

A neural model of preattentional and attentional visual search

Nabil Hassoumi, Emmanuel Chiva and Philippe Tarroux¹

*AnimatLab - Département de Biologie - Ecole Normale Supérieure
46, rue d'Ulm
75230 Paris cedex 05 - France*

Abstract

Visual processes do not amount to a simple filtering process performed by a series of hierarchical modules. They allow to select the items immediately useful for the current action from the information included in the external scene. To perform this selection, attentional top-down controls must combine with bottom-up information issued from the retina. In the prospect to understand how these informations are fused together, a computational model of the first steps of the visual process able to account for the pre-attentional and attentional mechanisms involved in visual search has been developed. This model, called Competitive Search, integrates the dynamical aspects of a local dynamical architecture. It accounts for 'pop-out' and attentional phenomena involved in the search for conjunctive targets without introducing ad hoc hypothetical mechanisms such as the attentional spotlight hypothesis. It suggests that such metaphors, issued from the conventional cognitive psychology, may in fact hide simple anatomo-functional arrangements of the circuits that process the visual information. The consideration of these metaphors as covering a specific mechanism is misleading. In the case investigated here, the dynamics of the neuronal circuits may by itself provide an explanation for the observed phenomena associated with the visual search of conjunctive targets.

Running head: Competitive Search Model

Introduction

Perception for action

The interpretation of the visual system as a tool designed to provide the animal with the information useful for action dates back from the onset of modern psychology (James, 1890). Several studies have illustrated this point of view (Allport, 1987; Gibson, 1979; Gibson, 1986). However, the influence of computational approaches has led to models in which the visual system is considered as a passive receptor, a set of filters organized in a modular way (Marr, 1982). This conception gave rise to computer vision systems organized as hierarchic and modular architectures (Milanese, 1993). It has also been very influential on our conception of the biological visual system itself. Unfortunately, artificial systems based on these principles do not compare favorably with natural ones. The common interpretation of this setback was the inadequacy of traditional computer architectures to the modeling of the massively parallel organization

¹To whom correspondance should be sent

of the nervous system. However, according to Sloman (1989), these difficulties uncover a misunderstanding of the real nature of natural vision mechanisms. Sloman indeed puts forward the conception that visual processes should be more integrated with the previously memorized informations, short term memory and current behavior than they are in the modular conception issued from Marr's work.

Attentional mechanisms and action

The modular approach promotes an analysis of the visual system independent from the functional context in which this system has evolved. James's conception, on the contrary, leads to consider that the visual system filters out the informations which are immediately useful for the action from the information flow reaching the retina. One consequence of this point of view is that the same scene will be differently perceived when it is viewed in different contexts. Attentional mechanisms are thus mandatory for the achievement of complex visual tasks as tools to select important cues in the flow of information reaching the visual areas.

Two main attentional theories have been proposed so far. The first one emphasizes the role of attention as a way to reduce the computational load of the visual system through the selection of the sole information useful for the achievement of the current task. The second one puts forward the idea that attention can be interpreted as a mean to set our perceptions in accordance with our internal expectancies (Allport, 1989; Allport, 1987). While the first interpretation has mainly been promoted by designers of artificial vision systems who are indeed confronted with this problem, the second one relies on behavioral considerations.

In this respect, attention should not be considered as the simple ability to focus towards potentially interesting stimuli to process them one at a time. It should rather be viewed as an active filtering process which tunes the perceptual filters in such a way that they become more efficient at detecting the targets related to the current goal. Thus, conflicting with the traditional informational theory approach, one has to consider a behavioral theory of attention. While the former considers that attention is required to solve some informational bottleneck through an internal processing, the later promotes the idea that a perceptual tuning controlled by the current goal is set up at the very early levels of information income.

Psychophysical and neurobiological framework

Early psychophysics experiments distinguished between preattentive and attentional mechanisms. Preattentive target identification seems to be an automatic, effortless and unavoidable effect mainly driven by bottom-up influences (pop-out). On the contrary, attentional mechanisms are related to voluntary search for complex stimuli that usually do not pop-out. These last phenomena are modulated by top-down controls.

Among these controls, the ones issued from memory seem to play such a major role that attention and memory involvements can be considered as two facets of the same mechanism (Desimone, 1996). During a behavioral task, the identification, by the visual system, of a target appropriate for the completion of the task first activates a working memory system in which a representation of the target is elicited. This memory trace is used to generate top-down controls that influence and prime the perceptual system.

Under these controls, this system dynamically acquires the ability to discriminate a specific target among complex and numerous stimuli. In this sense, and if we assume that vision is not the main modality which elicits the current goal, visual processes may always be bottom up, since the top-down influences always precede the visual processing itself.

An important issue is to identify to what extent are these top-down influences used to scan the external scene such that only a limited portion of the available space is enabled to be processed by a limited-capacity processor or to select relevant objects according to the current goal. This is the main question investigated in this paper.

Previous models

FIT

Treisman has shown that an attentional target defined by a conjunction of characteristics not only do not pop-out but also do exhibit a detection time proportional to the number of distractors. To explain these observations, she proposed the Feature Integration Theory (FIT). According to this theory, the different features composing an object may be bound together and sent to a retinotopically organized saliency map. An attentional spotlight is introduced that serially examines the salient locations on this map to determine the most interesting among them. As proposed by Koch and Ullman (1985), this target may be determined through the use of a Winner-Take-All (WTA) mechanism. This serial scan hypothesis, in which a spatially focused spotlight is supposed to screen out a saliency map, has been termed the spatial attention hypothesis (Treisman, 1996).

One of the main support of this theory is the fact that the detection time for conjunctive targets appeared to be roughly twice as high in blank trials than when the target was present. This result is effectively consistent with the hypothesis of a self-terminating scanning mechanism exploring the whole visual scene in search for the target.

Guided Search

Wolfe proposed the 'Guided Search hypothesis' to explain some anomalous results that Treisman's FIT fails to explain (Wolfe, Cave and Franzel, 1989). A reexamination of Treisman's experimental conditions has led Wolfe to conclude that the classical condition for conjunctive search is not as difficult as found by Treisman (Cave and Wolfe, 1990). On the contrary, in controlled conditions, it gives rise to shallow slopes for the relationship linking reaction times to the number of distractors. With high contrast stimuli, the apparent speed of the sequential process is very high (up to 6.1 msec./item compared with about 25 msec./item in Treisman's experiments). Wolfe found that among Treisman's conditions, the only ones that exhibit steep slopes are those corresponding to complex conjunctions (as the search for a T among L's). He concluded that one has first to explain why, in a vast majority of conjunctive experiments, the results are easier than reported by Treisman.

Other results that remain unexplained in Treisman's FIT relate to the high variability of the response of individual subjects. In the case of conjunctions between color and form used in Treisman's experiments, the response time in blank trials seems to be only

weakly correlated with the response time in target trials. Ratios between these durations far less than twice the response in the presence of the target are indeed very often observed. These ratios are strongly subject-dependent. For naive subjects, they range from 0.286 to 6.350. This dependency may reflect the strategy used by the subject to decide on the absence of the target. In our sense, this observation weakens the serial scan hypothesis, the self-termination of which makes on the contrary expect a strict 2:1 ratio.

Furthermore, Treisman's FIT fails to explain that the difficulty of a conjunction task is greatly reduced when the distractors of one type are kept constant while varying the number of the other type of distractors (unconfounded search: Egeth, Virsi and Garbart (1984)). It predicts that triple conjunctions are no more difficult than double conjunctions. As observed by Wolfe, the triple conjunction search is more difficult when target and distractors exhibit a great number of similarities, while this difficulty is reduced when the target shares only one characteristics with the distractors (Quinlan and Humphreys (1987), Treisman and Sato (1990)).

Although it has been suggested that these last results could be explained by the presence of double conjunction detectors (Treisman and Sato, 1990; Nakayama and Silverman, 1986), a purely serial scan seems to be inadequate to account for all the limitations of the standard FIT. Wolfe thus proposed to reconcile preattentive and attentional models by considering a coupling mechanism between the parallel, preattentive search and the serial attentional scanning. The hypothesis is based on the assumption that the parallel preattentive mechanism can be used to guide the spotlight in its search for the target, thus reducing the corresponding search time.

The need for a new model

We are going now to examine the limitations of these models that have led us to the idea that new proposals are required.

All these models are exclusively based on data collected from psychophysical experiments. They are then expressed in psychological terms with a few connections with the underlying neuronal mechanisms (Treisman, 1980). They use a series of black boxes reminiscent from the modular conception that remain to be explained in terms of the underlying neuronal circuitry. Thus, they address the question of the nature of the top-down signal in an *ad hoc* way. The modular view indeed leads to propose models based on the assumption that outside the considered module there is always a solution that affords the requirements needed to make this module functional. The attentional spotlight hypothesis is a significant example of this way of thinking, since it implies the existence of an external mechanism that drives the spotlight. As it was introduced, the attentional spotlight indeed appears to be an *ad hoc* hypothesis. Like every homonculus-based explanation, it only raises a new question (what mechanism drives the spotlight?) the answer to which is let to the top-down influences supposed to use high level information to control the system. As we will attempt to demonstrate it in this study, such a hypothesis may be unnecessary.

As the saliency map assumes the existence of a WTA, the routing mechanisms proposed by Olshausen, Anderson and Van Essen (1995) assumes the existence of shifter circuits and supposes an iconic memory of the object to be recognized coded in object-centered coordinates. All these hypothetical external mechanisms, albeit

endowing the model with its essential properties, are let to further studies. Thus, in these approaches, the essential part of the problem posed remains unsolved. Such models fail to propose an implementation for the top-down controls and do not take into account the numerous neurophysiological data concerning the identification of neurons modulated by attention.

FIT supposes the existence of a feature integration mechanism and thus poses the 'binding' problem at a very low integration level in the visual system (Treisman, 1988). In Treisman's early proposals, binding is necessary to achieve the integration of features. Alternative hypotheses make it a selection mechanism induced by attention (Eckhorn and Schanze, 1991; Fujii, Ito, Aihara, Ichinose and Tsukada, 1996) or a way to enhance the saliency of an item (Singer and Gray, 1995). In this respect, the most recent studies consider binding rather as a consequence than a cause of the attentional mechanisms.

