

## Chapter 4

# Extensions to the Model

### 4.1 Introduction

We have seen in the previous chapters that we can propose a model specific enough to ask very precise questions. This led us to an understanding of visual deprivation, an initial appreciation of the structure in natural scenes, and finally provided us enough insight to propose a simple, analytically tractable, model of deprivation.

There are many other questions we can ask about the role of the environment, some of these were asked in Chapter 2. In this chapter we present some answers to a few of them which don't fall neatly into the deprivation paradigm, and are unrelated to each other. This “grab-bag” of problems gives us an indication of some of the work that is yet to be done, and what potential the proposed models have.

### 4.2 X and Y Cells

In Section 2.2, where we outlined the visual system, we mentioned that there were at least two different types of retinal cells, called X and Y cells. X cells have smaller receptive fields, and respond slowly to moving stimuli. Y cells have larger receptive fields, and respond rapidly to moving stimuli. Though we ignore the response timings of these cells, it is clear that the size of the receptive field would influence the inputs into the LGN and the cortex. A smaller receptive field would be sensitive to smaller scales, and likewise for larger receptive fields.

Figure 4.1 shows the results of measurements of X and Y cell properties, fit to a center-surround difference of Gaussians(Linsenmeier et al., 1982), and converted to units of pixels. The conversion is done roughly, using  $0.5 \text{ deg} \sim 1 \text{ pixel}$ , to correspond with the DOG filters previously used. The mean center and surround values found are center = 0.83 pixels, surround = 3.2 pixels for X cells and center = 2.9 pixels, surround = 4.3 pixels for Y cells. Within each class of cell, there is a lot of variability. The classification of X and Y cells is really an artificial labeling of a broad range of cell

types, and shouldn't be considered a very strong distinction. We use the labels for convenience, with this in mind.

We notice that the Y cells are larger than the X cells, on average. There is a lot of variability in the data, and although there are some indications of correlations between center and surround sizes within each type of cell, the specific numbers are not important. We are interested in the qualitative effect of the retinal preprocessing on the cortical receptive field formation.

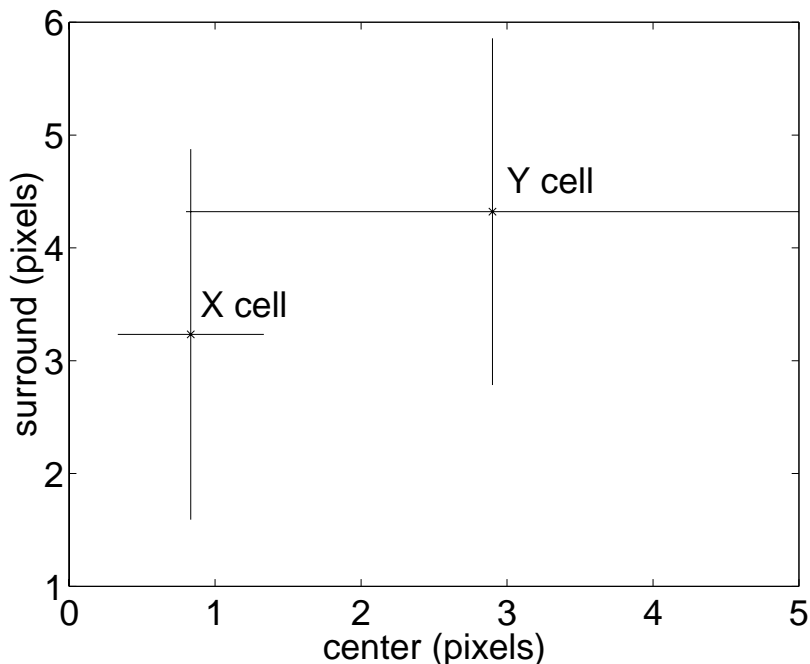


Figure 4.1: Size of X and Y cell receptive fields, as measured by Linsenmeier et. al. (1982), and fit to center-surround difference of Gaussians. (From Figure 8, and Paragraph 2 on Page 1179 in Linsenmeier). The mean center and surround values found are  $c = 0.83, s = 3.2$  for X cells and  $c = 2.9, s = 4.3$  for Y cells.

