COORDINATE TRANSFORMATIONS IN THE VISUAL SYSTEM: HOW TO GENERATE GAIN FIELDS AND WHAT TO COMPUTE WITH THEM

Emilio Salinas

Computational Neurobiology Laboratory
The Salk Institute
10010 North Torrey Pines Road
La Jolla CA 92037
e-mail: emilio@salk.edu

and

L. F. Abbott

Volen Center for Complex Systems
and
Department of Biology
Brandeis University
Waltham MA 02454-9110
e-mail: abbott@brandeis.edu

Introduction

Studies of population coding, which explore how the activity of ensembles of neurons represent the external world, normally focus on the accuracy and reliability with which sensory information is represented. However, the encoding strategies used by neural circuits have undoubtedly been shaped by the way the encoded information is used. The point of encoding sensory information is, after all, to generate and guide behavior. The ease and efficiency with which sensory information can be processed to generate motor responses must be an important factor in determining the nature of a neuronal population code. In other words, to understand how populations of neurons encode we cannot overlook how they compute.

Gain modulation, which is seen in many cortical areas, is a change in the response amplitude of a neuron that is not accompanied by a modification of response selectivity. Just as population coding is a ubiquitous form of information representation, gain modulation appears to be a widespread mechanism of neuronal computation. In particular, it allows information from different sensory and cognitive modalities to be combined. Gain modulated neurons simultaneously represent multiple forms of information in a population code. The responses of ensembles of neurons are necessary to understand what the population is representing, no single neuron is sufficient. The distributed, multi-modal representations of gain modulated neurons are ideally configured to facilitate certain kinds of computations, namely coordinate transformations. Functionally, the gain-modulated population code forms a distributed substrate for both information representation and processing.

Cortical areas that process visual information are subdivided functionally and anatomically into two pathways. The 'where' pathway runs dorsally from primary visual cortex into posterior parietal cortex, and the 'what' pathway runs ventrally from primary visual cortex into inferotemporal cortex (Ungerleider and Mishkin, 1982; Goodale and Milner, 1992). Parietal cortex is involved in the spatial analysis necessary for motor planning and for the localization of external objects (Andersen, 1989; Andersen et al., 1997), whereas inferotemporal (IT) neurons are important for object recognition (Goodale and Milner, 1992; Gross, 1992). In spite of their distinct functional roles, neuronal populations in the two streams are subject to similar forms of gain modulation. Gaze direction provides a strong gain control signal in the dorsal stream, while attention provides a similar signal in the ventral stream. Although gaze-dependent and attention-dependent gain modulation act preferentially on separate processing streams and, to a good extent, independently of each other, they seem to serve the same purpose, the computation of coordinate transformations. This functional interpretation of a widespread neuronal modulatory mechanism is the subject of this chapter. We review experimental evidence revealing gain modulatory processes in the dorsal and ventral visual pathways, focus-

ing on two questions that have been addressed through analytical and simulation methods: 1) How can gain modulation be implemented by cortical microcircuitry? and 2) How can gain modulation be used to perform behaviorally useful computations?

Retinocentered receptive fields

Neurons that respond to sensory stimuli are typically characterized by their selectivity, which is expressed in terms of a receptive field. In this context, we use a somewhat expanded definition of a receptive field. For example, the receptive field of a visually responsive neuron defines not only the location in the visual field where an image must be placed to trigger a response, but also the specific image pattern that elicits the maximal response at a given stimulus intensity. The receptive field thus defines both the preferred location and preferred visual stimulus for a given neuron. These two receptive field attributes change progressively as information flows more centrally in the visual system, so that receptive field sizes tend to increase while preferred images become more complex.

The receptive fields of neurons early in the visual pathway, such as those of retinal ganglion cells and cells in the lateral geniculate nucleus of the thalamus (LGN), are best described in retinal coordinates. This is because the locations of the receptive fields of these neurons are fixed to the eye or, equivalently, are always the same relative to the direction of gaze. Neurons in primary visual cortex or V1 are usually described in retinal coordinates as well, although recent reports suggest a more complex description (Guo and Li, 1997; Trotter and Celebrini, 1999; but see Sharma et al., 1999). A retinocentric receptive field is schematized in Figure 1. At this level of processing the actual gaze direction, which is determined by a combination of eye and head positions with respect to the body, does not influence neuronal activity by itself. If an image is shown at two different locations in visual space, the neuronal response does not change systematically as long as these locations correspond to the same position on the retina.

Figure 1. Early visual neurons operate in retinocentered coordinates. Thalamic visual neurons respond to contrasting center-surround patterns. The bars on the right represent the expected firing rate of a hypothetical thalamic neuron in response to the images shown on the left. The cross indicates the fixation point, the location to which gaze is directed. **a**, The image is aligned with the receptive field, and the neuron fires rapidly. **b**, If the pattern is shown at the same physical location but at a different position with respect to the fixation point, the neuron does not fire. **c**, If the pattern is moved to the original location with respect to the fixation point the neuron responds again, regardless of the actual gaze angle. In other words, when the gaze direction changes, the receptive field moves with the eyes.

3

Gain modulation in parietal cortex

Richard Andersen, initially working with Mountcastle (Andersen and Mountcastle, 1983) and later in collaboration with others (Andersen et al., 1985; Andersen, 1989; Andersen et al., 1990; Brotchie et al., 1995), showed that, in contrast to thalamic or retinal ganglion neurons, visual responses in parietal cortex depend on both the retinal location of a visual stimulus and on gaze direction. In these experiments, parietal neurons responded to spots of light located at various places within the visual field. If gaze direction is held fixed and the response is plotted as a function of the position of the spot, the resulting curve typically has a single peak and can be fitted to a Gaussian function. Figure 2c shows an example. Notice that the x axis represents retinal coordinates, referring to the stimulus position with respect to the location of gaze (the fixation point). If the set of measurements is repeated using a different fixation point and thus a different gaze direction, the neural response follows a curve with similar shape and preferred location, but with a different amplitude. Thus, the amplitude or gain of the receptive fields of these parietal neurons depends on gaze. As a comparison, the same experiment probing early visual neurons would produce the same peaked curve for all gaze directions. The term 'gain field' was coined to describe this gaze-dependent gain modulation and how it varies as a function of gaze direction. The dependence on gaze direction is fairly linear, although sometimes it is closer to sigmoidal due to saturation effects.