The existence of a conspicuity map raises the question of how the information is selected to build it up (see for instance the critics of this blackboard approach in Green (1991)). It is difficult to imagine that this selection is made only on the basis of the bottom-up flow. Top-down control signals generated by the higher areas holding a representation of the goal seem to be necessary to select the interesting target. This selection process is not compatible with the selection of multiple locations on the basis of their relative saliency. We have to choose between two unrelated mechanisms: one, bottom-up, building the conspicuity map on the basis of the information extracted from the data flow; the second, top-down, generated by the current goal. One needs the elaboration of a conspicuity map only in the first case where several salient elements are identified in the external scene to be further filtered through a WTA mechanism. If we suppose that the active tuning of the perceptual filters is elaborated from the centers in close relation with the aim of the animal, there is no reason to suppose that each location on the visual scene is screened at the same level of precision in search of the suitable target. Only a few positions are likely to correspond to this target and the most salient one (i.e. the most easily reachable) is identified first with a very simple WTA implemented 'on the flow'.

These apparently contradictory points of view result from a confusion between the two roles devoted to the visual system: the first one is to provide information to the brain as any other sensory modality, the second to select the information useful for the current action from the external scene. In these two situations, the saliency of an element has different meanings. In the first one, it derives from the integration of bottom-up information, while in the second, it results from a top-down goal directed control.

Though it represents a computationally implementable model, 'Guided Search' has its own drawbacks. To our opinion, it fails to propose a biologically relevant mechanism to explain the variation in response time with the number of distractors. It assumes the existence of a stochastic search mechanism which is not rooted to any biological implementation. Besides, it adopts the same controversial explanation as Treisman for the serial search: it explains the delay observed in conjunction searches by a serial scan mechanism. Thus, this hypothesis remains to be validated in this model too. There is no proposal concerning how and where in the visual system the global information concerning the saliency of the stimulus is computed. We show here that it can be very simply derived from the local information reaching each receptor field.

Single electrode recording experiments have allowed to identify neurons with a strong behavioral component in various areas of the visual system. These neurons are the targets for top-down attentional modulations. Since the princeps experiment by Moran and Desimone (1985) identifying such neurons in the area V4, several works (Fuster, 1990; Motter, 1993; Desimone, 1996; Maunsell and Ferrera, 1995) have reported such effects in various visual areas (V1, V2, V4, IT, MT, PP). Surprisingly, some of these neurons have been found in early layers of the visual system. Such locations are difficult to concile with the late processing hypothesis (Duncan, 1980; Palmer, Ames and Lindsley, 1993, Usher and Niebur, 1996) implicitly adopted by FIT and Guided Search models. Another contradictory finding with the saliency map hypothesis is the recent report by Humphreys, Romani, Olson, Riddoch and Duncan (1994) that even when attentional mechanisms exhibit a spatial component, they seem to be based on the nature of the objects rather than on their location.

We thus designed the present model to test alternative hypotheses accounting for the observations made with visual search paradigms and neurophysiological records concerning behavioral modulations of neurons in the early stages of visual processing. We tried to avoid the introduction of *ad hoc* hypotheses (such as the presence of a spotlight acting on an internal screen) unless they seemed absolutely necessary to obtain a self-consistent model.

We started from the point of view that the model should represent a neurobiologically plausible implementation of the first steps of the attentional controls.

Model description

General outlines

The present model is based on a multilayer recurrent neural network architecture the overall organization of which is shown figure 1. It assumes that feature extraction is achieved through an automatic wired process compatible with the algorithmic operations of filtering performed from the retina to the striate cortex. The model thus assumes the existence of a collection of parallel filters for the extraction of the different features. However, information processing is supposed to be similar in each one of these feature (or modality) channels. Thus, the model can be easily extended to the extraction of any set of basic modalities suitable for the description of a visual scene. The consideration of any basic features to which a particular set of cells in the visual cortex is known to be sensitive would have led to similar conclusions.

The resulting feature maps separately project onto a set of two coupled F1 and F2 neuronal maps.

F1 neurons send excitatory connections to F2 neurons which send back recurrent inhibitory connections to F1. The activity in F1 is also modulated by an aspecific inhibitory back connection.

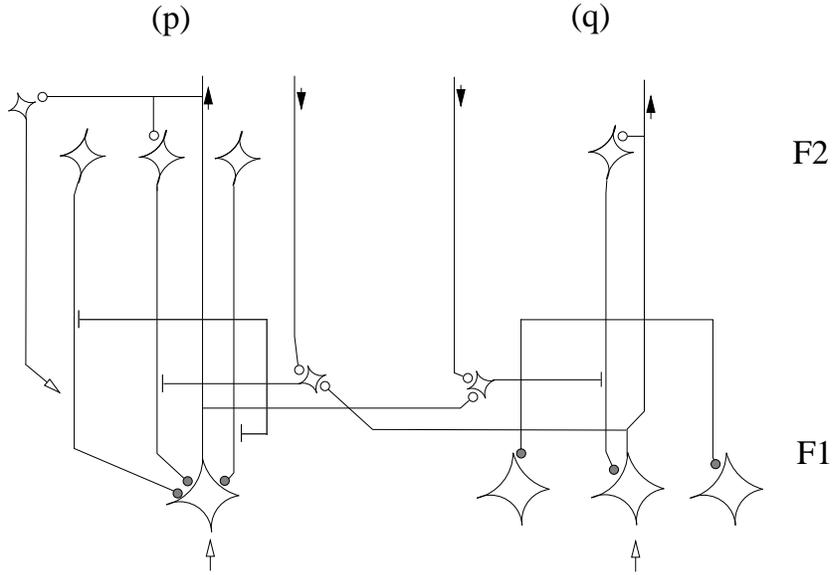


Figure 1 - The organization of the neuronal maps of the Competitive Search model. Neurons in F1 map are excitatory while neurons in F2 map are inhibitory. Additional globally inhibiting interneurons received inputs from F1 neurons and send back inhibitions to F1. Control signals bearing the description of the target are fed back to F1 neurons through top down connections. These activities modulate the inhibition of F1 maps by F2 maps through the modulation of gating interneurons. The maps are coupled through lateral connections issued from F1 neurons.

As a good compromise between biological realism and computational tractability, we used 'leaky integrator' neurons. The membrane potentials of F1 and F2 neurons are computed from the following equations:

$$\tau_1 \frac{dV_i^{(1p)}}{dt} = -V_i^{(1p)} + w^{(21)} f_2(V_j^{(2p)}) + R \left(I_i(t) - K I_{(p)}^{inhib} \left(1 - f_1(V_i^{(1p)}) \right) \right)$$

$$\tau_2 \frac{dV_i^{(2p)}}{dt} = -V_i^{(2p)} + w^{(12)} f_1(V_j^{(1p)})$$

where $V_i^{(kp)}$ denotes the membrane potential of the i th neuron of the k th (F1 or F2) map corresponding to the modality p . $I_i(t)$ is the input current on F1 maps superimposed with a blank noise and the w_{ij} represent the synaptic weights connecting F1 and F2 maps. R is the input resistance of the F1 neuron and K a normalization constant.

The mean firing rate of each neuron is computed from the logistic non linear functions f_1 and f_2 . The parameters of these functions have been adjusted so that the integration constant of inhibitory neurons is slower than the one of excitatory neurons (table 1).

The global inhibitory current is directly obtained from the activities in F1 through the following equation:

$$I_{(p)}^{Inhib} = \sum_j f_1(V_j^{(1p)})$$

without any explicit modeling of the activity of this inhibitory interneuron. This current is thus proportional to the sum of the outputs of F1 neurons, i.e. to the number of detected items on a given map.

Unless otherwise stated, the main parameters have been set to the values given in table 1.

1.1	R
0.4	K
0.8	$w^{(12)}$
-0.8	$w^{(21)}$
12.0	τ_1
8.0	τ_2
0.3	Cp
0.3	Cq
0.2	b
0.08	t_1
0.1	t_2
0.2	s_1
0.3	s_2

Table I - Parameter values. Unless otherwise stated, the main parameters of the model are adjusted to the values given in this table. t_i and s_i are respectively the threshold and the slope for the f_i logistic functions. The expression of time in msec. has been obtained through an appropriate scaling of the integration constants.

Results

General behavior

When stimulated with a noisy input, the model exhibits a complex dynamic behavior: due to the presence of inhibitory feedbacks, F1 neurons enter an oscillatory state. The presence of a global inhibitory activity applied on each feature map introduces a competition between the items. This inhibitory mechanism implements a dynamical winner-take-all the action of which is to force the F1 neurons to be activated only one at a time. Thus, neurons coding for every stimulus item are activated in turn, implementing a sequential stochastic exploration of the different items present in the stimulus. However, this sequential activation is not the consequence of an external serial mechanism as in FIT or Guided Search. It only results from the competition between the activities of the different neurons. For this reason this mechanism has been termed Competitive Search.

Pop-out

We tried to reproduce first the experimental results corresponding to pop-out situations. In this case, the stimulus used consists in a target differing from the distractors by a unique characteristics. Thus, the target projects alone on its feature map.

The corresponding neurons rapidly exhibit a high level of sustained activity (Fig 2). On the distractor map, the competition enables the items to be active only one at a time, thus forcing the neurons to enter a bursting intertwined activity. The speed at which target neurons reach their high level of activity does not depend on the number of distractors. There is no competition between target and distractor activities (Fig 3).

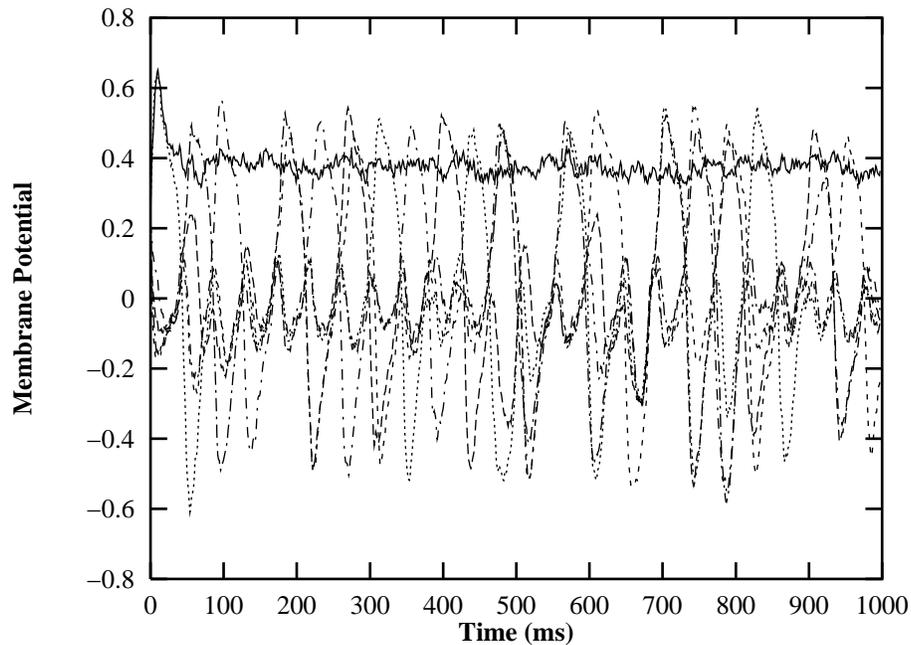


Figure 2 - Mean firing rates of neurons corresponding to the target (solid line) and to some distractors (dashed lines) in a pop-out condition.

