Efficient Stereo Coding in the Multiscale Representation [1]

To be published in *Network:* Computation in Neural Systems

Zhaoping Li and Joseph J. Atick
The Rockefeller University
1230 York Avenue
New York, NY 10021, USA

Abstract

Stereo images are highly redundant; the left and right frames of typical scenes are very similar. We explore the consequences of the hypothesis that cortical cells — in addition to their multiscale coding strategies (Li and Atick 1994a) — are concerned with reducing binocular redundancy due to correlations between the two eyes. We derive the most efficient coding strategies that achieve binocular decorrelation. It is shown that multiscale coding combined with a binocular decorrelation strategy leads to a rich diversity of cell types. In particular, the theory predicts monocular/binocular cells as well as a family of disparity selective cells, among which one can identify cells that are tuned-zero-excitatory, near, far, and tuned inhibitory. The theory also predicts correlations between ocular dominance, cell size, orientation, and disparity selectivities. Consequences on cortical ocular dominance column formation from abnormal developmental conditions such as strabismus and monocular eye closure are also predicted. These findings are compared with physiological measurements.

[1] Work supported in part by a grant from the Seaver Institute.

1. Introduction

It has become clear from recent works that specific properties of natural scenes can predict many neural computational strategies in the visual pathway (Atick and Redlich 1990, 1992; Atick et al. 1992, 1993; Field 1987). Unlike random collections of pixels, natural images possess a multitude of regularities that can be quantified by well-defined statistical measures. For example, due to the morphological consistency of objects, nearby pixels in natural images tend to be very similar in their visual appearance, giving a luminosity profile which changes gradually in space and only abruptly at edges. Such gradual change in the signal also occurs in the temporal and chromatic domains where there is continuity and smoothness. This means that natural images possess a high degree of spatio-temporal and chromatic correlations, and that a pixel by pixel representation of such scenes by the photoreceptors is highly redundant and thus inefficient.

One can argue that efficiency of information representation has evolutionary and cognitive advantages (Barlow 1961, Atick 1992, 1994). This leads to a predictive principle for sensory processing, namely, the principle of *redundancy reduction* which advocates that a major goal of early visual processing is to recode incoming signals into a redundancy-reduced representation, subject to identifiable hardware constraints.

This principle, modified appropriately to take noise into account, has been shown to provide a quantitative theory of retinal processing (Atick and Redlich 1990, 1992). Actually, as the first stage in the visual pathway, the retina can only eliminate the simplest type of redundancy — the pairwise pixel correlations — in an image. This predicts the spatio-chromatic receptive fields of ganglion cells that compare well with experimental data (Atick and Redlich 1992; Atick et al. 1992).

There are other types of regularities in natural images that we believe the visual system beyond the retina takes advantage of to build more efficient representation of the world[2]. One such regularity manifests itself in the laws of perspective transformation — an image of a scene at one distance can predict much of the image of the same scene at another distance. This means as we move in the 3D world, the successive images entering the visual system are highly redundant in the sense that most of the changes in them are predictable from perspective transformations, and they do not represent genuine changes in the input.

In view of this "perspective" redundancy, we have recently proposed that the preprocessing goal of the early visual cortex is to produce a representation where the action of perspective transformations (scaling) is manifest[3] (Li and Atick 1994a). In the retina, the changes in the neural response with the viewing distance are very complicated. We have shown that the retinal output, without compromising the efficiency achieved by elimination of pairwise correlations, can be transformed into a representation where neural response changes with viewing distance are very simple (Li and Atick 1994a). This, so called *multiscale* representation, requires remapping the visual field into multiple retinotopic maps identical in all respects except for the densities and (receptive field) sizes of their sampling nodes. As

[2]For a more complete discussion of such regularities and how the visual system may take advantage of them, see Atick 1994.

[3]Such representation — although not strictly redundancy-reduced in itself — is a major step towards redundancy reduction when followed by an attentional mechanism that takes advantage of the manifest action of the spatial group to compensate for viewing conditions, and hence to produce a representation where the same intrinsic neural activation pattern represents an object as it recedes or approaches the viewer.

an object recedes or approaches the viewer, the neural activation pattern it evokes in this representation remains intrinsically the same but shifts its locus from one cell or scale group to another.

Our previous work on cortical processing ignored the binocular nature of the visual input for simplicity. However, it is at the cortex where inputs from the two eyes are first combined. Binocular vision introduces another input regularity since the left and right images of the world correspond to slightly shifted views of the same scene and are thus highly correlated. Although the retina eliminates pairwise correlations within a given eye, correlations between eyes persist as the signal enters the cortex. One then expects that accounting for the binocular redundancy to lead to additional computational strategies for cortical cells. In this paper, we make the hypothesis that cortical cells — in addition to their multiscale computational strategies predicted earlier in Li and Atick (1994a) — combine signals from the two eyes in such a way to eliminate the inter-ocular correlations. This redundancy-reduced stereo coding is thus efficient [4].

A precise theoretical prediction of stereo coding strategies requires knowledge of the ocular correlation function in the input ensemble. Ocular correlations depend not only on the inter-eye distance, eye alignment, and the distribution of object distances in the input ensemble, but also on whether the visual system actively fixates on visual objects. We have measured the inter-ocular correlation function for a stereo camera system with a distance between the two lenses close to human inter-eye distance (see appendix for details of this measurement). However, this camera has a static fixation distance and can not imitate the active human fixation. Without available information, we have to model the human inter-ocular correlation by extrapolating the results from the stereo camera with some assumptions on the active fixation process.

We find from the measurement that the ratio inter-ocular correlation/intra-ocular correlation depends on orientation and decays as spatial scale gets smaller. Accordingly, a coupling between stereo coding and spatial scale/orientation emerges in a multiscale representation. In particular, we identify cells that are selective to near zero, divergent, and convergent disparities, as well as correlations between cell receptive field size, orientation, disparity selectivity, and ocular dominance. For example, the theory predicts that for disparity sensitive cells, the smaller ones are selective to smaller disparities, and that the larger one are selective to larger disparities and are more likely to be monocular. These results have been observed in some experiments (Poggio 1992, Ferster 1981). While this work derives cortical stereo coding from the principle of redundancy reduction, other relevant works on computational and development models of cortical stereo codings can be found in (Marr and Poggio 1979, Poggio and Poggio 1984, Miller et al. 1989, Miller and Stryker 1990, Blake and Wilson 1991, Berns et al. 1993).

2. Efficient binocular coding in the multiscale representation

Let $S^L(\underline{x}_n)$ and $S^R(\underline{x}_m)$ be the photoreceptor activities at retinal locations $\underline{x}_n$ and $\underline{x}_m$ in the left and right eyes respectively. A light source at the fixation plane will induce $S^L(\underline{x}_n)$

[4] It is important to point out that redundancy reduction is not inconsistent with combining signals from both eyes to extract depth information. Although the inputs from the two eyes are correlated, it is the difference between them that carries the stereo information. The coding that reduces inter-ocular redundancy highlights the non-redundant stereo information. In other words by not coding "sameness" in the inputs from the two eyes, more computational resources can be used to code stereo information.

and $S^R(\underline{x}_m)$ at the same location $\underline{x}_n = \underline{x}_m$ in the two eyes; otherwise photoreceptors at different locations in the two eyes are excited. The autocorrelator of the signals is

$$R_x^{ab}(\underline{x}_n, \underline{x}_m) \equiv < S^a(\underline{x}_n) S^b(\underline{x}_m) >, \tag{1}$$

where $a, b = L, R$ and brackets denote ensemble average. Assuming a uniform sampling grid and translation invariance within a local area, in the sense that $R_x^{ab}(\underline{x}_n, \underline{x}_m) = R_x^{ab}(\underline{x}_n - \underline{x}_m)$, then the correlation can be captured by a simpler quantity $R^{ab}(\underline{f}) \equiv < \tilde{S}^a(\underline{f}) \tilde{S}^b(-\underline{f}) >$, where $\tilde{S}^a(\underline{f})$ is the Fourier transform of $S^a(x)$ at frequency $\underline{f}$. $R^{ab}(\underline{f})$ is the *power spectrum* of visual inputs for $a = b$, and the *bispectrum* when $a \neq b$. The power spectrum for natural scenes has been measured by many people (Field 1987, Ruderman and Bialek 1993). These measurements show that $R^{aa}(\underline{f}) \equiv R(\underline{f}) \propto 1/|\underline{f}|^2$. The new ingredient that is necessary here is the inter-eye correlation function $R^{LR}(\underline{f})$.

Thus we can write the correlation matrix in ocular space explicitly as:

$$\begin{pmatrix} R^{LL}(\underline{f}) & R^{LR}(\underline{f}) \\ R^{RL}(\underline{f}) & R^{RR}(\underline{f}) \end{pmatrix} \equiv \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} R(\underline{f}), \tag{2}$$

where symmetry between left and right eyes is assumed, and $r < 1$. In general we expect the parameter r to be a function of frequency $\underline{f}$ and it can be complex, $r = r(\underline{f})$. A complex[5] $r = |r| e^{i\psi}$ can be turned real by redefining the right eye signal with $\tilde{S}^R(\underline{f}) = e^{-i\psi} \tilde{S}^R(\underline{f})$. This ψ can be understood as the mean phase disparity of the visual inputs. However, stereo information is captured in the variations of disparities around this mean, which is analogous to the mean light level and does not by itself convey the 3D information in the inputs. Hence, analogous to adaptation to mean light level, the visual system should be able to adapt to this mean disparity, by, e.g., vergence control to align the two images in the eyes, such that ψ could be effectively brought close to zero [6]. This paper only deals with stereo coding strategies assuming that the image alignment is achieved and $r(\underline{f})$ is real.

