SELECTIVE ATTENTION MODEL OF MOVING OBJECTS

Roman Borisyuk, David Chik*, Yakov Kazanovich†*

Abstract: Tracking moving objects is a vital visual task for the survival of an animal. We describe oscillatory neural network models of visual attention with a central element that can track a moving target among a set of distracters on the screen. At the initial stage, the model forms the focus of attention on an arbitrary object that is considered as a target. Other objects are treated as distracters. We present here two models: 1) synchronisation based model designed as a network of phase oscillators and 2) spiking neural model which is based on the idea of resource-limited parallel visual pointers. Selective attention and the tracking process are represented by the partial synchronisation between the central unit and a subgroup of peripheral elements. Simulation results are in overall agreement with the findings from psychological experiments: overlapping between the target and distractors is the main source of errors.

Key words: *Partial synchronization, visual pointer, object tracking, phase oscillator, spiking neuron model*

Received: January 30, 2009
Revised and accepted: July 10, 2009

1. Introduction

Selective visual attention is a mechanism that gives a living organism the possibility to extract from the incoming visual information a part that is most important at a given moment and that should be processed in more detail. This mechanism is necessary due to limited processing capabilities of the visual system which does not allow rapid analysis of the whole visual scene.

The important property of attention is its metastability. This means that after being fixed, the focus of attention does not change for some time even when objects in a scene gradually vary their parameters (shape, brightness, position). In particular, the metastability of attention makes it possible to track a moving object.

*Roman Borisyuk, David Chik
University of Plymouth, Plymouth, PL4 8AA, UK, E-mail: `r.borisyuk@plymouth.ac.uk`, http://www.tech.plym.ac.uk/soc/staff/roman/home2.htm
†Yakov Kazanovich
Institute of Mathematical Problems in Biology of the Russian Academy of Sciences, Pushchino, 240290 Russia

Special conditions should be fulfilled for attention to be switched from one object to another. These conditions include (a) an abrupt change of parameters of an object in the focus of attention, (b) appearance and disappearance of objects in the scene, (c) overlapping or hiding of objects due to their movements, (d) termination or voluntary break of object processing.

In recent years attention has become a popular field for neural network modelling. The models of attention can be subdivided into two categories. Connectionist models (for example, [1-4]) are based on a winner-takes-all procedure and are implemented through a proper modification of the weights of connections in a hierarchical neural network. Such models are difficult to use in the case of moving objects since the networks have to work in the space of the visual field, therefore for any new position of the objects the weights of the connections should be recomputed. Moreover, these models are incompatible with modern hypotheses about distributed representations in the brain and experimental findings about the relation between attention and synchronous oscillatory activity in the cortex [5-9].

Another type of attention modelling is based on the synchronisation principle in oscillatory neural networks (for example, [10-17]). Such kind of modelling is more suitable for object-based visual attention because in this case the network operates in phase-frequency space which makes the attention focus invariant to the locations of objects in physical space. Though this approach provides the potential for working with non-stationary images, no application of oscillatory neural networks in this case is known to us.

In this paper we present two oscillatory neural networks with a central element for automatic object-oriented attention. The main advantage of these models is that they can work with non-stationary objects. It is assumed that objects in the input image can move and change their intensity and form. For simplicity we consider the case of greyscale images. The focus of attention is represented by those oscillators that work synchronously with the central oscillator. It is shown that a proper synchronisation regime can be obtained by a suitable choice of parameters of oscillators and coupling strengths.

Wang and Terman [13, 18] developed a network LEGION of locally interacting Van der Pol type oscillators whose activity is globally controlled by an inhibitory neuron. The architecture of connections in LEGION is similar to the one that we use in our models (the inhibitory neuron plays the role of the central element in LEGION). Still the choice of Van der Pol oscillators as components of the model results in some difficulties when spreading LEGION operation on non-stationary objects. Phase oscillators and Hodgkin-Huxley type neurons that are used in our models seems to be more suitable for this purpose.

The first model is a modification of our Attention Model with a Central Oscillator (AMCO) comprising a layer of the so-called peripheral oscillators interacting with a special central oscillator [10-12]. Advanced phase oscillators have been used as the elements of this model. The state of such an oscillator is described by three variables: phase, amplitude, and natural frequency of oscillations. The functioning of the model has been based on the following main principles: 1) partial synchronisation between the central oscillator and some subset of peripheral oscillators and 2) resonant increase of the amplitude during partial synchronisation. The phase-locking mechanism used to synchronise oscillators allows them to achieve similar frequencies and the focus of attention is assumed to be formed by those peripheral oscillators whose activity is partially synchronous with the activity of the central oscillator. The functioning of the system is based on the principles of phase-locking, adaptation of the natural frequency of the central oscillator, and the resonance influence of the central oscillator on the assembly of oscillators that work in-phase with the central element.

The second model of object tracking is built from biologically derived Hodgkin-Huxley model neurons [19]. In this model we use the synchronisation hypothesis [5, 6] for implementation of a 'pointer' that identifies the neural assembly, coding a specified object. Pylyshyn [20] postulated that pointers are picked out and stay attached to individual objects in the visual field independently of object features. The neural mechanism of visual pointers is unclear, but there are studies which provide some suggestions about its nature. In Intriligator's study [21], it was found that the spatial resolution of visual attention is different from (coarser than) that of the receptive field, suggesting the existence of a separate 'pointer field' in the parietal cortex. In the study by Egly et al. [22], it was shown that attention seems to start at one point, then propagates and fills the object. Another study by Alvarez and Scholl [23] showed that when tracking a line, attention seems to concentrate in the middle of the line. This implies that the brain may be capable of quickly and roughly calculating the centre of an object in the receptive field.

Model descriptions and results of simulations are presented in the following sections.

