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

Kris De Meyer and Michael W. Spratling

Department of Informatics, King’s College London, London WC2R 2LS, UK

Address correspondence to Kris De Meyer, Department of Informatics, King’s College London, Strand, London WC2R 2LS, UK. Email: kris@corinet.org.

In multimodal integration and sensorimotor transformation areas of the posterior parietal cortex (PPC), neural responses often appear encoded in spatial reference frames that are intermediate to the intrinsic sensory reference frames, for example, eye-centered for visual or head-centered 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—similar to how complex cells in the visual cortex are thought to pool the responses of simple cells. The proposed model simulates 2 types of mixed-frame responses observed in the 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 2 types of mixed-frame responses from a third type in which sensory responses 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 organization of presynaptic responses.

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

Introduction

Posterior parietal cortex (PPC) is crucially involved in spatial awareness and the sensory guidance of actions toward spatial goals (Stein and Stanford 2008; Andersen and Cui 2009). Sub-regions of the 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, for example, eye-centered for visual and head-centered for auditory stimulation. In the course of multisensory 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 the 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 the 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 nonlinear functions. The responses of gain-modulated neurons have the required characteristics to form BF sets and hence may form the basis for the nonlinear remapping required in multisensory integration and sensorimotor transformation (Pouget and Sejnowski 1997). In BF networks, mixed-frame responses have been shown to arise naturally in the BF layer in which 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 analyzed 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. There are 2 types: 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 “nonproportional” 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 the mechanism thought to be employed by complex cells in the 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 nonproportional 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, nonproportional 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 modeling study and previous work with BF networks (Deneve et al. 2001; Avillac et al. 2005) is that proportional and nonproportional 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 analyze the spatial encoding of neural responses may be adapted to distinguish between proportional and nonproportional mixed-frame responses. Moreover, drawing further hypotheses from our simulation studies, we argue that nonproportional mixed-frame responses may be revealing of the local organization of the presynaptic, gain-modulated responses.

The model used here is the nonlinear predictive coding/biased competition (PC/BC) network, an implementation of the PC theory of cortical function that is consistent with the 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 2 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 1 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 group of model neurons (the disjunctive nodes) using a weighted-max operation. This pooling function has previously been used to model how complex cells in the 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 nonproportional mixed-frame responses in the PPC. We discuss how the model makes predictions about local cortical organization and propose how existing measures used to quantify intermediate frames of reference could be adapted to distinguish between proportional and nonproportional mixed-frame responses. We also contrast the mixed-frame responses in the PC/BC network with earlier modeling 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 3 different types of nodes: error nodes, prediction nodes, and disjunctive nodes. Error nodes and prediction nodes are reciprocally connected through feed-forward 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.

We focus on the transformation of “visual” information from an eye-centered 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 setup used in many of the physiological studies mentioned here. We do not make a distinction between head-centered and world-centered frames of reference. In such experimental settings, eye position or direction of gaze (in head-centered 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 “craniotropic,” and to intermediate or mixed reference frames as “mixed R/C.”

**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 organized 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 $b_i$ of unit $i$ was generated by:

$$h(r_x, r_y) = h_{\text{max}} \exp \left( -\frac{(r_x - a_x)^2 + (r_y - b_y)^2}{2\sigma^2} \right),$$

(1)

with $(a_x, b_y)$ the center of the Gaussian response profile, $(r_x, r_y)$ the retinal location of the visual stimulus, $\sigma$ the standard deviation, and $h_{\text{max}}$ the amplitude (maximum) of the Gaussian curve. Gaussian centers $(a_x, b_y)$ were spaced evenly in both dimensions from $-40^\circ$ to $40^\circ$ in steps of $5^\circ$, meaning that the visual input was encoded by $17 \times 17 = 289$ input units. $\sigma$ was set to $6^\circ$ and $h_{\text{max}}$ 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:

$$h(e_{\text{xy}}) = h_{\text{max}} \exp \left( -\frac{(e_{\text{xy}} - c_{\text{xy}})^2}{2\sigma^2} \right),$$

(2)

where $c_{\text{xy}}$ is the center of the Gaussian response profile and $e_{\text{xy}}$ the value of either horizontal ($e_x$) or vertical ($e_y$) eye position. For both horizontal and vertical eye position input units, $c_{\text{xy}}$ values were evenly spaced from $-40^\circ$ to $40^\circ$ in steps of $10^\circ$. Horizontal and vertical eye positions were thus each encoded by 9 input units, with $c_x = 10^\circ$ and $h_{\text{max}} = 1$. 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 $289 + 9 + 9 = 307$.  

