Journal of Computational Neuroscience 1, 141–158 (1994).

A Model for the Neuronal Implementation of Selective Visual Attention Based on Temporal Correlation Among Neurons

ERNST NIEBUR* ERNST@CALTECH.EDU.

CHRISTOF KOCH KOCH@KLAB.CALTECH.EDU.

Computation and Neural Systems Program, California Institute of Technology, Pasadena CA 91125, USA

Received November 2, 1993; Revised March 10, 1994; Accepted (in revised form) March 24, 1994

Action Editor: R. Andersen

Abstract. We propose a model for the neuronal implementation of selective visual attention based on temporal correlation among groups of neurons. Neurons in primary visual cortex respond to visual stimuli with a Poisson distributed spike train with an appropriate, stimulus-dependent mean firing rate. The spike trains of neurons whose receptive fields do *not* overlap with the "focus of attention" are distributed according to homogeneous (time-independent) Poisson process with no correlation between action potentials of different neurons. In contrast, spike trains of neurons with receptive fields within the focus of attention are distributed according to non-homogeneous (time-dependent) Poisson processes. Since the short-term average spike rates of all neurons with receptive fields in the focus of attention covary, correlations between these spike trains are introduced which are detected by inhibitory interneurons in V4. These cells, modeled as modified integrate-and-fire neurons, function as coincidence detectors and suppress the response of V4 cells associated with non-attended visual stimuli. The model reproduces quantitatively experimental data obtained in cortical area V4 of monkey by Moran and Desimone (1985).

1 Introduction

Selective attention is the process whereby a particular piece of information is selected from a sensory array for further processing, in particular for recognition or mnemonic tasks. Visual attention usually manifests itself within a single, circumscribed spatial area that can vary in size and that scans objects in the visual field at a rate of about 30–50 msec per object. Visual attention has been postulated to operate by either suppressing all distracting information, by enhancing the selected (attended) information, or by both. At the neuronal level, the cortical areas thought to be involved in the control and expression of attention include the object recognition pathway, in particular extrastriate areas V4 and IT, as well as the posterior parietal cortex and parts of prefrontal cortex. Among extra-cortical sites, the pulvinar nuclei of the thalamus and the superior colliculus in the midbrain have been implicated in the control of attention (for reviews see Treisman, 1988, Braun and Sagi, 1990, Julesz, 1991, Colby, 1991, Posner and Petersen, 1990, Kanwisher and Driver, 1992, Posner and Driver, 1992).

Behaviorally refined selective attention can be observed and quantified at the level of single neurons. Motter (1993) recorded in cortical areas V1, V2 and V4 of highly-trained macaque monkeys while they were attending one of several stimuli in their visual field. All stimuli except one were located outside the receptive field of the cell whose activity was recorded. It was found that the average firing rate of the cell depended on whether the animal directed its attention to this one stimulus or to other stimuli. In the early cortical areas V1 and V2, the firing rate of about one third of the ob-

*To whom correspondence should be addressed at Caltech, 139-74, Pasadena, CA 91125, USA. Phone: (818) 395-2880, fax: (818)796-8876. e-mail: ernst@caltech.edu

served cells was modified—either suppressed or enhanced—under the influence of focal attention. This effect occurred only for stimuli close to the cell's preferred orientation, and usually (in about 90% of all cells which showed a statistically significant change caused by attentional modulation) only in the presence of competing stimuli outside the receptive field. In contrast, modulation of cell responses in extrastriate area V4 was not restricted to stimuli close to the preferred orientation of the cells. Furthermore, attentional modulation was reliably observed in V4 not only for stimulus configurations comprising arrays of several stimuli in the surround of the receptive field but also for isolated stimuli (Motter, 1993).

While in these experiments all stimuli but one were placed outside the (classical) receptive field of the recorded cell, the functional structure of the attentional mechanism has been further investigated by another series of experiments in which more than one stimulus was located in each receptive field (Moran and Desimone, 1985, Desimone et al., 1991). When two different objects, say a red and a green bar, are both located within the receptive field of a V4 neuron selective for red, the neuron will respond vigorously if the monkey attends to the red stimulus, but respond much less if the monkey is attending to the green stimulus. Functionally, the receptive field has shrunk around the attended stimulus.

In the present work, we use computer simulations to study a model of the physiological implementation of selective visual attention compatible with the above mentioned experimental results. Our model is based on a hypothesis by Crick and Koch who proposed that selective visual attention manifests itself at the single cell level via "temporal tagging" (Crick and Koch, 1990a,b). It is assumed that different features of a perceived object are represented by correlated activity of neurons in different cortical and subcortical areas (von der Malsburg, 1981, von der Malsburg and Schneider, 1986, Crick and Koch, 1990a). It is this correlated activity which provides the system (i.e., the observer) with a unitary percept of the object across feature dimensions and modalities. It is commonly assumed that selective attention restricts the access of sensory input to conscious perception and working memory, and we will make the assumption that at any given time, only one stimulus is chosen for such access. This selection is performed in our model by a "saliency map" à la Koch and Ullman (1985; see also the "master map" of Treisman, 1988) which encodes information on *where* salient (conspicuous) objects are located in the visual field, but not *what* these objects are. After a short time, the location of the presently most salient object becomes inhibited in the saliency map and attention switches to the next most conspicuous location (we will not discuss the interior dynamics or mechanism of the saliency map in this report; see Koch and Ullman, 1985 for this). In the present work, we assume that salient objects have been selected in the visual field by such a mechanism and that they are "tagged" by modulating the temporal structure of the neuronal signals corresponding to attended stimuli.

Possible anatomical sites for this topographic map include the superior colliculus, the posterior parietal cortex, and the dorsomedial region of the pulvinar. The superior colliculus is intimately involved in the control of overt attention, i.e., saccadic eye movements. The response of cells in posterior parietal cortex can be modulated by attention factors, and this modulation can be completely dissociated from eye movements (Mountcastle et al., 1981, Colby, 1991). There is also strong experimental evidence linking the pulvinar to the control of selective attention. The pulvinar is phylogenetically the most recent nucleus of the dorsal thalamus. It may still be absent in rodents, it is relatively small in carnivores, and it increases in size progressively from monkeys to apes, until in man it reaches enormous proportions (several times the size of, e.g., the much more studied lateral geniculate nucleus). Among the six pulvinar nuclei whose function are strongly related to visual processing, three are organized retinotopically. Petersen, Robinson and Morris (1987) reported that reversible chemical de-activation of the dorsomedial portion of the pulvinar in monkeys causes an increased reaction time in switching attention from the ipsilateral to contralateral vi-

sual field. Studies of humans with thalamic lesion involving the pulvinar have shown impairments in the ability to engage attention (Rafal and Posner, 1987). Positron emission tomography studies in humans have shown increased activation of the pulvinar in attentional tasks (LaBerge and Buchsbaum, 1990). Single cell recordings showed that pulvinar neurons relate to the saliency of visual objects (Robinson and Petersen, 1992).

How can a signal from a saliency map be used to lead to the changes in V4 receptive fields observed by Desimone and colleagues? According to an idea expressed by Desimone (1992), competition between cells in area V4 could be biased in favor of cells representing attended stimuli. Following Crick and Koch (1990b), our basic hypothesis is that this is accomplished using different temporal structures (but identical average spike rates) of the spike trains generated by neurons presynaptic to V4 inside and outside the focus of attention. No change of the average firing rate of these neurons is needed; attentional modulation by the saliency map influences the timing of spikes so that V2 neurons within the focus of attention tend to fire together (within a few milliseconds) more frequently than neurons outside the focus of attention. In contrast, at the level of V4 and beyond, signals along the tagged pathway compete with signals in the untagged pathway, leading to an inhibition or reduction in the response of neurons in the untagged pathway.

2 The Model

The gross architecture of our model is shown in Figure 1A. We here simulate two interacting cortical areas in the hierarchy of visual cortical processing. We will first assume that these areas are among the first (i.e., closest to the periphery) of those areas which have been shown to be subjected to attentional influence. As was discussed previously, activity in area V4 is strongly and systematically modified by selective attention. Although this area receives some direct input from V1, this projection is weak and limited to the central field representation (Nakamura et al., 1993), and V4 receives its major input from V2. Therefore, we will label the first area modelled in this work as V2, and the area it projects to as V4. We emphasize, however, that the model does not rely on details of the anatomy of these areas, and we will, in fact, argue later on (see Figure 4) that our model can be "cascaded" over more than one stage, the output area of one stage becoming the input area of the next stage.

