Using visual latencies to improve image segmentation

Ralf Opara and Florentin Wörgötter*

Dept. of Neurophysiology, Ruhr-Universität, 44780 Bochum, Germany

An artificial neural network model is proposed that combines several aspects taken from physiological observations (oscillations, synchronizations) with a visual latency mechanism in order to achieve an improved analysis of visual scenes.

The network consists of two parts. In the lower layers which contain no lateral connections the propagation velocity of the activity of the units depends on the contrast of the individual objects in the scene. In the upper layers lateral connections are used to achieve synchronization between corresponding image parts. This architecture assures that the activity which arises in response to a scene containing objects with different contrast is spread out over several layers in the network. Thereby adjacent objects with different contrast will be separated and synchronization occurs in the upper layers without mutual disturbance between different objects. A comparison with a one-layer network shows that synchronization in the latency dependent multi-layer net is indeed achieved much faster as soon as more than 5 objects have to be recognized. In addition, it is shown that the network is highly robust against noise in the stimuli or variations in the propagation delays (latencies), respectively.

For a consistent analysis of a visual scene the different features of an individual object have to be recognized as belonging together and separated from

*To whom correspondence should be addressed.

Email: worgott@neurop.ruhr-uni-bochum.de, Phone: +234 700 4940, FAX: +234 709 4192

other objects. This study shows that temporal differences, naturally introduced by stimulus latencies in every biological sensory system, can strongly improve the performance and allow for an analysis of more complex scenes.

1 Introduction

The segmentation of a visual scene is a fundamental process of early vision, where elementary features like orientation, color, motion, texture, disparity must be grouped together to form discrete objects which have to be separated from each other and from the background. In the brain of the higher vertebrates this could be achieved by utilizing the temporal structure of the neuronal signals (McClurkin *et al.* 1991, Hopfield 1995). In particular it has been suggested that synchronization between cells could subserve this purpose (Eckhorn et al 1988, Gray *et al.* 1989, v.d. Malsburg 1981). According to this suggestion, different objects can be distinguished by the firing phases and the phase-relations of cells associated with them. As a consequence of such a synchronization process, cells which are coding one object will be forced to fire at the same time, while cells coding a different object will fire at different times. Thus ideally no phase difference exists between cells belonging to the same assembly, while the phase differences to cells of another assembly should be much larger. The temporal separation of two objects achieved by such a mechanism, however, cannot exceed more than one half-cycle of the underlying oscillation period which is about 10-30 ms in the visual cortex (Eckhorn *et al.* 1993). If we assume that the cell assemblies oscillate with 40 Hz, and if the temporal jitter for synchronously firing cells is about 5 ms, then a system that uses phase information to distinguish different objects is only able to separate less than 3 objects. To improve the performance of oscillatory networks used for complex visual scene analysis some authors employ the so called *focus of attention* (Milanese 1993, Niebur *et al.* 1993, Olshausen *et al.* 1993, Fellenz 1994, Niebur and Koch 1996, for a recent review see Posner and Dehaene, 1994 and

Desimone and Duncan 1995). The idea is to label a certain salient part of a visual scene and to restrict the segmentation to objects included in this area. This way only a few objects are present in the labeled part. The time differences (phases) between cells coding different objects are, therefore, sufficient to analyze such a restricted area. After a certain time or after recognizing all objects the attention is shifted to other parts of the visual scene and the synchronization process starts again. Although this is a very intuitive way to analyze complex visual scenes, it should be realised that in this case the segmentation is processed serially and the processing time is strongly depending on the number of objects and the complexity of the scene.

To avoid such a serial processing we suggest an alternative approach. The basic idea presented in this study relies on a contrast dependent propagation delay in a multi-layer network. The physiological background we take into consideration is based on the fact that visual latencies are depending on the stimulus contrast. They can span large ranges of 30-130 ms in the cortex (Levick 1973, Bolz *et al.* 1982, Sestokas *et al.* 1987), measured after a significant proportion of the response has built up (e.g., 4th spike, Bolz *et al.* 1982). The visual latencies in the cortex (Ikeda and Wright 1975, Raiguel *et al.* 1989) could, therefore, be used for a contrast dependent grouping of the neuronal activity. Similar to the visual system, in our model high contrast stimuli need less time to reach higher network levels than low contrast stimuli. This leads to a large temporal separation of the activity of different cell assemblies which strongly improves undisturbed internal synchronization.

