Effects of feature-based attention: simulation of an fMRI experiment.

Silvia Corchs and Gustavo Deco

Siemens AG, Corporate Technology, CT IC 4, Otto-Hahn-Ring 6, 81739 Munich, Germany

silvia.corchs@mchp.siemens.de   gustavo.deco@mchp.siemens.de

(February 10, 2003)

Abstract

We use a computational neuroscience approach to study the role of feature-based attention in visual perception. This model is used to numerically simulate a visual attention experiment. The neurodynamical system consists of many interconnected modules that can be related to the dorsal and ventral paths of the visual cortex. The biased competition hypothesis is taken into account within the model. From the experimental point of view, measurements exist which confirm that feature-based attention influences visual cortical responses to stimuli outside the attended location. These measurements show that attention to a stimulus feature (in the present case "color") increases the response of cortical visual areas to a spatially distant, ignored stimulus that shares the same feature. Our neurodynamical model is used to numerically compute the neural activity of area V4 corresponding to such ignored stimulus, giving a good description of the experimental data.

Keywords: visual attention, neural models, fMRI, biased competition

I. INTRODUCTION

Visual attention can function in two distinct modes: spatial focal attention and object or feature attention. Both modes of operation are manifestations of a top-down selection process. In spatial attention, the selection is focused in the spatial dimension and spread in the feature dimension; while in object attention, the selection is focused in the feature dimension and spread in the spatial dimension.

Observations from numerous cognitive neuroscience experiments provide insights into the neural basis of spatial attention and feature attention. There is evidence for different attentional modulation effects in the two "what-where" streams produced during the performance of object or feature recognition as well as spatial tasks.

Regarding spatial attention, it is known that objects presented at an attended location are discriminated more accurately and produce stronger blood flow response in ventral regions than targets presented at unattended locations. For example, fMRI studies in humans[1] show that directing attention to a stimulus counteracts the suppressive influence of nearby stimuli.

On the other hand, and regarding object or feature attention, recent fMRI studies in humans [2] indicate that feature-based attention can globally influence visual cortical responses to stimuli outside the attended location: attention to a stimulus feature (such as color) increases the response of cortical visual areas to a spatially distant, ignored stimulus that shared the same feature.

The experimental observations made with single-cell recording and with functional imaging provide strong evidence for the biased competition hypothesis [3,4]. This hypothesis says that attention modulates visual processing by enhancing the responses of the neurons representing the features or locations of the attended stimulus and by reducing the suppressive interactions from neurons representing nearby distractors. This attentional effect is produced by generating signals within areas outside the visual cortex which are then fed back to extrastriate areas, where they bias the competition such that when multiple stimuli appear

in the visual field, the cells representing the attended stimulus "win", thereby suppressing cells representing distracting stimuli. In this way, attention appears as an emergent property of competitive interactions that work in parallel across the visual field.

A neurodynamical model that takes into account both the segregation of the visual system into dorsal and ventral streams and the biased competition hypothesis has been formulated by Deco and Zilhl[5] (see also Rolls and Deco [6]). The neurodynamical model consists essentially of several interconnected network modules which can be related to the different areas of the dorsal and ventral path of the visual cortex. Each of the modules consists of a population of cortical neurons, where the temporal evolution of the system is described within the framework of a mean-field-approximation; i.e. an ensemble average of the neural population is calculated in order to obtain the corresponding activity.

Using a simplified version of the above mentioned model, we now propose to simulate the fMRI data of Saenz *et al.* [2] that is concerned with feature attention.

In the following sections we first summarize the experimental results Saenz *et al.* and then we present the neurodynamical model and apply it to simulate the measurements.