2. Phase Oscillator Model of Object Tracking

2.1 Model description

The network consists of a central oscillator (CO) that has feedforward and feedback connections to a set of oscillators located in the vertices of a square grid. We refer to these oscillators as peripheral oscillators (PO). Besides the CO, each PO is coupled with its four nearest neighbours except on the boundaries where no wrap-around is used.

An oscillator is described by three variables: the oscillation phase, the oscillation amplitude, and the natural frequency of the oscillator. The values of these variables change in time according to prescribed rules of interaction between oscillators. The input to the network is an image on the plane that contains several connected greyscale objects on the white background. The objects are non-stationary, they can continuously change their form, brightness and position. Due to movements, objects can temporarily overlap with each other. While two (or several) objects are overlapping, they are considered as a single object. In the case of overlapping it is always specified which object lies "on the top". Therefore, at any moment each pixel in the image has a single grey-level associated with it.

The input to the network is an image on the plane grid of the same size as the grid of peripheral oscillators. So there is a one-to-one correspondence between the pixels in the image and the peripheral oscillators. Each PO receives an external input from the corresponding pixel. This signal determines the natural frequency of the PO. The darker is the pixel, the better is its contrast relative to the background, and the higher is the value assigned to the natural frequency. It is assumed that the image contains several non-overlapping greyscale objects on the white background of the constant brightness. We call an *assembly* a connected set of POs that are stimulated by a single object. We say that an object is *coded* by the corresponding assembly of POs.

Depending on the input signal and previous dynamics, a PO can be in one of three states: *active*, *resonant*, and *silent*. If a PO receives zero input (corresponding to the signal from the background), it is in the silent state. In this state the oscillator does not participate in the network dynamics and is not included in the dynamics equations shown below. A PO becomes active as soon as it receives a non-zero signal from a pixel of an object. If the amplitude of oscillations of an active oscillator exceeds a certain threshold, it changes its state to resonant one. Being in the resonant state is interpreted as the fact that this oscillator corresponds to the pixel included in the focus of attention. If attention is switched to another object, the oscillator's amplitude drops down and the oscillator returns to the active state.

The dynamics of the network is described by the following equations:

$$\frac{d\theta_0}{dt} = \omega_0 + \frac{w}{n}\sum_{i\in N} a_i q(\omega_i)\, g(\theta_i - \theta_0), \tag{2.1}$$

$$\frac{d\theta_i}{dt} = \omega_i - a_0 w_0 \sin(\theta_0 - \theta_i) + w_1 \sum_{j\in N_i} a_j \sin(\theta_j - \theta_i) + \rho, \quad i \in N, \tag{2.2}$$

$$\frac{da_i}{dt} = \beta_1(-a_i + \gamma f(\theta_0 - \theta_i))^+ + \beta_2(-a_i + \gamma f(\theta_0 - \theta_i))^-, \quad i \in N, \tag{2.3}$$

$$\frac{d\omega_0}{dt} = -\alpha\left(\omega_0 - \frac{d\theta_0}{dt}\right). \tag{2.4}$$

In these equations, θ_0 is the phase of the CO, θ_i are the phases of POs, $\frac{d\theta_0}{dt}$ and $\frac{d\theta_i}{dt}$ are the current frequencies of oscillators, ω_0 is the natural frequency of the CO, ω_i are the natural frequencies of POs, a_0 is the amplitude of oscillations of the CO, a_i are the amplitudes of oscillations of POs, w, w_0, w_1 are positive parameters (constants) that control the strength of interaction between oscillators, N is the set of non-silent oscillators of size n, N_i is the set of non-silent oscillators in the nearest neighbourhood of the oscillator i, ρ is the normally distributed noise with mean 0 and standard deviation σ, the function g controls the influence of POs on the CO (g is 2π-periodic, odd, and unimodal in the interval of periodicity), f is the function that controls the amplitude of oscillations and transition to the resonant state (f is 2π-periodic, even, positive, unimodal in the interval of periodicity with the maxima in the points $2\pi k$), q is a positive non-decreasing function, $\alpha, \beta_1, \beta_2, \gamma$ are network parameters (positive constants). By definition,

$$(x)^+ = \begin{cases} x & if\ x \geq 0 \\ 0 & if\ x < 0 \end{cases}, \quad (x)^- = \begin{cases} x & if\ x \leq 0 \\ 0 & if\ x > 0 \end{cases}. \tag{2.5}$$

More details on the functions g and f can be found in [12].

Equations (2.1)–(2.2) describe the mechanism of phase-locking. Note that the interaction between oscillators depends not only on the coupling strength but also on the amplitude of oscillations: an oscillator with greater amplitude has stronger influence on other oscillators. The function q is used to make the influence of a PO on the CO dependant on the frequency of the PO. This results in the fact that an assembly of oscillators that codes a darker object will have greater influence on the CO than other oscillators. Thus a darker object (which is more salient on the white background) will have a better chance to be included in the attention focus. Also objects of greater size have a priority in being included in the attention focus just because of greater contribution to the sum of equation (2.1). We use the function q of the form

$$q(\omega) = q_0 + \mu(\omega - q_1)^+,$$

where q_0, q_1, μ are parameters.

The coupling strength between oscillators is constant. The connections from POs to the CO and local connections between POs are synchronising. The connections from the CO to POs are desynchronising. Due to synchronising connections from POs to the CO, the latter can be phase-locked by an assembly of POs. Due to synchronising connections between POs, the oscillators from an assembly of active POs become phase-locked and work nearly in-phase after they reach the resonant state. Desynchronising connections from the CO to POs are used to break the coherence between different assemblies of POs so that at each moment the system tends to include only one object in the attention focus.

