A distributed computational model of spatial memory anticipation during a visual search task

Jérémy Fix, Julien Vitay, Nicolas P. Rougier

A distributed computational model of spatial memory anticipation during a visual search task

Jérémy Fix and Julien Vitay and Nicolas P. Rougier

Loria, Campus Scientifique, BP239
54506 Vandoeuvre-les-Nancy, France

Abstract. Some visual search tasks require the memorization of the location of stimuli that have been previously focused. Considerations about the eye movements raise the question of how we are able to maintain a coherent memory, despite the frequent drastic changes in the perception. In this article, we present a computational model that is able to anticipate the consequences of eye movements on visual perception in order to update a spatial working memory.

1 Introduction

In the most general framework of behavior, the notion of anticipation is intimately linked with the possibility to predict the consequences and the outcomes of a given action. If we consider that any action is goal-motivated, then an action is carried out in the first place because it is anticipated that this action will lead to a situation in which the goal can be reached more directly. In this framework, anticipation can be viewed as a prediction of the future and is tightly linked to the notion of goal-directed behavior. However, there also exists more structural reasons why anticipation is necessary.

For example, when dealing with both accurate and very fast movements like catching a ball or scanning a visual scene, brain representations should be updated very quickly (even in advance in some cases) in accordance with the task that is carried out. The problem in this context is that the time scale required for carrying out such tasks may be dramatically smaller than the time scale of a single neuron. Moreover, those neurons are also in interaction with other neurons in the network and the resulting dynamic may be even slower. One solution to cope with this problem is to use a forward predictive model that is able to anticipate the consequences and outcomes of a motor action. The resulting dynamic at the level of the model is then faster than the dynamic of its components.

Let us consider the ability to anticipate changes in the visual information resulting from an eye saccade. This anticipation is known to be largely based on unconscious mechanisms that provide us with a feeling of stability while the whole retina is submerged by different information at each saccade; producing a saccade results in a complete change in the visual perception of the outer world.

If a system is unable to anticipate its own saccadic movements, it cannot pretend to obtain a coherent view of the world, because each image would be totally uncorrelated from the others. One stimulus being at one retinal location before a saccade could not be easily identified as being the same stimulus at another retinal location after the saccade. Consequently, the saccadic eye movements should be anticipated in order to keep the coherence of the scene and to be able to track down interesting targets. A number of works have already addressed the specific problem of visual search of a target among a set of distractors. However, most of the resulting models do not deal with the problem of saccadic eye movements that produce drastic changes in the available visual information.

Using neural fields introduced by Amari [1] for the one dimensional case and later extended to higher dimensions by Taylor [34], we would like to address in this paper the specific problem of anticipation during visual search using a purely distributed and numerical neural substrate. After briefly reviewing literature related to visual search in the first section, we introduce a very simple visual experiment that helps to illustrate the underlying mechanisms of the model that is detailed in that same section.

2 Visual search

Visual search is a cognitive task that most generally involves an active scan of a visual scene to find one or several given targets among distractors. It is deeply anchored in most animal behaviors, from a predator looking for a prey in the environment, to the prey looking for a safe place to avoid being seen by the predator. Psychological experiments may be less ecological and may propose, for example, to find a given letter among an array of other letters, measuring the efficiency of the visual search in terms of reaction time (the average time to find the target given the experimental paradigm). In the early eighties, [35] suggested that the brain actually extracts some basic features from the visual field in order to perform the search. Among these basic features, which have been recently reviewed by [40], one can find features such as color, shape, motion, or curvature. Finding a target is then equivalent to finding the conjunction of features, which may be unique, that best describes the target. In this sense, [35] distinguished two main paradigms (a more tempered point of view can be found in [6]).

Feature search refers to a search where the target differs from all distractors by exactly one feature.
Conjunction search refers to a search where the target differs from distractors by at least one of two or more features.

What characterizes feature search best is a constant search time that does not depend on the number of distractors. The target is sufficiently different from the distractors to pop out. However, in the case of conjunction search, the mean time needed to find the target is roughly proportional to the number of distractors

that share at least one feature with the target (cf. Figure 1). These observations lead to the question of how a visual stimulus could be represented in the brain. The explanation given by Treisman and Gelade [35], the so-called *Feature-Integration Theory*, proposes that elementary features are processed in separated feature maps. Competition inside one map would lead to feature search, based on the idea that the item differing the most from its background would win the competition and be represented. For targets differing from distractors by more than two features, there cannot be any global competition. This would mean that finding the target requires successively scanning every potential candidate until the correct target is found. This explains the dependence of the search time on the number of similar distractors in conjunction search tasks.

The main prediction of this theory is that processing visual inputs is not a global feed-forward processing, but more an iterative and sequential process on sensory representations. We describe below the strategies used by the brain to achieve this sequential search, by putting emphasis on saccadic eye movements and visual attention. The scope of this article is therefore to model the cognitive structures involved in the sequential processing of visual objects, and not the visual processing of the features alone.

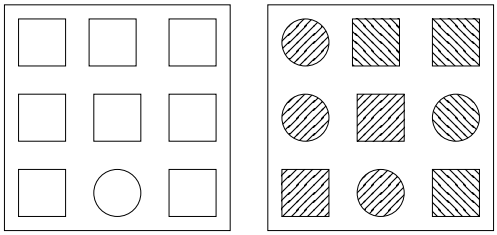

Fig. 1. Feature search can be performed very quickly as illustrated on the left part of the figure; the disc shape pops out from the scene. However, as illustrated in the right figure, if the stimuli share at least two features, the pop out effect is suppressed. Hence, finding the disc shape with the stripes going from up-left to down-right requires an active scan of the visual scene.

2.1 Saccadic eye movements

The eye movements may have different behavioral goals, leading to five different categories of movements: saccades, vestibulo-ocular reflex, optokinetic reflex, smooth-pursuit and vergence. However, in this article we will only focus on saccades (for a detailed study of eye movements, see [17], [3]).

Saccades are fast and frequent eye movements that move the eye from the current point of gaze to a new location in order to center a visual stimulus on the fovea, a small area on the retina where the resolution is at its highest. The

velocity of the eyes depends on the amplitude of the movement and can reach up to 700 degrees per second at a frequency of 3 Hz. The question we would like to address is how the brain may give the illusion of a stable visual space while the visual perception is drastically modified every 300 ms.

