Invariant Visual Responses From Attentional Gain Fields

EMILIO SALINAS AND L. F. ABBOTT
Volen Center for Complex Systems, Brandeis University, Waltham, Massachusetts 02254-9110

Salinas, Emilio and L. F. Abbott. Invariant visual responses from attentional gain fields. *J. Neurophysiol.* 77: 3267–3272, 1997. Inferotemporal (IT) neurons exhibit a substantial degree of invariance with respect to translation of images used as visual stimuli. Through theoretical and computer-modeling methods, we show how translation-invariant receptive fields, like those of IT neurons, can be generated from the responses of V4 neurons if the effects of attention are taken into account. The model incorporates a recently reported form of attention-dependent gain modulation in V4 and produces IT receptive fields that shift so they are centered at the point where attention is directed. Receptive fields of variable, attention-controlled spatial scale are obtained when the mechanism is extended to scale-dependent attentional gain fields. The results indicate that gain modulation may play analogous roles in the dorsal and ventral visual pathways, generating transformations from retinal coordinates to body- and object-centered systems, respectively.

INTRODUCTION

The ability to recognize an object regardless of the precise location and scale of its retinal image is a striking feature of visual perception. Inferotemporal (IT) neurons in monkeys provide a neuronal correlate of this phenomenon by displaying translation- and scale-invariant responses to complex visual stimuli (Desimone et al. 1984; Hasselmo et al. 1989; Logothetis et al. 1995; Lueschow et al. 1994; Miyashita and Chang 1988; Tovee et al. 1994). Neurons at high levels of the object-recognition pathway of the visual system act as complex filters selective for specific patterns of shape and color (Desimone et al. 1984; Fujita et al. 1992; Gallant et al. 1993; Schwartz et al. 1983). For these cells to exhibit invariant responses, their filters need to be translated from a fixed retinal coordinate frame to a coordinate frame centered on an attended object (Anderson and Van Essen 1987; Hinton 1981a,b; Olshausen et al. 1993). Despite some interesting suggestions (Olshausen et al. 1993), a neuronal mechanism capable of producing this shift has not been verified experimentally.

Lesion studies indicate that area V4 plays an important role in the recognition of visual objects subject to a variety of spatial transformations (Schiller 1995; Schiller and Lee 1991). Attention produces a number of effects in this area (Connor et al. 1996; Desimone and Duncan 1995; Moran and Desimone 1985; Motter 1993). Recent observations (Connor et al. 1996) indicate that the visual responses of many V4 neurons are modulated by a multiplicative gain factor that depends on where attention is being directed. The gain modulation for each cell is maximal when attention is focused on a point that we call the preferred attentional locus, and it decreases when attention moves away from this point (Connor et al. 1996, 1997). Although the neurons were not tested with attention focused directly at the center of their receptive fields, in several cases the responses were shown to increase as attention was directed further away from the receptive field center. Interestingly, the preferred attentional loci were found in directions that appear to be unrelated to the preferred orientations or receptive field locations of the cells and that are uniformly distributed (Connor et al. 1996, 1997). As will be shown below, this surprising feature is the crucial element that allows V4 neurons to generate object-centered receptive fields further down the visual processing stream.

Model

Our model consists of a population of V4 neurons driving a single model IT neuron through feed-forward synaptic connections. In accordance with the data, the firing rates of the model V4 neurons are represented by the product of two terms: the output of a nonlinear filter acting on the luminance distribution of the visual scene and a gain field that depends on the location where attention is being directed. The detailed structure of the V4 receptive fields is not critical for the results, but the model works better when the visual responses are nonlinear in the luminosity, for reasons given below. To satisfy this requirement, visual responses of the model V4 neurons are generated using an ''energy'' model (Heeger 1991, 1992), similar to that used to describe the receptive fields of complex cells in primary visual cortex. The effect of contrast normalization (Carandini and Heeger 1994; Heeger 1991, 1992) is included by dividing all visual responses by the total power present in the image. Receptive field centers for the V4 neurons are distributed uniformly across the visual field. To keep the total number of model cells reasonable, the V4 receptive fields have four orientation and three spatial frequency preferences. The output of the visual filter for cell i is denoted by $F_i(\mathbf{a}_i; I)$, where $\mathbf{a}_i$ is the center of the cell's receptive field and I is the image shown.