Conjunctions of characteristics

On the contrary, when a target is defined by a conjunction of characteristics, the presence of a signal issued from the distractors lying onto the same map as the target tends to inhibit the activation of the corresponding neurons. Both target and distractor neurons produce bursts, and the target is no longer distinguishable from the distractors (Fig. 3A). This results reproduces the situation of a conjunctive search experiment in the absence of an indication given to the subject on the composition of the target. It is assumed here that this information is provided to the layers on which the present mechanism is implemented as a top-down neuronal control. To account for such top-down controls, we assumed the existence of a shunting connection acting on the F2 neurons responding to the characteristics of the target (Fig 1).

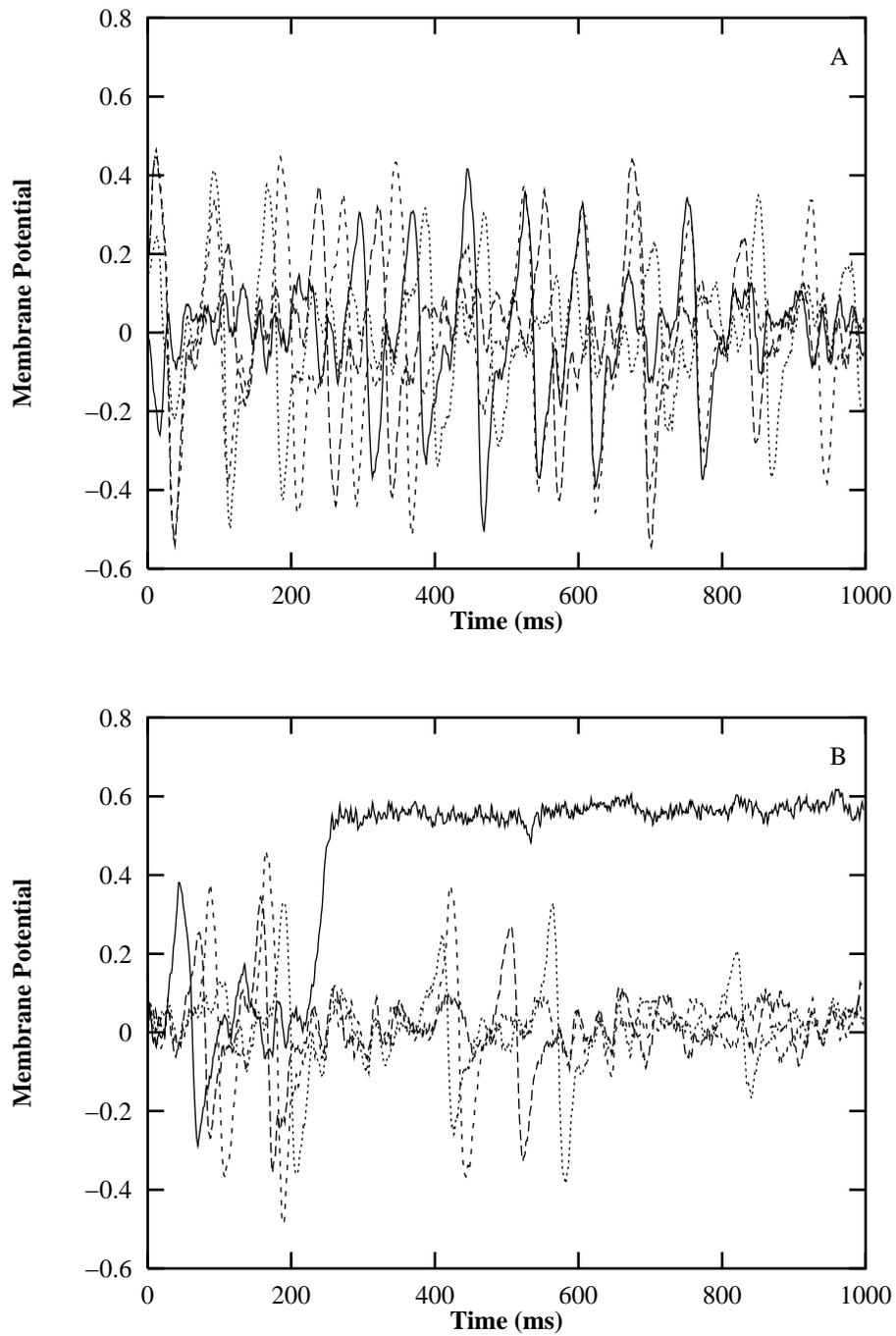


Figure 3 - Membrane potential of F1 neurons when the target is defined by a conjunction of characteristics. Target is thus characterized by the features p and q while distractors are either \bar{p}, \bar{q} or \bar{p}, q . A: no top down modulation was introduced: $C_p = C_q = 0$. B: in the presence of a modulation: $C_p = C_q = 0.3$

Such connections are introduced through the following modification of the initial equations governing the activity of the F1 neurons:

$$\tau_1 \frac{dV_i^{(1p)}}{dt} = -V_i^{(1p)} + w^{(21)} f_2(V_j^{(2p)}) \left(1 - C_p \left(\sum_k f_1(V_i^{(1k)}) \right) \right) + R \left(I_i(t) - KI_{(p)}^{inhib} \left(1 - f_1(V_i^{(1p)}) \right) \right)$$

The inhibitory activity from F2 neurons is thus modulated by a coupling term (corresponding to the activity of the inhibitory gate interneuron reprinted on Fig 1). This term depends on a top-down control (C_p) and on the activity of the other maps at the same location. A high value for this parameter can be interpreted as the fact that the feature p is present in the target to be found. Thus, we avoid the use of explicit coupling coefficients, which are otherwise unrealistic for combinatorial reasons. As it is modeled here, the top-down control increases the coupling of the p map with every other map activated by the display. However, we verified that, for every k such that C_k is set to 0, maps p and k do not succeed in synchronizing their activity. Consequently, in the expression, we omitted every k term for which $C_k = 0$.

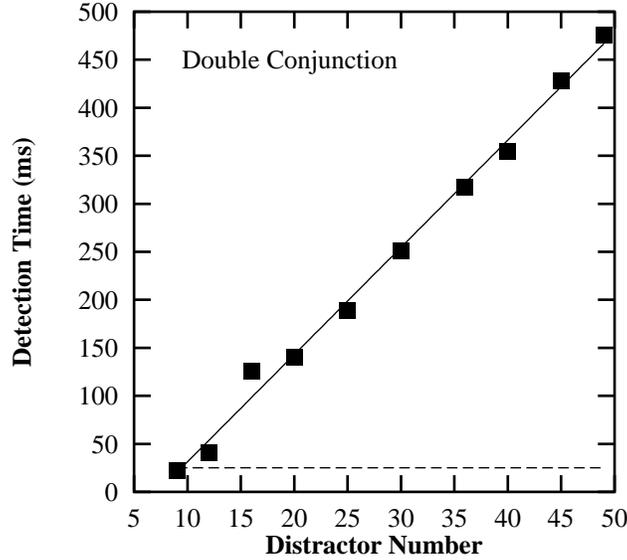


Figure 4 - Reaction times according to the number of distractors in a double conjunction experiment. The dashed line gives the behavior of the pop out targets. The slope of the fit line is 11.16 msec/distr.

In the presence of the top-down modulation signal, neurons receiving inputs from the target exhibit an ability to synchronize their bursting behavior such that they escape the oscillatory process and get locked to a high firing rate (fig 3B).

When simultaneously activated, neurons corresponding to the target on both p and q maps tend to maintain their activity at a high rate. However, due to the competition between target and distractors, the synchronization of these neurons is delayed according to the number of distractors. In these conditions, the model exhibits increasing time lags before F1 neurons get locked into a sustained activity due to the coupling mechanism between p and q maps. This delay spans over a time period of 10

msec to 0.6 sec for a density of distractors ranging from 10 to 50. It is approximately linear over the entire distractor density range used in this study (Fig 4).

We assumed here that this delay holds for the variable part of the reaction time as determined in psychophysics experiments. It does not include the fixed part of the delay due to the elaboration of the subject's response. Reported to the number of distractors, the detection time leads to a value of 11.16 msec/distr. which is close to the values reported by Wolfe *et al.* (1989). According to Cave *et al.* (1994) and to Wolfe (1992), we assumed that the canonical values for conjunctive search of basic feature conjunctions are those corresponding to experiments which give shallower slopes than those initially reported by Treisman and Gelade (1980) and Treisman (1986). Thus, Treisman's reports of steep slopes for conjunctive searches still remain to be explained.

One possible preliminary explanation can be provided by the observation that, in our model, the slope values strongly depend on the intensity of the modulation between the characteristics maps. This result can explain both the high variability of the detection rates for different subjects (Wolfe and al., 1989) and the results obtained by Treisman and Gormican (1988). These rates could indeed depend on the basic connection level between the characteristics maps in the different subjects. They can also be affected by low contrast conditions if these conditions interfere with the coupling mechanism. We will discuss this point later.

As shown Fig.3B, attention to a conjunctive target comprises a first competitive phase and a second sustained phase of activity. The first phase is responsible for the increasing detection time with the number of distractors. The second, during which the selection of the target is established, shows a parallel partial suppression of the distractor activity.

It is worth noting that a linear dependency between the reaction time and the number of distractors can unexpectedly be obtained albeit the competition between the target and the distractors is based on a strongly non-linear mechanism. A large corpus of data reported that RTs linearly depend on the number of distractors. However, some authors argued that this linearity is far from being established, whether there are too few experimental points for each curve, or non linear data have been effectively observed.

Guided Search is intrinsically linear, since the probability that an item is the target increases linearly with its saliency. On the contrary, no such implicit linearity is present in Competitive Search. Thus non linear dependencies as those reported in the literature are more likely to occur in Competitive Search than in other models. Actually, it seems that for some values of the parameters, conjunctive searches are slightly facilitated when the number of distractors increases (data not shown). This effect, reminiscent of the facilitation observed by Wolfe *et al.* (1989), can be explained if one assumes that, for a small number of distractors, the competition mechanism is very efficient, while when this number increases it enables two or more items to be simultaneously active. Besides, the competition mechanism greatly depends on the parameter values. How these parameter settings reflect the external conditions of the various psychophysics experiments published in the literature remains to be investigated.

