A Model of Partial Reference Frame Transforms through Pooling of Gain-Modulated Responses

Authors

Kris De Meyer and Michael W. Spratling

Contact Details

Kris De Meyer, Michael Spratling
Department of Informatics
King's College London
Strand
London WC2R 2LS
United Kingdom
Tel.: + 44 20 7848 2437

Fax: + 44 20 7848 2932

Email: kris@corinet.org, michael.spratling@kcl.ac.uk

Running Title

Partial Reference Frame Transforms through Response Pooling

Abstract

In multimodal integration and sensorimotor transformation areas of posterior parietal cortex (PPC), neural responses often appear encoded in spatial reference frames that are intermediate to the intrinsic sensory reference frames, e.g., eye-centred for visual or head-centred for auditory stimulation. Many sensory responses in these areas are also modulated by direction of gaze. We demonstrate that certain types of mixed-frame responses can be generated by pooling gain-modulated responses – similarly to how complex cells in visual cortex are thought to pool the responses of simple cells. The proposed model simulates two types of mixed-frame responses observed in PPC: in particular, sensory responses that shift differentially with gaze in horizontal and vertical dimensions; and sensory responses that shift differentially for different start and end points along a single dimension of gaze. We distinguish these two types of mixed-frame responses from a third type in which sensory reposes shift a partial yet approximately equal amount with each gaze shift. We argue that the empirical data on mixed-frame responses may be caused by multiple mechanisms, and we adapt existing reference-frame measures to distinguish between the different types. Finally, we discuss how mixed-frame responses may be revealing of the local organisation of presynaptic responses.

Keywords

frame of reference, gain field, partially-shifting receptive field, neural network, predictive coding

Introduction

Posterior partietal cortex (PPC) is crucially involved in spatial awareness and the sensory guidance of actions towards spatial goals (Stein and Stanford 2008; Andersen and Cui 2009). Subregions of PPC are thought to integrate signals that come from different sensory modalities but are generated by the same object or event in the world. They also play an important role in the transformation of sensory information into motor responses, such as saccades and visually-guided reaching movements. Sensory signals from different modalities are encoded with respect to certain intrinsic frames of reference, e.g., eye-centred for visual and head-centred for auditory stimulation. In the course of multi-sensory integration and sensorimotor transformation, these different spatial encodings may require remapping across different frames of reference (Pouget et al. 2002). How the brain performs this remapping is a matter of considerable debate. It has been observed repeatedly, however, that within PPC the interaction of sensory signals from different modalities (and hence encoded in different frames of reference) may give rise to neural responses that appear encoded in intermediate or mixed frames of reference (Stricanne et al. 1996; Duhamel et al. 1997; Avillac et al. 2005; Mullette-Gillman et al. 2005; Schlack et al. 2005; Chang and Snyder 2010).

Many sensory responses in PPC are also modulated by variables such as eye, head or hand position (Andersen and Mountcastle 1983; Andersen et al. 1990; Galletti et al. 1995; Bremmer et al. 1997; Chang and Snyder 2010). For example, the sensitivity of a parietal cell to a visual stimulus may change as a function of the direction of gaze, without any changes in its spatial selectivity. This modulatory, often multiplicative interaction between signals has been termed Gain Field (GF) (Andersen and Mountcastle 1983), in parallel to the concept of Receptive Field (RF) which describes the spatial extent and profile of the sensory selectivity itself.

Both intermediate frames of reference and gain modulation have previously been explained in the context of basis function (BF) networks (Deneve et al. 2001; Pouget et al. 2002). BF decomposition is a generic mathematical method for approximating non-linear functions. The responses of gain-modulated neurons have the required characteristics to form BF sets, and hence may form the basis for the non-linear remapping required in multi-sensory integration and sensorimotor transformation (Pouget and Sejnowski 1997). In BF networks mixed-frame responses have been shown to arise naturally in the BF layer where sensory signals encoded in multiple frames of reference converge and are integrated with other signals such as eye position (Pouget *et al.* 2002). A change in eye position or gaze shift may result in sensory responses that shift only a fraction of the gaze shift when analysed in the intrinsic sensory reference frames. Crucially, these partial RF shifts are *proportional* to the magnitude of the gaze shift, regardless of the position of its start or end point.

Empirical data on mixed-frame responses have shown idiosyncrasies that are not easily explained within this framework. More specifically, neural responses that shift differentially with gaze in horizontal and vertical dimensions (Galletti et al. 1993; Duhamel *et al.* 1997), and neural responses that shift differentially for different start and end points along a single dimension of gaze (Stricanne *et al.* 1996; Mullette-Gillman *et al.* 2005). In the current article we propose an explanation for these *non-proportional* mixed-frame responses: they can be generated by pooling gain-modulated responses that are encoded in a single frame of reference. The pooling mechanism is analogous to that thought to be employed by complex cells in primary visual cortex (Spratling 2011). The model thus proposes that the same computational mechanisms operate across cortical regions.

We postulate a clear distinction between proportional and non-proportional mixed-frame responses: the former are generated as a consequence of the interaction of signals encoded in different spatial frames of reference; within a BF network they arise in the BF layer. In contrast, non-proportional mixed-frame responses are generated by the pooling of gain-modulated responses; in a BF network they may arise in the network layer pooling responses from the BF layer, rather than in the BF layer itself. One immediate hypothesis following from the contrast between the present modelling study and previous work with BF networks (Deneve et al. 2001; Avillac et al. 2005) is that proportional

and non-proportional mixed-frame responses are generated by different physiological mechanisms. In the current neurophysiological literature no such distinction is made. As a first step to identify these different mechanisms, we propose how existing measures used to analyse the spatial encoding of neural responses may be adapted to distinguish between proportional and non-proportional mixed-frame responses. Moreover, drawing further hypotheses from our simulation studies, we argue that non-proportional mixed-frame responses may be revealing of the local organisation of the presynaptic, gain-modulated responses.

The model used here is the non-linear predictive coding/biased competition (PC/BC) network, an implementation of the predictive coding (PC) theory of cortical function that is consistent with the biased competition (BC) theory of attention (Spratling 2008). PC provides an elegant theory of how perceptual information can be combined with prior experience in order to compute the most likely interpretation of sensory data. It is based on the principle of minimizing the residual error between bottom-up, stimulus-driven activity and top-down predictions generated from the internal representation of the world. We have previously demonstrated that multiplicative gain modulation arises naturally when two population-coded input signals converge in a PC/BC network (De Meyer and Spratling 2011). The gain modulation arises as a consequence of competitive interactions in one group of model neurons, the prediction nodes. Synaptic weights generating gain-modulated responses can be easily learned using an unsupervised learning rule. In the current study, the prediction node responses are pooled together by another class of model neurons (the disjunctive nodes) using a weighted-max operation. This pooling function has previously been used to model how complex cells in primary visual cortex may generate their response properties by pooling the responses of simple cells (Spratling 2011). The same pooling function has been used in a PC/BC network trained to perform reference frame transformations in a simplified problem setting (Spratling 2009). Here we use the PC/BC network to replicate neurophysiological data of non-proportional mixed-frame responses in PPC. We discuss how the model makes predictions about local cortical organisation, and propose how existing measures used to quantify intermediate frames of reference could be adapted to distinguish between proportional and non-proportional mixed-frame responses. We also contrast the mixed-frame responses in the PC/BC network with earlier modelling work on proportional mixed-frame responses in BF networks (Deneve *et al.* 2001; Avillac *et al.* 2005), and with the occurrence of mixed-frame responses in backpropagation networks (Xing and Andersen 2000; Blohm et al. 2009).

Materials and Methods

We used a single-area version of the nonlinear PC/BC model (Spratling 2008). The model – shown in Figure 1 – receives external input from a population of input units. It contains three different types of nodes: error nodes, prediction nodes and disjunctive nodes. Error nodes and prediction nodes are reciprocally connected through feedforward and feedback connections and together constitute the core PC/BC part of the model. Disjunctive nodes pool the responses of small subsets of prediction nodes in a strictly feedforward manner, and do not alter the responses of error or prediction nodes. A detailed explanation of the operation of all types of nodes follows below.