The visual responses are multiplied by gain fields that represent the influence of attention. For each neuron, the gain modulation decreases as the actual point where attention is being focused moves away from the preferred attentional locus with the dependence being roughly Gaussian (Connor et al. 1997). In accordance with these results, the gain fields in the model are represented by Gaussian functions, G. The modulatory term for cell i is denoted by $G(\mathbf{y} - \mathbf{b}_i)$, where $\mathbf{y}$ is the currently attended location and $\mathbf{b}_i$ is the preferred attentional locus of cell i. The Gaussian attentional gain fields are approximately twice the size of the visual receptive fields. According to the experimental findings, there is no alignment or correlation between receptive field centers and preferred attentional loci, other than the fact that they are to some degree near to each other. In particular, for a given neuron, the direction that the preferred attentional locus is displaced relative to the receptive field center is random.

The visual field in the model is a pixel grid representing an area of $64 \times 32°$ (32×32 in the case of scaling). An image I, corresponding to a pattern of activated pixels, determines the firing rates in an array of model V4 neurons. The response of cell i is denoted by v_i and is equal to the output of its visual filter times the corresponding modulatory factor

$$v_i = F_i(\mathbf{a}_i; I)G(\mathbf{y} - \mathbf{b}_i) \qquad (1)$$

The response of the single model IT neuron, termed V, is determined by computing a synaptically weighted sum of V4 responses, subtracting a threshold θ, and rectifying the result

$$V = \left[\sum_i W_i v_i - \theta \right]_+ \qquad (2)$$

where $[x]_+$ is equal to 0 if $x < 0$ and equal to x otherwise. In the simulations, the V4-to-IT synapses W_i are established by Hebbian learning (Hebb 1949; Hertz et al. 1991). During a training period, a selected image is displayed and translated across the visual field, the V4 neurons respond according to *Eq. 1*, and the model IT neuron is held in an active state. For each image location, the synaptic weight W_i is increased by an amount proportional to $v_i V$. During learning, the focus of attention is always fixed at the center of the training image. The particular choice of image to be used during training is entirely arbitrary; the model IT neuron becomes selective for the pattern of pixels chosen. We use this correlation-based procedure to generate the connections W_i but, as discussed below, other mechanisms might serve the same purpose equally well. The synaptic weights that give rise to shifting receptive fields are not unique and thus could be generated in a number of ways.

METHODS

Translation invariance

In Fig. 1, images appear on a 64×32 pixel array (1 pixel = $1°$), and receptive field centers $\mathbf{a}_i$ are distributed uniformly on a 32×16 grid, separated by 2 pixels in each direction. For each location, there are neurons with four orientation preferences, 0, 45, 90, and 135°, and three frequency selectivities, 1/8, 2/8 and 3/8 cycles per degree. Complex-cell-like responses $F_i(\mathbf{a}_i; I)$ are generated using an energy model (Heeger 1991, 1992) by adding the squared outputs of two linearly filtered versions of the image I

$$F_i(\mathbf{a}_i; I) = [S_i(\mathbf{a}_i; I)]^2 + [C_i(\mathbf{a}_i; I)]^2 \qquad (3)$$

S_i and C_i stand for the outputs of localized sine and cosine linear filters, i.e.,

$$S_i(\mathbf{a}_i; I) = \int d\mathbf{x}\, I(\mathbf{x}) f_i^S(\mathbf{x} - \mathbf{a}_i)$$

$$C_i(\mathbf{a}_i; I) = \int d\mathbf{x}\, I(\mathbf{x}) f_i^C(\mathbf{x} - \mathbf{a}_i) \qquad (4)$$

The linear filters are similar to Gabor functions (Field and Tolhurst 1986; Jones and Palmer 1987) except that, for reasons of computational efficiency, half-cosine envelopes (rather than Gaussian) are used

$$f_i^S(\mathbf{x}) = \text{hcos}(x_1/\lambda)\text{hcos}(x_2/\lambda)\sin\{k_i[\cos(\phi_i)x_1 + \sin(\phi_i)x_2]\}$$

$$f_i^C(\mathbf{x}) = \text{hcos}(x_1/\lambda)\text{hcos}(x_2/\lambda)$$
$$\times \cos\{k_i[\cos(\phi_i)x_1 + \sin(\phi_i)x_2]\} \qquad (5)$$