Input from the two-dimensional retina is fed via the lateral geniculate nucleus (LGN) and V1 (both not shown) into area V2, where the attentional modulation – originating in the saliency map – is added. We assume that cells in V2 are only selective to one of two different features, e.g., the colors red or green (our results do not depend on this simplification). The output of our model V2 projects into neuronal "stacks" (see below) in our model V4, where it excites pyramidal cells as well as inhibitory interneurons. These V4 interneurons, in turn, inhibit the pyramidal cells of opposing feature selectivity (Fig. 1B).

For the sake of simplicity of language, we will sometimes refer to neurons whose receptive fields are inside the focus of attention as "attended neurons".

2.1 Circumstriate Cortex (V2)

The visual input into each V2 cell is provided by a 10 by 10 array of pixels. The output of any V2 cell are pulses, generated using a Poisson process of mean firing rate λ in combination with a refractory period, which is chosen randomly from a uniform distribution with values between 2 msec and 5 msec. Bair, Koch, Newsome and Britten (1993) analyzing 216 cells recorded from extrastriate area MT in the awake and behaving monkey, showed that the stochastic properties of about one third of these cells – firing at high discharge frequencies – can best be described on the basis of a Poisson process in combination with a short refractory period (see also Softky and Koch, 1993).

The total firing rate λ of any neuron is the sum of the spontaneous firing rate λ_{spont}, and the stimulus-dependent rate λ_0. If no stimulus

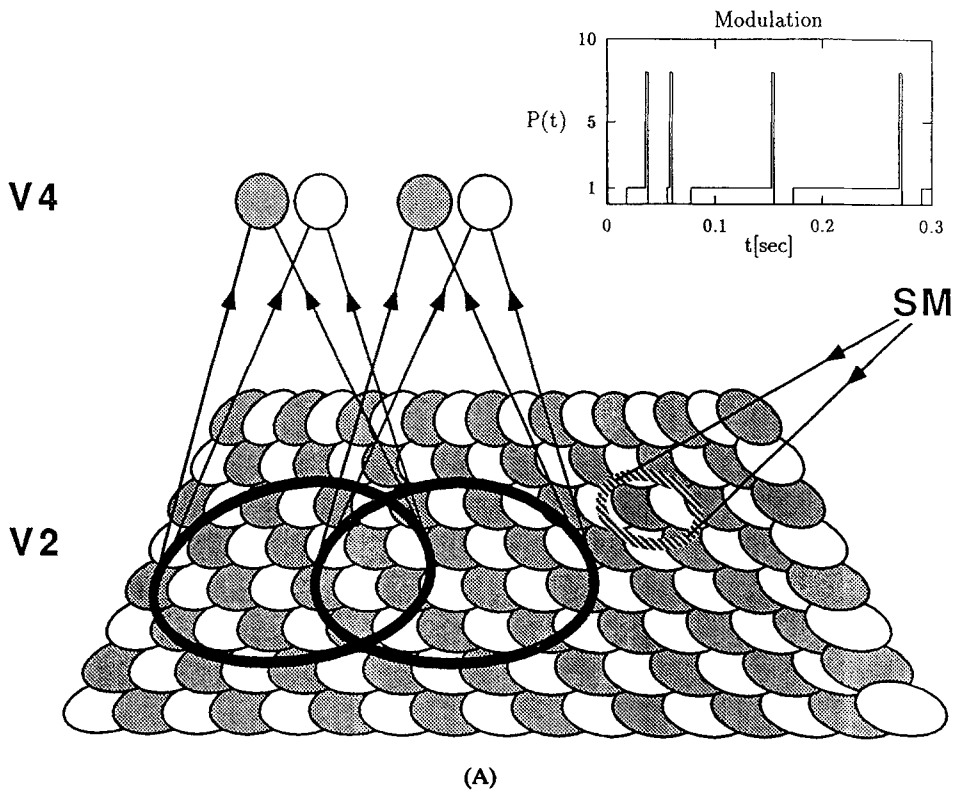


Fig. 1. (A) Architecture of the model. Receptive fields of V2 cells are represented by overlapping circles arranged in a two-dimensional tiling array (actual receptive fields in the model are square). White and gray circles represents cells receptive to the two features considered in the text. Actual overlap is larger than shown in the figure: every point is represented in the receptive fields of 4 cells of each feature type (i.e., by 4 white and 4 gray cells). The two unfilled black circles denote the receptive fields of V4 cells, i.e., all cells in a stack in V4 receive input from all V2 cell in the corresponding circle (arrows) with the same feature selectivity (See (B) for details of the connectivity). The shaded unfilled circle indicates the focus of attention; the activity of all V2 cells inside this circle is subjected to temporal modulation by the saliency map (SM). The inset shows the stochastic modulation of neuronal activity in attended cells. Shown is the modulation imposed on cells with receptive field inside the focus of attention. At random times (subject to a Poisson distribution), $P(t)$ deviates from its average value ($P = 1$) during events which consist of a 2.5 ms long elevation of the firing rate (in the case shown, to a level of $P = 8$), followed by a subsequent depression (to $P = 0$) of a duration of 17.5 ms. One can think of these events as the release of a neuromodulatory substance—which is quickly taken up again—that briefly increases the firing rate of all nearby neurons. The mean activity, averaged over any period of time which includes only entire events (i.e. no incomplete events) is unity. (B) Schematic connectivity of the model. The two types (white and gray) of V2 pyramidal cells (triangles) project to a stack of cells in V4. Strong excitatory synapses are made to V4 pyramidal cells (triangles) and smooth interneurons (circles) with the same preferred color (full lines), weaker connections are made to pyramidal cells of the other (non-preferred) color (dashed lines). The V4 smooth interneurons inhibit pyramidal cells of opposing feature selectivity.

or a non-preferred stimulus is present in the receptive field of the cell under study, $\lambda_0 = 0$. If a preferred stimulus is present, λ_0 is chosen to be proportional to the degree of spatial overlap of the receptive field of the cell with the stimulus,

$$\lambda_0 = \lambda_{max} \times \text{overlap(stimulus, receptive field)} \quad (1)$$

where $\lambda_{max} = 200$ Hz and the overlap varies between zero (no overlap) and unity (complete overlap).

The action of attention is to modulate this discharge *without* affecting its mean rate λ, since this is determined primarily by the stimulus conditions. We ignore here the occasional changes (which can be positive or negative) in V1 firing

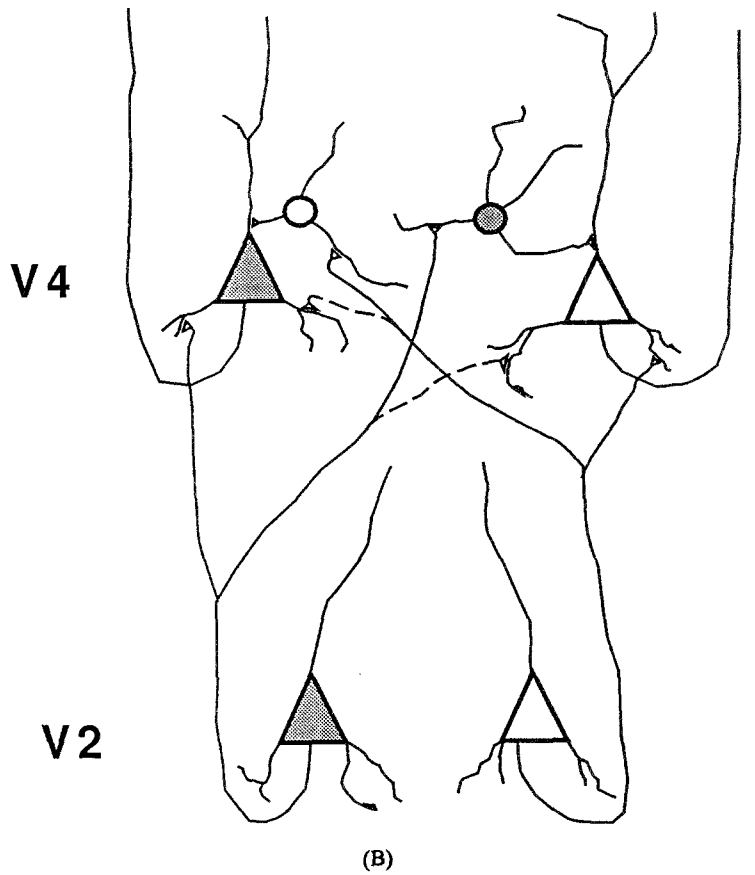


Fig. 1. (Cont'd.)

rates observed by Motter (1993). Such changes could be caused by the modulatory input from the saliency map which may influence the firing rate of some cells. The modulation is implemented using an inhomogeneous Poisson process, whose instantaneous firing rate $\lambda(t)$ varies over time. In the present model, we assume that $\lambda(t)$ is determined itself by a stochastic process with average rate $\lambda_{sm} = 20\ s^{-1}$ (see below for a discussion of this value). The events of this process do not correspond, however, to action potentials of V2 neurons but to a sequence of synchronized elevations and depressions of the spiking rates of all attended neurons. More concretely, we assume that the mean rate of an attended neuron is given by

$$\lambda(t) = \lambda_0 P(t) + \lambda_{\text{spont}} \qquad (2)$$

where $P(t) = 1$, except during a modulation event. If such an event of this Poisson process occurs at $t = t_0, P(t)$ instantaneously increases for 2.5 msec to 8, and then drops for 17.5 msec y to 0. This assures that the mean rate is always constant, independent of the level of attentional modulation (see inset in Fig. 1A).