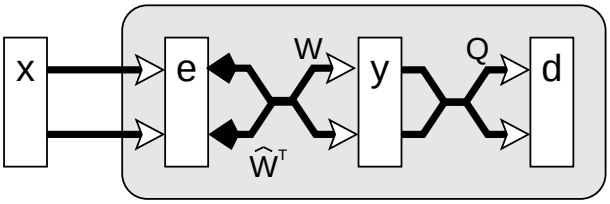

Figure 1. The single-area PC/BC model. Rectangles represent populations of input units (**x**), error nodes (**e**), prediction nodes (**y**) and disjunctive nodes (**d**). Open arrows signify excitatory connections, filled arrows indicate inhibitory connections. Crossed connections signify many-to-many connectivity; parallel connections indicate a one-to-one mapping.

We focus on the transformation of “visual” information from an eye-centred or retinal frame of reference to one that is invariant to eye movements, using eye position or direction of gaze. This is equivalent to measuring single-cell responses in awake animals with restrained head and body movements but unrestrained eye movements, which is the set-up used in many of the physiological studies mentioned here. We do not make a distinction between head-centred and world-centred frames of reference. In such experimental settings eye position or direction of gaze (in head-centred coordinates) coincides with the real position of the fixation points, and we therefore use these terms interchangeably. We refer to the retinal frame as 'retinotopic', to the eye-invariant frame as 'craniotopic', and to intermediate or mixed reference frames as 'mixed R/C'.

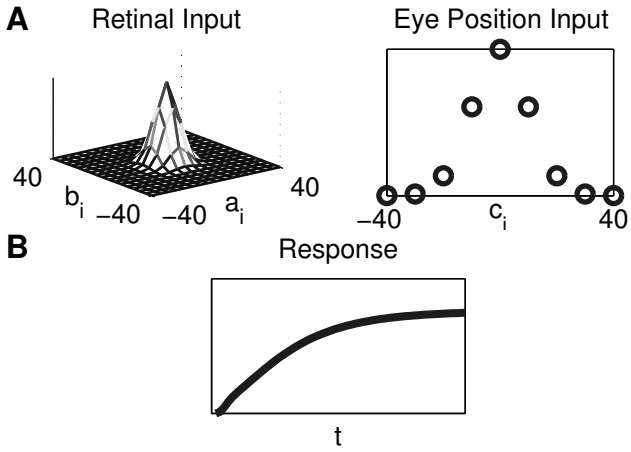

Figure 2. Network input stimuli and prediction node response properties. (A) Typical population input signals. A 2D Gaussian retinal input signal is shown for input values (r_x ,r_y) = (0,0)°, and a 1D Gaussian eye position signal (horizontal or vertical) for $e_{x|y}$= 0°. The strength of the input (and hence the scale of the vertical axes) depends on the value of h_{max} in Equations (4) and (5) which equalled 1 in all experiments reported here. The synaptic weight values of the prediction nodes were scaled copies of input stimuli and are therefore also represented by these graphs. (B) Typical temporal response of a node to its optimal stimulus. In this and all subsequent response graphs, the scale of the vertical axes is the same as for the input graphs.

Input

Similar to related models (Pouget *et al.* 2002) and to our previous work (De Meyer and Spratling 2011) input signals were generated by populations of topographically-organised input units with Gaussian response profiles. Their responses encoded the input variables: in particular, the retinal locations of visual stimuli and eye position or direction of gaze. Visual stimuli were encoded by units with 2D Gaussian response profiles. The response h_i of unit i was generated by:

$$h_i(r_x, r_y) = h_{max} \exp\left(-\frac{(r_x - a_i)^2 + (r_y - b_i)^2}{2\sigma_r^2}\right) \quad (1)$$

With (a_i ,b_i) the centre of the Gaussian response profile, (r_x ,r_y) the retinal location of the visual stimulus, σ_r the standard deviation and h_{max} the amplitude (maximum) of the Gaussian curve. Gaussian centres (a_i ,b_i) were spaced evenly in both dimensions from −40° to 40° in steps of 5°, meaning that visual input was encoded by 17 × 17 = 289 input units. σ_r was set to 6° and h_{max} was set to 1. A typical population signal for a given visual input is shown in Figure 2A (left).
Horizontal and vertical eye positions were encoded by separate populations of input units with 1D Gaussian response profiles:

$$h_i(e_{x|y}) = h_{max} \exp\left(-\frac{(e_{x|y} - c_i)^2}{2\sigma_e^2}\right) \quad (2)$$

Where c_i is the centre of the Gaussian response profile and $e_{x|y}$ the value of either horizontal (e_x) or vertical (e_y) eye position. For both horizontal and vertical eye position input units, c_i values were evenly spaced from −40° to 40° in steps of 10°. Horizontal and vertical eye position were thus each encoded by 9 input units, with σ_e = 10° and h_{max} = 1. A typical population signal encoding a given eye position can be seen in Figure 2A (right). The total number of input units (visual + horizontal and vertical eye position) is 289 + 9 + 9 = 307.
The choice of parameter values, here and in subsequent sections, is based on experience gathered in our earlier work (De Meyer and Spratling 2011). Within relatively broad limits, different values may lead to quantitative but not qualitative differences in the results discussed.

Model

The model is implemented by the following 3 equations:

$$\mathbf{e} = \mathbf{x} \oslash (\epsilon_2 + \hat{\mathbf{W}}^T \mathbf{y}) \quad (3)$$

$$\mathbf{y} \leftarrow (\epsilon_1 + \mathbf{y}) \otimes \mathbf{We} \quad (4)$$

$$\mathbf{d} = \max\{\hat{\mathbf{Q}} \otimes \check{\mathbf{Q}} \otimes \mathbf{Y}\} \quad (5)$$

x is an (m × 1) vector containing the input to the PC/BC area. **e** is an (m × 1) vector of error node activations. **y** is an (n × 1) vector of prediction node activations. **d** is a (q × 1) vector of disjunctive node activations. $\mathbf{W} = [\mathbf{w}_1,...,\mathbf{w}_n]^T$ is an (n × m) matrix representing synaptic weight values, each row of which, $\mathbf{w}_j{}^T = [w_{j1},...,w_{jm}]$, contains the weights of the synaptic connections arriving at prediction node j. $\hat{\mathbf{W}}$ is a scaled version of **W** with each row normalised such that its maximum value equals 1. **Q** is a (q × n) matrix representing the synaptic weight values from prediction nodes to disjunctive nodes. $\hat{\mathbf{Q}}$ and $\check{\mathbf{Q}}$ are scaled versions of **Q**, the first one scaled such that the maximum value in each row equals 1, and the second one scaled such that the maximum in each column equals 1. $\mathbf{Y} = [\mathbf{y},...,\mathbf{y}]^T$ is a (q × n) matrix, each row containing a copy of the prediction node activations. ⊘ and ⊗ indicate element-wise division and multiplication respectively. Function `max` returns the maximum in each row. ϵ_1 and ϵ_2 are parameters to prevent division-by-zero errors. They were set to values used previously, 0.001 and 0.05 respectively (De Meyer and Spratling 2011).

Equations (3), (4) and (5) are evaluated iteratively, with values of **y** calculated at time t used to obtain the node activations at time t + 1. After a number of iterations **e, y** and **d** generally approach steady-state values. For each new input **x** we evaluated the equations for 60 iterations, a value sufficiently large to reach steady state. Initially, **x** is set to the values generated by the input units and **y** values are set to 0. Initialising **y** to non-zero, randomised values has no effect on the steady-state values reached except in the case of “ambiguous” stimuli (Spratling and Johnson 2001). This situation did not occur in the experiments discussed here.