In order to explore this, we trained a neuron with images filtered using different sized difference of Gaussians. For simplicity, we are ignoring, here, any differences in temporal response properties between the X and Y cells. Examples of these images and filters, for X and Y cells, is shown in Figure 4.2. The X cells, with much smaller receptive fields, respond to much higher spatial frequencies in the images than the Y cells. A wide range of center and surround sizes were explored, in order to get an idea of the qualitative effects of these parameters on the cortical receptive fields.

All of the neurons trained achieved oriented receptive fields. Example RFs from X and Y cells are shown in Figure 4.3. The receptive fields were fit to Gabor filters (Jones and Palmer, 1987), which are merely sine gratings restricted by a Gaussian window (Figure 4.4). These, and the other RFs found, differed primarily in *spatial frequency*. Figure 4.5 shows the dependence of the resulting spatial frequency on the center and surround sizes of the retinal DOG. Those retinal cells which had *small* receptive fields yielded cortical cells with *high* spatial frequency. This could be a simple way for the

visual system to ensure that the cortex has a representation of many scales from the environment.

### 4.3 Direction Selectivity

Most simple cells in the cat striate cortex are both orientation and direction selective. We have dealt with orientation selectivity throughout this work, but have only touched upon the latter. At the preferred orientation, a cell which is direction selective responds to a drifting grating moving in one direction more strongly than the opposite direction. The ability of the cell to detect the direction of motion depends on the interaction of responses to at least two different points in the visual field at different times. This is to say, that it depends on the *spatiotemporal* receptive field of the cell(Reid et al., 1991).

Spatiotemporal (ST) receptive fields are generally characterized by their degree of separability. A spatiotemporal *separable* receptive field can be expressed as a product of a function which only depends on time and a function which only depends on space. A spatiotemporal *inseparable* receptive field is one that cannot be expressed as such a product.

In order for a cell to be direction selective, it must have a ST inseparable receptive field. This can be seen by a simple example. Consider the spatiotemporal contour plot of a one dimensional ST separable receptive field (Figure 4.6, left). A temporal cross section made vertically through the contour plot represents the temporal response function to flashed stimulus at a point  $x$ . This can be thought of as a probability of firing a time  $t$  after the flashed stimulus. A ST inseparable receptive field would have different response timings at different spatial positions, as in Figure 4.6, right. A space-time plot of a drifting grating can be superimposed with these contour plots (see Figure 4.7), from which we can easily see that only a ST inseparable receptive field can be direction selective.

A model of direction selectivity has been proposed(Feidler et al., 1997) which includes the effects of two types of LGN cells, called lagged and non-lagged cells, which differ only in their response timing. They conclude that simple Hebbian rules are inadequate to develop direction selectivity, and demonstrate more sophisticated Hebb-type learning rules, such as postsynaptic gating and BCM(Bienenstock et al., 1982), are sufficient to obtain direction selectivity. In this section we extend our model to include lagged and non-lagged LGN cells. This is done similarly to having left and right eyes.

In this model the cortical cell receives input from two sets of LGN cells which view the same area of space, but differ only in their timing. The first set has a delayed response to the input (lagged cells) relative to the second set (non-lagged cells). Essentially the cortical cell has a receptive field from two different times of the input at the same spatial location. We will refer to these as the lagged and the non-lagged RFs, because they arise from the lagged LGN and the non-lagged LGN cell inputs, respectively.

The input patches are chosen using a sequence of random *saccades* and *drifts*(Carpenter, 1977). A saccade is a large jump to a random part of an image, and a drift is a continuous motion within an

