Predictive coding accounts for V1 response properties recorded using reverse correlation.

M. W. Spratling

King's College London, Department of Informatics and Division of Engineering, London. UK.

Abstract

PC/BC ("Predictive Coding/Biased Competition") is a simple computational model that has previously been shown to explain a very wide range of V1 response properties. This article extends work on the PC/BC model of V1 by showing that it can also account for V1 response properties measured using the reverse correlation methodology. Reverse correlation employs an experimental procedure that is significantly different from that used in more typical neurophysiological experiments, and measures some distinctly different response properties in V1. Despite these differences PC/BC successfully accounts for the data. The current results thus provide additional support for the PC/BC model of V1 and further demonstrate that PC/BC offers a unified explanation for the seemingly diverse range of behaviours observed in primary visual cortex.

Keywords: primary visual cortex; predictive coding; receptive field; orientation-tuning; spatial-frequency tuning

1 Introduction

Typical experiments to assess the response properties of cortical neurons employ a "forward correlation" method (Jones and Palmer, 1987b). In such experiments, the activity of a neuron is recorded in response to a number of different stimuli, or to stimuli that differ in terms of a particular parameter value. These recordings are then used to calculate the *average response* evoked by each stimulus, or by a particular value of a stimulus parameter. Previous results (Spratling, 2010, 2011) have demonstrated that a simple functional model (PC/BC) can simulate a very wide range of response properties that have been recorded, in primary visual cortex (V1), using this forward correlation technique.

Reverse correlation is an alternative technique used by psychologists to test the perception of human subjects (Neri and Heeger, 2002), and by neuroscientists to test the behaviour of individual neurons in the brain (Ringach and Shapley, 2004). Reverse correlation has a long history in neuroscience, as reviewed by Ringach (2004). In such experiments, the activity of a neuron is recorded in response to a continuous stream of different stimuli, or to stimuli that differ in terms of a particular parameter value. These recordings are then typically used to calculate the *average stimulus* that generated a response, or the response-weighted average of the stimulus parameter values, although other analyses are also possible (*e.g.*, Nykamp and Ringach, 2002; Schwartz et al., 2006; Sharpee et al., 2008; Smyth et al., 2003). This article extends the previous work with the PC/BC model by showing that it can also simulate response properties of V1 cells recorded using the reverse correlation technique. The PC/BC model provides the first, functional, account of this neurophysiological data.

Reverse correlation has been used to explore similar response properties to those investigated using forward correlation (*e.g.*, orientation and spatial frequency tuning). However, the results generated using these two techniques are sometimes incongruous. PC/BC successfully accounts for these distinct results. Reverse correlation can also be used to investigate neurophysiological properties that can not be measured using forward correlation (*e.g.*, reconstructing the spatial RF of a neuron or measuring dynamical changes in tuning properties). PC/BC also successfully simulates these V1 response properties. Despite the relative simplicity of the PC/BC model, the current results coupled with the results of previous work (Spratling, 2010, 2011) shows that it is able to provide a very comprehensive account of the behaviour of V1.

2 Methods

2.1 The Computational Model

The PC/BC model of V1 is illustrated in Figure 1. Full implementation details are given in the Appendix or can be found in (Spratling, 2010, 2011)[1], however, very briefly the model operates as follows.

---

[1]The implementation used here was identical to that described in Spratling (2010) except that parameter ϵ_2 was given a value of 250 rather than 50. The implemented model was also exactly as that described in Spratling (2011), except that parameter κ was given a value of 2π rather
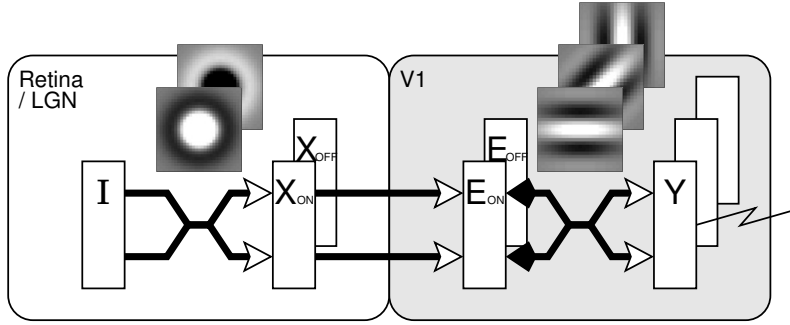
Figure 1: The PC/BC model of V1 (right) and the model of retina/LGN (left).

The input image I is preprocessed by convolution with a circular-symmetric on-centre/off-surround kernel (to generate the input, $\mathbf{X}_{on}$, to the ON channel of the V1 model), and a circular-symmetric off-centre/on-surround kernel (to generate the input, $\mathbf{X}_{off}$, to the OFF channel of the V1 model). The prediction neurons (labelled $\mathbf{Y}$) represent V1 simple cells. Responses from these neurons were recorded during the simulations. These responses were generated by convolving the outputs of the ON and OFF channels of the error-detecting neurons (labelled $\mathbf{E}$) with the ON and OFF channels of a number of kernels, defined using Gabor functions, representing V1 RFs. This convolution process effectively reproduces the same RFs at every pixel location in the image. The responses of the error-detecting neurons are influenced by divisive feedback from the prediction neurons, which is also calculated by convolving the prediction neuron outputs with the weight kernels. The size of the RF of a model prediction neuron is measured in pixels, rather than degrees of visual angle. The size of input stimuli are thus also specified in terms of pixels.

The PC/BC algorithm was applied recursively to continuously update the neural responses ($\mathbf{Y}$ and $\mathbf{E}$). During this time a stream of input images was presented to the model. Each image in this sequence was presented for a fixed number of iterations of the PC/BC algorithm: two iteration were used in the simulation described in sections 3.2 and 3.4, and three iterations were used in the simulations described in sections 3.1 and 3.3. In comparison, the presentation times used in the corresponding neurophysiological experiments ranged from 10ms to 20ms (for those experiments described in sections 3.2 and 3.4), and from 16.7ms to 25ms (for those experiments described in sections 3.1 and 3.3). At the end of the presentation period for one image, the input to the network was instantaneously changed to the next image in the sequence. During the presentation of the image sequence the values of $\mathbf{Y}$ were recorded for all 32 prediction neurons with RFs centred in the middle of the image. These values were used, as described below, to calculate the reverse correlation between the response of one particular prediction neuron and parameters of the input. For this purpose, the time varying response of a single model neuron will be denoted by the vector $\mathbf{r}$. In some simulations multiple trials were performed, in which neural responses were recorded to the same set of images presented in a different random order. Results are presented using the mean of the reverse correlation calculated in each individual trial.

2.2 Reverse Correlation For Mapping Spatial RFs

In section 3.1 reverse correlation is used to reconstruct the spatial RF of the recorded neuron. The reconstructed RFs were determined by calculating the cross correlation between the stimulus and the response at a particular time-lag, τ, between the iteration in which the stimulus was present and the iteration at which the response was recorded. For non-negative values of τ, the spatial reverse correlation was calculated as:

$$RF(\tau) = \frac{\sum_{t=1+\tau}^{T} \mathbf{r}_t \bar{I}_{t-\tau}}{T - \tau}$$

Where $\bar{I}$ is the input image normalised such that the mean luminance of the stimulus set is zero. It was also possible for there to be meaningful correlations at small, negative, time-lags. This occurs when images are presented for multiple iterations, meaning that the image presented at a small negative τ is not random, but has a high probability of being the same image as is presented at time zero. The same is also true in physiological experiments. However, in physiological experiments the long transmission delay between image presentation and the evoked response (which is not modelled here) is added to τ making it positive at all times for which there are meaningful

---

than 10 and the model complex cells were omitted.

correlations. For negative values of τ, the spatial reverse correlation was calculated as:

$$RF(\tau) = \frac{\sum_{t=1}^{T+\tau} \mathbf{r}_t \bar{I}_{t-\tau}}{T+\tau}$$

$RF(\tau)$ is effectively the mean of all stimuli weighted by the evoked response of the recorded neuron at time-lag τ. This is equivalent to the spike-triggered average (STA) method used in the neurophysiological literature, but adapted to work with rate-coded, rather than spiking, neurons.

Alternative methods to estimate the spatial RF are also employed in the neurophysiological literature. For example, the regularised pseudoinverse (Smyth et al., 2003), least-squares methods (Ringach et al., 2002; Smyth et al., 2000), maximally informative dimensions (Sharpee et al., 2008, 2004; Sharpee and Victor, 2009), and covariance analysis (Chen et al., 2007; Fairhall et al., 2006; Felsen et al., 2005; Schwartz et al., 2006; Touryan et al., 2005). The latter is applicable to the recovery of complex RFs composed of multiple sub-fields. The standard STA technique is adequate for the purposes of this paper, and hence, these more complex analysis methods are not applied to the PC/BC model.

