LETTER — **Communicated by Laurent Itti**

An Oscillatory Neural Model of Multiple Object Tracking

Yakov Kazanovich
yakov_k@impb.psn.ru
Institute of Mathematical Problems in Biology, Russian Academy of Sciences Pushchino, Moscow Region, 142290, Russia

Roman Borisyuk
rborisyuk@plymouth.ac.uk
Institute of Mathematical Problems in Biology, Russian Academy of Sciences Pushchino, Moscow Region, 142290, Russia, and Centre for Theoretical and Computational Neuroscience, University of Plymouth, Plymouth PL4 8AA, U.K.

An oscillatory neural network model of multiple object tracking is described. The model works with a set of identical visual objects moving around the screen. At the initial stage, the model selects into the focus of attention a subset of objects initially marked as targets. Other objects are used as distractors. The model aims to preserve the initial separation between targets and distractors while objects are moving. This is achieved by a proper interplay of synchronizing and desynchronizing interactions in a multilayer network, where each layer is responsible for tracking a single target. The results of the model simulation are presented and compared with experimental data. In agreement with experimental evidence, simulations with a larger number of targets have shown higher error rates. Also, the functioning of the model in the case of temporarily overlapping objects is presented.

1 Introduction

Selective visual attention is a mechanism that allows a living organism to extract from the incoming visual information the part that is most important at a given moment and that should be processed in more detail. This mechanism is necessary due to the limited processing capability of the visual system, which precludes the rapid analysis of the whole visual scene. Different types of attention are responsible for implementing different strategies of attention focus formation. Traditional theories characterized attention in spatial terms as a spotlight or a "zoom lens" that could move about the visual field focusing on whatever fell within that spatial region (Posner, Snyder, & Davidson, 1980; Eriksen & St. James, 1986). More recent theories of attention state that in some cases, the underlying units of selection are discrete objects whose selection into the focus of attention is independent of whatever location these objects occupy. This type of attention is called object-based attention (Egeth & Yantis, 1997, Scholl, 2001). For object-based attention, the limits of information processing concern the number of objects that can be simultaneously attended.

An important experimental paradigm in the study of object-based attention is multiple object tracking (MOT). In the canonical MOT experiments (Pylyshyn & Storm, 1988, Pylyshyn, 2001; Scholl, 2001), an observer views a display with m simple identical objects (up to 10 to 12 objects such as points, or plus signs, or circles). A certain subset of the objects (from 1 to $m/2$, m is supposed to be even) is briefly flashed to mark them as targets. Other objects are considered distractors. Then all objects begin moving independently and unpredictably about the screen without passing too near to each other and without moving off the display. The subjects' task is to track the subset of targets with their eyes fixed at the center of the screen. At various times during animation, one of the objects is flashed, and the observer should press a key to indicate whether this object is a target or a distractor. In other studies, the subject had to indicate all the targets using a computer mouse. It has been shown that trained subjects are quite efficient in performing MOT. Though the number of errors increases with an increasing number of targets, even for five targets, the performance level was about 85% correct target identifications. It has been argued that the results of the experiments are in agreement with the hypothesis about resource-limited parallel processing and cannot be explained entirely in terms of a serial attention scanning process, since in the latter case, sequential jumps of a single attentional spotlight from one target to another would have to be done with impossible velocities.

A large number of later studies (Yantis, 1992; Blaser, Pylyshyn, & Holcombe, 2000; Scholl & Tremoulet, 2000; Sears & Pylyshyn, 2000; Visvanathan & Mingolla, 2002; Oksama & Hyönä, 2004; Liu et al., 2005) confirmed that the early visual system is able "to individuate and keep track of approximately five visual objects and does so without using an encoding of any of their visual properties" (Pylyshyn, 2001). Visvanathan and Mingolla (2002) investigated MOT in the case when objects are allowed to overlap one another dynamically during a trial. The experiment included those with and without depth cues that signal occlusion. The results show that "although the tracking task does become more difficult ... it does not become impossible, even in the purely two-dimensional case." When occlusion cues are added to the display, MOT performance returns to the level observed for nonoverlapping objects. Observers are better in tracking targets than identifying them; the identity tends to be lost when targets come close to each other (Pylyshyn, 2004).

In the past few years, attention has become a popular field for neural network modeling. The models of attention can be subdivided into two categories. Connectionist models (Olshausen, Anderson, & Van Essen, 1993; Tsotsos et al., 1995; Moser & Silton, 1998; Grossberg & Raizada, 2000; Itti &

Koch, 2000, 2001) are based on a winner-take-all procedure and are implemented through 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. Another type of attention model is represented by oscillatory neural networks (Kazanovich & Borisyuk, 1999; Wang, 1999; Corchs & Deco, 2001; Borisyuk & Kazanovich, 2003, 2004; Katayama, Yano, & Horiguchi, 2004). They are more suitable for object-based 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.

In this letter, we present a neural network model of MOT based on the principles of oscillation synchronization and resonance. As far as we know, this is the first model of MOT and the first example of oscillatory neural network application to processing scenes with moving objects.

The MOT model design is based on our earlier published attention model with a central oscillator (AMCO) (Kazanovich & Borisyuk, 1999, 2002, 2003; Borisyuk & Kazanovich, 2003, 2004). Each element of AMCO is an oscillator whose dynamics are described by three variables: the phase of oscillations, the amplitude of oscillations, and the natural frequency of the oscillator. The interaction between oscillators is implemented in terms of phase locking, resonance, and adaptation of the natural frequency.

AMCO has a star-like architecture of connections. It contains a one-layer network of locally coupled oscillators, the so-called peripheral oscillators (POs), whose dynamics are controlled by a special central oscillator (CO) through global feedforward and feedback connections (Kryukov, 1991). The POs represent the columns of cortical neurons in the primary visual cortex (areas V1–V2) that respond to specific local features of the image. For simplicity, we use the contrast between the intensities of a pixel and the background as such features. A CO plays the role of the central executive of the attention system (Baddeley, 1996; Cowan, 1988). In AMCO, isolated objects are represented by synchronous assemblies of POs, and the focus of attention is formed by those POs that work synchronously with the CO.

In psychological literature, the central executive is considered as a system that is responsible for attentional control of working memory (Baddeley, 1996, 2002, 2003; Shallice, 2002). The question about localization of the central executive in brain structures is still obscure. For a long time, the functions of the central executive have been attributed mostly to a local region in the prefrontal cortex (D'Espasito et al., 1991; Loose, Kaufmann, Auer, & Lange, 2003), but later studies have shown that the central executive may be represented by a distributed network that includes lateral, orbitofrontal, and medial prefrontal cortices linked with motor control structures (Barbas, 2000; Andres, 2003). Recent neuroimaging data show that "different executive functions do not only recruit various frontal areas, but also depend upon posterior (mainly parietal) regions" (Collett & Van der Linden, 2002). Daffner et al. (1998) show the relative contribution of the frontal and posterior parietal regions to the differential processing of novel and target stimuli under conditions in which subjects actively directed attention to novel stimuli. The prefrontal cortex may serve as the central node in determining the allocation of attentional resources to novel events, whereas the posterior lobe may provide the neural substrate for the dynamic process of updating one's internal model of the environment to take a novel event into account. There is some evidence that in addition to neocortical areas, the hippocampus may play an important role in implementing central executive functions: the hippocampus has the final position in the pyramid of cortical convergent zones (Damasio, 1989), participates in controlling the processing of information in most parts of the neocortex (Holscher, 2003), and coordinates the work of the attention system (Vinogradova, 2001; Herrmann & Knight, 2001; Duncan, 2001).

An important property of a network with the star-like architecture is a relatively small number of connections (of the order n, where n is the number of elements in the system) in comparison to the systems with all-to-all connections (the number of connections is of the order n^2). This makes systems with a central element biologically plausible and technically feasible even in the case of large n.

Kazanovich and Borisyuk (2002) and Borisyuk and Kazanovich (2004) showed that by combining synchronization and resonance in AMCO, it is possible to select objects in the attention focus one by one in the order determined by the saliency of objects. More salient objects have the advantage in earlier selection. In section 2 we give a short description of AMCO and show how it can be used to track a single target moving among a set of distractors.