Although the exact value of λ_{sm} is to some extent arbitrary in our model, it is nevertheless constrained within relatively strict limits. The lower limit comes from the requirement that the experimentally observed time required for changing the location of the focus of attention is on the order of some tens of milliseconds (Saarinen and Julesz, 1991), which means that $1/\lambda_{sm}$ should be not smaller than this value. On the other hand, λ_{sm} is constrained from above by the required stochastic character of the simulated spike train. Since during each of the events, which occur with an average rate of λ_{sm},

the instantaneous firing rate is increased by a factor of eight for a duration of 2.5 ms and this increase is compensated by a decreased of the firing rate for 17.5 ms, each event takes 20 ms. If λ_{sm} comes too close to the inverse of this time, or if it becomes even larger (i.e., if $\lambda_{sm} \approx 50$ Hz or larger), the spike train will consist of immediately adjacent instances of events of elevated and following suppressed firing. The spike train is then basically a regular series of bursts, following one another with a repetition rate of 20 ms, which is not compatible with the stochastic character of most observed spike trains.

2.2 *Extrastriate Cortex (V4)*

Following Crick and Koch (1990a), we assume that selective attention activates competition within a "stack" or microcolumn of neurons in V4. In the presence of multiple stimuli, the neurons responding to the different stimuli will compete against each other and attentional modulation will bias this competition in favor of attended stimuli. In our model, there are two classes of neurons in every stack: inhibitory interneurons which function as coincidence detectors and detect the correlation in their input, and excitatory pyramidal neurons, which provide the output to higher cortical areas. Members of both classes receive overlapping input from 100 V2 cells. Thus, the receptive fields of V4 cells are much larger than those of V2 cells. The single-cell models for both classes are similar and will be discussed together.

The basic difference between the two neuron classes is the behavior of the firing threshold. While excitatory neurons have a time-independent threshold $\Theta(t) = 15$ mV, V4 interneurons ignore the average activity. Thus, their firing threshold is modelled as an elaborated version of a gliding average of the cell's membrane voltage with a time constant τ_Θ (parameters are listed in Table 1). We assure that the threshold is on average above the membrane voltage by adding a constant offset $\Theta_0 = 7$ mV to the differential equation for Θ. Since the voltage excursion during an action potential is not modeled explicitly (instead, spikes are represented by the variable $s(t)$, see below), we add

Table 1. Parameters used in the simulations. The first column shows the symbol of the parameter and the second the number of the equation where it was introduced. The third and fourth columns show the values used in the models for the inhibitory and excitatory V4 neurons, respectively. The synaptic weights (last 5 rows) and the constants b and d are dimensionless.

Parameter	Equation	Value inh.	Value exc.
τ_Θ	3	5 ms	*n/a*
Θ_0	3	7 mV	*n/a*
d	3	120	*n/a*
V_{Na}	3	100 mV	100 mV
V_K	4	−20 mV	−20 mV
V_{Cl}	8	−25 mV	−25 mV
τ	4	10 ms	30 ms
τ_K	5	3 ms	3 ms
τ_{exc}	6	1 ms	1 ms
τ_{inh}	7	60 ms	60 ms
b	5	70	300
W'_{exc}	6	9×10^{-3}	3.75×10^{-4}
$W^{(s)}_{\text{exc}}$	6	9×10^{-3}	7.5×10^{-5}
$W^{(o)}_{\text{exc}}$	6	0	1.25×10^{-5}
W'_{inh}	7	6×10^{-2}	1.5×10^{-6}
$W'^{(o)}_{\text{inh}}$	7	0	5×10^{-4}

a voltage-dependent term to the threshold each time an action potential occurs. We arrive at the following equation for the threshold Θ for inhibitory neurons:

$$\tau_\Theta \frac{d\Theta}{dt} = -\Theta + \Theta_0 + V + d(V_{Na} - V)s(t) \quad (3)$$

where d is a constant and $V_{Na} = 100$ mV is the sodium reversal potential.

The equation for the membrane voltage V is

$$\tau \frac{dV}{dt} = -V + I + g_K(V_K - V) \quad (4)$$

where τ is the membrane time constant. The first term on the right-hand side represents the transmembrane leakage current, I is the normalized synaptic current (having the units of voltage; see eq. 8), and the last term represents the repolarization of the membrane following action potential initiation. In this term, g_K is the

(dimensionless) conductance of the potassium-selective ion channels (see eq. 5), and $V_K = -20$ mV the potassium reversal potential.

Whenever $V > \Theta$ and the last spike was generated prior to the absolute refractory period τ_r, chosen randomly with uniform probability from the interval [2 ms, 5 ms], an action potential is generated, that is, we set $s(t) = 1$ for the neuron (otherwise, $s(t) = 0$). The cell is repolarized by the opening of calcium-dependent potassium channels whose conductance is given by

$$\tau_K \frac{dg_K}{dt} = -g_K + bs(t) \qquad (5)$$

with $\tau_K = 3$ ms and $b = 70$. As was mentioned previously, g_K as well as all other ionic conductances in our model is made dimensionless by dividing the actual potassium conductance by the unit of conductance.

Action potentials induce synaptic conductance changes in postsynaptic cells. A V4 neuron receives input from all V2 cell in its receptive field. Let $t_{ij}^{(s)}$ be the time of the j-th spike of the i-th presynaptic V2 neuron with the same preferred features as the V4 neuron under study (the superscript "s" stands for "same" preferred features), and let $W_{\text{exc}}^{(s)}$ be the weight of the synapse of this V4 neuron (we assume all these weights to be identical). Let $t_{ij}^{(o)}$ and $W_{\text{exc}}^{(o)}$ be the corresponding spike times and synaptic weights for V2 cells with different preferred features (the superscript "o" stands for "other" preferred features, and we assume always that $W_{\text{exc}}^{(s)} > W_{\text{exc}}^{(o)}$). We lump the input from all other areas into two stochastic (Poisson) spike trains, one conveying excitatory input (spikes at times $t_{E,j}$) and the other inhibitory input (spikes at times $t_{I,j}$). The rates of these processes were determined by the requirement of obtaining a spontaneous firing rate of a few spikes per second in the absence of stimulation. Modeling the time courses of the conductances by a decaying exponential, we obtain the (dimensionless) excitatory conductance g_{exc} from the equation,

$$\tau_{\text{exc}} \frac{dg_{\text{exc}}}{dt} = -W_{\text{exc}}^{(s)} \sum_{i,j} \delta(t - t_{ij}^{(s)}) + W_{\text{exc}}^{(o)} \sum_{i,j} \delta(t - t_{ij}^{(o)}) + W'_{\text{exc}} \sum_{j} \delta(t - t_{E,j}) \qquad (6)$$

The equation for inhibitory conductances is analogous, except that the input is from V4 and that there is only input from cells with different preferred features (plus noise):

$$\tau_{\text{inh}} \frac{dg_{\text{inh}}}{dt} = -W_{\text{inh}}^{(o)} + \sum_{i,j} \delta(t - t_{ij}^{(o)}) + W'_{\text{inh}} \sum_{j} \delta(t - t_{I,j}) \qquad (7)$$

We also choose a longer time constant ($\tau_{\text{inh}} =$ 60 msec) for inhibitory than for excitatory conductances ($\tau_{\text{exc}} = 1$ msec), because of the longer activation time constants of GABAergic synapses relative to glutaminergic (non-NMDA) synapses.

