J.C.A. & Cumming, B.G. (2007). Sensors for impossible stimuli may solve the stereo correspondence problem. *Nature Neuroscience*, 10(10), 1322-1328.
- Ringach, D.L. (2002) Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex, *Journal of Neurophysiology*, 88, 455-463.
- Ringach, D. L., Sapiro, G., & Shapley, R. (1997). A subspace reverse-correlation technique for the study of visual neurons. *Vision Research*, 37(17), 2455-2464.
- Roe, A.W., Parker, A.J., Born, R.T. & DeAngelis, G.C. (2007). Disparity channels in early vision. *The Journal of Neuroscience*, 27(44), 11820-11831.
- Shannon, C.E. (1949). Communication in the presence of noise. *Proceedings of the IRE*, 37(1), 10-21.
- Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. [10.1146/ANNUREV.NEURO.24.1.1193]. *Annu. Rev. Neurosci.*, 24, 1193-1216.
- Srinivasan, M., Laughlin, S. & Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 216(1205), 427-459.
- Tong, L., Guido, W., Tumosa, N., Spear, P.D. & Heidenreich, S. (1992). Binocular interactions in the cat's dorsal lateral geniculate nucleus, II: Effects on dominant-eye spatial-frequency and contrast processing. *Visual Neuroscience*, 8(06), 557-566.
- Tsao, D.Y., Conway, B. R. & Livingstone, M. S. (2003). Receptive Fields of Disparity-Tuned Simple Cells in Macaque V1. *Neuron*, 38(1), 103-114. doi: [http://dx.doi.org/10.1016/S0896-6273\(03\)00150-8](http://dx.doi.org/10.1016/S0896-6273(03)00150-8)
- Tyler, C.W. & Sutter, E. E. (1979). Depth from spatial frequency difference: An old kind of stereopsis? *Vision Research*, 19(8), 859-865.
- van der Schaaf, A. & van Hateren, J.H. (1996). Modelling the power spectra of natural images: Statistics and information. *Vision Research*, 36(17), 2759-2770.
- van Hateren, J.H. & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1394), 359-366.
- Wieniawa-Narkiewicz, E., Wimbome, B., Michalski, A. & Henry, G. (1992). Area 21a in the cat and the detection of binocular orientation disparity. *Ophthalmic and Physiological Optics*, 12(2), 269-272.
- Xu, L., & Jordan, M.I. (1996). On convergence properties of the EM algorithm for Gaussian mixtures. *Neural Computation*, 8(1), 129-151.
- Zoran, D. & Weiss, Y. (2012). Natural Images, Gaussian Mixtures and Dead Leaves, *Advances in Neural Information Processing Systems 25*, F. Pereira, C.J.C Burges, L. Boutou and K.Q. Weinberger (eds),
- Zhao, L. (2014). *Understanding vision: theory, models, and data*: Oxford University Press.