2.3 Reverse Correlation For Measuring Orientation Tuning

In section 3.2 reverse correlation is used to determine the orientation tuning of the recorded neuron. Following the procedure used in the corresponding neurophysiological experiments (Ringach et al., 1997a, 2003; Shapley et al., 2003; Xing et al., 2005), for each time-lag, τ, an array of counters is defined, one for each orientation. Each element (or bin) in this array is initialised to zero. At each iteration, the response of the recorded neuron, $\mathbf{r}_t$, was added to the bin corresponding to the orientation of the stimulus appearing at time-lag τ. In Ringach et al. (2003); Xing et al. (2005) each bin was normalised by the total number of times each orientation appeared in the stimulus sequence, to give values of $p(\tau, \theta)$. The orientation tuning curve, $OT(\tau, \theta)$, was then calculated as:

$$OT(\tau, \theta) = log_{10}\left(\frac{p(\tau, \theta)}{p(\tau, \text{blank})}\right)$$

Where $p(\tau, \text{blank})$ is the normalised response to a blank stimulus. However, since the PC/BC algorithm does not simulate spontaneous activity, the value of $p(\tau, \text{blank})$ approaches zero at certain values of τ, which results in $OT(\tau, \theta)$ approaching infinity. Rather than modify the PC/BC model to incorporate a spontaneous firing rate an alternative measure of baseline activity was employed. Hence, for the model the orientation tuning curve, $OT(\tau, \theta)$, was calculated as:

$$OT(\tau, \theta) = log_{10}\left(\frac{p(\tau, \theta)}{p(-2, \theta)}\right)$$

Where $p(-2, \theta)$ is the reverse correlation with an image that was presented two iterations *after* the response was recorded, and hence, did not have had any causal influence on the response. The results were similar for any sufficiently negative time-lag. Hence, in contrast to Ringach et al. (2003) and Xing et al. (2005) the positive (and negative) value of $OT(\tau, \theta)$ correspond to parameter values that enhance (or suppress) the response in comparison to the average response, rather than in comparison to the response to a blank stimulus. This change to the method of analysis does not affect the value of parameter A (see below), as this value is independent of method used to normalise response (Xing et al., 2005, p. 802).

Following the procedure used in the corresponding neurophysiological experiments (Ringach et al., 2003; Shapley et al., 2003; Xing et al., 2005), a number of features of the $OT(\tau, \theta)$ tuning curves were then calculated at each time-lag, τ: the orientation, $\theta_{max}(\tau)$, and magnitude, $OT_{max}(\tau)$, of the maximum of $OT(\tau, \theta)$; the magnitude, $OT_{min}(\tau)$, of the minimum of $OT(\tau, \theta)$; the magnitude, $OT_{orth}(\tau)$, of $OT(\tau, \theta)$ at an orientation, $\theta_{orth}(\tau)$, orthogonal to $\theta_{max}(\tau)$; and the "modulation depth", $A(\tau) = OT_{max}(\tau) - OT_{min}(\tau)$. The time at which the $A(\tau)$ reached its maximum value was referred to as τ_{peak}, the time preceding τ_{peak} at which $A(\tau)$ reached half its maximum value was referred to as τ_{dev}, and the time following τ_{peak} at which $A(\tau)$ reached half its maximum value was referred to as τ_{dec}. Following the procedure used in Ringach et al. (2003), population averages of $OT_{max}(\tau)$, $OT_{min}(\tau)$, $OT_{orth}(\tau)$, and $A(\tau)$ were calculated after aligning the response from each neuron to have the same τ_{peak}. Tuning curves for individual neurons at τ_{dev} and τ_{dec} were calculated by linearly interpolating $OT(\tau, \theta)$ between iterations of the PC/BC algorithm.

2.4 Reverse Correlation For Measuring Orientation Tuning Shifts

In section 3.3 reverse correlation is used to determine the effect the orientation of a preceding stimulus has on the orientation tuning of the recorded neuron. Following the procedure used in the corresponding neurophysiological

experiments (Felsen et al., 2002), for each time-lag, τ, a 2-dimensional array of counters is defined, one for each possible combination of orientations, θ_1 and θ_2, appearing at time-lags τ and $\tau + 1$. Each element (or bin) in this array is initialised to zero. At each iteration, the response of the recorded neuron, $\mathbf{r}_t$, was added to the bin corresponding to the orientation transition occurring between time-lag τ and $\tau + 1$. Each bin was normalised by the total number of times each orientation transition occurred in the stimulus sequence. The normalised orientation tuning curve $p(\tau, \theta_1)$, calculated as described in a preceding paragraph, was then subtracted from each column of the array so as to isolate the tuning to θ_2.

To quantify the effect of θ_1 on the neuron's tuning to θ_2, each of the tuning curves defined over θ_2 for a different value of θ_1 was fitted with a Gaussian[2], in order to determine the preferred orientation of the neuron (the preferred orientation was the orientation of the peak of the fitted Gaussian). The value of $\mathrm{shift}(\theta_1)$ was calculated as the difference between the orientation preference calculated for each tuning curve and the orientation preference for the tuning curve found by taking the average over all values of θ_1. This value gives the shift in orientation preference for each value of θ_1.

2.5 Reverse Correlation For Measuring Spatial Frequency Tuning

In section 3.4 reverse correlation is used to determine the spatial frequency tuning of the recorded neuron. Following the procedure used in the corresponding neurophysiological experiments (Bredfeldt and Ringach, 2002), which is analogous to that used for measuring orientation tuning (section 2.3), for each time-lag, τ, an array of counters is defined, one for each spatial frequency. Each element (or bin) in this array is initialised to zero. At each iteration, the response of the recorded neuron, $\mathbf{r}_t$, was added to the bin corresponding to the spatial frequency of the stimulus appearing at time-lag τ. Each bin was normalised by the total number of times each spatial frequency appeared in the stimulus sequence, to give values of $p(\tau, f)$. The spatial frequency tuning curve, $FT(\tau, f)$, was then calculated as:

$$FT(\tau, f) = log_{10} \left(\frac{p(\tau, f)}{p(\tau, \mathrm{blank})} \right)$$

As for orientation tuning, to analyse the results of the PC/BC model we substitute $p(\tau, \mathrm{blank})$ with $p(-2, f)$ to give:

$$FT(\tau, f) = log_{10} \left(\frac{p(\tau, f)}{p(-2, f)} \right)$$

2.6 Code

Software, written in MATLAB, which performs the experiments described below is available at http://www.corinet.org/mike/Code/v1_reverse_correl.zip.
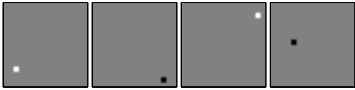
3 Results

3.1 Mapping Spatial RFs

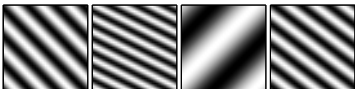Traditionally, reverse correlation has been used to map the spatial receptive fields of neurons in V1 and elsewhere in the nervous system. If a neuron can be considered to be a single linear filter, the output of which is transformed into a firing rate via a possibly nonlinear function, then the cross-correlation between the stimuli and the evoked response of the neuron (*i.e.*, the spike-triggered average (STA), see section 2.2) will provide an accurate estimate of the filter, as long as the stimulus set is spherically symmetric (Schwartz et al., 2006). Clearly this linear-nonlinear characterisation does not hold true for neurons in V1, due to the fact that there are multiple nonlinear operations performed along the visual pathway leading to V1, that V1 neurons interact with other neurons in the same population, and that their response properties are also influenced by neurons in other cortical areas via feedback connections. However, when STA is applied to V1 neurons, spatial filters are recovered that can be accurately described by a Gabor function (see below), and these filters are believed to provide an accurate estimate of the receptive field structure, except for a bias towards low spatial frequencies (Smyth et al., 2003). The underlying assumption is, therefore, that the deviations of V1 neurons from the linear-nonlinear characterisation have only a small influence on the recovered spatial receptive field when measured using reverse correlation with a Gaussian distributed stimulus set (Schwartz et al., 2006)[3].

---

[2]Gaussian curve fitting performed using the MATLAB function "fitgauss" by James R. Blake, available at http://www.mathworks.co.uk/matlabcentral/fileexchange/7489-fitgauss.