The main idea of tracking m targets is to use a network that consists of m copies of interactive AMCO with each copy tracking one particular target. When implementing this idea, the following problems have to be solved. First, one should prevent the situation where the same target is simultaneously tracked by several AMCO. Second, the model should be able to operate when objects intersect during their movements. Since objects are identical and assumed to be moving randomly and unpredictably, there is no possibility of errorless identification of a target after it has been occluded by a distractor. In this case, the best strategy for the attention system is to keep in the focus of attention on either of the two objects that have just been separated. Also, there is no possibility to recall the identity of two target objects after their occlusion. But the attention system should be able to track them both after separation. This strategy allows the attention system to permanently hold exactly m isolated objects. This is important in order to prevent the multiplication of attended objects that otherwise would take place.

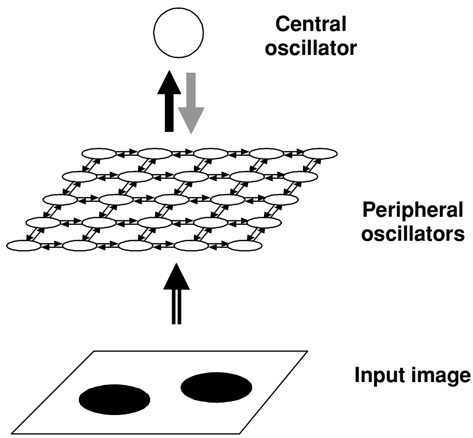

Figure 1: AMCO architecture. The hollow arrow shows the assignment of natural frequencies of POs. Black arrows show synchronizing connections that are used to (1) bind the POs coding an object into a synchronous assembly and (2) synchronize an assembly of POs with the CO. The gray arrow shows desynchronizing connections that are used to avoid simultaneous synchronization of the CO with several assemblies of POs.

In section 3 we describe how interactive AMCO layers are combined in the MOT model. In section 4 the results of computer simulation of MOT are presented. Section 5 contains discussion.

2 Tracking a Single Target

The architecture of AMCO is shown in Figure 1. The input to the network is a grayscale image on the plane grid of the same size as the grid of POs. Each PO receives an external signal from the pixel whose location on the plane is identical to the location of the PO. We use a conventional coding scheme where the external signal determines the natural frequency of oscillations (Niebur, Kammen, & Koch, 1991; Kuramoto, 1991). It is assumed that the external signal is formed in the lateral geniculate nucleus and depends on the contrast between the intensities of the pixel and the background. The value of the natural frequency of a PO is given by the formula

$$\omega_i = \lambda(B - I_i), (0 \leq I_i \leq B),$$

where ω_i is the natural frequency of the ith PO, I_i is the gray level of the ith pixel, B is the gray level of the background, and λ is a scaling parameter. Thus, the natural frequency is higher if the contrast is higher.

The POs corresponding to the pixels of objects are called active, and their dynamics are determined by equations A.1 to A.4 in appendix A. The POs corresponding to the pixels of the background are called silent. While an oscillator is silent, it does not participate in the dynamics of the system. The POs have synchronizing local connections with their nearest neighbors. These connections are used to bind the POs representing an isolated object into a coherent assembly. This is done in agreement with the synchronization hypothesis of feature binding (Singer, 1999). The CO has desynchronizing feedforward and synchronizing feedback connections to each PO. The synchronizing connections are used to phase-lock the CO by an assembly of POs. The desynchronizing connections are used to segregate different assemblies of POs in frequency space to prevent simultaneous synchronization of the CO with several assemblies of POs.

The interplay between the synchronizing and desynchronizing connections of the CO results in the competition of different assemblies of POs for the synchronization with the CO. Only one assembly of POs can win this competition; therefore, in AMCO at each moment, only one object can be attended (excluding short transitory periods). The POs representing this object work synchronously with the CO, which results in resonance: the amplitudes of these oscillators rapidly increase to a high level. The amplitudes of other POs (which do not work coherently with the CO) are shut down to a low level. A high enough amplitude of a PO is taken as a criterion that this oscillator is included in the focus of attention. The amplitude of the CO in AMCO is a constant.

Since the CO can be phase-locked only by those POs whose natural frequencies are in some range around the natural frequency of the CO, the natural frequency of the CO is adapted to its current value. As a result, the natural frequency of the CO becomes equal to the current frequency of an assembly of POs. Such adaptation allows the CO to "search" for an assembly of POs that is an appropriate candidate for synchronization.

Our previous publications on AMCO have dealt with stationary objects only. Images with moving objects are more difficult to process since the attention system should be able to properly react to the changes in object locations and also to the intersection and separation of objects during their movements. We illustrate the functioning of AMCO in the case of a visual field of size 25×50 pixels that contains nine black squares of size 7×7 on a white background.

All pixels that belong to the squares receive the same illumination and therefore have the same contrast relative to the background. To comply with the timescales that are typically used in the MOT experiments, we have selected the time unit equal to 100 msec and used gamma oscillations as the range of working frequencies. In computations, the natural frequencies of all active POs were set to $\omega_i = 5$ (5 oscillations in 100 msec), which corresponds to the frequency 50 Hz. The amplitudes of active POs have the initial value 2 and vary in the range (0, 11). The threshold for a resonant

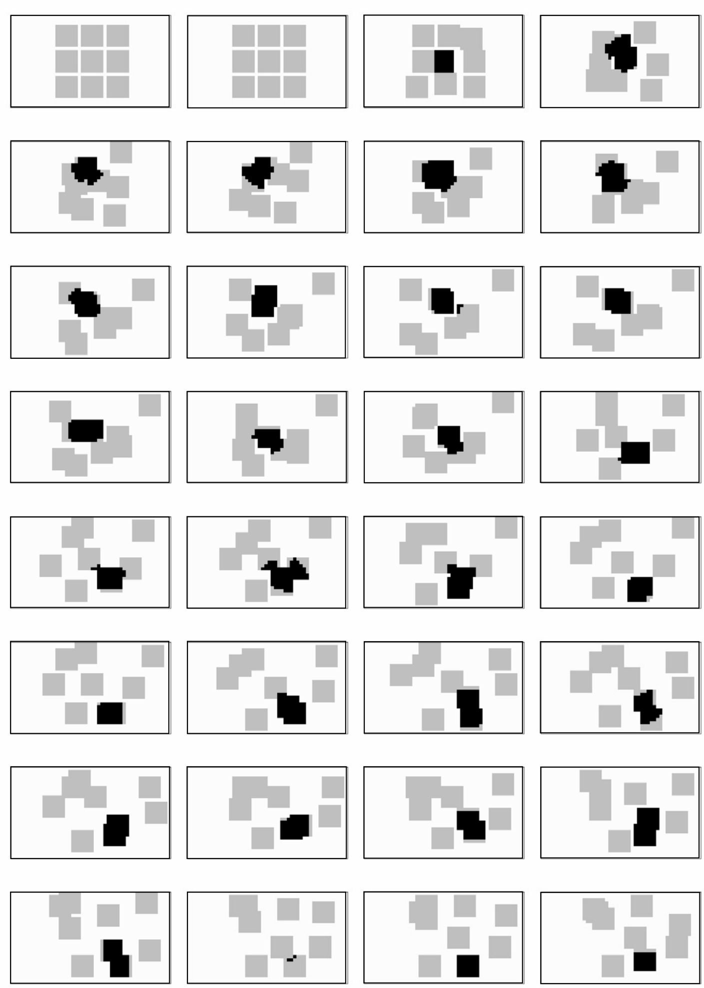

Figure 2: Single object tracking. Processing an image with a single target and eight distractors. Attended pixels are black, nonattended pixels of objects are gray, and pixels of the background are white.

amplitude is $R = 8.8$. If the amplitude of a PO exceeds R, the corresponding pixel is assumed to be included in the focus of attention.

Figure 2 displays movie frames of the dynamics of the image at the moments (in seconds) 0, 0.4, 0.8, ... (time interval between the frames is four time units). The frames are ordered from left to right and from top to bottom. The top-left frame shows the initial position of the squares. Later the squares move around in any of the four randomly chosen directions (up, down, left, right) with the probability 0.25. The choice of a new direction is taken at the moments (in seconds) 0.1, 0.2, 0.3, . . . The movement is omitted if it can lead to crossing the borders of the visual field.