Equation (3) describes the calculation of the activity of the error detecting nodes. These values are a function of the input to the PC/BC network divisively modulated by a weighted sum of the output of the prediction nodes. Equation (4) describes the updating rule for the prediction node activations. The response of each prediction node is a function of its activation at the previous iteration and a weighted sum of afferent inputs from the error nodes. The activation of the error nodes can be interpreted in two ways. First, **e** can be considered to represent the residual error between the input

x and the reconstruction of the input ($\hat{\mathbf{W}}^T\mathbf{y}$) generated by the prediction nodes. The values of **e** indicate the degree of mismatch between the top-down reconstruction of the input and the actual input (assuming ϵ_2 is sufficiently small). When a value within **e** is greater than 1, it indicates that a particular element of the input is underrepresented in the reconstruction, a value of less than 1 indicates that a particular element of the input is overrepresented, and a value of 1 indicates that the top-down reconstruction perfectly predicts the bottom-up stimulation. A second interpretation is that **e** represents the inhibited inputs to a population of competing prediction nodes. Each prediction node modulates its own inputs, which helps stabilise the response, since a strongly (or weakly) active prediction node will suppress (magnify) its inputs and, hence, reduce (enhance) its own response. Prediction nodes that share inputs (i.e., that have overlapping RFs) also modulate each other's inputs. This generates a form of competition between the prediction nodes, such that each node effectively tries to block other prediction nodes from responding to the inputs that it represents. According to this interpretation, therefore, prediction nodes compete to represent input.

We demonstrated previously that, when two or more population-coded input signals converge on a PC/BC area, the competitive interactions between the prediction nodes may give rise to multiplicative gain modulation in their response properties (De Meyer and Spratling 2011). The weight values **W** that generated multiplicative responses were learned under a wide range of input and training conditions using an unsupervised learning rule. For a single prediction node, weights generally assumed the shape of scaled copies of the population-coded input signals for particular values of the input variables. We call these the “preferred” stimuli of node *j* and refer to them as r_{xj},r_{yj}, e_{xj} and e_{yj}. At the network level, stimulus preferences tended to be evenly distributed over the entire input space (the space defined by the combined ranges of all input variables). Given such weight distributions, the prediction node responses of the entire network *tiled* the input space in a characteristic manner (see Results for further discussion).

In this article we did not train weight values but determined **W** in accordance with the results from (De Meyer and Spratling 2011) in order to generate a different but predictable tiling of the input space for each simulation. For prediction node *j* the weight value from retinal input *i* was calculated using Equation (1): $w_{ji} = h_i(r_{xj}, r_{yj})$ where (r_{xj}, r_{yj}) represented the node's preferred visual input. When the retinal w_{ji} values of node *j* are plotted as a function of (a_i, b_i) – the Gaussian centres of the presynaptic input units h_i – they form a 2D Gaussian distribution. The weight value from horizontal [vertical] eye position input *i* was calculated using Equation (2): $w_{ji} = h_i(e_{xj})$ [$w_{ji} = h_i(e_{yj})$] for the node's preferred eye position e_{xj} [e_{yj}]. When plotted as a function of c_i the w_{ji} values form a 1D Gaussian distribution. The weights of each prediction node *j* **($\mathbf{w}_j$)** were subsequently normalised such that their total sum equalled 1. This was done to reflect the self-normalising character of the learning rule used in (De Meyer and Spratling 2011). Each prediction node in the network was initialised to a different combination of preferred stimuli. How these are distributed over the entire input space is summarised in Figure 3A and Figure 3B. The details are network-specific and further explained in the next section, Experimental setup.

The activation of disjunctive nodes **d** is calculated by performing a weighted max operation over the activation of prediction nodes (Equation (5)). The stimulation of the disjunctive nodes depends on the activation of the prediction nodes multiplied by the disjunctive weight values. The max operation means that there is only ever one prediction node activating the disjunctive node. Disjunctive nodes were previously used in a single-area PC/BC model to simulate the responses of complex cells in primary visual cortex (Spratling 2011). Spratling (2009) also used disjunctive nodes in a hierarchical model performing sensory-sensory coordinate transformations. In (Spratling 2009) values for weight matrix **Q** were learned using an unsupervised learning rule that extracted temporal correlations across a sequence of input stimuli. During the training procedure randomly-generated visual and postural (eye and head position) stimuli were presented to the network, whereby visual stimulation changed more slowly than postural stimulation. Across a sequence of such input stimuli, disjunctive nodes learned to associate the activity of prediction nodes that were activated in close temporal succession when a visual input – stationary in real-position terms –

moved around the retina as a consequence of eye or head movements. We set the values of **Q** in accordance with the training results of (Spratling 2009) by applying the same principle for each network simulated: a disjunctive node received connections (with corresponding **Q**-values set to 1) from all prediction nodes whose preferred visual stimuli coincided in a craniotopic frame of reference. In other words, a disjunctive node pooled from all prediction nodes with the same value of (a_{xj} ,a_{yj}), where (a_{xj} ,a_{yj}) was calculated as:

$$(a_{xj}, a_{yj}) = (r_{xj}, r_{yj}) + (e_{xj}, e_{yj}) \quad (6)$$

Experimental setup

The experiment simulated 3 different single-area PC/BC networks. All networks used the same input and simulation parameters, but had different numbers of prediction nodes and values for the synaptic weight matrices **W** and **Q** (see Model). The weight values of each prediction node *j* were initialised to scaled copies of the population input signals for one unique combination of input values – the node's preferred inputs r_{xj},r_{yj}, e_{xj} and e_{yj} (see Model). Figure 3A and Figure 3B detail the spatial organisation of the prediction nodes as a function of their preferred inputs – one node for each unique combination of r_{xj},r_{yj}, e_{xj} and e_{yj}. In all 3 networks visual preferences (r_{xj} ,r_{yj}) ranged from −40° to 40° in steps of 20° in both dimensions. Eye position preferences e_{xj} and e_{yj} differ across the 3 networks, as summarised in Figure 3B. Each network contained a single disjunctive node. The values of **Q** were determined using Equation (6) for (a_{xj} ,a_{yj})=(0 ,0)°, i.e., the disjunctive node pooled the responses of all the prediction nodes for which the preferred visual stimulus falls at the centre in a craniotopic frame of reference.

Measurements

Network response properties were determined by presenting different combinations of visual and eye position stimuli and averaging the temporal response of the nodes (see Figure 2B) over 60 iterations of Equations (3), (4) and (5).

For the prediction nodes, the visual RF was mapped by systematically varying visual input (r_x ,r_y) from −30° to 30° in steps of 5° in both dimensions while setting eye position input to the nodes' preferred (e_{xj} ,e_{yj}) values. Gain Fields (GF) measure the sensitivity of the preferred visual response to changes in eye position, and were mapped by varying (e_x ,e_y) from −30° to 30° in steps of 10° in both dimensions while setting the visual input to the nodes' (r_{xj} ,r_{yj}) values.

We measured the response properties of the disjunctive nodes in multiple ways to enable a direct comparison with neurophysiological results. The first method consisted of applying two stimulus sets – each set consisting of 5 combinations of visual input and eye position – that would generate different responses for retinotopically- or craniotopically-organised cells. This format of testing and displaying response properties is equivalent to the one used in (Galletti *et al.* 1993) and allows to quickly identify which responses are retinotopic or craniotopic. It is further explained in Figure 4A.

The second method consisted of measuring one-dimensional RFs for different directions of gaze. These different gaze-dependent curves were then plotted in retinotopic and craniotopic coordinates to assess their possible shift in either of the two reference frames. This method is commonly used in physiological studies (e.g., Stricanne *et al.* 1996; Avillac *et al.* 2005; Mullette-Gillman *et al.* 2005). We mapped the horizontal RF of each disjunctive node by systematically varying the horizontal craniotopic position of the visual stimulus (a_x) from −50° to 50° in steps of 5° while keeping vertical stimulus position a_y = 0°, and repeating this for three different fixation points: (−20 ,0)°, (0 ,0)° and (20 ,0)°. To generate the retinal input to the network and to plot the results in a retinotopic frame of reference we calculated the reverse transformation from craniotopic to retinotopic coordinates:

$$(r_x, r_y) = (a_x, a_y) - (e_x, e_y) \quad (7)$$

The third method consisted of mapping the full 2D visual RF for different eye positions, a method previously used in (Duhamel, Bremmer et al. 1997). RFs were plotted in separate contour graphs for each eye position. Here we obtained such graphs by systematically varying the craniotopic position of the visual stimulus (a_x ,a_y) from −30° to 30° in steps of 5° in both dimensions, and repeating this RF mapping for 3 × 3 different fixation points, ranging from −20° to 20° in steps of 20° in both dimensions of eye position. The retinal location of the visual stimulus was again calculated from its craniotopic position by applying Equation (7).