Synaptic currents are products of a synaptic conductance and the difference between $V(t)$ and the respective reversal potential. We use the sodium reversal potential for excitatory synapses and the chlorine potential for inhibitory synapses. The total synaptic $I(t)$ current is then,

$$I = g_{\text{exc}}(V - V_{Na}) + g_{\text{inh}}(V - V_{cl}) \qquad (8)$$

3 Results

The activity of our V2 and V4 neurons is shown in Fig. 2. A key aspect of the "temporal tagging" hypothesis (Crick and Koch, 1990b) is evident here: in agreement with the data of Moran and Desimone (1985), the firing rate is modified by attention only in area V4, not in V2. Neurons in V2 respond to a stimulus at the same rate, whether they are unattended (Fig. 2A) or attended (Fig. 2B). Spike trains are not correlated between different trials (corresponding to different rows in Fig. 2A, B), except for the trivial correlation caused by the higher spiking frequency during stimulus presentation. Only the spike trains of attended neurons, recorded during the same trial, are correlated (see Fig. 4).

Figures 2C, D show the spike activity of two inhibitory V4 neurons which receive input from unattended and attended stimuli, respectively.

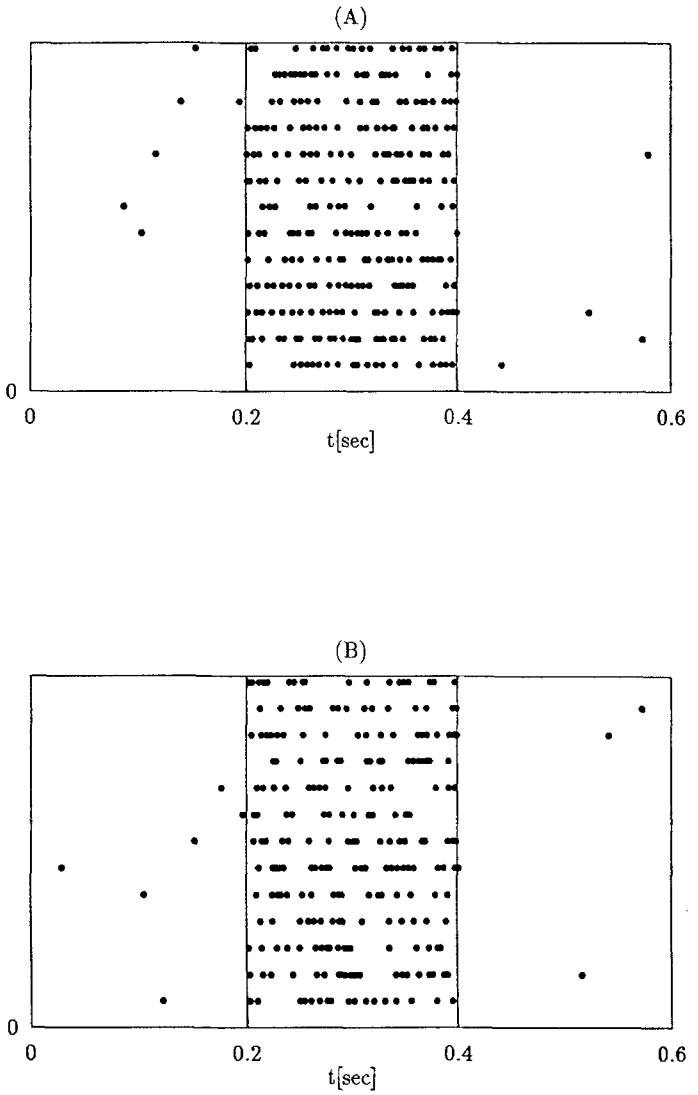


Fig. 2. Spiking activity of different cells. For each cell, 13 spike trains are shown, each action potential being represented by a dot. Stimuli are presented between $t = 0.2$ s and $t = 0.4$ s. Activity outside this period is due to spontaneous firing. (A) Response of a V2 cell to a preferred stimulus. The cell's receptive field has no overlap with the focus of attention ($P(t) = 1$ for all t). (B) Same, but for a cell in the focus of attention. The average spike rate for complete overlap with a preferred stimulus is the same as in (A), namely 202 Hz during stimulus presentation and 2 Hz otherwise. (C) Activity in inhibitory V4 cells. In contrast to (A, B), now two stimuli are presented simultaneously in the receptive field of the cell, one being of the preferred kind, the other of the non-preferred kind. The attention is directed to the non-preferred stimulus (as in the right part of Fig. 3A). After an initial bursts of activity due to the onset of the stimulus, the cell becomes nearly quiescent. Average spiking rate during stimulus presentation 7 Hz, spont. activity 0.7 Hz. (D) Same stimulus as in (C), but now the attention is directed on the preferred stimulus (as in the left part of Fig. 3A). The neuron continues to fire due to its correlated input. Average spiking rate during stimulus presentation 17 Hz, spont. activity 0.7 Hz. (E) Activity in excitatory V4 cells. Stimulus as in (C, D), attention directed on the non-preferred stimulus (as in C). (F) Same, but attention directed on the preferred stimulus (as in D). Note the higher firing rate in (F) *vs.* (E), due to the suppression of activity in the unattended neurons. For the stimulus shown, consisting of one 200 ms presentation per second, the average response to the stimulus (after subtraction of the spontaneous rate of 2.1 Hz) was suppressed from 16.9 Hz to 8.7 Hz.

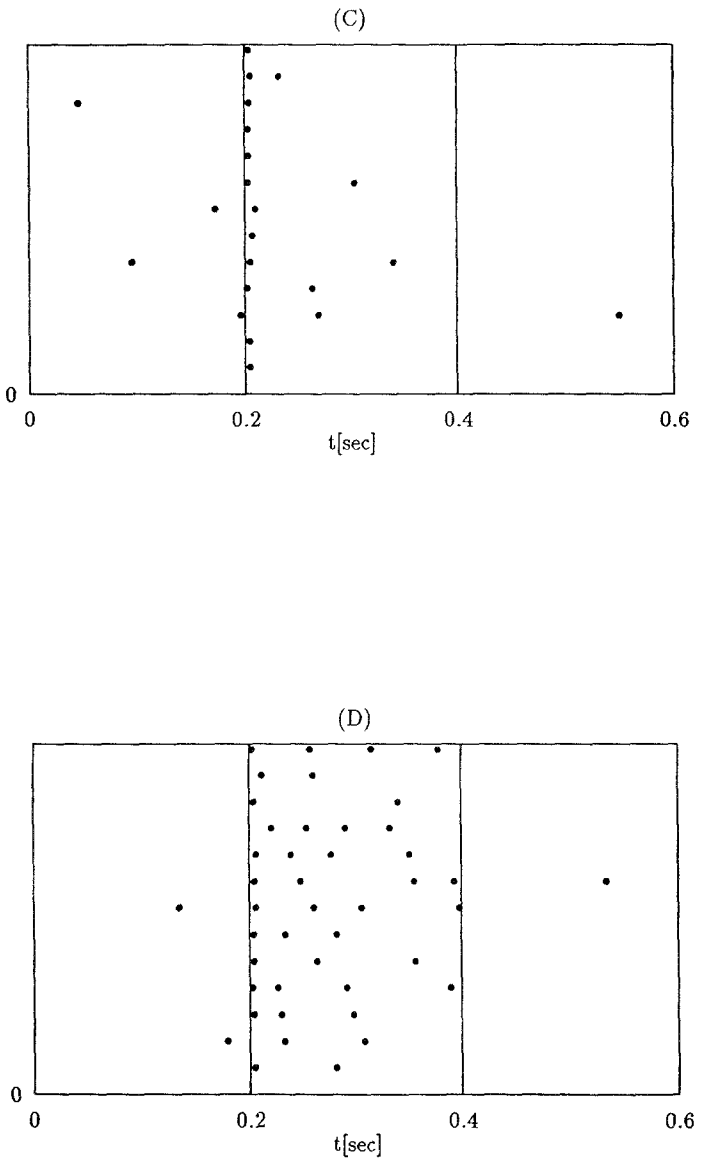


Fig. 2. (Cont'd.)