We have measured $r(\underline{f})$ for an ensemble of stereo images (see appendix); Figure 1 gives the results for a horizontal and vertical slices $\underline{f} = (f, 0)$ and $\underline{f} = (0, f)$. One can see that $r(\underline{f}) \approx 1$ for small $|\underline{f}|$, suggesting strong ocular correlation when one looks at the inputs in coarse spatial resolution. In particular, the mean light levels, i.e., $\underline{f} = 0$, to the two eyes are very similar. On the other hand, the inputs to the two eyes are quite different when one looks at the input in detail, causing $r(\underline{f})$ to decay to zero at large $\underline{f}$.

There are two kinds of redundancy that contribute to the matrix in equation (2), one is the binocular redundancy — largely in the inter-ocular correlations — and the other is the redundancy in space, represented in $R(\underline{f})$. The binocular redundancy can be reduced by eliminating the inter-eye correlations by introducing new variables:

$$S^+ = \frac{1}{\sqrt{2}}(S^L + S^R), \tag{3}$$

$$S^- = \frac{1}{\sqrt{2}}(S^L - S^R). \tag{4}$$

[5] In that case, the correlation matrix is $\begin{pmatrix} 1 & r \\ r^* & 1 \end{pmatrix} R(\underline{f})$.

[6] This vergence control to accommodate for the mean phase disparity ψ may have to be carried out by different amount when attention is directed at different spatial scales.

This linear transform $(S^L, S^R) \rightarrow (S^+, S^-)$ is consistent with biological constraints indicating linear combination of inputs from the two eyes in the striate simple cells (Ohzawa and Freeman 1986a). The ocular *summation* S^+ and the ocular *opponency* S^- signals are the building blocks from which cortical cell responses can be constructed, as we shall see below. Unlike S^R and S^L, S^+ and S^- are uncorrelated, as $< S^+S^- >= \frac{1}{2}(< S^LS^L > - < S^RS^R > + < S^LS^R > - < S^RS^L >) = 0$ assuming left-right symmetry. The power spectra for $S^\pm$ are

$$R^\pm(\underline{f}) = (1 \pm r)R(\underline{f}). \tag{5}$$

After eliminating inter-ocular correlations, one only needs to eliminate the spatial correlations in S^+ and S^-. As we know, pixels in a spatially white noise signal are uncorrelated. The power spectrum of such white noise is flat or independent of $\underline{f}$, by definition. The spatial correlations in $S^\pm$ are reflected in the non-flatness of the power spectrum $R^\pm(\underline{f})$. Therefore, spatial correlations can be eliminated by applying the whitening filters $K^\pm(\underline{f}) \propto (R^\pm(\underline{f}))^{-1/2} = |f|/\sqrt{1 \pm r}$ to the signals $S^\pm$:

$$O^\pm(\underline{f}) = K^\pm(\underline{f})S^\pm(\underline{f}), \tag{6}$$

such that the outputs $O^\pm(\underline{f})$ have flat power spectra $< O^\pm(\underline{f})O^\pm(-\underline{f}) >= (K^\pm(\underline{f}))^2 < S^\pm(\underline{f})S^\pm(-\underline{f}) >= (K^\pm(\underline{f}))^2R^\pm(\underline{f}) = constant$ (Atick and Redlich 1992). Intuitively, spatial redundancy is reduced by $K^\pm(\underline{f}) \propto |f|$ amplifying input spatial changes or edges; binocular redundancy is reduced by ocular filters $K^- > K^+$ amplifying the binocular "edge" S^-.

Before we can exhibit the predicted binocular receptive fields, there are three issues that have to be addressed. All three have been addressed in previous works (Atick et al. 1993, Atick and Redlich 1992, Li and Atick 1994a), here we briefly discuss the main conclusions. The first issue has to do with the *non-uniqueness* of the process of decorrelation (Atick et al. 1993). One can multiply the signal (O^+, O^-) by any rotation matrix and still maintain a decorrelated output. Although, all transformations constructed in this manner are equivalent in their decorrelation properties, they differ drastically in the degree by which they transform the original signal (S^L, S^R). In Atick et al. (1993) it was proposed that the transformation favored biologically is the one that requires the least deformation or the least change to the original signal (the so called most "local" transformation) while decorrelating the output. This was shown to give successful predictions for color as well as spatial processings and we will continue to make this proposal in the ocular space. It can be shown[7] that, within a constant factor, the new variables and the corresponding transformation are:

$$\begin{array}{lll} O^1 = (O^+ + O^-), & & K^1 = K^+ + K^-, \\ O^2 = (O^+ - O^-); & & K^2 = K^+ - K^-. \end{array} \tag{7}$$

The second issue has to do with *noise*. The filter $K^\pm(\underline{f}) = 1/\sqrt{R^\pm(\underline{f})} \propto |\underline{f}|$ performs gain control and amplifies the signals at high frequency. However, in reality, inputs contain noise from various sources. This leads to undesirable consequences since, at high frequencies the noise power, unlike the signal power (which is decaying like $\sim 1/|\underline{f}|^2$), is not becoming small. Whitening all the way to the highest frequency leads to unacceptable levels of noise in the output. Thus, for decorrelation to be useful, it must be combined with a noise smoothing strategy ensuring that no significant input noise is passed to the next stage. This strategy

[7]For derivation see Atick et al. 1993.

modifies the filters $K^{\pm}(\underline{f})$ such that $K^{\pm}(\underline{f})$ decays quickly at high frequencies before the noise overwhelms the signal (Atick and Redlich 1992), while at low frequency — where noise is not dominant — the filter still whitens the signal. The modified filters $K^{\pm}(\underline{f})$ look like those in Figure 2A (see Atick and Redlich 1992 for details). Their general features are:

$$\begin{array}{lcll} K_s^- & > & K_s^+ & \text{when } r(\underline{f}) \text{ is significant, } |\underline{f}| \text{ is small where signal/noise is high} \\ K_s^- & < & K_s^+ & \text{when } r(\underline{f}) \text{ is significant, } |\underline{f}| \text{ is large where signal/noise is low} \\ K_s^- & \approx & K_s^+ & \text{when } r(\underline{f}) \approx 0 \text{ at very large } |\underline{f}|. \end{array} \tag{8}$$

This is because the ocular opponency signal power $R^- = (1-r)R$ is always smaller than the ocular summation signal power $R^+ = (1+r)R$. Hence at low $|\underline{f}|$ when signal dominates the noise, K^- is relatively large to amplify the smaller opponency signal S^-; At high $\underline{f}$ when noise is relatively large, R^- is overwhelmed before R^+ by the noise, hence K^- decays before K^+, forcing $K^- < K^+$. At very high $\underline{f}$, $K^+ \approx K^-$ because the interocular correlation $r \approx 0$, giving $R^+ \approx R^-$. These features will dictate the details of stereo coding as will be shown in section 3 and 4.

The third issue is the *multiscale* coding of the cells. The cells have to perform other tasks that are a priori non-binocular in nature and that must be taken into account before a comparison with real cells can be made. In (Li and Atick 1994a) it was argued that cortical cells form what is called a multiscale representation in space. This means that instead of having one filter $K^{\pm}$ at each visual location, each location is covered by multiple filters of different sizes. In one dimensional space, these filters are obtained by breaking up the original filter into one lowpass filter and a set of bandpass filters covering the full range of frequency as shown in Fig. 2B. More explicitly:

$$\text{Lowpass: } \mathbf{K_0^{\pm}}(f) \equiv K^{\pm}(f) e^{-f^2/2f_l^2}, \tag{9}$$

$$\text{Bandpass: } \mathbf{K_s^{\pm}}(f) \equiv K^{\pm}(f) e^{i\theta^{\pm}} e^{-(\log(f/f_s))^2/2\sigma^2}. \tag{10}$$

where f_l models the cut-off of the lowpass filter, and f_s is the center frequency of the bandpass filter at scale $s = 1, 2, 3, ...$ (s is the scale index, and f_s increases with s), $\sigma = 1.6$ octave/2 sets the bandwidth of the bandpass filters to 1.6 octaves, as is observed in experiments (e.g., De Valois et al. 1982; Andrews and Pollen 1979), $\theta^{\pm}$ is the phase of the receptive fields for the bandpass cells, giving for example, even or odd receptive fields (bar or edge detectors in 2D) when $\theta^{\pm} = 0$ or 90^o (Li and Atick 1994a). The bold faced notation $\mathbf{K_s^{\pm}}$ is used to represent the filter as a vector, interpreting the phases $\theta^{\pm}$ as the vector direction, and the filter strength $|\mathbf{K_s^{\pm}}|$ the vector amplitude. Both $e^{-f^2/2f_l^2}$ and $e^{-(\log(f/f_s))^2/2\sigma^2}$ are models to give smooth tuning curves of spatial frequency sensitivities, instead of the sharp edges in tuning curves originally derived in Li and Atick (1994a); this is done for convenience. To obtain the two dimensional filters, we only need to apply a lowpass filter in the orthogonal spatial direction (Li and Atick 1994a), giving un-oriented cells from lowpass filters in both directions ($\mathbf{K_0^{\pm}}$) and oriented cells ($\mathbf{K_s^{\pm}}$) by a bandpass in one direction and a lowpass in the other. In our discussion, we will continue to suppress the lowpass in the orthogonal direction for simplicity.