The position vector $\mathbf{x}$ has components (x_1, x_2), and the half-cosine

function hcos(x) is equal to $\cos(x)$ if $-\pi/2 < x < \pi/2$ and to 0 otherwise. Here ϕ_i and k_i determine the preferred orientation and spatial frequency, respectively; λ determines the receptive field width at baseline, which is $4°$ ($= 4$ pixels) for all cells. Preferred attentional loci are located at 24 positions uniformly spaced throughout the 64 pixels in the x direction. Each visual filter output $F_i(\mathbf{a}_i; I)$ is combined with those preferred attentional loci within 8 pixels from the receptive field center $\mathbf{a}_i$, producing six or seven attention-modulated responses (originally we included all 24 combinations of preferred attentional loci for each visual filter output, but we found that only the 6 or 7 nearest the receptive field center actually were needed). The Gaussian attentional gain fields have a standard deviation $\sigma = 2°$, and therefore a baseline width of $\sim 4\sigma = 8°$. The result is a total of $32 \times 16 \times 4 \times 3 \times 6$ V4 responses. The threshold θ in *Eq. 2* serves to enhance the selectivity of the model IT neuron by eliminating the lower, typically broader, part of its response curve. It is set to 50% of the maximum response obtained when $\theta = 0$.

Scale invariance

In Fig. 2, images appear on a 32×32 pixel array, and receptive field centers are arranged uniformly on a 16×16 grid. The same variety of orientations and spatial frequencies as in Fig. 1 is used. Unlike Fig. 1, the visual responses are modeled as rectified linear filters. Two kinds of visual filters are considered

$$F_i^S(\mathbf{a}_i; I) = [S_i(\mathbf{a}_i; I)]_+ \quad F_i^C(\mathbf{a}_i; I) = [C_i(\mathbf{a}_i; I)]_+ \qquad (6)$$

where the brackets again indicate rectification. To model the modulation produced by the attended scale, each neuron is assigned a preferred attentional scale, analogous to the preferred attentional locus in the case of modulation by an attended location. The gain modulation is a Gaussian function of the difference between the current attended scale and the preferred attended scale of each neuron. A set of 20 preferred attended scales, varying from 3 to 30°, is used to modulate the visual responses; the standard deviation of the Gaussian gain field is $\sigma = 1°$. A total of $16 \times 16 \times 4 \times 3 \times 2 \times 20$ gain-modulated V4 responses are used. The threshold θ is set in this case to 40% of the maximum response obtained when $\theta = 0$ (this value is slightly smaller than in the case of translation, so the resulting response curves are not excessively narrow).

RESULTS

In the computer simulations, an image is shown at a particular location, the model V4 neurons respond according to *Eq. 1* and drive the model IT neuron as specified by *Eq. 2*. The synaptic connections are established first by translating the training image and enabling the Hebbian synaptic modification process described above. After training, the synaptic weights are not modified any more and the model then is tested. During training, the value of $\mathbf{y}$, corresponding to the position of the attentional locus, is equal to the position of the image presented; during testing, it is set to a variety of fixed locations. The results plotted in the figures show IT responses during the testing phase.

Figure 1 shows the results when the letter R was used as the training image. The model IT neuron is selective for this shape, firing at a maximum rate when the R is centered within its receptive field (*a, top*). A different letter, or a degraded version of the R, evokes less rapid firing (*c* and *e*). To test for receptive field translation, the locus of attention $\mathbf{y}$ was moved (*a, bottom*). The similar responses at the *top* and *bottom* of *a* and the results of *b* show that the receptive field shifts with attention. The firing rate varies with the type of image
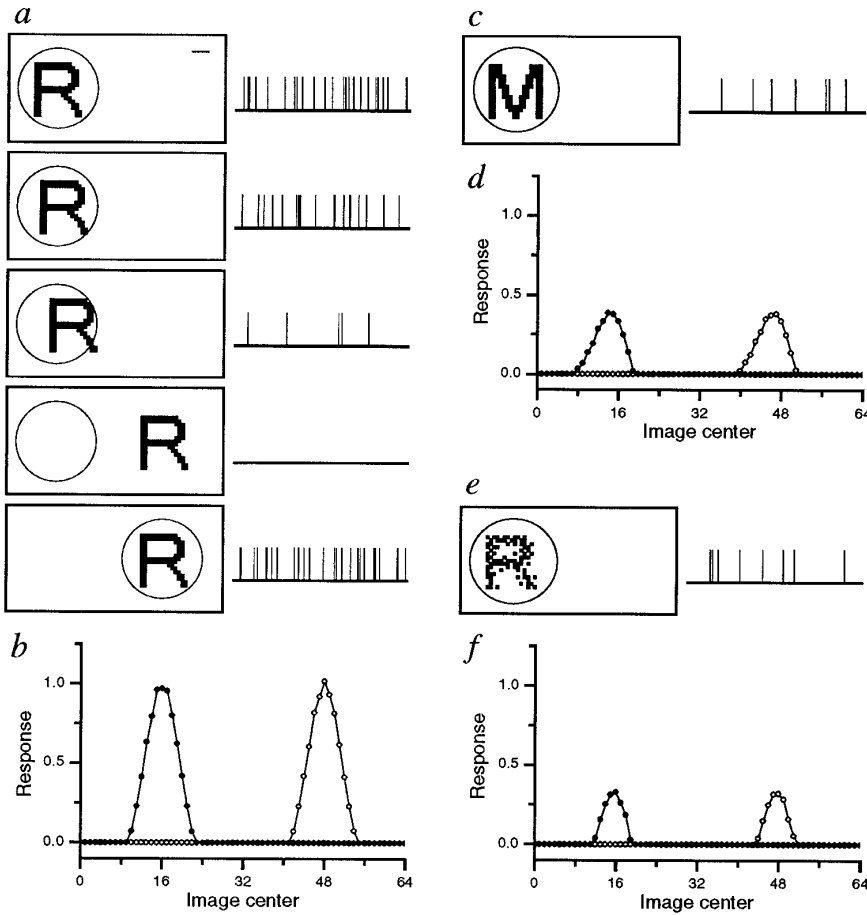
FIG. 1. Computer simulation of model network for images translated across visual field. Center of circle indicates locus of attention. *a*: inferotemporal (IT) responses to translated versions of a letter R, which was presented previously during learning. Spike trains on *right* are generated for visualization purposes using a Poisson process based on mean firing rate of model IT cell. IT neuron response depends on location of R relative to current attentional locus. Examples are shown with attention focused at pixels 16 and 48. Scale bar is 4 pixels. *b*: IT response (normalized mean firing rate) plotted as a function of image location. As in *d* and *f*, filled circles indicate attention centered at pixel 16; open circles indicate attention at pixel 48. *c* and *d*: responses of same model neuron to an M. *e* and *f*: responses to a degraded version of R. In all cases, IT receptive field moves with attention.