Figure 2. Gain modulation of parietal neurons. **a**, Visual stimuli are displayed at different locations while a monkey directs its gaze (fixates) straight ahead. The cross indicates the fixation point, the large circle indicates the location of the neuron's receptive field, and the small circles show the locations where stimuli are presented, one at a time. **b**, Visual stimuli are displayed at a similar set of locations while the monkey directs its gaze to the left. In **a** and **b** the stimuli are presented at the same locations as measured by retinal coordinates. **c**, Neuronal activity recorded at two different gaze angles during experiments like those in **a** and **b** (although both eye and head position were varied). A neuron's firing rate is plotted as a function of stimulus location in retinal coordinates, and the two data sets (filled and open symbols) correspond to two different gaze directions. The continuous lines are Gaussian fits. The peaks, which correspond to the preferred spot locations, are the same in the two cases, but the amplitude or gain is different. **d**, Hypothetical gain field (*i.e.* gain factor as a function of gaze angle) for the same neuron. For simplicity, only dependence on the horizontal direction is indicated. Full gain fields are two-dimensional. Diagrams redrawn from Andersen et al., 1985; data redrawn from Brotchie, et al., 1995.

A network-based mechanism for gain-modulatory interactions

One of the striking aspects of gain modulation is that the interaction between

gaze direction and retinal location of the visual image is very close to being multiplicative. Neither the baseline firing rate nor the shape of the response curves of the parietal neurons change as a function of gaze direction, and the gain modulated responses are well described by a product of two functions. One, $f(x)$, depends on stimulus location in retinal coordinates, x, and corresponds to the Gaussian response curve discussed above. The other function, $g(y)$, depends on gaze angle, y, and corresponds to the gain field. The firing rate can then be written as

$$r = f(x)\,g(y). \tag{1}$$

Neurons are typically modeled and thought of as integrators that compute weighted sums of their inputs. How can they achieve the type of nonlinear, multiplicative behavior seen in gain modulation? An early study (Mel, 1993) proposed that nonlinear cooperative interactions between neighboring synapses connected to the same neuron could generate responses that depended multiplicatively on the inputs. This explained multiplicative interactions between two input signals on the basis of nonlinear interactions between synaptic conductances. Although this is a plausible scenario, the extent to which these or other nonlinearities (Koch and Poggio, 1992) give rise to an effective gain-like multiplication remains a question.

Figure 3. A model circuit that produces multiplicative gain fields. **a**, A group of parietal neurons (filled circles) receives input from two external networks (open circles). The network on the left provides a signal $h^G(y)$ that depends on gaze angle, y, and is identical for all target cells. The network on the right provides visual signals $h_i^V(x)$ that depend on the location of the visual stimulus, x, and are different for different target cells. The $h_i^V(x)$ terms correspond to Gaussian receptive fields centered at different locations. As indicated by the equation, the firing rate of target cell i is determined by the sum of its external inputs plus a weighted sum of the activity of its neighbors minus a constant threshold. The square brackets with a plus subscript indicate rectification. As indicated by the upper plot, the weight w_{ij} is positive (excitatory) if neurons i and j have similar receptive fields and negative (inhibitory) otherwise. **b**, The upper graph shows the responses of a model parietal neuron. Each curve traces the firing rate as a function of stimulus location in retinal coordinates for a fixed gaze angle. The three curves correspond to three different gaze directions. The squares are the actual simulation results, and the lines are fits using Equation 1. The lower panel shows the total external input (the sum of the two h terms) for each neuron in the network. Because the gaze signal is the same for all neurons, the input curves simply shift up or down as a function of gaze. Nevertheless, the output curve shows a multiplicative interaction. Results modified from Salinas and Abbott, 1996.

Another mechanism proposed later (Salinas and Abbott, 1996), is based on the rich dynamics of networks with recurrent connections. Recurrent connectivity is a well described feature of cortical circuits (Gilbert and Wiesel, 1983, 1989; Ahmed et

al., 1994, 1997), and its dynamical properties have been implicated in many aspects of cortical function such as response selectivity (Ben-Yishai et al., 1995; Somers et al., 1995; Chance et al., 1999), signal amplification (Douglas et al., 1995), and sustained neuronal activity (Seung, 1996). Figure 3a shows a model network representing a small patch of parietal cortex where all cells have similar gain fields. Each of the model parietal cells receives two kinds of external inputs. One, $h^G(y)$, provides an eye-position signal and is the same for all target neurons. The other input, $h^V(x)$, provides visual information which can be thought of as coming from early visual cells responding to spots of light. This input is different for each target neuron, giving them different preferred stimulus locations.

The two external inputs to target neuron i, h^G and h_i^V, are added together. However, the response is not simply additive because, in addition to these external inputs, the model parietal cells receive recurrent input from their neighbors. This is a key property of the network, and the critical parameters are the connection weights w_{ij}. For the multiplicative responses to arise, similarly tuned neurons (with overlapping receptive fields) should excite each other, whereas neurons with different (non-overlapping) receptive fields should inhibit each other. As long as this general rule holds, the exact dependence of the connection strengths on receptive field properties does not affect the result.

The top graph of Figure 3b shows the responses of one model parietal neuron as a function of stimulus position. The curves match those of Figure 2 discussed above. The bottom graph shows the input to all of the neurons in the network. For different gaze angles all inputs are displaced up or down by the same amount because the gaze input signal is the same for all neurons. The resulting response curves on the top panel are almost exactly scaled versions of each other. The response curve changes in amplitude but does not change its shape. These model neurons are well described by Equation 1, so they have an almost perfectly multiplicative gain field. This result is quite robust. When noise is added to the inputs or when variability is introduced in the network parameters, the response functions are somewhat distorted and some variability across neuronal response curves arises, but the effect is still close to multiplicative. However, when the recurrent connections are turned off in the model, the responses are far from multiplicative. Thus, recurrent connections seem to be critical for generating multiplicative gain modulation (Salinas and Abbott, 1996).

This model has an additional property: if two inputs at two different retinal locations are presented to the network, the activity profile will have a single peak located at or close to the location of the strongest input. Thus the network also provides a mechanism for target selection, suppressing the activity driven by weak input stimuli. This is consistent with the finding that at least some parts of parietal cortex have a very sparse representation of the visual scene, coding faithfully the

locations of only those stimuli that are salient or behaviorally relevant (Gottlieb et al., 1998).

In summary, individual neurons in a recurrently connected network may add inputs driven by independent sources corresponding to different sensory variables, and nevertheless generate responses characterized by a a product of functions for each of those variables. In this mechanism there is no need to invoke explicit multiplication at the synaptic or cellular level. Gain modulation is an emergent property of the network. Since recurrent connections are ubiquitous in the cortex, this connectivity may also provide a basis for multiplicative interactions between other kinds of input signals across cortical areas and modalities (see below).

Figure 4. Schematic example of a simple coordinate transformation. Gaze is directed toward the left corner of the monitor and the task is to reach the mouse without shifting gaze. x is the location of the mouse in retinal coordinates and y is the gaze angle. To reach the mouse a movement in the direction specified by $x + y$ must be performed.