We are now ready to show what the actual filters look like. In ocular space, $\mathbf{K_s^+}$ acts on the summation of left and right eye signals $\mathbf{L} + \mathbf{R}$, while $\mathbf{K_s^-}$ acts on the difference $\mathbf{L} - \mathbf{R}$, where $\mathbf{L}$ and $\mathbf{R}$ are the basis vectors in the ocular space. The filters $K^{1,2} = K^+ \pm K^-$ in

ocular space and multiscale representation are explicitly:

$$\begin{array}{rcll} \mathbf{K_s^1} & = & \mathbf{L}(\mathbf{K_s^+} + \mathbf{K_s^-}) & +\mathbf{R}(\mathbf{K_s^+} - \mathbf{K_s^-}), \\ \mathbf{K_s^2} & = & \mathbf{L}(\mathbf{K_s^+} - \mathbf{K_s^-}) & +\mathbf{R}(\mathbf{K_s^+} + \mathbf{K_s^-}). \end{array} \tag{11}$$

Without loss of generality, from here on we concentrate on $\mathbf{K_s^1}$ to construct the cell types. The contribution of the left and right eyes respectively to this cell are $\mathbf{K_s^L} \equiv |\mathbf{K_s^L}|e^{i\phi^L} \equiv \mathbf{K_s^+} + \mathbf{K_s^-}$ and $\mathbf{K_s^R} \equiv |\mathbf{K_s^R}|e^{i\phi^R} \equiv \mathbf{K_s^+} - \mathbf{K_s^-}$. The receptive field forms in space in both eyes are

$$K_s^{L,R}(x) = \int df |K_s^{L,R}(f)| cos(2\pi f x + \phi^{L,R}). \tag{12}$$

The $\phi^{L,R}$ are the phases of the receptive fields. The optimal phase disparity of the cell $\Delta\phi \equiv \phi^L - \phi^R$ is the angle between the two vectors $\mathbf{K_s^L}$ and $\mathbf{K_s^R}$. Fig. 3 shows graphically how the receptive fields of both eyes are constructed as vector summations from $\mathbf{K_s^+}$ and $\pm\mathbf{K_s^-}$.

3. Binocular coding in a visual system with a constant fixation distance

In this section, we explore the predictions of (11) for a visual system which does not change its fixation distance. This system is like our stereo camera whose optical axes for the two lenses are fixed. Hence the ocular correlation $r(\underline{f})$ of such a system can be approximated by what we measured (see Fig. 1).

From equation (11) and Fig. 3, we can see that the relative sizes of $|\mathbf{K_s^+}|$ and $|\mathbf{K_s^-}|$, and their relative angle $\Delta\theta = \theta^- - \theta^+$, determine the relative strengths of the filters $|\mathbf{K_s^L}|$ and $|\mathbf{K_s^R}|$ as well as the optimal phase disparity $\Delta\phi$. In particular, since $\mathbf{K_s^{L,R}} = \mathbf{K_s^+} \pm \mathbf{K_s^-}$, a dominance of the $\mathbf{K_s^+}$ filter makes the contribution of the two eyes to such a cell relatively aligned in phase, while a dominance of the $\mathbf{K_s^-}$ filter makes them out of phase. Hence the optimal phase disparities are (see Fig. 3)

$$\begin{array}{rcll} \Delta\phi & > & 90^o & \text{if } \mathbf{K_s^-} > \mathbf{K_s^+} \\ \Delta\phi & < & 90^o & \text{if } \mathbf{K_s^-} < \mathbf{K_s^+}, \end{array} \tag{13}$$

Since the relative sizes of $\mathbf{K_s^\pm}$ depend on spatial scale, there will be a coupling between cell sizes and optimal disparities.

To obtain $\mathbf{K_s^\pm}$, we recall from (8) that there are usually three frequency regions of relative strengths for $|\mathbf{K_s^-}|$ and $|\mathbf{K_s^+}|$: When $r \not\approx 0$, $|\mathbf{K_s^-}| > |\mathbf{K_s^+}|$ at high signal (low $\underline{f}$) region where amplification of lower signal R^- is needed, $|\mathbf{K_s^-}| < |\mathbf{K_s^+}|$ at lower signal region (higher $\underline{f}$) where noise forces the filters $\mathbf{K_s^\pm}$ to decay, and finally, $|\mathbf{K_s^-}| \approx |\mathbf{K_s^+}|$ when the interocular correlation $r(\underline{f})$ diminishes. However, this simple visual system has $r(\underline{f}) \approx 0$ at $|\underline{f}| \geq 1.5$ cycle/degree, simply because the input depth variation for this visual system is of the order ~ 1 degree. Hence $|\mathbf{K_s^-}| \approx |\mathbf{K_s^+}|$ is enforced for $|\underline{f}| \geq 1.5$ cycle/degree. If we assume that this simple visual system resembles human central vision in normal illumination in the sense that at low spatial frequencies no higher than 1.5 cycle/degree, the signal is still large enough such that the $|\mathbf{K_s^-}| < |\mathbf{K_s^+}|$ region can not be realized (see Fig. 4A). Consequently, the binocular coding of this system can be divided into the following two different scale regions.

i: $|\mathbf{K}_\mathbf{s}^+| < |\mathbf{K}_\mathbf{s}^-|$ — *large tuned inhibitory, near-far cells.*

Small spatial frequency of this region leads to large cell sizes. From equation (13), the optimal phase disparity of this cell is always larger than 90^o. When $\mathbf{K}_\mathbf{s}^+ \parallel \mathbf{K}_\mathbf{s}^-$, the two receptive fields from the two eyes are completely out of phase (Fig.4Ba, 4Bb, 4Bc). These cells are thus inhibited by stimuli of zero disparities and are tuned inhibitory cells. They are likely to be ocularly unbalanced unless $\mathbf{K}_\mathbf{s}^-$ strongly dominates[8] $\mathbf{K}_\mathbf{s}^+$. When $\mathbf{K}_\mathbf{s}^+$ is not parallel to $\mathbf{K}_\mathbf{s}^-$, the cell is not completely out of phase and can resemble near/far cells (Fig. 4Bd, 4Be). Depending on the relative phases of the two receptive fields, some near/far cells can be binocularly balanced, as when $\mathbf{K}_\mathbf{s}^+ \perp \mathbf{K}_\mathbf{s}^-$ (see Fig.3, Fig.4Bd). (Fig.4B and 5 only exhibit the far cells, the near cells can be obtained by either using the $\mathbf{K}_\mathbf{s}^\mathbf{2}$ filter in (11) or using a negative $\Delta\theta$.)

ii: $|\mathbf{K}_\mathbf{s}^+| \approx |\mathbf{K}_\mathbf{s}^-|$ *when* $r(\underline{f}) \approx 0$ *at high* $\underline{f}$*— small monocular cells*

Higher optimal frequency gives smaller cell sizes to this group. When $\mathbf{K}_\mathbf{s}^+ \parallel \mathbf{K}_\mathbf{s}^-$, equation (11) suggests that these cells have $\mathbf{K^L} = \mathbf{K}_\mathbf{s}^+ + \mathbf{K}_\mathbf{s}^- \approx 2\mathbf{K}_\mathbf{s}^+$ and $\mathbf{K^R} = \mathbf{K}_\mathbf{s}^+ - \mathbf{K}_\mathbf{s}^- \approx 0$. The two eye contributions are thus very unbalanced and the cells are monocular (Fig. 4Bf, 4Bg). In fact, in this frequency region, the inter-ocular correlation $r(\underline{f})$ is so small that the binocular redundancy reduction step $S^\pm = \frac{1}{\sqrt{2}}(S^L \pm S^R)$ and the subsequent multiplexing (equation (7)) are essentially unnecessary. Monocular cells are therefore the most natural choice.

Although the predicted tuned inhibitory cells, near/far cells, and the variance in ocular dominance have been observed in some experiments (Hubel and Wiesel 1970, Poggio and Fischer 1977, Ferster 1981), the theory does not predict binocular cells tuned to near zero disparities in this visual system. Tuned-zero-excitatory cells are most excited by stimuli aligned in both eyes. They appear within this theory only in the scale region where $\mathbf{K}_\mathbf{s}^+ > \mathbf{K}_\mathbf{s}^-$ (see equation (13)). Intuitively, summing signals from the two eyes can help to combat noise, which is significant at high spatial frequencies. However, this noise smoothing strategy is only effective if the signals — the edges — in two eyes are spatially correlated, which is not the case in this visual system at high spatial frequencies or small scales. However, the signal integration by tuned excitatory cells will be effective when the images in the two eyes are aligned. This can be realized by a system which fixates dynamically on visual objects such that images are aligned at a high spatial resolution (i.e., high $\underline{f}$), as will be shown in the next section.

4. A dynamic fixating visual system — Classification of predicted cell types and comparison with experiments

We have so far explored the theory only in simple visual systems which do not actively fixate according to object distances. Object depths in such systems distribute randomly with a large standard deviation around a mean depth which may (or may not) be the static fixation distance. In reality, however, active fixation reduces the object depths to within a small range in a local visual area around the fixation point, giving much smaller standard deviation of depth. At this point, we do not know enough about what initiates the vergence

[8]The dominance of $\mathbf{K}_\mathbf{s}^-$ over $\mathbf{K}_\mathbf{s}^+$ can happen at very small $\underline{f}$ where a relatively small signal R^- compared to R^+ causes a much stronger amplification $\mathbf{K}_\mathbf{s}^-$ relative to $\mathbf{K}_\mathbf{s}^+$. However, there the cell sizes are large and thus such cells are rare.

eye movement. However, its effect is to modify the binocular input ensemble and we will use it to model the ocular correlations.

This dynamic fixating system has higher ocular correlations. In extreme cases, visual inputs from objects located exclusively on the horoptor, i.e., zero disparities, give complete inter-ocular correlation. If the input ensemble consists of visual objects of disparities within the range Δx, it is only manifested as misaligned stimuli in the two eyes at spatial frequency $|\underline{\mathrm{f}}| \geq 1/\Delta x$. Therefore, the ocular correlation $r(\underline{\mathrm{f}}) \ll 1$ only for $|\underline{\mathrm{f}}| \geq 1/\Delta x$. As argued above, the disparity range or disparity variation Δx in a dynamic fixating system should be much smaller than that of static fixating system[9]. Accordingly, the ocular correlation $r(\underline{\mathrm{f}})$ is substantial up to a much higher spatial frequency $\underline{\mathrm{f}}$.

