Knowing What and Where: A Computational Model for Visual Attention

Kaustubh Chokshi*, Christo Panchev, Stefan Wermter
Centre for Hybrid Intelligent Systems
University of Sunderland
Sunderland, SR6 0DD. United Kingdom.
E-mail:*kaustubh.chokshi@sunderland.ac.uk

John G. Taylor
Department of Mathematics
Kings College London
London, WC2R 2LS. United Kingdom.
E-mail: john.g.taylor@kcl.ac.uk

***Abstract*— We describe a model of invariant object recognition in the brain that incorporates feedback biasing effects of top-down attentional mechanism on a hierarchically organised set of visual cortical areas. The model displays a space based and a object based visual search by using a top-down attention feedback model from posterior parietal modules and interaction between the two processing streams dorsal and ventral.**

I. Introduction

Vision is a sufficiently complex problem that benefits from computational neuroscience approach that is closely linked to empirical neurophysiological investigations. A typical scene is far too complicated for the visual system to process it in all its detail in a single time step. Psychophysical experiments have found that when one searches for a target among distractors, the addition of homogeneous distractors typically increases the time needed to find the target, as does increasing the homogeneity of existing distractors. Thus, the visual system is limited in its capacity to process information [1]. However, it is equipped to overcome this constraint because it can direct this limited capacity channel to a location of the object of interest. Subjectively, attending to a given object reduces our awareness of other objects among the clutter. As a result, our ability to identify a given object maybe unaffected by the presence and number of distractors.

Visual attention is broadly divided into two parts, spatial and object-based attention. When an object is selected on the basis of information about its location in the visual field, the selection process is referred to as spatial attention. When the object is selected on basis of information about its identity (i.e. shape and colour) the selection process is referred to as object-based attention [2].

The model presented in this paper contains 6 areas such that they resemble two known visual paths of mammalian visual cortex [3], [4], [5]. Information from lateral geniculate nucleus (LGN) enters the visual cortex though V1 and proceeds into the two processing streams: ventral and dorsal. The ventral stream is responsible for object recognition invariant of positions and scaling. The dorsal stream is concerned with preservation of spatial information of the objects such as location. The model of spatial and object based attention, described in this paper incorporates interactions between the dorsal 'where' and ventral 'what' of the visual stream. It incorporates a feedback system, based on top-down (endogenous) goal-based competition. The assumptions made in this work are based on a number of previous models [1], [4], [5], [6], [7], [8].

II. Architecture

A. Basis of the model

The presented model uses goal-based competition between various objects representations, as it is based on related single cell observations in the anterior temporal cortex in monkeys performing saccade to object shapes matching previously presented ones among a set of distracters [6]. The bias may be bottom-up, such as brought about in exogenous attention control by sudden onset, or can arise from a top-down (endogenous) goal to observe some target feature such as specific coloured objects or to observe a particular object such as blue square in a spatial bias location. In this paper we are presenting a model that uses top-down attention approach.

In goal-based competition, the attention feedback biases this competition in favour of the attended stimuli. When the attention is directed to one of two stimuli, this should excite neurons activated by the attended stimuli and suppress the responses of the neurons activated by the ignored stimuli. As a result the neurons that respond to attended stimuli maintain higher mean firing rates compared to those responding to other stimuli.

fMRI studies show that when multiple stimuli are present simultaneously in the visual field, their cortical representations within the object recognition pathways interact in a competitive suppressive fashion [9], [10]. Further experiments conducted by Reynolds [1] show that attention can either increase or decrease neuronal responses depending on the change in response caused by addition of the ignored stimulus. It was seen that the biased competition model was observed in monkeys V4 and V2 neurons. According to Deco [3], attention has a greater influence for higher areas of visual cortex (IT,V4 and V2) compared to lower areas of visual cortex (V1). The goal-based competitive model presented in this paper is build in with accordance of existing hypotheses and constraints described in various attention model [3], [4], [6], [7].

B. Architecture

The attention model has 6 areas: LGN-V1,V4, LIP, IT and Spatial Goal and Object Goal (fig. 1). It was build using spiking neurons with active dendrites and dynamic synapses (ADDS IaF) [11], [12].

