

Activity–Gating Attentional Networks

Julian Eggert, Jan van Hemmen

2001

Preprint:

This is an accepted article published in Models of {N}eural {N}etworks IV. The final authenticated version is available online at: [https://doi.org/\[DOI not available\]](https://doi.org/[DOI not available])

Activity–Gating Attentional Networks

J. Eggert
J. L. van Hemmen

ABSTRACT In the visual system, “attention” selectively enhances and expedites the processing of a subset of the available stimuli vs. the rest. Attention can be directed to many different feature dimensions, such as location, form, color, texture and direction of movement. In this work, we present a model of attentional processing that makes extensive use of the feedforward, lateral and feedback connections known to exist in the visual cortex. The model uses local modulations of the activity of neuronal ensembles to superpose additional saliency and attentional information on top of the sensory data. The additional signals “gate” the information through the entire network and trigger response competition, resulting in an attentional concentration of the processing resources. At the network level, the model consists of two complementary information counterstreams that process separately sensory and attentional data: a sensory, feedforward stream directly analyses the features available in the stimulus, while an attentional stream provides expectations and global hypotheses about the stimulus. We explain the function of such a network as a hypothesis generating and confirming system. We also explain the architecture, components and dynamics necessary for the implementation of such an activity–gating network. The goal is to arrive at a consistent and unified model of attentional processing in the visual system that explains the different types of attention within a single framework.

1 Introduction

1.1 Different Types of Attention

The visual cerebral cortex of primates is composed of many distinct, functionally specialized processing areas. They can be roughly grouped into a hierarchy containing multiple levels that represent increasingly complex information about the visual scene (see e.g. [13, 53]). Although the different areas are often mainly dedicated to process visual information conveyed by feedforward connections from retinal sources, there is a large amount of evidence that neurons in the visual cortex can also be affected by extraretinal

or indirect visual influences.

“Attention” in the visual cortex can be characterized as a very special indirect influence. It acts in a *modulatory* way, in the sense that it enhances and/or suppresses already existing sensory information, and, in addition, this modulation is *selective*, meaning that at any single moment it acts only on a restricted portion of the visual scene [3, 12, 22, 23, 36]. Therefore the name “selective attention”.

Selective attention in the visual system is frequently implemented in the spatial domain, using a geometrically well-defined region of the visual field, the so-called “focus of attention” (FOA) [5, 12, 45] (see also the chapter by Niebur et al. in this volume). Models based on the FOA idea serve to explain some of the experimental evidence found about locationally guided attention.

In addition, there has been found abundant experimental evidence that attentional processing is not restricted to a purely spatial domain, but that there exist many different nonspatial, feature-based attentional mechanisms (with ‘feature-based’ meaning here any stimulus dimensions other than exclusively spatial ones). Attention can be directed to form features, color, texture or motion, to name only a few known.

1.2 Why Attentional Processing at All?

Why is attention necessary at all? Information processing in the visual cortex is organized along several hierarchically organized pathways. Along with the information flow in each pathway, a specialization takes place, meaning that the cells get more selective to particular stimulus features (see Fig. 1). This goes hand in hand with an increasing generalization capability and a loss of information that is irrelevant to the particular pathway. For example, along the form-processing ventral pathway of primates, cells get more selective to particular forms and acquire increasingly more translation invariance. The increasing selectiveness on one hand and the loss of information on the other constitutes a severe information bottleneck if the final result is to be extracted at the top of the pathways only. For example, consider a stimulus with two letters X and U at different locations. One pathway may extract the form information, arriving at the result that both X and U are present, but not knowing in detail where. A second pathway may extract the exact positions of the objects, without being able to identify their form. One of the problems that arises in such an architecture is the binding of the top information between different processing pathways, i.e., in the example, which form belongs to which position.

In some circumstances, like the case of conflicting stimuli, or in cases which require binding of information across pathways, not only the information at the top of the processing hierarchy is necessary, but a part of the entire tree of information flow along the hierarchy. This may comprise the route information takes in a single pathway, or it may involve two or more

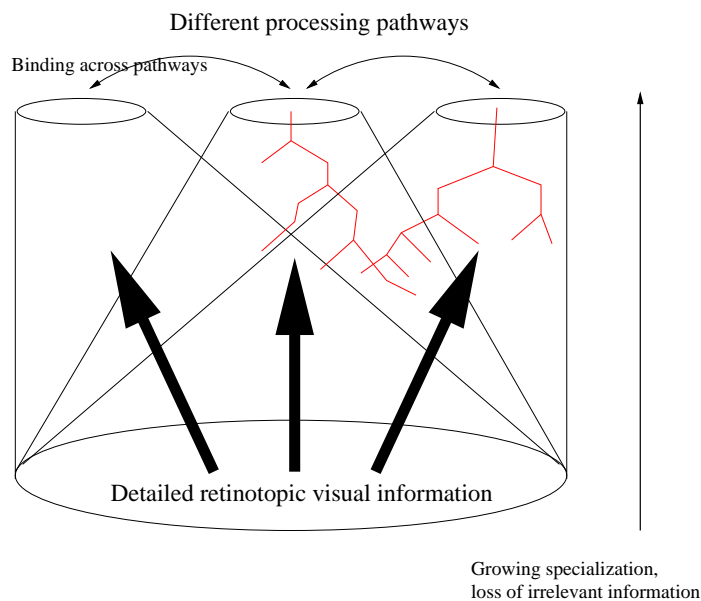


FIGURE 1. Diagram of the organization of visual processing into several pathways. In each pathway, a specialization and a loss of irrelevant information takes place. This causes a binding problem at the top of the pathways. Nevertheless, the necessary binding information is still contained in the common area of high-resolution at the bottom of the pathways. Attention may be used to solve the binding problem by reaccessing the low-level information and making it explicit at the top of the pathways. We propose a network architecture to accomplish this task. Also shown: partial trees of the information flow through the network. Solving the binding problem is equivalent to tracking the trees.

pathways, as in the X-U example. The binding problem is then equivalent to a tracking of the tree of information flow through the network. Such a tracking process relies heavily upon local operations, a characteristic that will be utilized in a later section when we develop a metaphor to illustrate the main points of our model.

We propose that the role of attention in the visual system is to solve this tracking problem, both along and across the processing pathways. This cannot be done for too many stimuli simultaneously, because the different trees of information flow interfere with each other. To remain tractable, the interference requires attention to act selectively concentrating resources on a subset of the available information, so that only the most important trees are tracked (in other words, attentional resources are limited). By tracking the information trees through the network, attentional processing allows to circumvent the information bottleneck mentioned above, binding and (re)accessing information that was lost when moving upwards in the processing hierarchies.

In fact, if the information contained in the entire network would always be *accessible*¹, there would be no such thing as a bottleneck or a binding problem, because at the high-resolution, low-level areas the binding information is present. Complete explicit access is not possible in a finite system, so that attention becomes necessary. In this work, we postulate that for *some specific tasks*, the access of low level information is necessary, and indicate how the binding problem can be solved in these cases using a special network architecture and attention to reaccess and make explicit the required low-level information.

1.3 Spotlight Models

Spotlight models (SM) of attentional processing constitute a special class of models that operate on a saliency map specifying *where* things and locations of interest are, but not exactly *what* they are. A focus of attention (FOA, see section 1.1) is computed preattentively, and attention then operates on the features inside the FOA for preferential processing. Features inside the FOA are routed to higher processing centers. Binding is accomplished by defining that all features inside the FOA belong together. This is an easy task for such models, because the stimuli outside of the FOA are neglected or suppressed.

Problems appear at several points in these models. They imply an architectural division into a preattentively working area (the saliency map area) and the object identification area, which relies on attentive processing. They also imply a strict order of operations, in the sense that spatial segregation (the computation of the FOA) has to precede binding and both have to precede object identification. This means that the attentional processing areas are obliged to wait for the outcome of the saliency map. Experimentally, the order of the operations as imposed by spotlight models is not always confirmed. It is known that identification cues can strongly influence the segregation of a stimulus (take e.g., the well-known picture of a dalmatian dog against a spotty black and white background). In addition, experimental data shows that targets can pop up in a visual search display before they are captivated by the focus of attention, even if they are defined by a conjunction of elementary attributes [8, 9, 28, 55]. This reverts the order of operations, since in this case binding occurs before attentional processing takes place. A similar question is that of detection and localization. Since the focus of attention has to have locked onto an object for attentional processing to occur, localization also precedes identification. But there is evidence that there can be preattentive detection that precedes localization [45]. Thus, the division between segmentation,

¹Meaning here that it can be processed or made explicit in such a way that other network areas can make use of it.

attentional processing and binding is not as clear cut as spotlight models suggest.

The basic question is how the attentional spotlight itself knows where to be directed to, i.e., how the saliency map computes the most sensible FOA. The computations in the saliency map are e.g. modeled using a series of feature detectors working along different feature dimensions and at several spatial scales, and then computing a “winner” (i.e., most salient) location (see also the chapter by Niebur et al. for an implementation of the FOA computation using a saliency map). Since the attentionally processing network itself is also assumed to work based on a series of maps containing similar types of feature detectors, spotlight models require a duplicate network architecture.

Another evidence found in psychophysical experiments that shows the complex nature of attentional processing indicates that more than one stimulus participates in attentional processing even if all of the stimuli can be bounded by nonoverlapping concave regions (see e.g. [11]). This suggests a simultaneous processing of a limited number of objects with attention distributed over a wider area that overlaps several stimuli, contrary to the object-per-object manner of strictly serial spotlight models. Similarly, an attentional interaction between several stimuli that are positioned in or near the focus of attention can be observed (see e.g. [31]). And considerable processing even continues to occur for non-attended objects when attention is directed elsewhere.

Finally, there is evidence for attentional mechanisms that operate both with a high spatial resolution (which is incompatible with the FOA notion) and on non-spatial, feature-based attentional mechanisms. The saliency maps of spotlight models have to segment objects based purely on low level feature analysis (i.e., without object identification). This makes it impossible to isolate a single interesting spot in the case of mutually overlapping objects. Further support for high spatial resolution in combination with attentional processing comes from experiments with spatially coincident patterns, such as transparent random dots [2, 4, 27, 34, 51]. Non-spatial, feature-based attentional mechanisms have also been observed in the case of color [30] and motion [35, 47, 48, 49]. In this case, the response of a cell to a stimulus is enhanced when the animal directs its attention towards features that the cell prefers.

1.4 The Discussion Forum Metaphor

Attentional processing as it is understood by spotlight models has been compared with a theater stage on which a play is taking place and which includes a single, bright spotlight that is directed towards the stage from outside and which serves to highlight a selected area. In this case, there is need for a central “spotlight manager” that directs the light always to one person or place on the stage at one time. This is the role taken by the

saliency map.

The model we propose is quite contrary. It is based on a handful of central assumptions. First of all, we think that the selective function common to the different types of attention (see section 1.1) expresses in a common functional architecture, meaning that all types of attention can be explained within a basic framework. Second, coarse locationally guided attentional processing like that of spotlight models is but one (special) type of attention that fits into this framework. Third, there is no single executive center that indicates where (in space, or to what features) attention is directed to. Instead, attention can originate in a series of different centers, explaining the different types of attention. Fourth, attention is not exclusively dedicated to a single object, totally suppressing all non-attended ones, but instead causes a concentration of resources on some aspects of the scene that are then processed preferentially.