The consequence of a higher ocular correlation is the creation of the region where $\mathbf{K}_{\mathbf{s}}^{+} > \mathbf{K}_{\mathbf{s}}^{-}$, at high $|\underline{\mathrm{f}}|$ where both $r(\underline{\mathrm{f}})$ and the noises are significant (see equation (8)). Hence, in addition to the large tuned inhibitory, near/far cells when $\mathbf{K}_{\mathbf{s}}^{-} > \mathbf{K}_{\mathbf{s}}^{+}$ and the small monocular cells when $\mathbf{K}_{\mathbf{s}}^{+} \approx \mathbf{K}_{\mathbf{s}}^{-}$ at high $|\underline{\mathrm{f}}|$, there will be following cell classes (Fig. 5).

iii: $|\mathbf{K}_{\mathbf{s}}^{+}| > |\mathbf{K}_{\mathbf{s}}^{-}|$ — *tuned excitatory cells.*

By (13), the optimal phase disparity of the cell is small. When $\mathbf{K}_{\mathbf{s}}^{+} \parallel \mathbf{K}_{\mathbf{s}}^{-}$, the two receptive fields are completely aligned (Fig. 5f,5g, 5j, 5k), giving tuned-zero-excitatory cells (Poggio 1992). If $\mathbf{K}_{\mathbf{s}}^{+} \gg \mathbf{K}_{\mathbf{s}}^{-}$ (which happens when $r(\underline{\mathrm{f}})$ is large enough) or $\mathbf{K}_{\mathbf{s}}^{+} \perp \mathbf{K}_{\mathbf{s}}^{-}$, then $|\mathbf{K}_{\mathbf{s}}^{\mathbf{L}}| \approx |\mathbf{K}_{\mathbf{s}}^{\mathbf{R}}|$ and the contributions from the two eyes are comparable, giving ocularly balanced cells. As the frequency in this region is higher than that of $\mathbf{K}_{\mathbf{s}}^{-} > \mathbf{K}_{\mathbf{s}}^{+}$ region, the tuned excitatory cell sizes are smaller than other disparity selective cells.

iv: $|\mathbf{K}_{\mathbf{s}}^{+}| \approx |\mathbf{K}_{\mathbf{s}}^{-}|$ *and* $r(\underline{f}) \not\approx 0$ — *monocular cells and near/far cells* .

This happens at the transition from $\mathbf{K}_{\mathbf{s}}^{-} > \mathbf{K}_{\mathbf{s}}^{+}$ to $\mathbf{K}_{\mathbf{s}}^{-} < \mathbf{K}_{\mathbf{s}}^{+}$, hence the ocular correlation is still significant. Therefore, unlike the monocular cells of class ii, the redundancy reduction step $S^{\pm} = \frac{1}{\sqrt{2}}(S^{L} \pm S^{R})$ and the subsequent multiplexing (equation (7)) are necessary. When $\mathbf{K}_{\mathbf{s}}^{+} \parallel \mathbf{K}_{\mathbf{s}}^{-}$, cells become monocular. When disparity tuning is concerned, there are still two distinct possibilities — before ($\mathbf{K}_{\mathbf{s}}^{-} > \mathbf{K}_{\mathbf{s}}^{+}$) and after ($\mathbf{K}_{\mathbf{s}}^{-} < \mathbf{K}_{\mathbf{s}}^{+}$) the curve crossing, giving tuned inhibitory (Fig. 5bc) and tuned excitatory monocular cells respectively. A simple geometric argument shows that $|\Delta\phi| \approx 90^{o}$ for almost all other phase angles $\Delta\theta$ between $\mathbf{K}_{\mathbf{s}}^{+}$ and $\mathbf{K}_{\mathbf{s}}^{-}$, giving near/far cells (Poggio 1992). The ocular dominance index of these near/far cells can vary from binocularity to monocularity, depending on $\Delta\theta$. For example, when $\mathbf{K}_{\mathbf{s}}^{+} \perp \mathbf{K}_{\mathbf{s}}^{-}$, $|\mathbf{K}_{\mathbf{s}}^{\mathbf{L,R}}| = \sqrt{|\mathbf{K}_{\mathbf{s}}^{+}|^{2} + |\mathbf{K}_{\mathbf{s}}^{-}|^{2}}$ and the cells are binocularly balanced, e.g., Fig. 5d, and $\Delta\theta \leq 45^{o}$ between $\mathbf{K}_{\mathbf{s}}^{+}$ and $\mathbf{K}_{\mathbf{s}}^{-}$ leads to an imbalance, e.g., Fig. 5e.

The classification — class i, ii, iii, iv in section 3 and 4 — gives the most representative properties of the predicted cells. In reality there is a gradual variation from $\mathbf{K}_{\mathbf{s}}^{-}$ dominance at low $\underline{\mathrm{f}}$, to the curve crossing, and then to $\mathbf{K}_{\mathbf{s}}^{+}$ dominance at high $\underline{\mathrm{f}}$, giving a continuous spectrum of optimal phase disparity and ocular dominance index, as observed in experiments (LeVay and Voigt 1988). In addition, there is another cause, due to the phase changes $\phi^{L,R}$ of each eye, for the receptive field variation when the ocular dominance index and optimal

[9]Our measured inter-ocular correlation should roughly approximate that of the visual periphery, where visual objects randomly distribute in front of, behind, and around the fixation distance. As one approaches from the visual periphery to the visual center, the inter-ocular correlation should continuously increase.

disparity $\Delta\phi$ are fixed. For instance, the left two columns of Fig. 4B and 5 give the same $\Delta\phi$, but different — even and odd — receptive field forms.

The predicted cell types, tuned excitatory/inhibitory, near/far, monocular/binocular, have been observed in physiological experiments (Poggio Fischer 1977, Fischer and Kruger 1979, Ferster 1981 Freeman and Ohzawa 1990, LeVay and Voigt 1988). Unfortunately, it is not possible at this stage to perform a detailed quantitative comparison between theory and experiment (see section 5). However, there are certain qualitative trends and correlations between different cell properties that the theory predicts and can be checked in experiments.

a: Cell size and disparity tuning range.

Within the multiscale representation, phase disparity $\Delta\phi$ and spatial disparity Δx are related by $\Delta\phi = 2\pi f \Delta x$. If we assume that all cells have roughly the same phase disparity range, the spatial disparity range Δx is then proportional to $1/f$ or the cell sizes, as observed in experiments (Ferster 1981). Fig. 6 shows the disparity tuning curves of the corresponding cells in Fig. 5 (note the differences in their horizontal scales).

b: Cell size and optimal disparity

Correlation between smaller (larger) cell size and smaller (larger) optimal phase disparity can be seen from equations (8) and (13), since cell size $\propto 1/|\underline{f}|$. This is consistent with the observed proportionality between cell size and disparity tuning range (Ferster 1981) and the observation that tuned excitatory cells have narrower tuning range than near/far cells (Poggio and Fischer 1977, Poggio 1992).

c: Optimal disparity and ocular dominance index

The theory predicts that the near/far cells (from class i and iv) can be ocularly unbalanced or balanced. The tuned inhibitory cells are likely to be ocularly unbalanced (Fig. 4Babc, 5abc). The tuned excitatory cells are binocularly balanced if $\mathbf{K}_\mathbf{s}^+$ strongly dominates, and can be balanced as well as unbalanced if $\mathbf{K}_\mathbf{s}^+$ dominance is not strong enough (in which case many cells resemble near/far cells). Experimentally, it was observed that most tuned excitatory cells are balanced, most tuned inhibitory cells and a large fraction of near/far cells are unbalanced (Poggio and Fischer 1977, Fischer and Kruger 1979, Ferster 1981).

d: Blob cells and ocular imbalance/opponency

This is so because the so called blob cells, which are unoriented and large (Livingstone and Hubel 1984, Silverman et al., 1989), can be seen as a special class of tuned inhibitory cells, which are large (class i, Fig. 4Ba, 5a). It has been observed that all tuned inhibitory cells in monkey area 17 are unbalanced (Poggio and Fischer 1977), and that the blob cells are monocular (Horton and Hubel 1981, Tootell et al. 1988, Livingstone & Hubel, 1984; Ts'o & Gilbert, 1988). It will be interesting to see in experiments if the blob cells do receive opponent signals from the two eyes.

e: Cell orientation and disparity tuning.

Natural scenes exhibit higher ocular correlation in vertical directions, i.e., $r(0,f) > r(f,0)$. Consequently, the ocular opponency signal power $R^- = (1-r)R$ is smaller for vertical frequencies, forcing the $K^-(\underline{f})$ filter to decay and cross the $K^+(\underline{f})$ curve sooner.

From equation (13), the curve crossing marks the optimal phase disparity crossing from $> 90^o$ to $< 90^o$. Hence, there should be a correlation between horizontally oriented cells and small optimal disparities, as observed in experiments (DeAngelis et al. 1991, see related work by Barlow et al. 1967).

Here, we present some intuitive reasons for the binocular coding properties. The shift from binocular opponency (large optimal disparity) for large cells to binocular summation (small optimal disparity) for smaller cells can be understood as a signal enhancing strategy. Ocular correlation makes the ocular difference signal S^- smaller than the summation signal S^+. Hence integration in space with large cells are used to enhance S^-, which is best extracted with ocularlly misaligned receptive fields. When the cell sizes are small, there is little spatial integration, summation in ocular space by aligned receptive fields is used to amplify the signal. Monocular cells and near/far cells of intermediate sizes can be seen as transition from large cell ocular opponency to small cell ocular summation. A smaller fraction of large disparity horizontal cells follows from sampling theory which allocates small sampling density for infrequent signals — large vertical disparities.