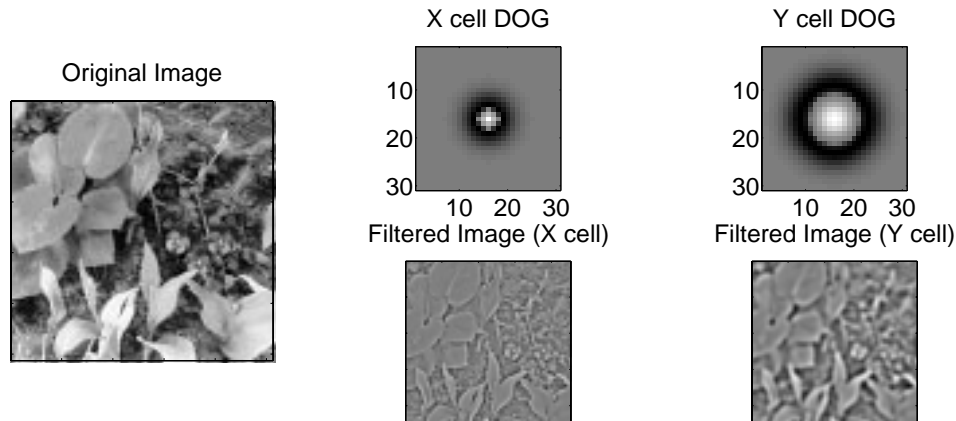


Figure 4.2: Filters and filtered images for X and Y cells. Shown is the original image (left), and the difference of Gaussian (DOG) filters for X and Y cells (above center and right). The center and surround values used are  $c = 0.83$ ,  $s = 3.2$  for X cells and  $c = 2.9$ ,  $s = 4.3$  for Y cells. Shown also are the images resulting from the application of these filters (below center and right). The X cells respond to much higher spatial frequencies in the images than the Y cells.

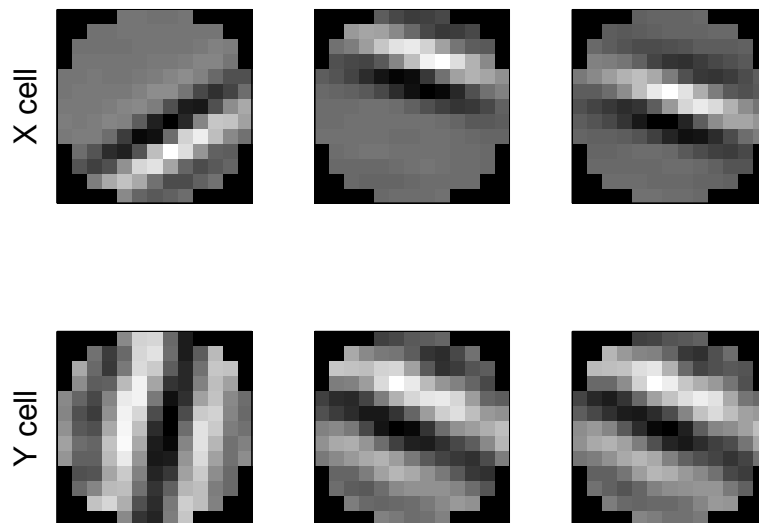


Figure 4.3: Example receptive fields trained cortical cells, with X cell input (above) and Y cell input (below).

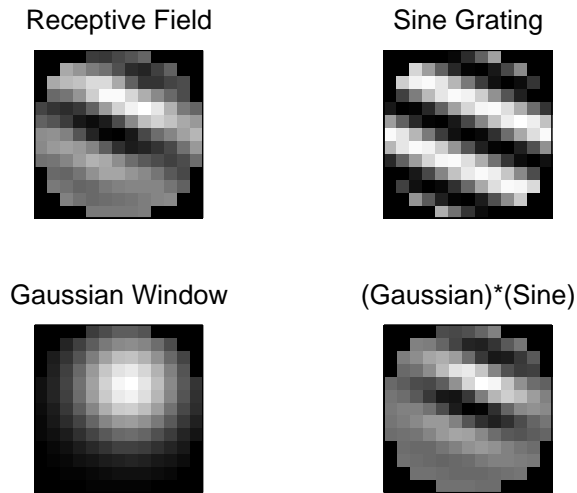


Figure 4.4: Fitting a receptive field to a Gabor filter. Shown is an example receptive field (upper left), the best fit sine grating (upper right), the best fit Gaussian window (lower left), and the product of the sine grating and the Gaussian window.