presented and with its location, but the neuronal response depends on the position of the image relative to the locus of attention, not on its absolute location. This is shown for two different attentional loci in Fig. 1, *a* and *b*, but is true for attention focused at any point in the visual field. Therefore, the neuron selectively reacts to an image in an attention-centered coordinate system. Equivalent results are obtained when other images are used during training: in all cases the IT neuron becomes selective for the training image, firing at higher rates than when other test images are shown and keeping its receptive field center in register with the attentional locus.

The model gives rise to translating receptive fields because collections of V4 neurons with similar preferred attentional loci act as separate pools to construct local IT filters centered at different locations. The modulatory gain fields select pools acting near the point of attention, interpolate seamlessly among them, and suppress irrelevant pools acting far from the attentional locus. The result is that the IT receptive fields filter the luminance distribution relative to the locus of attention, not to any fixed retinal location.

Analytic work supports the results shown and can provide some intuition into the mechanism at work in the model. The crucial elements in *Eq. 2* are the synaptic weights W_i, defined as the strength of the synapse connecting the model IT neuron to V4 neuron i. The requirement for attention to produce shifting receptive fields is that the weights W_i depend on the receptive field centers $\mathbf{a}_i$ and preferred attentional loci $\mathbf{b}_i$ only through their difference, i.e.

$$W_i = W(\mathbf{a}_i - \mathbf{b}_i) \tag{7}$$

This expression indicates that the synaptic weight from a particular V4 cell depends on the displacement between its preferred attentional locus and receptive field center but not on these two locations independently. It also implies that, viewed as functions of $\mathbf{a}_i$, the synaptic weights for two groups of neurons with different $\mathbf{b}_i$ are translated versions of one another. The weights also may depend on other parameters, such as preferred orientation, and no constraints are placed on those additional dependencies. For simplicity, we will ignore these additional dependencies in the following analysis. If condition *7* is satisfied, it can be shown, under fairly general assumptions, that gain modulation gives rise to shifting receptive fields. For clarity, we consider the simple case in which the visual responses are given by linear filters acting on the image I

$$F(\mathbf{a}_i; I) = \int d\mathbf{x}\, I(\mathbf{x}) f(\mathbf{x} - \mathbf{a}_i) \tag{8}$$

However, it should be stressed that nothing restricts the analysis to this case; similar results can be derived for nonlinear filters.