Equation (2.3) describes the dynamics of the amplitude of oscillations of POs. This equation provides the mechanism for the resonant increase of the amplitude of oscillations. We say that a PO is in the resonant state if its amplitude exceeds the threshold $R = 0.8a_{\max}$. The parameters β_1 and β_2 $(\beta_1 < \beta_2)$ determine the rate of amplitude increase and decay.

Equation (2.4) describes the mechanism of adaptation of the natural frequency of the CO. According to this equation, ω_0 tends to the current frequency of the CO. The adaptation is used to allow the CO to "search" for an assembly of POs with which the CO is going to synchronise. The parameter α determines the rate of adaptation. The value of α is chosen low enough for ω_0 to follow the main trend of the current frequency of the CO but not random fluctuations of this frequency.

In biological terms the model is interpreted in the following way. It is assumed that POs represent cortical columns and are constituted of locally interacting populations of excitatory and inhibitory neurons of the cortex. The CO represents the central executive of the attention system that is implemented by a distributed network in the prefrontal cortex and the septo-hippocampal system.

The selection of an object is realised through the following dynamics of the network. After network initialisation and start, the CO is synchronised by the assembly of POs that currently makes the greatest contribution into controlling the phase dynamics of the CO. This assembly corresponds to the most salient object. The relative importance of size and contrast for the saliency are controlled by the parameters of the function $q(\omega)$. The POs that work synchronously with the CO significantly increase the amplitude of their oscillations (switching to the state of resonance). The activity of other oscillators is shut down to a low level. This prevents the focus of attention from unwanted switching that could otherwise happen due to changes in object saliency.

The attention focus is kept stable until the object in the focus of attention, say A, is crossed (overlapped) by another object B. If this event happens, the combination of two objects is considered by the system as a connected object C. The assembly of oscillators that codes C will become synchronous (due to local connections between POs) and will synchronise the CO. After that the amplitudes of oscillations of the POs that code B will increase to the level of the resonance. Thus the whole object C will be included in the attention focus. The next change of the attention focus happens when C again separates into two isolated object A and B. At that moment the focus of attention will move to A or B depending on their current saliency.

2.2 Simulation results

In the following examples, the frames show the state of the network at integer moments of time. The frames are ordered from left to right and from top to bottom. The state of an oscillator is represented in the figures according to the following scheme: a pixel is white, grey, or black depending on whether the corresponding oscillator is silent, active, or resonant at the current moment of time. Thus, black pixels represent the focus of attention. The arrows show the direction of movement.

Example 1. Consider the case when the saliency of an object is solely determined by its size. Let an image contain two circles of a fixed size and intensity (Fig. 1). The circles move towards each other (the direction is fixed for each object). The circle of radius 5 moves to the right, the circle of radius 4 moves to the left. Denote by A and B the assemblies of POs coding the circles in the network. The natural frequencies of oscillators in A and B are 5 and 4, respectively. This reflects the fact that the circles have different grey-levels (not shown in the figure), but in this example this has no influence on the strength of interaction between A and B and the CO (the difference in grey-level is below the sensitivity of the attention system).

As can be seen from Fig. 1, after a short transitional period (two time units) the attention is focused on the larger circle (note that this takes place despite the fact that the initial frequency of the CO coincides with the natural frequency of oscillators in B). When the circles collide and overlap, the focus of attention is spread on both circles as if they form a complex object (see frames 9-15). Lately, when the circles separate, attention is again focused on the larger and therefore more salient circle (frames 16-20).

Example 2. This example presents the case when the saliency of an object depends on its intensity (Fig. 2). The example differs from the previous one by the value of the natural frequency of oscillators in the assembly B, which is now equal to 6. This means that the smaller circle is darker than the lager one and, hence, more salient on the white background. Therefore, the influence of the oscillators from B on the CO is increased in comparison to example 1. As a result, after a short transient period (two time units) attention is focused on the smaller circle (frame 3). After the collision (overlapping) of the circles, the focus of attention is temporarily spread on both circles (frames 10-15), but their separation results in attention returning to the smaller circle (frames 17-20).

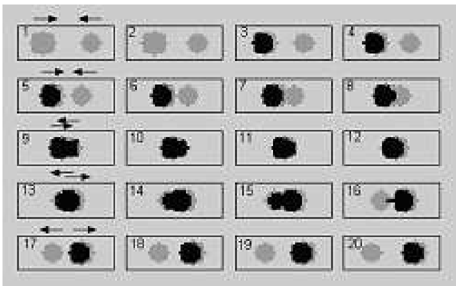


Fig. 1 *Selection of moving objects that is based on the size of the objects. Two circles of different size move towards each other. Attention is focused on the larger circle all the time except a short initial transitional period and the time when both circles collide and form a complex object.*

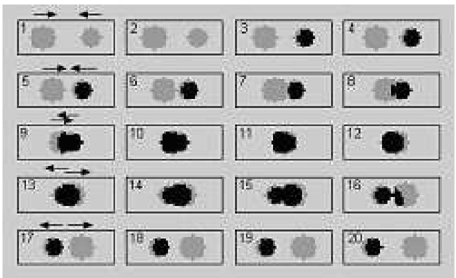


Fig. 2 *Selection of moving objects that is based on the intensity. Two circles of different size and intensity move towards each other. Attention is focused on the smaller object since it is darker (and has a better contrast) than the larger one.*

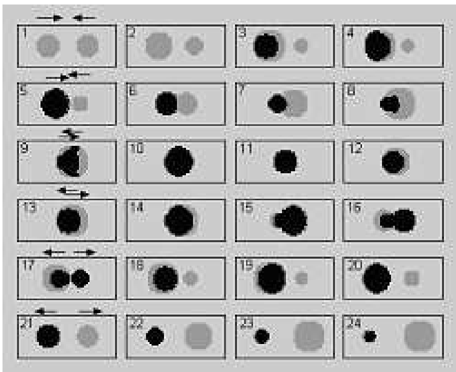


Fig. 3 *Selection of moving objects of varying size. Two circles of the same grey level move towards each other. Initially, attention is focused on the circle moving to the right but after collision and separation of circles it is switched to the circle moving to the left.*