II. FMRI DATA

Saenz *et al.* [2] used fMRI in humans to test whether feature-based attention can globally influence visual cortical responses to stimuli outside the attended location. Their fMRI results demonstrated spatially global neuronal modulation due to feature-based attention across multiple early stages of cortical visual processing. Attention to a stimulus feature (color or direction of motion) increased the response of cortical visual areas to a spatially distant, ignored stimulus that shared the same feature. The authors asked subjects to attend to one color (red) within two overlapping fields of stationary red and green dots on one side of a central fixation point. Subjects were asked to ignore a single field of red or green dots (Fig. 1a). Subjects were given a two-interval, forced-choice discrimination task at luminance threshold. They were instructed to attend to the red or green field of

dots (the target field) on the attended side in alternating blocks. The color of the ignored field of dots did not change during each scan. Each four minute fMRI scan consisted of six 40-s cycles in which attention alternated between the "same" condition, in which the color of the ignored stimulus matched the target field and the "different" condition, in which the color of the ignored stimulus was different from that of the target. During each trial, stimuli were presented for two sequential 1-s intervals separated by a 100-ms interval in which only the fixation point was present. Trials started every 3.3 s. The authors measured the fMRI response to the ignored stimulus in areas V1, V2, V3, V3A, V4 and MT+. They observed that response amplitudes to the ignored stimulus were stronger during the "same" condition, when its color matched the attended color. This means that the fMRI response to the ignored stimulus was modulated by feature-based attention in those areas. The results measured for area V4 are shown in Fig. 1b.

III. THEORY

A. Mathematical formulation of the model

Deco and Zihl ([5]; see also Rolls and Deco [6]) have developed a model to describe the cortical neurodynamics of visual attention. We will use a simpler version of this model in order to simulate the experimental data. Our system, shown in Fig. 2, is composed of four modules: 1) a visual cortex module (V4), 2) an inferior temporal cortex module (IT), 3) a posterior parietal module (PP) and 4) a prefrontal cortex module (PFC). They are structured such that they resemble the two known main visual paths of the mammalian visual cortex: the *what* and *where* paths.

Information from the retino-geniculo-striate pathway enters the visual cortex through areas V1-V4 in the occipital lobe and proceeds into two processing streams. The occipital-temporal stream (*what* pathway) leads ventrally through IT (inferotemporal cortex) and the occipito-parietal stream (*where* pathway) leads dorsally into PP (posterior parietal). The

4

what pathway is mainly concerned with object recognition, independently of position and scaling. The *where* pathway is mainly concerned with the location of objects and the spatial relationships between objects.

In our present model, the ventral stream consists of three modules: 1) a V4 module, 2) an IT module, 3) a module v46 corresponding to the ventral area 46 of the prefrontal cortex, which generates the target object in a visual search task.

The module V4 inputs spatial and feature information up to the dorsal and ventral streams. It contains PxP hypercolumns of neurons that cover a NxN pixel scene. Each hypercolumn contains two columns corresponding to colors *red* and *green*. There is one inhibitory pool interacting with each of the elements of the hypercolumns.

The module IT is concerned with the recognition of objects and consists of pools of neurons which are sensitive to the presence of a specific object in the visual field. In our present case the module IT will be taken into account only through the top-down bias corresponding to the condition of attention to a given color (red or green).

There is one inhibitory pool interacting with each one of pools. The inhibitory pool integrates information from all the excitatory pools within the module and feedbacks unspecific inhibition uniformly to each of the excitatory pools. It mediates normalizing lateral inhibition or competitive interactions among the excitatory cell pools within the module.

The dorsal stream consists of three modules: 1) the V4 module, 2) a PP module, 3) a module d46. The module PP consists of pools coding the position of the stimuli. It is responsible for mediating spatial attention modulation and for updating the spatial position of the attended object. A lattice of N x N nodes represents the topographical organization of the module PP. Each node on the lattice corresponds to the spatial position of each pixel in the input image. The module d46 corresponds to the dorsal area 46 of the prefrontal cortex that generates the attentional bias for spatial location.

Prefrontal cortex area 46 (modules d46 and v46) is not explicitly simulated in the current version of the model, but is part of the model. It is assumed that top-down feedback connections from these areas provide the external top-down bias that specifies the processing

conditions of earlier modules. Concretely, the feedback connection from area v46 with the IT module specifies the target object in a visual search task; and the feedback connection from area d46 with the PP module generates the bias to a targeted spatial location in a recognition task at a fixed prespecified location.