Analysis

In order to quantify RF shifts, we repeated 2 methods of analysis from the neurophysiological literature that are based on calculating the correlation between RF curves measured for different directions of gaze. The first method calculates the average of Pearson's correlation ρ for different RF curves aligned in both retinotopic and craniotopic frames of reference, a measure previously used by Mullette-Gillman et al. (2005):

$$C_{r|a} = \frac{1}{2}(\rho(R_l, R_c) + \rho(R_r, R_c)) \quad (8)$$

R_l, R_c and R_r are 1D response vectors of the nodes aligned in either retinotopic (r) or craniotopic (a) coordinates. The response vectors R were obtained by the 1D RF measurement procedure outlined above, for left (−20°), central (0°), and right (20°) horizontal eye position. The values of C_r and C_a range from −1 to 1, with a value of 1 indicating perfect alignment in that particular frame of reference.

The second quantitative method of analysis consists of calculating an average shift index (SI) by estimating the RF shift between each pair of RF graphs (ΔRF_{ij}) and normalising this by the corresponding difference in direction of gaze (ΔE_{ij}), as used by Duhamel et al. (1997). We applied this analysis to the 2D, craniotopic RF graphs obtained for 9 different eye positions as described above. Each ΔRF_{ij} was estimated by systematically shifting the two graphs, column-wise and row-wise, calculating the correlation between them, and taking as ΔRF_{ij} the value for which the cross-correlation reached its maximum. The average shift index was calculated separately for horizontal (SI_h) and vertical (SI_v) shifts, and pair-wise correlations for which $\Delta E_{ij} = 0°$ where discarded before calculating the mean. For cells which are organised craniotopically, the value of the average shift index would equal 1; for retinotopic cells it would equal 0.

Code

Software, written in MATLAB, which implements the experiments described in this paper is available at `http://www.corinet.org/mike/code.html`.

Results

Reference frame transformations in the PC/BC model

We demonstrate that disjunctive nodes in the PC/BC model can compute a transformation between reference frames by pooling responses from the prediction nodes in a strictly feedforward manner. We also replicate mixed-frame responses that have been observed in several areas of parietal cortex, namely, different types of non-proportional mixed R/C responses. The experiment simulates 3 networks with different visual and eye-position preferences for the prediction nodes, as explained in Figure 3A and Figure 3B (see also Experimental setup). The resulting differences in response properties of the prediction and disjunctive nodes are discussed below.

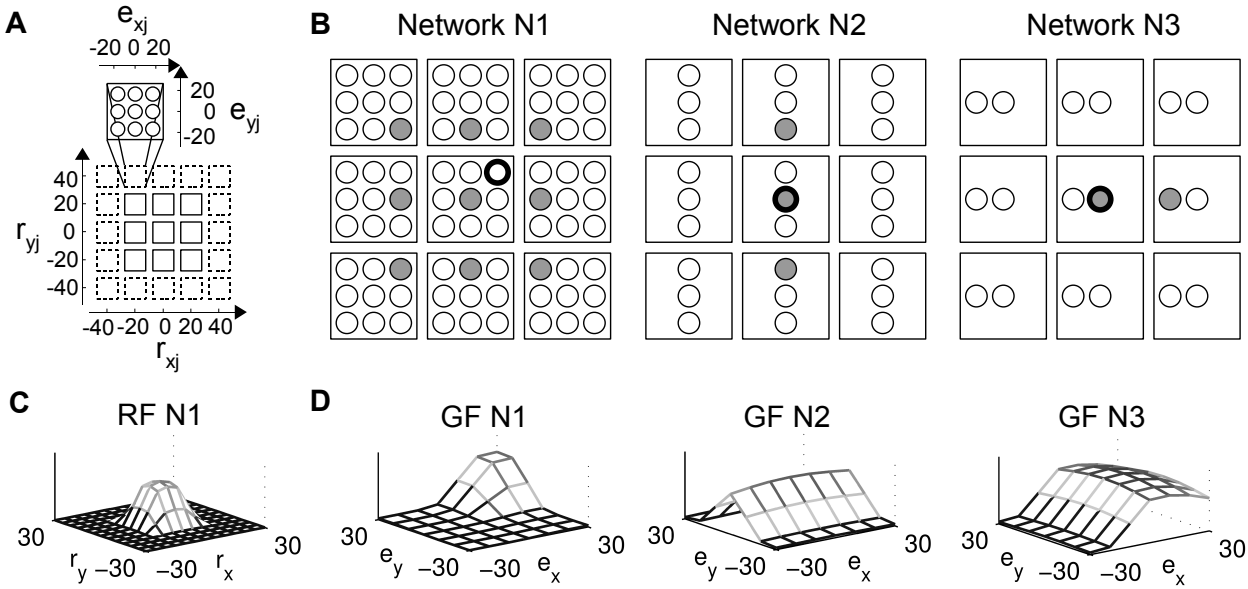

Figure 3. Prediction nodes: spatial organisation of preferred stimuli and response properties in 3 different networks. (A) All networks were simulated with a 5 × 5 array of visual preference combinations (r_{xj} ,r_{yj}). Squares group the prediction nodes with the same visual preferences. Within each square, prediction nodes (circles) are arranged according to eye position preferences (e_{xj} ,e_{yj}). (B) For each of the networks N1, N2 and N3 the central 3 × 3 (r_{xj} ,r_{yj}) section shows the layout of the eye position preferences. For each visual preference (r_{xj} ,r_{yj}), N1 encodes 3 × 3 different eye position preferences (e_{xj} ,e_{yj}); N2 encodes 3 vertical e_{yj} values and a single horizontal e_{xj} value; N3 encodes 2 horizontal e_{xj} values and only a single vertical e_{yj} value. The grey, filled circles indicate the prediction nodes that project to the single disjunctive node in each network. The bold circles indicate the prediction nodes whose RF and/or GFs are shown in (C) and (D). (C) RF of the bold prediction node in N1. This peak-shaped RF is representative of the RFs of the prediction nodes in all 3 networks. (D) GFs of the bold prediction nodes in, respectively, N1, N2 and N3. These are representative of the GFs in each network (see main text for further discussion).

Response properties of the prediction nodes

In each of the 3 networks the responses of the prediction nodes tile the input space – the 4D space defined by the input variables (r_x ,r_y ,e_x ,e_y) – in a characteristic way. In general terms, the tiling is generated because prediction nodes in a PC/BC network compete with one another in order to represent input (Spratling 2008). The precise shape of the response profile of a prediction node depends therefore partly on its own stimulus preferences (as determined by its weight values) and partly on suppression generated by other nodes in the network (De Meyer and Spratling 2011). Examples of prediction node responses in the 3 networks are shown in Figure 3C and Figure 3D. Figure 3C shows a typical visual RF. It was mapped by systematically varying visual input while fixing eye position to the node's preferred eye position (see Measurements for details). This RF shape is typical of all RFs across the 3 different networks: a bell-shaped curve peaking at the retinal location determined by the node's preferred visual input (r_{xj} ,r_{yj}).

Figure 3D displays typical GF shapes for each of the three simulated networks. The left-hand graph shows the GF of the same node whose RF is shown in Figure 3C. It shows the sensitivity of the node's preferred visual response to changes in eye position, and was obtained by systematically varying eye position while keeping visual input fixed to the node's preferred visual input (see Measurements). This GF is typical of network N1: the response to the preferred visual stimulus peaks at or near (e_{xj} ,e_{yj}), but is strongly suppressed for more distant eye position values. This suppression is caused by competition from the prediction nodes with the same visual preferences (r_{xj} ,r_{yj}) but different (e_{xj} ,e_{yj}) values (De Meyer and Spratling 2011).