Example 3. This example shows the case when the circles simultaneously move and change size (Fig. 3). It is presented to show that the focus of attention has the property of metastability. The image contains two circles whose radii periodically change in the range between 3 and 8. The radii vary in antiphase, that is while the first radius is increasing the other one is decreasing. At the initial moment both radii are identical. The natural frequencies of oscillators in the assemblies A and B are equal to 5. This means that both circles have the same grey level and, therefore, their saliency is determined by their size only. At the moment when the focus of attention is formed, the circle that moves to the right has larger size (frame 3), therefore attention is attracted by this circles and is fixed on it during some time (frames 3–8) despite the fact that the circle in the attention focus becomes smaller than the other circle (frames 7–8). Frames 9–17 correspond to the time when both circles are included in the attention focus. Then the largest circle (that moves to the left) wins the competition for attention (frame 18) and holds it even when this circle becomes small (frames 22–24).

3. Spiking Element Model of Object Tracking

3.1 Model description

The architecture of network connections is shown in Fig. 4. The tracking system consists of two stages of processing. First, visual objects enter a cognitive control module which represents attention pointers located in the parietal cortex as suggested by experiments [20–23]. Second, there is an attention formation module which implements partial synchronisation of "attended" neurons, in accordance with the experimentally found correlation of neural synchronous oscillations with various brain functions [7, 8, 9, 24].

The cognitive control module consists of a large grid of neurons covering the whole receptive field. Visual objects are represented separately in this receptive field (shown as black elements in the cognitive control grid in Fig. 4). Based on the location of the target, there is a smaller grid of neurons inside the receptive field (called "pointer field", shown as a grey square grid within the cognitive control grid in Fig. 4). The size of the pointer field is bigger than the size of the target. As the target moves, the location of the pointer field will be updated to catch up with the target. The updating process is assumed to be discrete (for example 10Hz). Inside the pointer field, we call some pixels "active" when they are representing some objects. The new location of the pointer field is calculated by averaging the co-ordinates of active pixels inside the pointer field. Mathematically,

$$X(t+1) = \frac{1}{N(t)} \sum_{i=1}^{N} x_i(t), \tag{3.1}$$

$$Y(t+1) = \frac{1}{N(t)} \sum_{i=1}^{N} y_i(t). \tag{3.2}$$

Here, (X,Y) is the co-ordinate of the centre of the pointer field inside the receptive field. N is the number of active pixels in the pointer field, and (x_i, y_i)

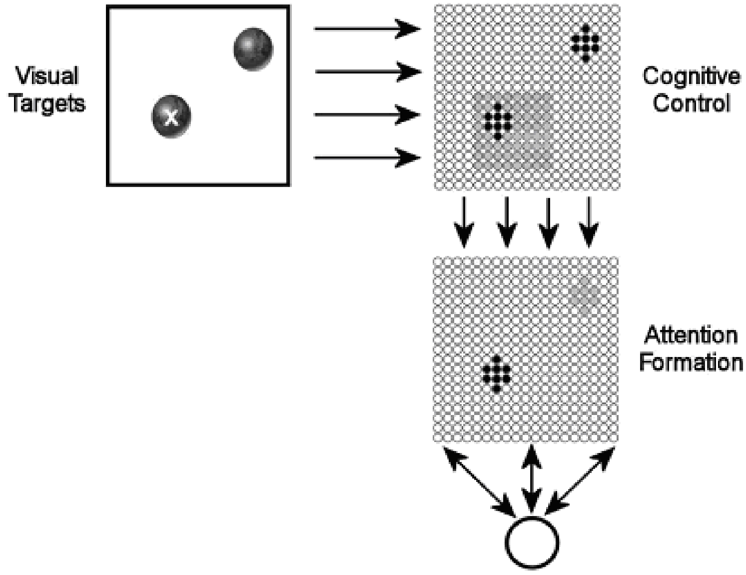


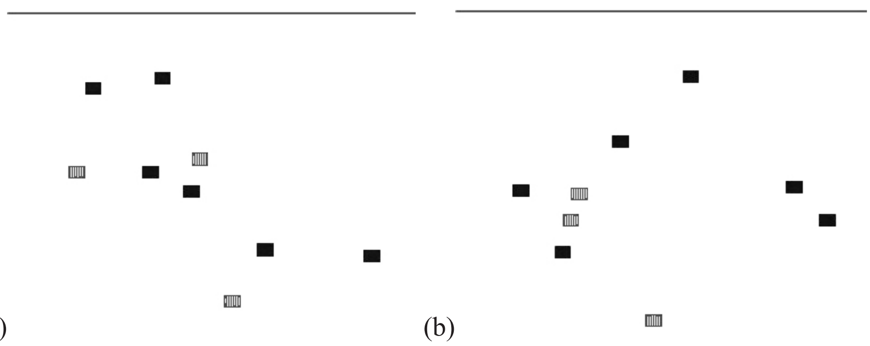
Fig. 4 *The architecture for the spiking element model. Visual objects enter the cognitive control module which provides pointer fields to track the targets. Then information is passed to the attention formation module which provides partial synchronisation for the attended object. The central neuron is shown by the circle. See the text for details.*

are the co-ordinates of active pixels. The new location of the pointer field (at time $t + 1$) is calculated from previous information (at time t). The same system can deal with the situation when there are more than one visual target. If there are k visual targets, there will also be k pointer fields operating in parallel (according to experimental findings $k < 6$).