While the debate to decide whether or not the brain is blind during a saccade has not been settled (see [18, 2, 14, 29] for the notion of saccadic suppression and [24] for a discussion about the necessity of a saccadic suppression mechanism), the coherence between the perception before and after a saccade cannot be established accurately solely based on perception. One solution to consider is that the brain may use an efferent copy of the voluntary eye movement to remap the representation it has built of the visual world. Several studies shed light on pre-saccadic activities in areas such as V4 and LIP where the locations of relevant stimuli are supposed to be represented. In [22], the authors suggest that "the pre-saccadic enhancement exhibited by V4 neurons [...] provides a mechanism by which a clear perception of the saccade goal can be maintained during the execution of the saccade, perhaps for the purpose of establishing continuity across eye movements." In [20], the authors review evidence that LIP neurons, whose receptive field will land on a previously stimulated screen location after a saccade, are excited even if the stimulus disappears during the saccade. In a recent study, Sommer and Wurtz [33] showed neurons in FEF that receive projections from the superior colliculus that could explain the origin of a corollary discharge signal responsible for the pre-saccadic activity exhibited by these neurons.

2.2 Visual attention

Focusing on a given stimulus of the visual scene is a particular aspect of the more general concept of attention that has been defined as the capacity to concentrate cognitive resources on a restricted subset of sensory information ([12]). In this context of visual attention, only a small subset of the retinal information is available at any given time to elaborate motor plans or cognitive reasoning (cf. *change blindness* experiments presented in [24], [32]). A visual scene is not processed as a whole but rather processed by successively focusing on interesting parts of it, possibly involving eye movements, but this is not necessary. The selection of a target for an eye movement is then closely related to the notion of spatial attention ([21]) that is classically divided into two types: **overt attention**, which involves a saccade to center a stimulus on the fovea, and **covert attention**, in which no eye movement is triggered. These two types of spatial attention were first supposed to be independent ([26]) but recent studies such as the premotor theory of attention proposed in [28] (see also [4], [16], [5]) consider that covert and overt attention rely on the same neural structures but the movement is inhibited in covert attention. A more general discussion about the covert and overt stages of action can be found in [13].

The deployment of attention on a specific part of the visual information can be the consequence of two phenomena. Firstly it can rely on the saliency of a

stimulus, compared to its surrounding (for example a sudden strong flash light); this is known as bottom-up attention. Secondly, it can also depend on the task in which the subject is involved, which may need to enhance some parts of the perception (for example, imagine that you have to find an orange among apples and bananas, the color information could be a good criteria to find the target rapidly).

In [23], the authors shed light on the neural correlates of attention on the response of neurons in the visual and temporal cortices. If we consider a specific neuron tuned to a given orientation in its receptive field, one can distinguish several cases:

- the response of the neuron is high when an oriented bar with the preferred orientation (called good stimulus) is presented in its receptive field
- the response of the neuron is low when an oriented bar with an orientation different from the preferred one (called bad stimulus) is presented in its receptive field
- the response is between the two preceding ones when both a good and bad stimulus are presented

When a monkey is involved in a task that requires to select one of the two stimuli, for example the good one, the response of the neuron is enhanced. The study of this suppressive interaction phenomena was extended by further authors ([19], [27], [36]).

As we will see in section 3.2, we do not deal with how the salience of the visual stimuli is computed, whether or not it is a bottom-up or top-down processing. The main points are that for each location in the visual space, we are able to compute its behavioral relevance, and that considering eye movements necessarily implies dealing with overt attention.

2.3 Computational models

Over the past few years, several attempts at modeling visual attention have been engaged ([15], [37], [41], [11], [10]). The basic idea behind most of these models is to find a way to select interesting locations in the visual space given their behavioral relevance and whether or not they have already been focused. The two central notions in this context have been proposed by [15] and [25]:

- saliency map
- inhibition of return (IOR).

The saliency map is a single spatial map, in retinotopic coordinates, where all the available visual information converge in order to obtain a unified representation of stimuli, according to their behavioral relevances. A winner-take-all algorithm can be easily used to find which stimulus is the most salient within the visual scene, and thus identify its location as the locus of attention. However, in order

to be able to go to the next stimuli, it is important to bias the winner-take-all algorithm in such a way that it prevents going backward to an already focused stimulus. The goal of the inhibition of return mechanism is precisely to feed the saliency map with such a bias. The idea is to have another neural map that records focused stimuli and inhibits the corresponding locations in the saliency map. Since an already focused stimulus is actively inhibited by this map, it cannot pretend to win the winner-take-all competition, even if it is the most salient.

The existence of a single saliency map is still not proved. In [10] the author proposes a more distributed representation of these relevances, making a clear anatomical distinction between the processing of the visual attributes of an object and its spatial position (according to the What and Where pathways hypothesized by [38], see also [9]). In this model, spatial competition occurs in a motor map instead of a perceptive one. It exhibits good performances regarding visual search task in natural scene, but is restricted to covert attention. In most of the previously proposed models, the authors do not take into account eye movements and the visual scene is supposed to remain stable: scanning is done without any saccade. During the rest of this article, we will keep the saliency map hypothesis, even if controversial, in order to illustrate the anticipatory mechanism.

3 A model of visual search with overt attention

The goal of our model is to show the basic mechanisms necessary to achieve sequential search in a visual scene using both overt and covert attention. Using a saliency map, we need to compute the location of the most interesting stimulus that will be processed to achieve recognition. This focus of attention on a stimulus has to be displaced in two situations. First, in covert attention this focus has to be dynamically inhibited to represent another stimulus. There is therefore a need for an inhibition-of-return mechanism than can inhibit the current focus of attention. Moreover, we have to memorize the locations of previously attended stimuli, by the means of a dynamic spatial working memory.

The second situation to consider is when eye movements can center the stimulus that is being attended to. The spatial working memory has to be updated by the eye movement so that its state corresponds to the post-saccadic locations of memorized stimuli. This is where an anticipatory mechanism is mandatory.