![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, and filled arrows indicate inhibitory connections. Crossed connections signify many-to-many connectivity, and parallel connections indicate a one-to-one mapping.](https://academic.oup.com/cercor/article/23/5/1230/806340)
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 quantitatively but not qualitative differences in the results discussed.

**Model**

The model is implemented by the following 3 equations:

\[ e = x \otimes (E + W'y), \]  
\[ y = (E + y) \odot W, \]  
\[ d = \max(Q \odot Q \odot Y), \]

where \( x \) is an \((m \times 1)\) vector containing the input to the PC/BC area, \( e \) an \((m \times 1)\) vector of error node activations, \( y \) an \((m \times 1)\) vector of prediction node activations, and \( d \) a \((q \times 1)\) vector of disjunctive node activations. \( W \) is an \((m \times n)\) matrix representing synaptic weight values, each row of which, \( w_j = [w_{j1}, \ldots, w_{jn}] \), contains the weights of the synaptic connections arriving at prediction node \( j \). \( W \) is a scaled version of \( W \) with each row normalized such that its maximum value equals 1. \( Q \) is a \((q \times n)\) matrix representing the synaptic weight values from prediction nodes to disjunctive nodes. \( Q \) and \( Q \) are scaled versions of \( Q \); the first one scaled such that the maximum value in each column equals 1 and the second one scaled such that the maximum in each column equals 1. \( Y = [y_1, \ldots, y^T] \) is a \((q \times n)\) matrix, each row containing a copy of the prediction node activations. \( \otimes \) and \( \odot \) indicate element-wise division and multiplication, respectively. Function \( \max \) returns the maximum in each row. \( e_1 \) and \( e_2 \) 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 (3–5) 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. Initializing \( y \) to nonzero, randomized 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 dividedly modulated by a weighted sum of the outputs 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 2 ways. First, \( e \) can be considered to represent the residual error between the input \( x \) and the reconstruction of the input \((W'y)\) 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 \( e_x \) is sufficiently small). When a value of \( e \) 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 input to a population of competing prediction nodes. Each prediction node modulates its own inputs, which helps stabilize 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 2 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 over 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 \( r_{xj}, r_{yj}, y_{xj} \) and \( e_{xj} \). 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 “tilted” 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): \( w_{ji} = h_b(r_{xj}, r_{yj}) \), 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 \( (a, b) \)—the Gaussian centers of the presynaptic input units \( h \)—they form a 2D Gaussian distribution. The weight value from horizontal [vertical] eye position input \( i \) was calculated using equation (2): \( w_{ji} = h_b(e_{xj}, e_{yj}) \) for the node’s preferred eye position \( e_{xj}, e_{yj} \), when plotted as a function of \( c_x, c_y \), the \( w_{ji} \) values form a 2D Gaussian distribution. The weights of each prediction node \( j \) \((W)\) were subsequently normalized such that their total sum equaled 1. This was done to reflect the self-normalizing character of the learning rule used in De Meyer and Spratling (2011). Each prediction node in the network was initialized to a different combination of preferred stimuli. How these are distributed over the entire input space is summarized in Figure 3A,B. The details are network-specific and further explained in “Experimental Setup.”