To illustrate the operation of an attentional network under such assumptions, we develop in the following a new metaphor, opposed to the (in our view, misleading) theater stage idea. We need no central “spotlight manager” because the decision, which parts of the scene are important, is taken locally by the actors of the play themselves. We say that attentional processing is like the organization of a public discussion in a democratic or artistic forum, with a mixed hierarchy of speakers and a number of available microphones to speak out loud. The microphones constitute a limited resource. Without having to rely upon a central moderator, groups of people with a fluctuating number of participants will concentrate around the microphones, compete with each other for statements and reorganize into new constellations. Sometimes a single person (perhaps, and very likely, of a high hierarchical ranking) will assert himself and dominate temporarily the discussion; at another time, a large group of persons will gather and coordinate themselves to a speaking choir, or different persons will speak one after another. This allows the system to dynamically organize according to the momentary situation, and to adjust optimally to the actual needs.

A consequence of this view is that there are two grouping processes going on at the same time at different levels of organization, one between single persons at the sub-microphone level and the other between the different groups of people around each microphone. These two processes correspond to preattentive and attentive processing. In addition, attention is never allocated exclusively, but always distributed among several contenders. The concentration of attentional resources in the brain that can be observed in experiments is only the manifestation of this dynamical organization process (in the discussion forum metaphor, the words that you can here aloud thanks to the microphones), while the organization itself occurs largely unnoticed. Such a view also implicates that there is no sharp boundary between preattentive and attentive processing in terms of segregated and consecutive processing stages. Instead, preattentive processing can be seen as the processing that takes place in the head of each of the participants

and between single persons of the discussion forum, while attentional processing is everything related to the organization of speakers around the microphones and their speaking out loud using them. Therefore, preattentive and attentive processing takes place *on the same substrate*, in this case, the participants of the discussion forum. The difference between preattentive and attentive resides in the modes of operation. In a model implementation, this means that the same architectural structures are used both for preattentive and attentive processing (with all the advantages for the network generation, learning, etc., and without requiring a duplicate network architecture, as in the FOA case, see section 1.3). Preattentive and attentive processes could be implemented using different neuronal codes, e.g., synchronized activity for attentional processing that is imposed on top of the non-synchronized, preattentive signal. The discussion forum metaphor also implies that attention is distributed along the entire visual processing system, and that there are multiple distributed areas where the attentional signal can have its origin. These multiple areas coordinate with each other so that the system keeps running smoothly and efficiently. In the following, we present a network model for the implementation of attentional processing in the visual system according to the “discussion forum” metaphor.

2 Activity–Gating Networks

2.1 Working Hypotheses About the Coding of Information

We propose that groups of neurons in the visual cortex can be joined into neuronal **ensembles** (also called **assemblies**, **pools** or **neuronal cliques**). This is motivated by the cortical columnar organization, which presents numerous dendrites running orthogonally to the layers and thus potentially enables neurons to collect input from all the layers they cross on their way and so to gather information available in all layers of their column. A further motivation to group neurons is the experimental observation that cortical neurons of the same type that are located near to each other tend to receive similar inputs. In experiments one often finds that neurons of the same type that are located close to each other are activated simultaneously, or in a correlated fashion. In cortical networks, this may be due to reciprocal connections and common convergent input.

In modeling studies it therefore seems sensible to consider all neurons of the same type in a small cortical volume as a building block of a neuronal network. All pool neurons have to be equivalent in the sense that they have the same input/output connection characteristics and, additionally, the same dynamics parameters. This is explained in Fig. 2. All neurons that constitute a pool feel a common synaptic input field, but still, each neuron evolves according to its own internal dynamics.

In the visual cortex, we can for example assume that all neurons of the

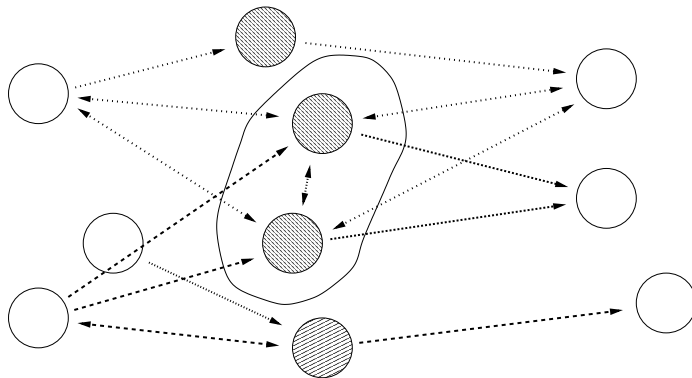


FIGURE 2. A “pool” or “assembly” of neurons. Neurons belonging to the same pool or assembly are characterized by having the same input/output connectivity pattern. Furthermore, all neurons of the same pool have the same parameters. In the figure, different types of neurons and connections are characterized by different textures (white neurons are of any type). According to the assembly definition, only the two neurons inside of the closed contour belong to the same pool.

same type located in the same layer and in the same cortical column, and which additionally have a similar stimulus selectiveness, form an assembly. Assemblies constitute one of the basic building blocks of our attentional network. In the following, we explain some of the dynamical properties of neuronal assemblies.

The response of neuronal assemblies is quantified as a macroscopic activity $A(t)$, which has the dimensions of spikes per time. The quantity $A(t)\Delta t$ is the assembly-averaged number of spikes elicited by the pool in the time interval $(t, t + \Delta t]$. Typically, isolated assemblies settle at a constant **stationary activity** if a constant external drive is applied. The activity vs. drive curve then resembles a sigmoidal gain function, commonly encountered in assembly-averaged pool models.

In addition to stationary solutions, self-coupled assemblies can develop stable, self-sustained **oscillatory activity**, e.g., in the case of assemblies composed of neurons of the spike–response or integrate–and–fire type (for a review, see [16]) if the neuronal parameters fulfill the requirements of the locking-theorem [17]. Oscillatory activity has been found in visual cortex in a variety of experiments (see e.g. [41]) and appears particularly in relation with states of pronounced cortical activation and attention (see e.g. [20]).

An **oscillation threshold** often prevents single neurons or groups of neurons to develop oscillatory activity below certain activations, such as it is the case for the “chattering cells”, which exhibit normal spiking behavior for low stimulation but elicit bursts if the stimulation is strong enough [18]. For groups of neurons, the existence of an oscillation threshold can be verified with simulations of assemblies of spike–response or integrate–and–

fire neurons, both for excitatory and inhibitory reciprocal couplings (see e.g. [17]).

In states of oscillatory activity, an **amplitude coding** scheme enables assemblies to remain sensitive to stimulus changes even if they are working in an oscillatory regime with strong coupling loops and high activity peaks. Amplitude coding means that increasing stimulation strengths are expressed in gradually increasing oscillation amplitudes. Amplitude coding can be observed again both at the single cell level for “chattering cells” which elicit bursts of increasing length in response to increasing stimulation [18], and in simulations of assemblies of spike–response neurons. The amplitude-coding scheme has been verified experimentally comparing optimally and suboptimally driven neurons [26].

The oscillation threshold above is defined for stationary stimulation. Interestingly, although assemblies of spiking neurons with oscillation threshold will not develop oscillatory activity if the stimulation is too weak, they will readily **follow an oscillatory stimulation** (see e.g. [15] for a theoretical account). This means that in this regime, assemblies develop oscillatory activity, e.g. if they receive a modulated input from another assembly with a pronounced oscillation amplitude.

Finally, connections between assemblies can be broadly classified into **driving** and **modulating** types. Driving connections have a direct effect on a cell’s response, enhancing or suppressing its activity. Modulating connections have an indirect effect, modifying a cell’s activity only if it has already been activated previously by some driving sources. Although the evidence is by no means conclusive, the two types of connections and the corresponding source neurons of driving and modulating signals seem to have been confirmed experimentally to some extent (see e.g. [6, 40]). They may serve to avert an activity explosion in a large network composed of neuronal assemblies (in particular the modulating connections, in a network with no strong driving loops, [6]).

2.2 Implementation of Neuronal Ensembles

There exists a variety of models for describing the activity of neuronal assemblies that have the properties enumerated in section 2.1. Since at the assembly level (and because of the computational costs), we cannot cope with all the details of neuronal arborization and dendritic structure, we neglect them and start at the descriptive level of threshold models of spiking neurons. Motivated by simulations and analytical calculations with assemblies of the spike–response type we introduce an abstract model for describing assembly dynamics. The model can be mapped quantitatively with assemblies of spiking neurons in the regime of oscillatory activity and it can be derived from phase oscillator neurons, but we will not go into the analytical details here.

The model is defined as follows. The dynamics of each assembly m is

determined by 3 variables. These variables are the **mean activity** $S_m(t) \in [0, 1]$, the oscillation **amplitude** $r_m(t)S_m(t) \in [0, 1]$ and the **phase** $\Psi_m(t) \in [0, 2\pi)$ ². The oscillation amplitude is composed by the oscillation **coherence** (or **centroid**) $r_m(t) \in [0, 1]$ and the mean activity $S_m(t) \in [0, 1]$. The mean activity, oscillation amplitude and oscillation phase correspond to the minimum of 3 variables needed to describe the joint activity of a neuronal assembly with periodic or quasi-periodic firing, i.e., that exhibits coherent oscillatory activity. In a stable oscillatory solution with period T_m , they correspond to the 0'th fourier-component and the coefficient and phase of the first fourier component of the activity $A_m(t)$.

The mean activity $S_m(t)$ indicates how many neurons participate in spiking during a time period $(t - T/2, t + T/2]$. A value of $S_m(t) = 0$ indicates that the entire assembly remains silent and no spikes are elicited during the period. A value of $S_m(t) = 1$ indicates that all neurons of the assembly release spikes. Neurons that do not spike may nevertheless participate in the assembly dynamics. They participate in the phase dynamics since they still exhibit a pronounced **subthreshold oscillation** in their synaptic input, but they will not participate in driving other neurons, since they do not fire. Therefore, they do not contribute to the macroscopic variables such as the activity or the amplitude and phase. The magnitude of $S_m(t)$ thus indicates the proportion of suprathreshold neurons in an assembly m , and also the mean spike-rate or mean activity of a pool during $(t - T/2, t + T/2]$. For experimental evidence of in-vivo subthreshold oscillations in cortical tissue see [43].

The coherence $r_m(t)$ indicates the proportion of suprathreshold neurons that participate in the oscillatory activity. A value of $r_m(t) = 0$ indicates that all suprathreshold neurons fire incoherently, so that no oscillatory activity is present. A value of $r_m(t) = 1$ indicates that all suprathreshold neurons fire coherently, so that the oscillation amplitude $r_m(t)S_m(t)$ is at its maximum for fixed $S_m(t)$. Therefore, $r_m(t)$ is the **degree of locking of the spiking neurons**. The phase of the oscillatory peak is expressed by $\Psi_m(t)$. Around $\Psi_m(t) = 0$ we find the maximum of the peak of the oscillatory activity.