A different approach which also utilizes latencies for image segmentation has been suggested by Burgi and Pun (1994). Their model is based on filter functions which include luminance dependent delays to separate a scene. The results from the different spatio-temporal filters are temporally integrated recombining some aspects of the scene. During the whole process only the most relevant primitives (e.g., edges) are used to recognize an object. Opposite to the model of Burgi and Pun (1994) our model is based on spiking neurons and its architecture contains a certain similarity to

that of the visual system of vertebrates. In addition, our model does not rely on the detection of primitives, instead - similar to the retina - all aspects of a scene are taken into account and the latency induced separation will be maintained throughout all layers. The final binding of objects will be achieved in our network by a conventional synchronization mechanism, and the conceptionally new latency algorithm is used to enhance the speed of synchronization for cell assemblies representing different objects.

We will first give an overview of the network, by describing the dynamic of the neurons and connection structure, then we will show some typical simulations and give a quantitative estimate of the performance of the network. Finally we will discuss the limitations of the approach and also the question of possible physiological relevance.

2 Description of the network

2.1 The dynamic of the neurons

The behavior of the network is characterized by the connectivity pattern (Fig. 1) and the dynamic behavior of the neurons which can be characterized as integrate and fire units. In our algorithm synchronization of neurons is achieved, applying a concave membrane potential time-function $U(t)$ ($U''(t) < 0$) (Mirollo and Strogatz 1990, Ernst *et al.* 1994; for more details about the synchronization see Appendix A).

The dynamics of each individual neuron without connections can be described by:

$$\begin{aligned} U(t_{i+1}) &= F[t_i + \Delta t] = F[F^{-1}(U(t_i)) + \Delta t]\ ; \quad if\ U(t_i) < \Theta \\ U(t_{i+1}) &= 0; \qquad\qquad\qquad\qquad\qquad\qquad else. \end{aligned} \tag{1}$$

with:

F : concave function that describes the membrane potential.
In our simulations we use $F(t) = 1 - \exp(-t/\tau)$,

F^{-1} : inverse function of F.

The notation which uses the inverse function will be needed as soon as connections between neurons are considered (Eq. 2). F describes a low-pass difference equation

with a time constant τ that in the first layer depends on the image contrast I. Neurons which are stimulated by a high contrast defined with respect to the background are firing with a higher frequency (small time constant) than neurons which are stimulated by a low contrast (large time constant).

Θ is the firing threshold. If the membrane potential U reaches the threshold Θ a spike is generated and propagated through the network. In the next time step U is reset to zero.

We used the following parameters for the network dynamics:

$$\begin{aligned} \Theta &= 0.9999, \\ \tau &= 0.15 \ / \ \text{contrast}; \quad \text{in layer 1}, \\ \tau &= 1; \qquad\qquad\quad \text{in layer 2-9}. \end{aligned}$$

Spikes which are propagated will lead to either an excitatory or an inhibitory action at the target cell(s) shifting the membrane potential $U(t)$ by a total of V_ϵ. Thus, the complete dynamic of a neuron is described by:

$$\begin{aligned} U(t_{i+1}) &= F[F^{-1}(U(t_i)) + \Delta t] + V_\epsilon \ ; \quad if \ U(t_i) < \Theta \\ U(t_{i+1}) &= 0; \qquad\qquad\qquad\qquad\qquad\qquad else. \end{aligned} \tag{2}$$

The total contribution of the connections from all neurons which spike at time t_i is given as:

$$V_\epsilon = \sum_k v_k \tag{3}$$

Thus, summation is performed over all neurons k that spike at time t_i. The individual values of v_k are given in Appendix B.

