



HAL
open science

A dynamic neural field approach to the covert and overt deployment of spatial attention

Jérémy Fix, Nicolas P. Rougier, Frédéric Alexandre

► **To cite this version:**

Jérémy Fix, Nicolas P. Rougier, Frédéric Alexandre. A dynamic neural field approach to the covert and overt deployment of spatial attention. *Cognitive Computation*, 2011, 3 (1), pp.279-293. 10.1007/s12559-010-9083-y . inria-00536374

HAL Id: inria-00536374

<https://inria.hal.science/inria-00536374>

Submitted on 16 Nov 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A dynamic neural field approach to the covert and overt deployment of spatial attention

Jeremy Fix · Nicolas Rougier · Frederic Alexandre

Received: date / Accepted: date

Abstract The visual exploration of a scene involves the interplay of several competing processes (for example to select the next saccade or to keep fixation) and the integration of bottom-up (e.g. contrast) and top-down information (the target of a visual search task). Identifying the neural mechanisms involved in these processes and in the integration of these information remains a challenging question. Visual attention refers to all these processes, both when the eyes remain fixed (covert attention) and when they are moving (overt attention). Popular computational models of visual attention consider that the visual information remains fixed when attention is deployed while the primates are executing around three saccadic eye movements per second, changing abruptly this information. We present in this paper a model relying on neural fields, a paradigm for distributed, asynchronous and numerical computations and show that covert and overt attention can emerge from such a substratum. We identify and propose a possible interaction of four elementary mechanisms for selecting the next locus of attention, memorizing the previously attended locations, anticipating

the consequences of eye movements and integrating bottom-up and top-down information in order to perform a visual search task with saccadic eye movements.

Keywords visual attention · eye movements · dynamic neural fields · emergence

1 Introduction

Several authors have proposed that the visual saccade has a central role in cognition [1, 2]. This elementary behaviour has been extensively studied, certainly because it includes most of the characteristics of a cognitive task and particularly its complexity and its great number of participating factors.

As such, the intrinsic parameters of a saccade are limited: its metric (direction and amplitude) and its latency (time from target appearance to beginning of movement). Moreover, for a given saccade with a specific metric, the trajectory and the dynamics are generally stereotyped. Hence, the complexity is elsewhere: as pointed out in [3], two main questions must be answered: when and where will be the next saccade. Accordingly, this latter paper proposes a framework where both questions are realized by interconnected information flows, implemented in five levels, from the most automatic one to the most cognitive one. Markedly, these information flows are characterized by two kinds of inputs: exogenous and endogenous. Exogenous inputs correspond to information coming from the outside of the agent (from the characteristics of the visual stimuli to the spoken instructions given during the behavioural task). The endogenous inputs correspond to information elaborated by various parts of the nervous system (like the memory of previously visited locations, the current goal of the task or internal needs that will make some targets preferable to others). Now, if one

Jeremy Fix
SUPELEC, 2 rue Edouard Belin, F-57070 Metz, France
Tel.: +33(0)387-76-47-79
Fax: +33(0)387-76-47-00
E-mail: Jeremy.Fix@Supelec.fr

Nicolas Rougier
INRIA Nancy - Grand Est research center, Bat C, CS 20101, 54603
Villers les Nancy Cedex, France
Tel. : +33(0)383-59-30-92
Fax : +33(0)383-27-83-19
E-mail: Nicolas.Rougier@Loria.fr

Frederic Alexandre
INRIA Nancy - Grand Est research center, Bat C, CS 20101, 54603
Villers les Nancy Cedex, France
Tel. : +33(0)383-59-20-53
Fax : +33(0)383-27-83-19
E-mail: Frederic.Alexandre@Loria.fr

considers the amount of such exogenous or endogenous parameters affecting the answer to the two questions, it is easy to understand that visual scene processing through gaze orientation is considered a complex cognitive task.

In most modelling approaches, the answer to the when question is subordinated to the where question. The processing time required to construct an answer to the where question plus the waiting for some potential trigger signal (endogenous or exogenous) is compared to the measured timing in some behavioural tasks that are classically addressed in most evaluations of models, like the gap effect, express saccades, anti-saccade task, etc. For example, in [3], the saccade is triggered as a function of the comparison between two levels: the fixation level that decreases as the fixated object becomes less interesting and the move level that increases as the next target becomes more desirable. In this framework, the answer to the where question is elaborated by a general scheme of information that can be summarized as a set of modules where the endogenous and exogenous information are shaped and incorporated into a topological substratum (possibly made of several maps) yielding the location of the next saccade.

In some works, the topological substratum is itself decomposed in several maps. For example, as evoked above, [3] distinguishes between the foveal processing of the current object and the peripheral processing of forthcoming targets, extracting from each a single scalar signal that are compared to trigger a saccade. In [4], saccades due to endogenous and exogenous cues are prepared on a distinct substratum (proposed to be situated respectively in the prefrontal and dorsal lobes of the cortex). The algorithm for decision making is reduced to the triggering of the first saccade, built on endogenous and exogenous cues, reaching a certain threshold. All these somewhat complex schemes have been supplanted by the very simple and elegant solution proposed in [5], where it is shown that all the behaviours reported in the previous models can also emerge from a single map where central and peripheral vision, exogenous and endogenous cues are merged. It is also proposed that this single map corresponds to the superior colliculus in the mammalian mid-brain. The superior colliculus receives exogenous and endogenous cues and also projects on the brainstem premotor circuits that trigger saccades [6]. Moreover, this so-called competitive integration model is consistent with the very influential model by Koch and Ullman [7, 8] that puts to the forth the principle of a saliency map as a basis for competition before decision in saccade programming.

However, as mentioned in [3], defining a framework does not necessarily "satisfy the formal requirements of a quantitatively testable model". A conceptual model can be made very attractive in a first approach but later proved impossible to be emulated *in silico*. In addition, if we are unable to bridge the gap between these two approaches, we may have

to question the validity of either the conceptual approach or the computational one. In this article, we propose to answer this question with a strongly constrained computational framework that helps to reconcile these two approaches.

Today, most computational models of attention are tightly linked to both anatomical and physiological data, taking into account a variety of structures known to be involved in visual attention. These models generally deal with neural computations using different mechanisms such as resonant ART formalism [9] or dynamic neural fields [10–14]. While we share a common framework with most of these latter models (integration of exogeneous and endogeneous information, feedback biasing effects) the proposed model differs since only few of these models address the complete sensorimotor loop taking into account the whole range of attentional related properties such as covert and overt attention, feature and spatial processing, the integration of exogenous and endogenous information, the selection and execution of a saccade as well as the memorization of the previously executed ones.

The model we propose here addresses this whole range of attentional related properties using a strongly constrained framework based on the dynamic neural field theory. It is consequently composed of a large set of different modules (or maps) that share a common definition of a computational unit. The specificity that we wish to highlight is related to its underlying mode of computation. Indeed, getting inspired from neuronal computation not only means defining a distributed, local block of computation. It also implies to get rid of such sequential computing principles as central clock, central executive, overseer and other centralized or symbolic representation. Instead, the observed behaviour is emerging from a fully distributed and numerical mode of computation where the semantic of the behaviour is solely governed by the interaction between the model and the external world.

The paper is organized as follows. We first introduce the model in the next section, in particular by explaining its individual components, the properties that emerge from each of them (selection, working memory and anticipation) and the way they are combined all together. We then illustrate in section 3 the behaviour of the model on a visual search task involving saccadic eye movements. The description of the model will be firstly disconnected from the biological facts that motivated it. A hypothetical binding between the model areas and cortical and subcortical structures of the primate brain, as well as the implications of the model are then discussed in section 4. All the simulations presented in this paper are written with the DANA library [15] available at <http://dana.loria.fr>.