The attention formation module consists of a large grid of "peripheral neurons". The grid has the same size as the receptive field. In addition, there is a "central neuron" which inhibits distracters (shown as grey elements in the attention formation grid in Fig. 4) and synchronises the spikes of attended peripheral neurons (shown as black elements in the attention formation grid in Fig. 4). The neurons in the attention formation module are simulated using a Hodgkin-Huxley model. The synaptic currents between neurons are modelled by the standard approach based on an alpha function. Equations and detailed descriptions can be found in Appendix.

3.2 Simulation results

Fig. 5 shows an example of a multiple object tracking task. All objects move randomly and the system needs to keep track with 3 targets out of 9 objects. Before the start of tracking, the subject would be told which targets to track. In our simulations, we assume that a top-down signal is provided to form the initial pointer fields in the cognitive control module. After the tracking task starts, the pointer fields update their locations according to equations (3.1)–(3.2). A visual object inside a pointer field is considered as the "attended" object. This information will be passed to the attention formation module. Peripheral neurons in the attention formation module corresponding to the attended object have higher natural frequencies. They will be synchronised by the system. Other peripheral neurons will be suppressed. This partial synchronisation represents selective attention being focused on a particular object.



(a) (b)

Fig. 5 *(a) and (b) illustrate two snapshots of the multiple object tracking task. The size of the receptive field is 400x400. Three (out of nine) objects are tracked by the system for a duration of 10 sec. All objects move randomly. Targets being tracked are indicated by grey vertical stripes in the figure, but in actual process targets and distracters are not distinguishable by their appearance.*

Fig. 6 shows the potential traces of the central neuron (CN) and some peripheral neurons (PNs) in the attention formation module. The top panel shows the trace of the CN. The next three panels show the potential traces of three attended PNs. The bottom three panels show the traces of three unattended PNs. The unattended PNs remain subthreshold (no spikes are produced by them), but the attended PNs form stochastic synchronisation with the CN. For individual PNs, the trace is quite irregular (due to noise), but as a whole population, the degree of synchronisation is strong.

Fig. 7 shows a series of snapshots during a tracking process. In the first snapshot (top left), the location of the centre of the pointer field is b, the location of the centre of the visual target is c. In the next snapshot, the location of the centre of the pointer field is updated to c, but the target moves to a new location d, and so on.

The system performs reasonably well in most situations. However, the system also produces some tracking errors in a way similar to those found in psychological experiments. As we can expect from Fig. 7, if the target moves too fast, the pointer field cannot catch up with the target. This may result in a tracking error, which is indeed observed in experiments [3]. The number of errors can be reduced by increasing the frequency of updating. We have assumed that this updating frequency relates to the operating brain rhythm (beta rhythm). Since object tracking is an

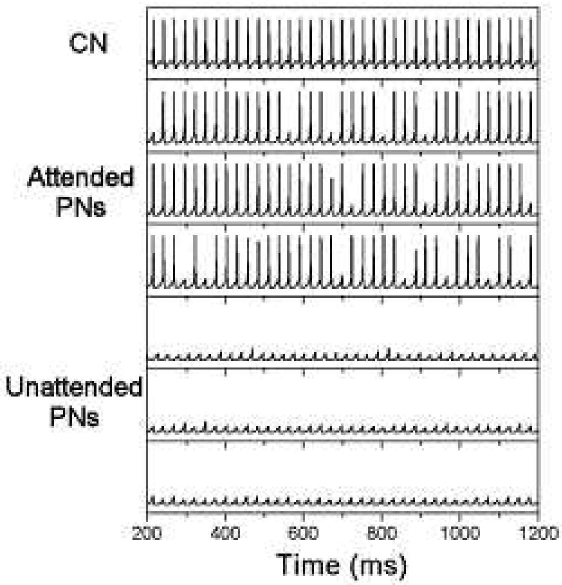


Fig. 6 *Partial synchronisation between attended PNs and the CN. The traces in 2^{nd}, 3^{rd}, and 4^{th} panels are the potential traces of attended PNs. They are stochastically synchronised with the CN (trace at top). The bottom three traces are the potential traces of unattended PNs which produce only sub-threshold fluctuations.*

attention-demanding task, an alert mind state (with higher brainwave frequency) may be associated with higher tracking accuracy.

Another source of errors comes from the confusion when a target overlaps with (or comes very close to) a distractor. For example, in the fourth snapshot, a target and a distractor join together, and the centre of location becomes f. In the fifth snapshot, the pointer field moves to f, there are now two groups of active pixels inside this pointer field. The pointer field will continue to use the average co-ordinates as its next location, but it may leave the target and attach to the distractor.

A movie showing the performance of the cognitive control module can be downloaded from `http://www.pion.ac.uk/binmam/mot3.avi`. Two snapshots are shown in Fig. 8. There are 9 objects (shown as small white rectangles). Three pointer fields (indicated as larger rectangles with horizontal stripes, vertical stripes, and woven texture respectively) are trying to track 3 targets respectively. At 0:04 (Fig. 8a), the woven-texture pointer field disappears due to confusion of two objects going in opposite direction. At 0:09 (Fig. 8b), the horizontal- and vertical-stripes pointer fields merge together. Finally, after 0:13, there is only one pointer field left. It is interesting to compare this simulation with psychological findings. Sometimes when human subjects are asked to track several fast moving objects, the subject will reduce the number of targets during the process. This tactic used by humans is automatically implemented in our model.