The uncorrelated input to the unattended neuron is usually unable to push the membrane voltage above the threshold, except at the stimulus onset when one or a few spikes are generated (Fig. 2C). Although the stimuli have identical strength, the attended interneuron receives modulated input, which enables it to fire during the presentation of the stimulus (Fig. 2D). It can therefore suppress the response of the unattended excitatory V4 neurons. Figures 2E, F illustrate the resulting response of unattended and attended excitatory V4 cells. The attended cell (Fig. 2F) spikes at a significantly higher rate than the unattended cell (Fig. 2E; see below and Table 2 for a more quantitative characterization). Note also the considerable spontaneous activity during the periods when no stimulus is present. Spontaneous activity is important in our model not only because a significant spontaneous firing rate is frequently observed exper-

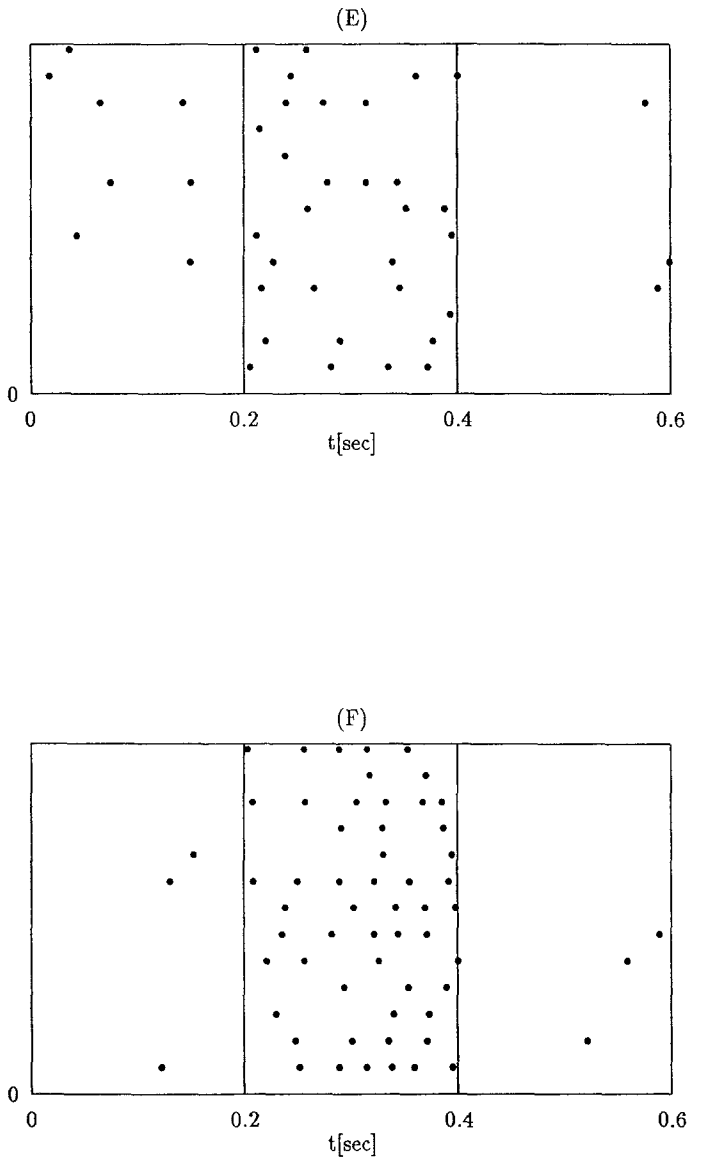


Fig. 2. (Cont'd.)

imentally in area V4, but also because the high level of added noise demonstrated the robustness of the temporal modulation mechanism we are proposing.

In order to compare our simulations against the experiments of Desimone and colleagues, we chose a representation analogous to that of Fig. 10 of Desimone and Ungerleider (1989). Figure 3A shows the activity of an excitatory V4 neuron with an effective and an ineffective stimulus in its receptive field. If attention is focused on the ineffective stimulus, the average rate of stimulus-generated action potentials (after subtraction of the spontaneous firing rate of 2.1 Hz) which are generated by the stimulus is reduced from 16.9 to 8.7 per second. This reduction in firing by a factor of about 1.9 in our model (bottom row) is comparable to the experimental results reported by Moran and Desimone (1985; middle row). They also found that fo-

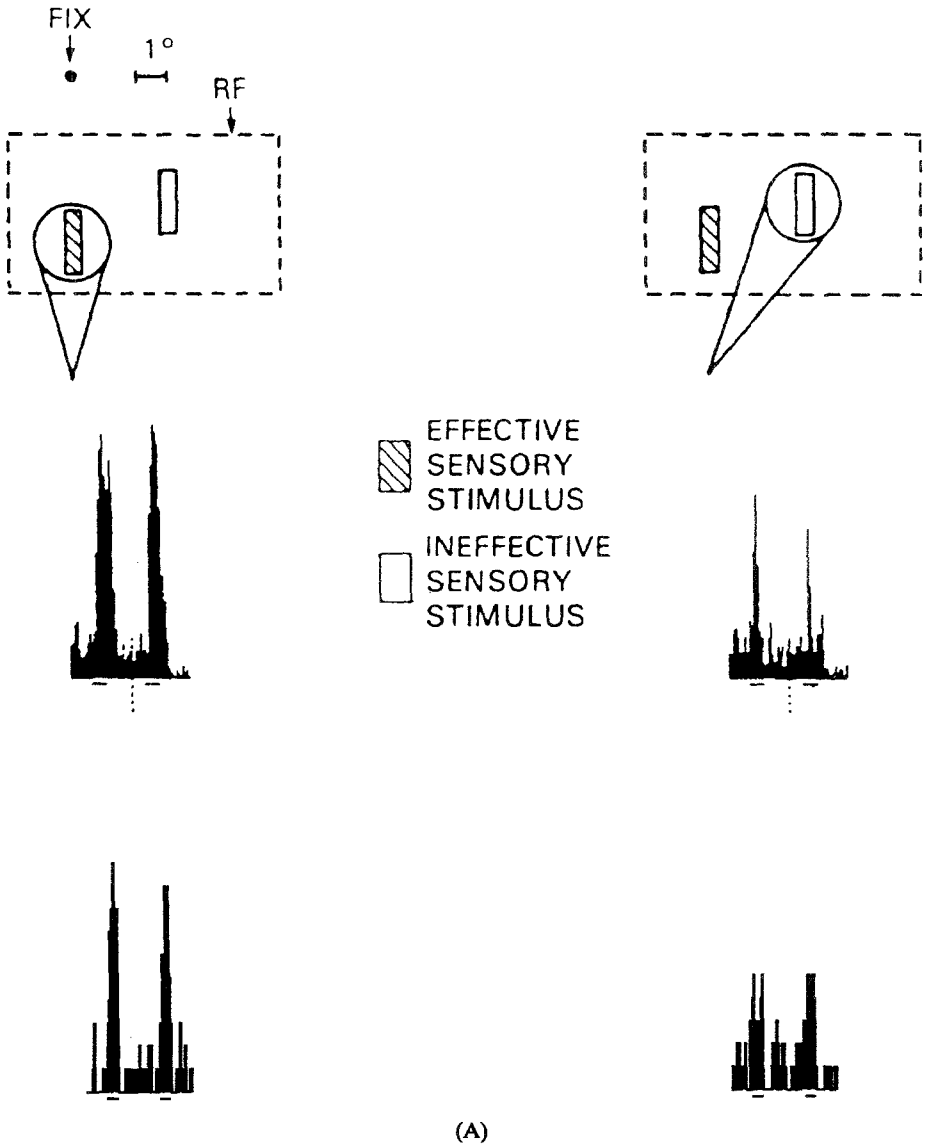
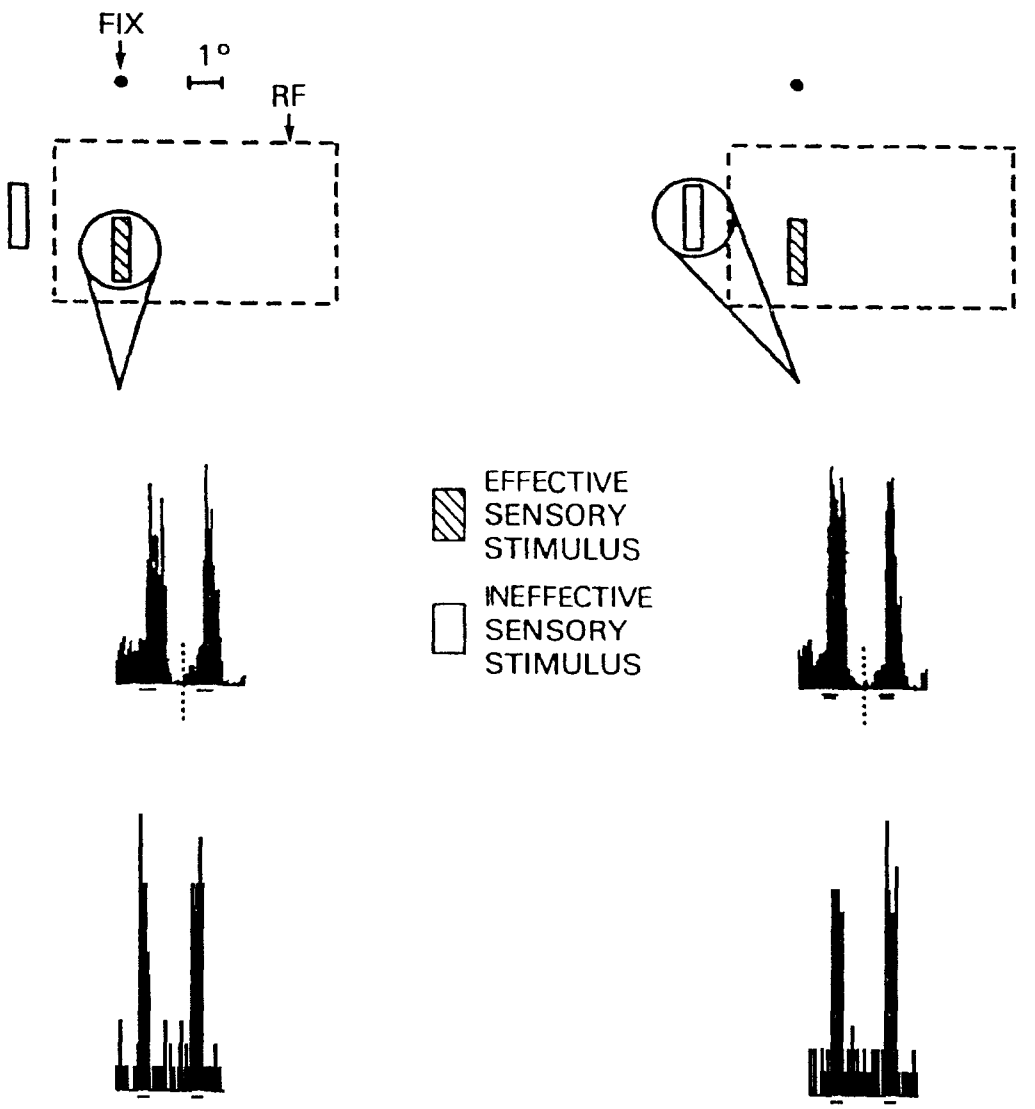