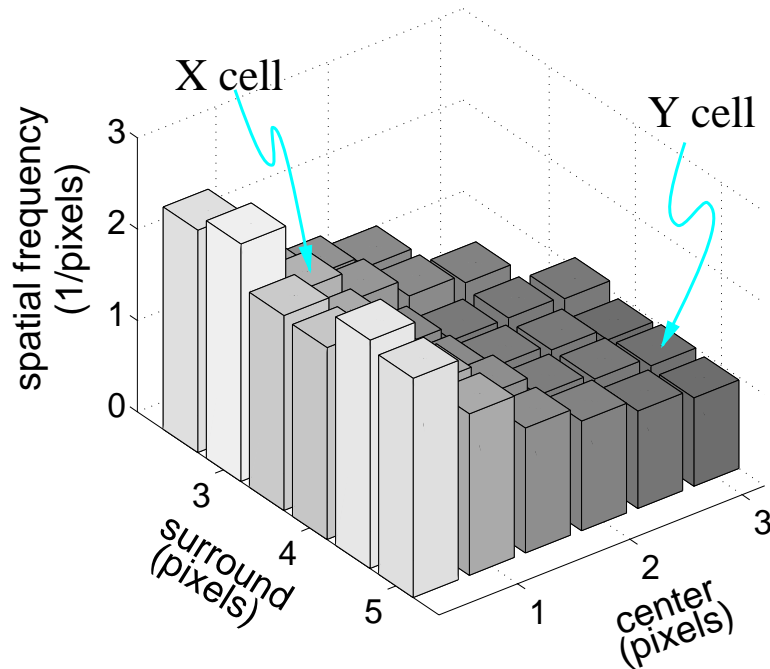


Figure 4.5: Spatial frequency of trained cortical neurons, as a function of the center and surround sizes of retinal receptive fields.

image in a particular direction at a particular velocity. In the model, the drift velocity is kept constant, and the drifts last a random time  $\tau$ . In between drifts are saccades to a different image or part of the same image.

We use some of the learning rules introduced in the previous chapter: PCA, QBCM,  $K_1$  and  $S_1$ . Example receptive fields and their orientation tuning, for a velocity of 2 pixels per iteration, are shown in Figure 4.8. The orientation tuning was obtained using drifting oriented sine gratings. It is clear that the PCA rule did not develop direction selectivity whereas the other rules did.

We measure direction selectivity using the DS index,

defined as

$$DS \equiv \frac{R_{(\text{preferred})} - R_{(\text{nonpreferred})}}{R_{(\text{preferred})} + R_{(\text{nonpreferred})}} \quad (4.1)$$

where  $R_{(\text{preferred})}$  and  $R_{(\text{nonpreferred})}$  are the responses to a sine grating, at optimum orientation and spatial frequency, moving in the *preferred* direction and *non preferred* direction, respectively. Figure 4.9 shows the index as a function of eye drift velocity, for a constant LGN lag of 1 iteration. Example receptive fields from the BCM learning rule for each velocity are shown. Other than the PCA rule, all rules show some velocity tuning: they exhibit direction selectivity for some velocities, but all lose it for either velocities which are too high or too low. If the velocity is zero, then we are in a case identical to the binocular training introduced in Chapter 2: both eyes see the same thing. In the high velocity case, the eyes see completely different, non-overlapping, patches of the environment. This is most comparable to the strobe light environment (Cynander et al., 1973), because in both situations the temporal correlations are lost. Thus, results in the high velocity case are consistent with those experiments.

There is a parallel between these results and the results on binocular cortical misalignment (Shouval et al., 1996). In the misalignment work, it was shown that the BCM rule in a natural scene environment, with varying degrees of binocular overlap, developed either identical receptive fields (for complete overlap), *monocular* receptive fields (for no overlap), or receptive fields formed only in the overlap region (for intermediate overlap). In the current work, if the lag time of the LGN lagged cells is kept constant, then a constant velocity would imply a constant amount of overlap of input patterns during eye drift. Thus, zero velocity would yield identical lagged and the non-lagged RFs, yielding no direction selectivity. Likewise, high velocity would give no overlap of input patterns during eye drift, and would yield either a completely lagged or a completely non-lagged receptive field, and again no direction selectivity.

#### 4.4 Structure Removal: Sensitivity to Outliers

Learning rules which are dependent on large polynomial moments, such as Quadratic BCM and kurtosis, tend to be sensitive to the tails of the distribution. This property implies that neurons are highly responsive, and sensitive, to the outliers, and consequently leads to a sparse coding of the input signal.

One should note that over-sensitivity to outliers is considered to be undesirable in the statistical literature. However, in the case of a sparse code the outliers, or the rare and interesting events, are what is important. The degree to which the neurons form a sparse code determines how much of the input distribution is required for maintaining the RF. This can be done in a straightforward and systematic fashion.