The activation of disjunctive nodes \( d \) is calculated by performing a weighted max operation over the activation of prediction nodes (eq. 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 1 prediction node ever active during the disjunctive node. Disjunctive nodes were previously used in a single-area PC/BC model to simulate the responses of complex cells in the 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 the 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 $\mathbf{Q}$ 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 $(a_x, a_y)$, where $(a_x, a_y)$ was calculated as:

$$(a_x, a_y) = (r_x, r_y) + (e_x, e_y).$$  (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 $\mathbf{W}$ and $\mathbf{Q}$ (see Model). The weight values of each prediction node $j$ were initialized to scaled copies of the population input signals for 1 unique combination of input values—the node's preferred inputs $r_x$, $r_y$, $e_x$, and $e_y$ (see Model). Figure 3A,B details the spatial organization of the prediction nodes as a function of their preferred inputs—1 node for each unique combination of $r_x$, $r_y$, $e_x$, and $e_y$. In all 3 networks, visual preferences $(r_x, r_y)$ ranged from $-40^\circ$ to $40^\circ$ in steps of $20^\circ$ in both dimensions. Eye position preferences $e_x$ and $e_y$ differ across the 3 networks, as summarized in Figure 3B. Each network contained a single disjunctive node. The values of $\mathbf{Q}$ were determined using equation (6) for $(a_x, a_y) = (0,0)^\circ$, that is, the disjunctive node pooled the responses of all the prediction nodes for which the preferred visual stimulus falls at the center 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 (Fig. 2B) over 60 iterations of equations (3–5).

For the prediction nodes, the visual RF was mapped by systematically varying the visual input $(r_x, r_y)$ from $-30^\circ$ to $30^\circ$ in steps of $5^\circ$ in both dimensions while setting the eye position input to the nodes' preferred $(e_x, e_y)$ values. GFs measure the sensitivity of the preferred visual response to changes in eye position and were mapped by varying $(e_x, e_y)$ from $-30^\circ$ to $30^\circ$ in steps of $10^\circ$ in both dimensions while setting the visual input to the nodes' $(r_x, r_y)$ 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 2 stimulus sets for 1 unique combination of input values—each pair of RF graphs (RF)$f$ and normalizing this by the corresponding difference in the direction of gaze ($\Delta E_p$), 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 earlier. Each RF was estimated by systematically shifting the 2 graphs, column-wise and row-wise, calibrating the correlation between them, and taking as $\Delta E_p$ the value for which the cross-correlation reached its maximum. The average SI was calculated separately for horizontal (SI$h$) and vertical (SI$v$) shifts, and pairwise correlations for which $\Delta E_p = 0^\circ$ were discarded before calculating the mean. For cells which are organized craniotopically, the value of the average SI would equal 1; for retinotopic cells, it would equal 0.

**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 $p$ for different RF curves aligned in both retinotopic and craniotopic frames of reference, a measure previously used by Mullette-Gillman et al. (2005):

$$C_{rf} = \frac{1}{2}(\rho(R, R_i) + \rho(R_i, R)),$$  (8)

where $R$, $R_i$, and $R$ are 1D response vectors of the nodes aligned in either retinotopic ($r$) or craniotopic ($c$) coordinates. The response vectors $R$ were obtained by the 1D RF measurement procedure outlined earlier, for left ($-20^\circ$), central ($0^\circ$), and right ($20^\circ$) horizontal eye positions. The values of $C_r$ and $C_c$ 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 (RF)$f$ and normalizing this by the corresponding difference in the direction of gaze ($\Delta E_p$), 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 earlier. Each RF was estimated by systematically shifting the 2 graphs, column-wise and row-wise, calibrating the correlation between them, and taking as $\Delta E_p$ the value for which the cross-correlation reached its maximum. The average SI was calculated separately for horizontal (SI$h$) and vertical (SI$v$) shifts, and pairwise correlations for which $\Delta E_p = 0^\circ$ were discarded before calculating the mean. For cells which are organized craniotopically, the value of the average SI 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 from 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 feed-forward manner. We also replicate mixed-frame responses that have been observed in several areas of the parietal cortex, namely, different types of nonproportional mixed R/C responses. The experiment simulates 3 networks with

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:

$$(r_x, r_y) = (a_x, a_y) - (e_x, e_y).$$  (7)
different visual and eye-position preferences for the prediction nodes, as explained in Figure 3A,B (see also Experimental Setup). The resulting differences in response properties of the prediction and disjunctive nodes are discussed subsequently.