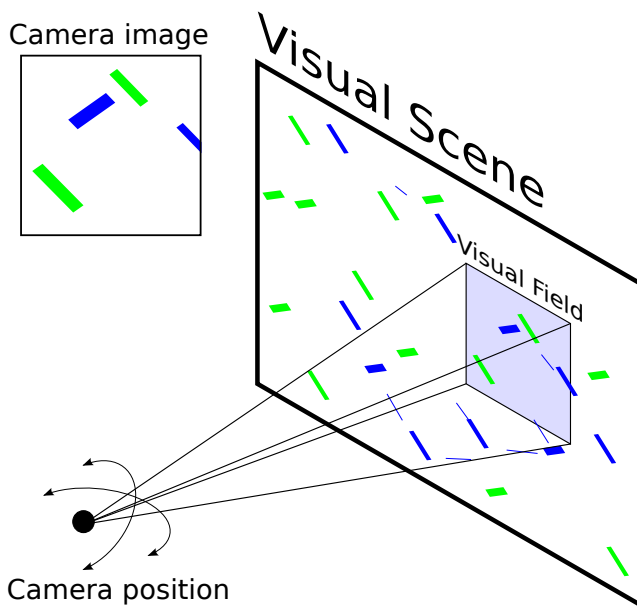


Fig. 1 The camera is placed in front of a visual scene and is able to pan and tilt. The visual display consists of several coloured and oriented bars. Since the visual field of the camera does not cover the whole scene, it is thus necessary to move it around to accurately explore the whole visual scene. It is to be noted that the perceived image is deformed because of the position of the camera and the projection of the visual scene onto the camera surface : a stimulus appears smaller as its eccentricity is increasing.

2 The Model

2.1 Experimental Setup

Through this whole section, we will use a simple experimental setup where a mobile camera (pan/tilt) is placed in front of a visual scene (fig. 1). The visual field of the camera does not cover the whole visual scene and it is thus necessary to move the camera to explore the whole visual scene depending on the task requirements. The visual scene is composed of a set of oriented ($+45^\circ$ and $+135^\circ$) and coloured (green and blue) bars on a neutral background. The task can be either to look for a specific orientation or colour or to look for a combination of such features as in conjunction search tasks (e.g. “look for a blue bar oriented at $+45^\circ$ ”).

2.2 Functional overview

As explained above, covert and overt visual attention can be summarized as bringing into a saliency map endogenous and exogenous information and dealing with critical temporal aspects for the consistency of decision making. The architecture of our model is composed of a set of maps, gathered in four processing poles, as depicted in figure 2. The biological validity of that architecture will be discussed at

the end of the paper. For the moment, we introduce its functional principles, that will be described through this section, and relate them to attention-related mechanisms that must emerge from local distributed computing to allow for those principles.

The *sensory pole* integrates both bottom-up and top-down visual information. The bottom-up information is provided by the visual input, processed along several dimensions (colour, orientation). The top-down information corresponds to the target template of the visual search task and to the current spatial focus of attention. These two information are respectively provided by the feature processing pathway and the spatial processing pathway, a division of processing related to the ventral/dorsal division of the primate brain [16]. The top-down signals multiplicatively modulate the bottom-up information, an influence that is consistent with current models of visual attention [17]. This influence allows to enhance the representation, within the *sensory pole*, of relevant features or spatial locations. The *sensory pole* therefore acts as an intermediate layer, through which one pathway indirectly influences the other.

The *feature processing pole* gathers a set of few units that are processing features with coarse-grained receptive fields. The units of *perceived features* map extract, within their wide receptive field covering the whole visual field, the maximally active features. The target template of the visual search task is held within the *target map*. The activities of these units are integrated by the units of the *motor pole* to trigger the execution of a saccade, when the focus of attention is on a target, or the disengagement of spatial attention when the focus of attention is on a distractor.

The *spatial processing pole* gathers a set of maps allowing to select the next attended location, to memorize the previously attended locations, to anticipate the consequences of eye movements on the memorized locations, and to disengage spatial attention when the currently attended stimulus is a distractor. The implementation of these mechanisms will be described in the next section. From a functional perspective, the *saliency map* provides a unified representation of the behavioural relevance of each spatial location within the visual field. It excites the *focus map* in which a competition is engaged to select the next attended location. The successively attended locations are memorized in a *working memory*. This memory allows to disengage attention when a distractor is attended and also to bias the exploration toward non-previously attended locations. This mechanism is analogous to the inhibition of return (IOR) [18].

As most of the primate brain areas involved in the control of saccadic eye movements have been shown to encode information in an eye-centred frame of reference, all the maps of the spatial processing pole also use an eye-centred representation. Then, as a visual scene is explored with sac-

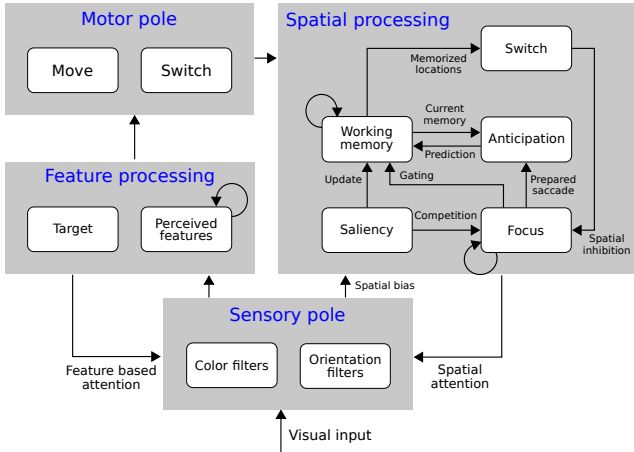


Fig. 2 Schematic illustration of the proposed architecture. The *sensory* pole integrates both bottom-up and top-down visual informations. The visual input is processed along several dimensions (colour, orientation) and multiplicatively combined with top-down informations indicating relevant feature or spatial locations, and respectively provided by the *feature processing* and *spatial processing* poles. The *feature processing* pole holds the target template and extracts from the *sensory* pole the maximal activity of their respective *sensory map*. The *spatial processing* pole is divided in several components ultimately leading to the selection of the attentional focus and its potential disengagement through a working memory circuit. The consistency of the working memory across saccades is ensured by anticipating the consequences of a planned eye movement on the position of the previously attended locations stored in working memory.

cadic eye movements, the working memory has to be updated accordingly. This update is obtained in our model by anticipating the consequences of the impending eye movement on the memorized position of stimuli in working memory. This mechanism ensures that the location of the previously attended stimuli is correctly transferred in the post-saccadic frame of reference.

Finally, and as it will be explained more clearly in the next sections, all these functions result from the interactions of distributed and dynamical units. The function, that we attribute to a field, results from the position of the field within the interconnected network of units and from the parameters that determine the evolution of their activity. There is therefore no supervisor observing and regulating the activities of the network. The visual exploration behaviour that is observed in the application in section 3 emerges only from local computations.

2.3 Computational paradigm

The model presented in this paper relies on dynamic neural fields (DNF), a model of the dynamic of a neural population [19–22]. The equation (1), a discretized equation in time of the DNF, states that the evolution of a field u_i depends on a relaxation term ($-u_i(t)$), an external input ($I_i(t)$), a baseline

firing (h) as well as lateral interactions within the field (w_{ij}). While some dynamical properties have been demonstrated mathematically (formation of stable patterns [20], travelling waves [22]), this framework has also been applied successfully on several sensorimotor tasks [23–26].

$$\frac{1}{\tau} \Delta u_i(t) = -u_i(t) + I_i(t) + \sum_j w_{ij} u_j(t) + h$$

$$u_i(t+1) = f(u_i(t) + \Delta u_i(t)) \quad (1)$$

$$w_{ij} = A_+ \exp\left(-\frac{d_{ij}^2}{2\sigma_+^2}\right) + A_- \exp\left(-\frac{d_{ij}^2}{2\sigma_-^2}\right) \quad (2)$$

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

The lateral influence w_{ij} introduces a topology in the neural field. It is usually chosen to be of a mexican-hat shape as in equation (2), where d_{ij} is the distance within the field between two locations i and j . The neural field will be one-dimensional in the illustrations of this section, but the full model presented in the result section involves two-dimensional neural fields. Figure 3 illustrates two lateral influences, function of the distance between two units in the field. A locally excitatory, broadly inhibitory influence (dashed line on figure 3) induces a competition between the excited units, for example to select a target for an eye movement among several candidates. On the figure 3, the excitatory and inhibitory gaussians have the same amplitude ($A_+ = A_-$) but the inhibitory component is wider than the excitatory component ($\sigma_- > \sigma_+$). This results in a combined effect that is only inhibitory. As we will see in the next section, this is sufficient to obtain the selection property of the field. A stronger lateral excitation (solid line on figure 3) can induce remanence : when close units get excited by an input, they can stay excited despite the removal of the input because the weaker external input is compensated by the local recurrent excitation.