Coordinate transformations for object localization

Imagine that you are facing a computer monitor (Figure 4), directing your eyes toward the left corner of the screen, and you want to reach the mouse without shifting your gaze. Arm movements are generated with respect to body position (see for example Georgopoulos, 1995), and to compute the direction to the mouse in body-centered coordinates, the retinal location of the mouse must be combined with the current eye position relative to the body. The angle x corresponding to the retinal location of the mouse and the gaze direction y must be added to obtain the angle $x + y$ describing the mouse location with respect to the body axis (Figure 4). How do neurons perform this addition needed to generate the coordinate transformation from retinal to body-centered reference frames?

The first indication that gain modulation could be useful for such coordinate transformations came from the work of Zipser and Andersen (1988). They trained a three-layered artificial neural network to perform the transformation just described. The network was presented with various target locations in retinal coordinates and with various gaze angles. The network was trained, using backpropagation and many examples of correct input-output associations, to compute the target locations in body-centered coordinates. Once the network had learned the correct transformations, they examined the properties of the neurons in the hidden layer. These responses, generated by connections that the backpropagation procedure had produced during training, were similar to the gain-modulated receptive fields found in the recorded parietal neurons. This result suggested that gain modulation provides an efficient solution to the coordinate transformation problem given the input

Figure 5. A model network that performs a coordinate transformation using gain modulation. **a**, The large, square network represents a population of parietal neurons with all combinations of the receptive field locations and gain modulation parameters described experimentally. These neurons respond to object location x and gaze angle y according to the equation appearing under the network. The network on the top represents an array of neurons that must encode a linear combination of x and y to generate a motor response given, as indicated by the equation above the linear array of neurons, as a function of the sum $x+y$. The output neurons are driven by the parietal model neurons through synaptic connections indicated by arrows. **b**, The bottom panel illustrates the responses of one gain-modulated parietal neuron, with the three curves corresponding to different gaze directions. The amplitude of the response changes with gaze, but the location of the peak response remains constant. The top panel shows the response of one output neuron. The tuning curve shifts when the gaze angle changes, but its amplitude remains constant. Results modified from Salinas and Abbott, 1995.

and output representations. This work revealed that the measured neurophysiological properties of real neurons could be understood to underlie a specific, nontrivial computation.

Zipser and Andersen imposed the computation of a coordinate transformation on a network and observed that gain-modulated responses resulted. Another approach is to put gain-modulated responses into a network from the start, and determine the conditions under which coordinate transformations arise. This provides insight into how and under what conditions gain modulation can perform coordinate transformation calculations (Salinas and Abbott, 1995). The network used for this purpose is shown in Figure 5. The bottom portion of Figure 5a represents a set of parietal neurons that have gain-modulated receptive fields like those described experimentally. Their responses are determined by Equation 1, with the set of neurons including combinations of receptive field locations and gain field modulations taken from the reported distributions across the population of parietal neurons. The output neurons at the top of the network figure represent an array of neurons that encode the target location in body-centered coordinates and can generate a motor response such as reaching to the target. They must have firing rates that are functions of the sum of stimulus location and eye position. The connections between the two layers allow the bottom array to drive the top one. For a particular target position in retinal coordinates, x, and gaze angle, y, some of the parietal neurons are activated, and they must drive the output neurons so that these encode $x+y$, the target location in body-centered coordinates. Given this setup, we can determine the synaptic connections that allow the parietal neurons, which combine x and y through a gain interaction, to drive downstream neurons that have responses that depend on $x + y$.

Our analytical and computer simulation calculations produced three results. First, all that is needed to guarantee that the output neurons encode $x+y$ is that the synaptic weights satisfy a simple mathematical condition (Salinas and Abbott, 1995). Figure 5b (top) shows the response function of one output neuron when this condition is satisfied. When gaze is directed straight ahead, the preferred location of the neuron (the peak of its response as a function of target location x) is at 0. When the gaze angle y changes, the receptive field shifts by an equal amount without changing shape, exactly as expected from a neuron responding as a function of $x+y$. Once the condition that guarantees a correct transformation is known, the question becomes what synaptic modification mechanisms can produce connections that satisfy such condition? The second result is that a simple Hebbian or correlation-based synaptic modification rule operating during a learning period in which self-generated movements are watched will automatically produce the appropriate connections. During learning, arm movements to different parts of space must be generated by the activity of the output neurons, and they must also be watched, thereby driving the responses of parietal neurons. Observations of these movements should occur for all possible combinations of retinal and gaze angles. These images provide examples of correct transformations because, when the hand acts as the target, the retinal and body-centered representations are automatically aligned, *i.e.* the arm angle always equals $x+y$. Indeed babies watch their own limb movements before they can control them (Van der Meer et al., 1995). Finally, a third result is that not only the sum but also any other linear combination of target location and gaze angle can be represented by the array of output neurons, as long as this linear combination corresponds to the output representation used during learning. This means that one downstream network may extract or read out $x+y$ from the activity of parietal neurons, while another downstream network may equally well read out another combination such as $x-y$ from the same responses using similar mechanisms.

This last point has been thoroughly elaborated by Pouget and Sejnowski (1997a). They showed that many psychophysical and lesion data in which the coordinate frame used for object localization could not originally be determined unambiguously, are actually consistent. The key is that the positions of objects are not encoded in one fixed coordinate frame, but through neural activity in parietal cortex from which any appropriate coordinate frame may be read out according to ongoing task requirements (Pouget and Sejnowski, 1997a). They also simulated the effects of a lesion in a model of parietal cortex (Pouget and Sejnowski, 1997b) and found that the model reproduced many of the typical effects found in patients. In particular, the deficit affected multiple frames of reference, including object-centered. The consequences of parietal lesions have been difficult to reconcile, perhaps because the flexibility of the gain-modulated spatial representation used by parietal cells led to widely different outcomes, depending on the combination of encoded

quantities being read out by downstream networks.

In conclusion, the advantage of a gain-modulated representation, consisting of a set of neurons tuned to a quantity x and gain modulated by a quantity y, is that a downstream network can easily extract any linear combination of x and y using correlation-based learning. This is particularly useful for sensory-motor control where movements can be practiced.

Additional evidence for modulatory gain control

Following the early studies of Andersen and colleagues, gaze-dependent gain modulation has been reported at numerous stages of the visual system (Galletti and Battaglini, 1989; Galletti et al., 1995; Van Opstal et al., 1995; Guo and Li, 1997; Snyder et al., 1999; Trotter and Celebrini, 1999) and in areas involved in motor functions (Boussaoud et al., 1998). Some studies have shown that even in primary visual cortex around 50% of the neurons display substantial gaze-dependent gain modulation of their orientation or disparity tuning curves (Guo and Li, 1997; Trotter and Celebrini, 1999; but see Sharma et al., 1999). Thus, gain modulation depending on eye position may operate simultaneously at different processing stages, possibly generating a larger final effect.