LIP is part of the dorsal stream and plays a key role in spatial attention related to saccadic eye movements. It receives inputs directly from LGN. and represents the location of an objects in the visual scene. Spatial attention is location specific and its effects are mediated by goal biassing inputs having effects via feedback into LIP. LIP is then further used to give spatial information to the ventral stream. The spatial attention is done with inhibition depending on the goal of the spatial attention. LIP plays a key role in control of spatial attention [13], [14]. When LIP neurons were studied they had at least one object in its receptive field. The response of the LIP neurons depended on the location of the target that matched the cue. The neurons did not respond when the cue identified the neurons outside their receptive field [5]. However when the identified target was in their receptive field the neuron started to respond after the presentation of the cue. There are evidences that show that the dorsal and the ventral streams interact with each other [15], [16], [5], [4].

There is evidence [17], [18], [14] indicating that the dorsal stream plays an important part in object-based attention, as the dorsal stream or LIP is referred in order to have the location of the objects. The ventral and dorsal stream work together, and V4 receives its inputs from V1 and LIP. V4 has shift invariant responses to various stimuli therefore each part of V4 has 3 sets of neurons that responds to shapes (fig. 1). This provides the information to the IT about the various objects. V4 can see all the objects, that are there in the visual field. The inputs from LIP would make the object neurons in the spatial attention field spike at a higher rate then neurons outside the object field.

There is evidence that at higher levels of the visual cortex spatial information about the object is lost [16]. In our model IT is where object attention takes place [19], [16]. LIP influences IT indirectly via V4. V4 gives inputs to IT about the entire object set in the visual field. There are nine neurons for IT one each for each figure. Here the attention is controlled by potentiation. The neuron spikes only when objects are received from V4 and when goal is met. IT will not spike unless the goal is set and it receives its inputs from V4.

III. Experiments and Results

Object based attention may have limiting effects in a overall visual scene, it proves to be very effective within a spatially attended region. Experiments have shown that the attention is first directed to the object's spatial region where the object is present, then it calls object based attention [20], [21]. Posner [21] shows that the attention is directed to a region of visual display where the target might appear [2]. These are the bases of various experiments presented in this paper.

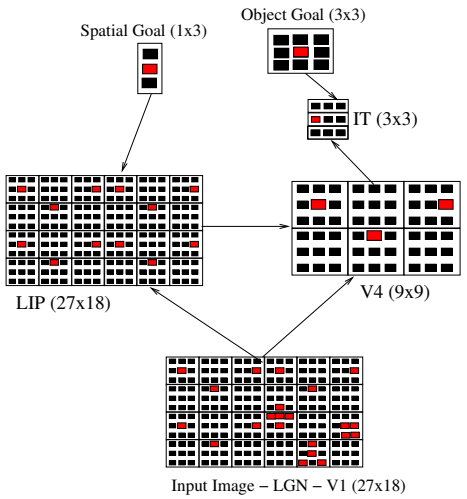


Fig. 1. A schematic representation of the neuron spike model.

A. Experimental Setup

For the experimental results presented in this paper, we had a set of nine different objects: 3 shapes (Square, Triangle and Circle) and 3 colours (Red, Green and Blue). All the experiments were done in MPGenesis Neural Simulator [22] and were performed on a Beowulf cluster of 56 nodes.

To investigate attention within our model, we had various conditions under which the model was tested. For the results that are presented in this paper, we have chosen to "attend" to four object in three spatial areas. At the bottom there are two objects a red square and green circle. At the middle of the visual field we have a green triangle and at the top right corner we have a blue circle. We also test our model, with a spatial field when an object is not presented.

The following section presents comprehensive results of the overall attention process in our model. This section begins with results a summary for attention under different conditions. The next sections deals with various other conditions under which the model was tested and further tested for parallel visual search and Posners paradigm test for the model.

B. Results

The model is presented with an image containing all four figures at 100ms (fig. 2). The response time for the model to the image is 115ms for LIP and V4. The input drives the neurons in LIP, V4 and IT to spike at 4Hz.

The experiment begins with no goals for the first second. Neurons in LIP and V4 respond to the objects in receptive fields with high firing rates. Then, at the 1sec. attention is turned on. Then, every second the attention goal is changed either spatial or object or both. The attention is first set to attend to the bottom of the visual scene and a red square, then to the middle and a green triangle and finally to top and a blue circle. We also have shown results of how model behaves when it is attending to a particular spatial region for an object and that the object is not present.