The architecture of the model used to perform a visual search task with covert and overt attention is depicted on figure 2. Each field is a two dimensional set of units whose activity evolves according to the first order differential equation (1). As introduced in section 2.2, it consists of mainly four functional components : a mechanism of selection in the *focus* map, a mechanism to sustain activities in the *working memory* maps and a mechanism anticipating the state of the working memory given a planned eye movement in the *anticipation* maps. Finally, endogenous (target template, current spatial locus of attention) and exogenous (visual input) information are combined within the *sensory* pole. Each of these components is described individually in the next sections, before presenting a simulation of the whole model

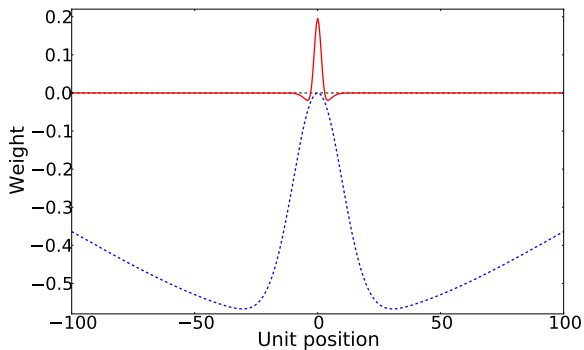


Fig. 3 Local-excitation, global inhibition lateral connectivity (dashed line) can induce competition between the excited units. $N = 100$, $A_+ = 0.6$, $\sigma_+ = 0.1.N$, $A_- = 0.6$, $\sigma_- = N$. The two gaussians being of equal amplitude, the net influence is only inhibitory. A local-excitation, local-inhibition lateral connectivity (solid line) can induce remanence. $N = 100$, $A_+ = 0.24$, $\sigma_+ = 0.03.N$, $A_- = 0.045$, $\sigma_- = 0.07.N$

on a visual search task with saccadic eye movements.

2.4 Selection

Visual attention has been defined as the capacity of the brain to focus on particular aspects of the visual information while ignoring distractors. The mechanisms involved in the deployment of visual attention remain unclear but it is proposed that the selection of the next attended location depends on the representation of the behavioural relevance of a visual information.

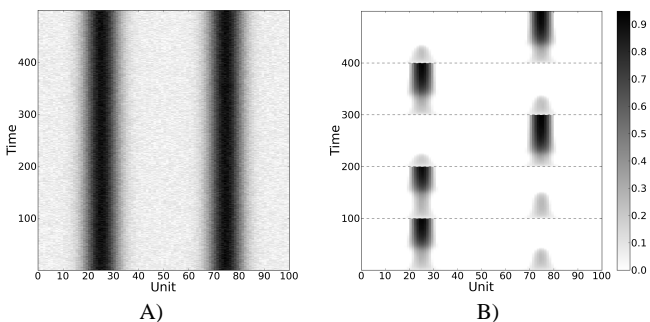


Fig. 4 Evolution of the activities in a 1D field (B), fed with a static input (A), with lateral influences defined as a local excitation-global inhibition (dashed line on figure 3). For illustration purpose, the activities of the field are reset every 100 time steps. This reset starts a new competition between the two excited locations. During each epoch, after a transient phase, the neural field settles in a state where only one of the excited regions is active. A small random noise, with an amplitude of 0.15, added to the input provides a stochastic competition.

While multi-layered neural networks have been introduced to perform a competition between several motor pro-

grams [27], the selection relies, in the presented model, on locally excitatory, globally inhibitory (dashed line on figure 3) lateral influences inside a single neural field [25, 28]. The ability of dynamic neural fields to select one among several excited regions has been used by [29] to address dynamical properties of saccadic target selection. Figure 4 illustrates the behaviour of such a 1D neural field in a space x time representation. The input (fig. 4A) consists of two excited regions of equal amplitude. As observed on figure 4B, there is first a transient phase where the two excited regions co-exist. As the competition is settled within the field, only one of the two excited regions remain active. For the purpose of the illustration, the activities within the field are reset every 100 time steps, which engages a new competition.

2.5 Dynamic working memory

While a strong competition is elicited by long-range lateral inhibition, a spatial information can be maintained with a local connectivity pattern (solid line on figure 3), with a stronger excitatory component and a weaker inhibitory influence (the inhibitory component preventing the activity to spread over the whole field) [30]. In addition, in the complete architecture (fig. 2), the working memory holds the stimuli that have been previously attended. This means that the emergence of a stimulus in working memory has to be gated by spatial attention. This gating is obtained by setting a negative baseline (h in equation 1) of the neural field. This implies that the excitation from the saliency map alone is not sufficient to drive the working memory but it needs also the excitation from the *focus* map, indicating that a stimulus is attended. Finally, when spatial attention is disengaged, a previously attended stimulus has to remain in working memory. The strong lateral excitation of the working memory compensates the decrease of the excitatory drive and therefore allows to keep a stimulus in working memory despite being later unattended.

An example of this behaviour is illustrated on figure 5. The input consists of three stimuli of initially weak amplitude (fig. 5A). The amplitude of the three stimuli is successively increased, leading to the successive emergence of the stimuli in the working memory field (fig. 5B). This increase of amplitude overcomes the negative baseline of the units. When the amplitude of the input stimuli is decreased back to its initial value, the memorized stimuli are kept in working memory because the decrease of the external input is compensated by the recurrent lateral excitation.

Despite strong lateral excitation, the units in the working memory remain sensitive to dynamical evolution of their external input, for example when the input stimuli are slowly moving. However, while the dynamical behaviour of the field

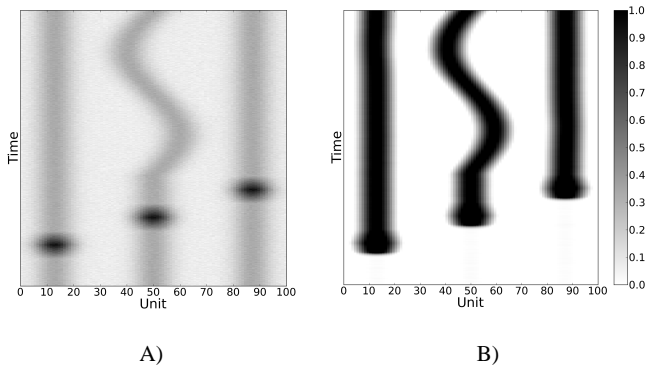


Fig. 5 Evolution of the activities of a 1D field (B) with strong lateral excitation and local inhibition as in figure 3. Three stimuli of initially equal amplitude feed the field. The amplitude of the stimuli are successively increased and decreased back to the initial amplitude (A). While strong lateral excitation allows to memorize the position of the stimuli, these lateral excitations are sufficiently weak to keep the field sensitive to dynamical evolutions of the external input.

allows to track slowly moving input stimuli, it is not sufficient to ensure the consistency of the working memory when saccadic eye movements are suddenly modifying the visual input. The next section introduces a mechanism by which consequences of saccadic eye movement, that are voluntary movements, are anticipated to satisfy this constraint.

2.6 Anticipation

The mechanism introduced in the previous section allows to keep in memory stimuli that are the target of an attentional bias. When saccadic eye movements are executed, and considering that the working memory holds the position of the previously attended stimuli in an eye-centred frame of reference, the working memory has to be updated.

We propose in our model that the consistency of the working memory between the pre and post saccadic perceptions is ensured by anticipating the consequences of an eye movement on the position of the memorized targets. When the eye movement is executed, the combination of this anticipation with the post-saccadic visual perception allows to update the working memory. The anticipatory activities are computed by combining the current state of the working memory with the motor command of the impending eye movement using sigma-pi units [31,32]. The multiplicative interactions of the afferences of the sigma-pi units are hardwired in the model, using equation (3), and lead to compute a translation of the input activities according to a motor vector through a convolution product.

$$I_k(t) = w \sum_j \text{input}_{k+j}(t) \cdot \text{command}_{\frac{N}{2}-j}(t) \quad (3)$$

where $\text{input}(t)$ represents the visual input and $\text{command}(t)$ a motor command. The subscripts refer to the indexes of the