[3]The nonlinearity of V1 cells does, however, influence the recorded responses, which enables reverse correlation to also be used to measure the shape of the nonlinear transfer function (Chichilnisky, 2001; Nykamp and Ringach, 2002; Schwartz et al., 2006; Sharpee et al., 2008).
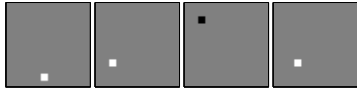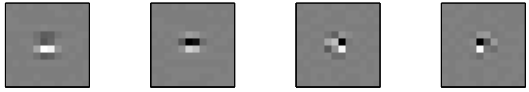
Stimulus type	Example stimuli	Reconstructed RFs	
Sparse Squares			
Sparse Bars			
M-Sequence			
Hartley			
RFs reconstructed from the recorded neuron's synaptic weights			

Table 1: RFs reconstructed for model prediction neurons. Each row shows typical example stimuli used to perform the reverse correlation analysis, and reconstructed RFs for two model prediction neurons. In each case images in the stimulus set were 32 by 32 pixels in size, and each image in the sequence was presented for three iterations of the PC/BC algorithm. The reconstructed RFs are for prediction neurons with even and odd symmetric horizontally aligned Gabor RFs. Reconstructed RFs are shown for the time-lag at which the variance of the reconstructed RF was greatest, which was at $\tau = 0$ for both sparse noise stimulus sets, and $\tau = 1$ for Hartley stimuli. For m-sequence stimuli the reverse correlation was only calculated at, and hence the RF is shown at, $\tau = 0$. The bottom row shows, for comparison, the RF reconstructed from the weights of the two prediction neurons. This reconstruction is obtained by subtracting the OFF weights from the ON weights after each set of weights have been convolved with the LoG filter used to simulate the processing performed by LGN. Each panel showing a reconstructed RF has been scaled separately to occupy the full range of gray values. In each case, mid-gray represents zero, lighter pixels represent negative values and darker pixels represent positive values.

Neurons in the PC/BC model are also not accurately characterised by the linear-nonlinear model. Specifically, the filters used to define the RFs of prediction neurons in the PC/BC model are not linear since neighbouring prediction neurons interact in a mutually suppressive and nonlinear manner. It is this nonlinear behaviour that gives rise to surround suppression and cross-orientation suppression when the PC/BC model is used to simulate forward correlation experiments (Spratling, 2010, 2011). It is therefore unknown whether or not the receptive field structure of these neurons can be accurately recovered using reverse correlation. The results presented below show that it is possible. Given that the prediction neurons in the model have been predefined to have Gabor RFs, it is not surprising that reverse correlation recovers Gabor-like RFs. What is significant is that these RFs can be reconstructed using reverse-correlation. Hence, the current results indicate that the PC/BC model contains appropriate mechanisms to simulate V1 response properties such as surround suppression and cross-orientation suppression in forward correlation experiments, while still being able to simulate V1 RF mapping experiments performed using reverse correlation.

A number of different stimulus types have been used for mapping V1 RFs (Ringach and Shapley, 2004). Four of these stimulus sets have been applied to the PC/BC model, as described below, and the results are summarised in Table 1.

Sparse-noise stimuli consist of a $p \times p$ grid of squares within which a single, randomly selected, light or dark rectangle (of width 1 square and length q squares) is present in each image. Examples of such images are shown in the first and second rows of table 1. Typically, an initial qualitative analysis of the neuron's RF is performed to enable the grid to be centred on the RF, scaled to the approximate size of the RF, and oriented to align with the RF (DeAngelis et al., 1993; Jones and Palmer, 1987b). A range of grid and stimulus sizes have been used, for example Yeh et al. (2009) used a $p = 12$, $q = 1$; Jones and Palmer (1987b) used $p = 16$, $q = 1$, 3, or 5; and DeAngelis et al. (1993) used $p = 20$, $q = 2$. For the simulations reported here $p = 16$ was used and each square was two pixels in size. Separate results for $q = 1$ and $q = 3$ are presented. The image set consisted of 512 ($16 \times 16 \times 2$) or 448 ($16 \times 14 \times 2$) images (for $q = 1$ and 3 respectively), and the reverse correlation was

Stimulus type	Example stimuli	Reconstructed RFs
Sparse Squares		
RFs reconstructed from the recorded neuron's synaptic weights		

Table 2: RFs reconstructed for model prediction neurons using stimuli similar to those used by Yeh et al. (2009). Reconstructed RFs are shown for the time-lag at which the variance of the RF was greatest, which was at $\tau = 0$. The format is identical to, and described in the caption of, Table 1.

averaged over 15 trails. For each trial all images in the data set were presented in a random order. Consistent with the majority of the neurophysiological experiments (DeAngelis et al., 1993; Jones and Palmer, 1987a,b) the reconstructed RFs were Gabor-like (table 1, rows 1 and 2). The RFs reconstructed from reverse correlation with sparse noise stimuli are slightly smaller than the true extent of the area from which the neuron receives non-zero afferent weights (as shown in the bottom row of table 1). This is to be expected since the sparse noise stimuli are unlikely to strongly activate the extremes of the RF.

Yeh et al. (2009) found that for cells in superficial layers of V1, sparse noise stimuli produced more circular RFs. However, Yeh et al. (2009) used a coarse, 12×12 square, grid that (in contrast to the experiments reported and simulated above) was not scaled or aligned to the RF of the recorded neuron. The RFs reconstructed from the PC/BC model using sparse noise are also more circular when measured using similar conditions ($p = 12$, $q = 1$, and each square equal in size to four pixels), see table 2. However, in the model the reconstructed RFs always contain separate ON and OFF sub-regions, whereas Yeh et al. (2009) found a single OFF region. The model, thus, does not explain the black-dominant responses observed in cortex (Xing et al., 2010; Yeh et al., 2009). However, given that the current model treats the ON and OFF channels identically, this is unsurprising.

An m-sequence is a pseudo-random sequence of binary digits which can be used to construct a pseudo-random sequence of images in which each pixel has high or low luminance (Reid et al., 1997). Examples of such images are shown in the third row of table 1. For this simulation an m-sequence of order 16 was used to generate a set of 65535 binary images. As in the corresponding neurophysiological experiments (Reid et al., 1997; Usrey et al., 2003), each image consisted of a 16×16 grid of black and white squares. Each square was two pixels in size. Each image was presented once (*i.e.*, only one trial was performed) and due to computer memory limitations only the correlation at a time lag of zero was calculated. The reconstructed RFs were Gabor-like (table 1, row 3), and were very similar to those produced using sparse noise, but they provided a more accurate estimate of the true size of the RF.

Ringach (2002) employed sinusoidal grating images with various spatial frequencies, spatial phases and orientations taken from a two dimensional Hartley basis function set (Ringach et al., 1997b) to estimate the RFs of cells in macaque V1. Examples of such images are shown in the fourth row of table 1. Using a similar Hartley basis function set containing 578 (32 by 32 pixel) sinusoidal gratings, produced reconstructed RFs for the PC/BC model that were Gabor-like (table 1, row 4). The shape of the reconstructed RF was closer to the true shape of the RF than was found using the other stimulus sets. However, the RFs reconstructed from reverse correlation with grating stimuli were also slightly smaller than the true extent of the area from which the neuron receives non-zero afferent weights (as shown in the bottom row of table 1). Mapping the RF using a forward correlation technique with circular gratings of different diameters also underestimates the true extent of the RF due to surround suppression (Spratling, 2010, 2011).

The neurons in the current model have been defined to have RFs that are all identical in terms of the ratio of width and length. These RFs therefore, by definition, do not show the range of RF shapes measured in V1 (Ringach, 2002). However, when the PC/BC model is trained on natural images, the RFs learnt by the model do show a close similarity to the distribution of RF shapes measured in V1 (Spratling, 2012).