To proceed further with this analytic approach, we must assume that preferred attentional loci corresponding to a given receptive field placement are uniformly distributed over the entire visual field, something not seen in the data. However, in computer simulations, we have found that neurons with attentional loci that are far away from the corresponding receptive field centers have a negligible impact. Thus this assumption can be relaxed without changing the performance of the model. To see that the IT response shifts
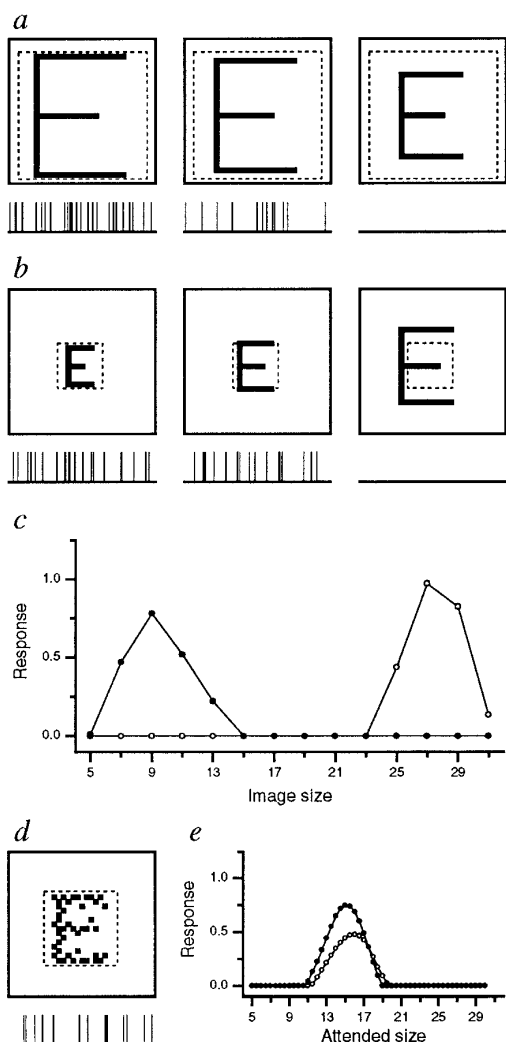
FIG. 2. Computer simulation of model network for images shown at different scales. Letters E, previously presented during learning, are shown. *a*: IT responses to images of sizes 27, 25, and 19 pixels when attended scale is set at 27 pixels (− − −). Spike traces are produced for visualization purposes using a Poisson process based on resulting IT firing rates. *b*: responses when attended scale is equal to 9 pixels, for images of sizes 9, 11, and 17 pixels. In both *a* and *b*, neuron responds strongly when attended scale closely matches size of image. *c*: mean normalized IT response plotted as a function of image size. Filled circles, attended scale of 9 pixels; open circles, attended scale of 27 pixels. *d*: degraded version of an E and neural response when attended scale is 15 pixels. *e*: mean normalized IT response vs. attended size. Filled circles, responses to original E of size 15 pixels; open circles, responses to degraded E shown in *d*.

with attention, all that is needed is to substitute expression *8* into *Eq. 1* and approximate the sum over cells in *Eq. 2* by an integral over their labels, assuming uniformity, high density, and independence. The synaptically weighted sum then becomes

$$\sum_i W_i v_i \propto \int d\mathbf{x} d\mathbf{a} d\mathbf{b}\; W(\mathbf{a} - \mathbf{b}) G(\mathbf{y} - \mathbf{b}) I(\mathbf{x}) f(\mathbf{x} - \mathbf{a}) \qquad (9)$$

Making the substitutions $\mathbf{a} \rightarrow \mathbf{a} + \mathbf{y}$ and $\mathbf{b} \rightarrow \mathbf{b} + \mathbf{y}$, the integral takes the following form

$$\int d\mathbf{x}\, I(\mathbf{x}) \mathcal{F}(\mathbf{x} - \mathbf{y}) \qquad (10)$$

with

$$\mathcal{F}(\mathbf{x} - \mathbf{y}) = \int d\mathbf{a} d\mathbf{b}\; W(\mathbf{a} - \mathbf{b}) G(\mathbf{b}) f(\mathbf{x} - \mathbf{y} - \mathbf{a}) \qquad (11)$$

Equation 10 is precisely a filtered version of *I* that shifts with the locus of attention, **y**. Thus the receptive field of the IT neuron, determined by the resulting filter $\mathcal{F}$, will move with the attentional locus. The simulations confirm this result because the simple Hebbian synaptic modification scheme used produces synaptic weights that satisfy *Eq. 7*; this too can be shown analytically (for a related example see Salinas and Abbott 1995). The particular values of the synaptic weights determine the precise form of the final shifting filter $\mathcal{F}$. This is not limited significantly by *Eq. 7* because the single-variable function on the right side of *Eq. 7* is entirely arbitrary. Furthermore, sets of weights W_i and w_i projecting to two different IT neurons can satisfy simultaneously the conditions $W_i = W_i(\mathbf{a}_i - \mathbf{b}_i)$ and $w_i = w_i(\mathbf{a}_i - \mathbf{b}_i)$ and still be completely different from each other. Thus the same array of gain-modulated V4 neurons can serve as a basis for multiple, arbitrary shifting filters.

