Behavioral/Systems/Cognitive

Visual Saliency Computations: Mechanisms, Constraints, and the Effect of Feedback

Alireza Soltani[1,2] **and Christof Koch**[1]

[1]Division of Biology and Computation and Neural Systems, California Institute of Technology, Pasadena, California 91125, and [2]Department of Neuroscience, Baylor College of Medicine, Houston, Texas 77030

The primate visual system continuously selects spatial proscribed regions, features or objects for further processing. These selection mechanisms— collectively termed selective visual attention—are guided by intrinsic, bottom-up and by task-dependent, top-down signals. While much psychophysical research has shown that overt and covert attention is partially allocated based on saliency-driven exogenous signals, it is unclear how this is accomplished at the neuronal level. Recent electrophysiological experiments in monkeys point to the gradual emergence of saliency signals when ascending the dorsal visual stream and to the influence of top-down attention on these signals. To elucidate the neural mechanisms underlying these observations, we construct a biologically plausible network of spiking neurons to simulate the formation of saliency signals in different cortical areas. We find that saliency signals are rapidly generated through lateral excitation and inhibition in successive layers of neural populations selective to a single feature. These signals can be improved by feedback from a higher cortical area that represents a saliency map. In addition, we show how top-down attention can affect the saliency signals by disrupting this feedback through its action on the saliency map. While we find that saliency computations require dominant slow NMDA currents, the signal rapidly emerges from successive regions of the network. In conclusion, using a detailed spiking network model we find biophysical mechanisms and limitations of saliency computations which can be tested experimentally.

Introduction

A crucial computational strategy of the primate visual system is to swiftly allocate processing resources to a region, feature or object to deal with the many overlapping and partially occluding objects in natural scenes. Attentional selection can be guided by exogenous signals from the environment, such as a red flashing light (bottom-up, saliency-driven attention), by endogenous signals such as when looking for a specific car in a parking lot (top-down, volitional-controlled attention), or by both. Nevertheless, the neural mechanisms underlying these processes are mostly unknown.

From a computational point of view, a purely feedforward model of bottom-up attention incorporating a saliency map successfully predicts a large fraction of fixated locations under free viewing conditions (Itti et al., 1998; Itti and Koch, 2001; Parkhurst et al., 2002; Cerf et al., 2008; Foulsham and Underwood, 2008; Mannan et al., 2009). At its heart there is a two-dimensional (2-D) topographic arrangement of neurons that represent stimulus saliency throughout the visual scene. Initially, prominent locations corresponding to regions with enhanced feature contrast are computed in individual maps (i.e., conspicuity maps) for different dimensions of the stimulus such as intensity, orientation, color, motion, etc. These computations are performed through a set of multiscale, center-surround and normalization operations. Finally, the conspicuity maps are combined to form a single saliency map. Activity in this map does not encode conspicuity in any one particular feature dimension, but encodes the overall conspicuity of a given location relative to its local and global neighborhood. Based on electrophysiological evidence from the monkey, the lateral intraparietal cortex (LIP) and the frontal eye fields (FEF) have been identified as possible saliency maps (Gottlieb et al., 1998; Kusunoki et al., 2000; Bisley and Goldberg, 2003; Moore and Armstrong, 2003; Thompson and Bichot, 2005).

While it is believed that LIP and FEF represent the saliency map, neurons in lower visual areas V1, V4, and V5 (MT) also show differential responses to a target stimulus depending on surrounding stimuli or its spatiotemporal context (Allman et al., 1985; Knierim and van Essen, 1992; Albright and Stoner, 2002; Hegdé and Felleman, 2003; Burrows and Moore, 2009). For example, Hegdé and Felleman (2003) measured the response of V1 neurons to oriented and colored bars in the receptive field (RF) that had different saliency values. In particular, they compared the response to popout targets—say a red bar among green bars that rapidly attracts the eye—to conjunction targets—say a red, vertical bar among red, horizontal and green, vertical and horizontal ones—targets that are defined by the combination of two or more feature dimensions. Such conjunction targets are not readily detectable. They found that V1 neurons do not distinguish between the popout and conjunction targets and therefore, that V1 neurons do not carry saliency signals. More recently, Burrows and Moore (2009) examined the response of V4 neurons to similar stimuli and concluded that these neurons can distinguish between the

Received March 24, 2010; revised July 9, 2010; accepted July 17, 2010.

We are grateful to the Mathers Foundations, the Office of Naval Research, and the Defense Advanced Research Projects Agency for financial support of the research reported here. We thank Zahra Ayubi, Brittany Burrows, and Tirin Moore for helpful discussions and comments on the manuscript.

Correspondence should be addressed to Alireza Soltani at the above address. E-mail: soltani@bcm.edu.

DOI:10.1523/JNEUROSCI.1517-10.2010

Copyright © 2010 the authors 0270-6474/10/3012831-13$15.00/0

popout and conjunction targets. Paradoxically, they also showed that the saliency signal in V4 diminishes if the monkey prepares a saccade to a location far from the RF of the neuron, indicating an important role for top-down attention in the formation of the bottom-up driven saliency signal in V4.

These findings raise some important questions. First, how are saliency signals formed in the visual cortex across the cascade of regions from V1, to V2, V3, V4 and so on? Second, how does top-down attention affect these computations?

To answer above questions and shed light on neural substrates of bottom-up attention and its interaction with top-down attention, we construct a 2-D, biophysically plausible spiking network model. The network contains three distinct layers of neural populations corresponding to three cortical regions—which we identify from here on as V1, V2, and V4—and a higher visual area assumed to instantiate the saliency map (either LIP or FEF). The model neurons receive realistic inputs which are generated from the actual stimuli used in the relevant experiment (Hegdé and Felleman, 2003; Burrows and Moore, 2009). Using our model, we consider biophysical mechanisms and constraints on saliency computations and how these are influenced by top-down attention. Our hypothesis is that feedback from a cortical area representing the saliency map to earlier visual areas improves saliency computations, while top-down attention interferes with these computations by disrupting the feedback through its influence on the activity in the saliency map.

Materials and Methods

We use leaky-integrate-and-fire (LIF) model neurons with realistic synapses as building blocks. Our spiking network model contains many 2-D populations of neurons (24 and 10 in the first and second set of simulations, respectively) with realistic inputs and synapses, making it computationally expensive. For example, simulating 200 trials of the response to a given stimulus takes ~10 h to run on a standard Unix system with a 3 GHz Intel CPU. We therefore had to adopt some simplifications.

First, we assume that inputs to the network are the outputs of lateral geniculate nucleus (LGN) and V1 neurons that are wavelength- and orientation-selective, respectively. We do not explicitly model these cells, using instead their RF properties to generate their response to visual stimuli. As a result, the inputs to the network have wavelength (here the colors red and green) and/or orientation selectivity (0, 45, 90, 135), and we only explicitly model the visual processing in the output layers of V1, cortical areas V2 and V4, and a higher area corresponding to the saliency map (LIP/FEF). Second, we use Cartesian coordinates and ignore the effect of cortical magnification.

For each trial, we simulate 300 ms of the network dynamics with $dt = 0.1$ ms using the improved RK2 integration algorithm (Hansel et al., 1998). We directly compare our model against two different electrophysiological experiments in the alert and behaving monkey (Hegdé and Felleman, 2003; Burrows and Moore, 2009) using the same visual stimuli (see below).

Spiking network model. The model consists of 3 regions of neural populations, each of which contains 4 excitatory and 4 corresponding inhibitory populations (not represented) of LIF units (Fig. 1). Each population consists of 28 × 28 neurons, covering 14° × 14° of the visual field. Therefore, each neuron spans 0.5° of the visual field. We assume periodic boundary conditions for connections between all neurons (i.e., each 2-D neural population is placed on a torus).

We examine two exclusive architectures for the network. In configuration *A* (Fig. 1*A*), individual features (i.e., color and orientation) are processed in separate neural populations, and neurons in different populations receive inputs selective to only one feature. That is, there are no cells which participate in saliency computations and are tuned to both color and orientation. To compare the activity of these model neurons with experimentally recorded neurons selective to two features (say a red bar at 0° orientation), we combine the outputs of neural populations selective to the color red and to 0° orientation (Fig. 1*A*). We formulate this combination by simply adding the spike trains of neurons with similar RF in the corresponding neural populations (to avoid further computations). In configuration *B* (Fig. 1*B*), a combination of features is jointly processed and neurons in different populations receive inputs selective to both orientation and color (e.g., red color and 0° orientation). That is, oriented neurons are also color-tuned and we can directly compare the activity of neurons in this configuration with the experimental data. Apart from the input organization, all parameters are similar for the two configurations.

Model parameters for all excitatory neurons are set to: threshold voltage $V_{th} = -50$ mV, reset voltage $V_{reset} = -55$ mV, leak voltage $V_{leak} = -70$ mV, refractory time period $t_{ref} = 2$ ms, capacitance $C_E = 0.5$ nF, and leak conductance $G_{leak,E} = 25$ nS. Inhibitory interneurons have similar parameters except that the capacitance and leak conductance are set to $C_I = 0.2$ nF and $G_{leak,I} = 20$ nS, respectively. Neurons in each population are connected to all other neurons with a circular Gaussian profile (i.e., with equal width, σ, in both dimensions). That is, most connections are local with synaptic weight falling off with distance.

Excitatory neurons project to their target neurons through two types of synaptic receptors; fast AMPA, with the time constant $\tau_s = 2$ ms, and slow saturating NMDA, with the time constant $\tau_s = 80$ ms. The spatial extent of the connectivity profile (i.e., σ in the Gaussian function) is the same for both receptor types (see supplemental Table 1, available at www.jneurosci.org as supplemental material, for connectivity parameters). Inhibitory neurons are connected to their target neurons through GABA synapses with the time constant $\tau_s = 10$ ms. The peak conductance for all synapses, g_{syn}, is set to 1 nS multiplied by the connection strength (see below). All synapses are modeled as having exponentially decreasing conductances with time.

To capture the observed response adaptation in visual areas, we include spike-rate-adaptation (SRA) current for all neurons. For neurons in feature-selective populations, we set the change in conductance and the time constant of the SRA current to $g_{SRA} = 0.6$ nS and $\tau_{SRA} = 50$ ms, respectively. To reduce the strong response to the single bar stimulus in the saliency population, we adopt a stronger SRA current for neurons in this population ($g_{SRA} = 4.0$ nS and $\tau_{SRA} = 50$ ms).

Every excitatory neuron in a given population is connected to all other neurons in that population and to all interneurons in the corresponding inhibitory population. Cross-orientation inhibition is implemented by subtracting 25% of the mean of all orientation inputs (0°, 45°, 90°, 135°) from a given orientation input. In addition, there are feedforward connections between excitatory neurons with similar feature selectivity in successive regions.