More refined experimental paradigms have given additional insights into the mechanisms at work in conjunctive search. These conditions deserve to be analyzed within the framework of the present model. This is done in the next three sections.

Triple conjunction experiments

As pointed out by Wolfe (1989), standard FIT does not account for the results obtained with triple conjunction targets. Treisman's theory predicts that the difficulty should be identical for targets defined by triple conjunctions than for targets defined by double conjunctions. However, when tested for their reaction time to find a target defined by a triple conjunction in which the target shares one common feature with the distractors, subjects are more efficient at finding this target than for the double conjunction experiments. This result can be explained by observing that, in this case, the modulation connects three maps together and that one third of the distractors are present on the maps where each target feature appears (table 2).

Feature	f_1	f_2	f_3	f'_1	f'_2	f'_3
Target	0	1	0	1	0	1
Distractors 1	1	1	1	0	0	0
Distractors 2	0	0	1	1	1	0
Distractors 3	1	0	0	0	1	1

Table 2 - Target and distractors share one common characteristics. The numbers in the columns indicate the presence (1) or the absence (0) of the target or distractors on the corresponding feature map. f_i and f'_i correspond to the complementary feature maps of the same modality.

For triple conjunctions such that the target shares two characteristics with the distractors, the modulation again connects three maps together and two third of the distractors are present on the maps on which the target appears (table 3).

Feature	f_1	f_2	f_3	f'_1	f'_2	f'_3
Target	0	1	0	1	0	1
Distractors 1	0	1	1	1	0	0
Distractors 2	0	0	0	1	1	1
Distractors 3	1	1	0	0	0	1

Table 3 - Target and distractors share two common characteristics. Same legend as in table I.

One expects that, in these conditions, the higher level of modulation can compensate for the presence of a higher density of distractors thus leading to a value similar to the 11.6 msec/distr. reported by Quinlan and Humphreys (1987). In the case of double conjunctions, two maps are connected by the modulation and only one half of the distractors appears on the same maps as the target. This situation can serve as a reference for the two situations reported above.

The model exhibits the same behavior as the one reported by Wolfe for the 'Guided Search' model (Fig 5). When target and distractors share one characteristics, the detection is easier than for the double conjunction experiment. When the target has two common characteristics with the distractors, the detection is much more difficult. As in Cave's and al. model, the double conjunction response curve lies in between the curves corresponding to the triple conjunction experiments.

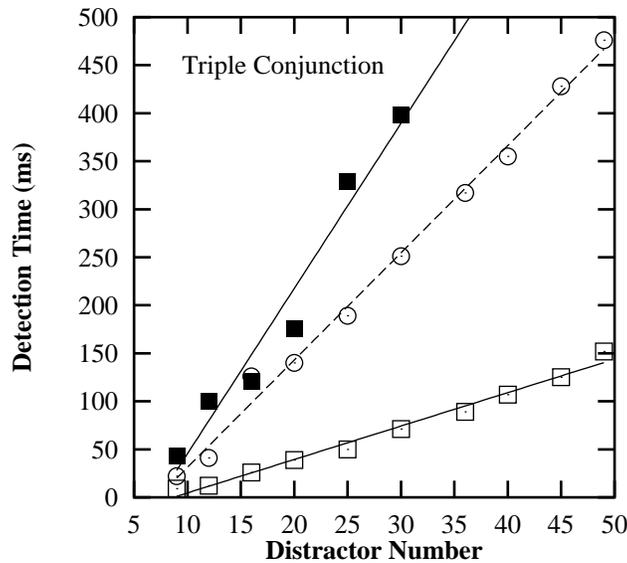


Figure 5 - Comparison between triple conjunction searches in which the target shares one feature (open squares) or two features (closed squares) with the distractors. The open circles correspond to the double conjunctions experiment reported in figure 4.

Cued search

In conjunctive search, a target bearing an additional feature absent from the distractors pops out and becomes easy to identify. This additional feature makes the conjunctive search easy. It can be considered as an internal cue drawing the attention of the subject toward the location of the target. Conversely, when a distractor bears this additional feature, it pops out. However, this pop-out does not significantly interfere with the search for the target.

The present model exhibits these properties. In the first case, the target is immediately detected due to its pop-out along the third dimension. In the second case (Fig. 6), the pop-out of the distractor along the third dimension is not sufficient to confound it with the target. On the feature map it shares with the target, it gives rise to longer bursts and to an enhanced activity. However, this activity is not sufficient to lock itself to a sustained high firing rate. This behavior enables the target to be detected as well as in a conventional conjunction search.

As stated above, Motter recently reported that V4 neurons exhibit attentional modulations mediated by the short term memory of a given cue (Motter, 1994). Luck, Chelazzi, Hillyard and Desimone (1997) reported similar results in a spatial-attention task. These observations suggest that top-down controls provide down to this layer an information on the nature of the feature relevant to the choice of the final item. Such an explanation is consistent with the implementation proposed here. The present model thus assumes that the top-down signals carry out the information describing the feature composition of the target and that these signals are present at early levels before the presentation of the display containing the target.

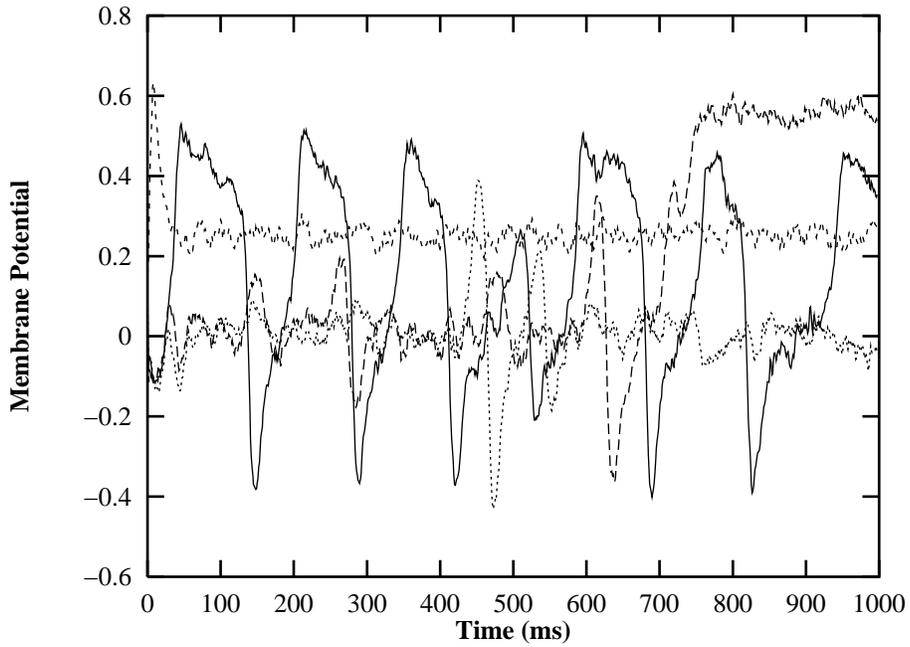


Figure 6 - Non competition between the target and a distractor distinguished in an irrelevant dimension. On the map corresponding to the feature it shares with the target, this distractor shows longer bursts (solid line) than the other distractors (dotted line). In the irrelevant dimension, it pops out (dashed sustained activity). The target is found around 750 msec in this run (long dashes)

However, in the form presented above, the model is adapted to the search for conjunctive targets. To cope with targets discriminated on the basis of a unique feature as in Motter paradigm, it needs to be modified as follow. We assume that the gating interneuron has a spontaneous activity that can be modulated by the top-down signal as previously described. This change leads to the following modification in the coupling term of the equation controlling the F1 neurons:

$$w^{(21)}f_2\left(v_j^{(2p)}\right)\left(1-C_p\left(\sum_k f_1\left(v_i^{(1k)}\right)+b\right)\right)$$

where b is a basic activity of the interneuron. We verified that this adaptation does not affect the results obtained in the previous section where b was set to 0.

In the absence of a signal from other feature maps, a top-down control bearing the information that the corresponding feature is relevant to target identification will produce an attenuation in the inhibition of the corresponding F1 map neurons. As shown on figure 7, the items bearing this feature will be selectively enhanced, while the other items will have a lower discharge rate.

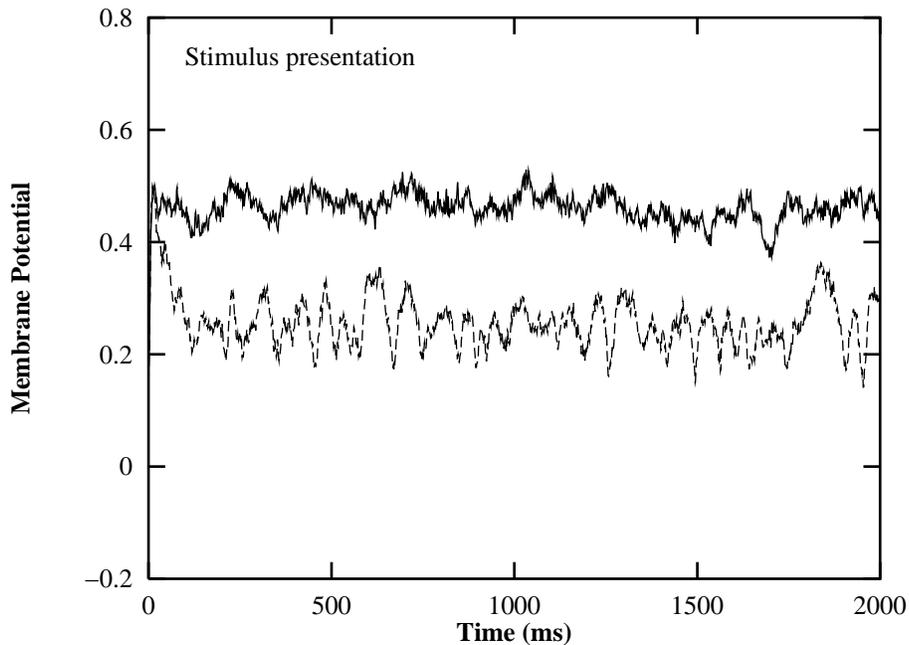


Figure 7 - Behavioral modulations during a conditional discrimination task. The conditions of a conditional discrimination task are simulated. The presence of the cue corresponds to the presence of a high top-down signal for the feature indicated by the cue ($C_p = 0.4$). A neuron selective for this feature (solid line) exhibits a sustained activity. A neuron selective for a non-attended feature shows a lower mean activity.