The procedure involves simply eliminating from the environment those patterns for which the neuron responds strongly. An example receptive field and some of the patterns which give that neuron strong responses is shown in Figure 4.10. These patterns tend to be the high contrast edges, and are thus the structure found in the image. The percentage of patterns that needs to be removed in order to cause a change in the receptive field gives a direct measure of the sparsity of the coding. The process of training a neuron, eliminating patterns which yield high response, and retraining can be done recursively to sequentially remove structure from the input environment, and to pick out the most salient features in the environment. The results of this are shown in Figure 4.11.

For Quadratic BCM and kurtosis, one need only eliminate *less than one half of a percent* of the input patterns to change the receptive field significantly. The changes which one can observe include orientation, phase and spatial frequency changes. This is a very small percentage of the environment, which suggests that the neuron is coding the information in a very sparse manner. For the skewness maximization rule, more than five percent are needed to alter the receptive field properties, which implies a far less sparse coding.

To make this more precise, we introduce a normalized difference measure between two different RFs. If we take two weight vectors,  $\mathbf{w}_1$  and  $\mathbf{w}_2$ , then the normalized difference between them is defined as

$$\mathcal{D} \equiv \frac{1}{4} \left( \frac{\mathbf{w}_1 - \bar{w}_1}{\|\mathbf{w}_1\|} - \frac{\mathbf{w}_2 - \bar{w}_2}{\|\mathbf{w}_2\|} \right)^2 \quad (4.2)$$

$$= \frac{1}{2}(1 - \cos \alpha) \quad (4.3)$$

where  $\alpha$  is the angle between the two vectors, and  $\bar{w}_i$  is the mean of the elements of the vector  $i$ . This measure is not sensitive to scale differences, because the vectors are divided by their norm, and it is not sensitive to scalar offset differences, because the mean of the vectors is subtracted. The measure has a value of zero for identical vectors, and a maximum value of one for orthogonal vectors.

Shown in Figure 4.12 is the normalized difference as a function of the percentage eliminated, for the different learning rules. Differences can be seen with as little as a tenth of a percent, but only changes of around half a percent or above are visible as significant orientation, phase, or spatial frequency changes. Although both skewness and Quadratic BCM depend primarily on the third moment, Quadratic BCM behaves more like kurtosis with regards to projections from natural images.

Similar changes occur for both the BCM and kurtosis learning rules, and most likely occur under other rules that seek kurtotic projections. It is important to note, however, that patterns must be

eliminated from *both* sides of the distribution for any rule that does not use the rectifying sigmoid because the strong *negative* responses carry as much structure as the strong positive ones. Such responses are not biologically plausible, so they wouldn't be part of the encoding process in real neurons.

It is also interesting to observe that the RF found after structure removal is initially of the same orientation, but of different spatial phase or possibly different position. Once enough input patterns are removed, the RF becomes oriented in a different direction. If the process is continued, all of the orientations and spatial locations would be obtained.