The full assembly dynamics for a network with L assemblies connected by coupling and feeding links with strengths J_{mn}^{coup} and J_{mn}^{feed} then reads

$$\begin{aligned} \frac{d}{dt}r_m(t) &= \begin{cases} \frac{r_m(t)}{\tau_m(t)} \left[1 - \frac{r_m(t)}{r_m^*(t)} \right] & \text{if } J_m^{\text{eff}}(t) > J_m^{\text{crit}} \\ \frac{r_m(t)}{\tau_m(t)} & \text{otherwise} \end{cases} \\ \frac{d}{dt}\Psi_m(t) &= \Omega_m - \frac{h_m^\Psi(t)}{r_m(t)} \end{aligned}$$

²All phases in this work are used modulo 2π .

$$\frac{d}{dt}S_m(t) = -\frac{1}{\tau_m^S}\{S_m(t) - G_m[h_m^S(t)]\} \quad (1.1)$$

with the effective self-coupling $J_m^{\text{eff}}(t) := h_m^r(t)/r_m(t)$, the fields

$$\begin{aligned} h_m^r(t) &= \sum_{n=1}^L J_{mn}^{\text{coup}} r_n(t) S_n(t) \cos[\Psi_m(t) - \Psi_n(t)] \\ h_m^\Psi(t) &= \sum_{n=1}^L J_{mn}^{\text{coup}} r_n(t) S_n(t) \sin[\Psi_m(t) - \Psi_n(t)] \\ h_m^S(t) &= \sum_{n=1}^L J_{mn}^{\text{feed}} F_m[S_m(t), r_m(t), \Psi_m(t), S_n(t), r_n(t), \Psi_n(t)] \end{aligned} \quad (1.2)$$

the ‘‘stationary’’ centroid strength (defined for $J_m^{\text{eff}}(t) > J_m^{\text{crit}}$)

$$r_m^*(t) = \sqrt{1 - \frac{J_m^{\text{crit}}}{J_m^{\text{eff}}(t)}} \quad (1.3)$$

and the time parameter for centroid growth

$$\tau_m(t) = \frac{1}{2} [J_m^{\text{eff}}(t) - J_m^{\text{crit}}] \quad (1.4)$$

The function $G_m(h_m^S)$ in Eq. (1.1) is a sigmoidal gain function. The function $F_m(\dots)$ in Eq. (1.2) depends on the *macroscopic* states of the pool m and the other pools n of the network. This completes the equations for the **Graded Oscillator Model (GOM)**.

The 3 fields $h_m^r(t)$, $h_m^\Psi(t)$ and $h_m^S(t)$ drive the centroid strength $r_m(t)$, the centroid phase $\Psi_m(t)$ and the mean activity $S_m(t)$, respectively. The centroid strength field $h_m^r(t)$ enters the dynamics equation for $r_m(t)$ in form of the effective self-coupling $J_m^{\text{eff}}(t)$. The oscillatory threshold is given by J_m^{crit} . Only if $J_m^{\text{eff}}(t) > J_m^{\text{crit}}$, the centroid strength or coherence $r_m(t)$ (and thus, the oscillation amplitude) grows. The frequency $\Omega_m = 2\pi/T_m$ is determined by the oscillation period T_m . Two different types of connections, quantified by the strengths of the links between assemblies, J_{mn}^{coup} and J_{mn}^{feed} , influence separately the oscillation dynamics (J_{mn}^{coup} , **coupling** connections) and the dynamics of the mean activity (J_{mn}^{feed} , **feeding** connections). In addition to the dynamics as defined by 1.1 and 1.2, noise plays an important role. We will not go into further details here but indicate that the effect of noise on $S_m(t)$, $r_m(t)$ and $\Psi_m(t)$ can be derived from microscopical considerations and mainly influences $r(t)$ and $\Psi_m(t)$ at small oscillation amplitudes $r_m(t)S_m(t) \rightarrow 0$. In this case, both the coherence $r_m(t)$ and the phase $\Psi_m(t)$ exhibit large fluctuations.

Pools modeled by (1.1) have a series of properties that are found in pools modeled explicitly using spiking neurons, and behave in a qualitatively similar manner. They develop coherent oscillations of the activity in response

to large inputs of the feeding field $h_m^S(t)$. They exhibit a pronounced oscillatory threshold, but nevertheless follow a modulating external stimulation (i.e., they develop a non-vanishing oscillation amplitude if another assembly with a pronounced oscillation provides input conveyed by means of the coupling connections J_{mn}^{coup}). In addition, the strength of the feeding signal is coded in the oscillation amplitude.

The dynamics 1.1 behave in a controllable way, since they can be demonstrated to have a Lyapunov function for symmetrical connections $J_{mn}^{\text{coup}} = J_{nm}^{\text{coup}}$ and $J_{mn}^{\text{feed}} = J_{nm}^{\text{feed}}$. They code explicitly an assembly's coherence and amplitude signals, thus allowing an easy decoding of such signals. And finally, they introduce in a natural way two different types of connections J_{mn}^{coup} and J_{mn}^{feed} , which access different types of macroscopic variables of the assembly.

2.3 Computational Units

The dynamics of a network composed of interconnected assemblies modeled according to (1.1) behave in a much more complex way than a single isolated pool. For a single pool m , the oscillatory behavior is determined by two parameters: its critical coupling strength J_m^{crit} , and its self-coupling J_{mm} . This is easy to understand since in the single pool case the effective self-coupling reduces to $J_m^{\text{eff}}(t) = J_{mm}S_m(t)$, and the critical condition for the development of coherence [$r_m(t) > 0$] is $J_m^{\text{eff}}(t) > J_m^{\text{crit}}$. The centroid $r_m(t)$ thus remains zero until the feeding field $h_m^S(t)$ raises a sufficient proportion of assembly neurons $S_m(t)$ to a suprathreshold level so that $J_{mm}S_m(t) > J_m^{\text{crit}}$, i.e., the oscillatory threshold is overcome. For more than one pool, the effective coupling strength $J_m^{\text{eff}}(t)$ of each pool is determined by the network connections and the centroids of the rest of the pools. In the case of one pool having a non-zero centroid strength, the effective coupling strengths of all other pools it connects to are modified. This has the consequence that a pool will practically always have a non-zero centroid if any other pool that is connected to it (regardless of the sign of the connections) also has a non-zero centroid. (Note that for $r_m(t) > 0$, we need $J_m^{\text{eff}}(t) = h_m^r(t)/r_m(t) > J_m^{\text{crit}}$, and this can always be fulfilled for a positive field $h_m^r(t) > 0$ if $r_m(t)$ is small enough.) A pool with non-zero centroid will influence all other pools in such a way, that although they would not have been able to develop a locking state by virtue of their internal couplings alone, they now have a non-vanishing centroid.

This effect is easy to understand intuitively since any non-constant external driving force pulls and pushes the neurons of a pool into some preferred phase, resulting in a coherence $r_m(t) > 0$. The critical coupling at which the entire network begins to have locking components is an implicit function of the network topology, i.e., it may be regarded as an emergent property of the network. Below that critical coupling, stationary activities dominate. Above it, oscillations can be used information processing.

After oscillations emerge, the fixed connections J_{mn}^{coup} and J_{mn}^{feed} are modified by amplitudes and phases of the presynaptic pools (1.2), generating temporarily new effective connections by strengthening or weakening the existing ones. This has two important consequences. First, it causes the appearance of “**dynamical assemblies**”, which are **groups of assemblies** that can be labeled as belonging together and separated from others because of their strong reciprocal effective connections³ (compare the organization into dynamical assemblies with the self-organization of persons around the microphones in section 1.4). The idea of dynamical assemblies that work in such a way goes back as far as Hebb [19]. The second consequence is that, since the new effective connection strengths modify the network’s hardware, they can remodel it in a specific, task- and goal-oriented way. Instead of an all-purpose machine, we then get a refined hardware that adapts dynamically to the momentary processing needs.

In terms of the variables $r_m(t)$ and $\Psi_m(t)$, the GOM dynamics (1.1) try to maximize the oscillation amplitudes and to adjust the relative phases of the assemblies to each other according to the couplings J_{mn}^{coup} . Assemblies connected by excitatory connections $J_{mn}^{\text{coup}} > 0$ “feel” each other stronger when their phases are near together, in the sense that they stabilize and enhance each other’s oscillation amplitudes. The same happens with inhibitory couplings $J_{mn}^{\text{coup}} < 0$ for phases that are far away from each other. The phase is therefore a sort of grouping label, indicating which assemblies should belong together in view of the internal network knowledge and the sensory input that is actually applied.

We differentiate four types of information that have to be processed in the network. First of all, **sensory input** about the external world arrives at each assembly by means of feeding connections J_{mn}^{feed} and influences mainly its mean activity $S_m(t)$. As a second type of very different information, **hypotheses, expectations and attentional signals**⁴ are conveyed to the assemblies, also mediated by feeding connections J_{mn}^{feed} . This is the information determined to a great extent by previously learned knowledge, stored in the network’s connections. Using the sensory and the attentional data, the assemblies try to reconcile the prior knowledge with the sensory data of the current task.

As a third type of information, **grouping hypotheses** (or grouping labels) are processed which are coded in the phase labels $\Psi_m(t)$ and communicated between assemblies using the coupling connections J_{mn}^{coup} . As a fourth and last type, **grouping certainty** (or grouping strength) is coded in the oscillation amplitude $r_m(t)S_m(t)$ and communicated between assem-

³Keep in mind the following difference. An assembly is **group of neurons** that process similar information. A dynamical assembly is **group of assemblies** that temporarily (for a specific processing task or sensory stimulation) organize themselves as belonging together.

⁴We will use this nomenclature interchangeably.

blies using the same coupling connections J_{mn}^{coup} as the grouping hypotheses. Grouping hypotheses as well as grouping certainty depend on both the sensory input and the expectational signals.

2.4 Working Hypotheses About the Network Function

We propose that the function of a network that processes and makes use all these different types of information is that of using the prior information to constrain the processing space, thus allowing the system to generalize, to eliminate conflicts and to resolve ambiguities. The network accomplishes this task by continually generating hypotheses which are tested against the sensory data and used to gate the information flow through the network.

In a network that uses both sensory and attentional data in the same architecture, it has to be made sure that the data is kept separately on a local scale and does not corrupt each other. Otherwise, the attentional data would be able to influence the sensory input to such a point that it cannot be discerned any more as such, meaning e.g. that the network would simply detect the objects it is hypothesizing of, corresponding to a situation of uncontrolled visual imagery. Therefore, a single computational unit of our network is basically composed of two functionally distinct assemblies. One is responsible for sensory data and the other one for attentional information. In addition, both assemblies receive information from other processing units about grouping processes. Figure 3 a) shows the basic computational unit and the four types of information that can influence its behavior. **SP** is the **sensory processing assembly** and **OG** the oscillation generating and **attentional processing assembly**. Two assemblies SP and OG always work together, processing data that corresponds to the same feature selectivity. (Meaning that, if an SP assembly shows a strong selectivity for a specific stimulus, its corresponding OG assembly manages the hypotheses about that particular stimulus).