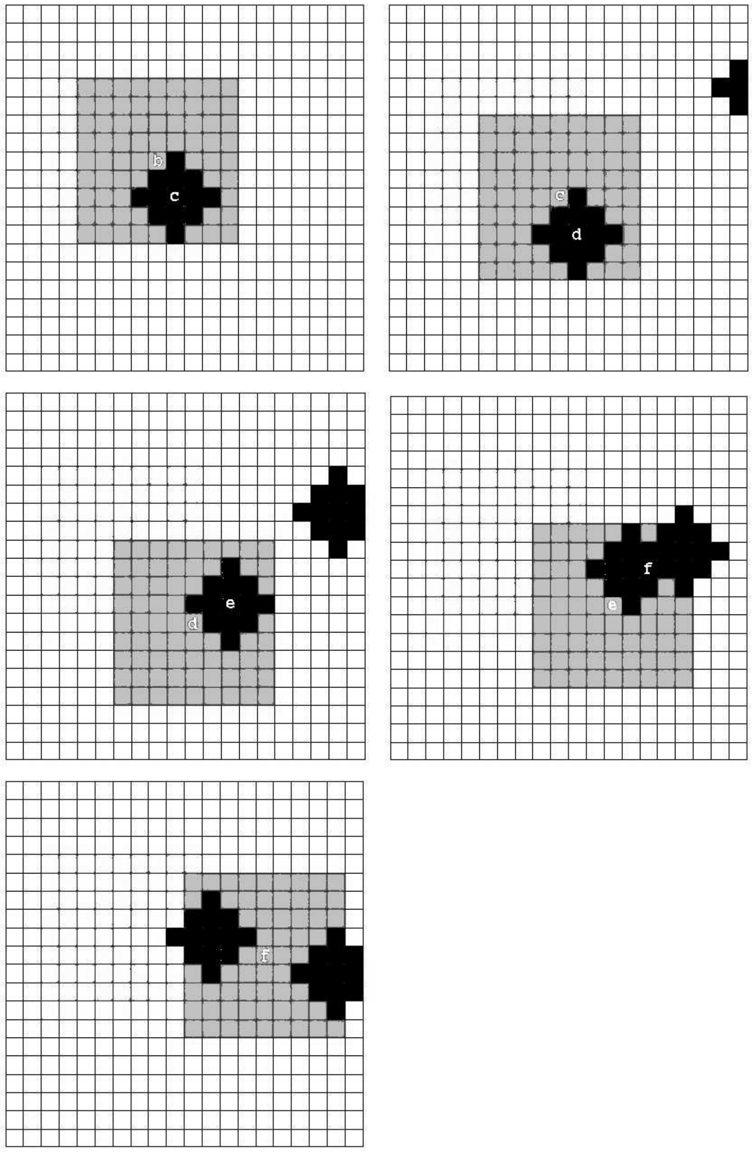
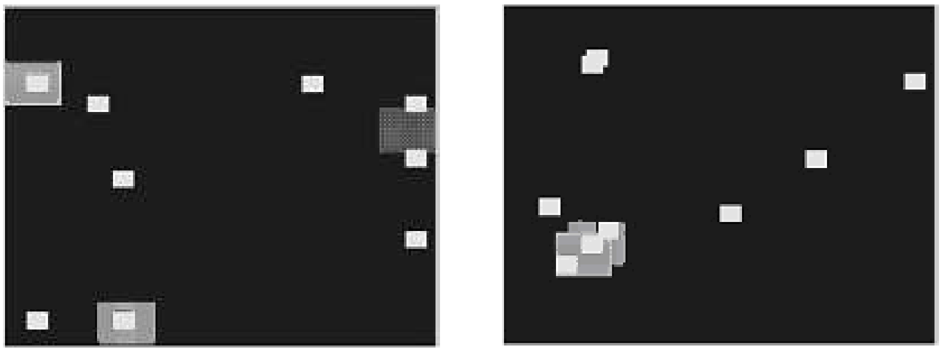


Fig. 7 *Snapshots of object tracking in the cognitive control module. Objects are shown as black. The pointer field is shown as grey.*

4. Discussion

Two models of visual attention applied to non-stationary images have been presented. The models have a similar architecture of connections but differ in the type of constructing units. The first model is built from phase oscillators, therefore it described the process of attention focusing in a rather abstract form as phase-locking of phase rotators. The second model is designed as a network of Hodgkin-Huxley



(a) (b)

Fig. 8 *Two snapshots from the movie (http://www.pion.ac.uk/binmam/mot3.avi), (a) at time 0:04; (b) at time 0:09.*

type neurons. It is more realistic from a biological point of view. In particular, the temporal and frequency characteristics of this model comply with experimental findings.

Both models allow the tracking of moving objects and attention switching from one object to another. The first model shows successful functioning even when objects change brightness or size (Fig. 1–3). When forming or switching attention the advantage is provided to more salient objects with higher intensity or larger size. The model demonstrates metastability of the focus of attention and automatic selection of a more salient object when moving objects collide and then separate again. The limitation of the model is that only one object can be tracked simultaneously.

The second model allows simultaneous tracking of several objects among a set of distractors which comply with the capabilities of human observers [20]. In this respect this model can be considered as an adaptation of our previous phase oscillator model for multiple object tracking [25] to the Hodgkin-Huxley neural network model. The model demonstrates realistic synchronisation behaviour in its spiking activity (Fig. 6).

The models capture some important biological and psychological findings, such as neural synchronisation, gamma oscillations, resource-limited parallel tracking mechanism, visual indexing, and so on. The simulation results are in overall agreement with the findings from psychological experiments: the models demonstrate the principal capability of tracking moving objects with overlapping between a target and a distracter being the main source of tracking errors. The detailed investigation of model performance is planned for the future. In particular, the dependence between tracking performance and neuron frequency should be elucidated.

Two main ideas are combined in the presented models: 1) key role of synchronisation in the formation of the attention focus and 2) control of the dynamics is realised through the central unit, therefore the number of connections in the network is of the same order as the number of elements (i.e. linearly dependant on the size of the visual field). The functioning of our models is invariant relative to locations of objects in the image. This is why the processing of non-stationary objects does not evoke any additional difficulties. Note that the situation is quite different for winner-take-all models because in these models the modification of connection strengths is attached to the definite positions of objects.