Depending on the value of the top-down parameter, different degrees of discrimination between the activities of neurons selective to attended and non attended features can be observed (Fig 7): While neurons sensitive to the cued feature show a sustained activity and neurons sensitive to non attended features exhibit a bursting behavior due to the competition between the items, lower C_p values give rise to a bursting behavior for the cued neurons too. However, the duration of the burst phase is increased by the presence of the top-down signal.

Unconfounded search

Egeth *et al.* (1984) showed that, in a conjunctive search, when the number of distractors from one type is held constant while the number of the other type increases, the slope of the reaction time curve vs the number of distractors is much shallower than in a classical conjunction experiment. This situation has been termed unconfounded search. A classical explanation of this phenomenon is that subjects are able to exclude one category of distractors, thus performing an easier search on the other.

Due to the increase of the inhibition issued from the increasing class of distractors, the model reproduces these results: In an unconfounded condition, the detection of a target is made easier than in a confounded condition (Fig 8).

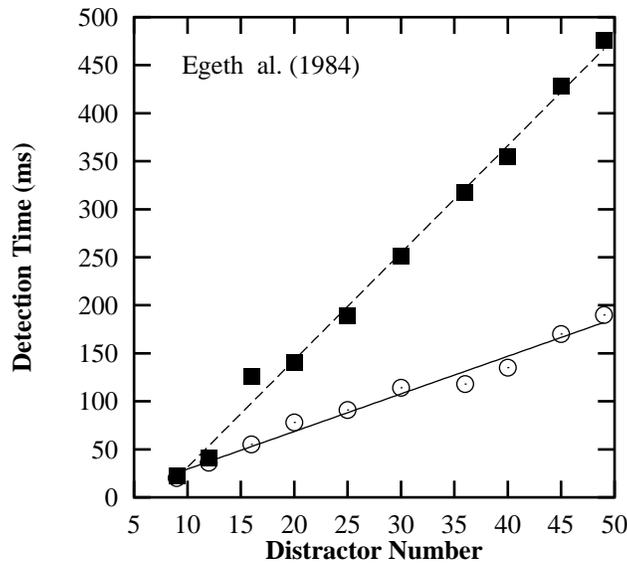


Figure 8 - Comparison between confounded and unconfounded search. Filled squares: unconfounded condition. Open circles: Confounded condition. Parameters values as in table 1 (b=0).

The introduction of a basic activity at the level of the interneuron offers the ability to modulate the responses to the different features in another way. Egeth *et al.* experiments can be interpreted as based on the ability of subjects to exclude one feature modality in the displayed data. Many examples of this ability to consider only a subset of the displayed items are reported in Wolfe (1992). This situation can be easily reproduced here by setting the coupling parameter corresponding to this modality to 0. This situation indeed mimics the exclusion of this modality by a top-down inhibitory control.

Comparison between SOA and RT paradigms

When subjects are tested on the basis of their reaction time (RT), the information they use for the elaboration of their responses is not strictly controlled. Besides, a complete model of the conditions of this paradigm would require to take into account the motor response of the subject. Since it depends on the time needed for the elaboration of this motor phase, the response latency observed in experimental plots cannot be accurately reproduced here. As well as for experimental conditions as for testing the model, stimulus onset asynchrony paradigms (SOA) give a more suitable evaluation of the performances concerning the visual search process in itself. Thus, to check for the coherence of the simulation data obtained in RT conditions, we evaluated the same simulations in an SOA paradigm.

We supposed here that the presence or the persistence of the display in the lower layers of the visual system is the sole source of information on which the decision is elaborated. As in other models (Cave *et al.*, 1990, Tsotsos, 1995), it has been assumed that the constancy of the inputs fits the conditions of the experiments the model has been designed to account for.

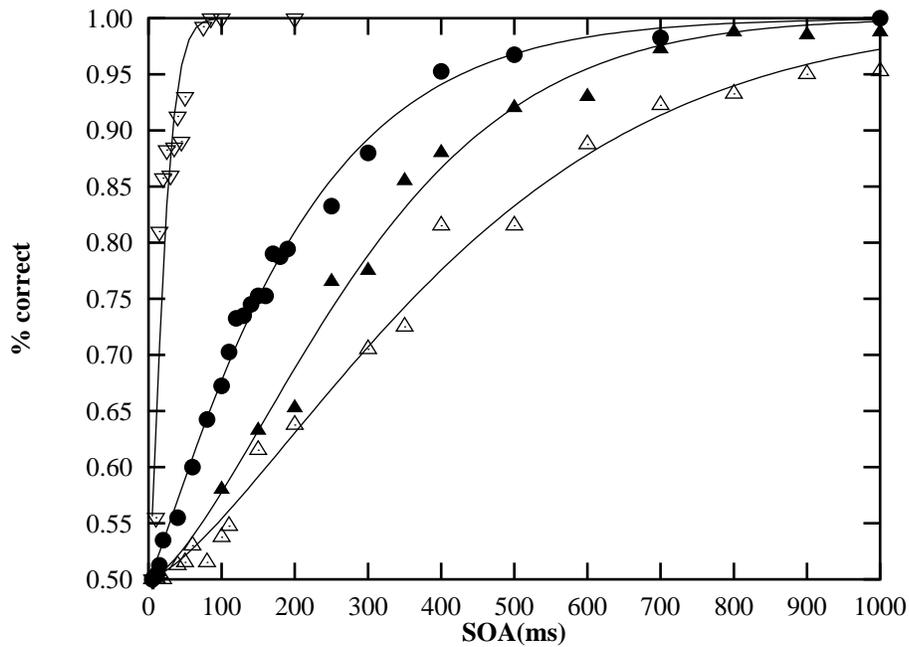


Figure 9 - Determination of threshold SOAs. For each set of trials, the percent of correct responses is plotted against the SOA duration. Each set of points is fitted with the psychometric function $f = (1 - \exp(-(x/\tau)^\sigma))$ (solid lines). The threshold SOA gives the mean performance of the model for each experimental condition (each number of distractors). The different curves correspond to an increasing number of distractors: 9 (open inverted triangles), 25 (closed circles), 36 (closed triangles), 49 (open triangles).

Thus we simulated SOA experimental conditions in the following way: the model is given a variable number of steps to determine the presence of the target. The percent of correct responses for each of these durations is obtained by summing the number of target detections. From these data, the threshold SOA (SOA for which a subject gives a 81.6% correct response) and the slope of the psychometric function are determined (Fig 9). It is observed that the threshold SOA increases with the number of distractors. Figure 10 shows that this increase is linear in the number of distractors. It leads to a very similar slope than the one obtained in RT conditions.

The obtained result fits well the values determined in RT paradigms (12.19msec./distr. in SOA compared to 11.6msec./distr. in RT).

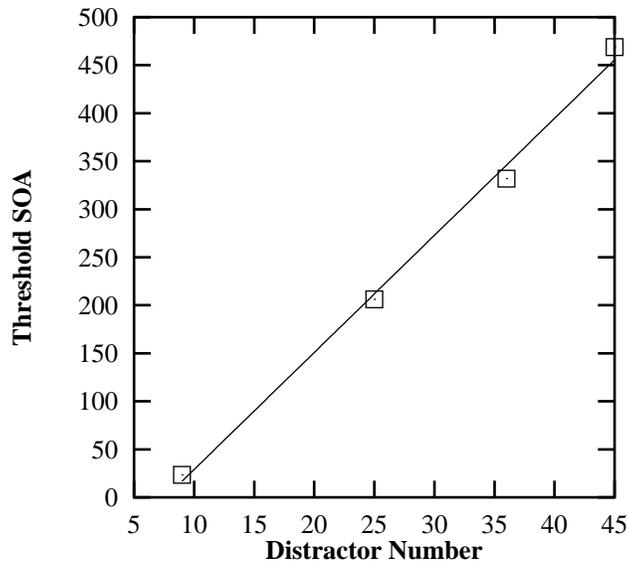


Figure 10 - Reproduction of SOA experimental conditions for a double conjunction search. The figure has been obtained by plotting the threshold SOAs as a function of the number of distractors. Each point correspond to the threshold percent correct responses for 200 trials. The slope of the fit line is 12.19msec./distr.

Discussion and conclusions

Discussion

Biological foundations of the model

Unlike the previous models, the present proposal assumes that the sites at which pop-out and detection of targets defined as conjunctions of basic features occur are located in the same layers of the visual cortex (Wolfe *et al.*, 1989). It also assumes that the attentional controls act down to a relatively low level in the visual system.

The model is based on the assumption that F1 neurons may correspond to neurons sensitive to basic features as are pyramidal neurons found in areas like V2 or V4. F2 neurons are local inhibitory cells fitting the main properties of inhibitory cells in cortical layers. Nevertheless, it is relatively meaningless to allocate a precise correspondance between model and actual cortical neurons. For instance, the coupling mechanism between F1 and F2 layers produces complex dynamical properties, e.g. a strong ability to oscillate. Notwithstanding, these properties can be produced by several alternative organizations, even at the level of the neurons themselves. Similarly, though the intermodality coupling mechanism is mediated here by gating inhibitory interneurons, it is likely that other mechanisms can exhibit equivalent properties.

The model suggests the existence of a crosstalk between pathways processing different features information. A recent study (Tamura and Sato, 1996) has shown that such crosstalk abilities are present to a greater extent in V2 than in V1. As suggested by this study, the existence in V2 of a generalized interaction between modalities may

confer to this area the adequate properties for an attentional enhancement of objects defined by a multiple combination of basic features. The cross correlation phenomena observed in the same study between neurons sensitive to different modalities suggest an underlying anatomical connection. Such a connection, via inhibitory interneurons, is present in the Competitive Search model. However, to propose that it can be sufficient for mediating the synchronization observed by the authors between neurons responding to different local modalities requires further investigations.

We assume here the existence of feedback connections tuning the activity of F1 neurons such that they become more responsive to the presence of a target occurring anywhere in the visual field. This proposal fits well with previous hypotheses concerning the role of feedback connections (Ullman, 1991; Mumford, 1992; Damasio, 1989, 1990; Rolls, 1990; Finkel and Edelman, 1989; Tononi, Sporn and Edelman, 1992; Rockland, 1994). As stated for instance by Damasio (1989), these reciprocal connections may support a kind of locking (through synchronization) between behavioral expectancies and early perception. However, most of these models (Ullman, 1991; Mumford, 1991, 1992; Finkel and Edelman, 1989) propose that this backward stream is essentially used to seek for a sequence of transforms that link source and target representations.