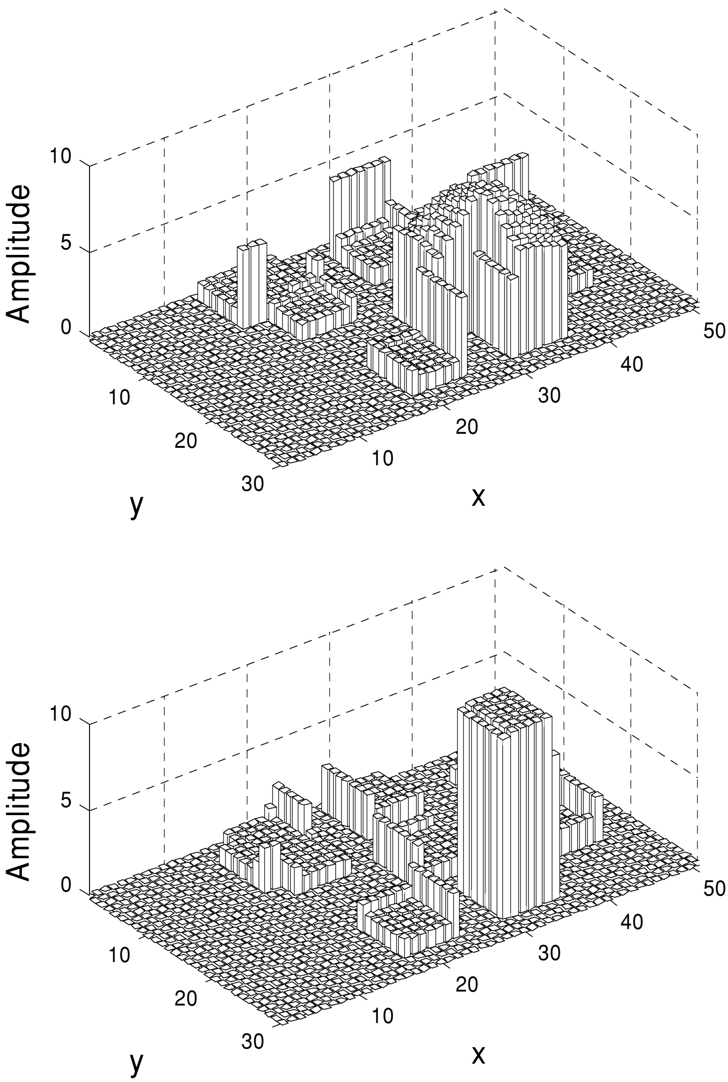

Figure 3: Amplitudes of POs at the moments (top) $t = 11.6$ (second frame in the last row of Figure 2) and (bottom) $t = 12.0$ (third frame in the last row of Figure 2).

In Figure 2, attended pixels are shown as black, nonattended pixels are gray, and pixels of the background are white. Initially the squares are regularly distributed in the image. Due to movements, the distribution of the squares around the field becomes random, and complex objects appear that represent different combinations of overlapping squares. At the initial moment, no object is under attention. After a short lag, attention is automatically focused on a randomly chosen square. In the case shown in Figure 3, it is the square in the middle of the image.

The focus of attention is firmly attached to an object while this object moves in isolation from other objects. But quite soon, the attended square crosses a complex object formed by several overlapping squares (this situation is shown in the fourth frame of the first row). Since an attended object is included in a cluster of overlapping objects, attention is gradually spread to the pixels surrounding the attended object. If a complex formed by attended objects is later split into two isolated objects, attention is focused on one of these objects. Such a movement of the attention focus along the image is reminiscent of passing a baton in a relay race, with the only difference being that the process of baton passing has a probabilistic nature.

One may think that the focus of attention nearly disappears in the second frame of the bottom row of frames and magically reappears again in the next frame. In fact, there is no magic here. The amplitudes of the oscillators in the attended area for a short time become a bit lower than the threshold R (see Figure 3, top). But soon afterward the amplitudes in the attended area again exceed the threshold (see Figure 3, bottom). At this moment, two attended squares become separate, and attention is focused on one of them. Note that in Figure 3, the columns in the diagram with the amplitudes abruptly rising over the surrounding pixels correspond to the pixels where squares have just recently moved.

3 The Model of Multiple Object Tracking

The architecture of the MOT network is shown in Figure 4. This architecture corresponds to the case of three targets (in general, the number of layers is the same as the number of targets); therefore, three layers of AMCO are shown as the components of the model. We consider these layers to be attentional subsystems. Each subsystem should track its own target. For convenience of reference, the layers are indexed by different colors: red, green and blue. The equations of dynamics of the network are shown in appendix B for the general case of m targets.

The POs that occupy the same location on the plane but belong to different layers form a column. The POs in a column are bound by strong all-to-all synchronizing connections. All POs in a column receive the same external signal from the corresponding pixel of the visual field. As in the case of a single AMCO, the external signal codes the contrast between the intensities of the given pixel and the background. This signal determines the values

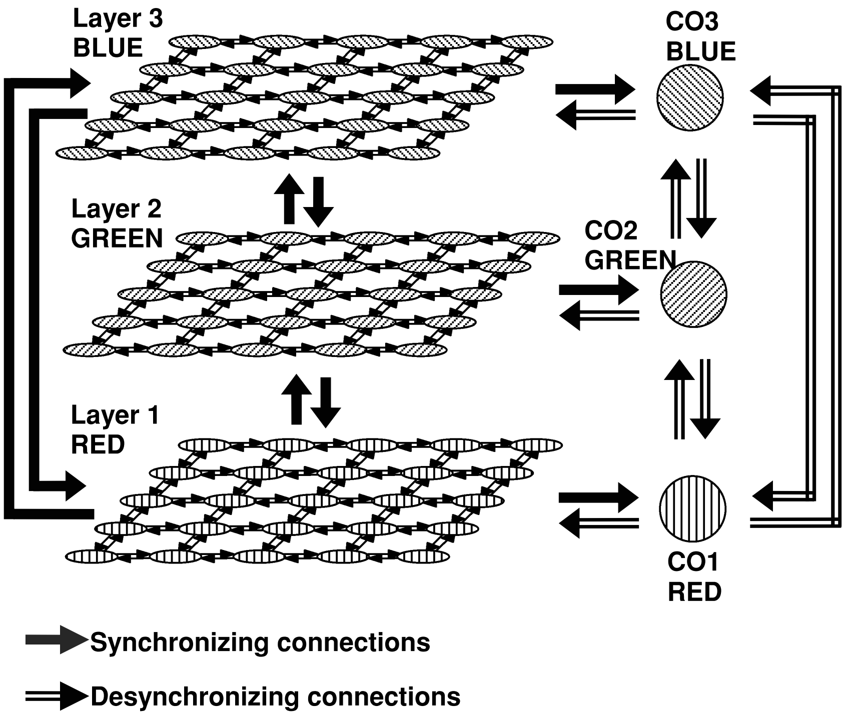

Figure 4: Architecture of the network for MOT. The layers of the network are indexed by different colors. Each layer of the network is responsible for tracking a single target.

of the natural frequencies of POs; therefore, the natural frequencies of the oscillators in a column are identical, which results in rapid synchronization of POs in a column.

The local connections between POs in a layer are restricted to the nearest neighbors. These connections are synchronizing. They are strong enough to synchronize the columns of POs that belong to an isolated object. Thus, a coherent assembly of POs is formed in the network in response to stimulation by an individual object in the visual field.

The COs belonging to different layers are bound by desynchronizing connections. Such connections are introduced to prevent the synchronization of different COs with the same synchronous assembly of POs. As a result, nonoverlapping targets are coded in the network by noncoherent oscillatory activity of different assemblies of POs.

As in the case of a single target, if an assembly of POs in the kth layer works synchronously with the CO in this layer, the amplitudes of these oscillators go up and exceed the threshold for the resonance. If all POs of the kth layer corresponding to the pixels of an object are in the resonant state, this is interpreted as the fact that this object is included in the focus of attention of the kth attentional subsystem.

If objects move slowly enough and do not overlap during their movements, the attention focus (after being formed) is rather stable due to the resonance of POs included in the focus of attention. Resonant oscillators have a much stronger influence on the CO of their layer, which prevents a jump of attention to an assembly of nonresonant oscillators. If the speed of object movements is high relative to the rate of the processes of synchronization and resonance, attention can spontaneously switch from one object to another. This results in errors in distinguishing between targets and distractors.

Consider what happens to the attention focus if two objects cross each other. If both objects are unattended, no change of the focus of attention takes place. Temporarily a complex distractor (composed of two objects) is formed, but this has no effect on objects in the focus of attention.