The expectational data is used to locally gate the information flow through the network. This is explained in Fig. 3 b), c) and d), which correspond to the situations of a unit receiving sensory data alone, of receiving both sensory and expectational data, and of receiving expectational data alone (thick and thin arrows indicate connections carrying weak and strong signals, respectively). In the first case (b), the SP assembly responds in a standard way, according to its input specificity. In the second case (c), the output of the processing unit is enhanced in relation to the first case, because *the expectation (or local hypothesis) has been confirmed by the sensory data*. The enhancement is coded as a modulation of the resulting SP activity, and indicated in the figure as an outgoing coupling connection carrying a strong signal. The confirmation of the SP activity occurs by modulation through coupling $\text{OG} \rightarrow \text{SP}$ connections, i.e., an enhancing occurs only if there is sensory data present. This is shown in the third case (d), where nothing occurs at the SP assembly, although the OG assembly indicates

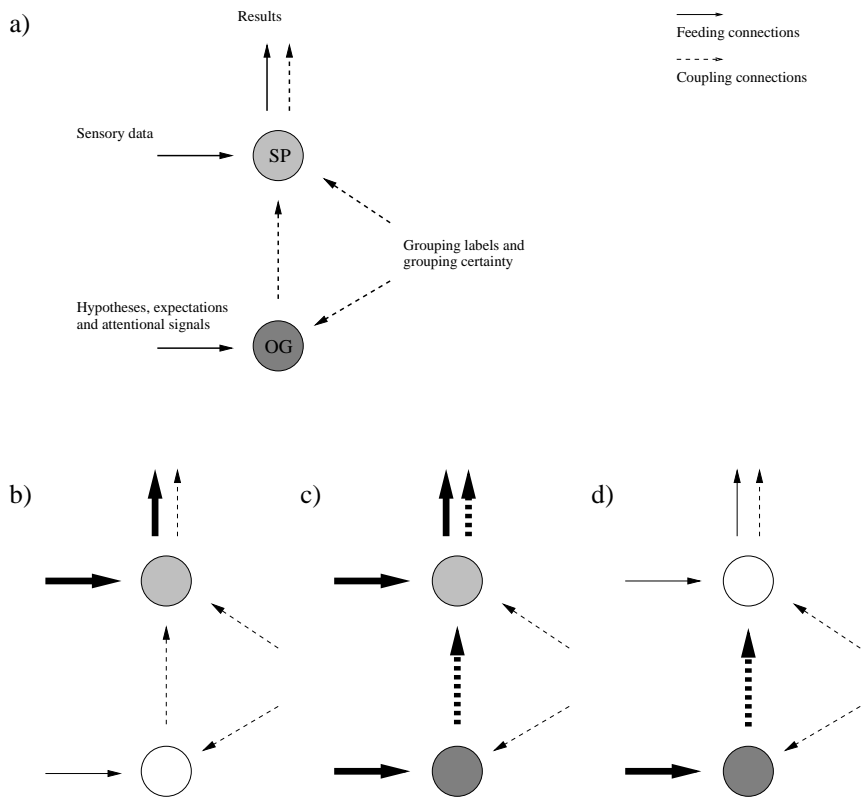


FIGURE 3. Four types of information are processed in a single computational unit. First of all, sensory data is processed independently from hypotheses, expectations and attentional data. This is accomplished by two functionally specialized assemblies, called SP (sensory processing) and OG (oscillation generating and attentional processing). In a), the two assemblies with the corresponding feeding connections are shown. In addition, both assemblies process grouping hypotheses and grouping certainties, provided by coupling connections, and coded in the oscillation phase and the oscillation amplitude, respectively. The attentional data gates the information flow through the network. The gating occurs locally at those computational units at which sensory data is confirmed by attentional data, as shown in c). In b), the unit receives sensory input only, in this case the data is propagated in a standard way through the SP assemblies exclusively. To the contrary, in d), the unit receives attentional input only. In this case, no further propagation of data occurs, since the expectations are not confirmed by sensory data.

that the network would expect input on the corresponding SP counterpart (strong coupling OG \rightarrow SP signal).

In case that the sensory and the expectational data confirm each other, the activity of the corresponding SP pool is modulated in such a way that postsynaptically connected units detect a stronger signal, indicating that the unit's output is to be processed preferentially. In our implementation, the modulation occurs using a **coherent oscillation of the assembly activity that is imposed on the otherwise non-modulated activity evoked by sensory input** on SP. This means that the information flow is gated through the network using an activity modulation, therefore we call this type of networks **activity-gating** networks. The strength of the activity modulation (and thus, the "gating saliency" of a unit) is coded in the oscillation amplitude $r_m(t)S_m(t)$. In short, from two units with the same sensory input but different oscillation amplitudes, the one with the largest amplitude will have a better chance to pass on the results of its computations to further processing levels.

Postsynaptically connected units are affected by the gating signal either directly using the coupling connections, or they can work explicitly as **coherence detectors**⁵, converting the modulated activity back to a field $h_m^S(t)$ that influences the mean activity $S_m(t)$ (for theoretical accounts on coherence detecting capabilities of spiking neurons see e.g. [24]). In this case, the function $F_m(\dots)$ from (1.2) expresses as:

$$\begin{aligned} h_m^S(t) &= \sum_{n=1}^L J_{mn}^{\text{feed, simple}} S_n(t) \\ &+ \sum_{n=1}^L J_{mn}^{\text{feed, intra}} S_n(t)r_n(t) \\ &+ \sum_{n=1}^L J_{mn}^{\text{feed, inter}} S_n(t)r_n(t) \cos[\Psi_m(t) - \Psi_n(t)]. \quad (1.5) \end{aligned}$$

The first term on the right detects the mean activity, while the second and the third terms are sensitive to the intrapool coherence and the joint intra/interpool coherence, respectively.

A single computational unit that works according to the activity-gating principle is basically composed of 2 GOM assemblies, corresponding to the SP and OG pools from Fig. 3. It is shown in Fig. 4 with its internal connections, and feedforward, lateral and feedback connections arriving from and going to other units of the system. This completes the microarchitecture of the network.

From the dynamics (1.1) and Fig. 4, we see that the effect of OG pool

⁵We speak of **coherence detection** instead of **coincidence detection** to indicate the population character.

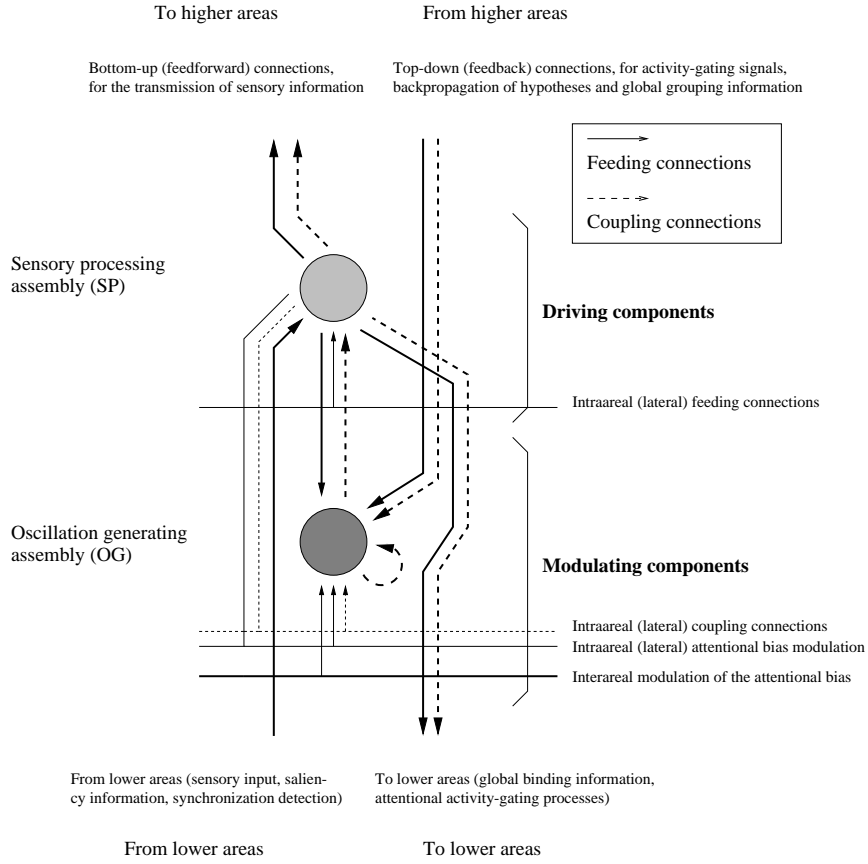


FIGURE 4. The computational unit of the activity-gating network as it is implemented in the model. The proposed microarchitecture is able to process separately sensory input on one hand and hypotheses, expectations and attentional signals on the other. It is composed of two GOM pools. The sensory processing (SP) pool receives feeding sensory input from lower areas and relays feeding and coupling information on to both higher and lower areas. The oscillation generating (OG) pool receives feeding and coupling attentional input, mainly from higher areas. Only the OG pool is able to intrinsically generate and sustain oscillations, if it is activated strong enough. The oscillatory modulation is then imprinted on the SP pool and used for activity-gating and grouping processes in higher and lower areas. The result is a modulated activity of the SP pool, with an oscillation amplitude $r_m^{OG}(t)S_m^{SP}(t)$, i.e., with a coherence determined by the hypotheses, expectations and attentional signals, multiplied by the mean activity determined by sensory input.

activity on its SP counterpart is to influence SP's oscillation amplitude $r_m^{\text{SP}}(t)S_m^{\text{SP}}(t)$ ⁶. More precisely, in the microcircuit architecture of Fig. 4 **it influences only the coherence** $r_m^{\text{SP}}(t)$ of the SP pool, so that its effect on the oscillation amplitude is **purely modulating**. Without sensory input (coded in the mean activity $S_m^{\text{SP}}(t)$ of the SP pool), the influence of the OG pool vanishes. With sensory input, it can be calculated from the dynamics that, for equal J_m^{crit} for the SP and the OG pools and equal strength of the coupling connections $\text{OG} \rightarrow \text{OG}$ and $\text{OG} \rightarrow \text{SP}$, it is $r_m^{\text{SP}}(t) = r_m^{\text{OG}}(t)$ and therefore

$$r_m^{\text{SP}}(t)S_m^{\text{SP}}(t) = r_m^{\text{OG}}(t)S_m^{\text{SP}}(t) , \quad (1.6)$$

so that the coherence of the SP pool is determined by that of its OG counterpart, and the amplitude of the SP pool is a **multiplication** of the sensory certainty, coded in $S_m^{\text{SP}}(t)$, and the confirmation of the sensory data by internal expectations, coded in $r_m^{\text{OG}}(t)$. This multiplicative character arises in our model as a consequence of the assembly dynamics and the microarchitecture, and not by the introduction of special multiplicative connections at the neuronal level (see e.g. [10, 25]).

2.5 Network Architecture: Complementary Processing Streams

The large-scale network architecture uses the function of the computational units as hypothesis generating and testing devices (additionally to the processing of sensory information), and provides the framework so that the different types of data can be accessed.