All connection matrices are normalized Gaussian functions, with width σ, multiplied by the weight of these connections, w (see supplemental Table 1, available at www.jneurosci.org as supplemental material). We assume identical values for σ and w in the three simulated regions. Because the connectivity matrices are normalized, the value of w for a given connection should not be taken by itself as the magnitude of that connection strength.

To study the importance of NMDA currents in saliency computations, we reduce NMDA currents while increasing AMPA currents such that their sum remains approximately the same. Specifically, for the two additional sets of simulations presented here, we set the connection strength parameters between excitatory neurons, $w^{AMPA}_{EE,i\to i}$ and $w^{NMDA}_{EE,i\to i}$, equal to [33, 22] and [44, 11], respectively (compare with the original values of [11, 44]) (see supplemental Table 1, available at www.jneurosci.org as supplemental material). As we reduce the NMDA currents between excitatory neurons we also need to reduce the strength of the NMDA currents from the excitatory to inhibitory neurons to avoid the slowly activated inhibitory neurons from suppressing the activity in the network after the onset response. For connections from the excitatory to inhibitory neurons we set $w^{AMPA}_{EI,i\to i}$ and $w^{NMDA}_{EI,i\to i}$, equal to [180, 45] and [202.5, 22.5], respectively (compare with the original values of [135, 90]). Note that the connection widths, σ, are similar for AMPA and NMDA synapses. These and the rest of the model parameters are kept the same for simulations on the role of NMDA receptors.

Finally, to study the effect of feedback from the saliency map on the saliency computations in V4, we use a simplified version of our model

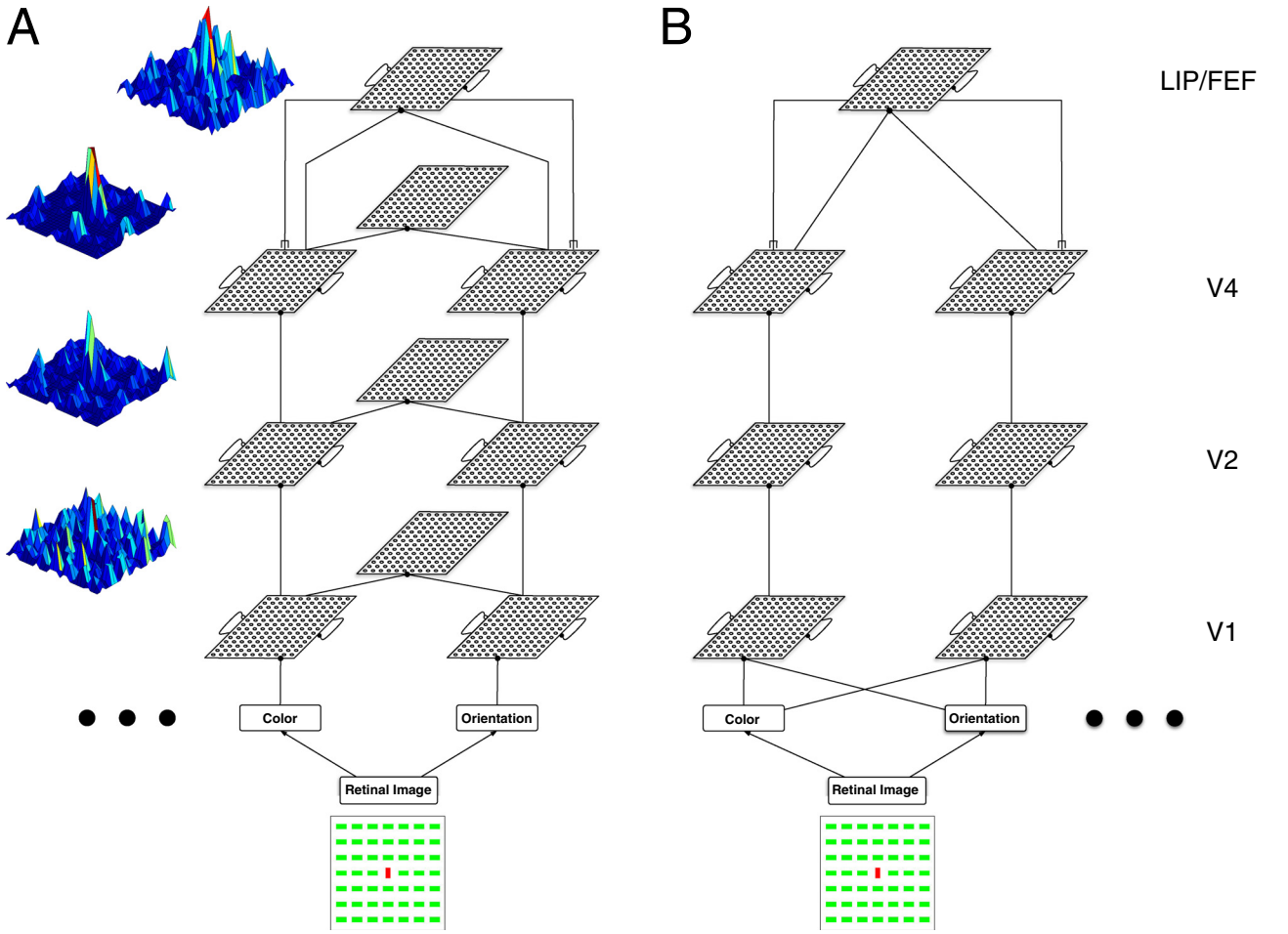

Figure 1. Two alternative architectures of the spiking network model of saliency. The model consists of 3 cortical regions—V1, V2 and V4. Each one includes 8 populations of excitatory and inhibitory interneurons, each modeled by 28 × 28 LIF units. Neurons in each population are connected to all neurons in the same population and to their corresponding interneurons (not represented). The first layer of populations represents neurons in the output layers of V1. Excitatory neurons in V1 (respectively V2) provide excitatory inputs to excitatory neurons at corresponding topographic locations in V2 (respectively, V4). For some simulations, we also included an explicit saliency map that receives inputs from all color and orientation-selective neurons in V4, and provides feedback to both excitatory and inhibitory neurons in V4. ***A***, In configuration *A*, each excitatory population receives an input selective to one of two orientations (0° and 90°) or one of two colors (red and green), and projects to neurons with similar selectivity in the next processing stage. These inputs are generated by the RF properties of LGN and V1 neurons. To compare the results with experimental data, we add the activity of these populations to construct different color- and orientation-selective populations. The insets show the average response (during a sample trial) of the constructed neural populations selective to red-horizontal bar and of the saliency population, to the combined popout display. ***B***, In configuration *B*, each excitatory population receives a combination of inputs selective to orientation and color (red-horizontal, red-vertical, green-horizontal, and green-vertical).

(for the sake of computational efficiency). This simplified model contains two layers of neural populations: the first layer corresponding to feature-selective neurons in V4 and the second layer corresponding to spatially selective saliency neurons in the LIP/FEF. The saliency map consists of one population of excitatory and one population of inhibitory neurons with strong lateral excitation and inhibition. Each excitatory cell in V4 projects to both excitatory and inhibitory neurons at its corresponding location in the saliency map, and receives feedback from excitatory neurons in this map (see supplemental Table 2, available at www.jneurosci.org as supplemental material, for model parameters). The saliency map also receives an input equal to 20% of the sum of the inputs to all feature-selective populations in V4 to account for direct inputs from the LGN and input layers of V1 to the LIP/FEF, which is supported by observation of an earlier response onset in the FEF with respect to V2 and V4 (Schmolesky et al., 1998). For simulation of the saccade preparation experiment, we use the same parameters while introducing extra inputs to all populations due to the presence of the saccade target (see below).

Inputs to the network. We use the same visual stimuli as used by Hegdé and Felleman (2003) and Burrows and Moore (2009) to generate the inputs to our model. These visual stimuli consist of arrays of 7 × 7 oriented, colored bars with six different arrangements: singleton (the single bar), homogeneous, color popout, orientation popout, combined popout, and conjunction (supplemental Fig. S1, available at www.jneurosci.org as supplemental material). Popout and conjunction displays contain one colored and oriented bar, the target, which can be distinguished from the rest of the colored and oriented bars, the 48 distractors, by either one or two features, respectively. In the above experiments, preferred and nonpreferred color and oriented bars were determined for each recorded neuron, and then used to construct different stimuli. For convenience, we construct the visual stimuli from four types of bars: green or red and vertical or horizontal. Moreover, we always place the target bar in the center of the array.

All neurons in the network receive two classes of synaptic inputs; background input and feature-selective inputs. The background input represents projections from surrounding cortical neurons and are modeled by Gaussian noise currents. For each excitatory (respectively inhibitory) neuron, this input is equal to the current generated by 1000 cortical neurons firing at 4.0 Hz (respectively 3.0 Hz), through AMPA synapses (with the peak conductance $g_s = 1$ nS). This spontaneous synaptic barrage provides the model with realistic noise and brings neurons near their firing threshold. The feature-selective inputs represent the outputs of color-selective neurons in the LGN, and of orientation-selective neurons in V1 (see below).

Responses to the orientation bars are computed using RF properties of orientation-selective cells in layer 4 of V1 (Dayan and Abbott, 2001).

More specifically, the input to location (x, y) at time t is equal to the following:

$$I_o(x,y,t) = I_o + kL_o(x,y,t), \quad (1)$$

where k is an arbitrary constant, and $L_o(x, y, t)$ is the linear response estimate of neuronal activity in space and time,

$$L_o(x,y,t) = \int_0^t dt' \int_{x_{min}}^{x_{max}} dx' \int_{y_{min}}^{y_{max}} dy' D_o(x',y',t')\, s(x-x', y-y'), \quad (2)$$

where $s(x, y)$ is the visual stimulus at location (x, y) and the kernel $D_o(x', y', t')$ defines the space-time RF of the neuron. As the input is stationary, s does not depend on time but is equal to the average intensity of that stimulus at a given location. The kernel can be decomposed into spatial and temporal RF components (Dayan and Abbott, 2001), as follows:

$$D_o(x',y',t') = D_{o,s}(x',y')\, D_{o,t}(t'). \quad (3)$$

The spatial RF of orientation-selective neurons is modeled with a Gabor function:

$$D_{o,s}(x',y') = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left(-\frac{x'^2}{2\sigma_x^2}-\frac{y'^2}{2\sigma_y^2}\right)\cos(kx'), \quad (4)$$

where σ_x and σ_y determine the extent of the RF, and $k = 1/8°$ is the preferred spatial frequency. In Equation 4, the neuron responds most strongly to 0° orientation. In our model, we employ four types of input neurons selective to different orientations (at 0°, 45°, 90°, 135°). We choose $\sigma_x = 0.5°$ for the nonpreferred direction, and $\sigma_y = 1°$ for the preferred direction. The temporal RF of the orientation-selective neurons is given by the following (Maex and Orban, 1996; Dayan and Abbott, 2001):

$$D_{o,t}(t') = \alpha\exp(-\alpha t')\left(\frac{(\alpha t')^5}{5!} - b_o\frac{(\alpha t')^7}{7!}\right), \quad (5)$$

where $\alpha = 1/(7.5\text{ ms})$ and $b_o = 0.85$. We introduce a bias factor, b_o, to make the integral of $D_{o,t}(t')$ nonzero, avoiding a vanishing response after the initial onset. Because the image input is stationary, the temporal component of the RF response can be integrated independently of the spatial component.