The connection strength which determines the absolute value of v_k mainly depends on the distance between source and target cell and the type of connection made (see below).

Because U describes a concave function, synchronization occurs necessarily as soon as V_ϵ is unequal to zero (Mirollo and Strogatz 1990, Ernst *et al.* 1994). The coupling

V_ϵ within and between layers is symmetrical and is only a function of distance. It leads to an excitatory interaction within layers over short distances, while inhibitory coupling is predominant at long distances. The coupling strength decays exponentially with increasing distance. For details see Appendix B.

Since the firing frequency is not constant in the lower layers, we can not define the phase in a conventional way.

To derive at a measure suited for our purposes, we define the "phase" Φ of a neuron i with respect to the actual time t as the minimum time-difference between its own firing of two subsequent spikes and the temporal average of the firing of all neurons, divided by the firing period. Thus, the phase Φ of a neuron i is:

$$\Phi_i = \Phi_i(t) = 360^\circ \frac{t^{mean} - min(t - t_i^{fire}, T_i - (t - t_i^{fire}))}{T_i}, \tag{4}$$

where t_i^{fire} is the time, when neuron i has fired the last time, T_i is its firing period and t is the actual time. The temporal average of the firing of all active neurons is $t^{mean} = \frac{1}{n} \sum_j min(t - t_j^{fire}, T_j - (t - t_j^{fire}))$. This yields a consistent phase measurement in all layers and is defined for all values t on the time axis.

2.2 The network

In the network two functionally different areas can be distinguished. The first area is mainly responsible for the contrast dependent latency differences. In Fig. 1A it is labeled as *Retina and LGN*. The second part labeled *visual cortex* causes fast synchronization and feature binding which is achieved by lateral connections.

The network consists of nine layers. Neurons of each layer are part of a three dimensional grid. For simplicity in Fig. 1A only a two-dimensional frontal view is shown. The central column of Fig. 1A shows the connectivity pattern of the multilayer neural network. Other columns are wired up in the same way, and the same connectivity also applies for the third dimension which is not shown in Fig. 1A.

The input of the network consists of a two-dimensional grey level image which is

fed into layer one. The process of image segmentation consists of different processing stages which are arranged in a hierarchical way. In the first processing stage the activity of cells coding certain objects needs different times to activate cells in the higher layers, due to their different latencies. The next layers (above 3) include local interactions between adjacent cells. Because of the locality of the interactions the activity of several objects can coexist in these layers. As the activity is propagated to higher layers, larger cell populations are involved in the synchronization/desynchronization process as a consequence of increasingly wider lateral connections. At the read-out layer (in Fig. 1A layer 9) a representation of global context is achieved. Note, however that we are not explicitly concerned with the read-out itself.