The system operates in two different modes: the learning mode and the recognition mode. During the learning mode the synaptic connections between V4 and IT are trained by means of Hebbian learning. In the present work we will be interested only with the recognition mode. In the recognition mode, there are two possibilities of running the system. First, an object can be localized in a scene (visual search) by biasing the system with an external top-down component at the IT module which drives the competition in favor of the pool associated with the specific object to be searched. Then, the intermodular attentional modulation V4-IT will enhance the activity of the pools in V4 associated with the features of the specific object to be searched. Finally, the intermodular attentional modulation V4-PP will drive the competition in favor of the pool localizing the specific object. Second, an object can be identified (object recognition) at a specific spatial location by biasing the system with an external top-down component at the PP module. This drives the competition in favor of the pool associated with the specific location such that the intermodular attentional modulation V4-PP will favor the pools in V4 associated with the features of the object at that location. Intermodular attentional modulation V4-IT will favor the pool that recognized the object at that location.

Each pool of neurons will be described within the mean field approximation [7] which consists of replacing the temporal averaged discharge rate of a cell with an equivalent activity of a neural population (ensemble average). After solving the system of differential equations, the activity corresponding to module V4 is obtained and it is then integrated over space. In order to compare our theoretical calculations with the fMRI data we should consider that the hemodynamic response to sensory input (evoked changes in neuronal activity) is transient, delayed and dispersed in time. Friston *et al.* [8] assumed a Poisson distribution function which was later found to be consistent with measurements in human visual area

6

V1 (Boynton *et al.* [9]). Following Friston and colleagues and Horwitz *et al.*[10] we will also use a Poisson function to represent the hemodynamic delay and convolve this with our simulated neural activity.

<p align="center">**B. Mathematical Formulation**</p>

Let us denote by I_{pql}^{V4} the sensory input activity to a neuronal pool A_{pql}^{V4} in V4 which is sensitive to a determined color l ($l=0,1$ corresponds to *red* and *green* respectively) and to stimuli with receptive fields at locations specified by the indices pq. This sensory input is given by

$$I_{pql=0}^{V4} = Is \quad I_{pql=1}^{V4} = 0 \quad presence\ of\ red\ in\ receptive\ field\ position\ pq \tag{1}$$

$$I_{pql=0}^{V4} = 0 \quad I_{pql=1}^{V4} = Is \quad presence\ of\ green\ in\ receptive\ field\ position\ pq \tag{2}$$

$$I_{pql=0}^{V4} = 0 \quad I_{pql=1}^{V4} = 0 \quad nor\ red\ nor\ green\ at\ receptive\ field\ position\ pq \tag{3}$$

where the distribution of the input I_{pql}^{V4} over the indices pq is done with a random function and the constant Is is a parameter to be adjusted in the calculations.

Let us now define the neurodynamical equations that regulate the evolution of the whole system. The evolution of activity level A_{pql}^{V4} is given by

$$\tau \frac{\partial}{\partial t} A_{pql}^{V4}(t) = -A_{pql}^{V4}(t) + a\ F(A_{pql}^{V4}(t)) - bF(I^{V4,I}(t)) + I_{pql}^{V4}(t)$$
$$+ c\ I_{pq}^{V4-PP}(t) + d\ I_{pql}^{V4-IT}(t) + I_0 + \nu \tag{4}$$

where I_{pq}^{V4-PP} and $I_{pql}^{V4-IT}(t)$ denote the input currents to the V4 module from the PP and IT modules respectively and the term $I^{V4,I}(t)$ reflects the operation of inhibitory neuronals pools defined below. The non linearity function F denotes the response function, transforming current into discharge rate. The parameters appearing in eq. 4 are taken as: $a = b = 0.95$, $c = 0.1$, $d = 1$; the diffuse spontaneous background input I_0 is taken equal to

<p align="center">7</p>

0.025 and the Gaussian noise $\nu = 0.01$. The choice of these parameters is uncritical and is based on biological parameters.

The attentional biasing due to the intermodular *where* connections with the pools in the PP module, I_{pq}^{V4-PP}, is given by

$$I_{pq}^{V4-PP} = \sum_{i,j} W_{pqij} F(A_{ij}^{PP}(t)) \tag{5}$$

The connections between pools in the V4 module and pools in the PP module are specified such that topographically corresponding centers (in PP and V4) are connected with maximal strength and the connections to neighbouring pools decay with a Gaussian function of distance. The connections are therefore defined by

$$W_{pqij} = \exp[-dist^2/2\sigma_{V4}^2] \tag{6}$$

where *dist* represents the distance from spatial localization (i, j) to the position of the receptive field (p, q).