A subset of V1 cells are found to have RFs that shift position over time, *i.e.*, RFs that are tilted in the space-time domain (De Valois et al., 2000; DeAngelis et al., 1995). Such neurons are motion sensitive. Given that the current model does not include any mechanisms for modelling direction selectivity, or for processing temporal information of any kind, it is not surprising that RFs in the current model do not shift over time, as shown in Fig. 2. Non-direction sensitive neurons in V1 form two sub-classes: monophasic and biphasic (De Valois et al., 2000; DeAngelis et al., 1995). Both types of cell have RFs that do not shift position over time. However, for the
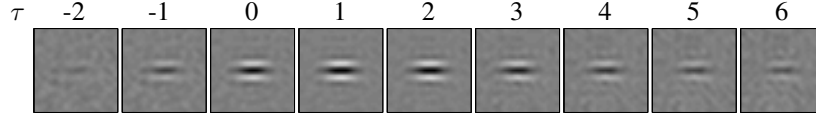
Figure 2: The RF of a single model prediction neuron reconstructed different time-lags (from $\tau = -2$ on left to $\tau = +6$ on right). Reverse correlation was performed using the Hartley subspace stimulus set.

smaller sub-population of biphasic simple cells the polarity of the RF reverses over time. The biphasic response of this class of V1 cell is believed to be inherited from the, transient, biphasic temporal response of magno-cellular cells in the LGN (De Valois et al., 2000). Whereas the monophasic response of the other sub-population of V1 cells is believed to be driven by the sustained response of parvo-cellular cells in the LGN (De Valois et al., 2000). Since the current model does not simulate separate sub-populations of LGN cell, it is unsurprising that it does not simulate both monophasic and biphasic responses in V1. However, the biphasic responses of LGN cells can also be explained by a predictive coding account (Jehee and Ballard, 2009), and hence, it should be possible to extend the PC/BC model, using similar mechanisms, to model biphasic LGN and V1 responses.

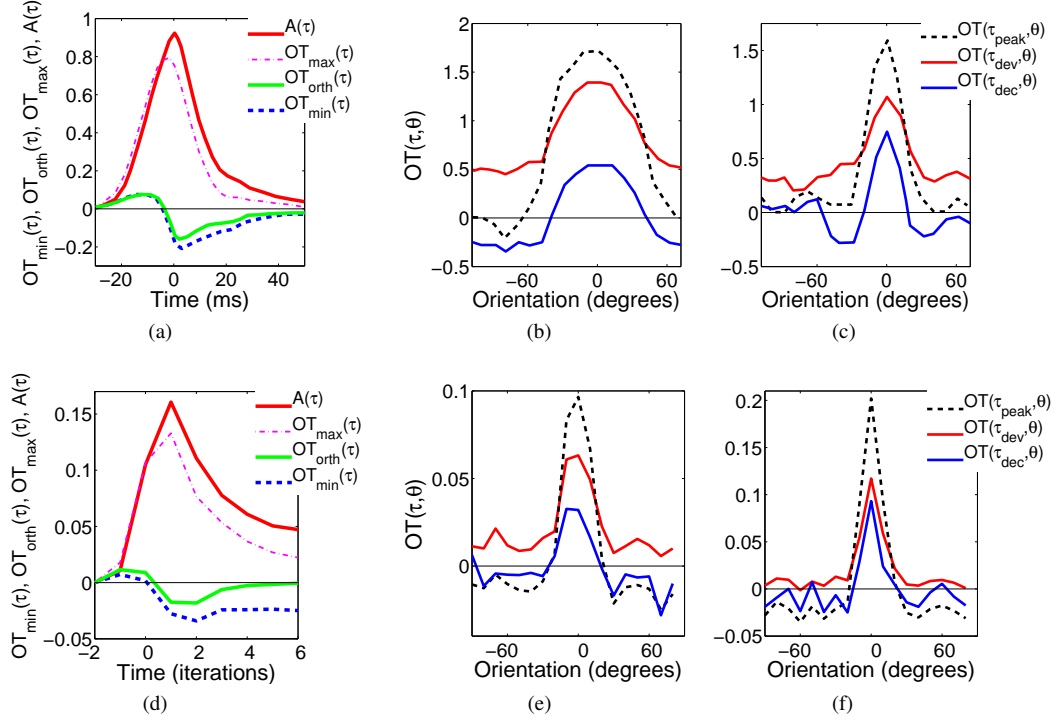## 3.2   Measuring Orientation Tuning

In the previous section, reverse correlation with the pixel intensity values has been used to map the spatial RF properties of neurons. It is also possible to perform reverse correlation with other properties of the stimuli. Particularly, it is possible to calculate the reverse correlation with the orientation of grating stimuli in order to measure the orientation tuning of cells in V1.

To assess orientation tuning, and the changes in orientation tuning that occur over time, Ringach *et al* (Ringach et al., 2003; Shapley et al., 2003; Xing et al., 2005) used a stimulus set consisting of 146 images of high-contrast sinusoidal gratings (at 18 orientations and 8 spatial phases), plus 8 blank stimuli. All gratings had the same spatial frequency (which was the preferred spatial frequency of the recorded neuron), and were 2 to 4 times (Ringach et al., 2003) or 2 to 5 times (Xing et al., 2005) the diameter of the optimally-oriented circular grating that generated the peak response from the recorded neuron (*i.e.*, on average 3 to 3.5 times the size of the summation field). An identical stimulus set was used in the simulations. The diameter of each grating was 45 pixels, which is approximately 3.5 times the summation field of prediction neurons in the model (see Spratling, 2010, Fig. 5b). Figure 3 shows that the orientation tuning dynamics of the model are in close agreement with those found in the corresponding neurophysiological experiments (Ringach et al., 2003; Shapley et al., 2003; Xing et al., 2005). Definitions of the abbreviations used in the figures and in the text below can be found in section 2.3.

Key features of the neurophysiological data identified by Ringach et al. (2003) were as follows:

- $OT(\tau_{dev}, \theta) > 0$ for all orientations, indicating that a grating at any orientation initially generated an increase in response. This can be seen in Fig. 3a by the initially positive values of OT_{min} and OT_{orth}, and in the example tuning curves shown in Figs. 3b and c in which the red/upper solid curves are positive for all orientations.

- $OT_{min}(\tau_{dec}) < 0$ indicating that suppression was generated at longer time-lags at some orientations. This can be seen in Fig. 3a by the negative values of OT_{min} at longer time-lags, and in the example tuning curves shown in Figs. 3b and c in the negative values of the blue/lower-solid tuning curves.

- Suppression could be untuned, meaning it was strong at all orientations away from the preferred one (see the blue/lower-solid tuning curves in the example in Fig. 3b).

- Suppression could be tuned, resulting in Mexican-hat profile to the tuning curve (see the blue/lower-solid tuning curves in example in Fig. 3c). This tuned suppression is also visible in the divergence of OT_{min} and OT_{orth} at longer time-lags in Fig. 3a.

- Suppression was rapid, reaching its peak only shortly after τ_{peak}, as shown in Fig. 3a by the values of OT_{min} and OT_{orth} reaching their most negative values just after $A(\tau)$ reaches its maximum value.

Many of these characteristics are shared by the prediction neurons in the PC/BC model (Fig. 3d–f). Specifically, suppression was greater (and hence, OT_{min} more negative) during the decay phase, compared to the development phase, of the response. The tuning curves show features of both untuned and tuned suppression, although these co-occur in the same neurons, rather than in separate neurons. Suppression was rapid. However, in the model, in contrast to the neurophysiological data, suppression is almost as strong at τ_{peak} as it is at τ_{dec}. This suggests that
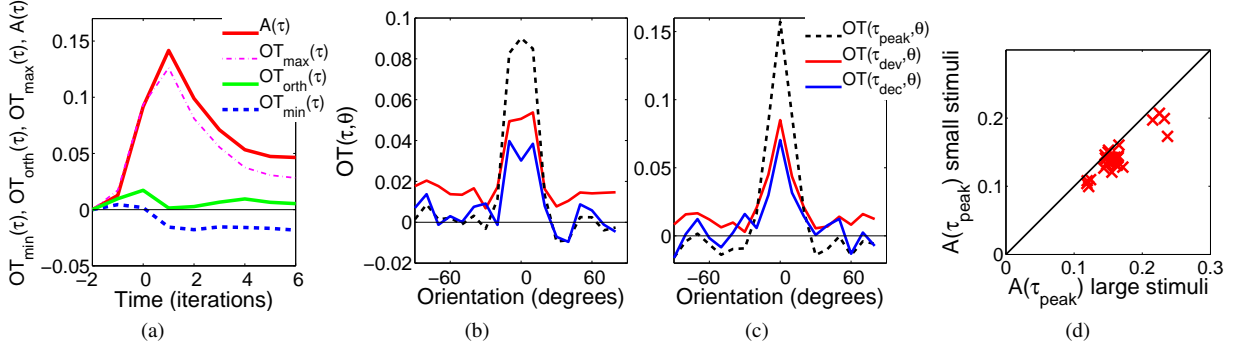
Figure 3: The temporal dynamics of orientation tuning, measured using stimuli with diameter significantly larger than the diameter of the summation field. Top row shows neurophysiological data recorded from macaque V1 (adapted from Shapley et al., 2003, Figs 4 and 5). Bottom row shows simulation data. First column shows the change over time in four parameters of the orientation tuning curve: $A(\tau)$ (red/upper-solid line), $OT_{max}(\tau)$ (magenta/dash-dot line), $OT_{orth}(\tau)$ (green/lower-solid line), and $OT_{min}(\tau)$ (blue/dashed line). Results are population averages for both the neurophysiological and simulation data. For the neurophysiological data time is measured relative to time τ_{peak}. For the simulation data time is measured in terms of the time-lag, τ. Second and third columns show tuning curves for two typical neurons from the V1 population and for two example neurons from the PC/BC model. In each case, the tuning curve is shown for three different times: at τ_{dev} (red/upper-solid line), τ_{peak} (black/dashed line), and τ_{dec} (blue/lower-solid line). The orientation is measured relative to the preferred orientation of the recorded neuron. See section 2.3 for definitions of $A(\tau)$, $OT_{max}(\tau)$, $OT_{orth}(\tau)$, $OT_{min}(\tau)$, $OT(\tau, \theta)$, τ_{dev}, τ_{peak}, and τ_{dec}.

in the model either inhibition increases too rapidly, or decays too rapidly. Given that the model does not include any factors, such as transmission delays, to simulate the temporal dynamics of cortical neurons it is surprising that the dynamics of the model are as close as they are to the neurophysiological data.