In the network, each unit receives sensory input and passes it on to other units. This means that sensory information is directly propagated through the network using only the SP assemblies as relay stations, and in a **fast, feedforward sweep of sensory activity**. To the contrary, the influence of the OG assemblies on processing is indirect, in the sense that they can only influence other units by their SP intermediary, so that the information flow goes $\text{OG} \rightarrow \text{SP} \rightarrow \text{OG} \rightarrow \text{SP} \rightarrow \text{OG} \dots$. In addition, they use oscillatory activity for their modulatory signals, which take time to build up. It must also be considered that the attentional signals are mainly sent to hierarchically lower subsystems, because they provide global hypotheses. Therefore, opposed to the fast, feedforward sweep of sensory activity, we have a **slower building up of expectational and attentional signals using lateral and feedback connections**. In summary, **two complementary processing streams** process differently sensory vs. attentional information, and influence each other mainly at those units at which the information flows meet. Fig. 5 shows the implementation of the complementary processing streams using the computational units from Fig. 4.

⁶From now on, we label the computational units with m, n and indicate additionally the assembly with SP and OG.

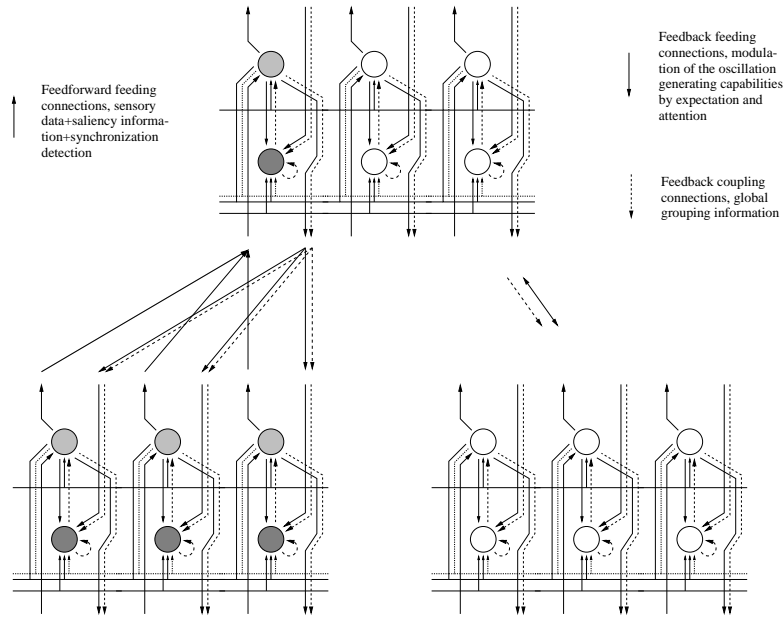


FIGURE 5. Two complementary counterstreams serve to process sensory vs. attentional information. The sensory information is processed in a mainly feedforward stream that comprises the SP pools (light grey circles). The attentional information is processed in a stream that uses lateral and feedback connections and comprises the OG pools (dark grey circles). While the feedforward stream directly communicates different hierarchical areas, in the feedback stream, the attentional information (represented in the activity of the OG pools) has to be confirmed by the activity of the corresponding SP pools. There is *no direct* link of feeding connections in the feedback stream, instead, the feedback information always crosses at least one modulatory $OG \rightarrow SP$ connection, meaning that there are no purely feeding loops in the system, and preventing a *hypothesis explosion*.

The relay on the intermediary SP units for attentional processing prevents, in combination with the modulatory (i.e., multiplicative) character of the $OG \rightarrow SP$ connections, a proliferation of hypotheses that are not confirmed by sensory input. This is a sensitive side effect of the architecture, since it averts a **hypothesis explosion** on the attentional processing stream. Without such an architecture, a hypothesis explosion occurs because there are many different and ambiguous ways to compose a high-level feature from low-level features.

The overall picture then is the following. Instead of introducing a division between preattentive and attentive processing on an architectural scale (see section 1.3 on FOA models), the difference between preattentive and attentional processing is more subtle. It resides in the different coding

schemes and the different time scales at which the processing of information occurs. On one hand, preattentive processing is inherently fast, occurs in parallel and serves to trigger hypotheses and categorize sensory input in the entire network. It is responsible for detecting fast onsets, novel appearing stimuli and imposed synchronization (e.g., by external sources), and for coding the presence of sensory features. On the other hand, attentive processing modulates existing sensory data, testing hypotheses and organizing the network into dynamical assemblies that are specialized for task-dependent processing. Since attentional processing uses grouping labels it is intrinsically capacity-limited, imposing a serial processing at the highest hierarchical stages of the network. If there is not sufficient time for the generation of modulatory signals, attentional processes cannot develop, (re)access to low-level information and grouping processes are impeded and ambiguity errors (such as false conjunctions) appear (see e.g. [46]).

In summary, in the activity-gating framework, attention controls, both locally and globally, the information flow through the network.

2.6 Overall Network Organization

The overall network is organized into several different processing pathways (see also section 1.2). Each pathway concentrates on specific stimulus attributes, such as location, form, color or texture, and is composed of a hierarchy of areas which show an increasing specialization for higher hierarchical levels. Between pathways, crossconnections exist. Most connections are bidirectional, i.e., there are both feedforward and feedback links between areas as well as between pathways.

This means that the network has neither a strictly serial hierarchical organization, nor that it is strictly feedforward. It is rather a hierarchical organization of areas in a broad sense, with higher and lower areas, concurrent processing streams, parallel processing among the streams and extensive feedback.

The organization of the network crudely mimics the so-called “what”, or ventral, and “where”, or dorsal, pathways of the visual cortex of primates with some of its identified processing areas, such as V1 and V2. The information arrives the network through a common area (corresponding to the LGN) and is then processed in parallel by the different pathways. Feature detectors of increasing complexity are implemented in all areas of the dorsal and the ventral pathways. In a pathway, each area innervates its successor with feedforward connections, and sends back information to its predecessor using feedback connections. In the upper left diagram of Fig. 6 a scheme of the main pathways, areas and connections is shown.

In our model, the areas denominated as V1 and V2 serve as the detecting areas for such simple features as points, simple lines, end-stopped lines and edges, while areas C1, C2 and C3 serve to detect color/texture. In all these areas, the visual field is sampled into discrete subregions, and for

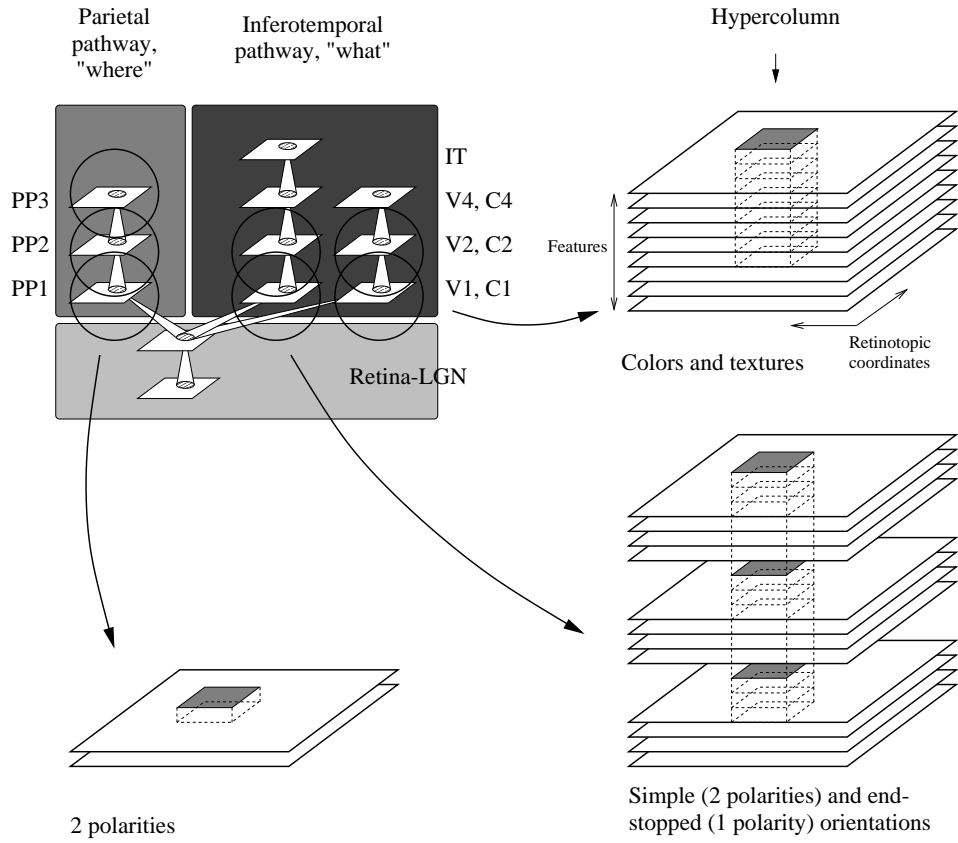


FIGURE 6. Scheme of the attentional network. There are different processing pathways (“what” and “where”), specialized to analyze location (where-pathway, areas PP1, PP2 and PP3), form (what-pathway, areas V1, V2, V4 and IT) , and color/texture (areas C1, C2 and C3) of the stimulus. Each pathway has an ascending hierarchy of areas with retinotopical organization. In each area, “hypercolumns” of feature detectors analyze in parallel the stimulus at the same retinal position. Also shown: The convergence/divergence pattern of feedforward/feedback connections. Moving towards higher processing areas, the convergence causes a gradual increase of generalization capabilities and a loss of low-level information that is irrelevant for each pathway.

each subregion there exists a number of units that detect simple features. Units with the same feature selectivities repeat over the entire visual field. The result is a sort of grid arrangement of feature detectors. All feature detectors responsible for the same retinotopic position form a *hypercolumn*. All processing units that compose a single feature detector are grouped into a *column*. The organization of areas V1 and V2 into columns and hypercolumns is shown on the left half of Fig. 6. A similar organization is implemented for the areas of the dorsal pathway.

The specialization of the processing pathways increases with increasing hierarchical level. At the same time, each area gets increasingly invariant against information that it is not specialized to process. For example, in the form processing pathway, the features get more form-specific but less specific to the exact location of the stimulus. This occurs by an increase of the complexity and size of receptive fields, a pooling of the inputs from lower levels and a coarser retinotopic sampling of the visual field. To the contrary, in the where pathway, only the receptive field sizes increase, whereas the form and the retinotopic sampling of the receptive fields keep unchanged. The result is an increasingly translation-invariant form detection along the what pathway, and an increasingly form-invariant location detection along the where pathway.

In the entire network, an area is composed of the units of Fig. 4. The corresponding feedforward and feedback connectivity is as depicted in Fig. 5. In addition, lateral connections connect units of a single area. We will not explain it in detail here, but in short, lateral connections have an intermediate connectivity pattern between feedforward and feedback links. They serve the multiple purposes of transmitting hypotheses laterally inside an area, redundancy reduction, regulation of biased competition and enhancement of rare feature conjunctions.

3 Results

3.1 *Biased Competition*

The typical suppression of responses seen in experiments related with attention is caused in our network by biased competition effects. Attentionally gated signals not only are processed preferentially using activity modulation and coherence detectors, but they also favor the affected units in the sense that attention enables them to elicit stronger suppressing signals on other units they compete with. This attentionally biased competition causes processing resources to be focused on less items, which is necessary because of the limited resources of the system (see section 1.2).