The middle graph in Figure 3D shows a GF representative of network N2. Only a single e_{xj} value is

represented in N2 (see the middle graph in Figure 3B), meaning that there is no competition between prediction nodes along the e_x dimension. The resulting GF shows that the visual response is largely unaffected by shifts in horizontal eye position, but strongly affected by vertical eye position – peaking along the horizontal midline. A similar result would have been obtained by making all horizontal eye position weights equal to 0.

The right-hand graph in Figure 3D is representative of GFs in network N3. The node's response to its preferred visual stimulus is only weakly modulated by vertical and by horizontal, right-of-centre gaze shifts, but is strongly suppressed for horizontal, left-of-centre gaze shifts. This node (with e_{xj} = 0°) does not experience competition from other prediction nodes for positive values of e_x because N3 does not contain prediction nodes with positive e_{xj} preferences (see the right-hand graph in Figure 3B). It does, however, experience strong competition for negative e_x values from the node with the same (r_{xj} ,r_{yj}) values but with an e_{xj} value of −20°.

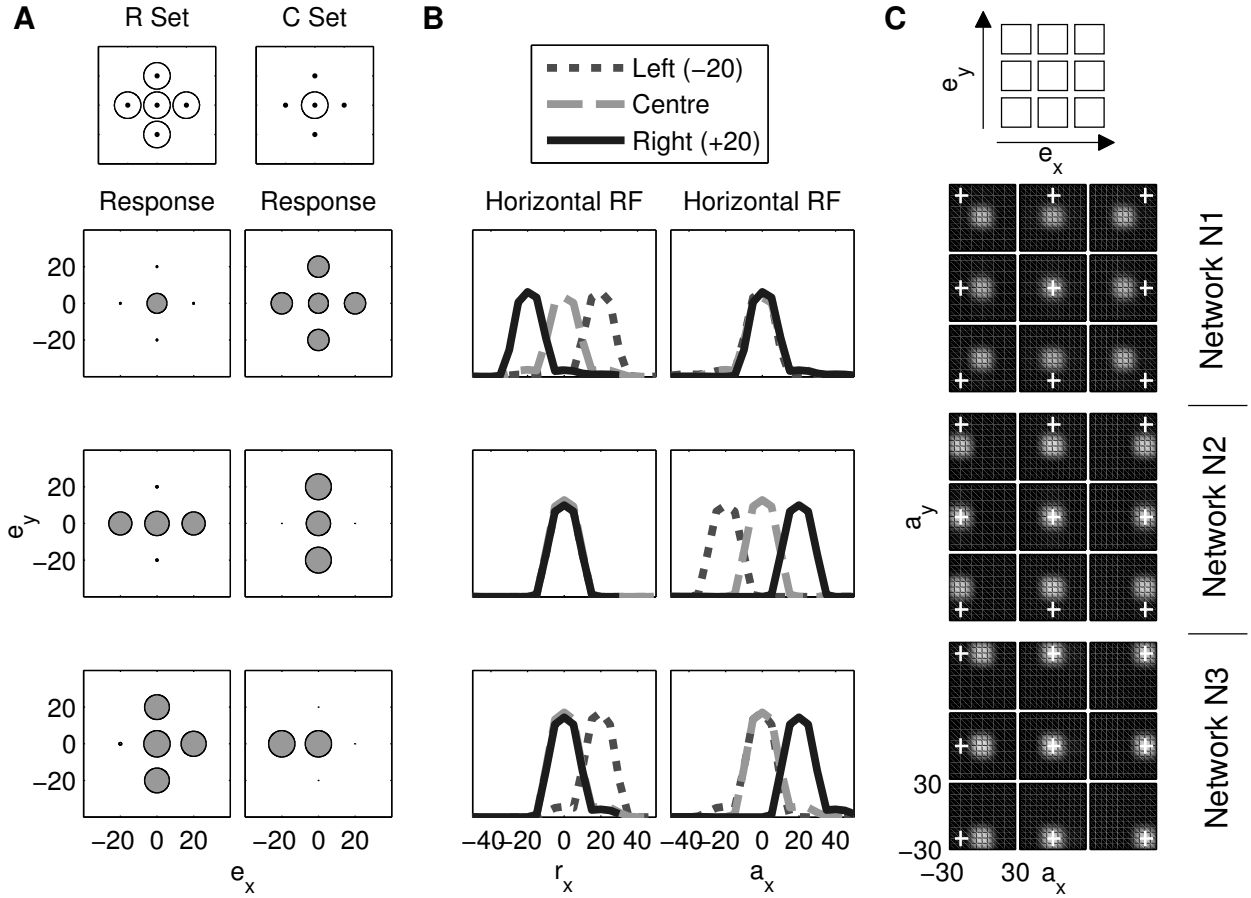

Figure 4. Disjunctive nodes: response properties of the single disjunctive node in the 3 different networks from Figure 3. For each panel (A), (B) and (C) the top row contains a legend, and the bottom 3 rows show responses properties of the disjunctive nodes in networks N1, N2 and N3 respectively. (A) Response to a retinotopic (first column) and a craniotopic (second column) stimulus set. In the R Set, the visual stimulus (circle) moves with each fixation point (dot). In the C Set, the visual stimulus remains in the same central spatial location for all fixation points. In the Response graphs, the radius of the dark, filled circles indicate the strength of the response of the disjunctive node, plotted at the location of the corresponding fixation point. (B) Horizontal 1D RFs plotted in a retinotopic (r_x) and craniotopic (a_x) coordinate frame. RFs were measured for 3 different eye positions. (C) Intensity plots of the full 2D RFs for a set of 3 × 3 different fixation points – plotted in craniotopic coordinates. The white crosses indicate the spatial location of the fixation point for each graph. For those graphs where the cross is difficult to see, they coincide with the high-intensity part of the visual response.

Response properties of the disjunctive nodes

We measured the response properties of each disjunctive node in 3 different ways to enable a direct comparison with neurophysiological results (see Measurements). The first method consisted of 2 sets of 5 stimulus configurations (see Figure 4A – top row): a retinotopic set (R Set) in which the visual stimulus moved together with eye position; and a craniotopic set (C Set) in which the visual stimulus remained in the same spatial location for the 5 different fixation points. A fully retinotopic cell with central RF would respond to all stimuli of the R Set, whereas a fully craniotopic cell would respond to all stimuli of the C Set. The central stimulus configuration was the same in both sets, hence elicits the same response. The responses of the disjunctive nodes in the 3 different networks are shown in Figure4A. In each network the disjunctive node pools from gaze-modulated prediction nodes for which $(a_{xj}, a_{yj})=(0, 0)°$ (see Equation (6)), the grey nodes in the different graphs of Figure 3B. In network N1 this means that the disjunctive node pools from 9 strongly-modulated prediction nodes. Its response is fully craniotopic. In N2, the 3 prediction nodes satisfying the pooling condition are modulated by vertical eye position only. The disjunctive node has craniotopic response properties for vertical gaze shifts, but has retinotopic response properties for horizontal gaze shifts. The disjunctive node in N3, pooling responses from 2 prediction nodes, is craniotopic for the horizontal, left-of-centre fixation point, and retinotopic for all other stimulus configurations.

The second test measured response properties for 3 different eye positions along the horizontal midline. The resulting RFs were plotted in both retinotopic and craniotopic reference frames. Retinotopic RFs would line up in a retinotopic reference frame and shift with eye position in a craniotopic frame. Craniotopic RFs would line up in a craniotopic reference frame and shift in opposite direction of eye position in a retinotopic frame. Figure4B shows the results for the 3 disjunctive nodes. When analysed along the horizontal midline, the RFs of the first node line up in the craniotopic graph, indicating a craniotopic response. The RFs of the second node line up in the retinotopic graph, indicating a retinotopic response. Had this node been analysed along the vertical midline, the response would have been the same as for the first node and it would have been classified as craniotopic. This node thus displays a mixed R/C response across the two gaze dimensions. For the third node, 2 RFs line up in the retinotopic graph, and 2 RFs line up in the craniotopic graph. This node displays a mixed R/C response along a single dimension of gaze.