Xing et al. (2005) repeated the reverse correlation experiments with oriented gratings, using gratings matched to the size of the recorded neurons summation field. It was found that the peak "modulation depth", $A(\tau_{peak})$, was smaller for gratings matched to the RF size compared to that measured for gratings larger than the RF size. Repeating the simulations using gratings with a diameter equal to the diameter of the summation field (*i.e.*, 13 pixels), produced results consistent with the neurophysiological data. The reduction in the magnitude of A can be seen by comparing the peak of the red/upper-solid line in Fig. 4a with that in Fig. 3a. The reduction in amplitude of $A(\tau_{peak})$ is small, as in V1 (see Xing et al., 2005, Fig. 7c), but it is consistent: Fig. 4d shows a scatter plot of the two values of $A(\tau_{peak})$ measured from the model prediction neurons when using the two different sized stimuli. As with the corresponding neurophysiological data (Xing et al., 2005, Fig. 4), for the majority of neurons $A(\tau_{peak})$ is larger when measured using larger stimuli, and hence, most of the points fall below the diagonal.

A key feature of the neurophysiological data identified by Xing et al. (2005) was that the response (in terms of both $OT_{max}(\tau)$ and $A(\tau)$ at, and prior to, τ_{peak} was greatest for large rather than small gratings, when measured using reverse correlation. This is the opposite of the results obtained when using forward correlation: in such experiments response to large gratings is smaller than that to small gratings (due to surround suppression). Exactly the same reversal of results is seen in the model: in the experiments reported here, $OT_{max}(\tau)$ and to a greater

Figure 4: The temporal dynamics of orientation tuning, measured using stimuli with diameter equal to the diameter of the summation field. (a), (b) and (c) show simulation results corresponding to those shown in the bottom row of Fig 3. The format is identical to, and described in the caption of, Fig 3. (d) shows a scatter plot of peak modulation depth, $A(\tau_{peak})$, measured with large and small gratings for each of the 32 prediction neurons recoded in the simulations. See section 2.3 for definitions of $A(\tau)$, $OT_{max}(\tau)$, $OT_{orth}(\tau)$, $OT_{min}(\tau)$, $OT(\tau, \theta)$, and τ_{peak}.

extent $A(\tau)$ at, and prior to, τ_{peak} was largest for large stimuli, whereas, in experiments using forward correlation response is suppressed for large stimuli (see Spratling, 2010, Fig. 5b). In the reverse correlation experiments the prediction neuron's response increases as the optimally oriented grating stimulates more of the neuron's RF. In the forward correlation experiments, as the size of the optimally oriented grating increases, so does the competition between the recorded prediction neuron and its neighbours resulting in a suppression of the response. Hence, the current results indicate that the PC/BC model contains appropriate mechanisms to simulate surround suppression in forward correlation experiments, while still being able to successfully simulate the lack of surround suppression in experiments performed using reverse correlation.

Similar neurophysiological experiments to those described above have reported a sharpening of orientation tuning over time, particularly when measured using large stimuli (Chen et al., 2005; Ringach et al., 1997a), and changes in the preferred orientation over time, particularly when measured using large stimuli (Chen et al., 2005; Ringach et al., 1997a). Neither effect is seen in the simulations, which are therefore more consistent with neurophysiological experiments that failed to replicate these effects in V1 (Mazer et al., 2002).

Similar neurophysiological experiments to those described above have also reported that orientation tuning was more selective for large stimuli compared to small stimuli (Chen et al., 2005). The PC/BC model does show less selective orientation tuning for small stimuli, but only when those stimuli are smaller than the summation field, and hence, smaller than those used in the neurophysiological experiments (Chen et al., 2005). Figure 5 shows the values of $p(\tau, \theta)$ at various time-lags, for a single prediction neuron when measured using stimuli with three different diameters. The PC/BC model does show more selective orientation tuning for large diameter gratings (> 13 pixels, Fig. 5a and b) compared to small diameter gratings (*e.g.*, 7 pixels, Fig. 5c). This is consistent with the results from forward correlation experiments with the same model (see Spratling, 2010, Fig. 9).

3.3 Measuring Orientation Tuning Shifts

By measuring the reverse correlation with the joint distribution of orientations θ_1 and θ_2, appearing at time-lags τ and $\tau + 1$, Felsen et al. (2002) were able to show that orientation of the preceding stimulus affects the tuning to the following stimulus. The effect is a repulsive shift in orientation tuning away from the orientation of the preceding stimulus, Fig. 6a. The stimulus set consisted of 48 images of 100% contrast sinusoidal gratings (at 12 orientations and 4 spatial phases). All gratings had the same spatial frequency (which was the preferred spatial frequency of the recorded neuron), and "completely covered the RF" (Felsen et al., 2002). An identical stimulus set, with gratings of diameter 45 pixels, was used in the simulation. Results of the simulation are in agreement to the neurophysiological results, Fig. 6b. However, just as orientation tuning width in the model is smaller than that recorded in V1 (compare Fig. 5a with Felsen et al., 2002, Fig. 2b), so is the range of orientations inducing a shift.
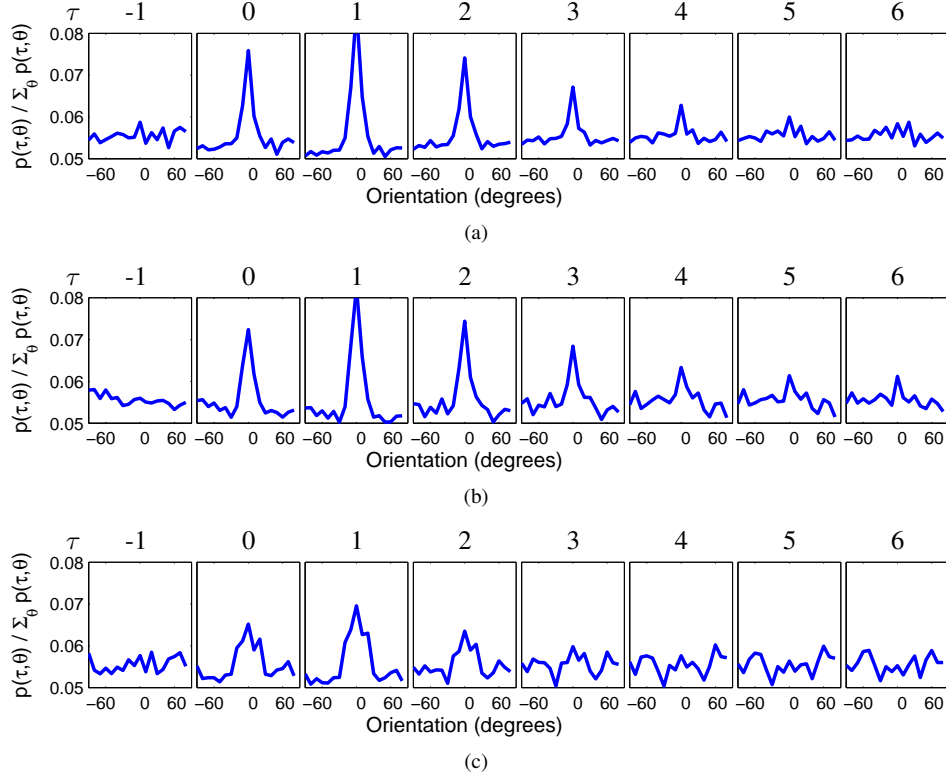
Figure 5: The orientation tuning of a single prediction neuron measured using reverse correlation with different sized stimuli: (a) 45 pixels, (b) 13 pixels, (c) 7 pixels. The $OT(\tau, \theta)$ values for the same neuron are shown in Figure 3c and Figure 4c. Values of $p(\tau, \theta)$ are normalised by the sum across all orientation values at that time-lag. Orientation tuning curves are shown for time-lags from $\tau = -1$ on left to $\tau = +6$ on right. See section 2.3 for the definition of $p(\tau, \theta)$.
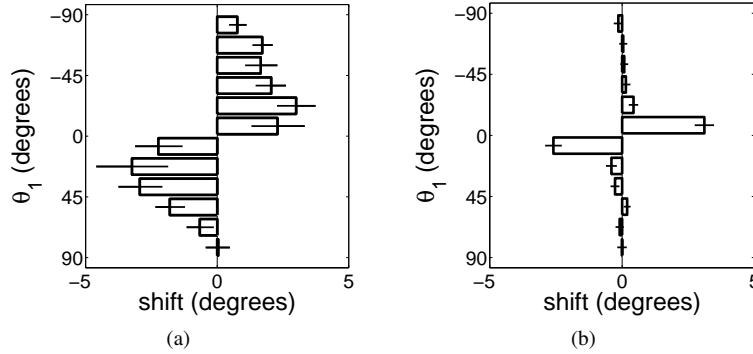
Figure 6: Shift in preferred orientation as a function of the orientation of the preceding stimulus relative to the orientation of average preferred stimulus. (a) Neurophysiological data recorded from cat V1 (adapted from Felsen et al., 2002, Fig 3b). (b) Simulation results. Bars show mean shift over all recorded neurons, lines show $\pm$ standard error of mean.