To describe these mechanisms, we first present an experimental setup for which previous computational models would fail to achieve efficient sequential search. We then present the architecture of our model and report simulated results.

3.1 Experiment

In order to accurately evaluate the model, we setup a simple experimental framework in which some identical stimuli are drawn on a blackboard and are observed by a camera. The task is to successively focus (i.e. center) each one of the stimuli without focusing twice on any of them. We estimate the performance of the model in terms of how many times a stimulus has been focused. Hence, the point is not to analyze the strategy of deciding which stimulus has to be focused next (see [7, 8] for details on this matter). In the context of the proposed model, the strategy is simply to go from the most salient stimulus to the least salient one, and to randomly pick one stimulus if the remaining ones are equally salient.

Figure 2 illustrates an experiment composed of four identical stimuli where the visual scan path has been materialized. The effect of making a saccade from one stimulus to another is shown and underlines the difficulty (for a computational model) of identifying a stimulus before and after a saccade. Each one of the stimuli being identical to the others, it is impossible to perform an identification based solely on features. The only criteria that can be used is the spatial location of the stimuli.

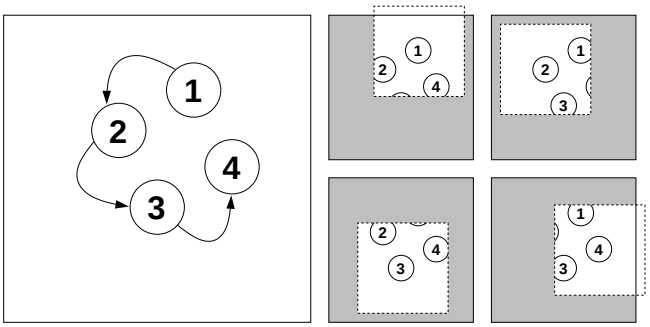

Fig. 2. When scanning a visual scene, going for example from stimulus 1 to stimulus 4, as illustrated in the left figure, the image received on the retina is radically changed when each stimulus is centered on the retina, as illustrated in the right figures. The difficulty in this situation is to be able to remember which stimuli have already been centered in order to center another one. The figures on the stimuli are shown only for explanation purpose and do not appear on the screen; all the stimuli are identical.

3.2 Model

The model is based on three distinct mechanisms (cf. Figure 3 for a schematic view of the model). The first one is a competition mechanism that involves potential targets represented in a saliency map that were previously computed according to visual input. Second, to be able to focus only once on each stimulus,

the locations of the scanned targets are stored in a memory map using retinotopic coordinates. Finally, since we are considering overt attention, the model is required to produce a camera movement, centering the target on the fovea, used to update the working memory. This third mechanism works in conjunction with two inputs: current memory and parameters of the next saccade. This allows the model to compute quite accurately a prediction of the future state of the visual space, restricted to the targets that have already been memorized.

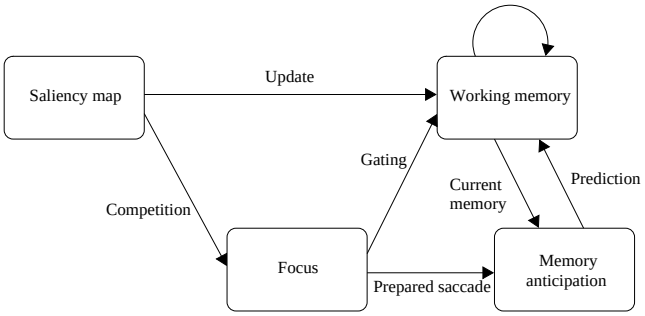

Fig. 3. Schematic view of the architecture of the model. The image captured by the camera is filtered and represented in the saliency map. This information feeds two pathways: one to the memory and one to the focus map. A competition in the focus map leads to the most salient location that is the target for the next saccade. The anticipation circuit predicts the future state of the memory with its current content and the programmed saccade.

The model is based on the computational paradigm of two dimensional discrete neural fields (the mathematical basis of this paradigm can be found in [1] for the one dimensional case, extended to a two dimensional study in [34]). The model consists of five $n \times n$ maps of units, characterized by their position, denoted $\mathbf{x} \in [1..n]^2$ and their activity as a function of their position and time, denoted u($\mathbf{x}$,t). The basic dynamical equation that follows the activity of a unit at position $\mathbf{x}$, depends on its input $I(\mathbf{x}, t)$. Equation (1) is the equation proposed in [1], discretized in space.

$$\tau . \frac{\partial u(\mathbf{x}, t)}{\partial t} = -u(\mathbf{x}, t) + baseline + \frac{1}{\alpha} I(\mathbf{x}, t) \tag{1}$$

We distinguish two kinds of units. The first are sigma units that compute their input as a weighted sum of the activity of afferent neurons, where afferent neurons are defined as neurons in other maps. We also consider lateral connections that involve units in the same map. If we denote w_{aff} the weighting function for the afferent connections and w_{lat} the weighting function for the lateral connections, the input $I(\mathbf{x}, t)$ of a unit $\mathbf{x}$ at time t can be written:

$$I(\mathbf{x},t) = \sum_{aff} w_{aff} u_{aff}(t) + \sum_{lat} w_{lat} u_{lat}(t), \tag{2}$$

where equations 3 and 4 define the lateral and afferent weighting functions as a Gaussian and difference of Gaussians, respectively.

$$w_{aff}(\mathbf{x},\mathbf{y}) = A.e^{\frac{\|\mathbf{x}-\mathbf{y}\|^2}{a^2}} \text{ with } A, a \in \mathbb{R}^{*+}, \mathbf{x}, \mathbf{y} \in [1..n]^2 \tag{3}$$

$$w_{lat}(\mathbf{x},\mathbf{y}) = B.e^{\frac{\|\mathbf{x}-\mathbf{y}\|^2}{b^2}} - C.e^{\frac{\|\mathbf{x}-\mathbf{y}\|^2}{c^2}} \text{ with } B, C, b, c \in \mathbb{R}^{*+}, \mathbf{x}, \mathbf{y} \in [1..n]^2 \tag{4}$$

The second kind of units we consider are sigma-pi units ([31]), which compute their input as a sum of the product of the activity of afferent neurons. We also consider the lateral connection term so that the input of a unit $\mathbf{x}$ at time t can be written:

$$I(\mathbf{x},t) = \sum_{i \in I} w_{aff_i} \prod_{j \in E_i} u_{aff_j}(t) + \sum_{lat} w_{lat} u_{lat}(t). \tag{5}$$

All the parameters of the previous equations used in the simulation are summarized in the appendix.
We now describe how the different maps interact. Since the scope of this article is the anticipation mechanism, the description of the saliency map, the focus map and the working memory will not be accurate; a more detailed explanation, with the appropriate dynamical equations, can be found in [39].