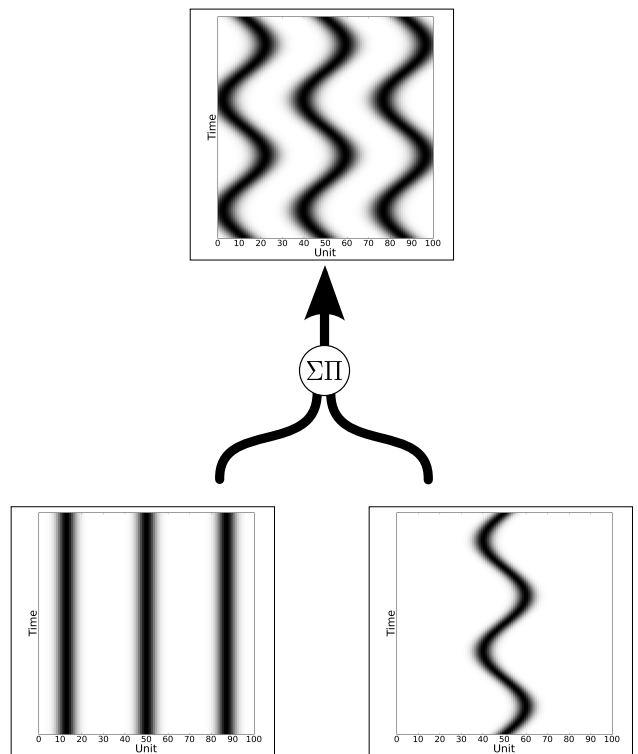


Fig. 6 The input $i(t)$ consists of three stationary stimuli (bottom left) and a time-varying motor signal $c(t)$ (bottom right). Combining these two signals with sigma-pi units according to the equation 3 allows to translate the input activities by the command signal. Since saccadic eye movements produce shifts of the visual input in a direction opposite to the eye movement, the projections are defined such that the input stimuli are translated in the opposite direction of the vector pointing from the centre of the field to the peak of the command. In this experiment, the weighting factor of equation 3 is set to $w = 0.2$.

discretized neural field. An illustrative example is shown on figure 6. The input consists of three stationary stimuli and a time-varying motor command. The activities of the sigma-pi units are the activities of the input units translated by a vector defined by the command. This computation is performed in "a single step" if we compare it to previous architectures performing continuous remapping [36]. In these architectures, the activities are translated dynamically during the execution of the movement. The mechanism we propose here allows to directly compute the shifted positions of the input before the execution of the movement. It allows to perform a specific sensorimotor transformation. Gain fields have been proposed in the literature to perform sensorimotor transformations [37], following in-vivo observations that neurons in different brain areas use such a coding scheme [38,39]. It has been shown that any linear combination of the information encoded in the input can be decoded from a gain field combining these two inputs. The sigma-pi mechanism we propose here can be understood as a reduction of a gain field to a unique sensorimotor transformation, the com-

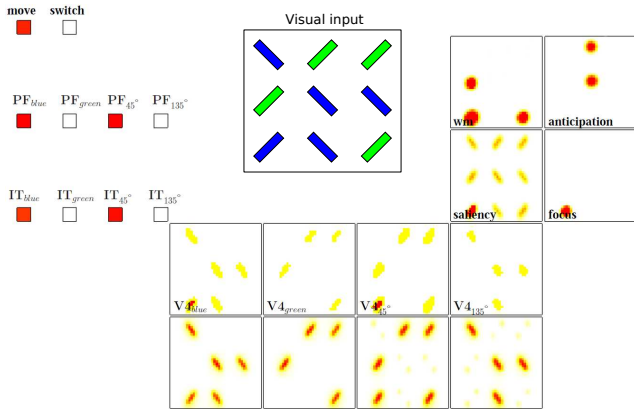


Fig. 7 Snapshot of the activities of the model performing a visual search task and looking for a blue target oriented at 45° (stimulus on the bottom left of the visual input), just before the execution of an eye movement. An array of 3×3 stimuli is presented, the target is a blue bar oriented at 45° and the activities of the *PF* maps are set accordingly. As all the stimuli share at least one feature with the target, they are initially equally represented within the *saliency* map. The lateral competition engaged within the *focus* determines the spatially attended location that is exciting in feedback the intermediate layer. This explains the increased activity of one of the stimuli. As the sensory maps (*V4*) feed the *IT* maps, one can decode from their activity the features under the attentional focus. The working memory (*wm*) contains all the stimuli that were previously attended and the *anticipation* map predicts their position, within the visual field, when the saccade will be executed. The discrete neural fields are two dimensional and made of 40×40 units.

putations being transferred from the gain field units directly into the synaptic connections between the input and the output. While the transformation is not learned in our model, one can note however that recent works successfully used self-organized neural networks of sigma-pi units in order to compute spatial transformations [33–35]

2.7 Combining bottom-up and top-down information

In the model we propose, the visual information is processed along two pathways; one is specifically processing non spatial visual attributes (colour, orientation, *Feature processing* in figure 2), the other is processing only spatial information (*Spatial processing* in figure 2). These two pathways are fed by a common intermediate layer (*sensory pole* in figure 2). The cross-talk between the two pathways follows the reentry hypothesis [40,41]. Each pathway, driven by the intermediate layer, also projects in feedback onto it. This ensures the consistency between the coarse-grained non-spatial feature representation of one pathway and the spatial representation of the other pathway. This computational principle of coarse-grained processing units fed by fine-grained processing units and sending feedback projections on them has been successfully employed in previous models of visual attention [41, 14, 42].

The figure 7 illustrates the activities within the model during a visual search, just before the execution of an eye movement. The intermediate layer (the four *V4* maps) is fed by the visual input and two reentrant signals originating from the *PF* and *focus* maps. These four neural fields encode four different features, from left to right two colours (green and blue) and two orientations (45° , 135°). The target template is encoded within the *PF* maps and is set on the illustration as the blue target oriented at 45° . This feature based feedback signal enhances the representation of all the stimuli that share at least one feature with the target. In the considered simulation, the display is made of stimuli that all share at least one feature with the target. In order to be consistent with the experimental observation that the time to perform a conjunction search (searching for a target among a display of stimuli that share at least one feature with the target) is linearly increasing with the number of distractors similar to the target, the saliency map of the spatial processing pathway integrates its inputs from the *V4* maps through a maximum receptive field. This means that all the stimuli are initially represented within the *saliency* map with the same amplitude. When the competition within the *focus* map (excited by the *saliency* map) is settled, one the stimuli is spatially attended. This leads to an excitatory feedback signal on the *V4* maps enhancing the representation of that spatial position. This enhancement ultimately biases the representation of the *IT* maps. At the steady state, the *focus* map encodes the spatially attended stimulus and the *IT* maps encode the feature of that stimulus. Therefore, the reentry mechanism ensures the consistency of the representations within the feature and spatial processing pathways and allows to retrieve the features of the attended stimulus. The features of the attended stimulus and of the target are integrated within the *motor pole* (*move* and *switch* units of figure 7) to trigger the execution of a saccade or the disengagement of spatial attention.

3 Results

In this section, we simulate the complete model during a visual search task with saccadic eye movements. The task for the model is to orient the camera toward targets defined by one feature (the blue bars), without focusing twice the same target, using the experimental setup presented in section 2.1. The target template is introduced in the model by setting appropriately the activities of four dedicated units, each representing a specific visual attribute (*PF* units on figure 7). The performances of the model are measured both at the behavioural level and at the unit level. At the behavioural level, we report the saccadic scanpath (filled circles on figure 8A). At the unit level level, we report the activities within the feature processing pathway (the activities of the four *IT* units

on figure 7) and spatial processing pathway (the mean activity of the *focus* map). In addition to report the consistency between the representation of the two pathways, the recordings of the units' activities allow to keep track of the deployment of spatial attention. In particular, since the covert deployment of attention is not followed by the execution of a saccade, it cannot be observed from the behavioural level while it can be observed from the internal monitoring of the model's activities, as would be measured by electrophysiologists.

On figure 8A, the extent of the visual field and its initial position is represented by the dashed square, the successive gaze targets are represented by the filled circles, while the dashed circles indicate the stimuli covertly attended (spatially selected without further executing a saccade). Figure 8B illustrates the activities of the four feature processing units (*IT* units of fig. 7), and the mean activity within the *focus* map as the model is performing the task. The vertical filled lines on the activity recordings indicate the time of saccade onset, while the dashed lines indicate a covert disengagement of spatial attention. The indexes on figure 8A and below the recordings of the units' activities on figure 8B correspond to the successive saccades and covert disengagements of spatial attention.

The successful performance of the task is observed on figure 8A since the model never performs a saccade toward a distractor nor focuses twice the same target. In addition, the units' recordings allow to keep track of the features of the stimulus below the focus of attention during the performance of the task. As the competition within the *focus* map settles, the activities within the *IT* maps converge to represent the features of the attended stimulus. In particular, when a distractor is spatially attended, its colour attribute drives the *switch* unit whose excitation leads to a disengagement of spatial attention (steps 4c, 5c and 9c).