Reaching for objects requires the computation of multiple coordinate transformations similar to the one shown in Figure 4. The gain-modulated representations in parietal cortex may be the basis for the subsequent spatial representations needed for motor execution and object localization (Pouget and Sejnowski, 1997a, 1997b). The simple two-layer network model for such transformations shows that explicit visual representations of the world in head-centered or body-centered coordinates are not absolutely necessary, because the readout units could be the same neurons producing the motor commands. However, visual neurons that are invariant to eye position have indeed been found (Graziano et al., 1997; Duhamel et al., 1997). They represent the output readout layer of a transformation process. In fact, the full transformation from retinal to world-centered coordinates seems to be explicitly computed, and the underlying mechanism appears to be gain modulation. Recent recordings from parietal cortex have shown that area LIP has mostly gain fields that depend on gaze direction, leading to body-centered coordinates useful for gaze control and object reaching, whereas area 7a has mostly gain fields that depend on body position with respect to the world, which may lead to world-referenced responses invariant to eye, head, and body orientation (Snyder et al., 1998). This is consistent with the existence of place fields in the rat hippocampus that encode the animal's position with respect to its environment. Area 7a projects directly to this structure (Snyder et al., 1998) which is believed to be strongly involved in spatial computation (O'Keefe and Nadel, 1978). Thus, gain-modulated signals that depend on eye, head, and body position may be progressively combined until a map

of extra-personal space is formed that is fully invariant with respect to the subject's position.

Coordinate transformations for object recognition

We mentioned at the beginning that thalamic and retinal ganglion neurons operate in retinocentered coordinates because their receptive fields move with the eyes. This, however, poses a problem. A characteristic feature of our visual system is that we are able to recognize objects independently of their location and size. This is not true for the full visual field, since a familiar face may not be recognizable if it appears far in the periphery, but it is true for a large, central region where we can identify a familiar image or object even if we don't look directly at it.

A neural correlate of this phenomenon is provided by high-level visual neurons like those found in area IT. IT neurons are often selective for highly structured complex images (Desimone, 1991; Logothetis et al., 1995). They may respond strongly to faces, for example, producing little or no response for a large variety of other objects. The receptive fields of these neurons are large; diameters of 60° or more are not unusual. More important than sheer size is the property that a relatively small image can have approximately the same effect no matter where it is placed, as long as it is inside the receptive field perimeter (Schwartz et al., 1983; Desimone et al., 1984; Tovee et al., 1994). Thus, if a neuron is selective for faces, a face presented anywhere inside the receptive field will typically produce a much stronger response than a non-face image anywhere inside the receptive field. This is known as translation invariance.

Translation-invariant responses correlate with our capacity to perform location-independent object recognition, but how are these responses generated? This is not a trivial problem, since translation-invariant responses must be evoked by the activity of early visual neurons that are not themselves translation-invariant. Some models achieved invariance to location (and to other image parameters such as scale and perspective) through synaptic modification rules that link the images of objects appearing close together in time (Földiák, 1991; Wallis and Rolls, 1998). Buonomano and Merzenich (1998) also exploited temporal structure, in this case the spike patterns arising from a distribution of latencies, to generate position-invariant neural responses. Other models have been based on the hierarchical multi-layered structure and nonlinearities exhibited by the visual system (Fukushima, 1980; Wallis and Rolls, 1998; Riesenhuber and Poggio, 1998, 1999). Detailed experimental support for these mechanisms is not abundant, but it is possible that the visual system uses some or all of them in order to achieve the level of translation invariance exhibited behaviorally.

A different approach was taken by Olshausen, Anderson and Van Essen who, building on earlier ideas proposed by Hinton (1981a, 1981b), developed a model in

which the translation invariance problem is cast as a coordinate transformation, in this case from a retinocentered representation to an object-centered one based on attention (Anderson and Van Essen, 1987; Olshausen et al., 1993). Their model required precisely coordinated synaptic interactions between inputs carrying attentional and visual signals. It motivated an important experiment by Connor and collaborators in David Van Essen's laboratory (Connor et al., 1996, 1997). The results ended up providing only partial support for the proposed mechanism, but uncovered an unexpected and dramatic effect that gave rise to yet another framework for the translation invariance problem. The resulting model (Salinas and Abbott, 1997a, 1997b), like that of Olshausen et al. (1993), implemented a transformation from retinal to attention-centered coordinates, but it was based on gain modulation, very much like the coordinate transformations in parietal cortex.

Connor and collaborators (1996, 1997) studied the responses of neurons in V4, an area that projects directly to IT (Felleman and Van Essen, 1991). V4 neurons have receptive fields that are more complicated than those in V1 (Gallant et al., 1993, 1996), but they also respond well to oriented bars flashed inside their receptive fields. More importantly, their responses also depend on attention. This dependence was determined quantitatively, and the effect can be described as a gain field, an attentional gain field.

Attention is frequently described as a spotlight; an activity beam that enhances neural responses in the vicinity of a particular region of visual space. Whatever the accuracy of this analogy, it should be stressed that the locus of attention is not equal to gaze location. This is an important feature of attention; it can modulate visual responses even when the eyes are fixed, and in the absence of any apparent change in behavior. In the experiments by Connor et al. (1996, 1997), eye position was kept fixed and the location where attention was directed, $\mathbf{z}$, varied independently of eye position. In this context, attention is a 'spotlight' that can be directed to various locations and that moves independently of eye position.

According to the experimental data (Connor et al., 1996, 1997), the firing rate R of a V4 neuron can be expressed as a product of two factors. The first one, F, corresponds to a purely visual response and depends on the image I being shown. The function F describes the visual receptive field properties such as location, orientation preference, spatial frequency selectivity, and so on, and it can be characterized by filtering operations like those commonly used to describe the responses of V1 complex cells (Heeger, 1991). The second term, G, depends on the location $\mathbf{z}$ where attention is directed, relative to a point called the preferred attentional locus of the neuron, $\mathbf{b}$. The function G defines the attentional gain field, and $\mathbf{b}$ is the locus where directed attention produces the maximum gain (Figure 6a). When $\mathbf{z}$ is equal to $\mathbf{b}$, this modulatory term is maximal, whereas if attention is directed far away from the

preferred attentional locus, it goes to zero. Combining the two terms gives

$$R = F(I; \mathbf{a})\, G(\mathbf{z} - \mathbf{b}). \tag{2}$$

Here $\mathbf{a}$ corresponds to the receptive field center. This expression, which represents a reasonable mathematical fit to the data, summarizes the experimental findings and describes how the locus of attention controls the gain of V4 cells.