The same mechanism that we have described for generating shifting receptive fields also can produce receptive fields that are scaled to an image size specified by an attentional signal. This requires gain fields that depend on an attended scale. In this case, we model the gain field for each neuron as a Gaussian function of the difference between the attended scale and a neuron-specific preferred attentional scale. An analogous Hebbian mechanism is used to establish the synaptic weights. During learning, an image is presented at a variety of sizes while the attended scale is set to the size of each image, the model IT neuron is held in an active state, and the synaptic connections change by an amount proportional to pre- and postsynaptic activity. In this case, images are presented always at the same position (just as size was kept constant in the case of translation). After training, the model then is tested by computing the IT response evoked by different images. Figure 2 shows the results of a computer simulation in which letters E of different sizes were used during training. In this case, the resulting IT response depends on the match between the size of the presented image and the attended scale, but not on the absolute size of the image (Fig. 2, *a–c*). The responses are also selective for the image used during training, as shown in *e*; the degraded E elicits a weaker response than the original E used during training. Interestingly, this graph reveals that the optimal attended scale for the degraded E is slightly bigger than for the original E, consistent with the fact that the former is effectively one pixel wider than the latter.

DISCUSSION

There are two costs associated with a gain modulation mechanism for producing object-centered receptive fields. First, there is some loss of resolution in the relative placement of the different V4 filters because the synaptically weighted sum that determines the IT neuron response acts effectively as a convolution over the gain field profile (see *Eq. 9*). However, analytic calculations show that the attentional gain field causes no loss of resolution for features within the receptive field of a given V4 cell, provided that the visual filter is a nonlinear functional of the luminance distribution. Indeed we found that the simulated IT responses are more selective when the V4 neurons are modeled as nonlinear filters (like, for example, those of complex cells) than as linear filters. Nevertheless, not all nonlinearities are

equally resistant to the ''smearing'' caused by the convolution over the gain field profile. In the case of translation, the complex-cell-like responses used to generate Fig. 1 (*Eq. 3*) result in a more pronounced IT selectivity than the simple-cell-like filters of *Eq. 6*, although, because of rectification, these too are nonlinear. The opposite happens in the case of scaling; the rectified linear filters produce IT receptive fields that are more selective than those obtained through the energy model. Thus each invariance is best achieved using a particular type of matched nonlinearity.

Second, the number of V4 cells needed to cover the visual field with both receptive fields and attention gain fields is greater than the number required without attentional modulation. We estimate this redundancy factor at somewhere between 20 and 100. In the simulations, only six attentional loci near to a given receptive field center were needed to achieve full translation invariance; adding more loci had no effect on the results. The exact number required depends on the size of the image that needs to be translated (images used were 16° wide). If a factor of 6 corresponds to translation along a single dimension, a factor of 36 would be needed for two dimensions (scale invariance would require an additional factor between 10 and 50). The actual redundancy factor required may be higher, because of effects that are not included in the model: not all V4 cells are equally modulated (Connor et al. 1996) and IT neurons also show some degree of rotation and perspective invariance (Logothetis et al. 1995). The combinatorial growth could require attentional modulation acting through successive stages in the ventral visual pathway, such that a sequential transformation gradually accumulates. There is some evidence that attentional effects are present in early visual cortical areas (Moran and Desimone 1985; Motter 1993). B. Olshausen has pointed out (personal communication) that the modest shifting effect seen in V4 neurons (Connor et al. 1996) could be due to attentional gain modulation acting at visual stages before V4.

The gain modulation mechanism has the outstanding advantage that IT neurons with complex and specialized selectivities do not have to be duplicated across the visual field, because they can be shifted to the location where they are needed. Although we considered only a single model IT neuron, the same set of V4 neurons can project to other neurons that respond selectively to different images. Attentional gain modulation in V4 then will cause all of the different IT receptive fields to shift with attention. The price paid is the large number of V4 neurons required, but these have much simpler receptive fields and, once generated, can serve as a basis for an arbitrary set of highly selective receptive fields that then will be shifted by attention.