There is one notable mistake for the saccade number 8 that corresponds to an illusory conjunction of a blue bar oriented at 135° while the target in the display is a blue bar oriented at 45° . At that time, two stimuli share the same receptive field : a target oriented at 45° and a distractor oriented at 135° . The top-down feature based bias only specifies the colour of the target and therefore does not constrain the orientation. To correctly resolve the orientation of the target, additional competitive mechanisms would be required within or before the *V4* maps. One may also note that the reaction time for some saccades is longer than for others. In our model, this phenomena is simply explained by the size of the projection of the stimuli on the model's retina. As the camera position is fixed and only its rotation is varied, some targets appear at larger eccentricities along the scan-path until the saccade number 6. Between the two saccades 5 and 6, the projection on the retina of the next target is smaller than the projection of the distractor. The bottom-up

drive of the distractor is therefore stronger than the bottom-up drive of the next target even if the latter is multiplicatively amplified by the top-down bias. This leads the model to first covertly attending the distractor before selecting the target and also to a longer time for spatially attending to the target.

4 Discussion

4.1 Limits of the model

The first limit of the model is the spatial extent of the working memory. The information is encoded within the working memory in an eye-centred frame of reference and its spatial extent is limited to the size of the visual field. This means that if a target goes outside the visual field, it will not be kept in working memory. This issue could be solved by associating with a relevant stimulus the motor programs required to retrieve it. In addition, from a mechanistic point of view, the maintenance in working memory in our model requires a permanent excitatory drive (from the *saliency* map). Indeed, if the lateral excitation within the working memory is made stronger in order to decrease the dependency on the external input excitatory drive for the maintenance of an information, the working memory content will not be sensitive to dynamical evolutions of the input anymore. There is therefore a trade-off between the minimal amplitude of the excitatory external input of the working memory, and its sensitivity to dynamical evolutions of the input.

The correct updating of the working memory relies on the integration of the pre-saccadic anticipation of the future position of the memorized stimuli with the post-saccadic visual perception. As the entrance in working memory relies on the same mechanism, some stimuli may emerge in working memory because their position is both a position occupied by a stimulus before the saccade and a position that will be occupied by a previously attended stimulus after the saccade. This leads to the potential emergence of non-attended stimuli in working memory. In the present model, the successful performance of the task also requires the anticipatory activities to be computed before the execution of a saccade. Otherwise, the working memory is not updated correctly. This provides constraints on the dynamic of the neural fields. The above mentioned limits, mainly due to the coarse grained functional modelling we propose here, would benefit from a finer grained modelling of the oculomotor circuit, for example by expanding the working memory and selection circuits. As it will be discussed in the next section, there are growing evidences that these two functions involve the basal ganglia.

Finally, as the focus of the model was mainly on the representation and processing of spatial visual information, the representation of features within the ventral stream of the model was kept coarse. In the model, we considered only

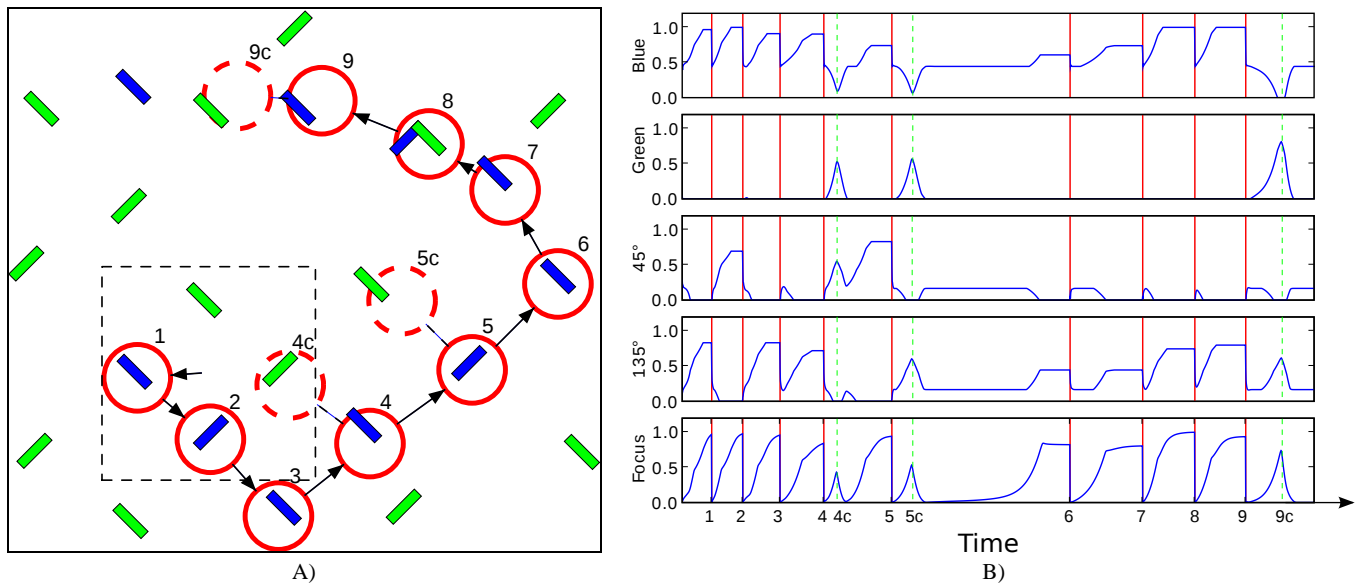


Fig. 8 A) Representation of the scanpath performed by the model during a visual search task where the target is a blue bar. The dashed square indicates the extent and initial position of the visual field. The dashed circles indicate the targets covertly selected by spatial attention (without further executing an eye movement toward them). These covertly attended locations are extracted from the recordings of the unit's activities. B) Activities recorded within the four *feature processing* units (selective for two colours and two orientations) and mean activity within the focus map while the model is performing the visual search task. The vertical lines indicate the onset of a saccade (solid line) or the disengagement of attention (dashed line). The indexes of the saccades below the activity recordings correspond to the figures indicated on the figure on the left. The three indexes 4_c , 5_c and 9_c correspond to covertly attended stimuli. A video of the model is available at <http://jeremy.fix.free.fr/demo.php?demo=CogComp2010>

four features, and used only eight units within this stream (four for extracting the most active features of the *sensory* pole and four for storing the target template of the visual search task). This could be easily extended by considering more detailed representations of visual features as in classical saliency models [8,43,44], or in modelling works on the ventral stream [45]. Using more sophisticated feature representation also implies to introduce additional competitive mechanisms within the model. Spatial based and feature based attention act in our model only as multiplicative gains on the afferences of the sensory units. It is known since the seminal work of Moran et al. [46] that attention modulates competition within the visual cortex, which led to introduce the biased competition framework [47]. One such mechanism has been proposed for example by Reynolds et al. [48]. Extending the feature processing pathway and introducing additional competitive mechanisms would lead to a model able to perform visual search tasks in real-world environments.

4.2 How the model relates to the primate brain areas

We would like now to lay emphasis on the relationship between the model and the primate brain areas. Recent reviews on the monkey brain areas involved in visual attention and the control of saccadic eye movements can be found in [49–51]. The neural fields of the model do not necessarily map

onto a single brain area as the functions we emphasized can be distributed among several brain areas. The visual attention literature stresses the requirement for a spatial saliency map integrating both endogenous and exogenous signals, and indicating the priority for processing a visual information. In fact, there is now strong evidences that the representation of visual saliency is shared among several cortical structures like the visual cortex, the parietal areas or the temporal areas [17]. There are also several works pointing out that different brain areas may contain a saliency map : pulvinar [52], V1 [53], lateral intraparietal area (LIP) [54], frontal eye field (FEF) [55], superior colliculus (SC) [10]. All these areas are part of the visual system or oculomotor circuit. Since they are strongly interconnected, it is not surprising that all these areas exhibit activities related to visual saliency.

The specificity of these different saliency maps remains to be clarified but the recurrent projections within this circuit may have one interesting function. The processing of visual information along two pathways (ventral and dorsal streams) [16] raises the question of the consistency between these representations, what has been referred to as the binding problem. The reentry hypothesis provides an elegant solution to the binding problem through recurrent projections within the brain areas [40,41]. Feedback projections have been identified both along the ventral visual stream [56] and the dorsal visual stream [57]. These projections allow to

propagate processing taking place in higher cortical areas to lower cortical areas, for example to enhance the representation of behaviourally relevant information.