In the experiment of Moran and Desimone ([29], see Fig. 7), single cells of monkey visual cortical areas V4 and IT were recorded. A cell responding effectively to one type of visual stimulus and ineffectively to another was

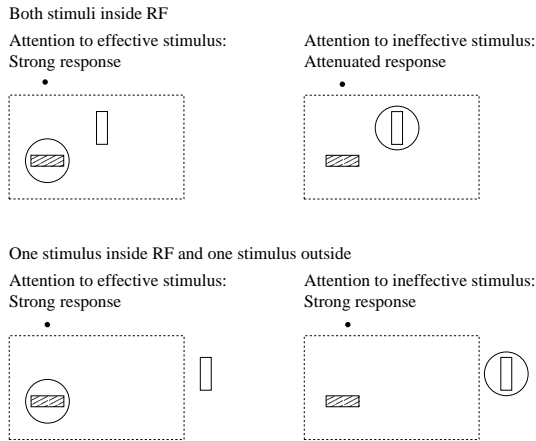


FIGURE 7. The experiment of Moran and Desimone [29]. When both the effective and the ineffective stimulus are inside the RF of a V4 cell, and the animal attends to the location of the ineffective stimulus, the cell gives almost no response. This effect does not appear if the ineffective stimulus is outside the RF. (Small black spot: fixation point.)

selected from one of the mentioned visual areas. In addition, the region of the visual field that influenced its response, its receptive field (RF), was determined. The monkey was trained to attend to stimuli at one location in the visual field and to ignore stimuli at another, while the fixation point in the visual field remained cued to the same location. The monkey was presented simultaneously an effective (one that the cell would selectively respond to under single stimulus presentations) and an ineffective sensory stimulus. Because the sensory conditions were identical for the two types of trials, during which the monkey directed its attention to any one of the two stimuli locations, the difference in the response characteristic of the selected cell has to be attributed to the effects of selective attention. When the locations of both stimuli fell inside the receptive field of the selected cell, and the animal attended to the location of the effective stimulus, the cell gave a strong response. But when the animal attended to the ineffective stimulus, the cell gave almost no response, although the effective stimulus was still within its receptive field. In this case, the response of the cell was greatly reduced (by more than half) by the attentional effects. On the other hand, if the ineffective stimulus was placed outside the receptive field of the selected cell, the neuron showed similar responses to the effective stimulus for both attentional conditions.

Figure 8 shows the simulation results of the Moran and Desimone experiment within our network. The attentional effect is based purely on the spatial location of the effective stimulus, so that attention biases all OG assemblies of those units that are close to its position. Fig. 8 a) shows a simplified network of the three units (corresponding to the effective unit, an ineffective unit inside of the effective unit's RF and an ineffective unit outside of the effective unit's RF) and the connections. In all cases, we concentrate on the left (i.e., the effective) unit. In Fig. 8 b), spatial attention is directed to the left unit's position ("attention to effective stimulus" in Fig. 7). Although the middle unit also receives a sensory input of the same

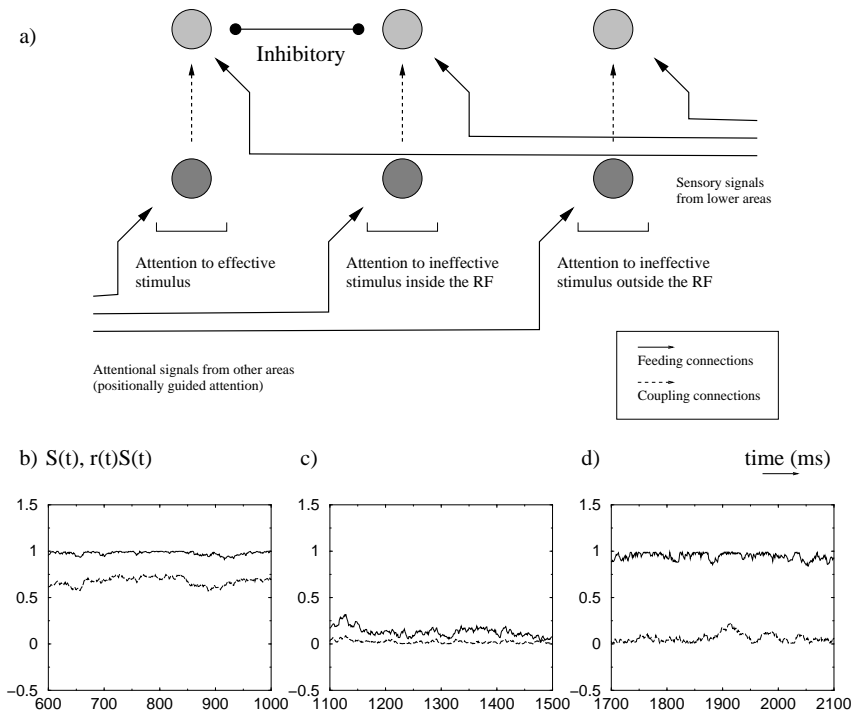


FIGURE 8. Suppression of activity that appears in relation with attental processing is a result of biased competition. Attentional processing evokes a modulation of the activity at attended units, which then get a competitive advantage. This causes the network to concentrate the previously broadly distributed processing resources on fewer units. At the top (a), a simplified diagram of the involved units and connections is shown. The bottom row of three graphs (b, c and d) shows the leftmost unit's SP assembly activity (dark solid line) and oscillation amplitude (dashed line), in different attentional conditions. The three conditions are: attention to the effective stimulus (b), to an ineffective stimulus located inside of the effective stimulus' RF (c), and to an ineffective stimulus located outside of the effective stimulus' RF (d). It can be seen that the unit's activity is suppressed in the second case (c). In this case, attention provides the middle unit of a) with a competitive advantage, so that it has greater suppressive effects on the left unit than viceversa.

strength, it loses the competition and is suppressed by the attended stimulus. The suppressive effect of the middle unit on the left unit is therefore diminished. In Fig. 8 c), the situation is reverted and spatial attention is directed to the middle unit's position ("attention to ineffective stimulus" in Fig. 7). Now the middle unit is favored in the competition, so that the suppressive effect on the left unit is strong, therefore its response is attenuated. In Fig. 8 d), spatial attention is directed to the right unit's position, outside of the left unit's receptive field. In this case, the units do not compete with each other, and there is no influence of attention on the left unit's response. This demonstrates the effects of biased competition using lateral inhibition that is triggered by activities with large amplitudes.

Of course, spatial attention may not only suppress but also enhance activity, because hierarchically lower areas are also influenced by attentional gating, rising the activity of units that are selective for features at the attended location. This in turn strengthens the activity of the attended units in the higher areas.

3.2 *Contour Integration by Laterally Transmitted Expectation*

As explained in section 2.6, in addition to enhancing biased competition, lateral connections serve to propagate hypotheses and expectations laterally inside an area. An example is contour integration in V1/V2. This is implemented by using lateral SP \rightarrow OG connections. If a feature detector receives a large sensory activation, indicating e.g. the presence of an oriented line segment, it sends a signal to all other feature detectors that it expects to be activated as well, according to the prior information about stimulus properties stored in the network. In the example shown in Fig. 9, the lateral connections reflect the psychophysical Gestalt laws of good continuation, similarity and proximity, so that continued smooth lines are preferred. At those units in V1/V2 at which the laterally transmitted expectations are met by sensory data, activity is enhanced, oscillation amplitudes grow and grouping processes take place. This can be seen in the time development of the network activity in Fig. 9. The lateral connections cause a grouping of those feature detectors that correspond to a "good" line according to the Gestalt laws. This in turn gates feedforward connections, so that grouped feature detectors now have a larger impact on further processing.

Of course, lateral propagation of hypotheses can only contribute proposals of limited range about grouping and segmentation, and may include conflicting signals. These cannot be resolved but by further analysis in higher processing stages and using the backpropagation of global hypotheses. This is accomplished by the connections of units as shown in Fig. 5.

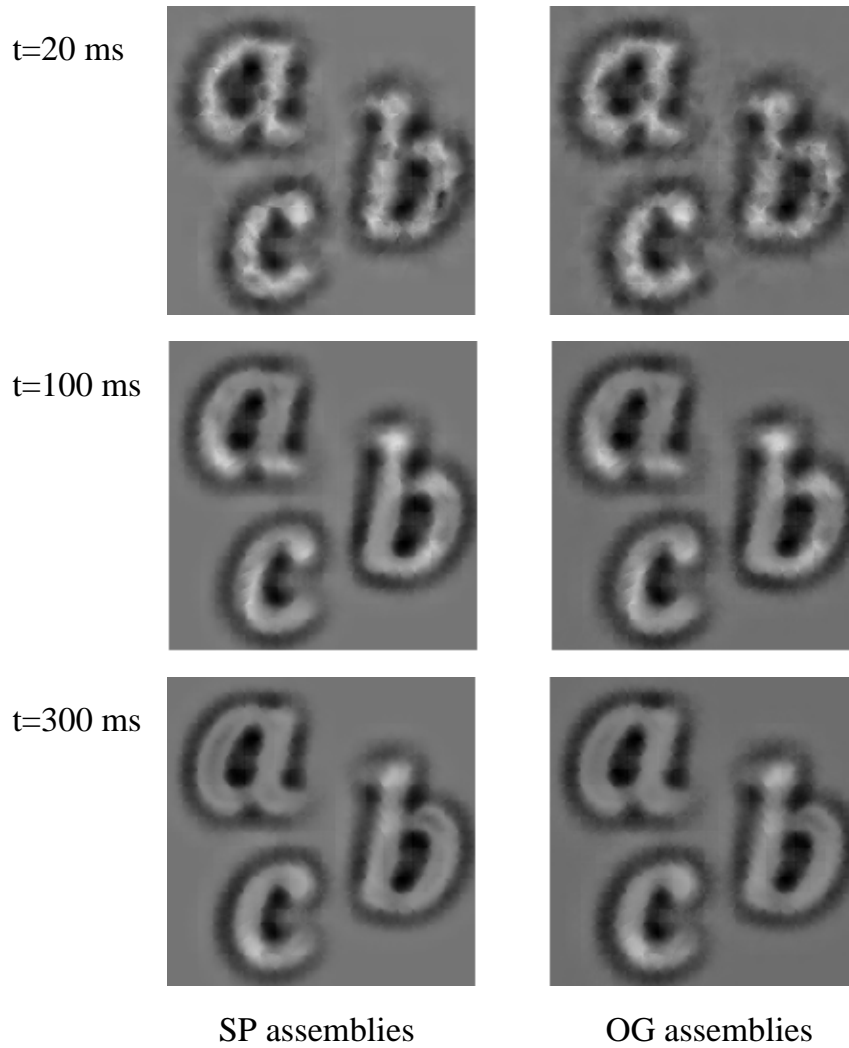


FIGURE 9. Contour integration by laterally transmitted expectation. The network continuously creates hypotheses that evoke expectations and attentional signals at other units, where they have to be confirmed by sensory data. Particularly in the lower areas V1/V2, lateral SP \rightarrow OG connections cause an increase of the amplitudes at attended units and provide grouping signals. The figure shows the amplitude and phase of the V1 units coded as color saturation and color hue. It can be seen that in the time course from top to bottom, the phases adjust so that a grouping of smooth contours occurs (in BW pictures, a smoothing of the roughness can be distinguished). Preferred contours are those in accordance with the Gestalt principles of good continuation, similarity and proximity, reflecting the stimulus statistics expected by the network.