The attentional biasing I_{kpql}^{V4-IT}, due to the intermodular *what* connections with the pools in the temporal module IT, is given by

$$I_{pql=0}^{V4-IT}(t) = 1 \quad I_{pql=1}^{V4-IT}(t) = 0 \tag{7}$$

for the condition of attention to color *red* and

$$I_{pql=0}^{V4-IT}(t) = 0 \quad I_{pql=1}^{V4-IT}(t) = 1 \tag{8}$$

for the condition of attention to color *green*.

Excitatory cell pools in each module are engaged in competition, mediated by an inhibitory pool which receives excitatory input from all the excitatory pools and provides uniform inhibitory feedback to each of the excitatory pools. The current activity of the inhibitory pool $I^{V4,I}$ obeys the following equation

$$\tau \frac{\partial}{\partial t} I^{V4,I}(t) = -I^{V4,I}(t) + \sum_{p,q,l} F(I_{pql}^{V4,I}(t)) - e\, F(I^{V4,I}(t)) \tag{9}$$

8

where the constant e is taken equal to 0.01. Again, this choice is uncritical and it is based on biological parameters.

Similarly, the current activity of the excitatory pools in the PP module evolves according to the following equation

$$\tau\frac{\partial}{\partial t}A_{ij}^{PP}(t) = -A_{ij}^{PP}(t) + a\,F(A_{ij}^{PP}(t)) - b\,F(I^{PP,I}(t)) +$$
$$+ I_{ij}^{PP-V4}(t) + I_{ij}^{PP,A} + I_0 + \nu \tag{10}$$

where $I_{ij}^{PP,A}$ is an external attentional spatial-specific top-down bias, the intermodular attentional biasing I_{ij}^{PP-V4} through the connections with the pools in the module V4 is

$$I_{ij}^{PP-V4}(t) = \sum_{p,q,l} W_{pqij}F(I_{pql}^{V4}(t)) \tag{11}$$

and the activity current of the common PP inhibitory pool evolves according to

$$\tau\frac{\partial}{\partial t}I^{PP,I}(t) = -I^{PP,I}(t) + \sum_{i,j} F(I_{ij}^{PP}(t)) - e\,F(I^{PP,I}(t)) \tag{12}$$

The integrated V4 activity over space at a given time t is given by

$$A^{V4}(t) = \sum_{p,q,l} I_{pql}^{V4}(t) \tag{13}$$

In order to compare with the experimental data of Saenz $et\ al.$ and to deal with the problem of hemodynamic delay we use a Poisson function to represent such delay and convolve this with our simulated neural activity [8,10].

$$f^{V4}(t) = \int A^{V4}(t - \tau)h(\tau)d\tau \tag{14}$$

where $h(\tau)$ is a Poisson distribution with delay parameter equal to 6.5 sec.

IV. RESULTS AND DISCUSSIONS

Following the conditions of Saenz $et\ al.$' work, we simulate their experiment as follows.

Within each trial the stimulus is shown during 1s, followed by a 100-ms interval where there is no input. Then, the same stimulus (but with a greater luminance that in our case

corresponds to a greater constant value Is) is shown followed by an interval of 1.2 s where there is no input. This procedure is repeated six times, what equals to an interval of 20 s. Another period of 20 s is followed, but now with a different condition of feature-attention. Two different values for the constant Is were used in the calculations: $Is = 0.07$ and $Is = 0.08$ to represent the two conditions of luminance of the experiment.

Two attention conditions are taken into account in our model. On one side, the spatial attention that corresponds to the top-down bias to module PP and it is representing the fact that the subject is asked to attend to the left field of Fig. 1a. If we consider an input image of 64 x 64 pixels, this corresponds to taking $I_{ij}^{PP,A}$ equal to 0.07 for all values of indices $i < 32$ and zero otherwise. On the other side, we have the top-down bias to module IT. Every 20 s the condition of attention is changing from *attention to red* ($I_{pql=0}^{V4-IT} = 0.02$ and $I_{pql=1}^{V4-IT} = 0$) to *attention to green* ($I_{pql=0}^{V4-IT} = 0$ and $I_{pql=1}^{V4-IT} = 0.02$).