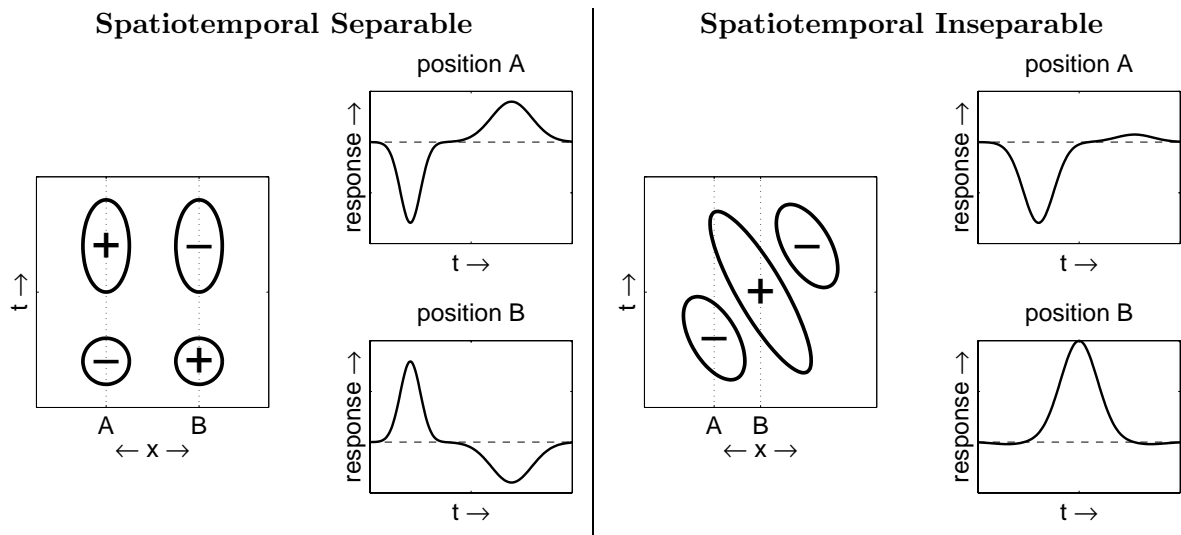


Figure 4.6: Spatiotemporal contour plots of a one dimensional ST separable receptive field (left) and ST inseparable receptive field (right). A temporal cross section made vertically through the contour plot represents the temporal response function to flashed stimulus at a point  $x$ . Example response functions, for points A and B, are shown on the right of each contour plot.

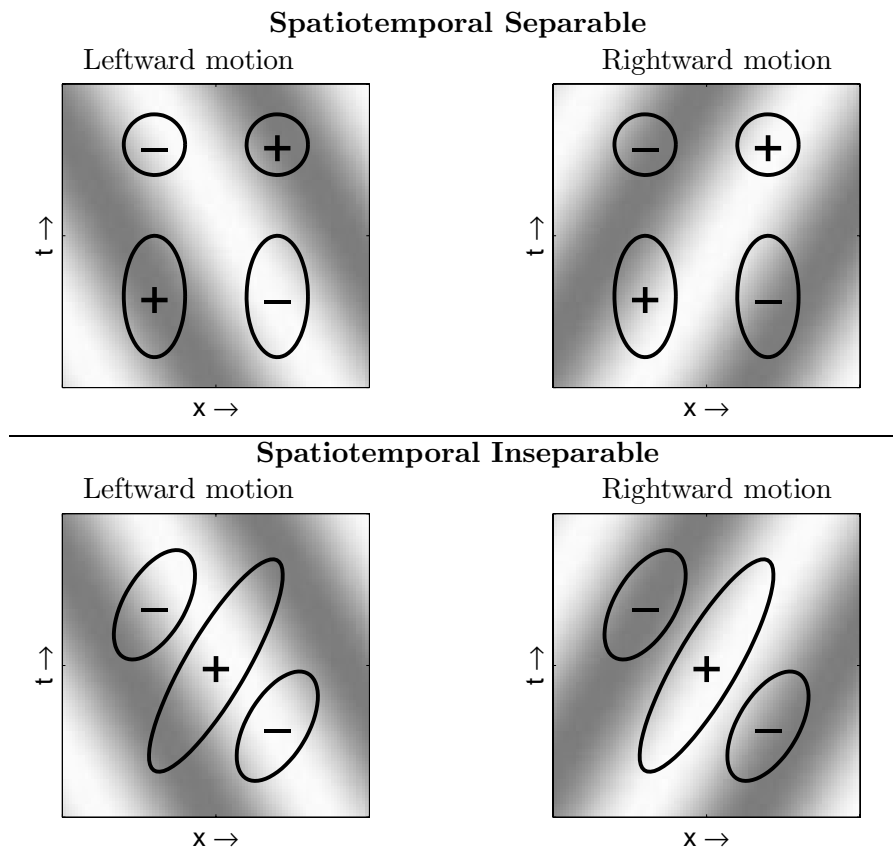


Figure 4.7: Spatiotemporal receptive field plots of separable (top) and inseparable (bottom) receptive fields, superimposed over spatiotemporal representations of drifting sinusoidal gratings. Gratings appear “oriented” because they shift in a particular spatial direction (x axis) as they move through time (y axis). The response of the cell is the product of the stimulus with the receptive field, summed over both space and time. Thus, for inseparable receptive fields, the stimulus at the preferred space-time “orientation”, i.e. moving the preferred direction of drift, is a more effective stimulus than one moving in the non-preferred direction.