We assume that the color inputs to our network are generated by the response of the most prevalent type of color-selective neurons in the LGN (Ts'o and Gilbert, 1988). That is, the center region of RF has color-opponency while the surround RF is nonchromatic, matching the modified type II neurons in Ts'o and Gilbert (1988). We assume that the form of spatial and temporal RF of these neurons is similar to ON- and OFF-center neurons in the LGN with the kernel,

$$D_c(x',y',t') = \frac{D_{c,t}^{cen}(t')}{2\pi\sigma_{cen}^2}\exp\left(-\frac{x'^2+y'^2}{2\sigma_{cen}^2}\right) - \frac{BD_{c,t}^{sur}(t')}{2\pi\sigma_{sur}^2}\exp\left(-\frac{x'^2+y'^2}{2\sigma_{sur}^2}\right). \quad (6)$$

Here B is a constant which determines the relative contribution of surround to center, $\sigma_{cen} = 0.5°$ and $\sigma_{sur} = 1.0°$ determine the extent of RF in the center and surround, respectively, and $D_{c,t}^{cen}$ (respectively, $D_{c,t}^{sur}$) determines the temporal RF of the center (respectively, surround). We only use ON-center neurons to generate the inputs. The temporal component of the RF for the center and surround are described as follows (Dayan and Abbott, 2001):

$$D_{c,t}^{cen,sur}(t') = \alpha_{cen,sur}^2 t'\exp(-\alpha_{cen,sur}t') - b_c\beta_{cen,sur}^2 t'\exp(-\beta_{cen,sur}t'), \quad (7)$$

where $\alpha_{cen} = 1/(10\text{ ms})$, $\beta_{cen} = 1/(32\text{ ms})$, $\alpha_{sur} = 1/(20\text{ ms})$, $\beta_{sur} = 1/32$ ms), and $b_c = 0.75$ is a constant introduced to avoid the integral of the temporal RF vanishing over time. Because the input is stationary, the temporal component of the RF can be integrated independently of the spatial component.

To simulate red and green bars, the response of neurons with an RF described above is generated using three color components of the input image. The red response equals the center response to the red component minus the center response to green plus the center response to the average of the three components. The surround response is nonchromatic and is equal to the average of the three color components (Ts'o and Gilbert, 1988). This way, neurons selective to red show a strong response to a red bar in their RF, but a weak response to a green bar in their RF (they prefer red and then anything but green) and vice versa.

Simulated V2 and V4 neurons receive color- and/or orientation-selective inputs which are 30% and 15% of the selective inputs to V1 (described above), respectively. These inputs are implemented to account for weak direct inputs from the LGN and input layers of V1 to areas V2 and V4 (Girard and Bullier, 1989; Girard et al., 1991). Note that we obtain similar results even in the absence of these direct inputs to V2 and V4.

Finally, in the saccade preparation experiment of Burrows and Moore (2009) the monkey is cued to make a saccade to a target while visual stimuli are presented in the RF of recorded neurons at a random time before the saccade initiation. To simulate this experiment, we assume that the onset of the saccade target introduces a strong input to the saliency population and a weak input (equal to 20% of the input to the saliency population) to all feature-selective populations, at the location of the saccade target. The strong input to the saliency map is assumed to originate from a working memory network which encodes the location of the saccade target. For simplicity, we assume that the onset of the saccade target is always 50 ms before the onset of the visual stimuli.

Data analysis. For the results presented here, the average response of a simulated neuron to a given visual display is computed by counting all spikes in the 200 ms interval following the onset of activity (defined as firing above 5.0 Hz) and averaging this number over 200 trials. Because neighboring neurons have overlapping RF and their activity is highly correlated due to all-to-all connectivity, we compute these averages over four neurons with overlapping RF (for both target-selective or for distractor-selective neurons). For convenience, we present most of the average responses after normalizing them by the response to the singleton display.

To quantify the saliency computations in successive layers of the network, we compute different quantities. First, we consider the difference between the response to popout and conjunction displays. These differences can be used to define a local saliency measure known as the popout selectivity index (PI) which has been reported in many experimental studies (Knierim and van Essen, 1992; Hegdé and Felleman, 2003; Burrows and Moore, 2009):

$$PI = \frac{R_{popout} - R_{conj}}{R_{popout} + R_{conj}}, \quad (8)$$

where R_{popout} and R_{conj} are the average responses of target-selective neurons to a given popout and conjunction display, respectively.

Second, as the saliency of a target depends on its contrast with nearby objects, the neuronal signature of the saliency of a target should depend on the response of target-selective neurons relative to the response of neurons selective to nearby distractors. Therefore, we use the average response of neurons selective to the target and to all 48 distractors surrounding the target to define a global measure of saliency.

We define the global saliency index (GSI) as the difference in the average response of target-selective neurons (R_{target}) and of all distractor-selective neurons (R_{distr}) divided by their sum, as follows:

$$GSI = \frac{R_{target} - R_{distr}}{R_{target} + R_{distr}}. \quad (9)$$

GSI, distributed between -1 and $+1$, is a measure for how easily the target can be distinguished among the distractors: the closer it is to 1, the larger the neuronal representation of the target relative to the distractors.

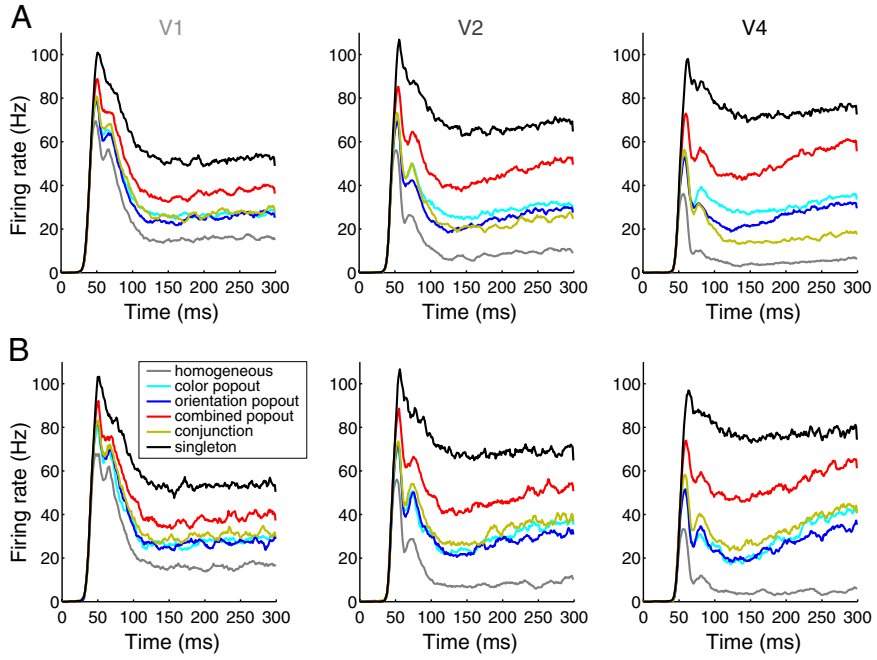

Figure 2. Simulated time course of the response of target-selective neurons in V1, V2 and V4 to the 6 displays. The responses are bounded from above by the response to the singleton (black curves) and from below by the response to the homogeneous array of bars (gray curves). ***A***, Configuration *A*, in which different features are processed independently. ***B***, Configuration *B*, in which combinations of features are jointly represented. The response in each condition is computed by averaging the response of four central neurons with overlapping RF for 200 trials of the simulated experiment. The response of target-selective neurons to the conjunction display falls below the response to popout displays in V4 for configuration *A* but not *B*.

Results

Saliency computations through lateral interactions in successive layers of spiking neurons

To understand the basic mechanisms underlying saliency computations, we examine two exclusive architectures (Fig. 1). In architecture or configuration *A*, individual features are processed in separate neural populations and neurons in different populations receive inputs selective to either orientation or color (Fig. 1*A*). We combine the outputs of such neural populations to construct a neural response selective to both features (see Materials and Methods for more details). In configuration *B*, a combination of features is jointly processed and neurons in different populations receive inputs selective to both orientation and color (Fig. 1*B*).

We first compare the response of neurons selective to the target (the central red-vertical bar) in V1, V2, and V4 and for the two configurations (Fig. 2). Note that the target is the same for all displays, while the surrounding distractors differ. The average response to the six displays are computed over 200 trials of network simulation. The onset of response (defined as firing above 5 Hz) to different stimuli occurs at ~40 ms for neurons in simulated V1, ~48 ms for V2, and ~54 ms for V4.

Soon after the activity onset, the responses to the different displays diverge due to lateral interactions in neural populations. Sometimes a smaller, secondary peak can be observed, a remnant of the shock oscillation caused by all-to-all connectivity. However, this oscillation is damped out quickly and does not contribute to the saliency computations. Because lateral interactions are dominated by inhibition, target-selective neurons show the weakest response to the homogeneous display, and simultaneously the strongest response to the singleton display (Fig. 2, compare black and gray traces). Note that when the singleton display is presented, the distractor-selective neurons receive the smallest inputs and so only weakly inhibit central neurons responding to the target. Yet when the homogeneous field of bars is presented, the distractor-selective neurons receive the largest inputs and so strongly inhibit target-selective neurons.

We also find that the response in V1 for both configurations for color and orientation popout displays (cyan and blue traces, respectively, in Fig. 2) are similar to the response to the conjunction display (yellow), while they are smaller than the response to the combined popout display (red). Thus, while target-selective V1 cells already show differential responses, they do not distinguish between popout and conjunction per se, similar to what has been observed in monkey V1 (see Hegdé and Felleman, 2003, their Fig. 10). As the signal propagates through V2 and V4, the response to the conjunction becomes smaller than the response to the orientation and color popout in configuration *A* but not in *B* (compare yellow traces in Fig. 2*A*,*B*). That is, neurons in higher areas show a differential response to the popout and conjunction displays only when individual features are processed independently.