**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,D. Figure 3C shows a typical visual RF. It was mapped by systematically varying the 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 3 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 the 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 \((e_{xj}, e_{yj})\) 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 Fig. 3B), meaning that there is no competition between prediction nodes along the \(e_y\) 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 horizontal, right-of-center gaze shifts, but is strongly suppressed for horizontal, left-of-center gaze shifts. This node (with \(e_{xj} = 0^\circ\)) 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 Fig. 3B). It does, however, experience strong competition for negative \(e_{xj}\) values from the node with the same \((r_{xj}, r_{yj})\) values, but with an \(e_{yj}\) value of \(-20^\circ\).

**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 (Fig. 4A, top row): a retinotopic set (R set), in which the visual stimulus moved together with eye position, and a cranioptopic 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 Figure 4. In each network, the disjunctive node pools from gaze-modulated prediction nodes for which \((a_{xj}, a_{yj}) = (0,0)^\circ\) (eq. 6), the gray 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-center 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 the opposite direction of eye position in a retinotopic frame. Figure 4B shows the results for the 3 disjunctive nodes. When analyzed 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 analyzed 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 2 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 (Fig. 4C). Pairwise 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 position shifts (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-center horizontal gaze shifts, but craniotopic for all left-of-center horizontal shifts.

Comparison with Neurophysiological Results

The simulation results described earlier 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, that is, cells whose...
visual RF remained in the same spatial location regardless of eye position. Such cells were later also reported in the ventral intraparietal (VIP) area (Duhamel et al. 1997) and medial intraparietal (MIP) and lateral intraparietal (LIP) areas (Mullethead et al. 2005). Galletti et al. (1995) proposed a schematic model to explain how real-position behavior 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 GIs 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 behavior can indeed be constructed from the responses of strongly gaze-modulated, retinotopically organized cells.

A curious idiosyncrasy in the responses of the V6A real-position cells was that half of the reported cells were craniotopic in only 1 of the 2 dimensions and retinotopic in the orthogonal dimension. Cells with the same type of mixed responses were also observed in VIP (Duhamel et al. 1997). The most parsimonious PC/BC model that could replicate this behavior 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 2 dimensions arise naturally when pooling from response fields that are gaze-modulated in 1 dimension only. We return to this prediction in Discussion.

Network N3 displayed another type of response that has also been observed in the parietal cortex. Stricanne et al. (1996) reported a cell in area LIP whose auditory RF showed the same mixed R/C RF alignment behavior along the horizontal dimension as the response in the bottom row of Figure 4B. Such mixed, irregular behavior has also been reported for visual cells in areas MIP and LIP (Mullethead et al. 2005) and has been observed in cortical area V6A (P Fattori, personal communication) and VIP (JR Duhamel, personal communication). The crucial model assumption underlying this behavior 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 eyef dimension gave rise to the mixed R/C responses along that dimension. This again generates a testable prediction which we will address in Discussion.

Characterizing RF 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 2 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 retinotopic and craniotopic frames of reference (eq. 8). The results are shown in Figure 5A. Both the craniotopic organization of the disjunctive node in N1 and the horizontal retinotopic organization of the node in N2 are apparent from their $(Cr, Ca)$ values. The node in N3 has $(Cr, Ca)$ values that are indicative of mixed or irregular behavior. 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 analyzed by calculating a horizontal $(SI_h)$ and vertical $(SI_v)$ average SI, and the results are summarized in Figure 5B. The craniotopic behavior of N1 and the mixed R/C behavior across the horizontal and vertical eye dimensions of N2 are apparent in this measure. For the node in N3, the mixed R/C behavior 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 2 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 tilings generated by the prediction nodes form approximative 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.