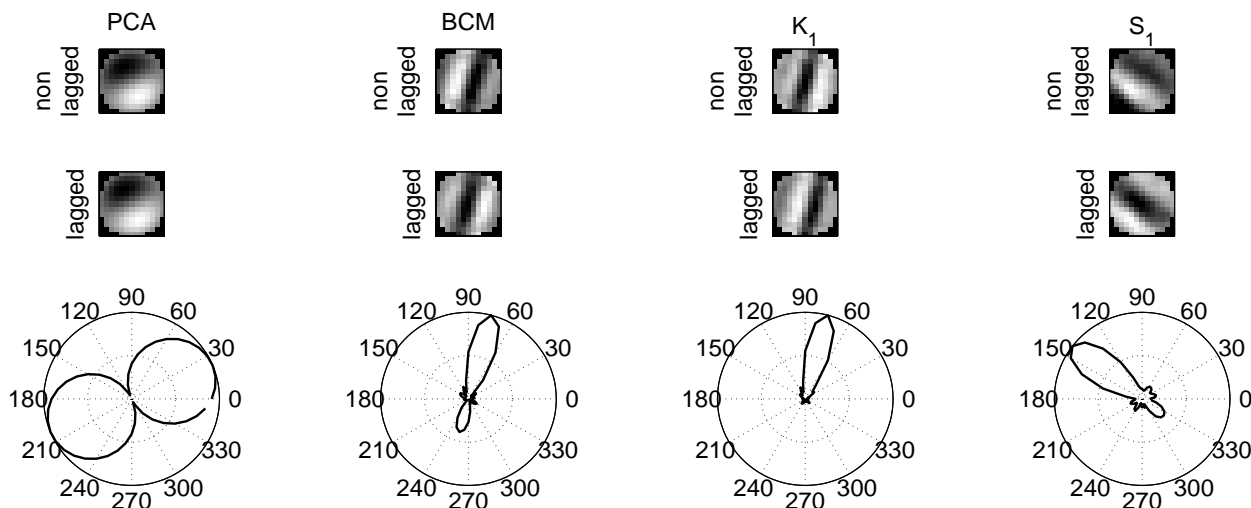


Figure 4.8: Example receptive fields and their orientation tuning, for a velocity of 2 pixels per iteration. The response of the cell, for a particular orientation of sine grating, is given by the radial component of the polar plots. The orientation tuning was obtained using drifting oriented sine gratings. Orientations larger than 180 degrees denotes motion in the opposite direction. Tuning curves which have a larger response for one direction than another are for direction selective cells. The PCA neuron is the only one which did not achieve direction selectivity.

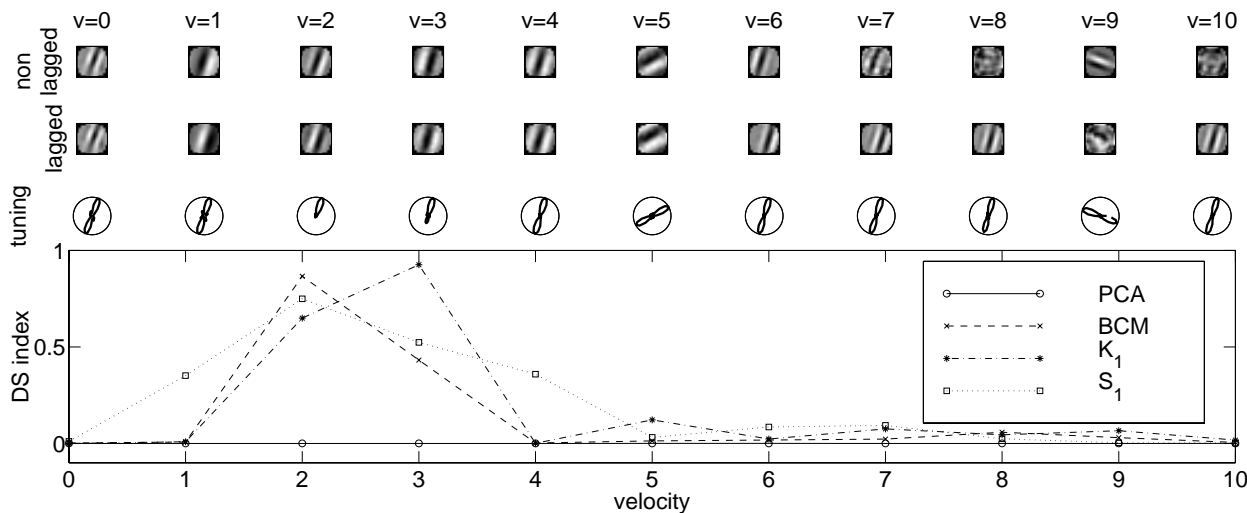


Figure 4.9: Example receptive fields with polar tuning plots (above) for BCM, for several eye drift velocities. The direction selectivity index as a function of velocity (below) for four different learning rules. The PCA learning rule did not develop direction selectivity. The other rules show some velocity tuning: they all lose direction selectivity for either velocities which are too high or too low. The LGN lagged cells had a constant 1 iteration lag.