This theory can also explain observations on ocular dominance column development (Li and Atick 1994b). When images in two eyes are misaligned (strabismus) or asynchronous during development, it gives a smaller inter-ocular correlation like the case in section 3. Hence a larger number of monocular cells will form, giving complete ocular dominance columns and nearly all cells in a column will be driven exclusively by a single eye, as observed experimentally (Hubel and Wiesel 1965, Van Sluyters and Levitt 1980, Miller and Stryker 1990). The opposite experimental condition where the two eyes receive much more correlated stimuli than normal gives very high ocular correlations. This makes the binocular opponency signal S^- very small and thus the $\mathbf{K}^-$ filter is negiligible. Consequently, $\mathbf{K}^+ \gg \mathbf{K}^-$ (class iii), binocular cells will dominate, or equivalently , ocular dominance columns will be weak or negiligible as observed in experiments (Stryker 1986). Starting from an ocular correlation matrix that is left-right asymmetric (cf. equation 2), this theory can also explain the unequal sizes of ocular dominance columns for the two eyes resulting from early monocular deprivations (Hubel Wiesel and LeVay 1977, Shatz and Stryker 1978).

Another theoretical prediction is that there can exist ocular disparities in orientation, optimal frequency or cell sizes, and tuning widths. This is because the tuning curves of the two eyes $|\mathbf{K}_{\mathbf{s}}^{\mathbf{L},\mathbf{R}}| = |\mathbf{K}_{\mathbf{s}}^{+} \pm \mathbf{K}_{\mathbf{s}}^{-}|$ can be slightly different, especially when $K^+(\underline{f})$ and $K^-(\underline{f})$ are not proportional to each other within a local scale region. For example, the cells in Fig. 5bcfg show different sizes or even shapes from the two eyes. However, since the left and right eyes are interchangable in this theory, the average disparity for the cell population as a whole, whether it is in optimal frequency, orientation, or in tuning width, will be zero. Experimentally, slight disparities in orientation, optimal frequency, and other properties have been observed, with average disparities for the cell population close to zero (Skottun and Freeman 1984).

5. Limitations

In this paper we have shown that the principle of efficient coding as implemented through multiscale representation and binocular decorrelation can explain many of the essential elements of binocular processing observed in the cortex (*e.g.* disparity selectivity, ocular dominance, their relationship with cell receptive field sizes, etc.). In our previous work (Li and Atick 1994a) multiscale coding and decorrelation in color was shown to account for many

features of spatial and chromatic cortical processing (spatial receptive field kernels at different scales, quadratures, color opponency, etc.). The current work further strengthens our belief in efficient coding as a framework for predicting and understanding neural processing in the cortex.

It is important to point out that the consequences of efficient coding have been derived making several simplifying assumptions that may not be true or may be only approximately true. These include the assumptions of active fixation and linear coding. We have also ignored feedback between binocular coding and vergence movement, which has been suggested by other researchers (e.g., Poggio 1992). This means that in detailed comparison with experiments, there will be disagreements for predictions that may be sensitive to such simplifying assumptions and that depend on the details of the ensemble (such as the property of true monocularity and the nonlinear effect of binocular facilitation). Our intent is to develop a framework that we can use to address the complex issues of binocular vision.

Another fundamental problem that we have to deal with in trying to make detailed comparison with experiments is the fact that there is no clear consensus in the experimental data when it comes to binocular cell properties. For example, the data of Hubel and Wiesel (1970) from anaesthetized monkey, differs in many respects from the data of Poggio and Fischer (1977) on awake behaving monkey. The former found no depth sensitive cells in area V1 of the monkeys while the latter did; and even in area V2 their observed cell depth selectivities differ qualitatively (Poggio and Fischer 1977). Another example is that while DeAngelis et al. (1991) found correlations between disparity selectivity and cell orientation, LeVay and Voigt (1988) failed to show a significant connection between the two. Although many experiments have studied correlations between cell disparity selectivities, ocular dominance indices, and cell orientations etc. (see Poggio 1992 and its references, LeVay and Voigt 1988) little effort has been made to link these properties to cell sizes (Freeman and Ohzawa 1990). Since our theory is based on a multiscale representation and many predictions are related to cell sizes, the comparison between the theory and experimental observations have to be indirect (e.g., in *b* of section 4). Freeman and coworkers (Freeman and Ohzawa 1990, DeAngelis et al. 1991) have proposed a stereo coding scheme in which the distribution of cell optimal phase disparities is the same for all scales (see also Marr and Poggio 1979). This scheme differs from the prediction of our theory (section 4 *b*) which allocates different cell optimal phase disparity ranges for different scales. Unfortunately, the relationship between the range of optimal phase disparities and cell size cannot be inferred from their published data (Freeman and Ohzawa 1990, DeAngelis et al. 1991).

We hope that our theory will facilitate more systematic exploration of binocular coding properties and their interaction with the environment. At this stage, more precise theoretical predictions require better knowledge and measurements on eye vergence movements and binocular input ensemble. Nevertheless there are some robust predictions, such as the correlation between cell sizes and binocular coding properties, that can be tested now. In particular, experiments measuring the sizes of different disparity selective cell types will be able to confirm or refute many aspects of the theory. The theoretical framework can also help organize and relate the seemingly large varieties of receptive field properties observed in the cortex under different noise or light levels and in different input ensembles for different animals (e.g., different inter-eye distances). At the same time, the theoretical development can benefit greatly from quantitative experimental input.

Appendix: Measurement on the inter-ocular correlation $r(\underline{\mathbf{f}})$

We used a special stereo camera, the so called *Stereo Realist*. It has two lenses of parallel optical axes separated by about 7 centimeters. Each shot takes two pictures simultaneously, one for each lense. The shutter speed, aperture size and the focal length are common for the two lenses. The camera focus ranges from 2.5 feet to infinity. The monocular field of view is about 34 degrees which is considerably less than the field of view of humans but nevertheless large enough to give us a preliminary measurement of the inter-ocular correlation function.

Our ensemble of stereo images consists of 127 shots taken mostly in Central Park of New York City, in bright enough daylight to allow for large depth of focus with small apertures. Most objects in most pictures are at least one meter away from the camera. Roughly 40-50 shots have the dominant objects in the scene approximately 2-4 feet from the camera.

All images were taken using Kodak T-MAX black and White film with exposure index 400. The films were developed and printed on contact sheets and scanned and digitized to 8 bit gray scale. The left and right images of each stereo picture were scanned to about 270×295 pixels each, and then cropped to 256×256 pixels for analysis. Controls were taken to ensure that the corresponding pixels between two images in a stereo pair are fixed from one pair to another within an error of 1 pixel horizontally and vertically. The power spectrum of the right and left images are both of the order f^{-2} up to $f \sim 0.4$ cycles /pixel (for higher f, noise in images play a large role), in close agreement with other measurements from (monocular) natural scenes (Field 1987, Ruderman and Bialek 1993).

The inter-ocular correlation $|r(\underline{\mathbf{f}})|$ is shown in Figure 1. One can see that $r \approx 1$ for $f \approx 0$, meaning that the average mean light inputs to two eyes are about the same. However, $r(\underline{\mathbf{f}})$ decays with f, much faster in the horizontal direction, to $r(\underline{\mathbf{f}}) \ll 1$ for $f \geq 1$ cycle/degree. This means that the images to the two eyes are similar only up to a spatial resolution of around 1 degree. This is not surprising as many objects in our visual inputs differ in disparity by that order. The $r(\underline{\mathbf{f}})$ is larger in the vertical direction since distribution of vertical disparities has a smaller mean and variations than that of the horizontal disparities. It can be shown to have significant consequence on the correlation between orientation and disparity selectivity of the cells (see section 4).

We would like to point out that the measurement on $|r|$ depends on the ensemble of scenes used, on the disparity range or distance of the objects in the scenes (and the distance between the eyes). We have a biased ensemble since the distribution of object distances sampled in our measurement is probably not the same as that under natural viewing conditions. In particular, limited by the camera focus range, we cannot take images with objects closer than 2 feet away from the camera.

References

[1] Andrews, B. W. and Pollen, D. A. 1979. Relationship between spatial frequency selectivity and receptive field profile of simple cells. *J. Physiol.* (London), **287**, 163–176.

[2] Atick, J. J. 1992. Could Information Theory provide an ecological theory of sensory processing? *Network*, **3**, 213-251.

[3] Atick, J. J. 1994. From natural scene regularities to neural coding: A predictive approach to visual processing. To appear.

[4] Atick, J. J. and Redlich, A. N. 1990. Towards a theory of early visual processing. *Neural Comp.*, **2**, 308–320.

[5] Atick, J. J. and Redlich, A. N. 1992. What does the retina know about natural scenes? *Neural Comp.*, **4**, 196–210.

[6] Atick, J. J., Li, Z. and Redlich, A. N. 1992. Understanding retinal color coding from first principles. *Neural Comp.*, **4**, 559–572.

[7] Atick, J. J., Li, Z., and Redlich, A. N. 1993. What does post-adaptation color appearance reveal about cortical color representation? *Vis. Res.*, **33**, 123-129.

[8] Barlow, H.B. 1961. Possible principles underlying the transformation of sensory messages *Sensory Communication* ed W. A. Rosenblith (Cambridge, MA: MIT Press). Unsupervised learning *Neural Computation* **1** 295-311 (1989).

[9] Barlow, H. B., Blakemore, C. and Pettigrew J. D. 1967. The neural mechanism of binocular depth discrimination. *J. Physiol.* **193** 327-342.

[10] Berns, G. S., Dayan, P., and Sejnowski, T. J. 1993. A correlational Model for the Development of disparity selectivity in visual cortex that depends on prenatal and postnatal phases. Submitted for Publication.

[11] Blake,R. and Wilson, H. R. 1991. Neural Models of Stereoscopic Vision *TINS* Vol. 14, p.445-451.

[12] DeAngelis, G. C., Ohzawa, I., and Freeman R. D. 1991, Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature* Vol. 352 p. 156-159.