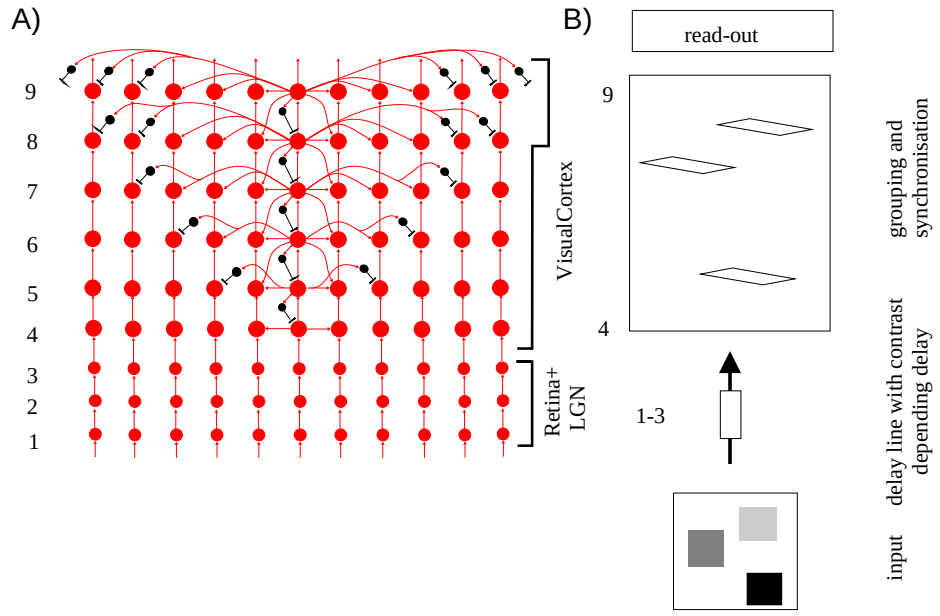

Figure 1: A) Schematic diagram of the network in a frontal view (third dimension omitted). Lateral connections are only shown for the central column. B) The functional components of the network. The input is piped into a delay line. The duration of the delay is depending on the stimulus contrast. In the next processing stage cell assemblies are formed and cell assemblies coding objects with different contrasts are spatial (i.e. temporally) separated.

2.2.1 Latency layers

The first three layers in the network produce the latencies which are depending on the stimulus contrast with respect to the background. Thus, these layers can be characterized as a contrast dependent delay line (Fig. 1B). To introduce delays in our system we use frequency coding at the input stage that depends on feature contrast. To excite neurons in each following layer long integration times (2-3 cycles) are necessary (Appendix B). The resulting latencies in computer time units are shown in Fig. 2 as a function of stimulus contrast. As parameter the integration time of the neurons is shown. The latencies in Fig. 2 and all other are shown in arbitrary computer time steps, because the network layout provides no constraints on the physiological parameter ranges which could be used to normalize the time-axis in (e.g.) milliseconds.

Neurons with a high firing frequency (high feature contrast) at the input stage need less time (short latency) to excite cells in layer two and three than neurons with a low frequency (low feature contrast). Propagation through the layers leads to a further amplification of the latency differences until the activity reaches layer 4 which is the first layer where the lateral connectivity sets in. Both effects (initial frequency differences and delay-amplification through the layers) effectively mimic the different latencies of a real retinal signal.

At this processing stage only feed-forward connectivity is used, to ensure a linear information flow to the *visual cortex* without interference between different cell assemblies.

2.2.2 Synchronization layers

The next stage synchronizes the neurons which leads to feature binding. The oscillator characteristics of the neurons assures that almost constant firing frequencies are achieved in the upper layers.

In the *visual cortex* four different types of connections can be distinguished which

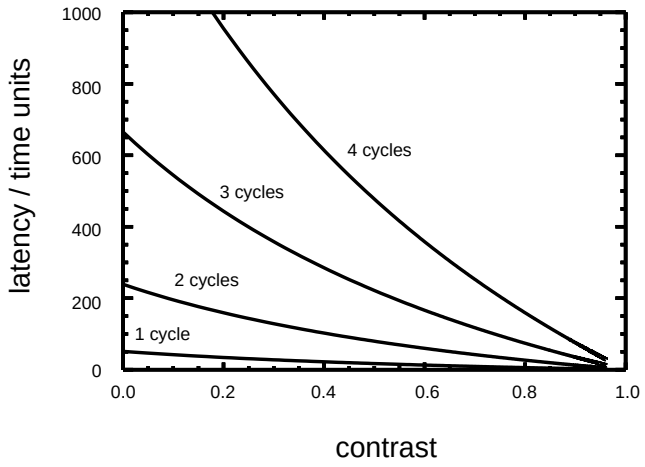

Figure 2: Latency of activity distributions to reach layer 4 as a function of stimulus contrast. Several curves are plotted according to the integration time (in cycles) necessary to fire the next layer.

will be described subdividing them in connections within a layer and those between layers.

connections within a layer:

1. Lateral excitatory connections between neighboring neurons. In the lower layers connections are only made between directly adjacent neurons. The locally restricted connections will prevent that objects disturb each other during the process of forming a cell assembly. This yields to a fast synchronization of cells belonging to a cell assembly coding a certain object, even if more than one object is represented. The radius of lateral excitation increases in the upper layers. This makes it possible that increasingly larger regions are involved in the synchronization process.