To quantify the evolution of the saliency signal in successive regions, we next consider the average response to the 6 different displays in successive layers of the network. As expected, we find that the average response to all five arrays of bars are suppressed compared with the singleton for neurons selective to the target (supplemental Fig. S2, available at www.jneurosci.org as supplemental material). Interestingly, the response in V1 is fairly similar for both configurations *A* and *B* and qualitatively matches the experimental results in V1 (Hegdé and Felleman, 2003, their Fig. 5).

The saliency of a target depends on its contrast with nearby objects, here the neighboring distractors. Likewise, the neuronal signature of target saliency should depend on the response of target-selective neurons relative to the response of neurons selective to nearby distractors. Therefore, we further analyze the average response of both target- and distractor-selective neurons in each region and for each configuration (Fig. 3).

When individual features are processed separately, the response to the target in popout displays is reduced by small amounts from one region to the next, which is much smaller than the reduction in the homogeneous and conjunction displays (Fig. 3*A*). As a result, the average response to popout targets exceeds the response to the conjunction target for configuration *A*. This is accompanied by an increase in the difference between the response to the target and distractors in successive layers for popout displays, but not for the conjunction display. On the other hand, the response to the target in both popout and conjunction displays is reduced in configuration *B* which is accompanied by an increase in the difference between the response to the target and distractors in successive layers for these displays.(Fig. 3*B*). The differential response of target-selective neurons to popout and conjunction stimuli can be quantified by the popout selectivity index (Knierim and van Essen, 1992; Burrows and Moore, 2009). We find that V4 popout selectivity indices for config-

uration *A* are qualitatively similar to those measured for V4 neurons in the monkey (Fig. 3*C*; cf. Burrows and Moore, 2009, their Fig. 2*B*).

This differential response in the two configurations is a consequence of the fact that for configuration *B*, lateral interactions take place between neurons selective to combinations of features. Thus, the response of target-selective neurons to either color or orientation popout or to conjunction displays is suppressed by active distractor-selective neurons, while this is not the case in configuration *A*. In the latter, the distractor-selective neurons are active in only one of the two populations for color and orientation popout displays. Consequently, the response to popout and conjunction displays is suppressed differentially for configuration *A* but to the same extent for *B*. This effect is compounded when ascending through the three regions (Fig. 3), and results in activity patterns in V4 similar to experimental observations (Burrows and Moore, 2009), but only for configuration *A*. Therefore, we conclude that saliency computations require independent processing of individual features in successive layers of neurons with lateral interactions.

Although popout selectivity has been used as a measure of saliency signals, it has been argued that the absence of popout selectivity may not be equivalent to the absence of saliency signals (Li 2002). To test this hypothesis we use the average response of neurons selective to the target and all 48 distractors surrounding the target to define the global popout selectivity indices (see supplemental material, available at www.jneurosci.org). Interestingly, we find that local and global popout selectivity indices are highly correlated in all regions (see supplemental Figs. S4, S5, available at www.jneurosci.org as supplemental material). Local and global saliency signals are correlated because they are generated through lateral interactions between neurons with similar selectivity. Therefore, assuming saliency computation is manifested in the brain through the same mechanism, we conclude that both local and global popout indices can be equally informative about the visual saliency of the target.

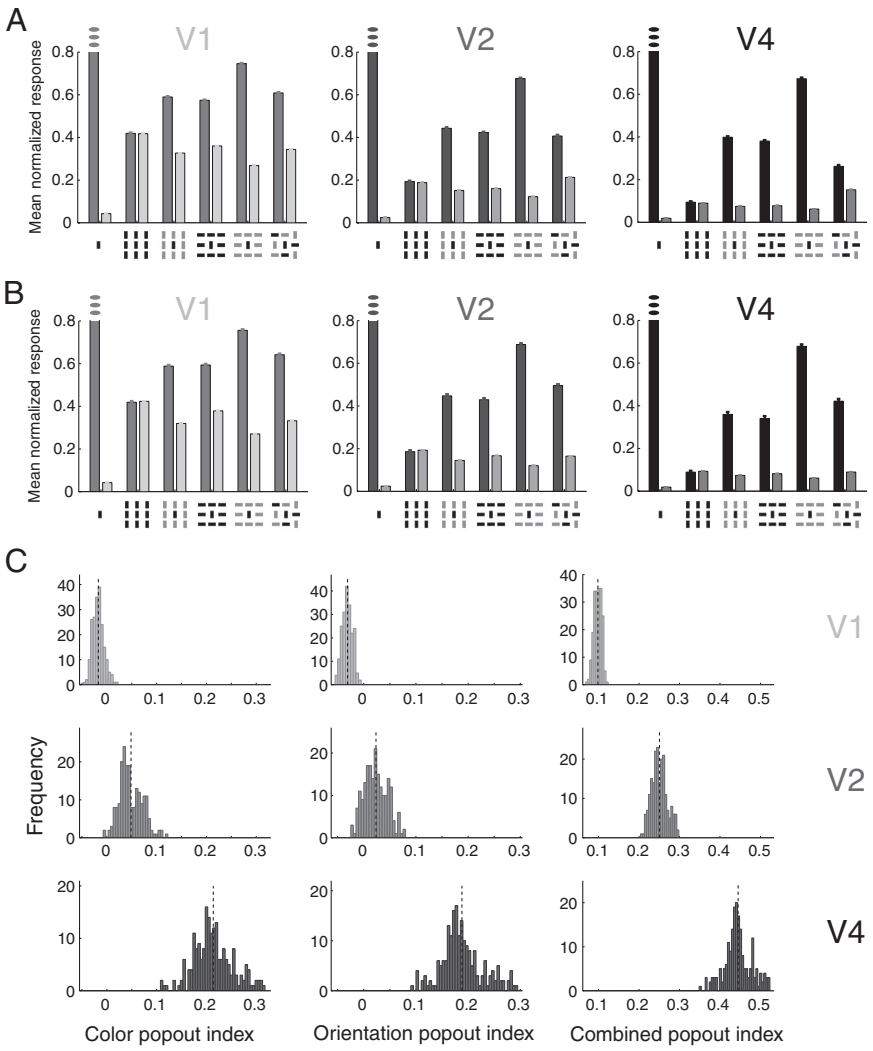

Figure 3. Evolution of response and the local saliency signal in feature-selective neurons in successive layers of the network. ***A***, ***B***, Average normalized response to different types of display is plotted, separately for neurons selective to either the target (central bar) or distractors (peripheral bars). In each panel, the average response of target-selective neurons is plotted with a darker shade than the average response of distractor-selective neurons. The error bars are the SEM. For illustrative purposes, we only show the values of responses between 0 and 0.8 (the response of target-selective neurons to the single bar is equal to 1). Results for configurations *A* and *B* are shown in ***A*** and ***B***, respectively. While in configuration *A* the response to the target decreases slightly, the difference between the response to the target and distractors increases only for popout displays in higher visual areas. In contrast, this difference increases for both popout and conjunction displays in configuration *B*. ***C***, Histograms of popout selectivity indices for popout displays in successive layers of the network. Histograms of popout indices are computed from the average response in the first 200 ms of the visual response in 30 randomly selected trials. Dashed lines show the mean value for each histogram.

Up to this point, we examined the response of target-selective neurons to different displays. We showed how a differential response to popout and conjunction displays is formed in successive layers of neurons through lateral excitatory and inhibitory interactions. The reason for examining the signal in target-selective neurons was to compare our results to electrophysiological recordings, although the presence of this signal is not equivalent to target detection per se (see Discussion). Instead, target selection can be performed by finding the most active location in a topographic map which is not selective to any feature (Itti and Koch, 2001). Therefore, we construct a "hypothetical" saliency map by adding the output of feature-selective neurons in each region to examine the signal related to target detection.

We find that the fictitious saliency neurons in early visual areas show little or no difference in response to the target and distractors. However, for the network architecture *A*, differential responses emerge in higher visual areas for the singleton and popout displays but not for the conjunction display (Fig. 4*A*). More specifically, the global differential response (i.e., the difference between the response of target- and distractor-selective neurons) increases for popout, while it fluctuates around zero for the conjunction display in all three regions (Fig. 4*B*). This is not the case for the network architecture *B*, where the global differential response also increases for the conjunction display (supplemental Figs. S6, S7, available at www.jneurosci.org as supplemental material).