Saliency map The saliency map, also referred to as INPUT in the following, is computed by convolving the image captured with the camera of a robot used for the simulation with Gaussian filters. The stimuli we use are easily discriminable from the background on the basis of the color information. This computation leads to a representation of the visual stimuli with Gaussian patterns of activity in a single saliency map. We do not deal with how this saliency map is computed, whether or not it is due to bottom-up or top-down attention. We only consider that we are able to compute a spatial map, in retinotopic coordinates, that represents the behavioral relevance of each location in the visual space. We point out again that this is one of our working hypothesis, detailed in section 2.3.

Focus The units in the FOCUS map have direct excitatory feedforward inputs from the saliency map. The lateral connections are locally excitatory and widely inhibitory so that a competition between the units within the map leads to the emergence of only one stimulus in the focus map. This mechanism is not just a dynamical *winner-take-all* algorithm because the winning stimulus will still be represented in this map, even if the other stimuli in the visual scene become comparatively more salient through time, but it has to be explicitly inhibited. This focused stimulus is considered the next target to focus on and the movement to perform to center it on the fovea is decoded from this map. This map then codes the parameters of the next saccade to make.

Working memory Once a stimulus has appeared within the focus map and because it is also present in the saliency map at the same location, it emerges within the working memory. Both the excitations from the focus map and the saliency map (at a same location) are necessary for the emergence of a stimulus in the working memory area. If the focused stimulus changes, it will not be present anymore in the focus map such that an additional mechanism is needed to maintain it in the memory. It is not shown on the schematic illustration (3) but the memory consists of two maps, WM and THAL_WM, that share excitatory connections in two ways: the first map excites the second and the second excites the first, weighted so that the excitation is limited in space.

Memory anticipation The memory anticipation mechanism aims at predicting what should be the state of the working memory after an eye movement centers another stimulus in the focus map before the movement is triggered. This map, filled with sigma-pi units, has two inputs: units of the focus map and units of the working memory. If we denote $wm(\mathbf{x}, t)$ the activity of unit $\mathbf{x}$ of the working memory at time t, and $f(\mathbf{x}, t)$ the activity of unit $\mathbf{x}$ of the focus map at time t, we define the input $I(\mathbf{x}, t)$ of unit $\mathbf{x}$ in the anticipation map as:

$$I(\mathbf{x}, t) = w_{sigma-pi} \sum_{\mathbf{y} \in \mathbb{R}^2} wm(\mathbf{y}, t) f(\mathbf{y} - \mathbf{x}, t) + \sum_{aff} w_{aff} u_{aff}(t) \qquad (6)$$

The input of each unit in the anticipation map is computed as a convolution product of the working memory and the focus map, centered on its coordinates. To make (6) clearer, the condition of the sum is weaker than the one that should be used: since the input maps are discrete sets of units, the two vectors $\mathbf{y}$ and $\mathbf{y}$-$\mathbf{x}$ mustn't exceed the size of the maps. The equation (6) should also take into account that the position *eye centered* is represented by a bell-shaped pattern of activity centered in the focus map, so that an offset should be included in the first sum when determining which unit of the focus map multiplies $wm(\mathbf{y}, t)$
From (1) and (6), the activity of the units in the anticipation map, without lateral connections, satisfies (7).

$$\tau . \frac{\partial u(\mathbf{x}, t)}{\partial t} = -u(\mathbf{x}, t) + baseline + w_{sigma-pi} \sum_{\mathbf{y} \in \mathbb{R}^2} wm(\mathbf{y}, t) f(\mathbf{y} - \mathbf{x}, t) \qquad (7)$$

Then, the shape of activity in the anticipation map converges to the convolution product of the working memory and the focus map. Since the activity in the focus map has a Gaussian shape and the working memory can be written as a sum of Gaussian functions, the convolution product of the working memory and the focus map leads to an activity profile that is the profile in the working memory translated by the vector represented in the focus map. This profile is the prediction of the future state of the working memory and is then used to slightly excite the working memory. After the eye movement, and when the saliency map is updated, the previously scanned stimuli emerge in the working

memory as a result of the conjunction of the visual stimuli in the saliency map and the prediction of the working memory, that is, the prediction is combined with the new perception. This is exactly the same mechanism as the one used when a stimulus emerges in the working memory owing to the conjunction of the activity in the saliency map and the focus map.

3.3 Simulation and results

The visual environment consists of three distributed but identical stimuli that the robot is expected to scan successively exactly once. A stimulus is easily discriminable from the background, namely a green lime on a white table. A complete activation sequence of the different maps is illustrated on Figure 4. The saliency map is filled by convolving the image captured from the camera by a green filter in HSV coordinates such that it leads to three distinct stimuli[1].

At the beginning of the simulation (Figure 4a), only one of the three stimuli emerges in the focus map, thanks to the strong lateral competition that occurs within this map. This stimulus, which present in both the focus map and the saliency map, emerges in the working memory. The activation within the anticipation map reflects what should be the state of the saliency map, restricted to the stimuli that are in the working memory after the movement that brings the next targeted stimulus into the center of the visual field. During the eye movement (Figure 4b), no visual information is available and the parameter τ in (1) and (7) is adjusted so that only the units in the anticipation map remain active, whereas the activity of the others approach zero. After the eye movement and as soon as the saliency map is fed with the new visual input, the working memory is updated thanks to the excitation from both saliency and anticipation map at a same location: the prediction of the state of the visual memory is compared with the current visual information. A new target can now be elicited in the focus map thanks to a switch mechanism similar to that described in [39], but not detailed here. This mechanism acts like the inhibition of return presented in section 2.3; the memorized locations in the working memory are inhibited in the focus map, therefore biasing the competition in it, so that only a stimulus that was not already focused can be the next target to focus.