3.4 Measuring Spatial Frequency Tuning

To assess spatial frequency tuning, and the changes in spatial frequency tuning that occur over time, Bredfeldt and Ringach (2002) used a stimulus set consisting of high-contrast sinusoidal gratings at eight spatial phases, covering a range of spatial frequencies centred at the preferred spatial frequency of the recorded neuron. All gratings had the same orientation (which was the preferred orientation of the recorded neuron), and were 1.5 to 3 times the diameter of the optimally-oriented circular grating that generated the peak response from the recorded neuron (*i.e.*, the size of the summation field). A similar stimulus set was used in the simulations. The diameter
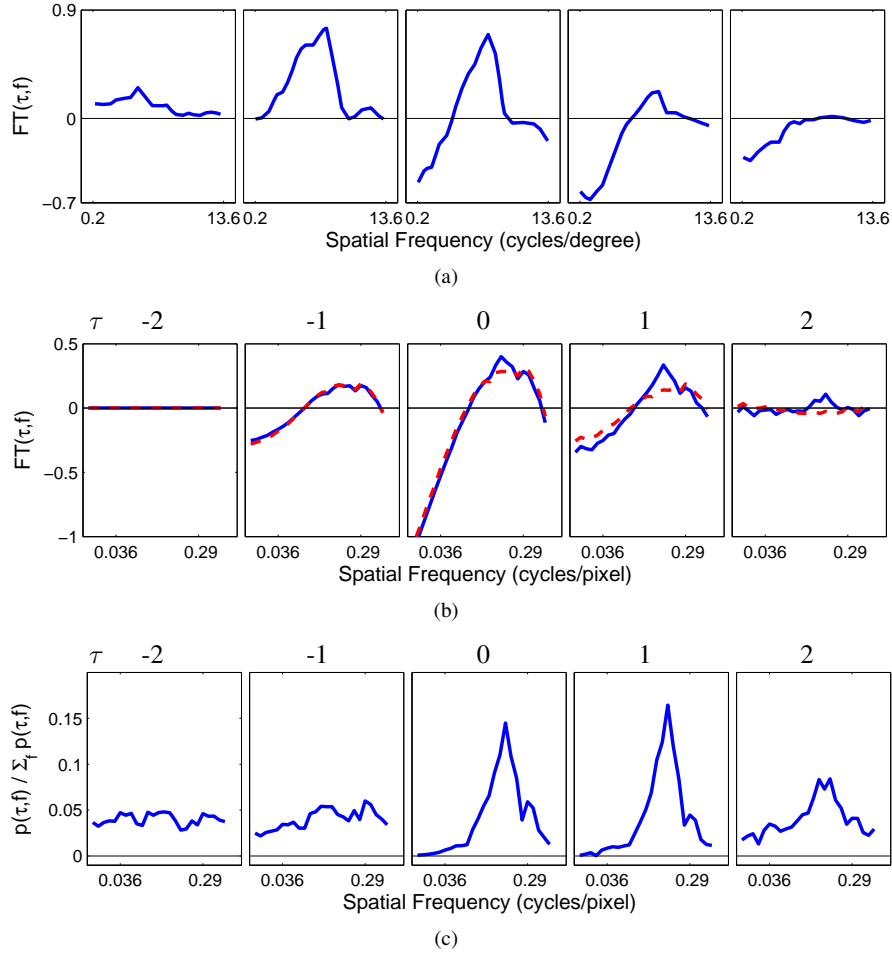
Figure 7: The temporal dynamics of spatial frequency tuning. Each subplot in a row shows the values of $FT(\tau, f)$ at a different time-lag τ. (a) Neurophysiological data recorded from primate V1 (adapted from Bredfeldt and Ringach, 2002, Fig 3a). $FT(\tau, f)$ shown (from left to right) for $\tau = $ 36, 48, 60, 72, and 84ms. (b) Simulation data showing $FT(\tau, f)$ at (from left to right) $\tau = $ -2, -1, 0, 1, and 2. The solid line shows the frequency tuning of a prediction neuron tuned to the orientation of the stimuli. The dashed line shows the frequency tuning of a prediction neuron tuned to an orientation perpendicular to that of the stimuli. (c) Values of $p(\tau, f)$ for the neuron tuned to the orientation of the stimuli after subtraction of $p(\tau, f)$ for the neuron tuned to the perpendicular orientation. Values were normalised by the sum of $p(\tau, f)$ across all orientation values at that time-lag. See section 2.5 for the definitions of $p(\tau, f)$ and $FT(\tau, f)$.

of each grating was 29 pixels, which is approximately 2.2 times the summation field of prediction neurons in the model (see Spratling, 2010, Fig. 5b). Figure 7 shows the spatial frequency tuning dynamics of a typical prediction neuron in the model. The simulation results share some features in common with the neurophysiological data. Specifically, there is strong suppression of low spatial frequencies and selectivity becomes sharper over time.

However, the simulation results suggest that prediction neurons in the model are broadly tuned to spatial frequencies, over the range 0.07 to 0.45 cycles/pixel. This is in contrast to the spatial frequency tuning measured using forward correlation which is more narrowly tuned, within the range 0.1 to 0.3 cycles/pixel, and shows a sharp peak at 0.17 cycles/pixel (see Spratling, 2010, Fig. 5d and e). However, the reverse correlation data seems to be dominated by the spatial frequency tuning of the linear filters used to simulate the LGN and which provide the input to the V1 model. The values of $FT(\tau, f)$ measured from the model were very similar irrespective of the orientation tuning of the recorded neuron. This can be seen from the dashed lines in Figure 7b, which show the spatial frequency tuning of a prediction neuron tuned to an orientation perpendicular to the orientation of the stimuli. Hence, the $FT(\tau, f)$ values recorded from the model seem to mostly uncorrelated with the orientation tuning, and hence, the spatial frequency tuning of the model simple cell.

To obtain the true spatial frequency tuning is is necessary to remove the component of the signal that is

uncorrelated with the orientation of the stimulus. This could be done by presenting stimuli at two orthogonal orientations, and subtracting the values of $p(\tau, f)$ obtained from stimuli that are orthogonal to the recorded neuron's preferred orientation, from the values of $p(\tau, f)$ obtained from stimuli with an orientation matched to the preferred orientation of the neuron. Given that the neurons in the model are perfectly uniform, we can perform an equivalent manipulation of the data by performing the reverse correlation experiment with stimuli at a single orientation, and subtracting the $p(\tau, f)$ obtained for a neuron with a preferred orientation at 90° to the stimuli from that obtained from the recorded neuron. This result is shown in Figure 7c. It can be seen that the orientation tuning measured using this technique is in much closer agreement to that measured using forward correlation: the tuning is narrower with response over the range 0.08 to 0.3 cycles/pixel, with a sharp peak at 0.16 cycles/pixel. Unless the baseline, $p(\tau, \text{blank})$, used to normalise the values of $FT(\tau, f)$ recorded in the neurophysiological data, also removes those components of the response that are uncorrelated to orientation, then the neurophysiological data might also suffer from the same issue.