![Figure 5](https://example.com/figure5.png)
(1995) proposed that real-position (i.e. craniotopic) responses might be built up in local networks within the parietal area V6A. Combining this idea of locality with our network layouts makes it possible to generate testable predictions about cortical organization.

**Prediction 1: Mixed R/C Responses for Vertical Versus 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 1 gaze dimension (4 for horizontal gaze shifts and 4 for vertical gaze shifts). In the model, such mixed R/C responses (the second row of graphs in Fig. 4) arose when prediction nodes were modulated by 1 eye-position signal only (see the middle graph in Fig. 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 (Galletti et al. 1993, 1995). It revealed that out of the 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 1-step transformation. The underlying reason for this is that BFs suffer from the curse of dimensionality, that is, the number of BFs needed to tile the input space increases combinatorially with the 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 (Stricaine 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 Fig. 4) was generated from prediction node responses modulated by left-of-center eye position, but not by right-of-center eye position (see the right-hand graph in Fig. 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 sections 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 organization, that is, cells that appear to encode information in one frame of reference for fixation points falling inside one part of the 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 toward representing, for example, left rather than right, or higher rather than lower gaze directions.

Several areas of the 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 (JR Duhamel, personal communication). The lack of published results makes it hard to determine at present whether these responses are generated by such systematic biases hypothesized earlier. An alternative explanation is that they are generated by nonsystematic irregularities in the tiling, nonsystematic 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 and nonsystematic nature of these mixed-frame responses.

**Quantitative Measures of Reference-Frame Transformations**

Both methods of analysis, taken from the physiology literature and applied to the simulated data (Fig. 5), 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 BF network (Deneve et al. 2001; Pouget et al. 2002). However, such measures do not allow making a distinction between proportional and nonproportional mixed frames and between systematic or nonsystematic origins of the mixed-frame responses.

The first type of nonproportional mixed-frame responses, that is, across 2 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) and 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 nonproportional mixed-frame responses can be made by assessing the distribution of the SIs calculated for the individual gaze shifts. This analysis becomes more feasible the more fixation points have been analyzed. For instance, for the 2D RF mapping procedure and the analysis method of Duhamel et al. (1997) (see also Fig. 4C), a total of 25 individual SI 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 nonsystematic 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 earlier. The latter case could be analyzed further by looking at the spatial distribution of the SI values: spatial structure would be indicative of systematic biases in the tiling rather than nonsystematic irregularities. Such analyses have, to date, not been performed but could in the future be used to make a distinction between proportional and nonproportional 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 nonsystematic 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 idealized 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), and 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 BF 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 nonsystematic 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 BF sets with increasing dimensionality by performing reference-frame transformations gradually or in multiple steps.

Comparison with Related Models

BF 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 2 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 the 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 BF layer), not from multimodal integration within the BF 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 2 different models. It is unlikely that the type of systematic biases in the GFs of N3 (see right-hand graph in Fig. 3D) can be easily generated with the BF 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 dimensions. Strikingly, 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 weights of the prediction nodes 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 2-stage hierarchical PC/BC model could also be learned through an unsupervised learning rule. Although the input to this hierarchical network was simpler than that 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 predetermined those weights. Extending these learning experiments to the current problem setting would allow us to compare, through simulation experiments, the potential effects of systematic biases (as reported here) and small, nonsystematic 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 BF networks (Deneve et al. 2001; Avillac et al. 2005). Combining multisensory integration with the results from this article would allow us, through simulation, to compare the properties of proportional and nonproportional mixed-frame responses.

Conclusion

We presented simulation results to demonstrate that certain types of mixed-frame responses may be generated by pooling gain-modulated responses. We argued that these nonproportional mixed-frame responses may be different from the proportional mixed-frame responses that arise through multisensory integration in BF networks. Nonproportional mixed-frame responses may be the result, and hence revealing, of systematic factors in the functional organization of certain parietal areas such as visuo-motor 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 no. EP/D062225/1).

Notes

We thank Prof. P. Fattori and Prof. J.R. Duhamel for sharing information about unpublished mixed-frame responses. Conflict of Interest: None declared.

References