3.3 Different Types of Attention

Attentional gating signals can originate at any point of the network at which large oscillation amplitudes of the SP pools are evoked by a convenient match of sensory and expectational input. According to Fig. 4, we see that feeding input arrives at the OG pool by means of a direct connection from its SP counterpart, and by indirect connections providing expectations and hypotheses using mainly lateral and feedback connections from other units. Therefore, there are two sources that facilitate the generation of large amplitudes. First, they can be evoked directly by sensory input, in this case the attentional signals are **saliency, or stimulus driven**. This is e.g. the case for highly detectable stimuli, such as a red spot against a background of another color.⁷ Second, facilitating signals can be evoked by extraneous sources, in this case the attentional processing is **task driven**. This is e.g. the case if the task demands the system to detect a particular feature. In this case, attentional resources are directed to the units that represent that feature, and thus gating signals in form of large oscillation amplitudes are more likely to originate at those units. Usually, both types of signals will be present, so that a strong task-determined attentional signal can overwhelm the stimulus saliency or viceversa.

According to the saliency characteristics of the stimulus and the task that the system has to solve, different areas will be the origin of the strongest attentional gating signals. Furthermore, according to the area of origin, we may distinguish the network behavior as corresponding to **different types of attention**. In our network of Fig. 6, the attentional signals can originate in any of the pathways, so that we can differentiate **locationally-guided attention** (areas PP1, PP2 and PP3), **feature-guided attention** (e.g., areas V1/V2, color and texture processing areas) and **form or memory-guided attention** (higher areas of the form-processing pathway). This list is by no means exhaustive, since further processing pathways, such as a pathway responsible for motion processing which is not implemented in the current version of our network, are known to exist and should provide further sources of attentional signals.

In Fig. 10, we show the results of network simulations for feature-guided attentional processing. Attention was drawn to a tilted segment (/), meaning that the corresponding OG pools in V1 were biased accordingly by top-down connections. As a consequence, attentional gating signals in form of large oscillation amplitudes are generated preferably in units with strong sensory input and the right orientation. This in turn allows the further propagation of attentional signals to other areas and pathways. In the fig-

⁷In section 1.4 we already indicated that, instead of separating architecturally the saliency map from attentive processing areas, it may be an advantage to integrate saliency and attentional computations into the very same network. In this case, the stimulus-driven gating signals are gained by internal saliency computations.

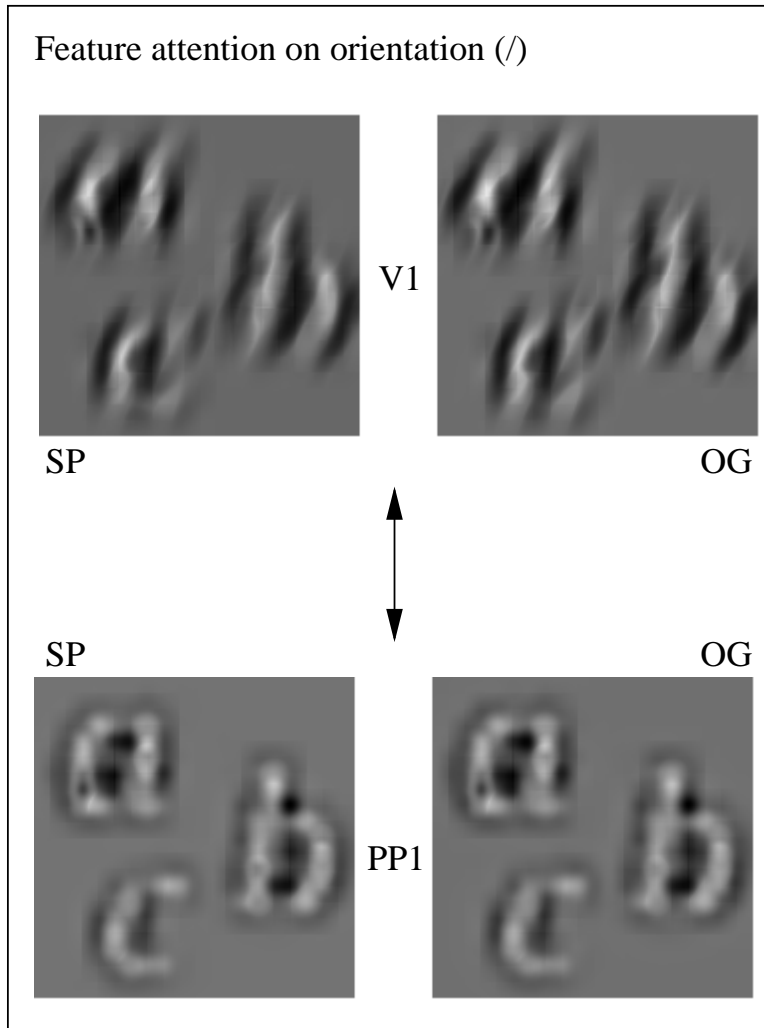


FIGURE 10. Feature-guided attention and top-down propagation of signals. In this figure, attention is drawn to a tilted segment (/) of a fixed orientation by using top-down signals that bias specific orientation-sensitive feature-detecting units. Shown are areas V1 and PP1 from the network of Fig. 6. Attentional gating signals are generated preferably in the feature-selective units of V1. This allows a propagation of attentional signals to other areas and pathways. In the figure, this is shown for area PP1, from where the location of the tilted segments can be extracted. Locationally-guided attention works in the opposite direction, with attentional signals that originate in the parietal areas and are propagated from there to the inferotemporal areas, say $PP1 \rightarrow V1$.

ure, this causes an enhancement of activity in area PP1, which is sensitive to the stimulus location. Area PP1 thus indicates the positions where tilted segments were found.

4 Discussion

4.1 Microarchitecture and Concurrent Processing Streams

The architecture of the system is based on physiological and anatomical data and on the dynamics of assemblies of spiking neurons observed in detailed simulations. Of course, the mapping on real biological structures is to a great extent a matter of debate and requires further experimental evidence, so that we intend this network as a first approximation to the very complex visual system. However, we have introduced a network which fully incorporates feedback processing in a realistic framework, which may have major implications for a further understanding of information flow in the cortex.

For simplicity, we will not discuss the experimental evidence in detail, but mention only a few important points. The organization of the visual system is fundamentally hierarchical [13, 38, 52, 54]. But instead of a strict pyramidal hierarchy, parallel processing in concurrent pathways seems to be the norm (see e.g. [53] for an account). The dorsal pathway along with the posterior parietal cortex has been implicated in the analysis of spatial relationships, movement and locationally-guided attention, while the ventral pathway seems to be involved in the analysis of form, color, texture and disparity. There are massive feedback projections to virtually any region of the visual cortex, they seem to be involved in gating, signal-to-noise enhancement and the improvement of figure-ground segmentation (see e.g. [21]). There is a marked asymmetry between feedforward and feedback connections regarding the layers of origin and destination of the connections [13], suggesting a functionally different role for the two types of connections. In addition, connections and neurons have been found that can be crudely categorized as having a driving and a modulating function [40].

A strategy of bidirectionally processing counterstreams with some characteristics similar to our model has been proposed previously by Ullman [50]. It has to be regarded, however, that no concrete model implementation is proposed, and the precise roles and functions of feedback, processing pathways and attention are not specified in his model.

A coherent modulation of the activity to enhance signals and increase feedforward inhibition at an attended location of the visual field is used in models of Niebur et al [33], and of Niebur and Koch [32]. Their models are based on the architectural division between preattentive saliency computation and attentive postprocessing as explained in section 1.3. Their work shows that very simple neurons are able to enhance coherent activity

if driven at the right parameters. In comparison, the proposed framework of our model does not rely on a segregated saliency map and the usage of activity modulation plays a fundamental, more complex role in gating signals in the entire network. Moreover, our model is able to account for a variety of effects related to attention (see section 3.3) and feedback.

4.2 *Biased Competition*

The model unifies two central characteristics of attentional processing that have been proposed previously. One is that attention acts by imposing a biased competition among attended units [7, 37]. The resulting effect of attention thus is to favor certain units against others.

The second characteristic is that the origin of the attentional signal need not be a single area, such as the saliency map of spotlight models, but that the entire network continuously produces saliency signals for the control of attention. R. Desimone already conjectured from his experimental results that ‘we believe that there is no single saliency map in the attentional control system, but rather a series of maps (or possibly a distributed map) ...’ [7].

In our network, competition is triggered and biased by attention because attention means increased oscillation amplitudes and it is assumed that coherent activities evoke stronger signals on the postsynaptic side than non-coherent ones. In the case of biased competition, these signals are inhibitory and the signals are indeed stronger because of the intra and interpool coherence detecting contributions to (1.5), which are sensitive to the amplitude $r_m(t)S_m(t)$.

This is a realistic assumption, since synchronized and coherent activity is more effective in driving postsynaptic neurons over their firing thresholds because of a temporal summation of postsynaptic potentials. Therefore, we consider biased competition in conjunction with coherent activity to be a biologically plausible contribution to attentional mechanisms.

In our model, attention expresses both as a suppression, caused by increased lateral competition among attended units, and enhancement, caused by activity-modulated signals arriving from lower areas that are gated using the coherence detectors. Both suppression and enhancement have been found in connection with attentional processing (see e.g. [48]), so that attentional effects cannot be restricted to either enhancement or suppression alone.

4.3 *Origin of the Attentional Signal*

The view of a distributed attentional control (or, in the sensory-driven case, distributed attentional saliency computation) implies that there may be different points of origin of an attentional signal. This different origin correlates in our model with the different types of attention. For example,

attentional signals can originate in the parietal area, in hierarchically high areas of the form processing pathway, or in an orientation, texture or color processing area, corresponding to locationally-guided, high-level form (or form-memory-guided), or feature-based attention, respectively.

The question remains how to determine the origin of the attentional signal for an applied stimulus. In our model, two criteria are used for this purpose. One is how well the functionality of an area matches the current task that has to be resolved. This implies e.g. that tasks that require localization of a stimulus slightly bias the parietal areas (in their entirety, non-specifically) so that attentional effects are enhanced and originated in these areas. The second is that, in a single area, attentional signals should originate in those units whose response selectivity is matched best by the stimulus, i.e., those that are strongest activated by sensory input. This means that, e.g., locationally-guided attention will originate in that special area of the parietal pathway with a spatial resolution that best matches the input.

4.4 Saliency and the Focus of Attention

As explained, the computation of salient parts of the scene is accomplished by the very same network units that also process attentional data. The focus of attention of the spotlight models relies mainly on the computation of a bounded region of interest in retinotopic space. It can therefore be compared with locationally-guided attention that originates in our network in the areas PP1, PP2 or PP3. These can be thought of constituting a spatial “saliency map” that indicates the position of interesting parts of the scene. Undoubtedly, locationally-guided attentional processing is crucial in scanning objects that are easy to segregate by their positions. Nevertheless, in our model this is but one source of attentional signals, and it may as well be the case that, instead of sending attentional signals to the other pathways it may receive attentional signals from them. This is e.g. the case if an object is identified by some prominent features, and afterwards its exact location has to be computed by the parietal areas.