We show our numerical results in Fig. 3. The units are arbitrary and we are only interested in describing the qualitative behavior of the neural activity. As in the experiments, our results show that the response activity to the ignored stimulus was modulated by feature-based attention to color. Our numerical simulations show a deflection in each peak that is attributed to the values of the parameters we have used in the calculations. As we have already stated, our goal is to *qualitatively* describe the global behaviour of the fMRI data. The amplitudes to the ignored stimulus are stronger during the *same* condition, when its color matched the attended color. Our theoretical results as well as the experimental data confirm the hypothesis that feature-based attention can globally influence visual cortical responses to stimuli outside the attended location. The experimental results are consistent with our neurodynamical model consisting of interconnected populations of cortical neurons distributed in different brain modules.

In our previous work [11] we have shown that our cortical architecture at the neurophysiological level of brain area activity shows the typical *spatial attentional* modulation effects. In the present paper we have also demonstrated that the same conclusion is achieved with respect to *feature-based attentional* modulation effects. The numerical results obtained from

our neurodynamical model are thus consistent with the two different kinds of fMRI measurements: those of Kastner *et al.* [1] and the corresponding ones to Saenz *et al.* [2]. Within our model, the two modes of attention emerge depending simply on whether a top-down bias is introduced to either the dorsal PP module (as in the experiment of Kastner *et al.*) or the ventral IT module (as in the experiment of Saenz *et al.*). The two apparently different processes such as spatial attention versus feature or object attention can be accounted for by a unitary system. In summary, these results demonstrate that our cortical attentional architecture shows and explains the typical dynamical competition and attention modulating effects found in attention experiments at the level of brain area activation as measured with fMRI.

V. REFERENCES

[1] S. Kastner, M. Pinsk, P. De Weerd, R. Desimone and L. Ungerleider, Increased activity in human visual cortex during directed attention in the absence of visual stimulation, *Neuron* **22** (1999) 751-761.

[2] M. Saenz, G.T. Buracas and G.M. Boynton, Global effects of feature based attention in human visual cortex, *Nature Neuroscience* **5** (2002) 631-632.

[3] J. Duncan and G. Humphreys, Visual search and stimulus similarity, Psychological Review **96** (1989) 433-458.

[4] R. Desimone, M. Wessinger, L. Thomas and W. Schneider, Attentional control of visual perception: cortical and subcortical mechanisms, *Cold Spring Harbor Symp. Quant. Biol.* **55** (1990) 963-971.

[5] G. Deco and J. Zihl, Top-down selective visual attention: a neurodynamical approach, *Visual Cognition* **8** (2001) 118-139.

[6] T. Rolls, and G. Deco, Computational Neuroscience of Vision. *Oxford University Press Inc. New York* (2002).

[7] H. Wilson and S. Cowan, Excitatory and inhibitory interactions in localised popula-

tions of model neurons, *Biological Cybernetics* **12** (1972) 1-24.

[8] K.J. Friston,P. Jezzard and R. Turner, Analysis of functional MRI time-series, *Human Brain Mapp* **1** (1994) 153-171.

[9] G.M. Boynton, S.A. Engel, G.H. Glover and D.J. Heeger, Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.* **16** (1996) 4207-4221.

[10] B. Horwitz, F.T. Husain, A.R. Braun and M-A. Tagamets, Simulating PET/fMRI studies of Visual and Auditory Pattern Recognition Using Biologically Realistic Large-Scale Neural Models, *Proceedings of the International Joint Conference on Neural Networks 2001* (2001) 878-883.

[11] S. Corchs, and G. Deco, Large-scale Neural Model for Visual Attention: Integration of Experimental Single-cell and fMRI Data, *Cerebral Cortex* **12** (2002) 339-348.

FIGURE CAPTIONS

FIGURE 1: (a) Stimuli used in the experiments. Stimuli are circular apertures of stationary red and green random dotes in the upper visual field. R or G, field of red or green dots on gray background. (b) fMRI time series measured in response to ignored stimulus for visual cortex V4. From Saenz *et al.*[2].

FIGURE 2: Architecture of the neurodynamical approach. The system is essentially composed of four modules structured such that they resemble the two known main visual paths of the visual cortex.

FIGURE 3: Our computer simulations of fMRI signals in visual cortex V4 based on the experiments of Saenz *et al.*