In the case when attended and unattended objects overlap, attention is spread to the whole composite object because all POs belonging to this object will be included in a common assembly of synchronized oscillators. This assembly will work synchronously with the same CO that was synchronous with the attended assembly of POs before overlapping. After becoming separate again, the objects renew their competition for being included in the focus of attention. Due to the desynchronizing influence of the CO on POs, only one of the two objects is able to win the competition. Since the whole composite object has been temporarily attended, the system has no information on how to detect which object had been attended before the intersection took place. In this situation, either of the two objects can be newly selected in the focus of attention. The choice is random; therefore, it may lead to an error in target identification with the probability 0.5.

If two attended objects overlap, both continue to be in the focus of attention despite the fact that there is a desynchronizing connection between the COs. This is achieved by making this desynchronization weak enough relative to the synchronizing influence that comes to both COs from the common assembly of synchronous POs. When objects move apart, the desynchronizing connection between the COs renews its influence on these oscillators. As a result, each object will again be tracked by its own AMCO layer. Of course, it is possible that objects, say A and B, that before the intersection have been tracked by, for example, "red" and "green" layers, will exchange their tracking systems: after separation, the "red" layer will be used to track the object B, and the object A will be tracked by the "green" layer. Whether this exchange will happen depends on how strong the overlapping has been and how long it has continued.

In fact, the interplay between the synchronizing and desynchronizing influences on a CO is even more intricate than the one we have just described. Computation experiments have shown that a constant desynchronizing interaction between COs cannot ensure the proper behavior of the COs. One type of error appears if the desynchronizing interaction between COs is too strong. In this case, a CO may lose the synchronization with the assembly of POs at the moment when two attended objects overlap. As a result, only one CO will maintain synchronization with a composite object formed by two previously attended objects. Another type of error may appear if the strength of desynchronization between the COs is too weak. In this case, two COs may maintain the synchronization with the same object after the separation of two simultaneously attended objects. In fact, no constant value of the connection strength between the COs allows the system to avoid one or other of these errors. But the problem can be solved if the interaction between the COs increases after two attended objects overlap. It has been done by using the idea of resonance but applying it now to the amplitudes of COs. Therefore in the MOT model, the amplitudes of COs are not constants anymore.

A resonant increase of the amplitude of a CO takes place if two attended objects cross each other (see equation B.5 in appendix B). According to the last term in equation B.1, this results in increasing the strength of desynchronization between the COs that track these objects. Therefore, at the moment of separation of the objects, the strength of desynchronization will be high enough to prevent the situation when both COs track one object leaving the other one outside the focus of attention.

To follow the experimental conditions of MOT, the model should not only be able to track moving objects but should also make a proper choice of a set of targets at the initial phase of MOT. In the experiments, targets are indicated to the observer by a brief flash of light. In the model, the notion of saliency is used to formalize the choice of flashed objects in the focus of attention. It is assumed that flashed objects are more salient than other objects and that this leads to automatic attraction of attention to these objects. In terms of the model, the salience of a pixel influences the strength of the influence of POs corresponding to this pixel on the COs. Thus, more salient objects have an advantage in being included in the focus of attention.

The idea of the saliency map was introduced by Koch and Ullman (1985) and has been intensively used in computational models of visual search (Itti, Koch, & Niebur, 1998; Itti & Koch, 2000, 2001; Olshausen et al., 1993). The saliency map is a two-dimensional table that encodes the saliency of objects in the visual environment and determines the priority of their choice in the focus of attention. In the MOT model, the saliency map is formed as a set of parameters s_i that correspond to the pixels of the image and determine the strength of influence of POs on COs as is shown in equation B.1 in appendix B. To reflect the difference in saliency between flashed and nonflashed objects, the saliency s_i takes one of two positive values: a higher value $S_{flashed}$ corresponds to the pixels of flashed objects, and a lower value $S_{nonflashed}$ corresponds to the pixels of nonflashed objects. For the pixels of the background, $s_i = 0$. The value of $S_{flashed}$ should be several times higher than $S_{nonflashed}$ to provide the assemblies of POs that represent flashed targets with a much better chances of winning the competition for synchronization with the COs than the assemblies corresponding to nonflashed distractors. In computer simulations, we put $S_{flashed} = 5$ and $S_{nonflashed} = 0.2$.

When flashing is over, all objects become equally salient. This is reflected in the model by making all values of s_i for the pixels of objects identical. In simulations, $s_i = 1$.

4 Model Simulation and Comparison with Experimental Data

Two types of MOT model simulations are considered below that correspond to movements with and without overlap, respectively. Computations of the fist type are used to compare the performance of the model with recent experimental data of Oksama and Hyönä (2004). In simulations of the second type, overlapping objects are used to demonstrate that the model follows the procedure described in section 3.

Oksama and Hyönä (2004) experimented with a set of 12 objects. To accelerate computation, we reduced the number of objects to 10 as in the experiments of Pylyshyn and Storm (1988). This does not significantly affect the results. In simulations, objects are black squares of size 7×7 pixels on a white background in a field of size 30×60 pixels. As in section 2, the pixels of the squares are coded in the network by the natural frequencies of POs equal to 50 Hz. Tracking of k targets ($2 \leq k \leq 5$) implies that a network with k layers is used.

The timescale for simulations has been chosen as in section 2. A single run of the model takes 7.2 sec and consists of three phases. The first phase takes 0.7 sec and is used for marking the targets. During this period, objects are motionless; the only distinction between targets and distractors is their saliency, as explained in section 3. The desirable result at the end of this phase is the focusing of attention on the targets so that each target is attended by one attentional subsystem. An error may appear if two attentional subsystems are focused on the same target or if a distractor is chosen in the focus of attention of one of the subsystems. Computations have shown that the probability of such errors is less than 0.005; therefore, the initial acquisition of targets in the focus of attention is nearly errorless.

The second phase continues 6 sec. This is when objects move in a random manner. The speed of motion is 1 pixel per 50 msec, that is, every 50 msec, all squares make a movement of 1 pixel length in one of the four directions: up, down, left, or right. For each square, the direction of the movement is chosen randomly and independently with the probability 0.25. To prevent collisions, the motion is subject to the restriction that the squares should be always separated by at least one pixel of the background. First, the horizontal or vertical direction of the movement is chosen with the probability 0.5. Then, for horizontal movements, the left or right direction is taken with the probability 0.5, and for vertical movement, the up or down direction is taken with the probability 0.5. If there is a danger of collision, the direction of the movement is reversed. If the danger of collision exists for both opposite directions, no movement is undertaken at the current moment. The same rules are applied to prevent the squares from crossing the borders of the visual field.

The third phase takes 0.5 sec. During this phase, all movements are stopped. This time is given to the system to resolve an ambiguous situation when several objects are simultaneously attended by an attentional subsystem. Such a situation appears from time to time if objects are moving. When objects are stationary, the time interval of 0.5 sec is long enough for the attentional subsystem to choose which of these objects should be kept in the focus of attention. Other objects are automatically excluded from the focus of attention as a result of desynchronizing influence of the CO on POs in the corresponding layer.

At the final moment of the third phase, the number of object identification errors is registered. According to the principles of system design and functioning, at this moment exactly k squares are attended by the network with k layers. Some of the attended squares are targets, but some of them may be distractors due to errors in attention focusing during object movements. Therefore, two types of error may appear: a target is identified as a distractor, or a distractor is identified as a target. According to the strategy implemented in the model, the number of errors is always even: if there is an error in attending a target (a target is missed by all attentional subsystems), this inevitably results in taking a distractor in the focus of attention.

To estimate the performance of the model, we executed 50 runs of the system for each target set size $k = 2, 3, 4, 5$. The results of computations are shown in Table 1. The analysis of variance test (ANOVA) has been used to check whether the means of 50 trials in four groups (corresponding to different number of targets) differ. The null hypothesis is that all the groups are drawn from the same probabilistic distribution (or from different distributions but with the same mean). The result is: $F = 43.7$, and the p-value is less than 0.0001; therefore, the results of our simulations do not support the null hypothesis. Further analysis by use of the pair-wise T-test gives the following results: $T_{23} = 4.6$, $T_{34} = 3.7$, $T_{45} = 3.1$. These values of the T-test do not support the null hypothesis that the mean of group k equals the mean of group $(k - 1)$ for $k = 3, 4, 5$. Therefore, the alternative hypothesis that the mean in group k is larger than the mean in group $(k - 1)$ is supported.

In experiments, Oksama and Hyönä (2004) estimated human performance by using a probe object that should be identified by the subject as a target or a distractor. To exclude a bias in guessing, a probe object was selected with the probability 0.5 from the set of targets and from the set of distractors. In contrast to Oksama and Hyönä, we have not made direct computational experiments with probe objects but estimated the

Table 1: Results of Identification of Targets and Distractors in the MOT Simulation.