The results then further explain the parallel search capabilities of the model presented and we also have applied the results to the Posner's paradigm.

1) No attention: In figure 2, in the first second, there is no attentional goal and therefore even though LIP and V4 neurons are spiking at higher rates there is no response in IT. IT in our model is based on potentiation and needs input from both V4 and Object Goal.

2) Attending to bottom and red square (figure 2 1-2sec and figure 3): At 1sec we set the first goal (red square in the bottom area of the visual field). This inhibits the LIP neurons which are not in the attended areas (middle and top), while the firing frequency of the ones in the bottom area remains unaffected (fig 2 **(a)** and **(b)**). Since the LIP neurons influence V4, this effect is forwarded to that area. V4 neurons which represent objects in the attended area maintain their high firing rates. The neurons representing objects outside the attended area receive much lower input from LIP, and as a result their firing rates are reduced.

The effects of inhibition in terms of reduced firing rates are seen around 1.08sec for LIP and 1.12secs for V4. (fig. 2 **(a)**, **(b)**). We see responses in IT around 1.14 seconds (fig. 2**(c)**).

The goal here is to attend to a red square and at the bottom of input. We can see that there is a change in the firing rate of the neurons. The LIP neurons that are outside the spatial attention have their firing rate decreased and this can also be seen in V4. There is a response in IT also to the object that we are attending to. There is no response in other IT neurons as IT neurons require high frequency spikes from both V4 and the Object Goal. Neurons in V1 and LIP that are responding to the object at the bottom of the visual scene, are firing at a higher frequency than the neurons which are responsible for other objects present in other parts of the visual scene (fig. 3).

3) Attending to Top and Blue Circle (figure 2 sec 3-4 and figure 4): The goal was to attend to the top of the visual scene for a green circle. As observed in previous section III-B.2, as soon as there is a change in attention there is a significant difference in the firing rate. We can observe, as the attention is changed, and that there is a difference in firing rate of neurons. The neurons that have been firing at a high frequency for the previous goal, now fire at a lower frequency. This can be noted from the mean firing rate graphs (fig. 2b and 4c). When the attention changed, there is a delay between the change in the attention compared to when for the first time attention is set. This delay in the LIP neuron exists because neurons need time to recover from inhibition from the previous attention. The delay is a very small period of approx. 20ms. In our model, when spatial attention is in extreme corners, it can still inhibit the whole of LIP and reduce the firing rate of LIP and V4 neurons. This is an important factor for visual search. If the object is not found in the current spatial region, visual search would look at spatial regions with high levels of activations in V4 and LIP in order to attend to that spatial region and search for the object.

4) Attending to bottom for blue circle - not present (figure: 2: 4-5sec. and 5): We have seen how the our attention model behaves at neuronal levels when there is a object present in the spatial region that is being attended. For the final attention goal, we keep the object attention to the space where the desired object is not present in the visual area. Here we see that there is a high frequency of the neurons that are firing in V4 and LIP, but the IT does not respond.

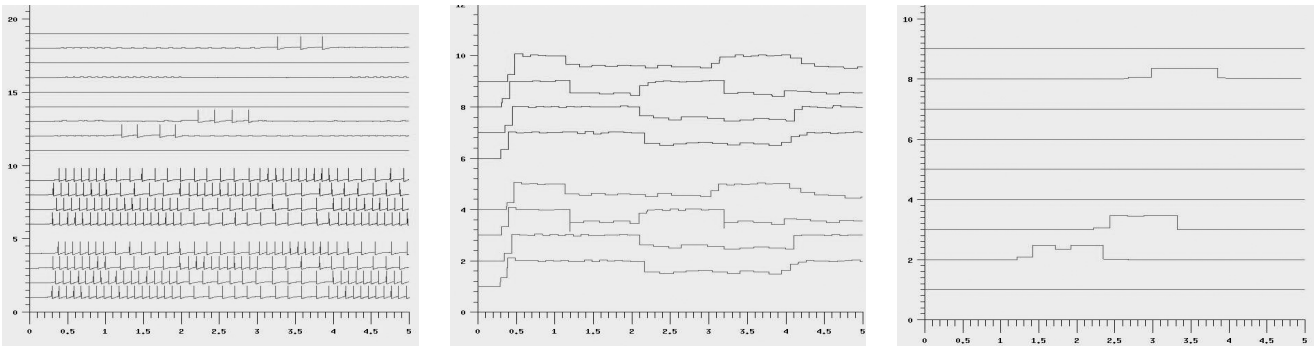
This is very interesting from a lot of perspectives. If the object goal is not found in the spatial region i.e. neurons in IT have not responded, the spatial attention goal should change in order to look for the object in different space. In other words, this would trigger a search task to find the object. It also helps in the serial search algorithm, that the bias of the region is inhibited with a decaying to return to zero [23]. Using this method it prevents the spatial attention from returning to the same spatial location for the same object.