Our model suggests a rather different role for these connections which, in our case, are used to tune the input filters before the arrival of the incoming signal. As suggested by e.g. Motter's or Luck *et al.* works, short-term memory may hold an information on the stimulus to be seek that can be projected back to the extrastriate areas before the appearance of the stimulus. So the visual system may be ready to efficiently filter the relevant information from the current behavior as soon as this information becomes available. Anatomical data are not contradictory with the existence of feedback connections from the areas concerned with a short-term memory representation of the objects (namely the inferotemporal cortex (IT)). As stated by Rockland (1994), area TEO seems to send widespread recurrent connections to area V4, while V4 sends similar connections to V2.

From a functional point of view, several authors reported activations in early layers of the extrastriate cortex in attention-related conditions. Some of these modulations have been monitored through single cell recordings (Moran and Desimone, 1984; Luck *et al.*, 1997; Motter, 1993). More global insights into the involvement of early visual areas in attentional modulations have been obtained by Sakai and Miyashita (1994), Shulman, Corbetta, Buckner, Raichle, Fiez, Miezin and Petersen (1997), and Le Bihan, Turner, Zeffiro, Cuénod, Jezzard and Bonnerot (1993) using PET or fMRI techniques. They all conclude that these modulations are present back to V1. They show at least that the anatomical circuitry can be functionally used to activate lower areas from higher areas. Our model fits well with the above proposal of Sakai and Miyashita that, during a mental imagery task, early visual areas receive top-down attentional signals to modulate their activity.

The model also predicts that the information given back to the feature maps is diffuse and not spatially organized. The recurrent connections described from PIT to V4, and from V4 to V2 seem to have these topological properties (Rockland, Salem and Tanaka, 1994). It also predicts that the target layer of this recurrent information depends on the nature of the feature to be identified in the incoming signal. Shulman study shows that

modulatory effects from higher areas exhibit several task dependancies (Shulman *et al.*, 1997). It suggests that they are supported by feature-specific underlying mechanisms.

As pointed out by Luck *et al.* (1997), the attentional modulations observed in V2 or V4 could be interpreted as a late effect of post-perceptual processing. However, Luck *et al.* experiments seem to lead to the conclusion that these effects are consistent with attentional mechanisms operating before the stimuli have been identified. The conclusion drawn by these authors fits very well with our hypothesis of an early effect of attention.

Perceptual learning

Spatially selective learning poses a serious problem to the spotlight and spatial attention hypotheses. To account for spatially specific learning effects within the framework of the spatial attention theory, it is indeed necessary to assume that the target of learning is the mechanism which drives the spotlight. In this hypothesis, the neurons tuned through learning are those of the master map of location. By construction, this master map of location is a high level map. This implies that in pop-out experiments learning would occur at a high level in the visual system, while a lot of converging evidences demonstrate that it occurs at a low level: As pointed out in a recent work, Ahissar and Hochstein (1996) showed that pop-out performances improve dramatically when performed attentively. The authors conclude that the level of pop-out occurrence must be accessed by some top-down attentional controls. On the other hand, it is well established that pop-out is mainly based on local interactions. Such interactions are just likely to be present in V1 or V2.

According to the conclusions of works reporting pop-out effects related to mechanisms not present in V1 (Ramachandran, 1988; Enns and Rensink, 1990; Wolfe, Yee and Friedman-Hill, 1992; He and Nakayama, 1992), we assume that the vast majority of pop-out phenomena based on the detection of simply characterized targets occur immediately after V1. This assumption has received a general agreement in the past few years (Nothdurft and Li, 1985; Knierim and Van Essen, 1992; Merigan, Nealey and Maunsell, 1993). Another arguments can be derived form the observation that there seems to be an almost complete interocular transfer of these perceptual learning effects (Ahissar and Hochstein, 1996). For that reasons, we suppose that the presence of attentional, top-down controls as early as in area V2 is likely, or at least not contradictory with the present available data.

The present model offers an explanation to these perceptual learning effects. If we assume that attentional controls are sent back to V2, they can both be used to prime feature detectors and be involved in learning and adaptation of these feature detectors. The specificity of the transfer observed by Ahissar and Hochstein (1996) within basic visual dimensions seems to be in favor of the feature-specific attention hypothesis investigated here.

Object- or location-based attention

Two types of attentional mechanisms have been proposed so far. The first one is based on attention to locations (Posner, Snyder and Davidson, 1980) while the second relates to the attention to objects (Duncan, 1984). In spite of a recent reexamination of

this classification (Vecera and Farah, 1994), it seems that some points have to be clarified.

Due to the hypothetical presence of a spotlight visiting each spatial position of the display, attentional mechanisms concerning most of the visual search paradigms have been classified into spatial attention. It has also been considered that spatial cuing paradigms belong to the same class of spatial mechanisms. A hypothetical spotlight could indeed be involved in both the rapid scanning of items in visual search paradigms and the enlightening of a particular position in spatial cuing paradigms.

The present model shows that classical visual search could be based on a mechanism that is not genuinely spatial. The Competitive Search is basically a competition between objects. On the contrary, the competition that occurs at the level of a saliency map to identify the winner is a competition between locations, even if these locations are labeled by the integrated conspicuity of each object. This last competition leads to the elimination of the features belonging to the distractors in a similar way than the one proposed by Treisman and Sato (1990). The competitive mechanism implemented in our model leads to an apparent serial selection that can appear a selection of locations. However, it is a competition between active neurons, that is neurons coding for the presence of objects features. Thus we consider that search for an item the characteristics of which have been previously given to the subjects is a matter for object-based attentional mechanisms. On the contrary, identification of an unknown item at a precued position belongs to location-based attentional processes.

It is worth noting that this spatial aspect of precuing, as it is revealed by the observations of Luck *et al.* (1997) and Motter (Motter, 1993), seem to spatially modulate neurons in the same layers as neurons modulated by object-based attention. These observations suggest that a similar mechanism is used for location-based attention and for object-based attention. It only assumes that the control signals can carry both feature -specific and position-specific- information.

Complex searches

The model fails to explain conjunctive search results when targets are made of conjunctions from the same modality. As stated by Wolfe (1992), in this case (i.e. a T among distracting Ls), this search is often serial, though, in some cases, these conjunctions produce pop-out detections (a cross among vertical and horizontal bars). The explanation given by Wolfe is that T and L stimuli do not exhibit orientation differences that can be discriminated on orientation maps. Consequently, no elementary characteristics enables to distinguish between these stimuli. This explanation rules out the possibility that T and L may be distinguished by higher level characteristics not extracted by V2 neurons or that they can only be defined by multiple conjunctions of elementary characteristics. The similarity between the Ts and the Ls suggests that this last hypothesis could be the most convenient.

Since target and distractors differ only by a few characteristics and share a lot of common features, our model predicts that search will be much more difficult than for a target defined by a simple conjunction of basic features. Thus, at this stage, there are at least two explanations for the steep slopes of these last searches: (i) the characteristics that distinguish the items can only be extracted at a higher level than the layers including basic feature detectors; (ii) the adequate feature detectors exist at a low level

in the visual system but the presence of a lot of common features makes the search much more difficult. The pop out of an 'X' among 'Ts' and 'Ls' (Julesz, 1990) can be explained by the generation by these stimuli of complex features able to selectively activate hypercomplex cells the presence of which is attested in V1 as well as in V2.

We thus follow Wolfe's conclusion that visual search is easy in general (it gives rise to shallow slopes in an RT paradigm). Difficult cases should be explained by specific mechanisms or by the fact they fail to be processed by the mechanism that makes easy the other cases. The present model provides a potential explanation of these discrepancies. Conjunctive search will be easy as long as it will be possible to couple neurons responding to the different features of the conjunction. Neurons responding to simple features can be easily coupled in maps such V1 or V2. However, the model predicts that the search becomes more difficult when the number of shared features between the target and the distractors increases. For those characteristics that are not simultaneously present at the level to which top-down control signals are sent, it is not possible to improve the detection of the target. It should be an explanation to the results reported by Nakayama and Silverman (1986) who showed that in a multiple comparison between basic conjunctions, the only difficult task is to find a target defined by a conjunction of color and movement. If we assume as emphasized by Logothetis (1991), that neurons sensitive to the direction of movement cannot be tuned by top-down controls back beyond MT and that there are no neurons sensitive to color in this area, it is easy to conclude that there is no simple way to couple the informations concerning these two characteristics.

It can be derived from recent studies (Duncan, Humphreys and Ward, 1997) that attentional effects should be viewed as gradual effects affecting most of the visual areas rather than the result of the action of a central executive. The classical interpretation that prevails in cognitive psychology models is based on the existence of an attentional center toward which all the attentional effects eventually converge. It leads to the proposal of a master map of locations which is not but an 'attentional center'. This interpretation is misleading. A more convenient view is that of a progressive involvement of attentional controls in the hierarchy of visual areas. This progressive mechanism seems to correlate with an increase in complexity of the filtering function of these areas (Van Essen and Gallant, 1994). This view is consistent with the neurophysiological data showing a behavioral modulation of neurons located in various visual areas. It is also in accordance with the present model which assumes that only pop-out, low-level perceptual priming and attentional cuing for conjunctive searches based on low-level features occur at a low-level (V2, V4). More complex searches can be performed at higher levels. Furthermore, they can be all based on the kind of competitive mechanism proposed in this study. The dynamic behavior of the model allows to conclude that a proportional relationship between the reaction time and the difficulty of the task (i.e. the number of distractors and the number of shared features between the target and the distractors) will be obtained each time such a competitive model will be involved in the selection of the relevant signals.

Our results do not imply the non existence of a master map of location somewhere in the visual system. They simply suggest that there is no need for this hypothesis to interpret most of the visual search experiments. It does not mean that in more complex situations -e.g. processing of complex objects in real environments- a map of the salient regions is not involved. In this sense, it is compatible with the integrated competition hypothesis put forward by Duncan *et al.* (1997). However, conventional psychophysics

experiments conducted with targets defined by a few characteristics are misleading. As long as the number of features involved in the definition of a target is small, it is indeed easy to assume that they can be integrated together on a master map of location. Actual objects can be defined by multiple conjunctions and even if the number of features involved in the definition of an object is small, the set of modalities from which it is defined can be very large. In these conditions, it is difficult to assume that every modality the visual system is able to extract from the visual stream additively projects to the saliency map to combine with the other modalities. As pointed out by Van der Heijden (1995), the selection of the features relevant for the current task, independently of the irrelevant features characterizing an object seems to be more efficient. The independance of the early representation of these features seems to facilitate this selection.