2. Lateral inhibitory connections over larger distances. The inhibitory connections within a layer have two effects. At first they will force cell assemblies which are active at the same layer at a certain time to fire with a phase shift within

the oscillator period. The second effect is a competition of cell assemblies to reach the next layer if they are coding different objects. Due to the competitive character of this connectivity pattern some objects will reach the next layer faster than other object. The competition continues up to the read-out layer thereby reducing the number of objects represented at a certain layer and as a consequence the complexity of the visual scene that is processed at a certain time.

The magnitude of the inhibitory connections is much smaller as compared to the excitatory connections between adjacent neurons within a layer, so the synchronization of an already formed cell assembly remains nearly unaffected.

connections between layers:

1. Direct feedback inhibition between layers. If the activity distribution of cells coding one object reaches a certain layer l, the activity of the corresponding neurons at layer l' (with $l' < l$) is inhibited, to prevent the continuous arrival of input. For example, a continuous arrival of input activity from a high contrast stimulus in the lower layers would disturb adjacent low contrast stimuli which just reached the same layer. Feedback inhibition leads to the restriction of the propagating signal for each object to one or two layers (in V1 (Fig. 3)) at a certain time.

2. Lateral excitatory feedback connections. These connections are introduced to speed up the propagation of lagging activity which is caused by noise in the feature intensity or the latency. The weights of these connections are smaller than the weights of the lateral connectivity pattern within a layer. This leads to the effect that locally formed cell assemblies are only little influenced while lagging activity of a few individual cells is pulled forward to be included into its cell assembly.

3 Results

3.1 Example

As an example, Fig 3C shows the activity and phase distribution of the network response with parameters shown in appendix B to a stimulus shown in Fig. 3A. The stimulus consists of two square objects (α,β) with a high contrast and a third (γ) low-contrast object. The squares α and γ are adjacent, while the square β is spatial separated from them.

In Fig. 3C (middle) each pixel represents the firing phase of one neuron. The phases are defined for active neurons which fire more than once and are calculated according to Eq. 4. White pixels represent inactive neurons. Columns show the phases of all active neurons at selected times for all layers of the network, while rows represent different times but a fixed layer. In the first three layers (rows 1,2 and 3 in Fig. 3C middle) the lack of lateral interactions leads to a random phase relation Φ between all neurons. Due to the latency mechanism, features with a high contrast (α and β in Fig. 3A) need less time ($\approx$ 1000 iterations) to reach layer 4 (row 4 and column 3), while features with a low contrast (γ in Fig. 3A) need longer times ($\approx$ 1500 iterations) (row 4 and column 8).

It should be noted that we assume zero background luminance for all simulations presented in this study which prevents "background-neurons" from firing. If luminance levels different from zero would be used the background would be regarded as an additional object and background-neurons would synchronize.