In order to illustrate more explicitly the role of the anticipatory signal, we now consider a second experiment. In this experiment, the visual scene consists of only two identical stimuli (Figure 5).

The task is the same as the previous one, namely, the robot must scan each stimulus only once, but the experimental conditions are slightly different: we enforce the robot to scan these targets in a predefined order. To bias the spatial attention toward one of the two targets, we first increase the intensity of the leftmost target. Then, when the saccade to center that target is performed, we

[1] A video of the model is available at http://www.loria.fr/ fix/publications.php

refresh the display and increase the intensity of the rightmost target. In that way, the scenario is as follows:

1. Select the leftmost target.
2. Focus on that target.
3. After the saccade, when the display is refreshed, select the rightmost target.
4. Perform the saccade to center the rightmost target.

The visual bias we add makes us able to get the same experimental conditions over the trials. During a trial, we record the activity of the neurons whose receptive field covers one of the five positions, denoted x0, x1, x2, x3, and x4 in the figure, in the four maps: visual, focus, wm and anticipation. In a typical trial, we will have a target at x1 and x3, then, after the first saccade, the targets will be at x2 and x4, to finally occupy, after the last saccade, the positions x0 and x2 (Figure 6, top).

Moreover, two conditions are considered; in the first one (Figure 6), the anticipation is enabled, whereas in the second one (Figure 7), the anticipatory signal is disabled. At the beginning of the trial, the targets are at positions x1 and x3 so that the neurons in the visual map at these positions are excited (dashed line) whereas the neurons at the other positions remain silent. The two positions x1 and x3 compete for the spatial attention. Since we added a bias toward the target at position x1, the spatial attention is on target x1, rather than on target x3, so that the activity of the neuron at position x1 in the focus map (solid line) grows, whereas the activity of the neuron at position x3 in the same map decreases to zero. The attention on target x1 enables it to emerge in the working memory (dash-dot line). The task is now to produce an eye movement that will center that target. The anticipatory mechanism predicts that when that target is centered, it will occupy the position x2; the activity of the neuron at position x2 grows (dotted line). As soon as the saccade is performed, we refresh the display. The two targets now occupy the positions x2 and x4. The bias toward the rightmost target enforces that target to be attended. The activity of the neurons at position x4 in the focus map and the working memory grows. Whereas the target at position x4 emerges in the working by the conjunction of an activity in the visual input and the focus map, the target at position x2 emerges thanks to the visual input and the anticipatory signal. As we can see in Figure 7, in which the anticipatory signal was disabled, the position of the first attended target cannot be updated at position x2. Finally, a saccade to center the target at position x4 is performed. In the case that the anticipation is present, the new positions of the two targets are in the working memory at x0 and x2, whereas when there is no anticipation, only the last attended target is in the working memory.

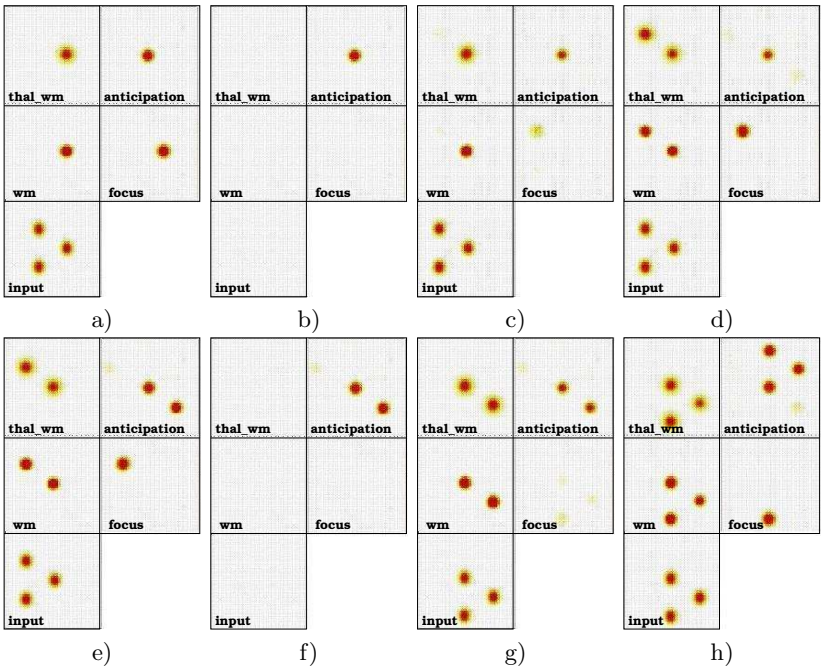

Fig. 4. A sequence of evolution of the model during an overt visual scan trial. a) One of the three stimuli emerges in the focus map and the anticipation's units predict the future state of the visual memory (the maps wm and thal_wm). b) During the execution of the saccade, only the units in the anticipation map remain active. c) The focused stimulus emerge in the memory since it is both in the saliency map and the anticipation map at the same location. d) A new target to focus is elicited. e) The future state of the memory is anticipated. f) The saccade is executed and only the prediction remains. g) The two already focused stimuli emerge in the memory. h) The attentional focus lands on the last target.

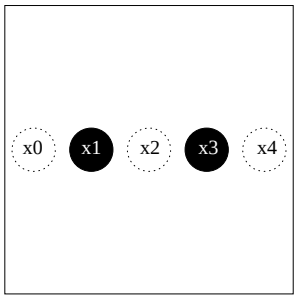

Fig. 5. The scene consists in two identical stimuli, the black blobs, initially symmetrically positioned around the center of gaze. The task is to successively focus on each target. During a trial, we measure the activity of neurons whose receptive field covers the five positions represented by the dashed circles and denoted as x0, x1, x2, x3 and x4, in several maps.