Routing and the nature of the decisional center

One important question raised by the present model is how the information corresponding to attended objects is routed toward higher areas. The present model enables to conclude that a competition mechanism could produce a temporal modulation of the signals corresponding respectively to attended and non attended objects. Due to the very simple neuron model adopted in this study, it is not possible to further examine to what extent this competition is able to produce temporal correlations of spike trains such as those invoked in Niebur and Koch (1994) model. However, attention enhanced activity and activity due to pop-out both produce sustained activities of the concerned neurons. It is thus possible to predict that whatever the modulation by which these objects are tagged, a similar modulation could distinguish pop-out and attended objects from non-salient ones.

In the tagging hypothesis (Niebur, Koch and Rosin, 1993), the discrimination between 'tagged' attended stimuli is achieved through the low pass filter properties of V4 neurons. This hypothesis has been introduced to explain the fact that the mean firing rate of V2 neurons seemed to be unaffected by attentional modulation. Since the publication of this work, Motter (1994a,b) showed that V2 neurons exhibit modulation of their firing rate under the influence of attention. This finding has motivated the present work and our model shows such mean-firing rate modulation by attentional controls. However, it is not possible to exclude the additional involvement of a synchronization or oscillations as proposed by Niebur *et al.* (1994) and the role of these synchronizations in the 'binding by attention' mechanism proposed by Fujii *et al.* (1996).

Except if one assumes a kind of intrinsic tagging mechanism which produces the 40Hz modulation at the level of V1 or V2 in the absence of top-down signal for salient targets, the signal issued from these targets will be blocked in V4 by the low pass filtering process. However, it is difficult to assume that there are distinct pathways for attended and unattended stimuli. More likely, both stimuli are projected onto higher areas through V4. The way unattended as well as attended stimuli, in spite of their different mechanism of selection, are routed through these areas remains to be explained. Unlike the '40Hz tagging hypothesis' introduced by Niebur and Koch (1994), the present work puts forward the role of a modulation issued from the signal itself. The study of the conditions for which this modulation could represent a complex

temporal code multiplexing different aspects of the visual information including attentional tagging requires further investigations.

Conclusion

Our model suggests that the attentional mechanisms supporting visual search of simple conjunctive targets are located as early as in area V2 or V4. It assumes that the search for the target is easy in general, as proposed by Wolfe *et al.* (1989). It can be made difficult in two ways: (i) when the establishment of a coupling between cells coding for the characteristics of the target is difficult, (ii) when the target shares several characteristics with the distractors, such that it becomes less distinguishable. Besides, the present model suggests that at least a part of the preattentive and attentive mechanisms is implemented in the same visual area. These conclusions are similar to the proposals discussed in Green (1991).

The present model shows that the integration of features does not necessarily precede the selection of a particular item. Thus the saliency map becomes a location map and no selection mechanism needs to be invoked at its level. The conspicuity of a conjunctive stimulus is built up through the coupling mechanisms without the need of any specialized map to perform the task. In this sense, our model agrees with Van der Heijden proposal that rejects the existence of a 'final level of perceptual coding'. According to this ideas, attention is not a mechanism to solve the modularity problem by the integration of separable features in unitary objects, but, on the contrary, modularity brings solutions to solve the attention problem (Van der Heijden, 1995). We also agree with Green (Green, 1991; Green and Odom, 1991) which propose that attention would be a mask rather than a beam acting through the inhibition of the non-relevant features instead of through the enhancement of the relevant ones. It must be emphasized that in her more recent papers, Treisman seems to promote a very similar idea of an early control of feature binding by attention (Treisman, 1996).

One important difference with previous FIT-based models is that they provide the lower layers of the visual system with both spatial and descriptive information about the target. Competitive Search assumes that top-down information deals only with the description of the potential target. The location of the target is a bottom-up mechanism. This organization, which does not require a feed back from an attentional center, ensures a faster response time than previous models and seems to be in accordance with the most recent experimental findings (Humphreys *et al.* 1994; Duncan *et al.*, 1997).

It seems clear from the present model that the serial effects do not depend on the existence of a spotlight. It is the result of a competitive mechanism induced by the mutual inhibition of the items to be selected. The present model suggests that this mechanism is due to a mutual inhibition of the neurons within each feature map. The 'serial' effect is observed only for those characteristics which requires the conjunctive activation of two cell types, each sensitive to a given feature. When the conjunctive feature is coded as the activation of a single cell type sensitive to both characteristics of the stimulus, our model predicts a 'pop-out' behavior (the cells corresponding to the target are isolated on their map). This is exactly what is observed for stereo disparity (Nakayama and Silverman, 1986). Thus, the present model supports an alternative explanation to the parallel/serial distinction between preattentive and attentive mechanisms. Preattentive filtering may be essentially a parallel static process while

attentive mechanisms may rather be based upon temporal switching toward the attended item.

The present work shows that the introduction of a coupling induces a temporal discrimination between signals coding for target and distractors. In the lack of such a coupling, there is no possible distinction between these signals. Signals corresponding to an attended item show a sustained firing mode giving rise to an attenuation of unattended items. This behavior is more consistent with the recent finding of Duncan, Ward, Shapiro (1994) who showed that attention seems to be more likely related to a sustained state during which relevant objects can be processed than to a high-speed switching mechanism searching for relevant objects. It is also consistent with the hypothesis that attention is essentially a competition mechanism biased in favor of the relevant objects (Duncan, *et al.*, 1997)

This organization accounts for the attentional modulations of V2 neurons reported by Moran and Desimone (1985), and Motter (1993). It explains why these modulations are not spatially defined and precedes the presentation of the target. It also explains how they are used as a tuning signal suitable for the adaptation of perceptual filters to internal expectancies.

One major difficulty of FIT models is to explain how the identification of an object can be achieved on the saliency map if this saliency is build from a bottom-up information, while the selection is made upon a top-down basis. One possible interpretation is that, in these models, the saliency of the objects computed on the basis of bottom-up information is assumed to be sufficient to determine the interesting object. In this case, they fail to explain how the top-down information interferes in this process. An alternative explanation is that the top-down information is used to build the saliency map, top-down and bottom-up information combining together at a low level. In this case, the saliency map is deprived of these most specific properties. It is not but a location map. As pointed out by Van der Heijden (1995), if attention is used to solve the modularity problem through a binding of the relevant as well irrelevant features composing an object, it is a very difficult to explain how a given task can determine what properties are relevant in its context. On the contrary, in early selection models like the present one, the control of these properties by the task at hand is straightforward.

As stated above and pointed out by several authors (see the "blackboard" hypothesis in Green, 1991), saliency maps act more or less as an internal screen which copies out the external information. As emphasized by O'Regan (1992), there is no need for such resource consuming hypothesis: the external world can be used as an external memory by the visual system. This is exactly what is done in the present model. However, one of the major argument against the existence of such internal screen, from which the spotlight proceeds, is drawn from behavioral considerations. If we replace vision in a behavioral framework, we have to understand how attentional mechanisms are used during behavior. In these conditions, the visual system acts in a closed loop way, and the processing of a visual scene is part of a behavior. Certainly a lot of areas are preset before one reach the situation at which the scene under consideration is processed (Laberge, 1995). The visual system does not operate on a *tabula rasa*. Instead, it is driven by the complex activity existing in higher cortical areas before processing the visual scene. Our hierarchical and isolated view of the visual system leads us to the

erroneous interpretation that the higher level cortical areas are only fed up from the visual input. This is exactly the contrary which is likely to happen in a natural situation.

In the present work we have shown that a simple but dynamical model is able to explain both the existence of a pop-out phenomenon and the response-time variation observed in the presence of distractors. The model shows that there is no need to postulate the existence of top-down controls of an hypothetical spotlight of attention to account for a time-variation in the response time in the presence of distractors. This variation could be explained by the dynamical properties of the system.

In conclusion, the pop-out and the 'sequential' mechanisms could be essentially 'data-driven'. However while the former do not usually require top-down controls, the latter use these controls to confer to the visual filters an improved ability to detect the target. After this tuning mechanism has been set up by a behavioral requirement such as the memory of a cue or any given task-specific requirement, the detection of the target occurs in parallel. If there are more than one competing items, the competition is biased in favor of the target by the presence of the top-down signal. The complexity of the competition depends solely on the coupling intensity which is more or less difficult to establish. Thus the model suggests that the filtering process is mainly goal-driven and governed by the dynamical properties of the system.

The serial scan hypothesis and its related mechanisms are typical of the problem raised by Van der Heijden about "Cognitive Psychology". Van der Heijden emphasized the distinction between *explananda* and *explanans*, i.e. the characteristics of the data that have to be explained and the theoretical entities used to explain them. The serial scan hypothesis is such an *explanans*. Unfortunately, a lot of authors have considered it as an *explanandum* too. It has resulted a lot of models that aim at explaining how this serial scan could work, while the existence of this mechanism is not even proved. The serial scan hypothesis and its *explanans*, the attentional spotlight, have to be considered as metaphors, not as explanations.

The same is true for the notion of master map of location. The existence of this map used in conjunction with a specific winner-take-all mechanism remains a hypothesis. However, its existence is taken for granted in the cognitive psychology community. Is it also an *explanans* and does not deserve to be considered as something to be explained. On the contrary, the approaches based on the bottom-up construction of models starting from the neuronal level and improved in a progressive way do not exhibit this difficulty. There is no need to invoke any metaphor as an explanation of the observed phenomena. We argue that our explanation based on the hypothesis that targets are discriminated in attentional mechanisms on the basis of a time delay due to a competitive mechanism is by no way similar to the searchlight hypothesis. The former is a computational neuroscience plausible explanation which gives rise to testable predictions while the latter is a cognitive psychology hypothesis which is by essence not refutable.

We wish to promote the idea that the behavioral observations should not give rise to untestable psychological theories. On the contrary, it seems to us that the only way to explain them is to try to understand their neurobiological foundations.

Acknowledgments

The authors thank all the members of the AnimatLab for helpful discussions during the course of this work. They are particularly indebted to P.Andrey, J.Kodjabachian and O.Trullier for their help in developing the model. This work was supported by a grant from the Direction Générale de l'Armement (DGA - 93070)).