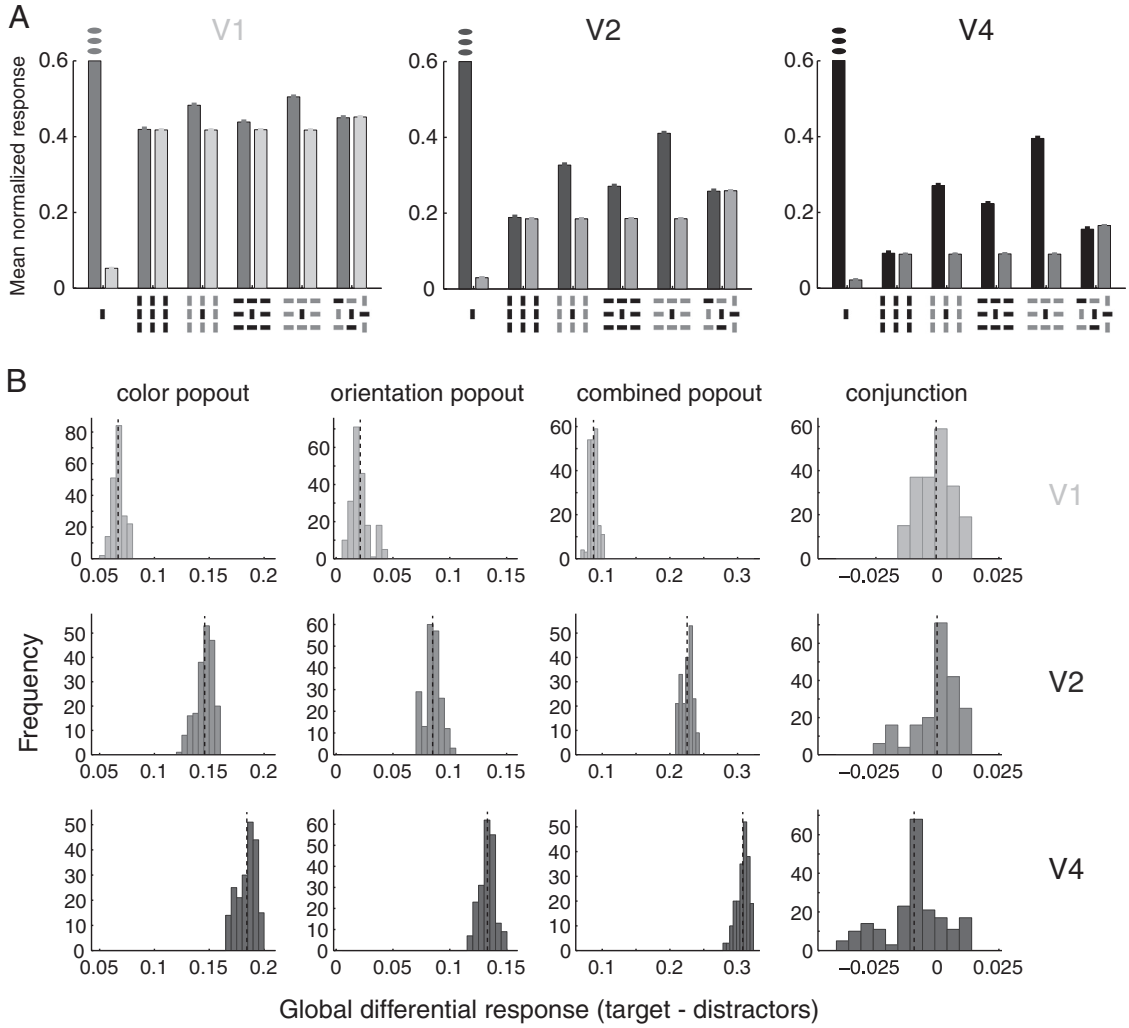

Figure 4. Evolution of the global saliency signal in the constructed saliency populations in successive regions for configuration *A*. ***A***, Average normalized response of the constructed saliency populations to different displays is plotted separately for neurons selective to either the target (central bar) or distractors (peripheral bars). The response of target-selective neurons is plotted with a darker shade than the response of distractor-selective neurons. The error bars are the SEM. For illustrative purposes, we only show the values of responses between 0 and 0.6 (the response to the single bar is equal to 1). ***B***, Histograms of the global differential response (i.e., the difference between response of target- and all distractor-selective neurons), for different displays. Histograms of the differential response are computed from the average response in the first 200 ms of the visual response in 30 randomly selected trials. The dashed lines show the mean in each histogram. The differential response increases for popout displays in successive regions but slightly decreases for the conjunction display.

Compatible with these observations, we find that when individual features are processed in distinct populations of neurons, the detection of popout but not conjunction targets is improved in successive layers in configuration *A* but not *B* (Fig. 5). Overall, these results indicate that a feature-independent saliency signal can be formed by convergence of outputs of different feature-selective neural populations in higher areas of the visual cortex, but this mechanism is effective only if feature processing is kept separated in lower visual areas.

When do local and global saliency signals first emerge? We examined the time course of the local saliency signal by calculating the difference between the activity of target-selective neurons in response to a given popout and conjunction displays: not only does this difference increase but it emerges earlier relative to the activity onset, as signals propagate from V1 to V2 and V4 (Fig. 6*A*).

We likewise examined the time course of the global saliency signal by calculating the difference between the response of target-selective neurons and of the most active distractor-selective neurons on each trial (for neurons in the constructed saliency populations). This difference increases and occurs earlier relative to the response onset in higher regions, but only for popout displays (Fig. 6*B*). Note that at the beginning of a trial before saliency computations are formed through lateral interactions, neurons are mainly driven by external inputs and because of noise many neurons can have higher activity than target-selective neurons. This results in early negative values for global saliency signals in our model.

Role of NMDA in saliency computations

The excitatory currents in our network model is transmitted through two types of synapses, fast AMPA and slow saturating NMDA. Generally, we find that the recurrent excitation should be dominated by NMDA currents but to test the role of these synapses in saliency computations, we reduce NMDA currents while increasing AMPA currents such that the overall recurrent excitation stays at the same level (see Materials and Methods for more details).

We find that increasing the AMPA to NMDA current ratio disrupts the formation of both local and global saliency signals (Fig. 7). That is, the response to popout displays is not different

from the response to the conjunction display in higher visual areas (Fig. 7*A*), and similarly, the difference in response of target- and distractor-selective neurons in the constructed saliency populations is reduced (Fig. 7*B*).

Moreover, as the NMDA currents are reduced, the amount of increase in the local and global differential response in successive regions is also reduced. Similarly we find that the probability of maximum response at the target location is reduced, especially for color and orientation popout displays (supplemental Fig. S8, available at www.jneurosci.org as supplemental material). Finally, we compute the temporal dynamics of the saliency signal for different values of the AMPA to NMDA current ratio. We find that an increase in the AMPA to NMDA current ratio delays and further eliminates the formation of the saliency signal in higher visual areas and moreover, introduces a strong oscillation in the response in these regions (supplemental Figs. S9, S10, available at www.jneurosci.org as supplemental material).

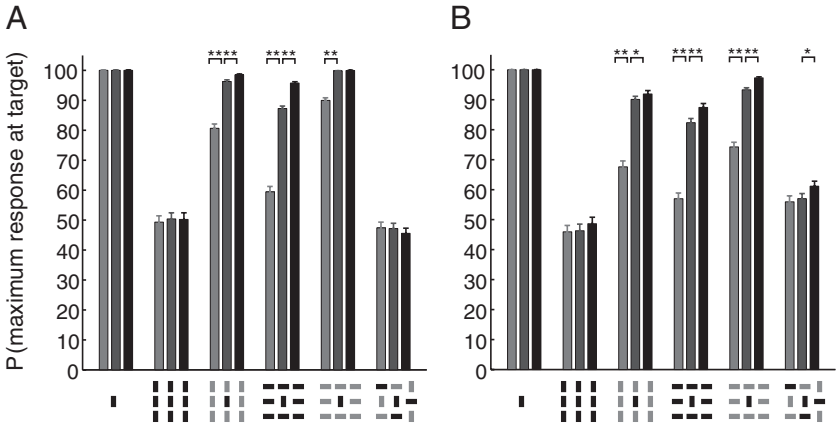

Figure 5. Improvement in target detection in successive layers of the network. The probability that the response of target-selective neurons is the maximum response (*P*(maximum response at target)) in the constructed saliency population is plotted with different shades of gray corresponding to three regions for different displays: light gray (V1), medium gray (V2), and black (V4). This probability is computed as the fraction of distractor-selective neurons which show smaller response than the target-selective neurons on each trial. The error bars are the SEM. One (respectively two) asterisk shows the statistical test that the described probability increases from one region to the next is significant at $p < 0.01$ (respectively, $p < 0.001$). Results for configurations *A* and *B* are shown in ***A*** and ***B***, respectively. In configuration *A*, *P*(maximum response at target) for popout displays reaches to 100% in the third layer, while this quantity is not different from 50% for the conjunction display. On the other hand for configuration *B*, *P*(maximum response at target) does not reach to 100% for color and orientation popout displays and furthermore, it increases in higher visual areas for the conjunction display.

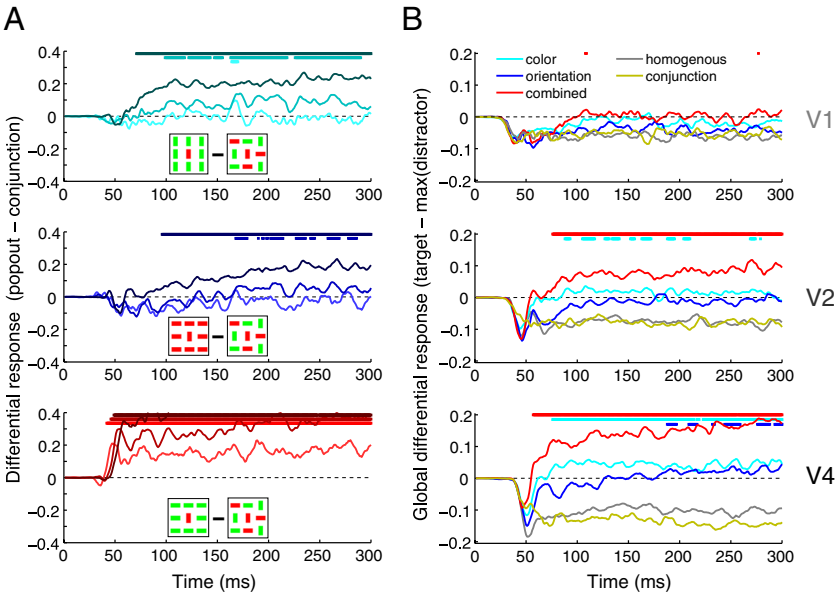

Figure 6. Temporal dynamics of local and global saliency signals. ***A***, The difference in the normalized response to the popout and conjunction displays is plotted for different types of popout displays (indicated by insets). Different shades of color (light to dark) correspond to responses in successive regions (V1 to V4, respectively). ***B***, The difference in the normalized response of the constructed saliency map neurons selective to the target and of neurons selective to the distractor that exhibit the maximum activity on a given trial is plotted for different types of displays and in three regions. The point at the top of each panel shows whether the differential response is statistically significant (at $p < 0.01$) for each 10 ms time interval. Both local and global saliency signals are larger in higher visual areas and emerge earlier in these areas relative to the response onset in these areas.

An explicit saliency map and its action onto earlier stages

To study the effect of feedback from the saliency map on the saliency computations in lower visual areas, we use a smaller version of our model with only two layers of neural populations corresponding to neurons in V4 and LIP/FEF (due to computational expenses). As we show in the previous sections, the saliency signal can be formed in successive layers of neurons when individual features are processed separately. Therefore, here we use the same architecture for neurons in V4 (see Materials and Methods for more details).

We first assure ourselves that approximate saliency signals can form in a single simulated cortical area (compared with three as in the above simulations). We use a stronger and more widely projecting regional connectivity matrix than before (compare model parameters in supplemental Tables 1, 2, available at www.jneurosci.org as supplemental material) to reproduce our basic result: the response of color/orientation-selective neurons to all popout displays is larger than the response to the conjunction display (Fig. 8*A*). Nevertheless, we observe a small positive but significant global saliency index for the conjunction display, which does not appear in saliency computations in successive regions of V1, V2 and V4 (compare Fig. 8 and results for V4 in Fig. 4*B*).

Because the input to the saliency map is the sum of the outputs of feature-selective neurons, this input carries a smaller saliency signal than the population of neurons selective to the target. That is, the global differential response is larger in the V4 population selective to red-vertical bars than in the saliency map (compare the response to the target and distractors for each case in Fig. 8*A*). We will return to this issue in the Discussion.