Ultimately, the distributed representations of saliency, or at least the result of the competition engaged between them, have to converge onto the superior colliculus to trigger the execution of a saccade. Where may the competition between all these excitatory signals take place? The tonic and selective inhibition from the basal ganglia onto the superior colliculus places them in a good position for such a role of mediating the competition within the saliency maps [51].

The role of the basal ganglia in mediating the competition between motor programs has been proposed since a long time and the involvement of the different nuclei, for example in oculomotor control, is getting clearer [58]. It is now also accepted that they are involved in other high level cognitive functions (executive, motivation). In particular, part of the basal ganglia are reciprocally connected with the dorsolateral prefrontal cortex (dlPFC), an area where sustained activities during memory phases of behavioural paradigms have been observed [59,60]. The loop formed by the basal ganglia with dlPFC involves the mediodorsal (MD) nucleus of the thalamus [61,62]. It is interesting to note that MD is also participating in the recently identified corollary discharge pathway SC to FEF [63], and its inactivation leads to a partial suppression of anticipatory responses observed in FEF. Therefore, this nucleus of the thalamus is at a place of convergence of spatial short-memory signals as well as motor signals, around the time of the execution of a saccade. This circuit may then provide the neuronal basis of the working-memory introduced in our model whose consistency between the saccades is ensured by anticipating the consequences of an eye movement. Anticipatory responses are also observed in the parietal lobe [64]. It has been suggested that these anticipatory responses are a consequence of the local circuitry of LIP, under the influence of FEF [65]. Given that LIP and FEF are strongly interconnected, it may also be possible that the anticipatory responses observed in LIP are the consequence of anticipatory responses already present in FEF.

4.3 Large scale simulation of embodied cognitive architectures

This work is primarily an attempt to show that complex cognitive functions can emerge from a large set of distributed, asynchronous, numerical computations. Particularly, it is related to the exploration of visuomotor functions, including difficult questions about the coordination between bottom-up and top-down information flows and decisions from several kinds of criterion in space and time. As mentioned in the introductory part, visual attention is a particularly interesting function, since it remains relatively simple while pre-

senting an apparently linear, sequential behaviour (the scan-path) resulting from fully distributed computations. Nevertheless, this basic behaviour has also been presented as deeply linked to consciousness; also, it was mentioned above that this model can also apply to similar circuits, involving the dorsolateral prefrontal cortex and responsible for non-motor and more abstract and cognitive functions.

Our model shares with several other authors the choice for an approach in computational neuroscience and the elaboration of a saliency map where the different information flows converge and interact to elaborate the decision. Its deep interest is to simulate not only the saliency map but also a whole system, including a large number of other maps. This underlies that many functions related to the classical principles of saliency maps might be shared with other structures, which could explain why many neuronal structures have been proposed as candidates to this role. Also, this multi-map complex system was a good opportunity to illustrate the usefulness of computational neuroscience in this domain: not only such a complete system can be used to emulate a complex behaviour in an artificial agent, as a robot, but also some hidden parameters, not observed at the behavioural level, can be computed and monitored, for comparison with some physiological data or just to better understand how a cognitive property emerge from the interplay between several maps.

Indeed, the most original aspect of our model is related to its fully distributed nature and to the fact that such important mechanisms as selection, dynamic working memory and anticipation are obtained by emergence, from a set of identical processing units, without any central clock nor central executive. All the knowledge is brought in the design of the architecture of the network and other functional hypotheses coming from neuroscience, as explained above.

This architecture is also an illustration of the premotor theory of attention [66], showing an overlapping substratum between visual attention and saccadic eye movements. In our model, the same parieto-frontal circuit is exploited in the overt and in the covert case: its role is to decide on which stimulus to focus on; then the move order (the saccade) is triggered only if the characteristics of the stimulus are positively compared to the current instruction, otherwise the system is asked to switch to the next candidate: in this view also, covert attention is a pre-saccadic behaviour. In this perspective, our work is very much comparable with that of Trappenberg et al. [10]. Not only, it shares a similar view about the use of Neural Fields but also it refers to the same neuronal substratum. The work by Trappenberg concentrates on the place of integration of endogenous and exogenous information (the saliency map), proposed to be in the intermediate layer of the superior colliculus and on the temporal dynamic of its neurons. The specificity of our work is to implement a larger part of the primate visual network,

generating anticipation and working memory mechanisms, to explore the behaviour of such a multi-map system and to exploit it in the framework of autonomous robotics.

Our model also exploits the reentry hypothesis [40,41], showing that interaction between two information flows can originate from successive deposits of information on a common substratum. This is the typical case where a more symbolic centralized system would have built a structured representation; here it is shown that feedback and modal representation can propose a robust and distributed way to represent information and maintain its consistency.

Among its strongest points, this model demonstrates that it can cope not only with ascending perceptive information but also with top-down more cognitive instructions. This is more realistic and makes it close to real world applications, as we have begun to study, using our robotic platform and real images. Our next goal is to make more complex the nature of top-down instructions and to go beyond the simple search of a target defined from its visual characteristics. Instead, the target could be discovered from more subtle interactions with the environment and the choice of the action could be more contextual. This implies to study more closely neuronal structures in the basal ganglia and the limbic system and their interactions with the structures present in the model.

References

1. D. Ballard, M. Hayhoe, P. Pook, R. Rao, Deictic codes for the embodiment of cognition, *Behavioral and Brain Sciences* 20 (4) (1997) 723–42; discussion 743–67.
2. F. Alexandre, Cortical basis of communication: local computation, coordination, attention, *Neural Netw* 22 (2) (2009) 126–33.
3. J. Findlay, R. Walker, A model of saccade generation based on parallel processing and competitive inhibition, *Behav Brain Sci* 22 (4) (1999) 661–74.
4. A. Kramer, D. Irwin, J. Theeuwes, S. Hahn, Oculomotor capture by abrupt onsets reveals concurrent programming of voluntary and involuntary saccades, *Behavioral and Brain Sciences* 22 (1999) 689–690.
5. R. Godijn, J. Theeuwes, Programming of endogenous and exogenous saccades: evidence for a competitive integration model, *J Exp Psychol Hum Percept Perform* 28 (5) (2002) 1039–54.
6. T. Isa, Intrinsic processing in the mammalian superior colliculus, *Current Opinion Neurobiology* 12 (6) (2002) 668–77.
7. C. Koch, S. Ullman, Shifts in selective visual attention: towards the underlying neural circuitry., *Human Neurobiology* 4 (4) (1985) 219–27.
8. L. Itti, C. Koch, Computational modeling of visual attention, *Nature Review Neuroscience* 2 (3) (2001) 194–203.
9. V. Cutsuridis, A cognitive model of saliency, attention, and picture scanning, *Cognitive Computation* 1 (2009) 292–299.
10. T. Trappenberg, M. Dorris, D. Munoz, R. Klein, A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus, *J Cogn Neurosci* 13 (2) (2001) 256–71.
11. S. Schneider, W. Erlhagen, A neural field model for saccade planning in the superior colliculus: speed-accuracy tradeoff in the double-target paradigm, *Neurocomputing* 44-46 (2002) 623–628.
12. J. Johnson, J. Spencer, G. Schoner, Moving to higher ground: The dynamic field theory and the dynamics of visual cognition., *New Ideas Psychol* 26 (2) (2008) 227–251.
13. C. Faubel, G. Schoner, Learning to recognize objects on the fly: a neurally based dynamic field approach., *Neural Networks* 21 (4) (2008) 562–76.
14. G. Deco, E. Rolls, A neurodynamical cortical model of visual attention and invariant object recognition., *Vision Research* 44 (6) (2004) 621–42.
15. N. Rougier, J. Fix, Dana, distributed asynchronous numerical and adaptive modeling framework, *Frontiers in Neuroinformatics* submitted.
16. M. Goodale, A. Milner, Separate visual pathways for perception and action., *Trends in Neurosciences* 15 (1) (1992) 20–5.
17. J. Reynolds, L. Chelazzi, Attentional modulation of visual processing, *Annu Rev Neurosci* 27 (2004) 611–47.
18. M. Posner, Y. Cohen, Attention and performance X, Lawrence Erlbaum Associates, 1984, Ch. Components of visual orienting, pp. 531–556.
19. H. Wilson, J. Cowan, A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue., *Kybernetik* 13 (2) (1973) 55–80.
20. S. Amari, Dynamics of pattern formation in lateral-inhibition type neural fields, *Biological Cybernetics* 27 (2) (1977) 77–87.
21. J. Taylor, Neural bubble dynamics in two dimensions, *Biological Cybernetics* 80 (1999) 5167–5174.
22. S. Coombes, Waves, bumps, and patterns in neural field theories, *Biological Cybernetics* 93 (2) (2005) 91–108.
23. W. Erlhagen, G. Schoener, Dynamic field theory of movement preparation, *Psychol Rev* 109 (3) (2002) 545–72.
24. W. Erlhagen, E. Bicho, The dynamic neural field approach to cognitive robotics, *J Neural Eng* 3 (3) (2006) R36–54.
25. N. Rougier, J. Vitay, Emergence of attention within a neural population, *Neural Network* 19 (5) (2006) 573–81.
26. E. Sauser, A. Billard, Dynamic updating of distributed neural representations using forward models., *Biol Cybern* 95 (6) (2006) 567–88.
27. K. Gurney, T. Prescott, P. Redgrave, A computational model of action selection in the basal ganglia. i. a new functional anatomy, *Biol Cybern* 84 (6) (2001) 401–10.
28. J. Vitay, N. Rougier, Using neural dynamics to switch attention, in: *International Joint Conference on Neural Networks (IJCNN 2005)*, 2005.
29. K. Kopecz, G. Schoner, Saccadic motor planning by integrating visual information and pre-information on neural dynamic fields., *Biol Cybern* 73 (1) (1995) 49–60.
30. J. Johnson, J. Spencer, S. Luck, G. Schoner, A dynamic neural field model of visual working memory and change detection., *Psychol Sci* 20 (5) (2009) 568–77.
31. J. Fix, J. Vitay, N. Rougier, A distributed computational model of spatial memory anticipation during a visual search task, in: M. Butz, O. Sigaud, G. Baldassarre, G. Pezzulo (Eds.), *Anticipatory Behavior in Adaptive Learning Systems: From Brains to Individual and Social Behavior*, Vol. 4520 of LNCS, Springer, 2007, pp. 170–188.
32. F. Alexandre, F. Guyot, Neurobiological inspiration for the architecture and functioning of cooperating neural networks, in: *IWANN 1995*, 1995, pp. 24–30.
33. S. Stringer, T. Trappenberg, E. Rolls, I. Araujo, Self-organizing continuous attractor networks and path integration: one-dimensional models of head direction cells, *Network: Computation in Neural Systems* 13 (2) (2002) 217–242.
34. S. Stringer, E. Rolls, T. Trappenberg, Self-organizing continuous attractor networks with multiple activity packets, and the representation of space, *Neural networks* 17 (2004) 5–27.
35. C. Weber, S. Wermeter, A self-organizing map of sigma-pi units, *Neurocomputing* 70 (2007) 2552–2560.