	Target Set Size			
	2	3	4	5
Number of errors	2	38	86	134
Mean number of errors per trial	0.04	0.76	1.72	2.68
Standard deviation	0.3	1.1	1.5	1.6
Probability of error checked by probe objects	0.006	0.09	0.179	0.268

probability of error in probe identification based on the number of errors in each run.

Let s be the number of objects ($s = 10$), k the number of targets ($k = 2, 3, 4, 5$), and e the number of the targets that have been mistakenly identified as distractors in a run of simulations (hence, the same number of distractors has been mistakenly identified as targets). Then the probability of error if checked by a probe object is

$$P = 0.5\left(\frac{e}{k} + \frac{e}{s-k}\right) = \frac{0.5se}{k(s-k)}.$$

Using this formula, we computed the values of P for each run and averaged these values through all simulation trials. The results are presented in the last row of Table 1 and in Figure 5. For comparison Figure 5 also shows the accuracy of humans in MOT (Oksama & Hyönä, 2004). Both error patterns in Figure 5 show similar behavior: the probability of error increases with a larger number of targets. The main difference is that the probability of error for the case of two targets is underestimated in simulations in comparison to the experimental data.

To illustrate the functioning of the system in the case when objects are permitted to overlap during their movements, we present a computational experiment where six identical objects (three targets and three distractors) are moving in the visual field. Again objects are black squares of size 7×7 pixels on a white background. The size of the field is 19×62 pixels.

A session of simulation is divided into two phases. The first short phase (1.2 sec) is used to mark the targets and select them into the attention focus. At this phase, all squares are isolated and motionless. The selection is done exactly in the same way as in the case with nonoverlapping objects.

During the second phase lasting for 9.6 sec, objects start moving according to the rules that are the same as in the case of a single target (see

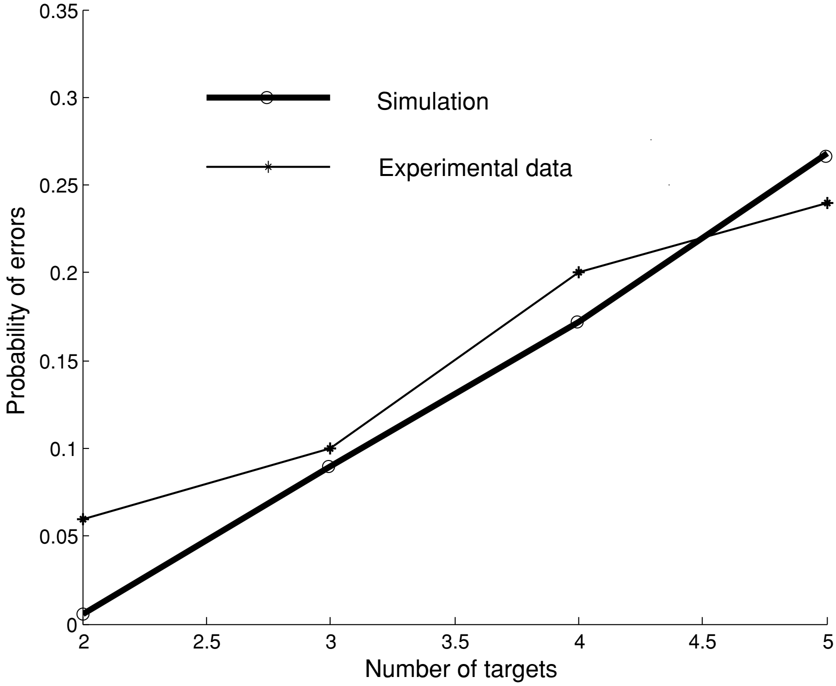

Figure 5: Accuracy of probe objects identification by the MOT model in comparison with humans. Error rates are shown as a function of number of targets tracked. Experimental data are taken from Oksama and Hyönä (2004).

section 2). The movie frames in Figure 6 illustrate the dynamics of the image and the process of attention focus formation and switching. The frames are ordered from left to right and from top to bottom. The time interval between the frames is four time units (0.4 sec). The top-left frame shows the initial position of objects in the image. The next frame shows a moment during exposition time when the attention focus with three targets is formed. Later, objects begin their random movements, which lead to the formation of different combinations of overlapping objects.

Colors in Figure 6 do not reflect the colors of the squares in the image (we have already mentioned that all squares in the image are black). Colors are used to show which objects are under the attention of different attentional subsystems (network layers). The color of a pixel depends on the state of the oscillators in the column that corresponds to this pixel. If a PO in the "red" layer is in the resonant state, then the pixel is red. A similar principle of color assignment is applied to green and blue colors. If several POs in a column are in the resonant state, the color of the pixel is a mixture of the basic colors. For example, the pixel is cyan if it is simultaneously attended

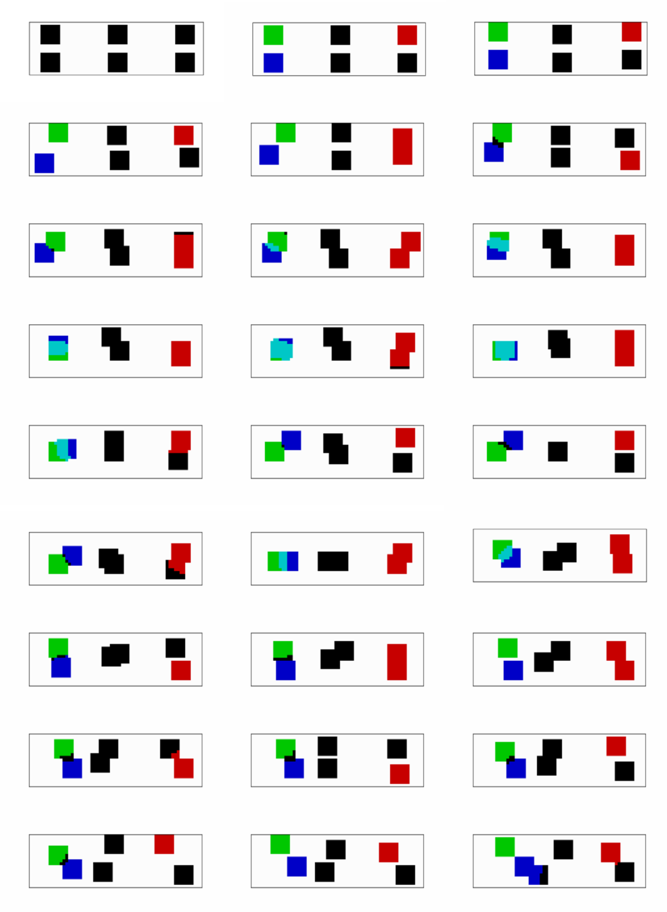

Figure 6: Multiple object tracking. Processing an image with three targets and three distractors. Each pixel is painted in red, green, blue, or a mixture of these colors. A pixel is red/green/blue if a "red"/"green"/"blue" oscillator in the column that corresponds to this pixel is in the resonant state. A pixel is cyan if both oscillators in the "blue" and "green" layers are in the resonant state. A pixel is black if no oscillator in the column is in the resonant state. A pixel is white if it belongs to the background.

in the "green" and "blue" layers. Black is used for nonattended pixels of objects, and white is used for the pixels of the background. The intensities of green, red, and blue colors are such that their combination in one pixel results in gray. (In fact, there are no such pixels in Figure 6.)

Consider what happens with the attention focus while squares are moving. The pair of squares in the middle of the image is always outside the attention focus. These squares are not attended in both cases when they move as isolated objects or temporarily form a composite object. The POs representing the pixels of these objects always work with low amplitude.

The pair of squares in the right part of the image shows the process of attention switching in the case when a complex object is formed due to overlapping of attended and nonattended squares. If the time of overlapping is short, after separation of the squares, attention focusing is correct in choosing the square that was under attention before. If the area of overlapping becomes large and the duration of existence of the complex object is not very short, attention is spread over the entire composite object. Therefore, after separation of the objects, either of them may be kept in the focus of attention, while the other is excluded from the focus of attention. In fact, the functioning of the MOT model in this case is not different from how AMCO works in the case of a single target. Computational experiments confirm that the much more intricate architecture of the MOT model does not lead to any additional complications in tracking a single target among distractors by a layer of the network.