Fig. 3. Effect of selective attention on neural response in V4 in experiment (center row) and model (bottom row). The stimulus configuration is shown in the top row. Simulated results are the sum of 50 simulated experiments. (A) A preferred (hatched) and a non-preferred stimulus were presented simultaneously in the receptive field in the cell during two period of length 200 ms each, indicated by horizontal lines below the histograms. The monkey attended to one location (circled) inside the receptive fields (RF) of the cell while maintaining fixation at the fixation point (FIX). When the animal attended to the location of the effective stimulus, the cell gave a good response (left), but when the animal attended to the location of the ineffective stimulus, the cell gave only a poor response (right). The two periods of elevated activity visible in the experimental date correspond to the presentation of the target and the test stimuli used in the "delayed match to sample task" used to control the monkey's attention (see Desimone and Ungerlieder, 19889 for details of the experimental procedure). In our model, this was simulated by presenting the same stimulus twice. Experimental data reproduced, with permission, from (Desimone and Ungerleider, 1989).



(B)

Fig. 3. (Cont'd.) (B) As in (A), but with different stimulus condition, only one stimulus bar being present in the receptive field. In the experiment shown to the left, the monkey attends to the object inside the receptive field, while in the case shown to the right, the attention is directed to an object outside the receptive field. In both cases, the cell response is comparable to the response to two stimuli when the monkey attended the preferred stimulus (A, left) and significantly larger than when the monkey attended the non-preferred of two stimuli (A, right).

cussing attention on the ineffective stimulus fails to reduce the response to the effective stimulus as long as the attended stimulus is not inside the receptive field of the recorded cell. Figure 3B shows that this behavior (middle row) is reproduced in our model (bottom row). As is seen in Table 2, attention increases the stimulus-related response of the inhibitory model V4 neurons by an even larger factor (from 4.9 Hz to 14.5 Hz) than it is the case for the excitatory neurons. However, these neurons are small, aspiny interneurons and we expect that microelectrode recordings which are not designed specifically to find such cells (for an example of such a study

Table 2. Influence of attentional modulation on spike rates in model area V4. The first row ("spontaneous") shows the average spike rates (impulses per second) of excitatory and inhibitory V4 neurons in the absence of external stimuli (second and third column, respectively. The stimulus configuration corresponding to the second and third column is as in Fig. 3A, where two stimuli are presented in the overlapping receptive fields of two neurons with different preferred stimuli. Only one of the two stimuli is attended. The second row shows the response of excitatory and inhibitory cells whose preferred stimulus is attended, and the third row the response of unattended cells (all in impulses per second). After subtraction of spontaneous activity, the response of the unattended cells is suppressed by about a factor of 2 for the excitatory cells and by a factor of 3 for the inhibitory cells.

Stimulus	Excitatory	Inhibitory
spontaneous	2.1	0.8
attended	19	14.5
unattended	10.8	4.9

see Llinas et al. 1991) would only rarely record their activity. Instead, generic recordings will select the larger pyramidal cells.

Cross-correlation functions between two long spike trains of attended and unattended cells in V2 and V4 are shown in Figure 4. It is easy to see that the cross-correlation function vanishes (except for statistical fluctuations) between two pure Poisson processes. This is found to be the case for the correlation function between unattended V2 cells (not shown). Attended V2 cells, however, show positively correlated activity if τ is on the order of a few milliseconds, as is seen in Fig. 4A. For $\tau \approx 15$ ms, a valley in the correlation function is observed, i.e., a negative correlation. For longer times, the correlation function vanishes. This behavior follows from the temporal structure of the modulation of the V2 neurons which imposes synchronous firing for short times (2.5 ms) and suppression of activity after each period of enhanced firing probability for 17.5 ms (see inset in Fig. 1A). There is also a peak is the cross-correlation function at about 30 ms which is caused by a maximum in the inter-spike interval distribution at 30 ms. The modulation-induced peak in V2 is reproduced approximately between attended V4 neurons, as is seen in Fig. 4B. On the other hand, only a very weak correlation is observed between two unattended V4 neurons (Fig. 4C). The importance of the correlation between attended pyramidal V4 cells, which is similar to the correlation imposed by the attentional modulation on V2 neurons, lies in the fact that the pyramidal cells represent the output stage of area V4. The proposed mechanism conserves the synchronicity structure across cortical areas, allowing the system to "cascade" the mechanism to the next cortical stage.

4 Discussion

Temporal modulation of sensory evoked activity in early cortical areas (like V2) is proposed as a labelling mechanism for neurons which respond to properties of an attended object. This is in spirit related to the hypothesis that the figure-ground distinction is similarly due to synchronous neuronal activity in all neurons firing in response to the "figure" (von der Malsburg and Schneider, 1986, Engel et al. 1992, Baldi and Meir, 1990). This labelling enables the system to suppress the cell responses to non-attended objects present in close vicinity of the attended object. The model is compatible with the known anatomy of the mammalian visual system and makes use of simple neuron models: neurons in V2 generate spikes with a Poisson distribution (with either a constant or a time-varying mean rate of fire) while V4 cells are of the leaky-integrate-and-fire type. By an appropriate choice of the details of the cell dynamics, the properties of the inhibitory cells enhance the coincidence-detection capabilities of integrate-and-fire cells while the excitatory cells are basically pure integrators.

One of the hypotheses underlying the present work is the presently popular idea that the nervous system might make better use of the bandwidth available to spiking neurons by using not only the average spike rate but also the details of the temporal structure of the spike trains (at a resolution of one or a few milliseconds; see von der Malsburg, 1981, McClurkin et al. 1991, Bialek et al., 1991). This allows the system to use a multiplexed code for