[13] De Valois, R. L, Yund, E. W., and Hepler, N. 1982. The orientation and direction selectivity of cells in macaque visual cortex. *Vision Res.* **22**, 531-544.

[14] Ferster, D. 1981. A comparison of binocular depth mechanisms in area 17 and 18 of the cat visual cortex. *J. Physiol.* **311** p.623-655.

[15] Field, D. J. 1987. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am.*, **A 4**, 2379–2394.

[16] Fischer, B. and Kruger, J. 1979. Disparity tuning and binocularity of single neurons in cat visual cortex *Exp. Brain Res.* 35, 1-8.

[17] Freeman, R. D. and Ohzawa, I. 1990. On the neurophysiological organization of binocular vision *Vis. Res.* Vol. 30 1661-1676.

[18] Horton, J. C. and Hubel, D. H. 1981. Regular patchy distribution of cytochrome oxidase staining in primary visual cortex of macaque monkey, *Nature* **292:** 762-764.

[19] Hubel, D. H. and Wiesel, T. N. 1965. Binocular interaction in striate cortex of kittens reared with artificial squint. *J. Neurophysiol* 28: 1041-1059.

[20] Hubel, D. H. and Wiesel, T. N. 1970. Cells sensitive to binocular depth in area 18 of the macaque monkey cortex. *Nature* 225: 41-42.

[21] Hubel, D. H. and Wiesel, T. N. LeVay, S 1977. Plasticity of ocular dominance columns in monkey striate cortex. *Philosophical Transactions of the Royal Society of London* Series B, 278: 377-409.

[22] LeVay, S., Voigt, T. 1988. Ocular dominance and disparity coding in cat visual cortex *Vis. Neurosci. 1*, 395-414.

[23] Li, Z. and Atick, J. J. 1994a. Towards a theory of the striate cortex. *Neural Comp.*, 6:127-146 1994.

[24] Li, Z. and Atick, J. J. 1994b. Understanding ocular dominance development from binocular input statistics, *in preparation.*

[25] Livingstone, M. S. and Hubel, D. H. 1984. Anatomy and physiology of a color system in the primate visual cortex. *J. of Neurosci.* Vol. 4, No.1. 309-356.

[26] Marr, D. and Poggio, T. 1979. A computational theory of human stereo vision. *Proc. R. Soc. London Ser. B* 204:301-28.

[27] Miller, K. D., Keller, J. B., and Stryker M. P. 1989. Ocular dominance column development: Analysis and simulation. *Science*, 245:605-615.

[28] Miller, K. D. and Stryker, M. P. 1990. The development of ocular dominance columns: mechanisms and models. In Hanson S. J. and Olson C.R., editors, *Connectionist Modeling and Brain Function: The Developing interface,* 255-350. MIT Press/Bradford, Cambridge, MA.

[29] Ohzawa, I. and Freeman, R. D. 1986. The binocular organization of simple cells in the cat's visual cortex *J. Neurophysiol.* **56** 221-242.

[30] Poggio, G. F. and Fischer, B. 1977. Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey *J. of Neurophysiol.* 40(6): 1392-1405.

[31] Poggio, G. F. 1992. Physiological basis of stereoscopic vision in J. R. Cronly-Dillon Ed. *Vision and Visual Dysfunction.* Vol. 9 *Binocular Vision* E. Regan Ed. CRC Press, Inc.

[32] Poggio, G. F. and Poggio T. 1984. The analysis of stereopsis. *Ann. Rev. Neurosci.* 7:379-412.

[33] Ruderman, D. L. and Bialek, B. 1993. Statistics of natural images: scaling in the woods. Conference presentation at *Neural Information Processing Systems — Natural and Synthetic, 1993 Denver, Colorado.*

[34] Shatz C.J. and Stryker M. P. 1978. Ocular dominance in layer IV of the cat's visual cortex and the effects of monocular deprivation. *J. of Physiology* 281:267-283.

[35] Silverman, M. S., Grosof, D. H., De Valois, R. L. and Elfar, S. D. 1989. Spatial-frequency organization in primate striate cortex, *Proc. Natl. Acad. Sci. U.S.A.* **86:** 711-715.

[36] Skottun, B. C. and Freeman, R. D. 1984. Stimulus specificity of binocular cells inthe cat's visual cortex: ocular dominance and the matching of left and right eyes. *Exp. Brain. Res.* 56:206-216.

[37] Stryker M. P. 1986 The role of neural activity in rearranging connections in the central visual system. In Ruben, R.J. Vater, T.R. Van De, and Rubel, E.W. (Eds.), The biology of Change in Otolaryngology, pp. 211-224, Elsevier Science B.V. Amsterdam.

[38] Tootell, R. B. H., Hamilton, S. L., Silverman, M. S., and Switkes, E. 1988. Functional Anatomy of Macaque Striate Cortex. I. Ocular Dominance, Binocular Interactions, and Baseline Conditions *J. Neurosci.* 8(5):1500-1530.

[39] Ts'o, D. Y. and Gilbert, C. D. 1988. The organization of chromatic and spatial interactions in the primate striate cortex. *J. of Neurosci.* 8(5):1712-1727.

[40] Van Sluyters, R.C. and Levitt, F.B. 1980. Experimental strabismus in the kitten. J. Neurophysiol. 43:686-699.

Figure Captions

Fig. 1. Measured $r(\underline{f})$ (data points with error bars) function from stereo images for slices in horizontal $\underline{f} = (f, 0)$ and vertical $\underline{f} = (0, f)$ frequencies.

Fig. 2 The ocular opponency K^- (dashed curves) and ocular summation K^+ (solid curves) filters before, A, and after, B, multiscale partitioning ((9) and (10)). They are generated from the equations $K^{\pm}(f) \propto M^{\pm}((M^{\pm})^2(R^{\pm}+1)+1)^{-1/2}$ where $M^{\pm} = R^{\pm}/(R^{\pm}+1)e^{-(|\underline{f}|/f_c)^{\alpha}}$, $R = 4.0/(|\underline{f}|^2 + f_{\mu}^2)$, $R^{\pm} = (1 \pm r)R$, with $f_{\mu} = 0.3\ c/deg$, $\alpha = 1.4$, $f_c = 22\ c/deg$ (see Atick and Redlich 1992 for details). The $r(\underline{f})$ is modeled as $r(f) = 0.96e^{-f/f_o}$, where $f_o = 15$ cycle/degree. The lowpass blob cell is obtained by convoluting $K^{\pm}(f)$ filters with frequency envelop $\exp(-f^2/(2f_l^2))$ where $f_l = 0.8\ c/deg$. The bandpass cells are obtained by convoluting $K^{\pm}$ with the envelops $\exp(-(log(f/f_s))^2/(2\sigma^2))$, where $f_s = 1, 3, 8, 25\ c/deg$ are the optimal frequencies and $\sigma = 1.6/2$ octaves. The same parameters are used in Figs. 5, 6.

Fig. 3: Demonstration of binocular receptive field construction. The receptive field in the left eye is $\mathbf{K_s^L} = \mathbf{K_s^+} + \mathbf{K_s^-}$, and that of the right eye is $\mathbf{K_s^R} = \mathbf{K_s^+} - \mathbf{K_s^-}$ (the subscript s is omitted in the plot for clarity). Depending on the relative strengths of $\mathbf{K_s^+}$ and $\mathbf{K_s^-}$, and on their angle $\Delta\theta$, different receptive field strengths $|\mathbf{K_s^L}|$ and $|\mathbf{K_s^R}|$, as well as their receptive field phase difference $\Delta\phi$ are formed. As $\Delta\theta$ changes from 0 to 90^o, the cell ocular dominance changes from the strongest monocularity to complete ocular balance, however, the optimal phase disparity changes from out of phase $\Delta\phi = 180^o$ to no less than 90^o when the $\mathbf{K^-}$ filter dominates, from zero phase $\Delta\phi = 0$ to no more than 90^o when the $\mathbf{K^+}$ filter dominates.

Fig. 4: $K^{\pm}$ sensitivities, A, and the receptive fields, B, predicted using the ocular correlation function $r(f) = 0.96e^{-f/f_o}$ with $f_o = 1$ cycle/degree, close to the correlation in our measurements. In B, the solid and dotted curves are receptive fields for the two eyes respectively. The top plot in B is for a large and unoriented cell type where only $\theta^{\pm} = 0$ is allowed (Li

and Atick 1993). The rest are oriented cells, and their receptive fields depend on $\theta^{\pm}$. Each row has the same optimal spatial frequency, and each column the same $\theta^{\pm}$, as indicated. The receptive field strengths are normalized such that the largest amplitude is the same for each cell in the same row. To reveal enough detail of the receptive field forms, the horizontal and vertical scales for different rows are different. Hence, the plots only convey *qualitatively* the larger and smaller receptive fields and higher or lower cell sensitivities. In reality, for instance, the larger cells in the plot should be relatively larger and the smaller cells smaller. Similarly, the receptive field amplitudes should be relatively smaller for, e.g., blob cells.

Fig. 5: Predicted receptive fields under ocular correlation $r(f) = 0.96e^{-f/f_o}$ with $f_o = 15$ cycle/degree. The $K^{\pm}$ filters in frequency space are plotted in Fig. 2B. The format is the same as in Fig. 4B.

Fig. 6: The disparity tuning curves for the corresponding cells in Fig. 5. The vertical line in each plot marks the zero disparity location. The horizontal scale is the same in all the plots, indicated by the bar at bottom. The tuning curve is normalized such that the maximum response for each cell is the same, and the bottom of the vertical line marks the zero response levels. Note that the tuning curves are shallow for cells strongly dominated by one eye (e.g., a, b, c, n, o); cells in d and e are tuned to same optimal disparity but have different modulations by depths by their different ocular dominance. If non-linearity such as thresholds are added, the tuning curves can be less shallow for the unbalanced cells. The cell responses at each disparity are obtained by simulating the sweeping of bars of widths $1/|\underline{f}|$, where $\underline{f}$ is the optimal frequency of the cell, over the receptive fields. The bars for the two eyes are separated by the measured disparity value.