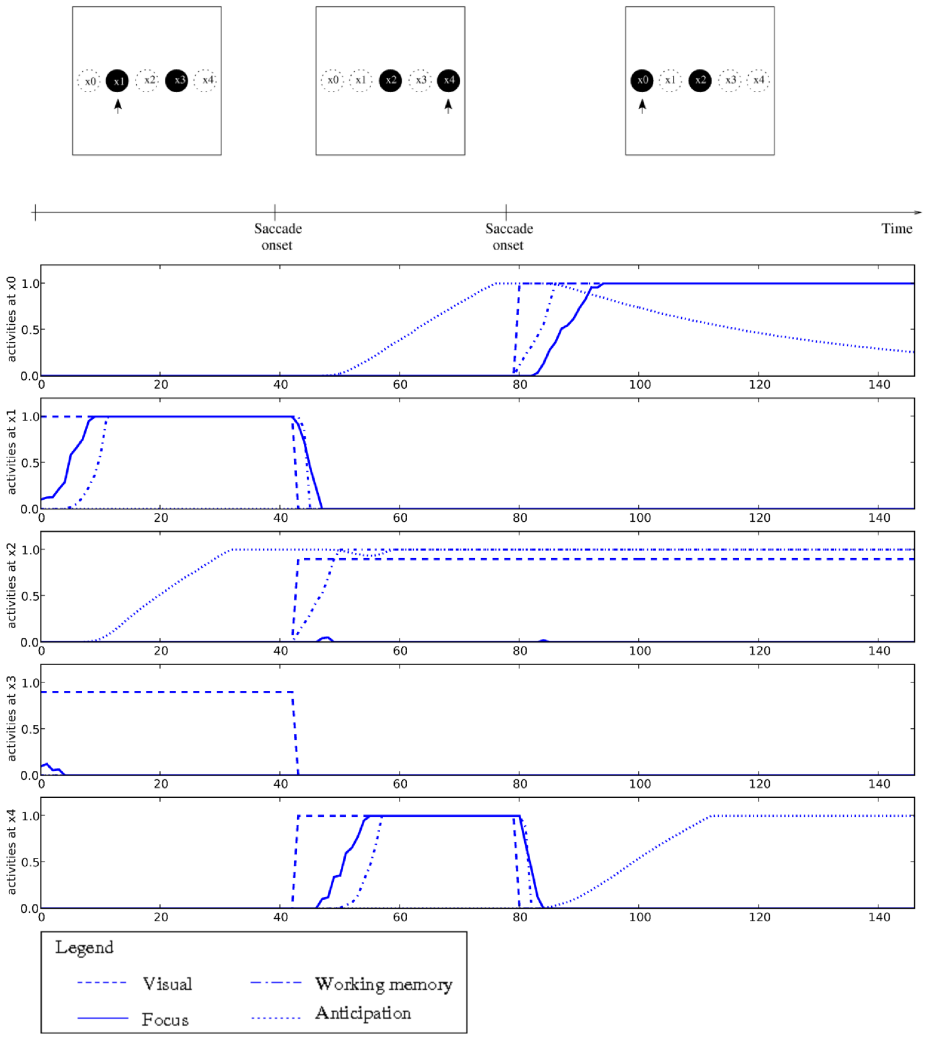

Fig. 6. Case with the anticipatory signal enabled. We record the activity of neurons whose receptive field covers one of the five positions x0, x1, x2, x3 and x4, in the four maps: visual, focus, wm and anticipation. During the trial, we add a bias toward one of the targets so that the attention directs to the biased target (that target is shown by the arrow). Each subplot represents the activity of the neurons in each map at a given position. The dashed line represents the activity of the neuron in the visual map, the solid line the activity of the neuron in the focus map, the dash-dot line the activity of the neuron in the working memory and the dotted line the activity of the neuron in the anticipation map. Please read the text for explanations on these curves.

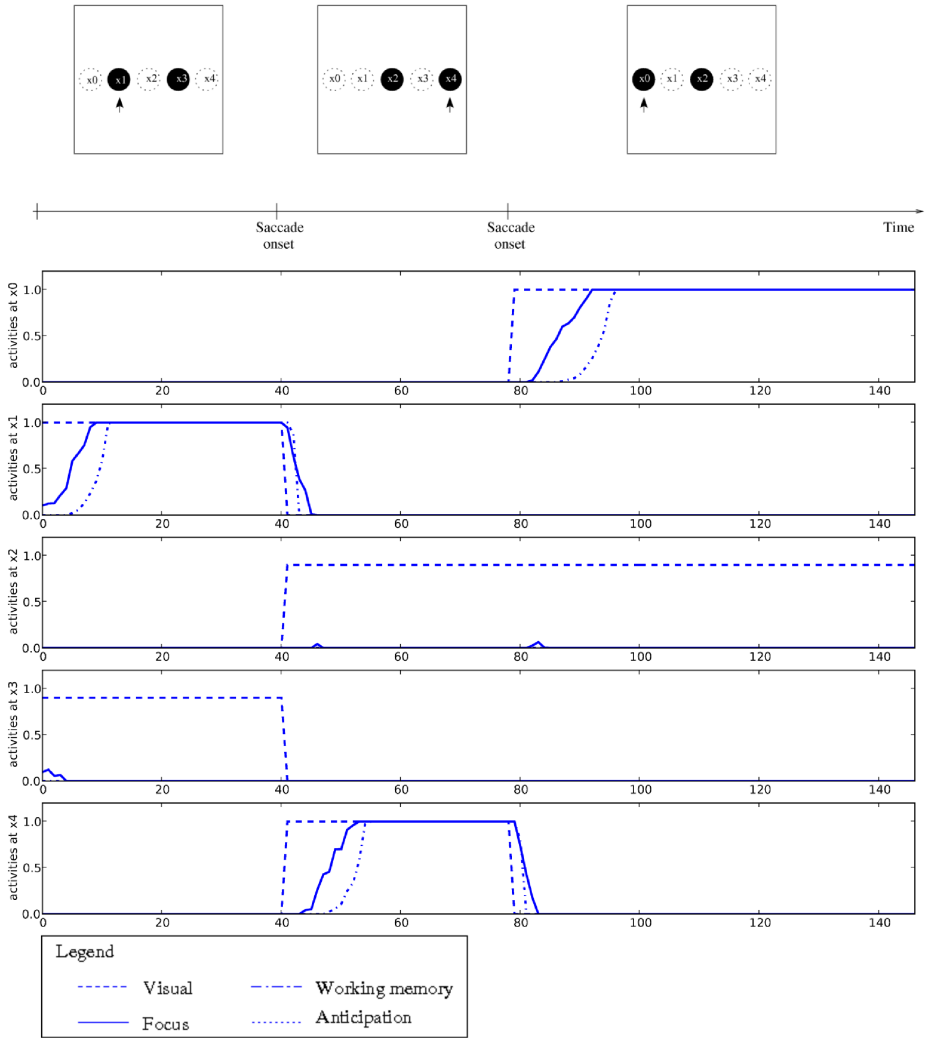

Fig. 7. The experiment is the same as in Figure 6 except that the anticipatory signal is disabled.