As we mentioned before, the architecture of our model is similar to the one of LEGION used by Wang and Terman [13, 18]. The main difference is that LEGION can only operate with static images while our models can be applied to both static and dynamically changing images including moving objects. Besides, in our models the selection of the most salient object does not require several stages of attention focus formation while LEGION requires sequential elimination of unwanted objects from the attention focus. LEGION is designed as a network of Van der Pol oscillators, therefore it is difficult to compare its temporal characteristics (such as the working frequency, the speed of attention focusing, etc.) with experimental data. On the other hand, LEGION is shown to be successful in processing realistic images while our models have been applied to simple artificial images where clear separation of targets and distracters from the background takes place. More efforts are needed to spread our models to real images from the video camera.

Acknowledgments

This research work was supported by the UK EPSRC (Grant EP/D036364/1), by the Russian Foundation of Basic Research (Grant 07-01-00218), and by the Ministry of High Education and Science of the Russian Federation (Grant 2.1.1/3876).

Appendix: Detailed mathematical formulation of the attention formation module

We use the Hodgkin-Huxley model for both peripheral neurons (PNs) and the central neuron (CN) in the attention formation module of the spiking element model of object tracking. Suppose that there are N peripheral neurons and one central neuron, a total of $N+1$ neurons. They are described by the following equations:

$$\frac{dV_i}{dt} = -I_{ion,i} + I_{ext,i} - I_{syn,i}, \tag{A1}$$

$$\frac{dX_i}{dt} = A_X(V_i)(1 - X_i) - B_X(V_i)X_i, \quad X_i \in \{m_i, h_i, n_i\}, \tag{A2}$$

$$A_m(V_i) = (2.5 - 0.1(V_i - V_{rest}))/(\exp(2.5 - 0.1(V_i - V_{rest})) - 1), \tag{A3}$$

$$A_h(V_i) = 0.07 \exp(-(V_i - V_{rest})/20), \tag{A4}$$

$$A_n(V_i) = (0.1 - 0.01(V_i - V_{rest}))/(\exp(1 - 0.1(V_i - V_{rest})) - 1), \tag{A5}$$

$$B_m(V_i) = 4\exp(-(V_i - V_{rest})/18), \tag{A6}$$

$$B_h(V_i) = 1/(\exp(3 - 0.1(V_i - V_{rest})) + 1), \tag{A7}$$

$$B_n(V_i) = 0.125\exp(-(V_i - V_{rest})/80), \tag{A8}$$

where $i = 1, 2, \ldots, N$ is used to index PNs; $i = N+1$ indexes the CN. The meaning of variables and the values of parameters are: $V_i(t)$ is the membrane potential of a neuron; $m_i(t)$ is the activation variable of the sodium conductance channel; $h_i(t)$ is the inactivation variable of the sodium conductance channel; $n_i(t)$ is the activation variable of the potassium conductance channel; $I_{ion,\ i}(t)$ is the total ionic current; $I_{ext,\ i}(t)$ is the external current to the neuron; $I_{syn,\,i}(t)$ is the synaptic current received by the neuron; V_{rest} is the resting potential of the neuron (equal to -65 mV). Notice that the membrane capacitance is equal to 1, therefore it is not shown in Equation A1.

The sum of ionic currents of the ith neuron is,

$$I_{ion,i} = g_{Na} m_i^3 h_i (V_i - V_{Na}) + g_K n_i^4 (V_i - V_K) + g_L (V_i - V_L), \tag{A9}$$

where $i = 1, 2, \ldots, N+1$. The meaning and the values of parameters are: V_{Na} is the reversal potential for the sodium current (equal to 50 mV); V_K is the reversal potential for the potassium current (equal to -77 mV); V_L is the reversal potential for the leak current (equal to -54.4 mV); g_{Na} is the maximum conductance for the sodium current ($g_{Na} = 120$ (1+0.02η) mS/cm2, η is uniformly distributed in [-1,1]); g_K is the maximum conductance for the potassium current ($g_K = 36(1 + 0.02\eta)$ mS/cm2, η is uniformly distributed in $[-1, 1]$); g_L is the maximum conductance for the leak current ($g_L = 0.3(1 + 0.02\eta)$ mS/cm2, η is uniformly distributed in $[-1, 1]$).

The following formulas define the external currents (I_{ext}) incoming to PNs and the CN respectively:

$$I_{ext,\ i}(t) = \begin{cases} 5(1 + 0.2\xi_i(t)) & background \\ 10(1 + 0.2\xi_i(t)) & unattended \\ 20(1 + 0.2\xi_i(t)) & attended \end{cases}, \quad i = 1, 2, ..., N, \tag{A10}$$

$$I_{ext,\ N+1} = 5\,mA. \tag{A11}$$

Here, $\xi(t)$ is a random noise uniformly distributed in [-1,1]. Different PNs receive different external currents depending on whether it is representing the background, "attended" object or "unattended" object. An attended object corresponds to a visual object inside a pointer field. This information comes from the cognitive control module.