Figure 6. A model of translation-invariant responses based on attentional gain modulation. **a**, The response of a V4 neuron depends on the product of its receptive field and its attentional gain field (Connor et al., 1996, 1997). The small open circle represents the receptive field of a V4 neuron with its center at position **a**. The large gray circle represents the attentional gain field of the neuron with its center at position **b**, the preferred attentional locus. To evoke a strong response, an image that matches the receptive field selectivity must appear at **a**, while attention is directed to a location near **b**. **b**, Network model for translation-invariant responses. The bottom grid is a pixel array on which images are displayed. The middle grid represents an array of V4 neurons that respond to the image and are gain modulated by attention according to Equation 2. Each crossing point in this grid represents a set of V4 neurons with the same receptive field location but with different combinations of preferred orientation, optimal spatial frequency, and preferred attentional locus. The topmost neuron represents an IT cell that is driven by the activity of the V4 layer through synaptic connections W_i. In the expression shown, θ is a constant threshold, R_i is the firing rate of V4 cell i, and the square brackets with a plus subscript indicate rectification. **c**, Response of the model IT neuron versus the location of a preferred image. The response is large because this is the same image used during the training of the network. Filled symbols correspond to attention located at pixel 16 and open symbols correspond to attention located at pixel 48. The IT response depends on the location of the image relative to the point where attention is focussed. **d**, Response of the same model IT neuron versus the location of a less effective image. The response is much reduced, because the cell is selective for the image used during training, not this image. As in **c**, the response depends on stimulus location in attention-centered coordinates. Images sizes were 16 by 16 pixels. Results modified from Salinas and Abbott, 1997a.

Neurons with visual responses modulated by attentional gain fields can generate translation-invariant responses in downstream neurons if theses are driven by synapses with appropriate strengths (Salinas and Abbott, 1997a, 1997b). This can be shown analytically if a few simplifications are allowed, and its validity under more general conditions can be verified by simulating a model network. The key for this result to be true is that the set of V4 neurons should include many combinations of receptive field properties and gain field centers that are not correlated or aligned. Thus, a given location in the visual field must be covered by neurons with different combinations of preferred orientation, preferred spatial frequency,

and other receptive field parameters. For each of these receptive fields, there must be several neurons with different preferred attentional loci that are independent of the receptive field parameters. The experimental data support these assumptions (Connor et al., 1997).

The model used in the simulations is schematized in Figure 6b. An image is projected on a pixel array to generate a set of model V4 responses determined by Equation 2. These responses are then synaptically weighted through feedforward connections to produce a model IT response. The array of V4 responses consists of a set of 32 by 16 receptive field centers spread uniformly. At each location, there are neurons with 4 orientation preferences, 3 frequency selectivities, and 6 different gain field centers (for horizontal translation only). This gives a total of about 37,000 model V4 responses. The crucial elements in the network are the synaptic weights W_i. We have found a mathematical condition that the weights must satisfy in order for the IT neuron to be translation-invariant. This condition can be satisfied if the weights develop through simple correlation-based learning, which is precisely how they were set in the simulations.

One simple training procedure that produces synaptic connections that satisfy the condition for translation invariance is the following. During a training period, a selected image is presented and translated to all locations while the IT neuron is set active (*i.e.* its firing rate, R_{IT}, is set to a high value throughout this period) while Hebbian learning takes place. Every time the training image appears at a given location, the V4 responses are computed and an amount proportional to the product $R_i R_{IT}$ is added to each synapse W_i, where R_i represents the firing rate of V4 neuron i and R_{IT} is the preset response of the IT neuron. During the training period, the location of attention is maintained at the center of the training image. Once the image has appeared at all locations, the weights are not modified any more, and the model is tested to determine how the IT cell responds. It is important to stress that this particular mechanism for establishing the connections is not crucial for the success of the model. Weights that satisfy the condition needed for translation-invariant responses are not unique and could thus be established in different ways.

Figure 7. Correspondence between invariant object recognition and activity in inferotemporal cortex according to the model based on attentional gain modulation. Rectangles correspond to visual displays. The small crosses represent the fixation point (the point to which gaze is directed), and the crosshairs, which are not part of the actual visual display, indicate the location where attention is directed. The bars on the right show the expected response of a hypothetical face-selective IT neuron in the three situations **a**, **b**, and **c**, according to the model.

Figure 6c shows the response of the model IT neuron to translated versions of the same image used during learning. The plot shows the firing rate of the IT cell

as a function of image location. The filled circles correspond to attention located at pixel 16, whereas the hollow circles correspond to attention directed at pixel 48. The IT neuron responds strongly whenever attention is focussed close to the center of the image, regardless of the location of the image on the retina or viewing screen. Figure 6d shows that when a different pattern is shown to the same IT cell the evoked response is smaller. This response falls off gradually as the image moves away from the location of attention. The receptive field of the model IT neuron thus shifts with attention, while retaining its selectivity for a specific image pattern; the model neuron operates in attention-centered coordinates.

Figure 7 schematizes the relationship between object recognition and IT activity according to the mechanism based on a coordinate transformation from a retinal to an attention-centered reference frame. This figure applies equally to the models by Olshausen et al. (1993) and Salinas and Abbott (1997a, 1997b). When an object for which the neuron is selective appears in the visual field, the IT neuron responds when attention is directed toward that object (Figure 7a). If a different object appears and attention is drawn to it, the neuron stops responding because, even though the new image is located at the center of the neuron's receptive field, the cell is not selective for it (Figure 7b). However, when attention is directed back to the first image, the neuron fires rapidly again (Figure 7c). Notice that the visual display is exactly the same in Figures 7b and 7c, only the locus of attention has changed. The models predict that an object may be recognized when attention is focused on it, but not when it appears far away from the attended location.

Mack and Rock and collaborators have performed psychophysical experiments that support this prediction (Mack and Rock, 1998). They designed a paradigm in which a test stimulus is displayed while subjects perform an attentionally demanding visual task unrelated to the test stimulus. In trials in which subjects performed the task and expected the appearance of the additional but unspecified stimulus, they were able to identify it reliably, without affecting performance of the primary task. In this condition, the subjects presumably divided their attention in such a way that they focused on the two stimuli simultaneously and effectively. In contrast, when subjects were engaged in the primary task and were not expecting the appearance of a test stimulus, a significant fraction of the time they were not even able to detect it; the test stimulus simply went unnoticed in many of these trials ($\sim$25%), presumably because attention was fully devoted to the primary task. This, in itself, was quite remarkable, but most importantly for our discussion, the subjects that did report seeing an additional stimulus in this 'inattention' condition could not identify its shape above chance levels. This is in marked contrast to other attributes of the test stimulus, such as color, orientation, location and numerosity, which were not correctly identified all the time, but still were identified more often than expected from random guessing. These results suggest that, if attention is far

away from a visual stimulus, its shape cannot be determined. This provides strong evidence for attention playing an essential role in object recognition, as suggested by the theoretical models.