To examine the result of lateral interactions in the saliency populations, we then compute *GSI* for the synaptic inputs that originate from V4 cells (to the saliency map) and for the response

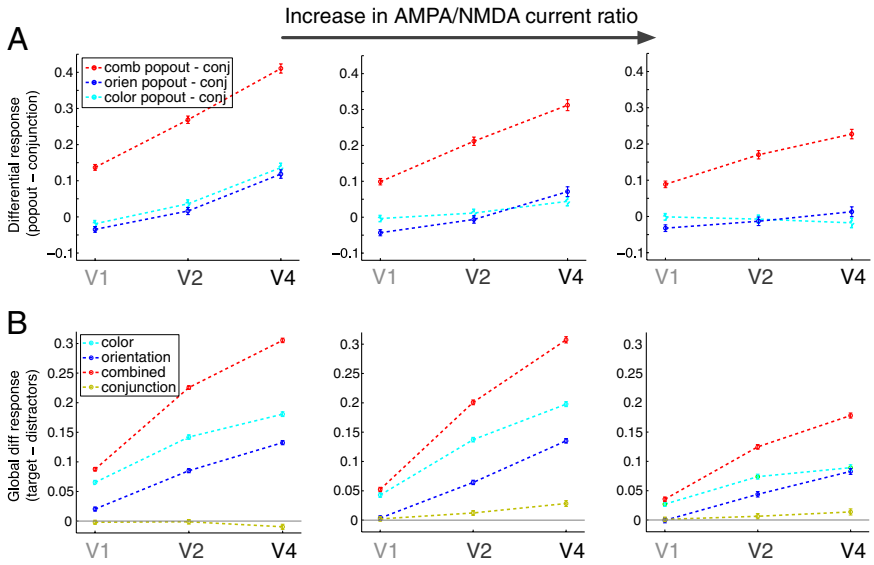

Figure 7. Local and global saliency signal formations are disrupted by increasing the AMPA to NMDA current ratio. ***A***, The difference between average response to different popout displays and to the conjunction display, for different AMPA to NMDA current ratios. The error bars are the SEM. ***B***, The average global differential response in the constructed saliency populations (i.e., difference between response of target- and distractor-selective neurons), for different types of displays and in successive regions of the network. As the AMPA to NMDA current ratio increases, the enhancement of both local and global saliency signals in successive regions is reduced or even diminished.

of neurons in the saliency map. As expected, we find that the lateral interactions within the saliency map enhance the *GSI* and therefore enhance the saliency signal (Fig. 8*B*). Note that the observed decrease in the normalized response in the saliency map population (compared with its inputs) is due to the large response to the singleton display in this population.

We next introduce feedback from the excitatory neurons in the saliency map to all excitatory and inhibitory populations in V4 at corresponding locations. These two types of feedback are adjusted such that they approximately modulate the response of the feature-selective neurons rather than driving them (Schwabe et al., 2006). Feedback differentially changes the response to each display, increasing popout selectivity while increasing variability (Fig. 9).

Finally, how does this model act in the presence of a top-down signal, as in the saccade preparation experiment of Burrows and Moore (2009)? In this experiment, the monkey was cued to make a saccade to a location far from the RF of the recorded V4 neuron, and was rewarded a drop of juice for making saccade to the cued location after the fixation spot disappeared. While planning such a saccade, a visual stimulus was presented at a random time before the initiation of the saccade. They found that under this manipulation, the observed differential response of V4 neurons to the popout and conjunction displays in the control passive viewing experiment, was eliminated. We hypothesize that interaction between bottom-up and top-down attentional signals occurs in the saliency map (LIP or FEF), where both signals are found (Thompson et al., 2005; Balan and Gottlieb, 2006; Ipata et al., 2006; Buschman and Miller, 2007). More specifically, we assume that saccade preparation induces an activity within the saliency map at the saccade target location. This corresponds to the introduction a highly salient target at that location through working memory, thereby altering feedback from the saliency map to earlier areas.

To test this hypothesis, we simulate the saccade preparation task by adding excitatory inputs to all populations at locations corresponding to the saccade target (to mimic working memory inputs). These inputs are strong enough so there is a representation of the saccade target in the saliency map on all trials, but are weak enough so they do not alter the response to the singleton display.

During the simulated saccade preparation the response to different displays changes slightly and differentially (Fig. 9), such that the popout selectivity indices for popout are reduced, in line with the experimental data (Burrows and Moore, 2009). That is, saccade preparation reduces the response to popout displays more so than the response to the conjunction display. This happens because for popout displays, the target provokes a strong response in the saliency map (Fig. 8*A*) which in turn results in a strong feedback to feature-selective neurons. The converse is true for the conjunction target. During saccade preparation, the saccade target also invokes a strong response in the saliency population which reduces the response to the target in this population through lateral inhibition (supplemental Fig. S11, available at www.jneurosci.org as supplemental material) and consequently, the response to the target in feature-selective neurons. Therefore, the amount of reduction in the response of feature-selective neurons depends on the target response in the saliency map during the control trials, and on the response to the saccade target in this map.

Discussion

Despite a large body of literature on the psychology of bottom-up attention and how it operates within visual scenes, much less is known about its neural substrates. Here we design a biophysically plausible spiking network model to investigate the representation and formation of saliency signals in the visual cortex and its interaction with top-down attention.

We focus on lateral excitatory and inhibitory interactions as the main mechanism for saliency computations. By comparing two distinct network architectures, we find that local and global saliency signals emerge and increase in successive layers of neural populations only if individual features are processed in different neural populations (configuration *A*). That is, while the activity of target-selective neurons in the first visual area of our model (V1) does not discriminate between popout and conjunction displays, neurons in higher areas of the model (V2 and V4) show stronger response to popout than to conjunction displays, similar to experimental observations in V1 and V4, respectively (Hegdé and Felleman, 2003; Burrows and Moore, 2009). Moreover, the difference between the response to the target and distractors, as well as target detectability, increases in successive layers for popout but not for conjunction displays, compatible with the basic difference in detection of popout and conjunction targets (Treisman and Sato, 1990).

Similar to experimental data (Burrows and Moore, 2009) we obtain larger local and global saliency signals for the combined popout than for single popout displays; that is, the detection of the popout target is easier when it differs from distractor in two features rather than one feature. Our finding is also consistent with the so-called "redundant-signal effect" (shortening of the reaction time when the response is triggered by two rather than

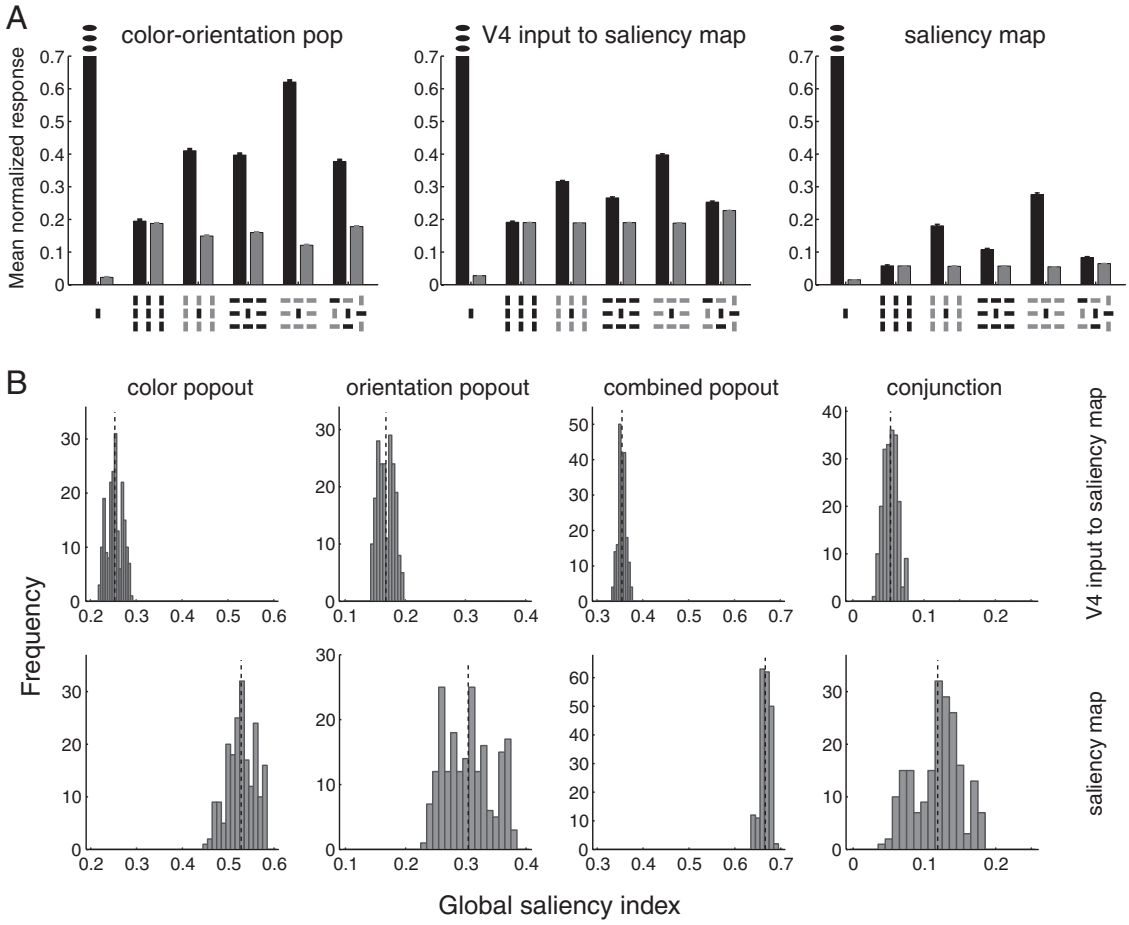

Figure 8. Formation of the saliency signal in a single, simulated cortical region V4 with stronger and wider lateral interactions, and in the saliency map which receives the output of these V4 neurons. ***A***, Average normalized response to different displays is plotted separately for neurons selective to either the target (black) or distractors (gray) in different populations: color/orientation-selective population, input to the saliency map from V4, and the saliency map population. The measured color-orientation neurons are selective to red-vertical bars, and the error bars show the SEM. For illustrative purposes, we only show the values of responses between 0 and 0.7 (the response to the single bar is equal to 1). ***B***, Histograms of the global saliency indices (*GSI*) for different displays are plotted for the input from V4 into the saliency map and for neurons in the saliency map. Dashed lines show the mean of the global saliency index for each display. The effect of lateral interactions in the saliency map is to increase the mean and variance of the global saliency indices.

one response-related target signal) demonstrated in a visual popout search task. This effect has been attributed to coactivation of different visual pathways and their subsequent convergence before response triggering (Krummenacher et al., 2001, 2002, 2010; Zehetleitner et al., 2008; Töllner et al., 2010). Similarly, in our model saliency signals for the combined popout is stronger because of independent input processing related to different features and their parallel processing in separate neural populations before convergence in the saliency map.