The third method measured the full 2D RF for 9 different directions of gaze. Each 2D RF was plotted in a separate intensity graph in craniotopic coordinates (see Figure 4C). Pair wise comparisons of these graphs reveal when nodes are retinotopic or craniotopic. The disjunctive node of N1 is fully craniotopic as all RFs remain in the same location in the craniotopic graphs, regardless of eye position. The node of N2 is invariant for vertical eye positions (i.e., craniotopic) but moves with the eye for horizontal eye position shifts (i.e., retinotopic). The node of N3, finally, is retinotopic for vertical and right-of-centre horizontal gaze shifts, but craniotopic for all left-of-centre horizontal shifts.

Comparison with neurophysiological results

The simulation results described above are qualitatively similar to sensory responses observed in several PPC areas. Galletti et al. (1993) reported the existence of “real-position” cells in parietal visuomotor area V6A, i.e., cells whose visual RF remained in the same spatial location regardless of eye position. Such cells were later also reported in the ventral intraparietal area (VIP) (Duhamel *et al.* 1997) and medial and lateral intraparietal areas (MIP and LIP) (Mullette-Gillman *et al.* 2005). Galletti et al. (1995) proposed a schematic model to explain how real-position behaviour may be constructed by locally pooling the responses of another type of cell found in area V6A: strongly-modulated, gaze-dependent visual responses (Galletti *et al.* 1993; Breveglieri et al. 2009). These retinotopic visual cells are visually responsive for a limited range of eye positions but not for others. Network N1 implements this schematic model, with the prediction nodes with their peak-shaped GFs in the role of strongly-modulated, gaze-dependent visual cells and the disjunctive node

constituting a real-position cell. These results demonstrate that in the PC/BC model real-position behaviour can indeed be constructed from the responses of strongly gaze-modulated, retinotopically-organised cells.

A curious idiosyncrasy in the responses of the V6A real-position cells was that half of the reported cells were craniotopic in only one of the two dimensions, and retinotopic in the orthogonal dimension. Cells with the same type of mixed responses were also observed inVIP (Duhamel *et al.* 1997). The most parsimonious PC/BC model that could replicate this behaviour was network N2 . Mixed-frame responses arose naturally when prediction node responses were modulated by vertical but not by horizontal eye position. The model thus generates a testable prediction: mixed R/C responses across the two dimensions arise naturally when pooling from response fields that are gaze-modulated in one dimension only. We return to this prediction in the discussion.

Network N3 displayed another type of response that has also been observed in parietal cortex. Stricanne et al. (1996) reported a cell in area LIP whose auditory RF showed the same mixed R/C RF alignment behaviour along the horizontal dimension as the response in the bottom row of Figure 4B. Such mixed, irregular behaviour has also been reported for visual cells in areas MIP and LIP (Mullette-Gillman *et al.* 2005) and has been observed in cortical area V6A (P. Fattori, personal communication) and VIP (J.R. Duhamel, personal communication). The crucial model assumption underlying this behaviour was that the disjunctive node pooled from prediction nodes that showed little or no gaze modulation for certain gaze shifts: a coarser tiling of the input space along the e_x dimension gave rise to the mixed R/C responses along that dimension. This again generates a testable prediction which we will address in the discussion.

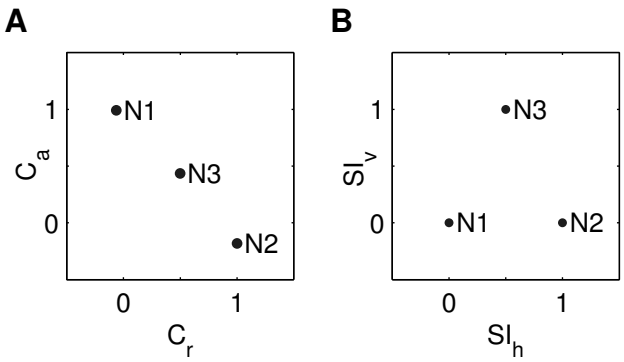

Figure 5. Quantitative analysis of RF shifts for the 3 disjunctive nodes in networks N1, N2 and N3. (A) Retinotopic (C_r) versus craniotopic (C_a) average correlation calculated from the 1D RF curves (see Figure 4B) aligned in a retinotopic, respectively craniotopic, reference frame (see Equation (8)). $C_r \gg C_a$ indicates predominantly a retinotopic response, $C_r \ll C_a$ indicates a response that is mostly craniotopic, and $C_r \approx C_a$ means that the response is mixed R/C. (B) Horizontal (SI_h) versus vertical (SI_v) average shift index calculated from the 2D RF curves (see Figure4C) aligned in a craniotopic reference frame (see Analysis). Values of 0 indicate a craniotopic response, whereas values of 1 indicate a retinotopic response.

Characterising Receptive Field Shifts

In addition to the qualitative assessment of reference frames that can be performed by visual inspection of RF shifts, past physiological studies have also proposed measures to quantify the reference frame analysis. Here we repeated two methods of analysis (see Analysis) to assess how well they capture the potential differences in mixed R/C responses.

For the 1D RF data shown in Figure 4B, we calculated the average correlation between the different RF curves aligned in both a retinotopic and craniotopic frame of reference (see Equation (8)). The results are shown in Figure 5A. The craniotopic organisation of the disjunctive node in N1, and the horizontal retinotopic organisation of the node in N2 are both apparent from their (C_r,C_a) values. The node in N3 has (C_r,C_a) values that are indicative of mixed or irregular behaviour. Crucially,

those values are similar to the ones that would have been obtained for proportionally shifting RFs that shift roughly half with each eye position shift.

The 2D RF data from Figure 4C were analysed by calculating a horizontal (SI_h) and vertical (SI_v) average shift index, and the results are summarised in Figure 5B. The craniotopic behaviour of N1 and the mixed R/C behaviour across the horizontal and vertical eye dimensions of N2 are apparent in this measure. For the node in N3, the mixed R/C behaviour for horizontal eye shifts is captured by its $SI_h = 0.5$ value, but this value would also have been obtained for proportionally-shifting RFs.

Discussion

Interpretation of the results

In the previous section we demonstrated that certain mixed-frame responses can be generated by pooling together gain-modulated, single-frame responses. We reported two types of such mixed-frame responses: those that are encoded in different frames for horizontal and vertical gaze shifts, and those that are encoded differently within a single dimension of gaze. In our model the hybrid nature of these responses is not caused by irregularities in the pooling process, because in all our networks the same principle was used to determine the connections from gain-modulated prediction nodes to response-pooling disjunctive nodes. Rather, the hybridity is the consequence of how the population of gain-modulated responses tiles the input space – the multidimensional space spanning the sensory and eye position input domains. In De Meyer and Spratling (2011) we investigated in detail how competition between prediction nodes causes their responses to tile the input space. We also established that the coarse tiling generated by the prediction nodes form approximative basis-function (BF) sets. We can thus rephrase our earlier observation in the language of BF networks: it is the precise structure of the BF set generated by the prediction nodes that determines the reference frame transformation generated by the pooling process. Phrased like this the principle behind our results appears to be self-evident, but there is more: Galletti et al. (1995) proposed that real-position (i.e., craniotopic) responses might be built up in local networks within parietal area V6A. Combining this idea of locality with our network layouts makes it possible to generate testable predictions about cortical organisation.

Prediction 1: mixed R/C responses for vertical vs. horizontal gaze shifts

In visuomotor area V6A 8 out of the 16 real-position cells reported in (Galletti *et al.* 1993) were craniotopic for only one gaze dimension (4 for horizontal gaze shifts, 4 for vertical gaze shifts). In the model such mixed R/C responses (the second row of graphs in Figure 4) arose when prediction nodes were modulated by one eye-position signal only (see middle graph in Figure 3D). The model thus leads to the prediction that in area V6A partial real-position cells are predominantly surrounded by cells that are only modulated by gaze shifts in the same dimension. Some preliminary evidence to support this idea comes from an analysis of the original microelectrode penetration trajectories of (Galletti *et al.* 1993, 1995). It revealed that, out of 5 penetrations that encountered partial real-position units and could be reconstructed, 3 fit the hypothesis: a partial real-position unit was surrounded by gaze-dependent cells modulated only by the same dimension of eye displacement (P. Fattori, personal communication). Although far from conclusive, this 3 out of 5 result is intriguing because a penetration trajectory could easily have sampled outside the local area or group of cells projecting to the real-position cell.