Finally, let us consider how the pair of squares in the left part of the image is processed by the system. Both squares are marked as targets and selected into the focus of attention. While these squares move separately, the attentional subsystem (network layer) tracking each object does not change; therefore, the colors of the squares (green and blue) in Figure 6 are kept unchanged during this period. The situation changes after a large enough area of overlapping between the squares appears. This area is painted in cyan because it is simultaneously attended by two attentional subsystems, "green" and "blue." Due to object movements, the size of the cyan area may increase or decrease; it may disappear if the overlapping area is too small and reappear again when the overlapping area becomes large enough. The important fact is that as soon as the squares separate in the image field, each of them is tracked by a single attentional subsystem. The attentional subsystems may exchange targets only if at some moment, the overlapping of the squares does not allow the reliable identification of each square in the composite object.

A thorough quantitative investigation of error rates in MOT with overlapping objects is in progress. Preliminary simulations have shown that error rates essentially depend on the speed of object movements. If the speed is low enough, as in the experiment presented above, the probability of error (except those errors that are inevitable after strong overlapping of a target and a distractor) is rather low.

5 Discussion

Three main ideas are combined in the presented model of attention. First, oscillations and synchronization are used as a key mechanism in attention focus formation. The evidence that oscillatory activity and long-range phase synchronization play a major role in the attentional control of visual processing has been provided in studies of EEG (Herrmann & Knight, 2001), MEG (Sokolov et al., 1999; Gross et al., 2004), and local field potential recordings (Steinmetz et al., 2000; Fries, Reynolds, Rorie, & Desimone, 2001; Fries, Schroeder, Roelfsema, & Engel, 2002; Niebur, Hsiao, & Johnson, 2002; Fell, Fernandez, Klaver, Elger, & Fries, 2003; Liang, Bressler, Ding, Desimone, & Fries, 2003). In particular, it has been shown that modulation of neural synchrony determines the outcome of an attention-demanding behavioral task (Gross et al., 2004; Tallon-Baudry, 2004). Second, it is assumed that attention focusing is controlled by a special neural system, the so-called central executive (Baddeley, 1996; Cowan, 1988). In terms of the model, visual attention is normally characterized by synchronous activity of an assembly of neurons that represent the central executive and an assembly of neurons in the primary areas of the visual cortex. MOT represents a special situation when attention is distributed among several isolated objects. It is supposed that in this case, the central executive is split into several subsystems whose activity is desynchronized. Thus, synchronous oscillations are used to label different objects that are simultaneously included in the focus of attention. Third, the resonance is used to formalize the assumption that attention enhances neural activity in attended areas and inhibits responses to nonattended stimuli.

The relation between the resonance as it is used in the model and attentional modulation of cortical activity should be explained in more detail. Though electrophysiological (Motter, 1993; Roelfsema, Lamme, & Spekreijse, 1998) and functional imaging studies (Somers, Dale, Seifert, & Tootel, 1999; Kanwisher & Wojciulik, 2000; Seifert, Somers, Dale, & Tootel, 2003) have shown that attentional modulation can be found as early as in the primary visual cortex, it is known that the strength of attentional effects increases as one moves up the cortical processing hierarchy (Treue, 2003). Moreover, Culham et al. (1998) in the fMRI studies of activation produced by attentive tracking of moving objects have found no enhancement in early visual cortex, but it has been shown that the signal more than doubles in parietal and frontal areas. How do these experimental results comply with the model? In the model, direct connections are used from POs to COs. This is a radical simplification of the real situation. In fact, there are many intermediate cortical structures in the pathway from the striate cortex to the higher regions occupied by the central executive. At the moment the flow of information reaches its final station, the difference in the activity of its attended and unattended components becomes quite clear. Still, this difference is not as high as the difference between resonant and nonresonant oscillations in the model. Therefore, one should not literally think that the amplitude of oscillations in the model is a relevant characteristic of the activity in the cortex as it is observed, for example, in fMRI studies. The amplitude of oscillations should be considered as a formal variable that positively correlates with the activity in the cortex and determines the strength of interaction between cortical oscillators and the central executive.

The theoretical explanation of MOT is based on the idea of preattentive assignment of an index to objects tracked (Pylyshyn & Storm, 1988; Pylyshyn, 2001). It is supposed that this indexing can occur independently and in parallel at several places in the visual field. In contrast to this theory, in our model indexing is implemented in two stages, which include both preattentional and attentional levels. During the first stage, an oscillatory label is assigned to each object. All information related to a given object is coded in the form of synchronous (coherent) oscillations, while oscillations corresponding to different objects are incoherent. Although the oscillatory label is not something constant but varies with time, it provides a reliable mechanism of distinguishing among identical targets. The oscillatory label is used in the second stage when an attentional subsystem is associated with a single target, the one that is tracked by this subsystem. The number of this subsystem can be considered as the index of the target. The difference between attended and unattended objects is realized in the form of synchronization or desynchronization of the activity of assemblies of POs with the COs.

The model has a rigid architecture of connections and predetermined interaction functions and parameters. Real biological systems for MOT must be much more flexible. A question that may arise is how the system with a fixed number of layers can adapt to tracking different number of targets. A trivial solution is to assume that the number of active COs is controlled by internal effort and is always set equal to the number of targets tracked. We think that a more plausible solution is a flexible type of interaction among the COs. Suppose that the number k of the COs is restricted (according to the experimental evidence $k = 5$ is a reasonable upper bound for the number of targets tracked simultaneously), but the type of interaction between the COs can vary depending on the task. If one target should be tracked, all the COs are bound into a synchronous assembly by synchronous interaction among the COs. In the case of two targets, two assemblies of COs are formed with synchronizing interaction among the COs in an assembly and desynchronizing interaction among the assemblies. This situation is equivalent to the case of the model with two COs. Similarly, an arbitrary number of targets below five can be tracked by grouping the COs into a proper number of assemblies.

The dependence of number of errors in MOT on the number of targets was the reason for associating MOT with resource-limited parallel processing (Pylyshyn & Storm, 1988). Our model presents an alternative explanation of this phenomenon. Although the processing of information in our model is purely parallel, it has been shown by computer simulations that tracking of the targets will be poorer if the number of targets increases. This is caused by the limited capacity of the phase space where several central oscillators have to operate simultaneously. Increasing the number of central oscillators makes it more and more difficult for them to avoid temporal synchronization, which may result in unpredictable jumps of attention to nontarget objects. If the number of targets is below five, the probability of such jumps is very low for stationary objects, but it significantly increases when objects start moving with high speed. Due to movements, the processes of synchronization and resonance have not enough time to fully proceed, and these result in the loss of synchronization of a CO with a previously selected object.

Although we have chosen the parameters of the model in such a way that it closely follows real-time relations in the experiments with MOT, one should not be too serious about this fact. The model is too simple to be realistic in this respect. A very small number of pixels used for objects representation, restriction of POs interaction to nearest neighbors, and many other features of the model are conditioned by the need to make computations in a reasonable amount of time. In fact, the model is rather flexible in its operation time. Another choice of parameters, or even the duration of the time unit, can lead to other time relations. Therefore, when comparing the performance of the model with human error rates, we used the experimental data averaged through time periods of 5, 9, and 13 sec. The data for 5 sec in the experiments with humans give lower values than those obtained in our simulations, but the pattern of error probabilities is the same. The improvement of the model in timescale representation is planned. In particular, the results of Oksama and Hyönä (2004) on nonlinear variation of error rates with the duration of trials are a challenge for the future development.

The decision-making strategy used in the model is also oversimplified. The model is forced to always track a fixed number of objects. The experiments show that human strategy is more clever (Pylyshyn & Storm, 1988; Oksama & Hyönä, 2004). If a subject has a feeling that correct identification of an object during tracking is doubtful, he or she is inclined to stop tracking this object and focus attention on tracking a smaller number of objects. This causes gradual reduction of human performance when the number of targets exceeds five and is probably the source of smaller differences in the probability of error between the cases with four and five targets observed in MOT experiments relative to those found in simulations (see Figure 5).

It is known that in MOT experiments, the quality of target tracking can be enhanced by grouping the targets into a virtual polygon and then tracking deformations of the polygon. This grouping can be done spontaneously or according to the instruction given to subjects (Yantis, 1992). These facts can be explained in terms of oscillatory neural networks by assuming that in this case, all targets are combined in a single visual object. The oscillators representing this object are synchronized along virtual borders of the polygon that are formed by some internal effort. As a result, attention is no longer divided, and the central executive operates in a standard manner as a single central oscillator. It is also possible that humans can follow some mixed strategy combining grouping with tracking individual objects. The strategy of grouping may be an explanation for the cases when more than five targets are tracked successfully. But according to the data of Oksama and Hyönä (2004), if the number of targets exceeds five, the subjects are inclined to ignore some targets and focus attention on a smaller target set.