4 Discussion

In this paper, we have presented a continuous attractor network model that is able to anticipate the consequences of its own movements by actually predicting the visual scene as it is supposed to be after the execution of an action. Furthermore, the model also illustrates how this information is used in the context of a serial search of a target among a set of distractors. Each already focused target is kept within a working memory area that is updated with regards to eye movements.

The model is of a completely distributed nature and does not require any central supervisor. All the units in the model satisfy a dynamical equation. When dealing with this kind of dynamic model, the integration time of the units is a critical factor as shown in [30], which shares some ideas with the present model. It means that in our case, even if we make the hypothesis that the perception is available during the saccade (ignoring also that the perception is smeared), the working memory could be updated dynamically with the perception only if the movement's speed doesn't exceed a critical limit. In the case of saccadic eye movements, it is then necessary to have an anticipatory mechanism. We are definitely speaking about anticipation since a prediction about the future perception is used to maintain a coherent memory which is mandatory to accomplish the task we designed. It is nonetheless not limited to that particular case since scanning several potential targets is one of the basic primitives we use when performing a visual search task.

The question of learning the underlying transformation of the anticipatory mechanism, namely the convolution product of the focus map and the working memory, remains open and is still under study. We did implement a learning mechanism, under restrictions and strong hypotheses that rely heavily on the difference between the pre-saccadic prediction and the post-saccadic actual perception. This self generated signal is able to measure to what extent the prediction is correct or not. Hence, it is quite easy to modify the weights accordingly. The main difficulty during learning remains the sampling distribution of examples within the input space, which is a well known problem in information and learning theory. Without any additional motivational system that could bias the examples according to a given task, it is quite unrealistic to rely on a regular distribution of examples.

Finally, the coherence of the visual world is solely based on an anticipatory mechanism that ultimately allows the identification of identical targets before and after a saccade, despite drastic changes in the visual perception. The prediction of the future state of the visual memory enriches the perception of the visual world in order, for example, to prevent focusing twice on the same stimulus. Of course, this model does not pretend to be complete nor accurate and does not tackle a number of problems that are directly related to visual perception. However, we think that the possibility to unconsciously anticipate our own actions using a dynamic working memory could be extended to other motor tasks involving other types of perception as well.

References

1. Amari, S.I.: Dynamical study of formation of cortical maps. Biological Cybernetics **27** (1977) 77–87
2. Burr, D.: Eye movements, keeping vision stable. Current Biology **14** (2004)
3. Carpenter, R.: Movements of the Eyes, 2nd edition. Pion Ltd London (1988)
4. Chelazzi, L., Miller, E.K., Duncan, J., Desimone, R.: A neural basis for visual search in inferior temporal cortex. Nature **363** (1993) 345–347
5. Craighero, L., Fadiga, L., Rizzolatti, G., Umilta, C.: Action for perception : a motor-visual attentional effect. Journal of Experimental Psychology **25** (1999)
6. Duncan, J., Humphreys, G.W.: Visual search and stimulus similarity. Psychological Review **96** (1989) 433–458
7. Findlay, J.M., Brown, V.: Eye scanning of multi-element displays: I. scanpath planning. Vision Research **46** (2006a) 179–195
8. Findlay, J.M., Brown, V.: Eye scanning of multi-element displays: Ii. saccade planning. Vision Research **46** (2006b) 216–227
9. Goodale, M.A., Milner, A.D.: Seperate visual pathways for perception and action. Trends in Neurosciences **15** (1992) 20–25
10. Hamker, F.H.: A dynamic model of how feature cues guide spatial attention. Vision Research **44** (2004) 501–521
11. Itti, L., Koch, C.: Computational modeling of visual attention. Nature Reviews Neuroscience **2** (2001) 194–203
12. James, W.: The principles of psychology. New York : Holt (1890)
13. Jeannerod, M.: Neural simulation of action : A unifying mechanism for motor cognition. NeuroImage **14** (2001) S103–S109
14. Kleiser, R., Seitz, R.J., Krekelberg, B.: Neural correlates of saccadic suppression in humans. Current Biology **14** (2004) 386–390
15. Koch, C., Ullman, S.: Shifts in selective visual attention : Towards the underlying neural circuitry. Human Neurobiology **4** (1985) 219–227
16. Kowler, E., Andersen, E., Dosher, B., Blaser, E.: The role of attention in the programming of saccade. Vision Research **35** (1995) 1897–1916
17. Leigh, R.J., Zee, D.S.: The Neurology of Eye Movements, 3rd edition. Philadelphia: FA Davis Company (1999)
18. Li, W., Martin, L.: Saccadic suppression of displacement : separate influences of saccadic size and of target retinal eccentricity. Vision Research **37** (1997)
19. Luck, S.J., Chelazzi, L., Hillyard, S.A.: Neural mechanisms of spatial attention in areas v1, v2 and v4 of macaque visual cortex. Journal of Neurophysiology **77** (1997)
20. Merriam, E.P., Colby, C.L.: Active vision in parietal and extrastriate cortex. The Neuroscientist **11** (2005) 484–493
21. Moore, T., Fallah, M.: Control of eye movements and spatial attention. PNAS **98** (2001) 1273–1276
22. Moore, T., Tolias, A.S., Schiller, P.H.: Visual representations during saccadic eye movements. Neurobiology **95** (1998) 8981–8984
23. Moran, J., Desimone, R.: Selective attention gates visual processing in the extrastriate cortex. Science **229** (1985) 782–784
24. O'Regan, J.K., Noë, A.: A sensorimotor account of vision and visual consciouness. Behavioral and Brain Sciences **24** (2001) 939–1031
25. Posner, M.I., Cohen, Y.: Components of visual orienting. In Bouma, H., Bouwhuis, D., eds.: Attention and performance X. (1984) 531–556