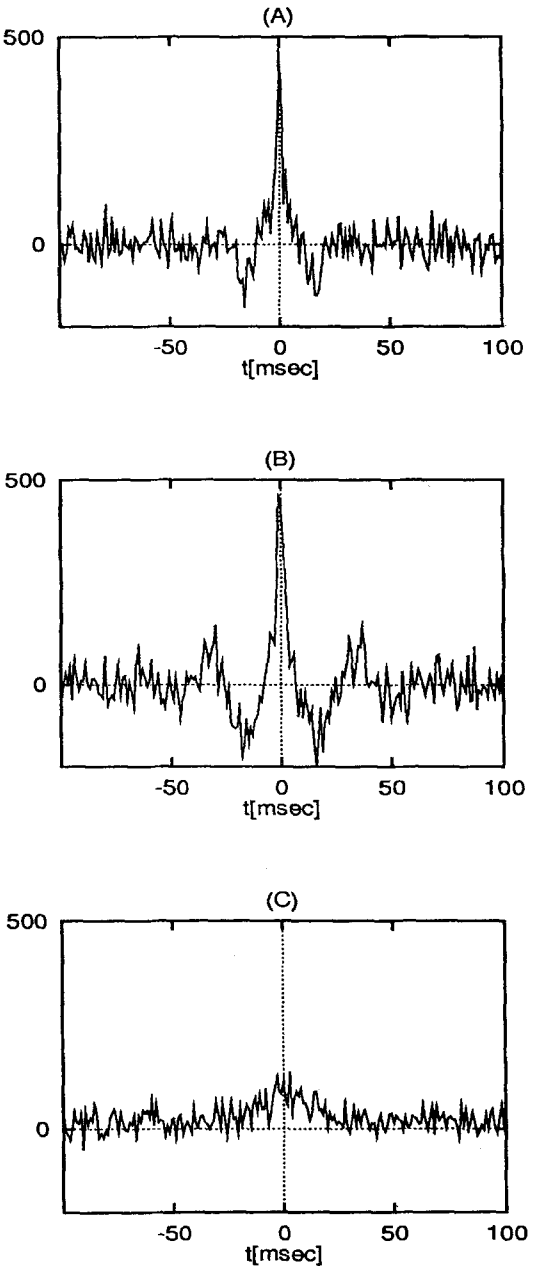


Fig. 4. Cross-correlations between (A) two V2 cell with receptive fields in the focus of attention, (B) two excitatory V4 cells in the focus of attention, (C) two excitatory V4 cells outside the focus of attention. The cross-correlation function $C_{f,g}(\tau)$ between the spike trains of two cells is computed as follows. We divide each of the spike trains, $f(t)$ and $g(t)$, in N adjacent pieces of length T each and number these segments as $f_i(t)$ and $g_i(t)$, respectively. Segments are defined by $f_i(t) = f(t - iT)$ for $0 \leq t < T$, and analogously for $g_i(t)$. We chose $N = 3$ in (A) and $N = 31$ in (B, C), the length of each segment being in each case $T = 32.768s$. Longer spike trains (length about 1000 s) were use for V4 neurons (in B,C) than for V2 neurons (length about 100 s; in A) in order to obtain comparable signal-to-noise ratios, since the average firing rate of V4 neurons is about 10 times smaller than that of V2 neurons. Assuming that correlations over very long times (longer than T) are accidental, we can then compute a shift predictor

$$S(\tau) = \frac{2}{N(N-1)} \sum_{i<j} \int f_i(t) g_j(t+\tau) dt \tag{9}$$

Fig. 4. (Contd.)

where the sum runs over the $N(N-1)/2$ permutations of (i,j) with $i<j$. The correlation function is then found as

$$C_{f,g}(\tau) = \frac{1}{N}\sum_{i=1}^{N}\int f_i(t)g_i(t+\tau)dt - S(\tau) \tag{10}$$

(In practice, the integrals are computed in Fourier space, using the fact that the Fourier transformation of the correlation function between two functions is the product of the Fourier transforms of the first function with the complex conjugate of the Fourier transform of the second function.). For a pure Poisson process, the correlation functions vanishes except for statistical fluctuations. The modulation-induced peak in V2 (A) is reproduced approximately between attended V4 pyramidal neurons (B), while only a very weak correlation is observed between unattended V4 pyramidal neurons (C). Valleys in (A) and (B) are due to the 17.5 msec long suppression of activity after each 2.5 msec long period of elevated firing.

different information dimensions. In the case considered, the stimulus properties are coded in terms of the average spike rate of neurons in primary cortical areas, whereas information about the attentional state of the animal (i.e., which stimulus is at any given time attended to) is provided in the fine temporal structure among spike trains. One of the crucial tasks of higher cortical areas would then be to decode this information and to make both kinds of information available. In the implementation suggested in the present work, the two kinds of V4 neurons described subserve this function, one of them (the inhibitory neurons working in the coincidence-detection regime) detecting correlated input while the other (the excitatory neurons of the integrator type) determining the average activity. Despite the simplicity of the model, it is capable of reproducing electrophysiological results quantitatively (Fig. 3). Noise in the form of realistic spontaneous activity was included in all stages to demonstrate the robustness of the proposed mechanism.

We here assume that synchronization is introduced by briefly and rapidly increasing the input from the saliency map in the form of $P(t)$ (see eq. 2 and Fig. 1A) to all neurons. By transiently increasing the firing rate of all neurons within the "focus of attention," cross-correlation among the neurons is induced for a short time. However, even this brief period of mutually enhanced firing probability is sufficient for the postsynaptic V4 interneurons to suppress activity of all non-attended cells.

In other modelling work (e.g., Olshausen et al. 1993), the suppression of unattended stimuli is effected by switching off the input of unattended neurons using specific synaptic connectivity schemes (thus the term "input-gating models" in the terminology of Desimone, 1992). In contrast, the requirements of our model regarding the details of the connectivity are minor, at the cost of a higher complexity in the temporal fine structure of the neuronal spike trains. Instead of switching off selected neurons by cutting off their *input*, temporal structure is used in our model to modulate the cells' *output* responses. Our model would therefore be classified as a "cell-gating model" by Desimone (op. cit.).

In a previous report (Niebur et al., 1993), we studied a model in which temporal tagging was implemented at the single cell level using an oscillatory signal in the 30–50 Hz range (referred to as "40 Hz" oscillations) observed in the visual system of the cat and macaque monkey (Eckhorn et al., 1988, Gray and Singer, 1989, Gray et al. 1990, Livingstone, 1991, Gray et al., 1993, Kreiter and Singer, 1992). The two models are based on the same overall anatomical architecture (compare Fig. 1 in Niebur et al., 1993 with Fig. 1 in this report), and in both models, competition between V4 cells is biased by temporal modulation in earlier cortices. The nature of this modulation is, however, different in the two models, and therefore different functional cell models are required. Temporal tagging was mediated in the previous work by oscillatory modulation of the average spike rate around a mean value. However, while the existence of single unit activity oscillating in the 30–70 Hz range is strong (for a review see Singer, 1993), a number of groups have failed to find such

oscillatory activity in single neurons in cortical areas V1, MT and IT in the anaesthesized monkey (Young et al., 1992, Tovee and Rolls, 1992, Bair et al., 1994). We therefore investigate alternative ways to implement temporal tagging in the present model in which no periodically recurring events are required. Instead, attended regions of the visual field are distinguished from unattended regions by the activity of V2 neurons which respond to attended stimuli by firing synchronously (within a few milliseconds) but not periodically. In both case, there is no correlation between unattended neurons. These different modulation schemes in early cortices are reflected in different detection schemes in higher areas. The oscillatory modulation used in (Niebur et al. 1993) was detected in area V4 by frequency selective neurons. In the present work, detection of attended stimuli relies on neurons which detect coincidences in their input, rather than specific frequencies.

Which one—if any—of the described two models corresponds to physiological reality will ultimatively have to be determined by experiment. But models predict relatively subtle changes of the temporal structure of spike trains inside the focus of attention compared to the response to the same stimulus outside the focus of attention. If the implementation of selective attention is based on oscillatory mechanisms, as proposed in (Niebur et al., 1993), periodic modulation should be visible in the recordings of cells with attended stimuli in their receptive fields (but their amplitude can be expected to be relatively weak, as was shown in the cited work). In contrast, modulation by synchronization, as proposed in the present work, would not be visible in any single spike train, but only in multi-unit recordings. Furthermore, both models predict correlations between spike trains generated by neurons in different cortical areas if an attended stimulus is present in their receptive fields. The discovery of stimulus-locked, long-distance correlations between cells in visual cortex of cat (Gray and Singer, 1989, Engel et al., 1991a, Engel et al., 1991b) and monkey (Kreiter and Singer, 1992, Livingstone, 1991) is certainly compatible with this prediction, insofar as it shows that the neural substrate is capable of producing such synchronization. However, these experiments cannot be taken as direct evidence for the validity of the presented model, since this would require careful control of the attentional state of the animals, which was not the case in the cited studies (most of which were performed in anesthetized animals).

Keele and collaborators used psychophysical methods to study temporal mechanisms for attentional binding (Keele et al., 1988). They found that binding of different features of an object to closely related to the common location of these features, whether the features are presented simultaneously or not. These findings provid little support for models in which external temporal information is used for attentional binding (e.g., von der Malsburg and Schneider, 1986, Horn et al, 1991), but they are consistent with models in which the temporal patterns are generated internally, as in our models. Indeed, the strong correlation of attentional binding with the spatial location of the object is consistent with both of our models, since they are based on a spatially defined focus of attention in which the modulation takes place.