Designing a MOT model, we intentionally tried to avoid the use of traditional image processing techniques such as shape analysis, connectivity testing, pattern recognition, and others. Therefore, the model can work equally well when objects in the visual field are not identical or even vary in shape. This is important, for example, if object movements take place in 3D space with the projection of objects on the retina constantly changing. The model design is also caused by the fact that complex procedures of information processing demand more time, and are assumed to be implemented by higher cortical structures. It was interesting to investigate whether MOT can be explained in terms of a simple network architecture where the only function of the top-down flow is to control the focus of attention.

Appendix A: Mathematical Description of AMCO

The oscillators comprising AMCO are described as generalized phase oscillators. The state of such an oscillator is described by three explicitly given variables: the phase of oscillations, the amplitude of oscillations, and the natural frequency of oscillations.

The dynamics of AMCO are described by the following equations.

$$\frac{d\theta_0}{dt} = 2\pi\omega_0 + \frac{w_0}{n}\sum_{i=1}^{n} s_i a_i g(\theta_i - \theta_0) \tag{A.1}$$

$$\frac{d\theta_i}{dt} = 2\pi\omega_i - a_0 w_1 h(\theta_0 - \theta_i) + w_2 \sum_{j \in N_i} a_j p(\theta_j - \theta_i) + \rho \tag{A.2}$$

$$\frac{da_i}{dt} = \beta(-a_i + \gamma f(\theta_0 - \theta_i)) \tag{A.3}$$

$$\frac{d\omega_0}{dt} = -\alpha\left(2\pi\omega_0 - \frac{d\theta_0}{dt}\right). \tag{A.4}$$

In these equations, θ_0 is the phase of the CO; $\theta_i \quad (i = 1, \ldots, n)$ 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 constant); a_i are the amplitudes of oscillations of POs; w_0, w_1, w_2 are constant positive parameters that control the strength of interaction between oscillators; s_i is the parameter that distinguishes between active and silent oscillators; $s_i = 1$ if the PO is active, otherwise $s_i = 0$; N_i is the set of active POs in the nearest neighborhood of the oscillator i; ρ is a gaussian noise term with mean 0 and standard deviation σ; functions g, h, p control the interaction between oscillators; f is a function that controls the amplitude of oscillations of POs and their transition to the resonant state; and α, β, γ are network parameters (positive constants). The values ω_i are determined by the input signal; $\theta_0, \theta_i, \omega_0, a_i$ are internal variables that characterize the state of the network.

The functions g, h, p are 2π-periodic, odd, and unimodal in the interval of periodicity. f is 2π-periodic, even, positive, and unimodal in the interval of periodicity. An exact description of these functions and the values of the parameters that are used in computations can be found in Borisyuk and Kazanovich (2004).

Equations A.1 and A.2 are traditional equations of phase locking. They correspond to the architecture of connections of Figure 1 and control the processes of synchronization and desynchronization in the network. Equation A.1 describes the dynamics of the CO. Equation A.2 describes the dynamics of POs. The noise ρ in equation A.2 is used as an additional source of desynchronization between assemblies of POs. It helps to randomize the location of different assemblies of POs in phase-frequency space, thus making them distinguishable by the CO.

Equation A.3 describes the dynamics of the amplitude of oscillations of POs. This equation provides a mechanism for the resonant increase of the amplitude of oscillations. Let the interval of variation of f be $(f_{\min}, f_{\max})$. The amplitude of a PO increases to the maximum value $a_{\max} = \gamma f_{\max}$ if the PO works synchronously with the CO. The amplitude takes a low value $a_{\min} = \gamma f_{\min}$ if the phase of the PO is significantly different from the phase of the CO. We say that a PO is in the resonant state if its amplitude exceeds the threshold $R = 0.8\gamma f_{\max}$. The parameter β determines the rate of amplitude increase and decay.

Equation A.4 describes the adaptation of the natural frequency of the CO. According to this equation, the value of $2\pi\omega_0$ tends to the current frequency of the CO. Such adaptation allows the CO to "search" for an assembly of POs that is an appropriate candidate for synchronization.

Appendix B: Mathematical Description of the MOT Model

The equations of the MOT model dynamics represent a modification of equations A.1 to A.4 according to the scheme of Figure 4:

$$\frac{d\theta_0^k}{dt} = 2\pi\omega_0^k + \frac{w_0}{n_{res}} \sum_{i=1}^{n} s_i a_i^k g\left(\theta_j^k - \theta_0^k\right) - w_3 \sum_{l=1}^{m} a_0^l q\left(\theta_0^l - \theta_0^k\right) \tag{B.1}$$

$$\frac{d\theta_i^k}{dt} = 2\pi\omega_i^k - a_0^k w_1 h\big(\theta_0^k - \theta_i^k\big) + w_2 \sum_{j \in N_i} a_j^k p\big(\theta_j^k - \theta_i^k\big) + \frac{w_4}{m}\sum_{l=1}^{m} a_i^l p\big(\theta_i^l - \theta_i^k\big) + \rho \tag{B.2}$$

$$\frac{da_i^k}{dt} = \beta\big(-a_i^k + \gamma f\big(\theta_0^k - \theta_i^k\big)\big) \tag{B.3}$$

$$\frac{d\omega_0^k}{dt} = -\alpha\left(2\pi\omega_0^k - \frac{d\theta_0^k}{dt}\right). \tag{B.4}$$

In these equations, upper and lower indices are used to number layers and oscillators in a layer, respectively; $k = 1, \ldots, m$, where m is the number of layers (the same as the number of targets in MOT). The normalizing parameter n_{res} is equal to the current number of resonant oscillators, but not less than 49 (the number of pixels in an object). The last term in equation B.1 describes the interaction between the central oscillators, and the function q determines the type of interaction (the negative sign before this term makes the interaction desynchronizing). The term before the noise in equation B.2 gives the interaction in a column of POs. Parameters w_3, w_4 are positive constants.

The parameters s_i in equation B.1 form a saliency map, $s_i > 0$, for the pixels of objects and $s_i = 0$ for the pixels of the background. During the stage when some objects are flashed, the values of s_i are made higher for the pixels of flashed objects than for the pixels of nonflashed objects. When objects are homogeneously illuminated, the values of s_i are made identical for all pixels of objects.

Equations B.3 and B.4 generalize equations A.3 and A.4 for the amplitudes of POs and the natural frequency of COs in the case of a multilayer network.

The amplitudes of COs vary according to an equation similar to equation B.3,

$$\frac{da_0^k}{dt} = \beta\left(-a_i^k + \gamma_1 r\left(\sum_{l=1,\, l\neq k}^{m} f\big(\theta_0^l - \theta_0^k\big)\right) + 1\right), \tag{B.5}$$

where

$$r(x) = \begin{cases} x, & x \le f_{\max}, \\ f_{\max}, & x > f_{\max}. \end{cases}$$

The function r is necessary to normalize the variation of the amplitudes of COs. In computations, the values of the resonant amplitude of a CO were about two times higher than the amplitude of a nonresonant CO.

Acknowledgments

This work was supported by the Russian Foundation of Basic Research (Grants 03-04-48482 and 06-04-48806) and by the UK EPSRC (Grant EP/0036364).

References

Andres, P. (2003). Frontal cortex as the central executive: time to revise our view. *Cortex, 39*, 871–895.

Baddeley, A. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology, 49A*, 5–28.

Baddeley, A. (2002). Fractionating the central executive. In D. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 246–260). New York: Oxford University Press.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*, 189–208.

Barbas, H. (2000). Connections underlying the synthesis of cognition, memory, and emotion in primate prefrontal cortices. *Brain Research Bulletin, 52*, 319–330.

Blaser, E., Pylyshyn, Z. W., & Holcombe, A. O. (2000). Tracking an object through feature space. *Nature, 408*, 196–199.

Borisyuk, R., & Kazanovich, Y. (2003). Oscillatory neural network model of attention-focus formation and control. *BioSystems, 71*, 29–36.

Borisyuk, R., & Kazanovich, Y. (2004). Oscillatory model of attention-guided object selection and novelty detection. *Neural Networks, 17*, 899–915.