5) Visual Search: Visual search is one of the main functions of attention. In visual search tasks, subjects are commonly asked to detect the presence or absence of a target display containing distractors elements. The response time is measured as a function of the number of elements in the display. The shape of this curve indicates something about how subjects perform the search tasks. Flat curves, in which response time does not increase with the number of elements is a suggestive of parallel search across the field. Curves with steep slopes, in which each additional element increases the response time [24], are suggestive of a serial search. For example, searching for a blue square among circles produces a flat response curve. Searching for a plus among horizontal and vertical distractors produces a positively sloped curve. Variety of promising computational models have been devised to replicate various aspects attentive visual search [23].

In this paper we are presenting a model of parallel search with our attention model. According to Deco and Zihl [15] it is plausible to build a neural system for visual search, which works across the visual field in parallel, due to the intrinsic dynamics of the attention system, visual attention can perform parallel search. Based on this, we demonstrate that it is plausible to perform visual search. Our visual attention can perform search across the visual field in parallel.

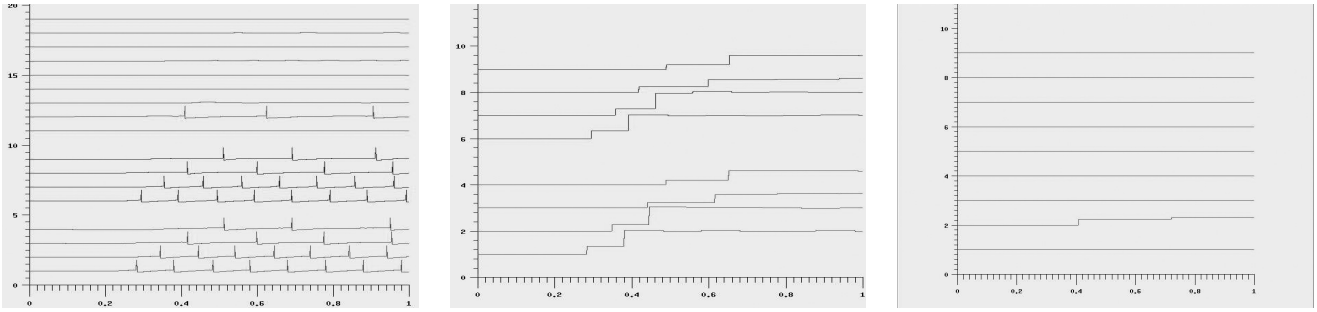
We assume that targets and distractors are known in advance i.e. all the display elements containing similar feature are discarded. In our case we search for a red square among the blue circle, green triangle and green square. Our search method is a parallel search method and therefore based on an endogenous attention method. With spatial attention present this enables the model to select a region. The selected region may contain one object or multiple objects. Neither explicit serial focal search nor saliency maps are used.