26. Posner, M.I., Petersen, S.E.: The attentional system of the human brain. Annual Review of Neurosciences **13** (1990) 25–42
27. Reynolds, J.H., Desimone, R.: The role of neural mechanisms of attention in solving binding problem. Neuron **14** (1999) 19–29
28. Rizzolatti, G., Riggio, L., Dascola, I., Ulmita, C.: Reorienting attention accross the horizontal and vertical meridians. Neuropsychologia **25** (1987) 31–40
29. Ross, J., Morrone, C., Goldberg, M.E., Burr, D.C.: Changes in visual perception at the time of saccades. Trends in Neurosciences **24** (2001) 113–121
30. Rougier, N.P., Vitay, J.: Emergence of attention within a neural population. Neural Networks **19** (2006) 573–581
31. Rumelhart, D.E., Hinton, G.E., McClelland, J.L.: A general framework for parallel distributed processing. In: Parallel Distributed Processing, Vol. 1, MIT Press (1987)
32. Simons, J.S.: Current approaches to change blindness. Visual Cognition **7** (2000)
33. Sommer, M.A., Wurtz, R.H.: Influence of the thalamus on spatial visual processing in frontal cortex. Nature **444** (2006) 374–377
34. Taylor, J.G.: Neural bubble dynamics in two dimensions. Biological Cybernetics **80** (1999) 5167–5174
35. Treisman, A., Gelade, G.: A feature-integration theory of attention. Cognitive Psychology **12** (1980) 97–136
36. Treue, S., Maunsell, J.H.R.: Attentional modulation of visual motion processing in cortical areas mt and mst. Nature **382** (1996) 539–541
37. Tsotsos, J.K., Culhane, S.M., Lai, W.Y.K., Davis, N.: Modeling visual attention via selective tuning. Artificial Intelligence **78** (1995) 507–545
38. Ungerleider, L.G., Mishkin, M.: Two cortical visual systems. In: Analysis of Visual Behavior. MIT Press (1982) 549–586
39. Vitay, J., Rougier, N.P.: Using neural dynamics to switch attention. In: International Joint Conference on Neural Networks, IJCNN (2005)
40. Wolfe, J.M.: Visual search. In: Attention, University College London Press (1998)
41. Wolfe, J.M.: Visual attention. In: Seeing : Handbook of Perception and Cognition, 2nd ed., De Valois KK (2000) 335–386

Appendix

Dynamic of the Neurons

Each sigma neuron *loc* in a map computes a numerical differential equation given by equation (8), which is a numerized version of that proposed in [1] and [34]:

$$act_{loc}(t+1) = \sigma(act_{loc}(t) + \frac{1}{\tau}(-(act_{loc}(t) - baseline) + \frac{1}{\alpha}\sum_{aff} w_{aff} act_{aff}(t) + \frac{1}{\alpha}\sum_{lat} w_{lat} act_{lat}(t))) \quad (8)$$

Each sigma-pi neuron *loc* in a map computes a numerical differential equation given by equation (9):

$$act_{loc}(t+1) = \sigma(act_{loc}(t) + \frac{1}{\tau}(-(act_{loc}(t) - baseline) + \frac{1}{\alpha}\sum_{lat} w_{lat} act_{lat}(t) + \frac{1}{\alpha}\sum_{(i,j)\in E_{loc}} w_{sigmapi} act_{aff_i}(t) act_{aff_j}(t)) \quad (9)$$

where $\sigma(x)$ is a semi-linear function assuring that $0 \leq \sigma(x) \leq 1$, τ is the time constant of the equation, α is a weighting factor for external influences, *aff* is a neuron from another map and *lat* is a neuron from the same map. To know how the set of afferent neurons E_{loc} is determined in the case of a sigma-pi map, please refer to the section 3.2 describing the model.

The size, τ, α and baseline parameters of the different maps are given in the following table:

Map	*Size*	*Type*	*Baseline*	τ	α
INPUT	40*40	Sigma	0.0	0.75	6.0
FOCUS	40*40	Sigma	-0.05	0.75	13.0
WM	40*40	Sigma	-0.2	0.6	13
THAL_WM	40*40	Sigma	0.0	0.6	13
ANTICIPATION	40*40	Sigma-Pi	0.0	2.0	5.0

Connections intra-map and inter-map

The lateral weight from neuron *lat* to neuron *loc* is:

$$w_{lat} = Ae^{-\frac{dist(loc,lat)^2}{a^2}} - Be^{-\frac{dist(loc,lat)^2}{b^2}} \text{ with } A, B, a, b \in \Re^{*+}, loc \neq lat\ . \quad (10)$$

where $dist(loc, lat)$ is the distance between *lat* and *loc* in terms of neuronal distance on the map (1 for the nearest neighbor). In the case of a "receptive field"-like connection between two maps, the afferent weight from neuron *aff* to neuron *loc* is:

$$w_{aff} = Ae^{-\frac{dist(loc,aff)^2}{a^2}} \text{ with } A, a \in \Re^{*+} \quad (11)$$

In the case of the sigma-pi connections, all the weights are the same:

$$w_{sigma-pi} = A \text{ with } A \in \Re^{*+} \quad (12)$$

The connections in the model are described the following table:

Source Map	*Destination Map*	*Type*	*A*	*a*	*B*	*b*
INPUT	FOCUS	receptive-field	0.25	2.0	-	-
FOCUS	FOCUS	lateral	1.7	4.0	0.65	17.0
INPUT	WM	receptive-field	0.25	2.0	-	-
FOCUS	WM	receptive-field	0.2	2.0	-	-
WM	WM	lateral	2.5	2.0	1.0	4.0
WM	THAL_WM	receptive-field	2.35	1.5	-	-
THAL_WM	WM	receptive-field	2.4	1.5	-	-
ANTICIPATION	ANTICIPATION	lateral	1.6	3.0	1.0	4.0
WM, FOCUS	ANTICIPATION	sigma-pi	0.05	-	-	-
ANTICIPATION	WM	receptive-field	0.2	2	-	-