If partial real-position cells are indeed generated by a systematic mechanism, then this raises an important question: what is their use? A first possibility is that the population of vertical and horizontal partial craniotopic responses forms an implicit, vectorial representation of craniotopic coordinate space. Alternatively, partial transformations could be an intermediate step in a hierarchy that calculates an explicit representation of craniotopic coordinate space with fewer resources than a one-step transformation. The underlying reason for this is that basis functions suffer from the curse

of dimensionality, i.e., the number of basis functions needed to tile and input space increases combinatorially with number of dimensions of the input space. In this context partial real-position cells can thus be seen as a sign of efficient computation. This principle of efficiency has been demonstrated previously for a different combination of transformations using a hierarchical version of the PC/BC model (Spratling 2009).

Cortical area V6A is not the only area where differences in horizontal and vertical reference frames have been demonstrated. The same effect has been observed in VIP (Duhamel *et al.* 1997), a multimodal integration area receiving visual, auditory and somatosensory information encoded in different frames of reference (Stricanne *et al.* 1996; Mullette-Gillman *et al.* 2005). It is likely that the interaction of those signals adds to the complexity of the mixed-frame responses observed there. We will return to this point in the discussion of quantitative measures of reference frames.

Prediction 2: mixed R/C responses along a single dimension of gaze

In the model, an example of a mixed R/C response along a single dimension of gaze (see the bottom row of graphs in Figure 4C) was generated from prediction node responses modulated by left-of-centre eye position, but not by right-of-centre eye position (see right-hand graph in Figure 3D). What determines the mixed-frame response in this case is that the disjunctive node pools from prediction nodes that are gain-modulated for fixation points falling inside some section of the visual field, but unmodulated for fixation points in other sections of the visual field. In the unmodulated section of the visual field the prediction node responses do not form a BF set, and hence a reference-frame transformation cannot be computed from their responses. This leads to another prediction about local cortical organisation, i.e., cells that appear to encode information in one frame of reference for fixation points falling inside one part of visual field, but encode information in another frame of reference for fixation points outside that area, pool from a population of cells that are systematically gain-modulated for some parts of the visual field, but not for others. In the PC/BC model such a representation can be learnt if the local eye position signal is biased towards representing, e.g. left rather than right, or higher rather than lower gaze directions.

Several areas of parietal cortex have been shown to contain this type of mixed-frame responses. Individual examples have been reported in areas LIP and MIP (Avillac *et al.* 2005; Schlack *et al.* 2005). Unpublished results have also placed them in areas V6A (P. Fattori, personal communication) and VIP (J.R. Duhamel, personal communication). The lack of published results makes it hard to determine at present whether these responses are generated by such systematic biases hypothesised above. An alternative explanation is that they are generated by non-systematic irregularities in the tiling, or non-systematic irregularities in the pooling, or simply by neuronal or measurement variability. However, a change in the reporting of quantitative reference-frame measures would allow a distinction to be made between the systematic or non-systematic nature of these mixed-frame responses.

Quantitative measures of reference-frame transformations

The two methods of analysis, taken from the physiology literature and applied to the simulated data (see Figure 5), both averaged an estimate of RF displacements over different eye displacements to provide a quantitative assessment of the encoding reference frame. This is a sensible approach for proportional mixed-frame responses, such as the partially-shifting RFs that follow from multisensory integration in a basis-function network (Deneve *et al.* 2001; Pouget *et al.* 2002). However, such measures do not allow making a distinction between proportional and non-proportional mixed-frames, and between systematic or non-systematic origins of the mixed-frame responses.

The first type of non-proportional mixed-frame responses, i.e., across two gaze-shift dimensions, can be measured by independently quantifying the reference frames of horizontal and vertical gaze shifts. Indeed, the results and the analysis of (Galletti *et al.* 1993; Duhamel *et al.* 1997) indicate

how common these mixed-frame responses may actually be.

Within a single dimension of eye displacement, a distinction between proportional and non-proportional mixed-frame responses can be made by assessing the distribution of the shift indexes calculated for the individual gaze shifts. This analysis becomes more feasible the more fixation points have been analysed. For instance, for the 2D RF mapping procedure and analysis method of (Duhamel *et al.* 1997) (see also Figure 4C), a total of 25 individual shift index values are available for analysis in both horizontal and vertical dimensions. If the distribution of these values appears unimodal around its mean value, then the mixed-frame response can be classified as proportional. If the distribution is indistinguishable from the uniform distribution, then the mixed-frame responses are likely to be generated by non-systematic irregularities in the tiling or the pooling, or by noise. Finally, if the distribution appears to be bimodal, then it may have been generated by systematic biases in the tiling of the input space as discussed above. The latter case could be analysed further by looking at the spatial distribution of the shift index values: spatial structure would be indicative of systematic biases in the tiling rather than non-systematic irregularities. Such analyses have, to date, not been performed but could in the future be used to make a distinction between proportional and non-proportional mixed-frame responses, and between the different mechanisms that are thought to underlie them: multimodal integration, systematic effects in the tiling and pooling mechanisms, or non-systematic irregularities and noise. They may give different results for different cortical areas: a predominantly visuomotor area such as V6A might have less proportional mixed-frame responses than multimodal area VIP.

The pooling process: complex-cell-like?

The pooling function (a weighted max operation) has previously been used as a model for complex cells in a PC/BC model of primary visual cortex (Spratling 2011). This model of complex cells is similar to that employed in “Hierarchical Model and X” (HMAX) (Riesenhuber and Poggio 1999) which, in turn, is an idealised version of the hierarchical model proposed by Hubel and Wiesel (1962). Physiological support for this model is found in (Gawne and Martin 2002; Lampl et al. 2004; Finn and Ferster 2007; Kouh and Poggio 2008). There is currently no such physiological evidence for parietal cells and the results we reported in this article do not depend on the max operation itself: they could also have been obtained by a weighted summation of the prediction node responses. However, we opted for the max operation to maintain consistency with previous work (Spratling 2009, 2011).

Implicit in the model of complex cells is that their responses are constructed by *locally* pooling the responses from simple cells. It is this assumption of locality that provides another element of analogy: the disjunctive nodes perform partial reference-frame transformations by locally pooling from gaze-modulated response fields that constitute sparse basis-function sets. Repeated multiple times over the cortical surface, such partial, incomplete and noisy reference-frame transformations could calculate a population-level coordinate transformation that is robust against neuronal noise (Deneve *et al.* 2001), systematic biases and non-systematic irregularities in the tiling and pooling. At the same time, it may reduce the computational requirements that follow from the combinatorial explosion in the size of basis-function sets with increasing dimensionality by performing reference-frame transformations gradually or in multiple steps.

Comparison with Related Models

Basis function networks

In BF networks with multimodal attractors, mixed-frame responses arise naturally within the hidden layer of the networks (Deneve *et al.* 2001). Signals encoded in a 1D visual, a 1D auditory frame of reference, and eye position information converge in the hidden layer and are integrated at the single-cell level. The resulting tiling of the input space (the BF set) is intermediate between the two

intrinsic frames of references. The node responses in the hidden-layer cells are strongly-modulated by eye position, and the partial RF shifts are proportional to gaze shift. Different shift ratios are obtained by varying the strength of the different sensory inputs (Avillac *et al.* 2005), but the RF shifts remain proportional to gaze shifts.

The main difference with our current work is to be found in the purported mechanism that generates the mixed-frame responses: in our model they arise from pooling gain-modulated responses from the prediction nodes (the equivalent of the basis-function layer), not from multimodal integration within the basis-function layer. This mechanism is complementary rather than antithetical to the multimodal integration mechanism proposed by Deneve et al. (2001). There are, however, additional differences in what type of tiling can be generated with the two different models. It is unlikely that the type of systematic biases in the GFs of N3 (see right-hand graph in Figure 3D) can be easily generated with the basis-function model of Deneve et al. (2001).