In our search task we have the target as a red square along with one to four distractors. Different objects were placed as distractors at random positions in the visual field. Searching for a element with distinctive features is easy [16], in our case these features are shapes and colour. Since the elementary features in the object chosen are distinct a red square would pop-out easily and can be localised independently without the number of other distracting objects. An attention network



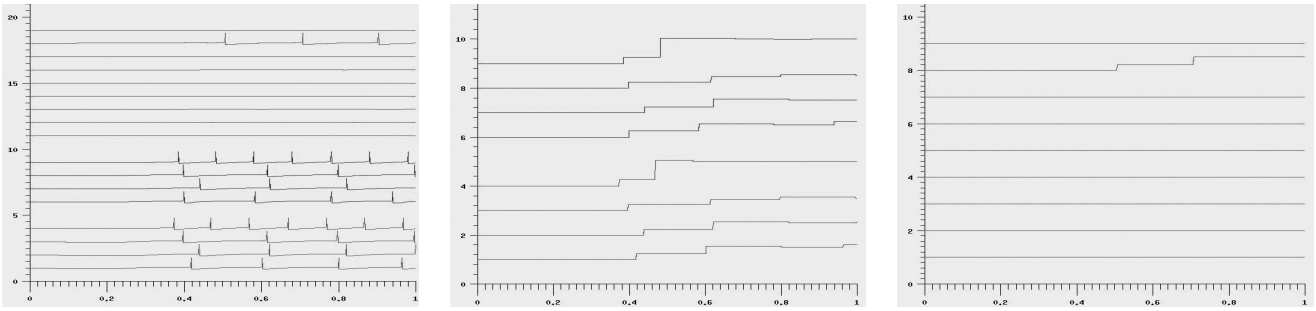
(a) **(b)** **(c)**

Fig. 2. **(a)** Summary of the attention experiments and its results. X-Axis represents time and Y-Axis represents neurons. 1-4 LIP Neurons, 6 to 9 V4 neurons the rest are the IT neurons. Of all the neurons these are the relevant neurons selected where the attention was taking place during experiments presented in this paper. **(b)**Mean Firing Rate of V4 and LIP neurons. Neurons from 1 to 4 are LIP neurons and 4 until 8 are V4 neurons. **(c)** Mean Firing Rate of IT neurons.



(a) **(b)** **(c)**

Fig. 3. This is neural behaviour of when the focus of attention is on the bottom towards the red square. **(a)** Firing rate of the neurons to attend to a red square at the bottom of the visual scene. X-Axis represents time and Y-Axis represents neurons. 1-4 LIP neurons, 6 to 9 V4 neurons, the rest are the IT neurons. **(b)** Mean Firing Rate of V4 and LIP neurons. Neurons from 1 to 4 are LIP neurons and 4 until 8 are V4 neurons. **(c)** Mean Firing Rate of IT neurons.



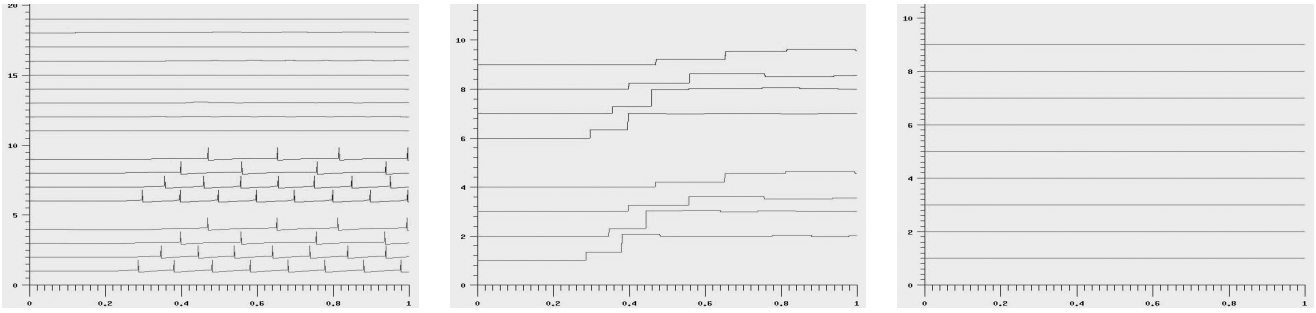
(a) **(b)** **(c)**

Fig. 4. **(a)** Summary of the attention experiments and its results. X-Axis represents time and Y-Axis represents neurons. 1-4 LIP Neurons, 6 to 9 V4 neurons, the rest are the IT neurons. **(b)** Mean Firing Rate of V4 and LIP neurons. Neurons from 1 to 4 are LIP neurons and 4 until 8 are V4 neurons. **(c)** Mean Firing Rate of IT neurons.

always acts to develop and select display elements, therefore in parallel search it is necessary to modulate over features to be selected otherwise, the target may be suppressed and may not be detected. In other words, response time is independent of the number of distractors. Our model shows similar results (fig. 6 and 7). In order for us to check parallel search in our model, we also presented all the four distractors and target at the same time. Similar results were observed.