A consequence of the brain's distributed coding of features is that the brain must identify and distinguish objects based on the conjunction of features. In the Crick and Koch (1990a,b) and in other models (von der Malsburg, 1981, Engel et al., 1992), this is achieved naturally by imprinting a temporal structure on the neural signals at the peripheral level, which can then be detected in all structures higher in the cortical hierarchy. The representation of features of the objects inside the focus of attention share a common trait throughout all cortical areas, namely their temporal structure. We suggest that imposing a temporal modulation on attended sensory signals is a plausible mechanism for producing unique percepts within the highly distributed architecture of the cortex.

Acknowledgments

We thank Francis Crick and Marius Usher for helpful discussions, Bob Desimone for providing us with part of Figure 3, and Wyeth Bair

for wrting the routines for computing the correlation functions. This work was supported by the Office of Naval Research, the Air Force Office of Scientific Research and the National Science Foundation.

References

Bair, W., Koch, C., Newsome, W., and Britten, K. (1993). Power spectrum analysis of MT neurons in the awake monkey. In Bower J. and Eeckman F., editors, Computation and Neural Systems. Kluwer, Norwell, MA.

Bair, W., Koch, C., Newsome, W., and Britten, K. (1994). Power spectrum analysis of bursting cells in area MT in the behaving monkey.

Baldi, P. and Meir, R. (1990). Computing with arrays of coupled oscillators: an application to preattentive texture discrimination. Neural Computation, 2:458–471.

Bialek, W., Rieke, F., De Ruyter Van Stevenick, R.R., and Warland D. (1991). Reading a neural code. Science, 252:1854–1857.

Braun, J. and Sagi, D. (1990). Vision outside the focus of attention. Perception and Psychophysics, 48:45–58.

Colby, C.L. (1991). The neuroanatomy and neurophysiology of attention. J. Child Neurology, 6:S90–S118.

Crick, F. and Koch, C. (1990a). Some reflections on visual awareness. Cold Spring Harbor Symp. Quant. Biol., 55:953–962.

Crick, F. and Koch, C. (1990b). Towards a neurobiological theory of consciousness. Seminars in the Neurosciences, 2:263–275.

Desimone, R. (1992). Neural circuits for visual attention in the primate brain. In Carpenter G. and Grossberg S., editors, *Neural networks for vision and image processing*. MIT Press, Cambridge.

Desimone, R. and Ungerleider, L.G. (1989). Neural mechanisms of visual processing in monkeys. In Boller F. and Graffman J., editors, *Handbook of Neuropsychology*, pages 267–299. Elsevier, Amsterdam.

Desimone, R., Wessinger, M., Thomas, L., and Schneider, W. (1991). Attentional control of visual perception: cortical and subcortical mechanisms Symp. Quant. Biol., 55:963–971.

Eckhorn, R., Bauer, R., Jordan, W., Brosch, M, Kruse, W, Munk, M., and Reitboeck, H.J. (1988). Coherent oscillatios: a mechanism of feature linking in the visual cortex? Biol. Cybern., 60:121–130.

Engel, A.K., König, P., Kreiter, A.K., Schillen, Th.B., and Singer, W. (1992). Temporal coding in the visual system: new vistas on integration in the nervous system. Trends in Neurosciences, 15:218-226.

Engel A.K., König, P., Kreiter, A.K., and Singer, W. (1991a). Interhemispheric synchronization of oscillatory neuronal responses in cat visual cortex. Science, 252:1177–1179.

Engel, A.K., König, P., Kreiter, A.K., and Singer, W. (1991b). Synchronization of oscillatory neuronal responses between striate and extrastriate visual cortical areas of the cat. Proc. Nat. Acad. Sci., USA, 88:6048–6052.

Gray, C.M., Engel, A.K., König, P., and Singer, W. (1933). Synchronization of oscillatory neuronal responses in cat striate cortex: temporal properties. Visual Neurosci., 8:337–347.

Gray, C.M., Engel, A.K., König, P., and Singer, W. (1990). Stimulus-dependent neuronal oscillations in cat visual cortex: receptive field properties and feature dependence. Europ. J. Neurosci., 2:607–619.

Gray, C.M. and Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. Proc. Nat. Acad. Sci., USA, 86:1698–1702.

Horn, D., Sagi, D., and Usher, M. (1991). Segmentation, binding and illusory conjunctions. Neural Computation, 3:510–525.

Julesz, B. (1991). Early vision and focal attention. Rev. Mod. Physics, 63:735–772.

Kanwisher, N. and Driver, J. (1992). Objects, attributes and visual attention: which, what, and where. Current Directions in Psychological Science, 1:26–31.

Keele, S.W., Cohen, A., Ivry, R., Liotti, M., and Yee, P. (1988). Test of a temporal theory of attentional binding J. Experimental Psychology: Human Perception and Performance, 14:444–452.

Koch, C. and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. Human Neurobiol., 4:219–227.

Kreiter, A.K. and Singer, W. (1992). Oscillatory neuronal response in the visual-cortex of the awake macaque monkey. Europ. J. Neurosci., 4(4):369–375.

LaBerge, D. and Buchsbaum, M.S. (1990). Positron emission tomographic measurements of pulvinar activity during an attention task. J. Neurosci., 10:613–619.

Livingstone, M.S. (1991). Visually-evoked oscillations in monkey striate cortex. Soc. Neurosci. Abstr., 17(1):176.

Llinas, R.R., Grace, A.A., and Yarom, Y. (1991). In vitro neurons in mammalian cortical layer 4 exhibit intrinsic oscillatory activity in the 10- to 50-hz frequency range. Proc. Natl. Acad. Sci. USA, 88:987–901.

McClurkin, J.W., Optican, L.M., Richmond, B.J., and Gawne, T.J. (1991). Concurrent processing and complexity of temporally encoded neuronal messages in visual perception. Science, 253:675–677.

Moran, J. and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. Science, 229:782–784.

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. Neurophysiology, 70(3):909–919.

Mountcastle, V.B., Anderson, R.A., and Motter, B.C. (1981). The influence of attentive fixation upon the excitability of the light-sensitive neurons of the posterior parietal cortex. J. Neurosci., 1:1218–1232.

Nakamura, H., Gattass, R., Desimone, R., and Ungerleider, L.G. (1993). The modular organization of projections from area V1 and area V2 to area V4 and TEO in macaques. J. Neuroscience, 13(9):3681–3691.

Niebur, E., Koch, C., and Rosin, C. (1993). An oscillation-based model for the neural basis of attention. Vision Research, 33:2789–2802.

Olshausen, B., Andersen, C., and Van Essen, D. (1993). A neural model of visual attention and invariant pattern recognition. J. Neuroscience, 13:4700–4719.

Petersen, S.E., Robinson, D.L., and Morris, J.D. (1987). Contributions of the pulvinar to visual spatial attention. Neuropsychologia, 25:97–105.

Posner, M.I. and Driver, J. (1992). The neurobiology of selective attention. Current Opinion in Neurobiology, 2:165–169.

Posner, M.I. and Petersen S.E. (1990). The attention system of the human brain. Ann. Rev. Neurosci., 13:25–42.

Rafal, R.D. and Posner, M.I. (1987). Deficits in human visual spatial attention following thalamic lesions. Proc. Nat. Acad. Sci., USA, 84:7349–7353.

Robinson, D.L. and Petersen, S.E. (1992). The pulvinar and visual salience. Trends in Neurosciences, 15(4): 127–132.

Saarinen, J. and Julesz, B. (1991). The speed of attentional shifts in the visual field. Proc. Nat. Acad. Sci., USA, 88:1812–1814.

Singer, W. (1993). Synchronization of cortical activity and its putative role in information processing and learning. Annu. Rev. Physiol., 55:349–74.

Softky, W. and Koch, C. (1993). The highly irregular firing of cortical-cells is inconsistent with temporal integration of random EPSPS. J. Neurosci., 13(1):334–350.

Tovee, M.J. and Rolls, E.T. (1992). Oscillatory activity is not evident in the primate temporal visual cortex with static stimuli. Neuro Report, 3:369–372.

Treisman, A. (1988). Features and Objects: the fourteenth Bartlett memorial lecture. Quant. J. Exp. Psychol., 40A:201–237.

von der Malsburg, C. (1981). The correlation theory of brain function. Technical Report 81-2, Max-Planck-Institute for Biophysical Chemistry, D-3400 Goettingen, Germany.

von der Malsburg, C. and Schneider, W. (1986). A neural cocktail party processor. Biol. Cybern., 54:29–40.

Young, M.P., Tanaka, K., and Yamane, S. (1992). On oscillating neuronal responses in the visual cortex of the monkey. J. Neurophysiol., 67:1464–1474.