36. K. Zhang, Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory, *Journal of Neuroscience* 16 (6) (1996) 2112–26.
37. A. Pouget, T. Sejnowski, Spatial transformations in the parietal cortex using basis functions, *Journal of Cognitive Neuroscience* 9 (1997) 222–237.
38. R. Andersen, G. Essick, R. Siegel, Encoding of spatial location by posterior parietal neurons., *Science* 230 (4724) (1985) 456–8.
39. E. Salinas, P. Thier, Gain modulation: a major computational principle of the central nervous system., *Neuron* 27 (1) (2000) 15–21.
40. G. Tononi, O. Sporns, G. Edelman, Reentry and the problem of integrating multiple cortical areas: simulation of dynamic integration in the visual system., *Cereb Cortex* 2 (4) (1992) 310–35.
41. F. Hamker, The reentry hypothesis: the putative interaction of the frontal eye field, ventrolateral prefrontal cortex, and areas V4, IT for attention and eye movement., *Cerebral Cortex* 15 (4) (2005) 431–47.
42. G. Deco, T. Lee, A unified model of spatial and object attention based on inter-cortical biased competition, *Neurocomputing* 44-46 (2002) 775–781.
43. F. Hamker, The emergence of attention by population-based inference and its role in distributed processing and cognitive control of vision, *Computer Vision and Image Understanding* 100 (2005) 64–106.
44. S. Frintrop, VOCUS: A Visual Attention System for Object Detection and Goal-directed Search, Vol. 3899 of Lecture Notes in Computer Science, Springer-Verlag, 2006.
45. M. Riesenhuber, T. Poggio, Hierarchical models of object recognition in cortex., *Nature Neuroscience* 2 (11) (1999) 1019–25.
46. J. Moran, R. Desimone, Selective attention gates visual processing in the extrastriate cortex., *Science* 229 (4715) (1985) 782–4.
47. R. Desimone, J. Duncan, Neural mechanisms of selective visual attention., *Annual Review Neurosciences* 18 (1995) 193–222.
48. J. Reynolds, L. Chelazzi, R. Desimone, Competitive mechanisms subserve attention in macaque areas v2 and v4., *Journal of Neuroscience* 19 (5) (1999) 1736–53.
49. S. Shipp, The brain circuitry of attention., *Trends Cogn Sci* 8 (5) (2004) 223–30.
50. J. Lynch, J.-R. Tian, Cortico-cortical networks and cortico-subcortical loops for the higher control of eye movements., *Progress Brain Research* 151 (2005) 461–501.
51. O. Hikosaka, Y. Takikawa, R. Kawagoe, Role of the basal ganglia in the control of purposive saccadic eye movements, *Physiological Review* 80 (3) (2000) 953–78.
52. D. Robinson, S. Petersen, The pulvinar and visual salience., *Trends Neuroscience* 15 (4) (1992) 127–32.
53. L. Zhaoping, A saliency map in primary visual cortex., *Trends Cognitive Sciences* 6 (1) (2002) 9–16.
54. J. Gottlieb, M. Kusunoki, M. Goldberg, The representation of visual salience in monkey parietal cortex., *Nature* 391 (6666) (1998) 481–4.
55. K. Thompson, N. Bichot, A visual salience map in the primate frontal eye field., *Progress Brain Research* 147 (2005) 251–62.
56. K. Rockland, G. VanHoesen, Direct temporal-occipital feedback connections to striate cortex (V1) in the macaque monkey, *Cerebral Cortex* 4 (3) (1994) 300–13.
57. T. Moore, K. Armstrong, Selective gating of visual signals by microstimulation of frontal cortex., *Nature* 421 (6921) (2003) 370–3.
58. O. Hikosaka, Basal ganglia mechanisms of reward-oriented eye movement., *Annals New York Academy Sciences* 1104 (2007) 229–49.
59. S. Funahashi, C. Bruce, P. Goldman-Rakic, Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex., *Journal of Neurophysiology* 61 (2) (1989) 331–49.
60. C. Constantinidis, X. Wang, A neural circuit basis for spatial working memory., *Neuroscientist* 10 (6) (2004) 553–65.
61. Y. Watanabe, S. Funahashi, Neuronal activity throughout the primate mediodorsal nucleus of the thalamus during oculomotor delayed-responses. I. cue-, delay-, and response-period activity, *J Neurophysiol* 92 (3) (2004) 1738–55.
62. Y. Watanabe, S. Funahashi, Neuronal activity throughout the primate mediodorsal nucleus of the thalamus during oculomotor delayed-responses. II. activity encoding visual versus motor signal, *J Neurophysiol* 92 (3) (2004) 1756–69.
63. M. Sommer, R. Wurtz, Influence of the thalamus on spatial visual processing in frontal cortex, *Nature* 444 (7117) (2006) 374–7.
64. J. Duhamel, C. Colby, M. Goldberg, The updating of the representation of visual space in parietal cortex by intended eye movements, *Science* 255 (5040) (1992) 90–2.
65. C. Quaia, L. Optican, M. Goldberg, The maintenance of spatial accuracy by the perisaccadic remapping of visual receptive fields, *Neural Netw* 11 (7-8) (1998) 1229–1240.
66. G. Rizzolatti, L. Riggio, I. Dascola, C. Umiltà, Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention., *Neuropsychologia* 25 (1A) (1987) 31–40.