With extensions to the model, a wide variety of other data can be addressed, including response-time curve for effects of target distractors homogeneity. The goal here is to present parallel visual search which our model performs.

6) Posner paradigm simulations: The Posner paradigm as described in [6], explores the manner in which attention is moved either exogenously or endogenously. The subject views a lighted screen on which the stimuli appears. They are



(a) **(b)** **(c)**

Fig. 5. The neuronal response when the correct object is not presented in the spatial attention field. **(a)**X-Axis represents time and Y-Axis represents neurons. 1-4 LIP Neurons, 6 to 9 V4 neurons, the rest are the IT neurons. **(b)** Mean Firing Rate of V4 and LIP neurons. Neurons from 1 to 4 are LIP neurons and 4 until 8 are V4 neurons. **(c)** Mean Firing Rate of IT neurons.

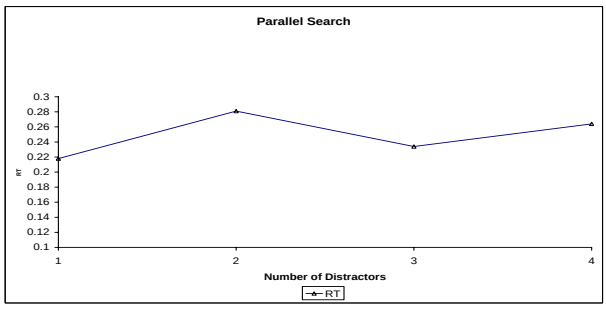
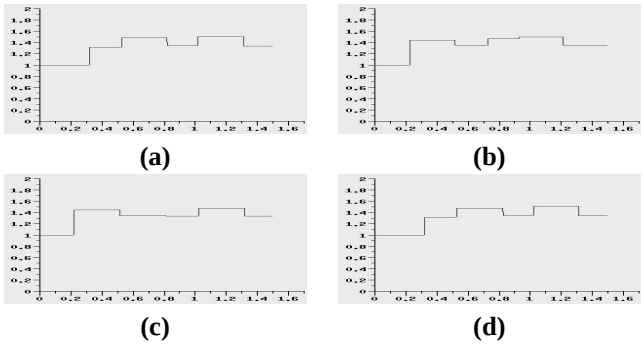


Fig. 6. Parallel search curve. The X-axis is the number of distractors and Y-axis is the response time of the IT neurons. The response time of the IT neurons fluctuates between 0.22sec to 0.28sec independent to the number of distractors in the visual field.



(a) **(b)** **(c)** **(d)**

Fig. 7. Here is the response of IT neurons for the objects being searched. In **(a)** Response of IT neuron with a single distractors. **(b)**Response of IT neuron with a two distractors. **(c)** Response of IT neuron with a three distractors. **(d)** Response of IT neuron with a four distractors. It is observed that the response time of the IT neuron remains the same. The response time if shown in figure 6.

required to fixate on a central cross throughout the experiment. A cue appears, either by brightening of the stimulus either to the left or the right of the fixation point (exogenous attention direction) or by central arrow pointing to left or right (endogenous attention direction), to direct attention to the appropriate side of the fixation point. After a subsequent stimulus appearing a certain time after the initial cue a target appears. This is either a cross or an X, and the subject has to respond to the presence of X by pressing the button as soon as possible and the response time is measured. It is found that if there are valid cues (when the cue position is the same as that of the target) there is a speedup of the response time compared to both of the boxes lighting up (exogenous) or a two way central arrow (endogenous). On other hand if there is an invalid cue (when target appears on the opposite side of to the cue direction) then there is a delay of response compared to the double cued cases. The difference of the reaction time in the invalid and valid cue cases is called as attention benefit. The paradigm is important since there is considerable experimental data associated with it, little of which has been simulated.