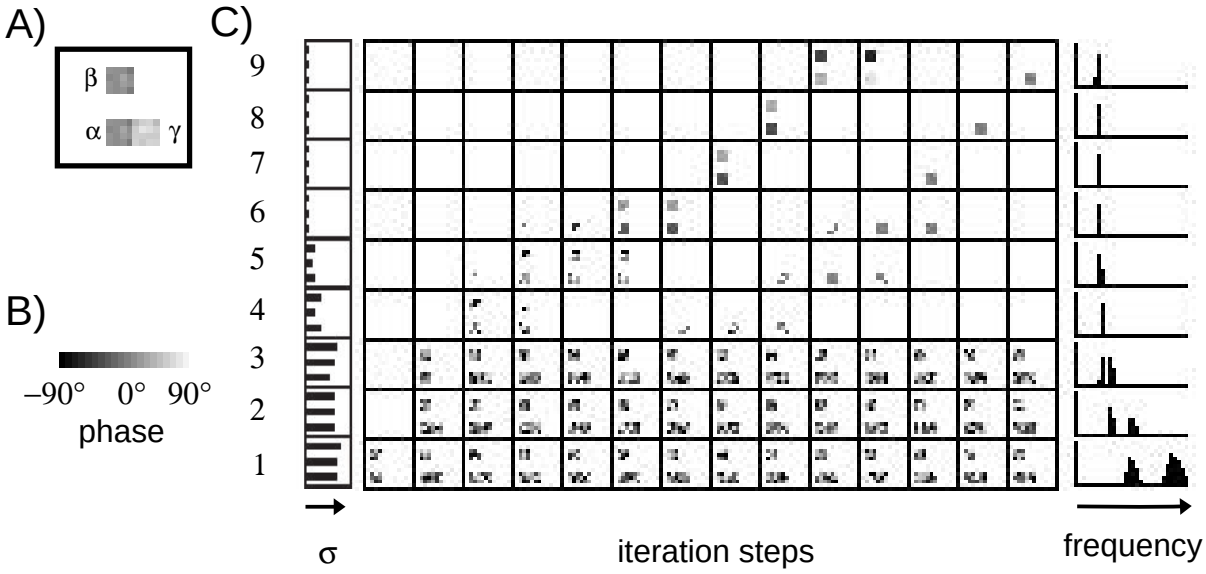

Figure 3: A) The stimulus which was given to the network consists of two square objects (α,β) with a high contrast and a third (γ) low-contrast object. The squares α and γ are adjacent, while the square β is spatial separated from them. B) Grey level coding of the phases. C) Left: Horizontal bars indicate the standard deviation of the firing phase σ of the neurons that belong to one object (top=α, middle=β, bottom=γ) averaged for the complete run. Standard deviation drops to zero in the higher layers. Middle: Time course of the activity distribution at selected iterations. The shown iterations steps are: 15, 670, 1078, 1253, 1258, 1294, 1326, 1511, 1648, 1672, 1838, 2020, 2192 for colums 1 to 14. Pixel grey levels represent the firing phase. Above layer three two objects (α,β) are separated in time from the third (γ) which has a longer latency. Neurons which represent one object acquire synchrony (i.e., constant phase which is indicated by constant grey levels) in the upper layers and neurons coding objects with the same contrast (α and β) are firing with different phases. Right: Average firing frequency for all cells. The frequency is higher in lower layers with two peaks representing the three objects (neurons coding α and β are firing with the same frequency) and constant at a single lower value in the model cortex.

The activity of cells coding the same object is initially spread out over several layers. At column 4 in Fig. 3C the activity of square α (Fig. 3A) is distributed from layer 1 up to layer 6. This occurs because the input images are noisy (25% noise). The processing in the grouping and synchronization layers will concentrate the activity distribution of cells coding a certain object to one layer at a certain time. In Fig. 3C the activity of square α is grouped at the same layer after passing layer 6 and the cell assemblies that are coding one object are firing synchronously, coded by the same grey level in Fig. 3 C.

Features with a high contrast difference are separated across the model layers after a few iterations (column 2 in Fig. 3C). This spatial (viz. temporal) separation of the activity permits an independent processing of features with different contrast, minimizing the interference between different features after the first processing stages. As a consequence the speed of the synchronization between neurons coding the same feature is strongly enhanced and in the higher layers individual features are bound together to form a consistent representation of the objects in the visual scene.

3.2 Comparison to networks without latency

In this section we will compare the multi-layer network to a one-layer network without latency. The one-layer network has the same dynamics and the same connectivity pattern as layer 9 in our network. The task of both networks is to recognize 2 up to 15 objects in a two dimensional input image. For each network the time is calculated until all cells coding a certain object are firing synchronously. The time for recognizing all objects is summarized in Table 1. If only a few (2 - 5) objects are presented to the networks, the one layer model is faster, because of the delays in the multi-layer network. The opposite effect is observed for many (> 5) objects. Here the interference between the different cell assemblies in the one-layer network leads to a long time until synchronization occurs. For more than 6 objects synchronization is altogether prevented in the one-layer network. On the other hand, in these cases the