5 Model's parameters

In the following, we denote $d \in \{blue, green, 45^\circ, 135^\circ\}$ a considered feature. For each feature d , we denote \bar{d} the antagonist feature of d (e.g. $green = \bar{blue}$). This dimension is relevant for the *Visual input*, *Sensory pole* and *Feature processing* modules only, as the *Spatial processing* module is not selective to visual features other than spatial position. When required, the activities of the units are superscripted by a unique map name and subscripted by a spatial position :

- Visual input : $u_{i,j}^{I,d}$
- Sensory pole : $u_{i,j}^{V4,d}$
- Target : $u^{PF,d}$
- Perceived features : $u^{IT,d}$
- Move : u^{mv}
- Switch : u^{sw}
- Saliency : $u_{i,j}^{sal}$
- Focus : $u_{i,j}^{focus}$
- Working memory : $u_{i,j}^{wm}$
- Anticipation : $u_{i,j}^{ant}$

In order to clarify the equations, we remove the map name superscript when the context is clear enough. In the following, we denote $\|(i, j), (k, l)\|$ the Euclidean distance between the positions (i, j) and (k, l) : $\|(i, j), (k, l)\| = \sqrt{(i-k)^2 + (j-l)^2}$. The activities $u_{i,j}$ of a unit at position (i, j) is updated using equation 4.

$$u_{i,j}(t + \Delta t) = f(u_{i,j}(t) + \Delta u_{i,j}(t)) \quad (4)$$

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

with $\Delta u_{i,j}(t)$ specific to each neural field and given below.

5.1 Sensory pole

The *Sensory pole* is made of four maps of 40x40 units. If we denote $d \in \{blue, green, 45^\circ, 135^\circ\}$ the considered feature, the activity $u_{i,j}^d(t)$ evolves according to the equation 5 :

$$\tau \Delta u_{i,j}^d(t) = -u_{i,j}^d(t) + (u_{i,j}^{I,d}(t) \cdot (0.25 + 0.15 \cdot u_{i,j}^{PF,d} + 0.5 \cdot u_{i,j}^{Focus})) \quad (5)$$

with $\tau = 0.75$.

5.2 Spatial processing

Saliency The *Saliency* map consists of 40x40 units. The activities $u_{i,j}(t)$ of the units evolve according to equation 6

$$\tau \Delta u_{i,j}(t) = -u_{i,j}(t) + \frac{1}{\alpha} (\max_d u_{i,j}^{V4,d}(t) + h) \quad (6)$$

with $\tau = 2.0, \alpha = 0.5, h = -0.1$.

Focus The *Focus* consists of 40x40 units. The activities $u_{i,j}(t)$ of the units evolve according to equation 7. The inhibitory bias provided by the *Working memory* avoids redeploying spatial attention on previously attended locations, after a saccadic eye movement has been executed.

$$\begin{aligned} \tau \Delta u_{i,j}(t) = & -u_{i,j}(t) + \frac{1}{\alpha} \left(\sum_{k,l} w^{foc}(|(i,j),(k,l)|) u_{k,l}^{foc}(t) \right. \\ & + \sum_{k,l} w^{sal}(|(i,j),(k,l)|) u_{k,l}^{sal}(t) \\ & - 4.0 \cdot u^{sw}(t) \cdot u_{i,j}^{wm}(t) \\ & \left. + \sum_{k,l} w^{wm}(|(i,j),(k,l)|) u_{k,l}^{wm}(t) \right) \quad (7) \end{aligned}$$

with $\tau = 7.0, \alpha = 4.0$ and :

$$w^{foc}(x) = \exp\left(-\frac{x^2}{25.0}\right) - 0.65 \exp\left(-\frac{x^2}{2n^2}\right)$$

$$w^{sal}(x) = 0.4 \cdot \exp\left(-\frac{x^2}{4.0}\right)$$

$$w^{wm}(x) = -0.10 \cdot \exp\left(-\frac{x^2}{4.0}\right)$$

Working memory The *Working memory* consists of 40x40 units. The activities $u_{i,j}(t)$ of the units evolve according to equation 8 :

$$\begin{aligned} \tau \Delta u_{i,j}(t) = & -u_{i,j}(t) + \frac{1}{\alpha} \left(\sum_{k,l} w^{wm}(|(i,j),(k,l)|) u_{k,l}(t) \right. \\ & + \sum_{k,l} w^{sal}(|(i,j),(k,l)|) u_{k,l}^{sal}(t) \\ & + \sum_{k,l} w^{foc}(|(i,j),(k,l)|) u_{k,l}^{foc}(t) \\ & + \sum_{k,l} w^{ant}(|(i,j),(k,l)|) u_{k,l}^{ant}(t) \\ & \left. + h \right) \quad (8) \end{aligned}$$

with $\tau = 0.6, \alpha = 13.0, h = -0.2$ and :

$$w^{wm}(x) = 3.0 \cdot \exp\left(-\frac{x^2}{4.0}\right) - 0.5 \exp\left(-\frac{x^2}{16.0}\right)$$

$$w^{sal}(x) = 0.3 \cdot \exp\left(-\frac{x^2}{4.0}\right)$$

$$w^{foc}(x) = 0.2 \cdot \exp\left(-\frac{x^2}{4.0}\right)$$

$$w^{ant}(x) = 0.3 \cdot \exp\left(-\frac{x^2}{4.0}\right)$$

Anticipation The *Anticipation* map consists of 40x40 units. These units integrate their input provided by the *Working memory* and *Focus* maps by a weighted sum of the product of the activities of one unit from each map. Namely, the activities $u_{i,j}(t)$ of the units evolve according to equation 9 :

$$\tau \Delta u_{i,j}(t) = -u_{i,j}(t) + 0.01 \sum_{k,l} u_{i+k,j+l}^{wm}(t) \cdot u_{n/2-k,n/2-l}^{foc}(t)$$

with $\tau = 4.0$.

5.3 Feature processing

Perceived features The *Feature processing* map is made of four units, one per feature. The activity $u^d(t)$ of each unit evolves according to equation 9 :

$$\begin{aligned} \tau \Delta u^d(t) = & -u^d(t) + \frac{1}{\alpha} \left(\max_{k,l} u_{k,l}^{V4,d}(t) \right. \\ & + 0.6 \cdot u^d(t) \\ & \left. - 0.6 \cdot u^{\bar{d}}(t) \right) \end{aligned}$$

with $\tau = 0.75, \alpha = 1.5$.

Target The *Target* map is made of four units, one per feature. The activity of these units is clamped manually to define the target of the visual search task.

5.4 Motor pole

Move The activity of the *Move* unit reflects when the features of the stimulus below the focus of attention has all the features of the target. It then has to detect a match between the perceived features and the target's features. The activity of this unit therefore evolves according to equation 9:

$$\tau \Delta u(t) = -u(t) + \alpha \sum_d u^{PF,d}(t) \cdot (u^{T,d}(t) - u^{T,\bar{d}}(t)) \quad (9)$$

with $\tau = 0.75$. We set $\alpha = 0.5$ when the target is defined by a single feature (e.g. color) and $\alpha = 1.0$ when the target is defined by two features (color and orientation).

Switch The *Switch* unit which modulates inhibitory projections between the *Working memory* and the *Focus* map to disengage spatial attention, has to detect a mismatch between the perceived features and the target's features. Therefore, the activity $u(t)$ of this unit evolves according to equation 10

$$\tau \Delta u(t) = -u(t) + \sum_d u^{PF,d}(t) \cdot u^{T,\bar{d}}(t) \quad (10)$$

with $\tau = 0.75$.

6 Interfacing the model with the virtual environment

The model is embedded in a virtual environment written in OpenGL™. Two interfaces are considered : the extraction of the visual features feeding the input maps of the model and the decoding of the motor command to execute the saccadic eye movement. The extraction of the visual features (two color filters and two orientation filters) are performed with classical HSV filter and Sobel filters.

The parameters of the saccadic movement to execute are determined by first extracting the center of mass of the activities within the *Focus* map (see fig. 2) using equation 11 :

$$\begin{aligned} \Delta x = & \frac{\sum_{i,j} (i - n/2) \cdot u_{i,j}}{\sum_{i,j} u_{i,j}} \\ \Delta y = & \frac{\sum_{i,j} (j - n/2) \cdot u_{i,j}}{\sum_{i,j} u_{i,j}} \quad (11) \end{aligned}$$

Given the previous eye movement is defined in the neuronal space, we need to translate it in the physical space. If we denote (\hat{h}, \hat{v}) respectively the horizontal and vertical field of view angles (in degrees), the horizontal and vertical eye movements $(\Delta h, \Delta v)$ are computed as :

$$\Delta h = \tan^{-1}\left(\Delta x \cdot \frac{\tan(\hat{h}/2)}{n/2}\right)$$

$$\Delta v = \tan^{-1}\left(\Delta y \cdot \frac{\tan(\hat{v}/2)}{n/2}\right)$$