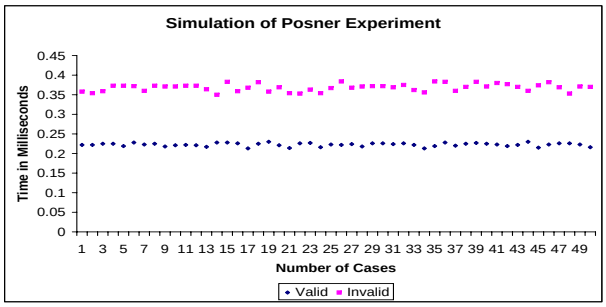


Fig. 8. Firing times of IT neuron in Posner experiments. We have measured response time for valid and invalid cues for 50 presentations of each. The IT neuron for valid cue responded between 0.2sec and 0.25sec whereas for invalid cue the response time was between 0.35sec to 0.4sec.

In this paper we simulated a test of the attention model. Qualitatively, humans benefit most from a valid exogenous cue when it is presented approx 100-200ms before the target, while the endogenous cues are most beneficial with an onset of asynchrony of at least 300ms [25], [6]. The benefits provided by the cues declines rapidly after 200ms, while endogenous cues continue to remain beneficial for several seconds at least.

We have simulated experiments for valid and invalid cues for endogenous attention. Invalid cues were placed in the bottom of the visual field while the valid ones and the objects were in the middle of the field (fig. 8). The reaction times of IT neurons were different for valid and invalid cues. There are two main contributions to the time difference between the

two cases. The first one is the reaction time, i.e. time take to switch attention between different fields. Following [25], [6] we assumed it to be 100ms. The second contribution to the time difference is the time taken for the inhibition from the spatial goal to LIP to ware off and neurons in a new attended area to recover their frequency of firing. In our experiments this time was found to be between 40ms - 60ms.

IV. Comparison with other models

There is an increasing number of models of attention, and it is important to relate our model to them and in particular compare and contrast the underlining approaches to assess the benefits from our model of visual attention.

Velde and Kamps ([5]) present a model of visual object based attention. It is based on neural activity performed by monkeys during visual search task. The model is trained to identify each object in the object array on each location in the array using back error propagation. This results in feed forward network developing a distributed representation of the objects in the hidden layers. Their model does not posses competitive network for determining the focus of attention. Only exogenous attention has been considered. Our model has got inhibition from Spatial Goal to LIP and also IT works on potentiation. Their model lack visual search paradigm. In contrast, we are using more biologically plausible neurons.

Deco and Rolls [3], [4] have recently developed a similar model for visual attention. This work can be seen as complimentary to their model. Their model has implemented visual search tasks, in both serial and parallel search. However, their model does not simulate the paradigm of Posner as presented in this paper. Their model also use a different overall architecture to our model. Mainly it is more of bottom-up influenced model only arising from non-goal sites.

V. Conclusion

The simulations of the model presented here show that information about the location of an identified object (goal) is retrieved by an interaction of ventral and dorsal streams. Our model is based on the fact that the ventral and dorsal streams work together for invariant object recognition. In the model described, the ventral stream can select the goal in a specific spatial attended areas. It was also seen that the IT neurons did not respond if there was not object present in the spatial attended areas.

There are further paradigms and architectures to which this model can be extended to. The next extension to this paradigm is to include exogenous attention. This would enable us to simulate paradigms of serial or conjunction search tasks where targets are distracted in a set of target + distractors. This is especially of interest in the case of rapid serial visual presentation.

References

[1] J. H. Reynolds and R. Desimone, "Competative mechanism subserve selective visual attention," in *Image, Language, Brain: Papers from First Mind Articulation Project Symposium* (A. Marantz, Y. Miyashita, and W. O'Neil, eds.), (London), pp. 233–247, The MIT Press, 2000.

[2] N. G. Mller and A. Kleinschmidt, "Dynamic interaction of object- and space-based attention in retinotopic visual areas," *Journal of Neuroscience*, vol. 23, pp. 9812 – 9816, 2003.

[3] G. Deco and E. T. Rolls, "A neurodynamical theory of visual attention: comparisions with fmri and single neuron data," in *Artifical Neural Networks: ICANN 2002* (J. R. Dorronsoro, ed.), 2002.

[4] G. Deco and E. T. Rolls, "A neurodynamical cortical model of visual attention and invariant object recognition," *Vision Research*, vol. 44, pp. 621–642, 2004.