Two important points should be stressed about the mechanism for translation invariance that we have described. First, we have modeled only a single IT neuron, but other units could be included in the network, driven by the same population of V4 neurons. Different IT neurons could then be selective for different images. The major constraint imposed by the simple learning mechanism that we have proposed is that only small numbers of neurons should be active at the same time during learning. But regardless of the specific mechanism, if the synaptic weights satisfy the proper condition, any number of receptive fields can operate in attention-centered coordinates. The second point is that the model requires on the order of 250,000 V4 neurons modulated by attention for full two-dimensional translation. It thus seems that the cost of invariance is a large number of driving neurons. However, these V4 cells have simpler receptive fields than IT neurons, and the same population of V4 neurons can provide the basis for any number of complex IT receptive fields that need to be translated to an attention-centered system.

Desimone et al. (1983) found that face-selective neurons in IT cortex of anesthetized monkeys responded strongly to faces presented anywhere inside a large bilateral receptive field. This does not necessarily constitute evidence against an attention-centered reference frame as a basis for these responses, because attention could operate under the control of 'automatic' mechanisms. Eye movements can be controlled consciously but they need not be, and they can also take place during sleep. Similarly, we can consciously direct our attention, but this does not prove that an attentional locus is absent under anesthesia. Manipulations that are known to eliminate attentional effects selectively would be required to settle this question.

A word should also be said about learning mechanisms that could possibly give rise to synaptic weights satisfying the condition for translation invariance. We already mentioned that there is no unique set of such synaptic connections. Nevertheless, regardless of the synaptic modification rules considered, one aspect of the training procedure does put a general constraint on the model: an object can be recognized at a given position only if it has appeared previously at that position during learning. Therefore, if during learning an object appears only in the left hemifield, later the model IT cell will respond to that object strongly and independently of position only when it appears in the left hemifield. No response will be observed when it is shown anywhere in the right hemifield. The learning process itself is not translation invariant. This applies particularly to distances larger than a few degrees, because invariance over smaller distances could also be obtained through other mechanisms, for example from the properties of complex cells (Logothetis et al., 1995; Riesenhuber and Poggio, 1998, 1999). This restriction is consistent with

the results of psychophysical experiments. In certain visual tasks such as discrimination of unfamiliar images, in which increases in psychophysical performance can be controlled and quantified, learning is location-specific. Increases in performance are either absent outside the specific location used during learning (Karni and Sagi, 1991; Dill and Fahle, 1997), or acquired gradually for subsequent locations (Sigman et al., 1999). These results suggest that translation invariance may indeed require objects to appear at different locations during a training period.

Gain modulation as a generalized control mechanism

In the 'where' visual pathway, eye position affects activity at multiple points in the processing chain. Attention also seems to act in parallel at multiple sites, as effects have been found in many visual areas, including primary visual cortex (Motter, 1993; Connor et al., 1996, 1997; Luck et al., 1997; Vidyasagar, 1998; McAdams and Maunsell, 1999; Treue and Martínez-Trujillo, 1999). As noted for eye position, attentional control may be more effective when acting at different points along a hierarchical processing stream.

Visual neurons are typically selective for a number of stimulus attributes, such as orientation, color, and spatial frequency, and may be modulated by multiple quantities as well, such as eye position and attention. Previously, we argued that recurrent connections could give rise to gaze-dependent gain modulation, but could this mechanism account for the effects of other or even multiple modulatory influences? In the model, the same modulatory input $h^G(y)$ (see Figure 3) is added to a group of recurrently connected neurons. As long as this input is common to them and independent of other tuning properties, its modality or origin is irrelevant. An input that was a function of the location of attention would produce exactly the same scaling of a tuning curve as seen in the example using eye position modulation. If two modulatory inputs depending on quantities y and z act independently, so that the total modulatory input is $h^G(y) + h^G(z)$, the two influences would be combined additively to determine the total gain.

Recent studies are consistent with this prediction. McAdams and Maunsell (1999) investigated the effects of attention on the orientation selectivity of neurons in V4 and V1, and found that tuning curves were almost exactly scaled by attention. The authors pointed out the ubiquity of multiplicative interactions among many stimulus dimensions, noting that by virtue of its multiplicative effect attention is put on the same footing as many other sensory attributes. Contrast, for instance, provides a well-known example of a multiplicative influence on tuning properties (McAdams and Maunsell, 1999). In another study, Treue and Martínez-Trujillo (1999) showed that attending to a selected location and to a selected feature both have almost perfectly multiplicative effects on direction tuning in area MT. Furthermore, they showed that the effects combined additively. Thus, the model based on recurrent activity

fits well with a variety of measured interactions between tuning properties and modulatory inputs. Nevertheless, its tendency to select one stimulus over others when many of them are presented simultaneously may be incompatible with the many cortical areas that are subject to attentional modulation under a wide variety of conditions. Attention sometime does act as if it were suppressing irrelevant stimuli (Reynolds and Desimone, 1999), but not all observations are consistent with this (Treue and Martínez-Trujillo, 1999). Thus, although the recurrent network may capture some important properties of the circuits underlying gain modulation in general, the specific mechanisms that may produce attentional modulation consistent with the growing body of experimental observations need to be worked out in more detail.

Although we have discussed and modelled gain modulation that is multiplicative, this is not a critical feature. The key property is that gain modulation must combine information about the modulatory influence with information about the sensory stimulus in a nonlinear way. In simulations, we have found that even large deviations from a product relationship can produce results similar to those found with exact multiplication, as long as the two terms involved are combined nonlinearly. However, our results suggest that multiplicative gain control may be advantageous if it is combined with a Hebbian synaptic modification mechanism. In this case, neural representations in a new reference frame can be established through correlation-based learning. The repertoire of synaptic modification rules available to a circuit must place strong constraints on the neural representations that it can use.

Closing remarks

In conclusion, gain modulation may be a generalized mechanism by which populations of neurons may encode sensory stimuli and other kinds of information, with the advantage that such representation may greatly facilitate certain computations. The two models for coordinate transformations that we discussed involved separate anatomical structures in the dorsal and ventral visual pathways, as well as independent modulatory effects. However, as can be appreciated by comparing Equations 1 and 2, the two gain control signals exert identical effects, regulating the amplitude of a set of visual responses. Our findings support earlier conclusions by Andersen and collaborators, who suggested that gain fields are an efficient means to perform coordinate transformations in general. The presence of gain fields at one stage of a processing pathway suggests that responses at a downstream stage will be in a different coordinate system. Similarly, the presence of transformed or invariant responses at one area suggests that gain modulated responses will be found at points upstream from this area. We also found that gain fields arise naturally from the recurrent connectivity that is characteristic of cortical circuitry (Gilbert and

Wiesel, 1983, 1989; Ahmed et al., 1994, 1997). Further experiments should outline more precisely how gain fields are built and exploited by the nervous system, but they will probably remain a prime example of neural design serving a computational purpose.