A related point concerns the number of synapses used in the model. The model IT neuron potentially could make connections with all the neurons in the V4 array, ~40,000 of them. However, analysis of the weights produced by the computer simulations showed that, for the parameters used in Fig. 1, many of them were essentially 0 (i.e., <5% of the maximum weight). The reason is that, for a given image and for attention focused at a given point, only a few V4 cells respond strongly, namely those whose filters overlap significantly with the image and whose preferred attentional loci are close to the current attentional locus. Other simulations were done in which, after the learning period, only those weights equal to or greater than some cutoff fraction

of the maximum weight were kept, the rest being set to 0. In simulations analogous to those shown in Fig. 1, a cutoff equal to one-half the maximum weight eliminated all but ~3,300 connections. Rather than interfering with the shifting effect, or distorting the shape of the model IT tuning curve, this manipulation noticeably increased the selectivity of the IT neuron, leaving the shifting effect intact. These results indicate that, in the model, most of the highly selective part of the IT response is determined by relatively few V4 neurons. They also suggest that synaptic pruning might act as an effective mechanism to enhance the selectivity of neural responses.

Visual neurons with gain fields that depend on the location where attention is being focused thus can form an effective basis for receptive fields that shift across the retina. Similarly, visual responses that are gain modulated by the scale that is being attended could serve to generate receptive fields that zoom into or out of a region. Like others (Hinton 1981a,b; Olshausen et al. 1993), we envision that the responses of high-level visual neurons are fed back to guide the attentional signal, so that receptive fields are scaled accurately and centered on objects that produce robust responses. The mechanism described here is distinct from previous models that achieve translation invariance either through multilayered connectionist architectures engineered to produce ''grandmother-cell''-like responses (Fukushima 1980) or by specifying a hypothesized learning or recall dynamics at single synapses (Anderson and Van Essen 1987; Földiák 1991; Hinton 1981a,b; Olshausen et al. 1993; Wallis 1994). The present model exploits the mechanism of gain modulation within a neuronal array in a way that is consistent with reported observations (Connor et al. 1996, 1997) and places a much looser constraint on the individual synapses. Our model is related closely to ideas developed during the study of parietal cortex, where gaze-direction-dependent gain modulation of visual responses has been reported (Andersen et al. 1985, 1990; Brotchie et al. 1995). Theoretical work (Andersen et al. 1990, 1993; Pouget and Sejnowski 1995, 1996; Salinas and Abbott 1995; Zipser and Andersen 1988) suggests that gain-modulated parietal activity forms the basis for transformations from retinal to body-centered coordinates useful in visually guided motor tasks. We propose here that a similar mechanism acts to transform images from a retinal basis to an object-centered form useful for invariant perception. Thus gain modulation may be used in a similar manner to perform coordinate transformations in both the dorsal-''where'' and the ventral-''what'' visual pathways.

REFERENCES

ANDERSEN, R. A., BRACEWELL, R. M., BARASH, S., GNADT, J. W., AND FOGASSI, L. Eye position effects on visual, memory, and saccade-related activity in areas LIP and 7a of macaque. *J. Neurosci.* 10: 1176–1198, 1990.

ANDERSEN, R. A., ESSICK, G. K., AND SIEGEL, R. M. Encoding of spatial location by posterior parietal neurons. *Science Wash. DC* 230: 450–458, 1985.

ANDERSEN, R. A., SNYDER, L. H., LI, C.-S., AND STRICANNE, B. Coordinate transformations in the representation of spatial information. *Curr. Opin. Biol.* 3: 171–176, 1993.

ANDERSON, C. W. AND VAN ESSEN, D. C. Shifter circuits: a computational strategy for dynamic aspects of visual processing. *Proc. Natl. Acad. Sci. USA* 84: 6297–6301, 1987.

BROTCHIE, P. R., ANDERSEN, R. A., SNYDER, L. H., AND GOODMAN, S. J. Head position signals used by parietal neurons to encode locations of visual stimuli. *Nature Lond.* 375: 232–235, 1995.

CARANDINI, M. AND HEEGER, D. J. Summation and division by neurons in primate visual cortex. *Science Wash. DC* 264: 1333–1336, 1994.

CONNOR, C. E., GALLANT, J. L., PREDDIE, D. C., AND VAN ESSEN, D. C. Responses in area V4 depend on the spatial relationship between stimulus and attention. *J. Neurophysiol.* 75: 1306–1308, 1996.

CONNOR, C. E., PREDDIE, D. C., GALLANT, J. L., AND VAN ESSEN, D. C. Spatial attention effects in macaque area V4. *J. Neurosci.* 17: 3201–3214, 1997.

DESIMONE, R., ALBRIGHT, T. D., GROSS, C. G., AND BRUCE C. Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.* 4: 2051–2062, 1984.