[5] F. van der Velde and M. de Kamps, "From knowing what to knowing where: Modeling object-based attention with feedback disinhibition of activation," *Journal of Cognitive Neuroscience*, vol. 13, no. 4, pp. 479–491, 2000.

[6] J. G. Taylor and M. Rogers, "A control model of movement of attention," *Neural Networks*, vol. 15, pp. 309–326, 2002.

[7] J. G. Taylor, "A general framework for dunctions of the brain," in *IJCNN 2000*, 2000.

[8] J. G. Taylor, "The where, what and how of consciousness," in *Proceedings of the Emergence of Mind*, (Milan), 2000.

[9] M. Corbetta and G. Shulman, "Control of goal-directed and stimulus-driven attention in the brain," *Nature Review: Neuroscience*, vol. 3, no. 3, pp. 201–215, 2002.

[10] A. D. Mehta, I. Ulbert, and C. E. Schroeder, "Intermodal selective attention in monkeys. ii: Physiological mechanisms of modulation," *Cereb. Cortex*, vol. 10, no. 4, pp. 359–370, 2000.

[11] C. Panchev and S. Wermter, "Spike-timing-dependent synaptic plasticity from single spikes to spike trains," *Neurocomputing*, 2004. to appear.

[12] C. Panchev, S. Wermter, and H. Chen, "Spike-timing dependant competitive learning of integrate-and-fire neurons with active dendrites," in *Proceedings of the International Conference on Artificial Neural Networks*, (Madrid, Spain), August 2002.

[13] J. P. Gottlieb, M. Kusunoki, and M. E. Goldberg, "The representation of visual salience in monkey parietal cortex," *Nature*, vol. 391, no. 481-484, 1998.

[14] M. Kusunoki, C. L. Colby, J.-R. Dhumel, and M. E. Goldberg, "The role of lateral intraparietal area in the control of visospatial attention," in *Association Cortex: Structure and Function* (H. Sakata, J. M. Fuster, and A. Mikami, eds.), Harwood Acedemic Publisher, 1997.

[15] G. Deco and J. Zihl, "Top-down selective visual attention: A neurodynamical approach," *Visual Cognition*, vol. 8, no. 1, pp. 118 – 139, 2001.

[16] E. Rolls and G. Deco, *Computational Neuroscience of Vision*. New York: Oxford University Press, 2002.

[17] M. E. Goldberg, "Attentional and spatial mechanisms in the parietal cortex," in *Vision and movement mechanisms in the cerebral cortex* (R. Caminiti, K.-P. Hoffmann, F. Lacquaniti, and J. Altman, eds.), pp. 89–96, The Human Frontier Science Program, 1996.

[18] L. G. Ungerleider, "What and where in the human brain? evidence from functional brain imaging studies," in *Vision and movement mechanisms in the cerebral cortex* (R. Caminiti, K.-P. Hoffmann, F. Lacquaniti, and J. Altman, eds.), pp. 23–30, The Human Frontier Science Program, 1996.

[19] K. Tanaka, "Inferotemporal cortex and object recognition," in *Vision and movement mechanisms in the cerebral cortex* (R. Caminiti, K.-P. Hoffmann, F. Lacquaniti, and J. Altman, eds.), pp. 126–132, The Human Frontier Science Program, 1996.

[20] G. C. Baylis, "Visualparsing," in *Cognitive Neuroscience of attention: a developmental presepective* (J. E. Richards, ed.), (New Jersey), pp. 251–286, Lawrence Erlbaum Associates, 1998.

[21] M. I. Posner, "Orienting of attention," *Quarterly Jouranl of Experimental Psychology*, vol. 32, pp. 3–26, 1980.

[22] "Genesis (general neural simulation system)," *http://www.genesis-sim.org/GENESIS/*.

[23] M. C. Mozer and M. Sitton, "Compuatational modelling of spatial attention," in *Attention* (H. Pashler, ed.), (Philadelphia), pp. 341–393, Taylor & Francis Press, 1998.

[24] S. Corchs and G. Deco, "A neurodynamical model for selective visual attention using oscillators," *Neural Networks*, vol. 14, no. 8, pp. 981–990, 2001.

[25] R. D. Wright, ed., *Visual Attention*. USA: Oxford University Press, 1998.