4   Discussion

The procedure for carrying out a reverse correlation experiment is significantly different from that used in more typical forward correlation experiments. In the latter each stimulus is usually preceded by a blank screen, the stimulus is presented for a long duration (100s of milliseconds), and the recorded response is averaged over repetitions of the same stimulus parameters. Whereas in the former, each stimulus is preceded by another stimulus, each stimulus is presented briefly (10s of milliseconds), and the recorded response is used to calculate the response-weighted average of stimulus parameters (such as pixel intensity, orientation, *etc.*). Despite these significant differences, a single model has been shown, here and in previous publications (Spratling, 2010, 2011), to account for a wide-range of neurophysiological data recorded in V1 using both techniques.

The forward correlation experiments performed with the model have relied on the nonlinear behaviour of PC/BC to account for V1 response properties such as surround suppression and cross-orientation suppression. This nonlinear behaviour results from prediction neurons which provide an accurate representation of the stimulus, suppressing the response of other prediction neurons which are less able to accurately represent the stimulus. Such suppression takes several iterations (with an unchanging input) to become effective (Spratling, 2011). Hence in the PC/BC model, predictions activated by recent past experience, suppress incompatible responses to the current stimulus.

In reverse correlation experiments, the constantly changing input means that predictions activated by recent past experience are generally out-of-date. Assuming that there is no correlation between successive stimuli, the suppression received by the recorded neuron will be independent of the current stimulus parameters. Averaged over multiple presentations, the total suppression to any stimulus will be approximately the same (*i.e.*, the average suppression generated by a random set of preceding stimuli). Hence, in the experiments reported here, although suppression is active it goes un-noticed because it is approximately equal for each stimulus. The reverse correlations are thus dominated by the linear behaviour of the model. This behaviour is consistent with the behaviour observed in V1 when reverse correlation is used to map spatial RFs (see section 3.1), measure orientation tuning (see section 3.2), measure shifts in orientation tuning due to the orientation of preceding stimuli (see section 3.3), and measure spatial frequency tuning (see section 3.4).

The ineffectiveness of the nonlinear suppression mechanisms in the PC/BC model resulting from rapid stimulus presentation accounts for marked differences seen in the response properties of model V1 cells when measured using reverse correlation rather than forward correlation. Specifically, when orientation tuning is measured using forward correlation, response falls to zero for stimuli presented at orientations far from the preferred one (see Spratling, 2010, Fig. 5a). However, when orientation tuning is measured using reverse correlation, response remains strong at all orientations (Fig. 5). Similarly, surround suppression is evident in forward correlation experiments, resulting in a reduction in the response to large diameter stimuli (see Spratling, 2010, Fig. 5b). However, increasing the diameter of the stimulus in a reverse correlation experiment results in an increase in the strength of the response (Fig. 5). These changes in measured response properties resulting from changes to the experimental method are consistent with those seen in V1. The current results thus indicate that the PC/BC model contains appropriate nonlinear mechanisms to simulate V1 response properties such as surround suppression in forward correlation experiments, while also successfully simulating the lack of such nonlinear influences when V1 response properties are recorded using reverse correlation. PC/BC thus provides a single computational account for these seemingly inconsistent neurophysiological results.

Acknowledgements

References

Bredfeldt, C. E. and Ringach, D. L. (2002). Dynamics of spatial frequency tuning in macaque V1. *J. Neurosci.*, 22(5):1976–84.

Chen, G., Dan, Y., and Li, C.-Y. (2005). Stimulation of non-classical receptive field enhances orientation selectivity in the cat. *The Journal of Physiology*, 564(1):233–43.

Chen, X., Han, F., Poo, M.-M., and Dan, Y. (2007). Excitatory and suppressive receptive field subunits in awake monkey primary visual cortex (V1). *Proc. Natl. Acad. Sci. U.S.A.*, 104(48):19120–5.

Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Network*, 12:199–213.

Daugman, J. G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Res.*, 20:847–56.

Daugman, J. G. (1988). Complete discrete 2-D Gabor transformations by neural networks for image analysis and compression. *IEEE Trans. Acoust.*, 36(7):1169–79.

De Valois, R. L., Cottaris, N. P., Mahon, L. E., Elfar, S. D., and Wilson, J. A. (2000). Spatial and temporal receptive fields of geniculate and cortical cells and directional selectivity. *Vision Res.*, 40(27):3685–702.

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

DeAngelis, G. C., Ohzawa, I., and Freeman, R. D. (1995). Receptive-field dynamics in the central visual pathways. *Trends Neurosci.*, 18(10):451–8.

Fairhall, A. L., Burlingame, C. A., Narasimhan, R., Harris, R. A., Puchalla, J. L., and Berry, M. J. (2006). Selectivity for multiple stimulus features in retinal ganglion cells. *J. Neurophysiol.*, 96:2724–38.

Felsen, G., Shen, Y. S., Yao, H., Spor, G., Li, C., and Dan, Y. (2002). Dynamic modification of cortical orientation tuning mediated by recurrent connections. *Neuron*, 36(5):945–54.

Felsen, G., Touryan, J., Han, F., and Dan, Y. (2005). Cortical sensitivity to visual features in natural scenes. *PLoS Biol.*, 3(10):e342.

Jehee, J. F. M. and Ballard, D. H. (2009). Predictive feedback can account for biphasic responses in the lateral geniculate nucleus. *PLoS Comput. Biol.*, 5(5):e1000373.

Jones, J. P. and Palmer, L. A. (1987a). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiol.*, 58(6):1233–58.

Jones, J. P. and Palmer, L. A. (1987b). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J. Neurophysiol.*, 58(6):1187–211.

Lee, T. S. (1996). Image representation using 2D Gabor wavelets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(10):959–71.

Marcelja, S. (1980). Mathematical description of the responses of simple cortical cells. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.*, 70:1297–1300.

Mazer, J. A., Vinje, W. E., McDermott, J., Schiller, P. H., and Gallant, J. L. (2002). Spatial frequency and orientation tuning dynamics in area V1. *Proc. Natl. Acad. Sci. U.S.A.*, 99(3):1645–50.

Neri, P. and Heeger, D. J. (2002). Spatiotemporal mechanisms for detecting and identifying image features in human vision. *Nat. Neurosci.*, 5(8):812–6.

Nykamp, D. Q. and Ringach, D. L. (2002). Full identification of a linear-nonlinear system via cross-correlation analysis. *J. Vis.*, 2:1–11.

Reid, R. C., Victor, J. D., and Shapley, R. M. (1997). The use of m-sequences in the analysis of visual neurons: Linear receptive field properties. *Vis. Neurosci.*, 14:1015–27.

Ringach, D. and Shapley, R. (2004). Reverse correlation in neurophysiology. *Cogn. Sci.*, 28(2):147–66.

Ringach, D. L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *J. Neurophysiol.*, 88(1):455–63.

Ringach, D. L. (2004). Mapping receptive fields in primary visual cortex. *The Journal of Physiology*, 558(3):717–28.

Ringach, D. L., Hawken, M. J., and Shapley, R. (1997a). Dynamics of orientation tuning in macaque primary visual cortex. *Nature*, 387:281–4.

Ringach, D. L., Hawken, M. J., and Shapley, R. (2002). Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences. *J. Vis.*, 2:12–24.

Ringach, D. L., Hawken, M. J., and Shapley, R. (2003). Dynamics of orientation tuning in macaque V1: The role of global and tuned suppression. *J. Neurophysiol.*, 90(1):342–52.

Ringach, D. L., Sapiro, G., and Shapley, R. (1997b). A subspace reverse correlation method for the study of visual neurons. *Vision Res.*, 37:2455–64.

Schwartz, O., Pillow, J. W., Rust, N. C., and Simoncelli, E. P. (2006). Spike-triggered neural characterization. *Journal of Vision*, 6(4).

Shapley, R., Hawken, M., and Ringach, D. L. (2003). Dynamics of orientation selectivity in the primary visual cortex and the importance of cortical inhibition. *Neuron*, 38(5):689–99.

Sharpee, T. O., Miller, K. D., and Stryker, M. P. (2008). On the importance of static nonlinearity in estimating spatiotemporal neural filters with natural stimuli. *J. Neurophysiol.*, 99(5):2496–509.