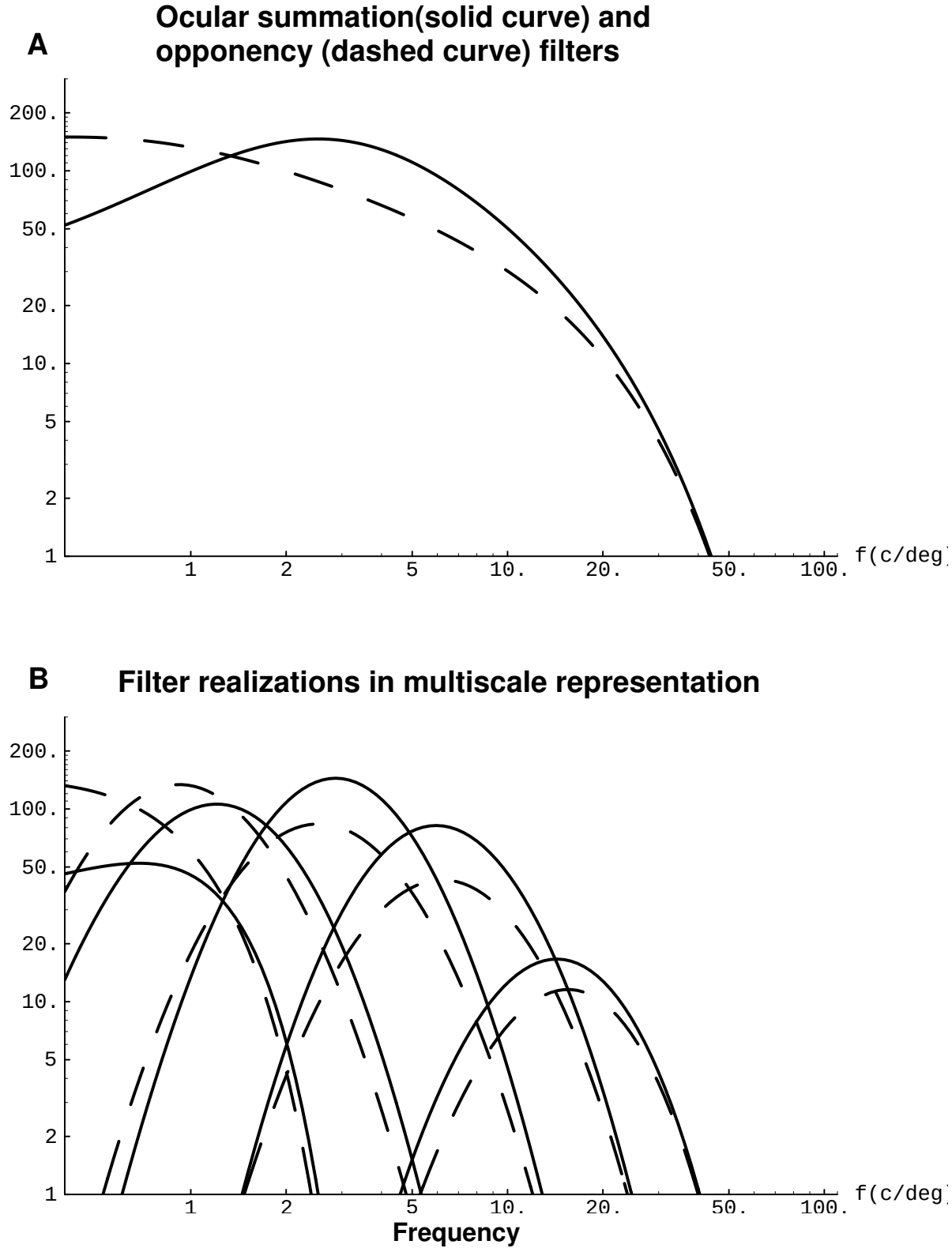
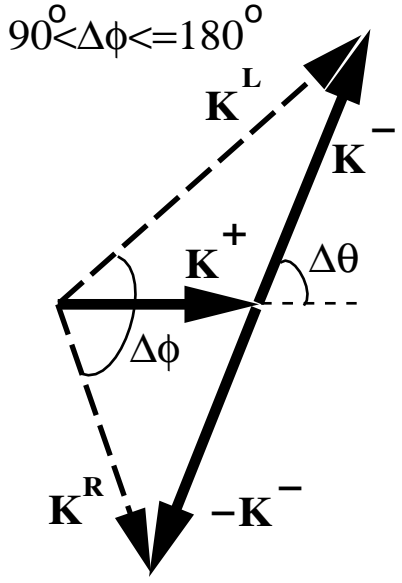

A
Ocular summation(solid curve) and
opponency (dashed curve) filters
200.
100.
50.
20.
10.
5
2
1
1
2
5
10.
20.
50.
100.
f(c/deg)
B
Filter realizations in multiscale representation
200.
100.
50.
20.
10.
5
2
1
1
2
5
10.
20.
50.
100.
f(c/deg)
Frequency

Fig.2

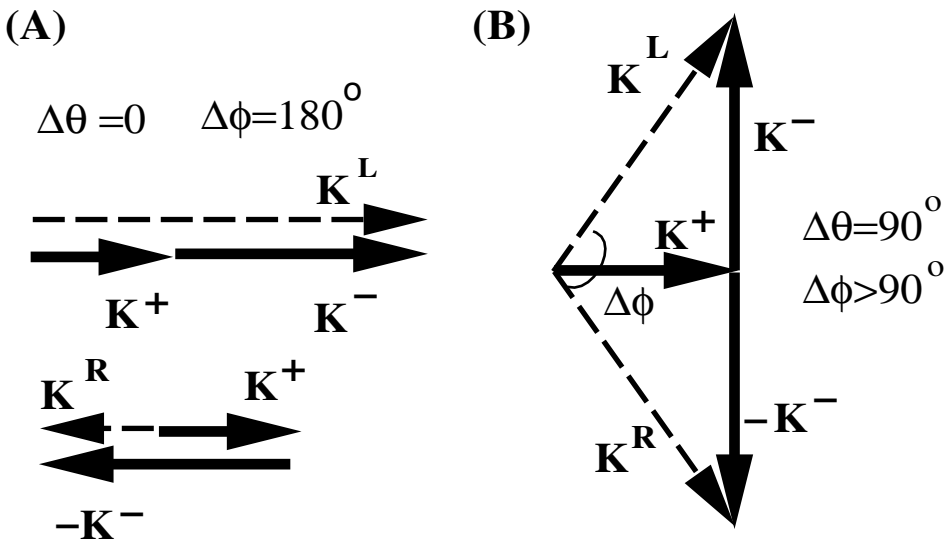
Binocular filter construction

$K^- > K^+$

General case:

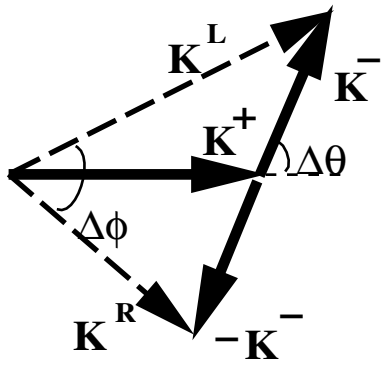


Extreme cases:



$K^- < K^+$

General case:



Extreme cases:

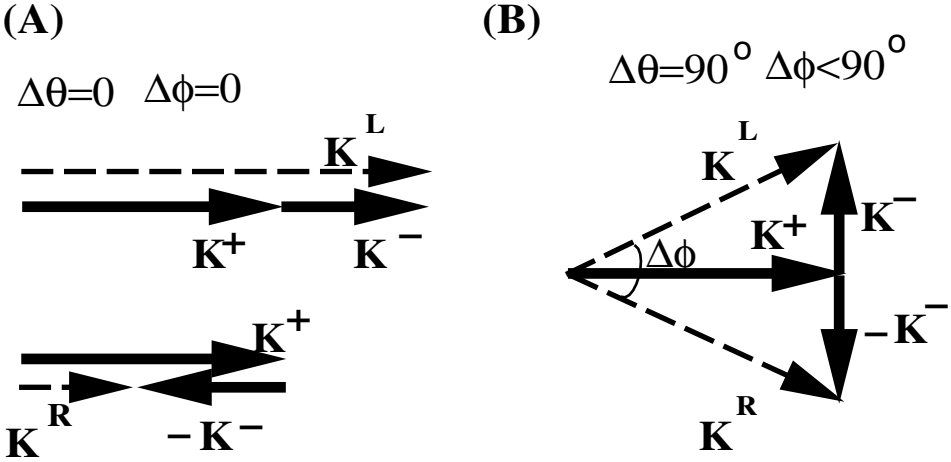
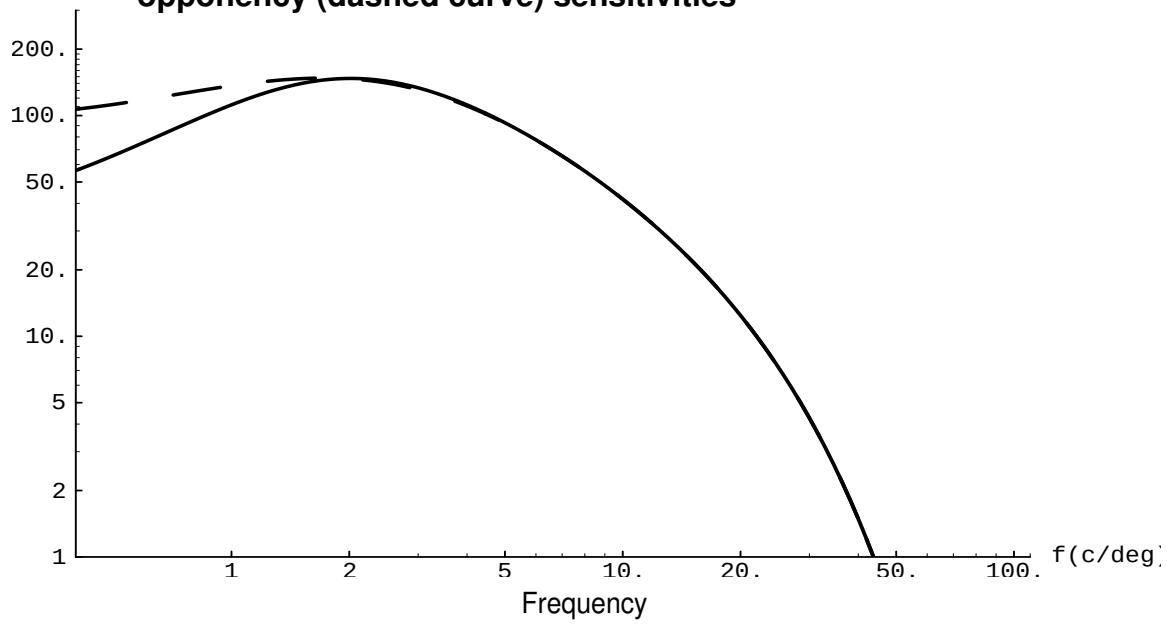