As an alternative to saliency computations in successive layers, we also consider wider and stronger lateral interactions in one layer of neural populations which process individual features separately. Even though we observe local saliency signals (i.e., positive popout selectivity indices), this signal was very small for orientation popout display (Fig. 9*B*). Moreover, this mechanism results in a small positive but significant global saliency index for the conjunction display (Fig. 8) and in a noisier detection of the target (data not shown).

Therefore, we conclude that moderate lateral interactions in successive layers of neurons selective to individual features provide a suitable mechanism for early saliency computations. Furthermore, neurons that process individual features separately are more likely to contribute to bottom-up saliency than neurons that are simultaneously selective to both color and orientation (Livingstone and Hubel, 1987).

Our saliency computations can be compared with the standard Itti-Koch computational saliency model (Koch and Ullman, 1985; Itti et al., 1998; Itti and Koch, 2001). The saliency model exploits center-surround computations (i.e., subtracting a filtered image at a lower spatial resolution from the image at a higher spatial resolution) to capture local feature contrasts in the image and to form feature maps, as well as normalization to enhance (respectively suppress) those maps with a few (respectively many) active locations. Lateral excitation and inhibition between neural populations enables our model to approximately perform center-surround and normalization computations without using a multi-resolution representation of an input image at different scales. An important requirement for this to happen is that inhibitory connections should be wider than excitatory connections.

Normalization of sensory inputs by the sum of inputs has been used in a few models of top-down attention (Reynolds et al., 1999; Lee and Maunsell, 2009; Reynolds and Heeger, 2009). For example, Reynolds and Heeger (2009) proposed that top-down attention improves sensitivity to faint stimulus through multiplicative interaction of inputs and the "attention field," followed by

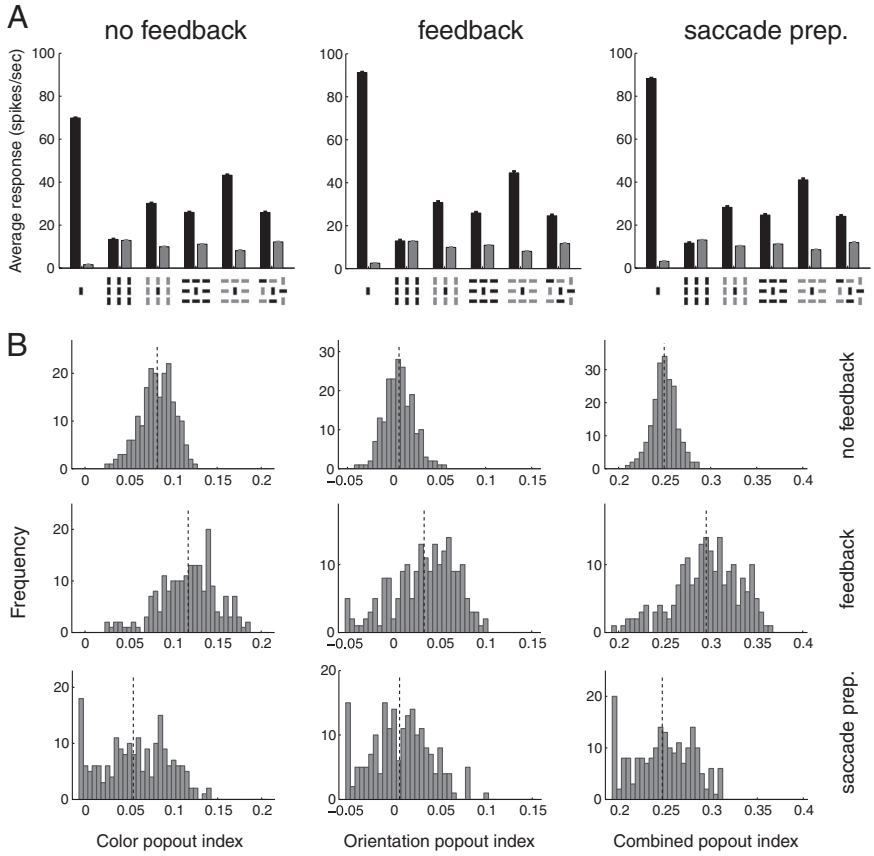

Figure 9. The effects of feedback from the saliency map and saccade preparation on saliency computations. ***A***, Average response to different displays is plotted separately for neurons selective to either the target (black) or distractors (gray) in three different cases: with no feedback, with feedback from the saliency map to all feature-selective populations, and simulated saccade preparation experiment. The measured neurons are selective to red-vertical bars, and the error bars show the SEM. ***B***, Histograms of the popout selectivity index for three different types of popout in three cases described above. The dashed lines show the mean of the popout index for each histogram. All popout selectivity indices are increased due to feedback ($p < 0.0001$) and are then decreased during the saccade preparation ($p < 0.0001$).

normalization of the response by the activity in the "suppressive field." Lateral interactions proposed here can provide a biophysical mechanism for normalization due to the suppressive field in this model. Alternatively, Lee and Maunsell (2009) proposed that top-down attention affects neural response solely through changing the strength of normalization and not the inputs. If such a mechanism is implemented through lateral interactions, then top-down attention should mostly affect the activity in inhibitory neurons to change the strength of normalization.

While lateral interactions between spiking neurons through realistic synapses can approximate center-surround and normalization computations, these computations are limited by biophysical properties of neural elements in the network. This happens because neurons in the network should integrate, in every spike, the noisy background inputs plus excitatory and inhibitory inputs from neighboring neurons and subsequently, transmit this signal to other neurons in the network through realistic synapses with different time constants. Due to these factors, some conditions should be met in order for the network to perform efficient center-surround and normalization computations. First, we find that recurrent excitations between excitatory neurons should be dominated by slow NMDA currents. This enables neurons to integrate the noisy input signals over a longer timescale and to increase the signal-to-noise ratio. Second, excitatory to inhibitory connections, which drive the lateral inhibition through inhibitory interneurons, should have both AMPA and NMDA currents. This is because if these connections are also dominated by NMDA currents, the inhibitory interneurons become active slowly which can suppress the activity in the network after the response onset. These results emphasize the crucial role of the NMDA receptor in the saliency computations, similar to its role in working memory and decision making (Wang, 1999, 2002; Compte et al., 2000). Based on our results, we predict that inactivation of NMDA receptors in the visual cortex results in a nosier and weaker saliency signal, and impairs the performance in the visual search task.

Although we find that recurrent excitation in our spiking network should be dominated by slow NMDA currents, this does not translate to slow emergence of the saliency signal. Interestingly, we find that both local and global saliency signals emerge earlier in successive regions of the model (relative to the response onset in each region). Similarly, Buffalo and colleagues recently showed that top-down attentional effects appear earlier and stronger in V4 than in V2 and V1 (Buffalo et al., 2010). This result may also explain how the saliency signal can appear in higher visual areas such as the LIP and FEF even before the appearance of this signal in striate and extrastriate cortices. For example, compare the emergence of this signal in V4 (Burrows and Moore, 2009) with the LIP and FEF (Buschman and Miller, 2007). Interestingly, Buschman and Miller (2007) found an earlier saliency signal in the LIP than in the FEF during popout search, while this signal emerges earlier in the FEF during a visual search task which requires top-down attention (but see Schall et al., 2007). These observations suggest different roles for the LIP and FEF in saliency computations, a topic which requires further investigations.

V1 has been implicated as the site of early saliency signals (Li, 2002) because V1 neurons are influenced by the stimulation of regions outside their classical RF in a nonlinear fashion (DeAngelis et al., 1994; Albright and Stoner, 2002; Cavanaugh et al., 2002a,b). For example, activity of V1 neurons in the alert monkey is only weakly suppressed (with respect to the singleton display) when the surrounding bars are in orientation perpendicular to the orientation of the central bar (Knierim and van Essen, 1992). However, an experiment specifically designed to detect the presence of saliency signals in monkey V1 came up empty handed; Hegdé and Felleman (2003) found that V1 neurons do not distinguish between popout and conjunction displays; rather, they signal the existence of center-surround discontinuity. Similarly, we find that moderate lateral interactions between neurons with similar feature selectivity can result in a response pattern which depends on the display but does not distinguish between popout and conjunction. In addition, we show that the absence of local saliency signals is indicative of the absence of global saliency signals.

Conversely, single cell data indicates that V4 neurons are selective to bottom-up attentional signals such as popout display, and are modulated by top-down attention as well as the activity of neurons in LIP and FEF (Schiller and Lee, 1991; Hupé et al., 1998; Reynolds et al., 2000; Tolias et al., 2001; Moore and Armstrong, 2003; Reynolds and Desimone, 2003; Armstrong et al., 2006; Armstrong and Moore, 2007; Gregoriou et al., 2009). Compelling evidence for the presence of saliency signals in V4 is found by Burrows and Moore (2009). They demonstrated that V4 neurons, considered as a single population, respond stronger to popout than to conjunction displays. Furthermore, this difference is eliminated when the monkey prepares a saccade to a location far from the RF of the recorded neuron, which indicates that this bottom-up saliency signal is influenced by top-down attentional signals. There is converging evidence that these signals possibly originate in the FEF (Moore and Armstrong, 2003; Armstrong et al., 2006; Armstrong and Moore, 2007; Monosov et al., 2008; Gregoriou et al., 2009). We find in our model that feedback from the saliency map to earlier regions enhances the saliency signal while saccade preparation reduces this signal by altering the feedback.

Interestingly, our model predicts that the effect of saccade preparation on the response to a given stimulus depends on the response of target-selective neurons in the saliency map during the control passive viewing task. This prediction can be tested by recording from neurons in the saliency map (e.g., LIP/FEF) and in feature-selective populations which receive feedback from the saliency map (e.g., V4). Such recording can be used to compute the correlation between the responses of neurons in the saliency map and the reduction in the popout selectivity index for different displays in the control and saccade preparation tasks, respectively. We predict a positive correlation between these two quantities.

By combining the outputs of neural populations which process individual features (in configuration *A*) we construct different color/orientation-selective neural populations. Among these populations the one which is selective to the target features carries a saliency signal stronger than the signal in the saliency population (Fig. 8). So why should the brain bother with a distinct saliency map in the first place? However, in the absence of an explicit saliency map, the brain needs to detect the target by first identifying this feature-selective population. This is of course quite difficult in dense natural scenes with many, partially occluded targets, which is why the strategy of a saliency map that labels the sites of potential objects is an attractive computational option. Similarly, feedback to neural populations in V1 from V2 or V4 populations with similar selectivity does not improve saliency signals as this feedback does not contain any information about the most salient location in the visual scene, and only acts as a stronger recurrent input from the same area. Instead, feedback from the saliency map improves saliency signal as it can enhance the signal in neurons selective to the most salient location. Finally, a feature-independent saliency map, formed by the convergence of outputs of neural populations selective to different features, is consistent with the observation that the saliency signal in the FEF appears in the spiking activity before the local field potential (Monosov et al., 2008).