The following formulas define the synaptic currents (I_{syn}) received by PNs and the CN respectively:

$$\begin{aligned} I_{syn,i} &= w_2(V - V_{syn,inh}) \sum_{k=1}^{M_2} \alpha(t - T_k) + w_3(V - V_{syn,exc}) \sum_{j \in \mathrm{N}_i} \sum_{k=1}^{M_j} \alpha(t - S_{j,k}), \\ i &= 1, 2, \ldots, N \qquad (A12) \end{aligned}$$

$$I_{syn,\ N+1} = w_1(V_{N+1} - V_{syn,\ exc}) \sum_{j=1}^{N} \sum_{k=1}^{M_j} \alpha(t - S_{j,k}). \qquad (A13)$$

Each PN receives strong inhibitory synaptic currents from the CN, and weak local excitatory synaptic currents from 8 neighbouring neurons. In Equation (A12), the connection strength between CN and a PN is $w_2 = 9$; $V_{syn,\ inh} = -80$ mV is the synaptic reversal potential of inhibitory coupling; M_2 is the total number of spikes from the CN; T_k is the time of the kth spike generated by the CN; $\alpha(t) = at\exp(-bt)$ for $t \geq 0$ and zero for $t < 0$ is the alpha-function of synaptic conductance, with parameters $a = 6$ msec^{-1} and $b = 0.3$ msec^{-1}; $w_3 = 0.5/\mathrm{N}_i$ is the connection strength of local excitatory coupling between neighbouring PNs; $V_{syn,\ exc} = 0$ is the synaptic reversal potential of excitatory coupling; N_i is the neighbourhood of the ith PN (for a PN in the middle it has 8 neighbours, PNs at edges and corners have mirror boundary condition); M_j is the total number of spikes of the jth PN; $S_{j,k}$ is the time of the kth spike generated by the jth PN. The CN receives excitatory synaptic inputs from all peripheral neurons. $w_1 = 1/N$ is the connection strength where N is the total number of PNs.

References

[1] Olshausen B. A., Anderson C. H., Van Essen D. C.: A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. J. Neurosci., **13**, 1992, pp. 4700–4719.

[2] Tsotsos J. K., Culhane S., Wai W., Lai Y., Davis N., Nuflo F.: Modeling visual attention via selective tuning. Artificial Intelligence, **78**, 1995, pp. 507–547.

[3] Deco G., Rolls E. T.: Neurodynamics of biased competition and cooperation for attention: A model with spiking neurons. J. Neurophysiol., **94**, 2005, pp. 295–313.

[4] Itti L., Koch C.: Computational modeling of visual attention. Nature Reviews Neuroscience, **2**, 2001, pp. 194-203.

[5] Gray C. M.: The temporal correlation hypothesis is still alive and well. Neuron., **24**, pp. 31–47.

[6] Singer W.: Neuronal synchrony: A versatile code for the definition of relations. Neuron, **24**, 1999, pp. 49–65.

[7] Fries P., Schroeder J.-H., Roelfsema P. R., Singer W., Engel A. K.: Oscillatory neural synchronisation in primary visual cortex as a correlate of stimulus selection. J. Neurosci., **22**, 2002, pp. 3739–3754.

[8] Jensen O., Kaiser J., Lachaux J.-P.: Human gamma-frequency oscillations associated with attention and memory. Trends Neurosci., **30**, 2007, pp. 317–324.

[9] Steinmetz P. N., Roy A., Fitzgerald P., Hsiao S. S., Johnson K. O., Niebur E.: Attention modulates synchronised neuronal firing in primate somatosensory cortex. Nature, **404**, 2000, pp. 187–190.

[10] Kazanovich Y. B., Borisyuk R. M.: Dynamics of neural networks with a central element. Neural Networks, **12**, 1999, pp. 441–454.

[11] Borisyuk R., Kazanovich Y.: Oscillatory neural network model of attention focus formation and control. BioSystems, **71**, 2003, pp. 29–38.

[12] Borisyuk R., Kazanovich Y.: Oscillatory Model of Attention-Guided Object Selection and Novelty Detection. Neural Networks, **17**, 2004, pp. 899–910.

[13] Wang D. L.: Object selection based on oscillatory correlation. Neural Networks, **12**, 1999, pp. 579–592.

[14] Corchs S., Deco G.: A neurodynamical model for selective visual attention using oscillators. Neural Networks, **14**, 2001, pp. 981–990.

[15] Grossberg S., Versace M.: Spikes, synchrony, and attentive learning by laminar thalamocortical circuits. Brain Research, **1218**, 2008, pp. 278–312.

[16] Katayama K., Yano M., Horiguchi T.: Neural network model of selective visual attention using Hodgkin-Huxley equation. Biol. Cybern., **91**, 2004, pp. 315–325.

[17] Ursino M., La Cara G. E.: Modeling segmentation of a visual scene via neural oscillators: fragmentation, discovery of details and attention. Network, **15**, 2004, pp. 69-89.

[18] Wang D.-L., Terman D.: Image segmentation based on oscillatory correlation. Neural Computation, **9**, 1997, pp. 805–836.

[19] Hodgkin A. L., Huxley A. F.: A quantitative description of membrane current and its applications to conduction and excitation in nerve. J. Physiol., **117**, 1952, pp. 500–544.

[20] Pylyshyn Z. W.: Visual indexes, preconceptual objects, and situated vision. Cognition, **80**, 2005, pp. 127–158.

[21] Intriligator J., Cavanagh P.: The spatial resolution of visual attention. Cogn. Psychol., **43**, 2001, pp. 171–216.

[22] Egly R., Driver J., Rafal R. D.: Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. J. Exp. Psychol., General, **123**, 1994, pp. 161–177.

[23] Alvarez G. A., Scholl B. J.: How does attention select and track spatially extended objects? New effects of attentional concentration and amplification. J. Exp. Psychol., General, **134**, 2005, pp. 461–476.

[24] Usher M., Niebur E.: Modeling the temporal dynamics of IT neurons in visual search: a mechanism for top-down selective attention. J. Cogn. Neurosci., **8**, 1996 pp. 311–327.

[25] Kazanovich Y. B., Borisyuk R. M.: An oscillatory neural model of multiple object tracking. Neural Computation, **18**, 2006, pp. 1413–1440.