Collett, F., & Van der Linden, M. (2002). Brain imaging of the central executive component of the working memory. *Neuroscience and Behavioral Review, 26*, 105–125.

Corchs, S., & Deco, G. (2001). A neurodynamical model for selective visual attention using oscillators. *Neural Networks, 14*, 981–990.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention and their mutual constraints within the human information processing system. *Psychological Bulletin, 104*, 163–191.

Culham, J., Brandt, S. A., Cavanagh, P., Kanwisher, N. G., Dale, A. M., & Tootell, R. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. *Journal of Neurophysiology, 80*, 2657–2670.

Daffner, K. R., Mesulam, M. M., Scinto, L. F. M., Cohen, L. G., Kennedy, B. P., West, W. C., & Holcomb, P. J. (1998). Regulation of attention to novel stimuli by frontal lobes: An event-related potential study. *NeuroReport, 9*, 787–791.

Damasio, A. (1989). The brain binds entities and events by multiregional activation from convergent zones. *Neural Computation, 1*, 123–132.

D'Espasito, M., Detre, J. A., Alsop, D. C., Shin, R. R., Atlas, S., & Grossman, M. (1991). The neural basis of central executive system of working memory. *Nature, 378*, 279–281.

Duncan, J. (2001). An adaptive coding model of neural functions in prefrontal cortex. *Nature Reviews Neuroscience, 2*, 820–829.

Egeth, H., & Yantis, S. (1997). Visual attention: Control, representation, and time course. *Annual Review of Psychology, 48*, 269–297.

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception and Psychophysics, 40*, 225–240.

Fell, J., Fernandez, G., Klaver, P., Elger, C. E., & Fries, P. (2003). Is synchronized neuronal gamma activity relevant for selective attention? *Brain Research Reviews, 42*, 265–272.

Fries, P., Reynolds, J. H., Rorie, A. E., & Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science, 291*, 1560–1563.

Fries, P., Schroeder, J.-H., Roelfsema, P. R., Singer, W., & Engel, A. K. (2002). Oscillatory neural synchronization in primary visual cortex as a correlate of stimulus selection. *Journal of Neuroscience, 22*, 3739–3754.

Gross, J., Scmitz, F., Schnitzler, I., Kessler, K., Shapiro, K., Hommel, B., & Schnitzler, A. (2004). Modulation of long-range neuronal synchrony reflects temporal limitations of visual attention in humans. *Proc. Natl. Acad. Sci. (USA), 101*, 13050–13055.

Grossberg, S., & Raizada, R. (2000). Contrast sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research, 40*, 1413–1432.

Herrmann, C. S., & Knight, R. T. (2001). Mechanisms of human attention: Event related potentials and oscillations. *Neuroscience and Biobehavioral Reviews, 25*, 465–476.

Holscher, C. (2003). Time space and hippocampal functions. *Review of Neuroscience, 14*, 253–284.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*, 1489–1506.

Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience, 2*, 194–203.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid screen analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*, 1254–1259.

Kanwisher, N., & Wojciulik, E. (2000). Visual attention: Insights from brain imaging. *Nature Reviews Neuroscience, 1*, 91–100.

Katayama, K., Yano, M., & Horiguchi, T. (2004). Neural network model of selective visual attention using Hodgkin-Huxley equation. *Biological Cybernetics, 91*, 315–325.

Kazanovich, Y. B., & Borisyuk, R. M. (1999). Dynamics of neural networks with a central element. *Neural Networks, 12*, 441–454.

Kazanovich, Y., & Borisyuk, R. (2002). Object selection by an oscillatory neural network. *BioSystems, 67*, 103–111.

Kazanovich, Y. B., & Borisyuk, R. M. (2003). Synchronization in oscillator systems with phase shifts. *Progress in Theoretical Physics, 110*, 1047–1058.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4*, 219–227.

Kryukov, V. I. (1991). An attention model based on the principle of dominanta. In A. V. Holden & V. I. Kryukov (Eds.), *Neurocomputers and attention I. Neurobiology, synchronization and chaos* (pp. 319–352). Manchester: Manchester University Press.

Kuramoto, Y. (1991). Collective synchronization of pulse coupled oscillators and excitable units. *Physica D, 50*, 15–30.

Liang, H., Bressler, S. L., Ding, M., Desimone, R., & Fries, P. (2003). Temporal dynamics of attention-modulated neuronal synchronization in macaque V4. *Neurocomputing, 52–54*, 481–487.

Liu, G., Austen, E. L., Booth, K. S., Fisher, B. D., Argue, R., Rempel, M. I., & Enns, J. T. (2005). Multiple-object tracking is based on scene, not retinal coordinates. *Journal of Experimental Psychology, 31*, 235–247.

Loose, R., Kaufmann, C., Auer, D. P., & Lange, K. W. (2003). Human prefrontal and sensory cortical activity during divided attention tasks. *Human Brain Mapping, 18*, 249–259.

Moser, M. C., & Silton, M. (1998). Computational model of spatial attention. In H. Pashler (Ed.), *Attention* (pp. 341–393). London: UCL Press.

Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *Journal of Neurophysiology, 70*, 909–919.

Niebur, E., Hsiao, S. S., & Johnson, K. O. (2002). Synchrony: A neuronal mechanism for attentional selection? *Current Opinion in Neurobiology, 12*, 190–194.

Niebur, E., Kammen, D. E., & Koch, C. (1991). Phase-locking in 1-D and 2-D networks of oscillating neurons. In W. Singer & H. Schuster (Eds.), *Nonlinear dynamics and neuronal networks*, (pp. 173–204). Berlin: Vieweg Verlag.

Oksama, L., & Hyönä, J. (2004). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? An individual difference approach. *Visual Cognition, 11*, 631–671.

Olshausen, B. A., Anderson, C. H., & Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience, 13*, 4700–4719.

Posner, M. I., Snyder, C. R. R., & Davidson, D. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*, 160–174.

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition, 80*, 127–158.

Pylyshyn, Z. W. (2004). Some puzzling findings in multiple object tracking (MOT): I. Tracking without keeping track of object identities. *Visual Cognition, 1*, 301–322.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3*, 179–197.

Roelfsema, P. R., Lamme, V., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature, 395*, 376–381.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition, 80*, 1–46.

Scholl, B. J., & Tremoulet, P. D. (2000). Perception causality and animacy. *Trends in Cognitive Science, 4*, 299–309.

Sears, C. R., & Pylyshyn, Z. W. (2000). Multiple object tracking and attentional processing. *Canadian Journal of Experimental Psychology, 54*, 1–14.

Seifert, A. E., Somers, D. C., Dale, A. M., & Tootel, R. (2003). Functional MRI studies of human visual motion perception: Texture, luminance, attention and aftereffects. *Cerebral Cortex, 13*, 340–349.

Shallice, T. (2002). Fractionation of the supervisory system. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 261–277). New York: Oxford University Press.

Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations. *Neuron, 24*, 49–65.

Sokolov, A., Lutzenberger, W., Pavlova, M., Pressl, H., Braun, C., & Birbauner, N. (1999). Gamma-band MEG activity to coherent motion depends on task-driven attention. *Neuroreport, 10*, 1997–2000.

Somers, D. C., Dale, A. M., Seifert, A. E., & Tootel, R. (1999). Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. *Proc. Natl. Acad. Sci. (USA), 96*, 1663–1668.

Steinmetz, P. N., Roy, A., Fitzgerald, P., Hsiao, S. S., Johnson, K. O., & Niebur, E. (2000). Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature, 404*, 187–190.

Tallon-Baudry, C. (2004). Attention and awareness in synchrony. *Trends in Cognitive Sciences, 8*, 523–525.

Treue, S. (2003). Visual attention: The where, what, how and why of salience. *Current Opinion in Neurobiology, 13*, 428–432.

Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y., Davis, N., & Nufl, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence, 78*, 507–545.

Vinogradova, O. S. (2001). Hippocampus as comparator: Role of the two input and two output systems of the hippocampus in selection and registration of information. *Hippocampus, 11*, 578–598.

Visvanathan, L., & Mingolla, E. (2002). Dynamics of attention in depth: Evidence from multi-element tracking. *Perception, 31*, 1415–1437.

Wang, D. L. (1999). Object selection based on oscillatory correlation. *Neural Networks, 12*, 579–592.

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology, 24*, 295–340.

Received April 25, 2005; accepted September 15, 2005.