Backpropagation networks

Mixed-frame responses have also been shown to arise in backpropagation networks that learn to perform sensorimotor transformations through supervised learning (Xing and Andersen 2000; Blohm et al. 2009). For a network performing visually-guided reaching in 3D space, Blohm et al. (2009) reported RFs in the hidden layers that shifted differently in the horizontal and vertical directions, and shifted differently for horizontal and vertical eye displacements. In other words, there were cells whose RF would shift vertically (horizontally) for horizontal (vertical) eye displacements – albeit by a small amount. A detailed analysis performed for the vertical and horizontal RF shifts of one hidden-layer unit showed that these were almost proportional.

Future Work

Unsupervised learning of network weights

For the current simulations we predetermined the weights of the prediction nodes in order to generate different but predictable tiling of the input space, and the weights of the disjunctive nodes to pool the responses of the prediction nodes in a systematic manner. In (De Meyer and Spratling 2011) we demonstrated that prediction-node weights that give rise to gain-modulated RFs can be learned using an unsupervised learning rule. Spratling (2009) demonstrated that weights of the disjunctive nodes performing reference frame transformations in a two-stage hierarchical PC/BC model could also be learned through an unsupervised learning rule. Although the input to this hierarchical network was simpler than for the networks reported here (there was no spatial extent in the input signals), the pooling principles generated by the learning rule in the structure of the disjunctive weights were similar to the principle used here to predetermine those weights. Extending these learning experiments to the current problem setting would allow comparing, through simulation experiments, the potential effects of systematic biases (as reported here) and small, non-systematic irregularities that may arise in the learned network weights.

Multisensory integration in the PC/BC model

Preliminary results indicate that, when sensory signals encoded in different intrinsic frames of reference converge in the prediction node layer of the PC/BC model, the competition between the prediction nodes may give rise to proportional mixed-frame RFs within that layer, as was shown for other basis-function networks (Deneve *et al.* 2001; Avillac *et al.* 2005). Combining multisensory integration with the results from this article would allow, through simulation, to compare the properties of proportional and non-proportional mixed-frame responses.

Conclusion

We presented simulation results to demonstrate that certain types of mixed-frame responses may be generated by pooling of gain-modulated responses. We argued that these non-proportional mixed-frame responses may be different from the proportional mixed-frame responses that arise through multisensory integration in basis function networks. Non-proportional mixed-frame responses may be the result, and hence revealing, of systematic factors in the functional organisation of certain parietal areas such as visuomotor area V6A. Finally, we suggested how existing measures of reference frame encodings may be adapted to further distinguish between these different putative mechanisms that give rise to mixed-frame responses.

Funding

This work was supported by the Engineering and Physical Sciences Research Council (grant number EP/D062225/1).

Notes

We thank Prof. P. Fattori and Prof. J.R. Duhamel for sharing information about unpublished mixed-frame responses. Address correspondence to Dr. Kris De Meyer, Department of Informatics, King's College London, London WC2R 2LS, United Kingdom. Email: kris@corinet.org.

References

Andersen RA, Bracewell RM, Barash S, Gnadt JW, Fogassi L. 1990. Eye position effects on visual, memory, and saccade-related activity in areas LIP and 7a of macaque. Journal of Neuroscience. 10:1176-1196.

Andersen RA, Cui H. 2009. Intention, action planning, and decision making in parietal-frontal circuits. Neuron. 63:568-583.

Andersen RA, Mountcastle VB. 1983. The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. Journal of Neuroscience. 3:532-548.

Avillac M, Deneve S, Olivier E, Pouget A, Duhamel JR. 2005. Reference frames for representing visual and tactile locations in parietal cortex. Nat Neurosci. 8:941-949.

Blohm G, Keith GP, Crawford JD. 2009. Decoding the cortical transformations for visually guided reaching in 3D space. Cereb Cortex. 19:1372-1393.

Bremmer F, Ilg UJ, Thiele A, Distler C, Hoffmann KP. 1997. Eye position effects in monkey cortex. I. Visual and pursuit-related activity in extrastriate areas MT and MST. Journal of Neurophysiology. 77:944-961.

Breveglieri R, Bosco A, Canessa A, Fattori P, Sabatini SP. 2009. Evidence for Peak-Shaped Gaze Fields in Area V6A: Implications for Sensorimotor Transformations in Reaching Tasks. In: Mira JFJMAJRDFTFJ, ed. Bioinspired Applications in Artificial and Natural Computation, Pt Ii p 324-333.

Chang SW, Snyder LH. 2010. Idiosyncratic and systematic aspects of spatial representations in the macaque parietal cortex. Proc Natl Acad Sci U S A. 107:7951-7956.

De Meyer K, Spratling MW. 2011. Multiplicative gain modulation arises through unsupervised learning in a predictive coding model of cortical function. Neural Computation. 23:1536-1567.

Deneve S, Latham PE, Pouget A. 2001. Efficient computation and cue integration with noisy population codes. Nature Neuroscience. 4:826-831.

Duhamel JR, Bremmer F, BenHamed S, Graf W. 1997. Spatial invariance of visual receptive fields

in parietal cortex neurons. Nature. 389:845-848.

Finn IM, Ferster D. 2007. Computational diversity in complex cells of cat primary visual cortex. J Neurosci. 27:9638-9648.

Galletti C, Battaglini PP, Fattori P. 1993. Parietal neurons encoding spatial locations in craniotopic coordinates. Exp Brain Res. 96:221-229.

Galletti C, Battaglini PP, Fattori P. 1995. Eye position influence on the parieto-occipital area PO (V6) of the macaque monkey. Eur J Neurosci. 7:2486-2501.

Gawne TJ, Martin JM. 2002. Responses of primate visual cortical V4 neurons to simultaneously presented stimuli. J Neurophysiol. 88:1128-1135.

Hubel DH, Wiesel TN. 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol. 160:106-154.

Kouh M, Poggio T. 2008. A canonical neural circuit for cortical nonlinear operations. Neural Comput. 20:1427-1451.

Lampl I, Ferster D, Poggio T, Riesenhuber M. 2004. Intracellular measurements of spatial integration and the MAX operation in complex cells of the cat primary visual cortex. J Neurophysiol. 92:2704-2713.

Mullette-Gillman OA, Cohen YE, Groh JM. 2005. Eye-centred, head-centred, and complex coding of visual and auditory targets in the intraparietal sulcus. J Neurophysiol. 94:2331-2352.

Pouget A, Deneve S, Duhamel JR. 2002. A computational perspective on the neural basis of multisensory spatial representations. Nat Rev Neurosci. 3:741-747.

Pouget A, Sejnowski TJ. 1997. Spatial transformations in the parietal cortex using basis functions. J Cogn Neurosci. 9:222-237.

Riesenhuber M, Poggio T. 1999. Hierarchical models of object recognition in cortex. Nat Neurosci. 2:1019-1025.

Schlack A, Sterbing-D'Angelo SJ, Hartung K, Hoffmann KP, Bremmer F. 2005. Multisensory space representations in the macaque ventral intraparietal area. J Neurosci. 25:4616-4625.

Spratling MW. 2008. Predictive coding as a model of biased competition in visual attention. Vision Res. 48:1391-1408.

Spratling MW. 2009. Learning posture invariant spatial representations through temporal correlations. IEEE Transactions on Autonomous Mental Development. 1:253-263.

Spratling MW. 2011. A single functional model accounts for the distinct properties of suppression in cortical area V1. Vision Res. 51:563-576.

Spratling MW, Johnson MH. 2001. Dendritic inhibition enhances neural coding properties. Cereb Cortex. 11:1144-1149.

Stein BE, Stanford TR. 2008. Multisensory integration: current issues from the perspective of the single neuron. Nat Rev Neurosci. 9:255-266.

Stricanne B, Andersen RA, Mazzoni P. 1996. Eye-centred, head-centred, and intermediate coding of remembered sound locations in area LIP. J Neurophysiol. 76:2071-2076.

Xing J, Andersen RA. 2000. Models of the posterior parietal cortex which perform multimodal integration and represent space in several coordinate frames. J Cogn Neurosci. 12:601-614.