Electrophysiological evidence suggests that a saliency map is instantiated in the response of neurons in the posterior parietal area 7a (Constantinidis and Steinmetz, 2001, 2005), area LIP (Gottliebet al., 1998; Kusunoki et al., 2000; Bisley and Goldberg, 2003), and in the FEF (Thompson and Bichot, 2005). Interestingly, in such a setting the activity of the saliency population and also the detection of the target can be adjusted by gating the inputs from different feature-selective populations (Rutishauser and Koch, 2007) and by top-down signals. To model saccade preparation, we assume that saliency neurons selective to the location of the saccade target become active and stay active during the stimulus presentation (at a fixed level of activity). Conceivably, trial-by-trial variability in the representation of the saccade target can alter the feedback to V4 neurons and consequently, the saliency signal. Such variability has been observed in areas LIP and FEF and was shown to be correlated with monkeys ability to ignore distractors (Thompson et al., 2005; Balan and Gottlieb, 2006; Ipata et al., 2006).

At the end, while it has been shown that saliency computations can be easily performed through center-surround and normalization computations (Itti et al., 1998), we find biophysical limits for performing these computations by spiking neurons and realistic synapses. These limits point to the general biophysical mechanisms which are used in other parts of the brain. Namely, due to response variability of cortical neurons, the integration of input signals need to be done through slow NMDA synapses and in successive layers of neural populations. Computation in successive layers of neural populations results in earlier emergence of the saliency signal in higher visual areas, which in turn can provide feedback to lower visual areas and improve the saliency computations.

References

Albright TD, Stoner GR (2002) Contextual influences on visual processing. Annu Rev Neurosci 25:339–379.

Allman J, Miezin F, McGuinness E (1985) Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. Annu Rev Neurosci 8:407–430.

Armstrong KM, Moore T (2007) Rapid enhancement of visual cortical response discriminability by microstimulation of the frontal eye field. Proc Natl Acad Sci U S A 104:9499–9504.

Armstrong KM, Fitzgerald JK, Moore T (2006) Changes in visual receptive fields with microstimulation of frontal cortex. Neuron 50:791–798.

Balan PF, Gottlieb J (2006) Integration of exogenous input into a dynamic salience map revealed by perturbing attention. J Neurosci 26:9239–9249.

Bisley JW, Goldberg ME (2003) Neuronal activity in the lateral intra parietal area and spatial attention. Science 299:81–86.

Buffalo EA, Fries P, Landman R, Liang H, Desimone R (2010) A backward progression of attentional effects in the ventral stream. Proc Natl Acad Sci U S A 107:361–365.

Burrows BE, Moore T (2009) Influence and limitations of popout in the selection of salient visual stimuli by area V4 neurons. J Neurosci 29: 15169–15177.

Buschman TJ, Miller EK (2007) Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science 315:1860–1862.

Cavanaugh JR, Bair W, Movshon JA (2002a) Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. J Neurophysiol 88:2530–2546.

Cavanaugh JR, Bair W, Movshon JA (2002b) Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons. J Neurophysiol 88:2547–2556.

Cerf M, Harel J, Einhäuser W, Koch C (2008) Predicting human gaze using low-level saliency combined with face detection. Adv Neur Inf Proc Sys 20:241–248.

Compte A, Brunel N, Goldman-Rakic PS, Wang XJ (2000) Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. Cereb Cortex 10:910–923.

Constantinidis C, Steinmetz MA (2001) Neuronal responses in area 7a to multiple-stimulus displays: I. Neurons encode the location of the salient stimulus. Cereb Cortex 11:581–591.

Constantinidis C, Steinmetz MA (2005) Posterior parietal cortex automatically encodes the location of salient stimuli. J Neurosci 25:233–238.

Dayan P, Abbott LF (2001) Theoretical neuroscience: computational and mathematical modeling of neural systems. Cambridge, MA: MIT.

DeAngelis GC, Freeman RD, Ohzawa I (1994) Length and width tuning of neurons in the cat's primary visual cortex. J Neurophysiol 71:347–374.
Foulsham T, Underwood G (2008) What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. J Vis 8:6.1–6.17.
Girard P, Bullier J (1989) Visual activity in area V2 during reversible inactivation of area 17 in the macaque monkey. J Neurophysiol 62:1287–1302.
Girard P, Salin PA, Bullier J (1991) Visual activity in macaque area V4 depends on area 17 input. Neuroreport 2:81–84.
Gottlieb JP, Kusunoki M, Goldberg ME (1998) The representation of visual salience in monkey parietal cortex. Nature 391:481–484.
Gregoriou GG, Gotts SJ, Zhou H, Desimone R (2009) High-frequency, long-range coupling between prefrontal and visual cortex during attention. Science 324:1207–1210.
Hansel D, Mato G, Meunier C, Neltner L (1998) On numerical simulations of integrate-and-fire neural networks. Neural Comp 10:467–483.
Hegdé J, Felleman DJ (2003) How selective are V1 cells for pop-out stimuli? J Neurosci 23:9968–9980.
Hupé JM, James AC, Payne BR, Lomber SG, Girard P, Bullier J (1998) Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. Nature 394:784–787.
Ipata AE, Gee AL, Gottlieb J, Bisley JW, Goldberg ME (2006) LIP responses to a popout stimulus are reduced if it is overtly ignored. Nat Neurosci 9:1071–1076.
Itti L, Koch C (2001) Computational modelling of visual attention. Nat Rev Neurosci 2:194–203.
Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Patt Anal Mach Intell 20:1254–1259.
Knierim JJ, van Essen DC (1992) Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. J Neurophysiol 67:961–980.
Koch C, Ullman S (1985) Shifts in selective visual attention: towards the underlying neural circuitry. Hum Neurobiol 4:219–227.
Krummenacher J, Müller HJ, Heller D (2001) Visual search for dimensionally redundant pop-out targets: evidence for parallel-coactive processing of dimensions. Percept Psychophys 63:901–917.
Krummenacher J, Müller HJ, Heller D (2002) Visual search for dimensionally redundant pop-out targets: parallel-coactive processing of dimensions is location specific. J Exp Psychol Hum Percept Perform 28: 1303–1322.
Krummenacher J, Grubert A, Müller HJ (2010) Inter-trial and redundant-signals effects in visual search and discrimination tasks: separable pre-attentive and post-selective effects. Vision Res 50:1382–1395.
Kusunoki M, Gottlieb J, Goldberg ME (2000) The lateral intraparietal area as a salience map: the representation of abrupt onset, stimulus motion, and task relevance. Vision Res 40:1459–1468.
Lee J, Maunsell JH (2009) A normalization model of attentional modulation of single unit responses. PLoS One 4:e4651.
Li Z (2002) A saliency map in primary visual cortex. Trends Cogn Sci 6:9–16.
Livingstone MS, Hubel DH (1987) Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. J Neurosci 7:3416–3468.
Maex R, Orban GA (1996) Model circuit of spiking neurons generating directional selectivity in simple cells. J Neurophysiol 75:1515–1545.
Mannan SK, Kennard C, Husain M (2009) The role of visual salience in directing eye movements in visual object agnosia. Cur Biol 19:R247–R248.
Monosov IE, Trageser JC, Thompson KG (2008) Measurements of simultaneously recorded spiking activity and local field potentials suggest that spatial selection emerges in the frontal eye field. Neuron 57:614–625.
Moore T, Armstrong KM (2003) Selective gating of visual signals by microstimulation of frontal cortex. Nature 421:370–373.
Parkhurst D, Law K, Niebur E (2002) Modeling the role of salience in the allocation of overt visual attention. Vision Res 42:107–123.
Reynolds JH, Desimone R (2003) Interacting roles of attention and visual salience in V4. Neuron 37:853–863.
Reynolds JH, Heeger DJ (2009) The normalization model of attention. Neuron: 61:168–185.
Reynolds JH, Chelazzi L, Desimone R (1999) Competitive mechanisms subserve attention in macaque areas v2 and v4. J Neurosci 19:1736–1753.
Reynolds JH, Pasternak T, Desimone R (2000) Attention increases sensitivity of V4 neurons. Neuron 26:703–714.
Rutishauser U, Koch C (2007) Probabilistic modeling of eye movement data during conjunction search via feature-based attention. J Vis 7:1–20.
Schall JD, Paré M, Woodman GF (2007) Comment on 'Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices'. Science 318:44 [author reply 44].
Schiller PH, Lee K (1991) The role of the primate extrastriate area V4 in vision. Science 251:1251–1253.
Schmolesky MT, Wang Y, Hanes DP, Thompson KG, Leutgeb S, Schall JD, Leventhal AG (1998) Signal timing across the macaque visual system. J Neurophysiol 79:3272–3278.
Schwabe L, Obermayer K, Angelucci A, Bressloff PC (2006) The role of feedback in shaping the extra-classical receptive field of cortical neurons: a recurrent network model. J Neurosci 26:9117–9129.
Thompson KG, Bichot NP (2005) A visual salience map in the primate frontal eye field. Prog Brain Res 147:251–262.
Thompson KG, Bichot NP, Sato TR (2005) Frontal eye field activity before visual search errors reveals the integration of bottom-up and top-down salience. J Neurophysiol 93:337–351.
Tolias AS, Moore T, Smirnakis SM, Tehovnik EJ, Siapas AG, Schiller PH (2001) Eye movements modulate visual receptive fields of V4 neurons. Neuron 29:757–767.
Töllner T, Zehetleitner M, Krummenacher J, Müller HJ (2010) Perceptual basis of redundancy gains in visual pop-out search. J Cogn Neurosci. Advance online publication. Retrieved May 25, 2010. doi:10.1162/jocn.2010.21422.
Treisman A, Sato S (1990) Conjunction search revisited. J Exp Psychol Hum Percept Perform 16:459–478.
Ts'o DY, Gilbert CD (1988) The organization of chromatic and spatial interactions in the primate striate cortex. J Neurosci 8:1712–1727.
Wang XJ (1999) Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. J Neurosci 19:9587–9603.
Wang XJ (2002) Probabilistic decision making by slow reverberation in cortical circuits. Neuron 36:955–968.
Zehetleitner M, Muller HJ, Krummenacher J (2008) The redundant-signals paradigm and preattentive visual processing. Front Biosci 13:5279–5293.