	1 layer		9 layer	
no. of objects	time/ time units	cpu time/s	time/time units	cpu time/s
2	98	39	2179	570
3	104	42	2193	579
4	218	83	2230	611
5	562	209	2281	652
6	2627	1010	2310	678
7	-	-		
9	-	-	2606	899
12	-	-	2861	1096
15	-	-	3148	1305

Table 1: Comparison of the performance of two networks with (9 layer) and without (1 layer) propagation delays. The task is to recognize 2 up to 15 objects in a visual scene. For each model the iterations and the CPU-time on a SUN Sparc 20 for recognizing all objects is shown. A total of 900 neurons per layer have been simulated.

synchronization of cell assemblies in the multi-layer network is still very fast, because of minimal interference between different objects. In addition, it should be noted that the time for recognition remains almost constant in the 9-layer network which is a desirable feature whereas a steep non-linear increase is observed in the 1-layer net.

4 Performance tests

The performance of the network is determined in a series of simulations, in which some basic parameters, contrast, noise, shading, latency parameter and the number of adjacent objects are varied. The performance is measured by the quality of object segmentation in layer 9. We avoid to use classical correlation methods like auto- and cross- correlation which are not very well suited to analyse the coherence in our simulations, because sychronous activity always occurs in short bursts in the read-out

layer neurons which are otherwise silent. In addition cross-correlation would here to be performed mutually between many (> 10) neurons which is computationally very expensive. Alternatively we propose a quantification which is achieved by the *internal coherence rate* within a cell assembly (c_{int}) coding one object and the *external coherence rate* between cell assemblies coding different objects (c_{ext}). This way we get two similarly derived values for each simulation which can be plotted simultaneously in a three dimensional diagram (see Figs. 4-7).

Coherence rates are calculated for layer 9 only. Thus the following description refers to this layer.

The internal coherence rate measures the relative part of synchronously active cells as compared to the maximally possible synchronous cell activity for each individual object in the scene. For one object α it is calculated in the following way.

Because our stimuli are rather simple we know how many cells should represent object α and in the first step all spikes from all those neurons are collapsed into one trace. Multiple spikes occuring at the same time step are considered. From this trace an event-histogram is compiled for each iteration step using a sliding window with width 10 iterations. The time window (10 iterations) defines the maximally allowed temporal jitter for synchronously firing cells. This window is less than 20% of an average oscillation cycle so that each cell can only fire once in the window. Thus, the local maxima of the histogram always represent combined activity across cells. Accordingly, we call each local maximum an *assembly activation*. We can now determine the occurance times t_i of these assembly activations. Note that the assembly activations can sometimes consist of a single spike if the spike train is highly dispersed in the case of non-synchronous firing, where no true cell assemblies exist.

In the next step we determine how many spikes are involved in generating each assembly activation. The number of spikes n_i^α included in an assembly activation is determined by counting all spikes in the original collapsed trace which occur close to the assembly activation, i.e., within the allowed distance of $t_i \pm 5$ iterations.

Because of the known dynamics we also know how often the neurons of object α should fire synchronously for each assembly activation (N^α). Thus, we can compute the internal coherence rate by summing the minima[1] of $(N^\alpha - n_i^\alpha)$ and n_i^α.

The minimum of $N^\alpha - n_i^\alpha$ and n_i^α is summed for each assembly activation over the whole spike train in layer 9 eliminating the time-dependency. If all cells of a cell assembly are synchronized the sum over the minimum of $N^\alpha - n_i^\alpha$ and n_i^α is zero, while in the case that no cells are synchronized it is $k \times N^\alpha = N_{max}^\alpha$ (if each neuron fires k times).

The normalized internal coherence rate c_{int} is thus given by:

$$c_{int} := 1 - \frac{1}{N_{max}^\alpha} \sum_i min(n_i^\alpha, N^\alpha - n_i^\alpha) \tag{5}$$

with:

N_{max}^α	:	all spikes in the collapsed spike train of object α ,
N^α	:	total possible number of spikes in each assembly activation,
n_i^α	:	actual number of spikes in the assembly activation,

A high coherence rate indicates synchronicity of cells in the assembly over the whole spike-train, while a low coherence rate indicates uncorrelated activity.