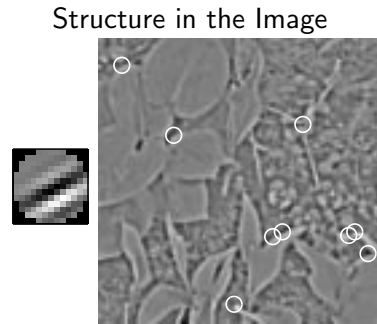


Figure 4.10: Patterns which yield high responses of a model neuron. The example receptive field is shown on the left. Some of the patterns which yield the strongest 1/2 percent of responses are labeled on the image on the right. These patterns are primarily the high contrast edges.

Structure Removal for BCM, Kurtosis, and Skew

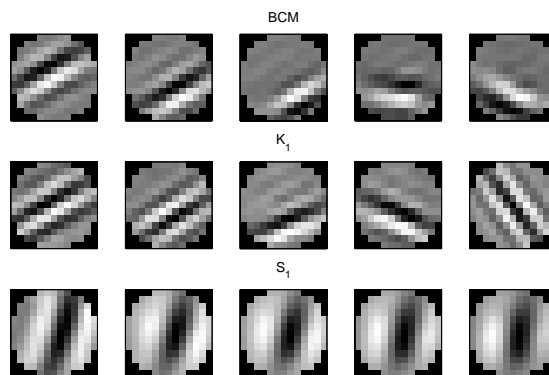


Figure 4.11: Receptive fields resulting from structure removal using the Quadratic BCM rule, the rule maximizing the multiplicative form of kurtosis and skewness. The RF on the far left for each rule was obtained in the normal input environment. The next RF to the right was obtained in a reduced input environment, whose patterns were deleted that yielded the strongest 1% of responses from the RF to the left. This process was continued for each RF from left to right, yielding a final removal of about five percent of the input patterns.

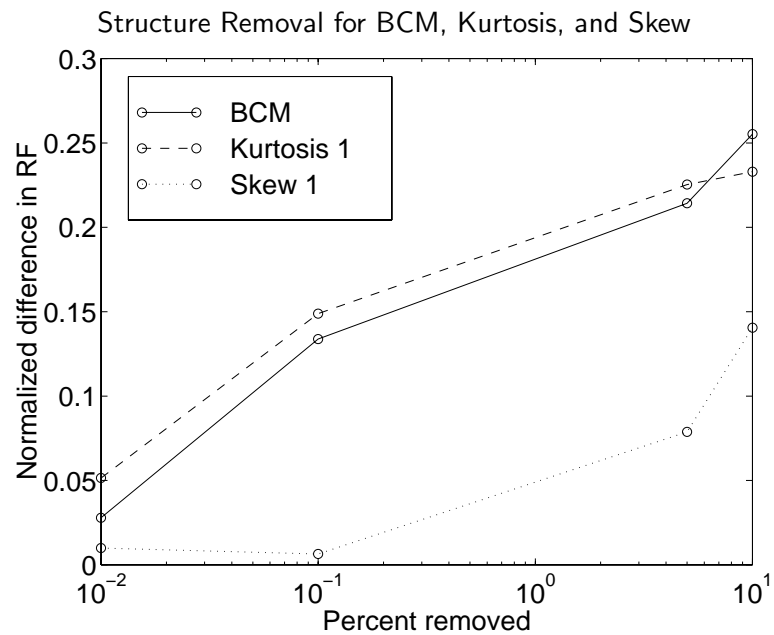


Figure 4.12: Normalized difference between RFs as a function of the percentage deleted in structure removal. The RFs were normalized, and mean zero, in order to neglect magnitude and additive constant changes. The maximum possible value of the difference is 1.