Fig.3

A system for inputs of low ocular correlation

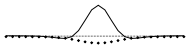
A **Ocular summation (solid curve) and opponency (dashed curve) sensitivities**



B **Cell receptive fields**

Unoriented Blob Cell

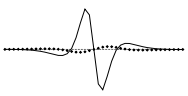
a



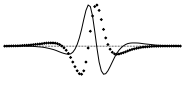
Oriented Cells

b

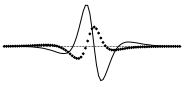
c



d



e



1c/deg

f

g

4c/deg

Fig.4

Cell receptive fields in an actively fixating visual system

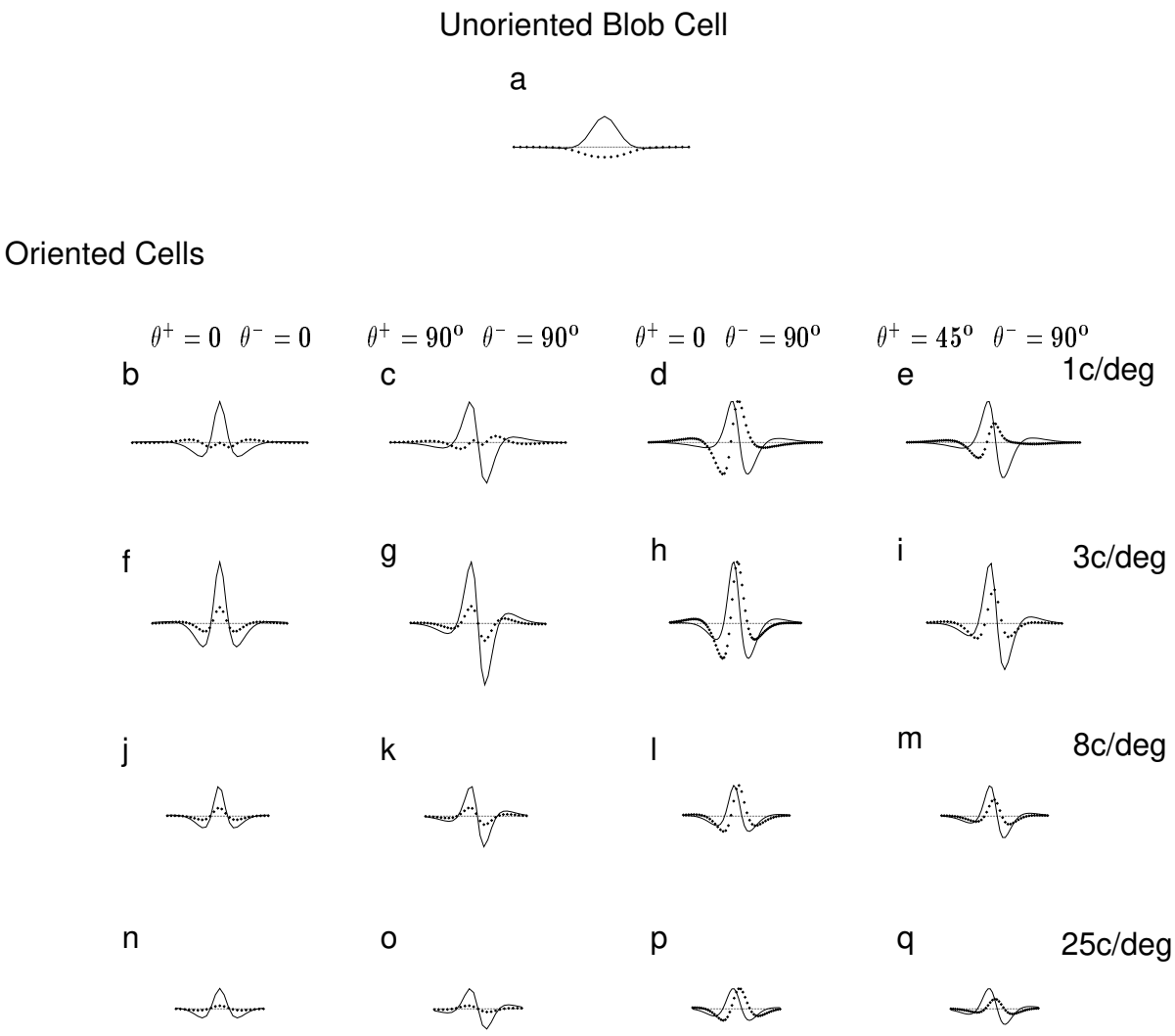


Fig.5

Cell disparity tuning curves

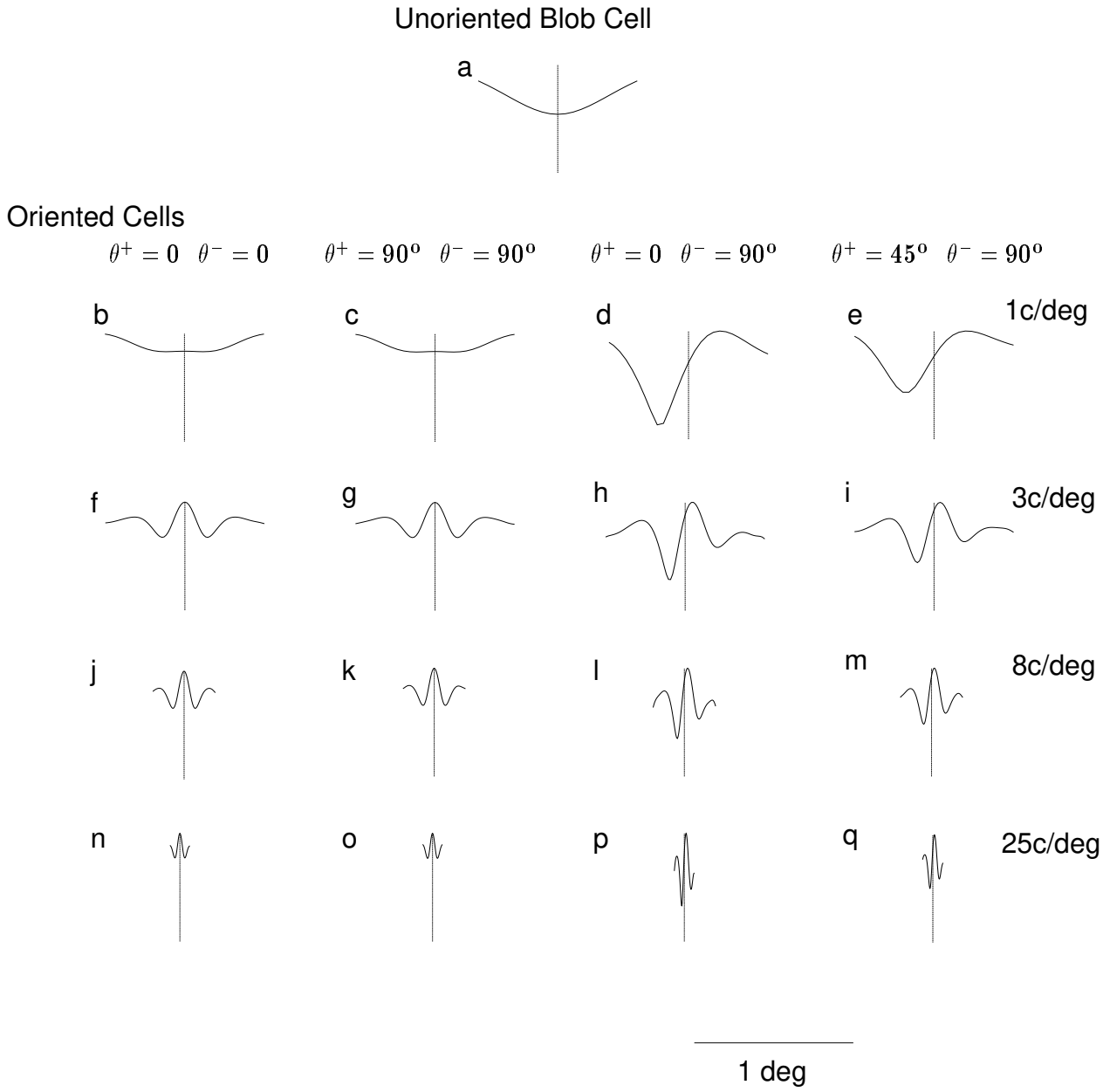


Fig.6