Acknowledgements

References

Ahmed B., Anderson, J.C., Douglas, R.J., Martin, K.A., Nelson, J.C. (1994) Polyneuronal innervation of spiny stellate neurons in cat visual cortex. J. Comp. Neurol., 341: 39–49.

Ahmed, B., Anderson, J.C., Martin, K.A., Nelson, J.C. (1997) Map of the synapses onto layer 4 basket cells of the primary visual cortex of the cat. J. Comp. Neurol., 380: 230–242.

Andersen, R.A. (1989) Visual and eye movement functions of posterior parietal cortex. Annu. Rev. Neurosci., 12: 377–403.

Andersen, R.A., Bracewell, R.M., Barash, S., Gnadt, J.W., Fogassi, L. (1990) Eye position effects on visual, memory, and saccade-related activity in areas LIP and 7a of macaque. J. Neurosci., 10: 1176–1198.

Andersen, R.A., Essick, G.K., Siegel, R.M. (1985) Encoding of spatial location by posterior parietal neurons. Science, 230: 450–458.

Andersen, R.A., Mountcastle, V.B. (1983) The influence of the angle of gaze upon the excitability of light-sensitive neurons of the posterior parietal cortex. J. Neurosci., 3: 532–548.

Andersen, R.A., Snyder, L.H., Bradley, D.C., Xing, J. (1997) Multimodal representation of space in the posterior parietal cortex and its use in planning movements. Annu. Rev. Neurosci., 20: 303–330.

Anderson, C.W., Van Essen, D.C. (1987) Shifter circuits: a computational strategy for dynamic aspects of visual processing. Proc. Natl. Acad. Sci. USA, 84: 6297–6301.

Ben-Yishai, R., Bar-Or, R.L., Sompolinsky, H. (1995) Theory of orientation tuning in visual cortex. Proc. Natl. Acad. Sci. USA, 92: 3844–3848.

Boussaoud, D., Jouffrais, C., Bremmer, F. (1998) Eye position effects on the neuronal activity of dorsal premotor cortex in the macaque monkey. J. Neurophysiol., 80:

1132–1150.

Brotchie, P.R., Andersen, R.A., Snyder, L.H., Goodman, S.J. (1995) Head position signals used by parietal neurons to encode locations of visual stimuli. Nature, 375: 232–235.

Buonomano, D.V., Merzenich, M.M. (1998) A neural network model of temporal code-generation and position-invariant pattern recognition. Neural Comput., 11: 103–116.

Chance, F.S., Nelson, S.B., Abbott, L.F. (1999) Compex cells as simple cells at high cortical gain. Nature Neurosci., 2: 277-282.

Connor, C.E., Gallant, J.L., Preddie, D.C., Van Essen, D.C. (1996) Responses in area V4 depend on the spatial relationship between stimulus and attention. J. Neurophysiol., 75, :1306–1308.

Connor, C.E., Preddie, D.C., Gallant, J.L., Van Essen, D.C. (1997) Spatial attention effects in macaque area V4. J. Neurosci., 17: 3201–3214.

Desimone, R. (1991) Face-selective cells in the temporal cortex of monkeys. J. Cognit. Neurosci., 3: 1–8.

Desimone, R., Albright, T.D., Gross, C.G., Bruce, C. (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. J. Neurosci., 4: 2051–2062.

Dill, M., Fahle, M. (1997) The role of visual field position in pattern-discrimination learning. Proc. R. Soc. Lond. B, 264: 1031–1036.

Douglas, R.J., Koch, C., Mahowald, M., Martin, K.A.C., Suarez, H.H. (1995) Recurrent excitation in neocortical circuits. Science, 269: 981–985.

Duhamel, J.R., Bremmer, F., BenHamed, S., Graf, W. (1997) Spatial invariance of visual receptive fields in parietal cortex neurons. Nature, 389: 845–848.

Felleman, D., Van Essen, D.C. (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex, 1: 1–47.

Földiák, P. (1991) Learning invariance from transformed sequences. Neural Comput., 3: 194–200.

Fukushima, K. (1980) Neocognitron: A self-organized neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybern., 36: 193–202.

Gallant, J.L., Braun, J., Van Essen, D.C. (1993) Selectivity for polar, hyperbolic, and cartesian gratings in macaque visual cortex. Science, 259: 100–103.

Gallant, J.L., Connor, C.E., Rakshit, S., Lewis, J.W., Van Essen, D.C. (1996) Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. J. Neurophysiol., 76: 2718–2739.

Galletti, C., Battaglini P.P. (1989) Gaze-dependent visual neurons in area V3A of

monkey prestriate cortex. J. Neurosci., 9: 1112–1125.

Galletti, C., Battaglini, P.P., Fattori, P. (1989) Eye position influence on the parieto-occipital area P.O. (V6) of the macaque monkey. Eur. J. Neurosci., 7: 2486–2501.

Georgopoulos, A.P. (1995) Current issues in directional motor control. Trends Neurosci., 18: 506–510.

Gilbert, C.D., Wiesel, T.N. (1983) Clustered intrinsic connections in cat visual cortex. J. Neurosci., 3: 1116–1133.

Gilbert, C.D., Wiesel, T.N. (1989) Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. J. Neurosci., 9: 2432–2442.

Goodale, M.A., Milner, A.D. (1992) Separate visual pathways for perception and action. Trends Neurosci., 15: 20–25.

Gottlieb, J.P., Kusunoki, M., Goldberg, M.E. (1998) The representation of visual salience in monkey parietal cortex. Nature, 391: 481–484.

Graziano, M.S.A., Hu, T.X., Gross, C.G. (1997) Visuospatial properties of ventral premotor cortex. J. Neurophysiol., 77: 2268–2292.

Gross, C.G. (1992) Representation of visual stimuli in inferior temporal cortex. Philos. Trans. R. Soc. Lond., 335: 3–10.

Guo, K., Li, C.Y. (1997) Eye position-dependent activation of neurones in striate cortex of macaque. NeuroReport, 8: 1405–1409.

Heeger, D.J. (1991) Nonlinear model of neural responses in cat visual cortex. In M. Landy and J.A. Movshon (Eds.), Computational Models of Visual Processing, MIT Press, Cambridge, Masachussetts, pp. 119–133.