DESIMONE, R. AND DUNCAN, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18: 193–222, 1995.

FIELD, D. AND TOLHURST, D. J. The structure and symmetry of simple-cell receptive-field profiles in the cat's visual cortex. *Proc. R. Soc. Lond. B Biol. Sci.* 228: 379–400, 1986.

FÖLDIÁK, P. Learning invariance from transformed sequences. *Neural Comput.* 3: 194–200, 1991.

FUJITA, I., TANAKA, K., ITO, M., AND CHENG, K. Columns for visual features of objects in monkey inferotemporal cortex. *Nature Lond.* 360: 343–346, 1992.

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

GALLANT, J. L., BRAUN, J., AND VAN ESSEN, D. C. Selectivity for polar, hyperbolic, and cartesian gratings in macaque visual cortex. *Science Wash. DC* 259: 100–103, 1993.

HASSELMO, M. E., ROLLS, E. T., BAYLIS, G. C., AND NALWA, V. Object-centered encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey. *Exp. Brain Res.* 75: 417–429, 1989.

HEBB, D. O. *The Organization of Behavior: A Neuropsychological Theory.* New York: Wiley, 1949.

HEEGER, D. J. Nonlinear model of neural responses in cat visual cortex. In *Computational Models of Visual Processing*, edited by M. Landy and J. A. Movshon. Cambridge, MA: MIT Press, 1991, p. 119–133.

HEEGER, D. J. Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* 9: 181–197, 1992.

HERTZ, J., KROGH, A., AND PALMER, R. G. *Introduction to the Theory of Neural Computation.* New York: Addison-Wesley, 1991.

HINTON, G. E. A parallel computation that assigns canonical object-based frames of reference. In: *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*. Vancouver, Canada: 1981a, vol. 2, p. 683–685.

HINTON, G. E. Shape recognition and illusory conjunctions. In: *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*. Vancouver, Canada: 1981b, vol. 2, p. 1088–1096.

JONES, J. P. AND PALMER, L. A. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiol.* 58: 1233–1259, 1987.

LOGOTHETIS, N. K., PAULS, J., AND POGGIO, T. Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* 5: 552–563, 1995.

LUESCHOW, A., MILLER, E. K., AND DESIMONE, R. Inferior temporal mechanisms for invariant recognition. *Cereb. Cortex* 5: 523–531, 1994.

MIYASHITA, Y. AND CHANG, H. S. Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature Lond.* 331: 68–70, 1988.

MORAN, J. AND DESIMONE, R. Selective attention gates visual processing in the extrastriate cortex. *Science Wash. DC* 229: 782–784, 1985.

MOTTER, B. C. Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J. Neurophysiol.* 70: 1–11, 1993.

OLSHAUSEN, B. A., ANDERSON, C. H., AND VAN ESSEN, D. C. A neurobiological model of visual attention and invariant pattern recognition based on dynamical routing of information. *J. Neurosci.* 13: 4700–4719, 1993.

POUGET, A. AND SEJNOWSKI, T. J. Spatial representations in the parietal cortex may use basis functions. In: *Advances in Neural Information Processing Systems 7*, edited by G. Tesauro, D. S. Touretzky, and T. K. Lean. Cambridge, MA: MIT Press, 1995.

POUGET, A. AND SEJNOWSKI, T. J. Spatial transformation in the parietal cortex using basis functions. *J. Cogn. Neurosci.* In press.

SALINAS, E. AND ABBOTT, L. F. Transfer of coded information from sensory to motor networks. *J. Neurosci.* 15: 6461–6474, 1995.

SCHILLER, P. H. AND LEE, K. The role of the primate extrastriate area V4 in vision. *Science Wash. DC* 251: 1251–1253, 1991.

SCHILLER, P. H. Effect of lesions in visual cortical area V4 on the recognition of transformed objects. *Nature Lond.* 376: 342–344, 1995.

SCHWARTZ, E. L., DESIMONE, R., ALBRIGHT, T. D., AND GROSS, C. G. Shape recognition and inferior temporal neurons. *Proc. Natl. Acad. Sci. USA* 80: 5776–5778, 1983.

TOVEE, M. J., ROLLS, E. T., AND AZZOPARDI, P. 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, 1994.

WALLIS, G. *Neural Mechanisms Underlying Processing in the Visual Areas of the Occipital and Temporal Lobes* (PhD thesis). Cambridge: Oxford University, 1994.

ZIPSER, D. AND ANDERSEN, R. A. A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature Lond.* 331: 679–684, 1988.