Sharpee, T. O., Rust, N. C., and Bialek, W. (2004). Analysing neural responses to natural signals: maximally informative dimensions. *Neural Comput.*, 16:223–50.

Sharpee, T. O. and Victor, J. D. (2009). Contextual modulation of v1 receptive fields depends on their satial symmetry. *J. Comput. Neurosci.*, 26:203–18.

Smyth, D., Tolhurst, D. J., and Thompson, I. D. (2000). Simple-cell receptive field reconstruction from natural-scene stimuli in ferret primary visual cortex. *J. Physiol. (Lond.)*, 527:93P–94P.

Smyth, D., Willmore, B., Baker, G. E., Thompson, I. D., and Tolhurst, D. J. (2003). The receptive-field organization of simple cells in primary visual cortex of ferrets under natural scene stimulation. *J. Neurosci.*, 23:4746–59.

Spratling, M. W. (2010). Predictive coding as a model of response properties in cortical area V1. *J. Neurosci.*, 30(9):3531–43.

Spratling, M. W. (2011). A single functional model accounts for the distinct properties of suppression in cortical area V1. *Vision Res.*, 51(6):563–76.

Spratling, M. W. (2012). Unsupervised learning of generative and discriminative weights encoding elementary image components in a predictive coding model of cortical function. *Neural Comput.*, 24(1):60–103.

Touryan, J., Felsen, G., and Dan, Y. (2005). Spatial structure of complex cell receptive fileds measured with natural images. *Neuron*, 45:781–791.

Usrey, W. M., Sceniak, M. P., and Chapman, B. (2003). Receptive fields and response properties of neurons in layer 4 of ferret visual cortex. *J. Neurophysiol.*, 89(2):1003–15.

Xing, D., Shapley, R. M., Hawken, M. J., and Ringach, D. L. (2005). Effect of stimulus size on the dynamics of orientation selectivity in macaque V1. *J. Neurophysiol.*, 94(1):799–812.

Xing, D., Yeh, C.-I., and Shapley, R. M. (2010). Generation of black-dominant responses in V1 cortex. *J. Neurosci.*, 30(40):13504–12.

Yeh, C.-I., Xing, D., Williams, P. E., and Shapley, R. M. (2009). Stimulus ensemble and cortical layer determine V1 spatial receptive fields. *Proc. Natl. Acad. Sci. U.S.A.*, 106(34):14652–7.

Appendix: Computational Model Details

A.1 The LGN Model

The input to the PC/BC model of V1, described below, was an input image (I) pre-processed by convolution with a Laplacian-of-Gaussian (LoG) filter (l) with standard deviation equal to one pixel. This is virtually identical to the Difference-of-Gaussians (DoG) filter that has traditionally been used to model circular receptive fields (RFs) in LGN and retina. The output from this filter was subject to a multiplicative gain (the strength of which was determined by parameter κ) followed by a saturating non-linearity, such that:

$$\mathbf{X} = \tanh\left\{\kappa(I * l)\right\} \tag{1}$$

A value of $\kappa = 2\pi$ was used in all experiments reported here.

The positive and rectified negative responses were separated into two images $\mathbf{X}_{ON}$ and $\mathbf{X}_{OFF}$ simulating the outputs of cells in retina and LGN with circular-symmetric on-centre/off-surround and off-centre/on-surround RFs respectively. This pre-processing is illustrated on the left of Figure 1.

A.2 The V1 Model

The PC/BC model of V1 is illustrated in Figure 1 and described by the following equations:

$$\mathbf{E}_o = \mathbf{X}_o \oslash \left(\epsilon_2 + \sum_{k=1}^{p}(\hat{w}_{ok} * \mathbf{Y}_k)\right) \tag{2}$$

$$\mathbf{Y}_k \leftarrow (\epsilon_1 + \mathbf{Y}_k) \otimes \sum_o (w_{ok} \star \mathbf{E}_o) \qquad (3)$$

Where $o \in [ON, OFF]$; $\mathbf{X}_o$ is a 2-dimensional array, equal in size to the input image, that represents the input to the model of V1; E_o is a 2-dimensional array, equal in size to the input image, that represents the error-detecting neuron responses; $\mathbf{Y}_k$ is a 2-dimensional array, equal in size to the input image, that represent the prediction neuron responses; w_{ok} is a 2-dimensional kernel representing the synaptic weights for a particular class (k) of neuron normalised so that sum of all the weights was equal to ψ; $\hat{w}_{ok}$ is a 2-dimensional kernel representing the same synaptic weights as w_{ok} but normalised so that the maximum value was equal to ψ; p is the total number of kernels; ϵ_1, ϵ_2, and ψ are parameters; $\oslash$ and $\otimes$ indicate element-wise division and multiplication respectively; $\star$ represents cross-correlation (which is equivalent to convolution without the kernel being rotated 180°); and $*$ represents convolution (which is equivalent to cross-correlation with a kernel rotated by 180°). Parameter values $\psi = 5000$, $\epsilon_1 = 0.0001$ and $\epsilon_2 = 250$ were used in the simulations reported in this article.

Equation 3 describes the updating of the prediction neuron activations. The response of each prediction neuron is a function of its activation at the previous iteration and a weighted sum of afferent inputs from the error-detecting neurons. Equation 2 describes the calculation of the neural activity for each population of error-detecting neurons. These values are a function of the activity of the input to V1 divisively modulated by a weighted sum of the outputs of the prediction neurons in V1. The activation of the error-detecting neurons can be interpreted in two ways. Firstly, $\mathbf{E}$ can be considered to represent the residual error between the input and the reconstruction of the input generated by the prediction neurons. The values of $\mathbf{E}$ indicate the degree of mismatch between the top-down reconstruction of the input and the actual input (assuming ϵ_2 is sufficiently small to be negligible). When a value within $\mathbf{E}$ is greater than $\frac{1}{\psi}$ it indicates that a particular element of the input is under-represented in the reconstruction, a value of less than $\frac{1}{\psi}$ indicates that a particular element of the input is over-represented in the reconstruction, and a value of $\frac{1}{\psi}$ indicates that the top-down reconstruction perfectly predicts the bottom-up stimulation. A second interpretation is that $\mathbf{E}$ represents the inhibited inputs to a population of competing prediction neurons. Each prediction neuron modulates its own inputs, which helps stabilise the response of the prediction neurons, since a strongly (or weakly) active prediction neuron will suppress (magnify) its inputs, and hence, reduce (enhance) its own response. Prediction neurons that share inputs (*i.e.*, that have overlapping RFs) will also modulate each other's inputs. This generates a form of competition between the prediction neurons, such that each neuron effectively tries to block other prediction neurons from responding to the inputs which it represents.

The RF of a simple cell in primary visual cortex can be accurately modelled by a 2-dimensional Gabor function (Daugman, 1980, 1988; Jones and Palmer, 1987a; Lee, 1996; Marcelja, 1980). Hence, the Gabor function was used to define the weights of each kernel w_{ok}. A definition of a Gabor function of the form proposed by Lee (1996) was used, which includes a term to remove the D.C. response of the filter:

$$g(\sigma, \gamma, \lambda, \phi, \theta) = \exp\left\{-\frac{x'^2 + (y'/\gamma)^2}{2\sigma^2}\right\}\left[\cos\left\{\frac{2\pi y'}{\lambda} + \phi\right\} - \cos(\phi)\exp\left\{-\left(\frac{\pi\sigma}{\lambda}\right)^2\right\}\right]$$

Where $\sigma = 4$ (pixels) was a constant that defined the standard deviation of the Gaussian envelope (which determines the spatial extent of the RF); $\gamma = \frac{1}{\sqrt{2}}$ was a constant that defined the aspect ratio of Gaussian envelope (which determines the ellipticity of the RF); $\lambda = 6$ (pixels) was a constant that defined the wavelength of the sinusoid; ϕ was the phase of the sinusoid; and $x' = x\cos(\theta) + y\sin(\theta)$ and $y' = -x\sin(\theta) + y\cos(\theta)$ where θ defined the orientation of the RF. A family of 32 Gabor functions with eight orientations ($\theta = 0$ to 157.5^o in steps of 22.5^o) and four phases ($\phi = 0^o$, 90°, 180°, and 270°) were used to define the RFs of the neurons in the model. Each Gabor kernel was 21×21 pixels in size. The weights were separated into distinct ON and OFF channels which represented the positive and negative parts of the Gabor function using separate sets of non-negative weights. The cross-correlation and convolution performed in Equations 2 and 3 mean that neurons with these RFs are reproduced at every pixel location in the image, and consequently, that the size of the population of V1 cells simulated varies with image size.