References

- Ahissar, M. and Hochstein, S. (1996). Learning pop-out detection: Specificities to stimulus characteristics. *Vision Research*, 36, 3487-3500.
- Allport, A. (1987). Selection for action: Some behavioral and neurophysiological considerations of attention and action. In Heuer, H. and Sanders, A. F. (Eds.), *Perspectives on Perception and Action*. Hillsdale, NJ: Erlbaum.
- Allport, A. (1989). Visual attention. In Posner, M. I. (Ed.), *Foundations of cognitive science* (pp. 631-682). Cambridge, MA: The MIT Press.
- Cave, K. R. and Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology*, 22, 225-271.
- Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural computation*, 1, 123-132.
- Damasio, A. R. (1990). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. In Eimas, P. D. and Galaburda, A. M. (Eds.), *Neurobiology of Cognition* (pp. 25-62). Cambridge, MA: The MIT Press.
- Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 13494-13499.
- Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, 87, 272-300.
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology*, 113, 501-517.
- Duncan, J., Humphreys, G. and Ward, R. (1997). Competitive brain activity in visual attention. *Current Opinion in Neurobiology*, 7, 255-261.
- Duncan, J., Ward, R. and Shapiro, K. (1994). Direct measurement of attentional dwell time in human vision. *Nature*, 369, 313-315.
- Eckhorn, R. and Schanze, T. (1991). Possible neural mechanisms of feature linking in the visual system: Stimulus-locked and stimulus-induced synchronizations. In Babloyantz, A. (Ed.), *Self-organization, emerging properties and learning* (pp. 63-81). Advanced Sciences Institutes. Series B: Physics, 260. New York: Plenum Press.
- Egeth, H. E., Virzi, R. A. and Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 32-39.
- Enns, J. and Rensink, R. A. (1990). Influence of scene-based properties on visual search. *Science*, 247, 721-723.
- Finkel, L. H. and Edelman, G. M. (1989). Integration of distributed cortical systems by reentry: a computer simulation of interactive, functionally segregated visual areas. *Journal of Neuroscience*, 9, 3188-3208.
- Fujii, H., Ito, H., Aihara, K., Ichinose, N. and Tsudaka, M. (1996). Dynamical cell assembly hypothesis - Theoretical possibility of spatio-temporal coding in the cortex. *Neural Networks*, 9, 1303-1350.
- Fuster, J. M. (1990). Inferotemporal units in selective visual attention and short-term memory. *Journal of Neurophysiology*, 64, 681-697.
- Gibson, E. and Rader, N. (1979). The perceiver as performer. In Hale, G. A. and Lewis, M. (Eds.), *Attention and cognitive development*. New York, NY: Plenum.
- Gibson, J. J. (1986). An ecological approach to visual perception. Erlbaum, Hillsdale, NJ.
- Green, M. (1991). Visual search, visual streams, and visual architectures. *Perception and Psychophysics*, 50, 388-403.
- He, Z. J. and Nakayama, K. (1992). Surfaces vs features in visual search. *Nature*, 359, 231-233.
- Humphreys, G. W., Romani, C., Olson, A., Riddoch, M. J. and Duncan, J. (1994). Non-spatial extinction following lesions of the parietal lobe in humans. *Nature*, 372, 357-9.
- James, W. (1890). Psychology (Briefer Course). In Anderson, J. A. and Rosenfeld, E. (Eds.), *Neurocomputing: foundations of research* (pp. 1-4). Cambridge, MA: The MIT Press.

- Julesz, B. (1990). Early Vision Is Bottom-up, Except for Focal Attention. *Brain*, Proceedings of the Cold Spring Harbor Symposia on Quantitative Biology Cold Spring Harbor Laboratory Press, Cold Spring Harbor. 973-978.
- Knierim, J. J. and Van Essen, D. C. (1992). Neuronal Responses to Static Texture Patterns in Area V1 of the Alert Macaque Monkey. *Journal of Neurophysiology*, 67, 961-980.
- Koch, C. and Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, 4, 219.
- Laberge, D. (1995). Computational and anatomical models of selective attention in object identification. In Gazzaniga, M. S. (Ed.), *The Cognitive Neurosciences* (pp. 649-664). Bradford Books. Cambridge, MA: The MIT Press.
- Le Bihan, D., Turner, R., Zeffiro, T., Cuénod, C., Jezzard, P. and Bonnerot, V. (1993). Activation of human primary visual cortex during visual recall: a magnetic resonance imaging study. *Proceedings of the National Academy of Sciences of the United States*, 90, 11802-11805.
- Logothetis, N. K. (1991). Is movement perception color blind? *Current Biology*, 1, 298-300.
- Luck, S. J., Chelazzi, L., Hillyard, S. A. and Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J Neurophysiol*, 77, 24-42.
- Marr, D. (1982). Vision: A computational investigation into the human representations and processing of visual information. Freeman, San Francisco, CA.
- Maunsell, J. H. R. and Ferrera, V. P. (1995). Attentional mechanisms in visual cortex. In Gazzaniga, M. S. (Ed.), *The Cognitive Neurosciences* (pp. 451-461). Bradford Books. Cambridge, MA: The MIT Press.
- Merigan, W. H., Nealey, T. A. and Maunsell, J. H. R. (1993). Visual effects of lesions of cortical area V2 in macaques. *The Journal of Neuroscience*, 13, 3180-3191.
- Milanese, R. (1993). Detecting salient regions in an image: from biological evidence to computer implementation. University of Geneva, Switzerland.
- Moran, J. and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, 229, 782-784.
- Motter, B. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2 and V4 in the presence of competing stimuli. *Journal of Neurophysiology*, 70, 909-919.
- Motter, B. C. (1994). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *The Journal of Neuroscience*, 14, 2178-2189.
- Motter, B. C. (1994). Neural correlates of feature selective memory and pop-out in extrastriate area V4. *The Journal of Neuroscience*, 14, 2190-2199.
- Mumford, D. (1991). On the computational architecture of the neocortex. I. The role of the thalamo-cortical loop. *Biological Cybernetics*, 66, 135-145.
- Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loop. *Biological Cybernetics*, 66, 241-251.
- Nakayama, K. and Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, 320, 264-265.
- Niebur, E. and Koch, C. (1994). A model for the neuronal implementation of selective visual attention based on temporal correlation among neurons. *Journal of Computational Neurosciences*, 1, 141-158.
- Niebur, E., Koch, C. and Rosin, C. (1993). An oscillation-based model for the neuronal basis of attention. *Vision Research*, 33, 2789-2802.
- Nothdurft, H. C. and Li, C. Y. (1985). Texture discrimination: representation and luminance differences in cells of the cat striate cortex. *Vision Research*, 25, 99-113.
- O'Regan, J. K. (1992). Solving the "Real" Mysteries of Visual Perception: The World as an Outside Memory. *Canadian Journal of Psychology*, 46, 461-488.
- Olshausen, B. A., Anderson, C. H. and Van Essen, D. C. (1995). A multiscale dynamic routing circuit for forming size- and position-invariant object representation. *Journal of Computational Neurosciences*.
- Palmer, J., Ames, C. T. and Lindsley, D. T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 108-130.
- Posner, M. I., Snyder, C. R. R. and Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology*.
- Quinlan, P. T. and Humphreys, G. W. (1987). Visual search for targets defined by combinations of color, shape, and size: An examination of the task constraints on features and conjunction searches. *Perception and Psychophysics*, 41, 455-472.
- Ramachandran, V. S. (1988). Perceiving shape from shading. *Scientific American*, 259, 76-83.
- Rockland, K. S. (1994). The organization of feedback connections from area V2 (18) to V1 (17). In Plenum Press, N. (Ed.), *Cerebral Cortex* (pp. 261-299), 10. .
- Rockland, K. S., Saleem, K. S. and Tanaka, K. (1994). Divergent feedback connections from areas V4 and TEO in the macaque. *Visual Neuroscience*, 11, 579-600.

- Rolls, E. T. (1990). The Representation of Information in the Temporal Lobe Visual Cortical Areas of Macaques. In Eckmiller, R. (Ed.), *Advanced Neural Computers* (pp. 69-78). Amsterdam: North Holland.
- Sakai, K. and Miyashita, Y. (1994). Visual imagery: an interaction between memory retrieval and focal attention. *Trends in Neuroscience*, *17*, 287-289.
- Shulman, G. L., Corbetta, M., Buckner, R. L., Raichle, M. E., Fiez, J. A., Miezin, F. M. and Petersen, S. E. (1997). Top-down modulation of early sensory cortex. *Cerebral Cortex*, *7*, 193-206.
- Singer, W. and Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neurosciences*, *18*, 555-586.
- Sloman, A. (1989). On designing a visual system. *Journal of Experimental and Theoretical Artificial Intelligence*, *1*, 289-337.
- Tamura, H., Sato, H., Katsuyama, N., Hata, Y. and Tsumoto, T. (1996). Less segregated processing of visual information in V2 than in V1 of the monkey visual cortex. *European Journal of Neuroscience*, *8*, 300-309.
- Tononi, G., Sporns, O. and Edelman, G. M. (1992). Reentry and the problem of integrating multiple cortical areas: simulation of dynamic integration in the visual system. *Cerebral Cortex*, *2*, 310-335.
- Treisman, A. (1986). Properties, parts, and objects. In Boff, K. R., Kaufman, L. and Thomas, J. P. (Eds.), *Handbook of perception and human performance*, II. New York: Wiley.
- Treisman, A. (1988). Features and Objects: The Fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology*, *40A*, 201-237.
- Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*, 171-178.
- Treisman, A. and Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, *12*, 97-136.
- Treisman, A. and Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, *95*, 15-48.
- Treisman, A. and Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 459-478.
- Tsotsos, J. K. (1995). Toward a computational model of visual attention. In Pappathomas, T. V., Chubb, C., Gorea, A. and Kowler, E. (Eds.), *Early vision and beyond* (pp. 207-218). A Bradford Book. Cambridge, MA: The MIT Press.
- Ullman, S. (1991) Sequence-seeking and counter streams: A model for information processing in the cortex. Technical Report, MIT - Cambridge, MA. AI Memo 1311.
- Usher, M. and Niebur, E. (1996). Modeling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention. *Journal of Cognitive Neuroscience*, *8*, 311-327.
- Van der Heijden, A. H. C. (1995). Modularity and attention. *Visual Cognition*, *2*, 269-301.
- Van Essen, D. C. and Gallant, J. L. (1994). Neural mechanisms of form and motion processing in the primate visual system. *Neuron*, *13*, 1-10.
- Vecera, S. and Farah, M. (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology: General*, *123*, 146-160.
- Wolfe, J., Yee, A. and Friedmann-Hill, S. R. (1992). Curvature is a basic feature for visual search task. *Perception (England)*, *21*, 465-480.
- Wolfe, J. M. (1992). "Effortless" texture segmentation and "parallel" visual search are not the same thing. *Vision Research*, *32*, 757-763.
- Wolfe, J. M., Cave, K. R. and Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology*, *15*, 419-433.