Hinton, G.E. (1981a) A parallel computation that assigns canonical object-based frames of reference. In Proceedings of the seventh international joint conference on artificial intelligence, Vol. II, Vancouver, Canada, pp. 683–685.

Hinton, G.E. (1981b) Shape representation in parallel systems. In Proceedings of the seventh international joint conference on artificial intelligence, Vol. II, Vancouver, Canada, pp. 1088–1096.

Karni, A., Sagi, D. (1991) Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. Proc. Natl. Acad. Sci. USA, 88: 4966–4970.

Koch, C., Poggio, T. (1992) Multiplying with synapses and neurons. In T. McKenna, J.L. Davis and S.F. Zornetzer (Eds.), Single Neuron Computation, Academic Press, Cambridge Masachussetts, pp. 315–345.

Logothetis, N.K., Pauls, J., Poggio, T. (1995) Shape representation in the inferior temporal cortex of monkeys. Curr. Biol., 5: 552–563.

Luck, S.J., Chelazzi, L., Hillyard, S.A., Desimone, R. (1997) Neural mechanisms of

spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. J. Neurophysiol., 77: 24–42.

Mack, A., Rock, I. (1998) Inattentional Blindness. MIT Press, Cambridge, Masachussetts.

McAdams, C.J., Maunsell, J.H.R. (1999) Effects of attention on orientation tuning functions of single neurons in macaque cortical area V4. J. Neurosci., 19: 431–441.

Mel, B.W. (1993) Synaptic integration in an excitable dendritic tree. J. Neurophysiol., 70: 1086–1101.

Motter, B C. (1993) Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. J. Neurophysiol., 70: 909-919.

O'Keefe, J., Nadel, L. (1978) The Hippocampus as a Cognitive Map. Oxford University Press, Oxford.

Olshausen, B.A., Anderson, C.H., Van Essen, D.C. (1993) A neurobiological model of visual attention and invariant pattern recognition based on dynamical routing of information. J. Neurosci., 13: 4700–4719.

Pouget, A., Sejnowski, T.J. (1997a) Spatial tranformations in the parietal cortex using basis functions. J. Cog. Neurosci., 9: 222–237.

Pouget, A., Sejnowski, T.J. (1997b) A new view of hemineglect based on the response properties of parietal neurones. Philos. Trans. R. Soc. Lond. B, 352: 1449–1459.

Reynolds, J., Desimone, R. (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. J. Neurosci., 19: 1736–1753.

Riesenhuber, M., Poggio, T. (1998) Just one view: invariances in inferotemporal cell tuning. In M.I. Jordan, M.J. Kearns and S.A. Solla (Eds.), Advances in Neural Information Processing Systems 10, MIT Press, Cambridge, Masachussetts, pp. 215–221.

Riesenhuber, M., Poggio, T. (1999) Hierarchical models of object recognition in cortex. Nature Neurosci., 2: 1019–1025.

Salinas, E., Abbott, L.F. (1995) Transfer of coded information from sensory to motor networks. J. Neurosci., 15: 6461–6474.

Salinas, E., Abbott, L.F. (1996) A model of multiplicative neural responses in parietal cortex. Proc. Natl. Acad. Sci. USA, 93: 11956–11961.

Salinas, E., Abbott, L.F. (1997a) Invariant visual responses from attentional gain fields. J. Neurophysiol., 77: 3267–3272.

Salinas, E., Abbott, L.F. (1997b) Attentional gain modulation as a basis for translation invariance. In J. Bower (Ed.), Computational Neuroscience: Trends in Re-

search 1997, Plenum, New York, pp. 807–812.

Schwartz, E., Desimone, R., Albright, T.D., Gross, C.G. (1983) Shape recognition and inferior temporal neurons. Proc. Natl. Acad. Sci. USA, 80: 5776–5778.

Seung, H.S. (1996) How the brain keeps the eyes still. Proc. Natl. Acad. Sci. USA, 93: 13339–13344.

Sharma, J., Dragoi, V., Miller, E.K., Sur, M. (1999) Modulation of orientation specific responses in monkey V1 by changes in eye position. Soc. Neurosci. Abstr., 25: 677.

Sigman, M., Westheimer, G., Gilbert, C.D. (1999) The role of perceptual learning in object oriented attention. Soc. Neurosci. Abstr., 25: 1049.

Snyder, L.H., Grieve, K.L., Brotchie, P., Andersen, R.A. (1998) Separate body- and world-referenced representations of visual space in parietal cortex. Nature, 394: 887–891.

Somers, D.C., Nelson, S.B., Sur, M. (1995) An emergent model of orientation selectivity in cat visual cortical simple cells. J. Neurosci., 15: 5448–5465.

Tovee, M.J., Rolls, E.T., Azzopardi, P. (1994) Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert macaque. J. Neurophysiol., 72: 1049–1060.

Treue, S., Martínez-Trujillo, J.C. (1999) Feature-based attention influences motion processing gain in macaque visual cortex. Nature, 399: 575–579.

Trotter, Y., Celebrini, S. (1999) Gaze direction controls response gain in primary visual-cortex neurons. Nature, 398: 239–242.

Ungerleider, L.G., Mishkin, M. (1982) The two cortical visual systems. In D.J. Ingle, M.A. Goodale and R.J.W. Mansfield (Eds.), Analysis of Visual Behavior, MIT Press, Cambridge, Masachussetts, pp. 549–586.

Van der Meer, A.L.H., Van der Weel, F.R., Lee, D.N. (1995) The functional significance of arm movements in neonates. Science, 267: 693–695.

Van Opstal, A.J., Hepp, K., Suzuki, Y., Henn, V. (1995) Influence of eye position on activity in monkey superior colliculus. J. Neurophysiol., 74 :1593–1610.

Vidyasagar, T.R. (1998) Gating of neuronal responses in primary visual cortex by an attentional spotlight. NeuroReport, 9: 1947–1952.

Wallis, G., Rolls, E. (1997) Invariant face and object recognition in the visual system. Prog. Neurobiol., 51: 167–194.

Zipser, D., Andersen, R.A. (1988) A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. Nature, 331: 679–684.

Figure 1

Figure 2

Figure 3

a

w_{ij}

0

Receptive field overlap

$r_i = [h^{\mathrm{G}}(y) + h_i^{\mathrm{V}}(x) + \Sigma\, w_{ij}\, r_j - \theta]_+$

b

Firing rate

Stimulus location (x)

External input

Preferred retinal location

Figure 4

Figure 5

a

$r = H(x + y)$

$r = f(x)\ g(y)$

b

Firing rate

Firing rate

Stimulus location (x)

Figure 6

a

b

IT

$R_{IT} = [\sum W_i R_i - \theta]_+$

V4

pixel
array

c

IT response

Image location

0   16   32   48   64

d

Image location

0   16   32   48   64

Figure 7