4.5 Predictions of the Model and Conclusions

A series of predictions arise from the model. One of the most important predictions is that there is spatially specific attentional modulation right down to the very low areas with high spatial resolution, where the different pathways originate. Furthermore, the model predicts that the attentional effects in these areas are not limited to locationally-guided attention, but can as well be feature or object based. In fact, experimental data seemed to support that attention modulates higher (extrastriate) object recognition areas, but not the primary visual cortex. Only recently, detailed studies

have shown that there are robust attentional effects in the primary visual cortex [14, 39, 42], confirming the predictions.

A second important prediction is that visual processing occurs on two timescales, corresponding to a first, mainly feedforward activity propagation which lasts for some tens of ms and a second processing step that deploys on a slower timescale, relying on feedback processing and involving attentional processing. This division into early and late responses, with the late responses acquiring an increasing specificity with time, should be observed experimentally. In a recent paper [44], this has been confirmed for neurons of the IT cortex. An information analysis showed that IT neurons encode different stimulus attributes in their early and late responses to the same stimulus.

Finally, in our model it makes no difference whether the source of synchronicity is internal to the network, or if it is imposed by externally induced time cues. All that the units feel is an input with a temporal modulation. Therefore, the external temporal structure can be as strong an indicator for visual processing as internally generated synchronicity. That external time cues can help the brain in object processing has been confirmed experimentally in [1, ?].

We do not claim that binding occurs exclusively using synchronicity cues, nor that coherent oscillations must be present in every task that involves binding. Our claim is more moderate, stating that the possibility to label units is a very powerful tool in the framework of the activity-gating architecture, and coherent oscillations are a simple method to implement such labels. Furthermore, binding by synchronicity is often understood as ‘every parts of an object must be in synchrony for the object to be identified’. This is not the case in our model. Instead, a great deal of the recognition processes can occur without synchronicity. Nevertheless, synchronicity is able to provide additional information required for specific tasks, specially those involving several processing pathways.

5 References

- [1] D. Alais, R. Blake, and S.-H. Lee. Visual features that vary together over time group together over space. *Nature Neuroscience*, 1(2):160–164, 1998.
- [2] H.-U. Bauer. Is there parallel binding of distributed objects? Preprint, 1994.
- [3] D. E. Broadbent. Task combination and selective intake of information. *Acta Psychologica*, 50:253–290, 1982.
- [4] A. Chauduri. Modulation of the motion aftereffect by selective attention. *Nature*, 344:60–62, 1990.

- [5] F. Crick. Function of the thalamic reticular complex: The searchlight hypothesis. *PNAS*, 81:4586–4590, 1984.
- [6] F. Crick and C. Koch. Constraints on cortical and thalamic projections: The no–strong–loops hypothesis. *Nature*, 391:245–250, 1998.
- [7] R. Desimone. Neural Circuits for Visual Attention in the Primate Brain. In G. A. Carpenter and S. Grossberg, editors, *Neural Networks for Vision and Image Processing*, Bradford Book, pages 344–364. MIT Press, Cambridge, Massachusetts, 1992.
- [8] R. Desimone and J. Duncan. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.*, 18:193–222, 1995.
- [9] J. Duncan and G. W. Humphreys. Visual search and stimulus similarity. *Psychol. Rev.*, 96:433–458, 1989.
- [10] R. Eckhorn, H. J. Reitboeck, M. Arndt, and P. Dicke. Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex. *Neural Comput.*, 2:293–307, 1990.
- [11] C. W. Eriksen. Attentional search of the visual field. In B. David, editor, *International Conference on Visual Search*, pages 3–19, 4 John St., London, WC1N 2ET, 1988. Taylor and Francis Ltd.
- [12] C. W. Eriksen and J. D. St. James. Visual attention within and around the field of focal attention: A zoom lens model. *Percept. Psychophys.*, 40:225–240, 1986.
- [13] D. J. Felleman and D. C. van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1:1–47, 1991.
- [14] S. P. Gandhi, D. J. Heeger, and G. M. Boynton. Spatial attention affects brain activity in human primary visual cortex. *PNAS*, pages 3314–3319, 1999.
- [15] W. Gerstner. Populations of spiking neurons. In W. Maass and C. M. Bishop, editors, *Pulsed Neural Nets*, chapter 10, pages 261–296. MIT Press, Cambridge, MA, 1998.
- [16] W. Gerstner and J. L. van Hemmen. Coding and information processing in neural networks. In E. Domany, J. L. van Hemmen, and K. Schulten, editors, *Models of Neural Networks II*, pages 1–93, Berlin, 1994. Springer.
- [17] W. Gerstner, J. L. van Hemmen, and J. D. Cowan. What matters in neuronal locking? *Neural Comput.*, 8(8):1653–1676, 1996.

- [18] C. M. Gray and D. A. McCormick. Chattering Cells: Superficial Pyramidal Neurons Contributing to the Generation of Synchronous Oscillations in the Visual Cortex. *Science*, 274:109–113, 1996.
- [19] D. O. Hebb. *The Organization of Behavior*. Wiley, New York, 1949.
- [20] S. Herculano-Houzel, M. H. J. Munk, S. Neuenschwander, and W. Singer. Precisely Synchronized Oscillatory Firing Patterns Require Electroencephalographic Activation. *The Journal of Neuroscience*, 19(10):3992–4010, 1999.
- [21] J. M. Hupé, A. C. James, B. R. Payne, S. G. Lomber, P. Girard, and J. Bullier. Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, 394:784–787, 1998.
- [22] W. James. *The Principles of Psychology*, volume 1. Henry Holt, New York, 1890.
- [23] B. Julesz. Toward an axiomatic theory of preattentive vision. In G. M. Edelman, W. E. Gall, and W. M. Cowan, editors, *Dynamic aspects of neocortical function*, pages 585–612. Neurosciences Research Foundation, 1984.
- [24] R. Kempter, W. Gerstner, J. L. van Hemmen, and H. Wagner. Extracting oscillations: Neuronal coincidence detection with noisy periodic spike input. *Neural Comput.*, 10:1987–2017, 1998.
- [25] C. Koch and T. Poggio. Multiplying with Synapses and Neurons. In T. McKenna, J. Davis, and S. F. Zornetzer, editors, *Single Neuron Computation*, chapter 12, pages 315–345. Academic Press, 1992.
- [26] P. König, A. K. Engel, P. R. Roelfsema, and W. Singer. How Precise is Neuronal Synchronisation? *Neural Comput.*, 7:469–485, 1995.
- [27] M. J. M. Lankheet and F. A. J. Verstraten. Attentional modulation of adaptation to two-component transparent motion. *Vision Res.*, 35:1491–1412, 1995.
- [28] P. McLeod, J. Drive, and J. Crisp. Visual search for a conjunction of movement and form is parallel. *Nature*, 332:154–155, 1988.
- [29] J. Moran and R. Desimone. Selective attention gates visual processing in the extrastriate cortex. *Science*, 229:782–784, 1985.
- [30] B. C. Motter. Neural correlates of feature selective memory and pop-out in extrastriate area V4. *Journal of Neuroscience*, 14:2190–2199, 1994.

- [31] T. D. Murphy and C. W. Eriksen. Temporal changes in the distribution of attention in the visual field in response to precues. *Perception and Psychophysics*, 42:576–586, 1987.
- [32] E. Niebur and C. Koch. A model for the neuronal implementation of selective visual attention based on temporal correlation among neurons. *Journal of Computational Neuroscience*, 1(1):141–158, 1994.
- [33] E. Niebur, C. Koch, and C. Rosin. An oscillation-based model for the neural basis of attention. *Vision Research*, 33:2789–2802, 1993.
- [34] K. M. O’Craven, B. R. Rosen, K. K. Kwong, A. Treisman, and R. L. Savoy. Voluntary attention modulates fMRI activity in human MT-MST. *Neuron*, 18:591–598, 1997.
- [35] D. R. Patzwahl, U. J. Ilg, and S. Treue. Switching attention between transparent motion components modulates responses of MT and MST neurons. *Soc. Neurosci. Abstr.*, 24:649, 1998.
- [36] M. I. Posner. Orienting of Attention. *Q. J. Exp. Psychol.*, 32:3–25, 1980.
- [37] J. H. Reynolds, L. Chelazzi, and R. Desimone. Competitive Mechanisms Subserve Attention in Macaque Areas V2 and V4. *The Journal of Neuroscience*, 19(5):1736–1753, 1999.
- [38] K. S. Rockland and D. N. Pandya. Laminar origins and termination of cortical connections of the occipital lobe in the rhesus monkey. *Brain Research*, 179:3–20, 1979.
- [39] P. R. Roelfsema, V. A. F. Lamme, and H. Spekreijse. Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, 395:376–381, 1998.
- [40] S. M. Sherman and R. W. Guillery. On the actions that one nerve cell can have on another: Distinguishing “drivers” from “modulators”. *PNAS*, 95:7121–7126, 1998.
- [41] W. Singer and C. M. Gray. Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, 18:555–586, 1995.
- [42] D. C. Somers, A. M. Dale, A. E. Seiffert, and R. B. H. Tootell. Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. *PNAS*, 96:1663–1668, 1999.
- [43] E. A. Stern, D. Jaeger, and C. J. Wilson. Membrane potential synchrony of simultaneously recorded striatal spiny neurons *in vivo*. *Nature*, 394:475–478, 1998.

- [44] Y. Sugase, S. Yamane, S. Ueno, and K. Kawano. Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, 400:869–873, 1999.
- [45] A. Treisman and G. Gelade. A feature–integration theory of attention. *Cognitive Psychology*, 12:97–136, 1980.
- [46] A. Treisman and H. Schmidt. Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14:107–141, 1982.
- [47] S. Treue and J. H. R. Maunsell. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature*, 382:539–541, 1996.
- [48] S. Treue and J. H. R. Maunsell. Effects of Attention on the Processing of Motion in Macaque Middle Temporal and Medial Superior Temporal Visual Cortical Areas. *The Journal of Neuroscience*, 19(17):7591–7602, 1999.
- [49] S. Treue and M. Trujillo. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399:575–579, 1999.
- [50] S. Ullman. Sequence Seeking and Counterstreams: A Model of Bidirectional Information Flow in the Cortex. In C. Koch and J. L. Davis, editors, *Large-Scale Neuronal Theories of the Brain*, chapter 12, pages 257–270. MIT Press, Cambridge, Massachusetts, 1994.
- [51] M. Valdez-Sosa, M. A. Bobes, V. Rodriguez, and T. Pinilla. Switching attention without shifting the spotlight: Object-based attentional modulation of brain potentials. *J. Cog. Neurosci.*, 10:137–151, 1998.
- [52] D. C. van Essen, C. H. Anderson, and D. J. Felleman. Information processing in the primate visual cortex: An integrated systems perspective. *Science*, 255:419–423, 1992.
- [53] D. C. van Essen and E. A. DeYoe. Concurrent Processing in the Primate Visual Cortex. In M. S. Gazzaniga, editor, *The Cognitive Neurosciences*, chapter 24, pages 383–400. MIT Press, Cambridge, MA, 1995.
- [54] D. C. van Essen and J. H. R. Maunsell. Hierarchical organization and functional streams in the visual cortex. *Vision Research*, 24:429–448, 1983.
- [55] J. M. Wolfe, K. R. Cave, and S. L. Franzel. Guided search: An alternative to the feature integration model for visual search. *J. Exp. Psychol.*, 15:419–433, 1989.