The other measure we use is the coherence *between* two cell assemblies α and β coding different objects defined by:

[1] The minimum is taken, because the temporally breaking up of an expected activity group into two groups of equal size (i.e. $n_i^\alpha = \frac{N^\alpha}{2}$) represents the worst case for a coherence estimation as compared to a situation where only one or two spikes are missing in a group.
Accordingly: $N^\alpha - 1 > N^\alpha - \frac{N^\alpha}{2}, \quad if\ N^\alpha > 3,$
but $min(N^\alpha - 1, 1) < min(N^\alpha - \frac{N^\alpha}{2}, \frac{N^\alpha}{2}), \quad if\ N^\alpha > 3.$

$$c_{ext} := \frac{1}{N_{max}^{\alpha,\beta}} \sum_{\substack{i,j \\ if \ | \ t_i - t_j| < \Delta t}} (n_i^\alpha + n_j^\beta) \tag{6}$$

with:

$N_{max}^{\alpha,\beta}$	:	all spikes of objects α and β,
n_i^α	:	number of firing cells in the i'th assembly activation of cell assembly α,
n_j^β	:	number of firing cells in the j'th assembly activation of cell assembly β,
t_i, t_j	:	occurance times of the assembly activations.
Δt	:	maximal time difference for synchronously firing cells (10 iteration steps),

This sum grows only if assembly activations for the objects α and β occur simultaneously, i.e., within 10 iteration steps. Thus, the external coherence rate indicates, if different cell assemblies are belonging together ($c_{ext} = 1$) or not ($c_{ext} = 0$).

The coherence rate is calculated for several simulations (mostly 12) and the results are averaged considering their errors. The central advantage of these two measures is that they are derived similarly and can, thus, be plotted in one diagram.

4.1 Coherence rate as a function of contrast and latency

To use the network in an image segmentation task it is necessary to know the network response to different kinds of stimuli as a function of some basic parameters. The most important parameter is the latency. It determines the extent of temporal separation at a certain layer or - which is equivalent - the spatial separation of activity at a certain time as a function of contrast. In several simulations a stimulus was presented to the system, consisting of to adjacent squares (Fig. 4A1-A3) with a different contrast varying from 3% to 43% between them. The latency parameter indicates the time (not normalized) between stimulus onset and the occurrence of the first response in layer 4 (Appendix B). To examine the behavior of the system the internal and external coherence rates are calculated according to equations 5 and 6. The external coherence

rate is in Fig. 4B shown as surface, while the internal coherence rate is presented as grey level distribution on the surface. The internal coherence rate is proportional to the amount of white in the grey level distribution. In Fig. 4B region 1 indicates little interference between the cell assemblies (low external coherence rate c_{ext}) and a high coherence rate within each cell assembly (high internal coherence rate c_{int}). Two assemblies are formed with an activity which is independently propagating through the network. Region 2 is characterized by an intermediate external c_{ext} and internal c_{int} coherence rate, where the relation between the cells is ambiguous. The network does not form distinct cell assemblies, but individual cells switch between different states. In this region the system cannot clearly decide if the stimulus consists of one or two objects. In region 3 the external c_{ext} and internal c_{int} coherence rate are high and all cells are firing synchronously, forming a single assembly. At short latencies the ambiguity range (region 2) is large because the transition from region 1 to region 3 is shallow. Up to longer latencies the transition is steep and the ambiguity range is small.

4.2 Robustness against noise

A system, utilizing the contrast differences of objects, must be tolerant against any unintentional disturbance caused by noise. The stimulus that was given to the network is shown in Fig. 5A1-A3. It consists of two adjacent squares with a contrast of 20% between them. The luminance I of the individual pixels varies according to:

$$I = I_0(1 + \frac{(rand(noise_{max}) - 0.5 * noise_{max})}{100}) \tag{7}$$

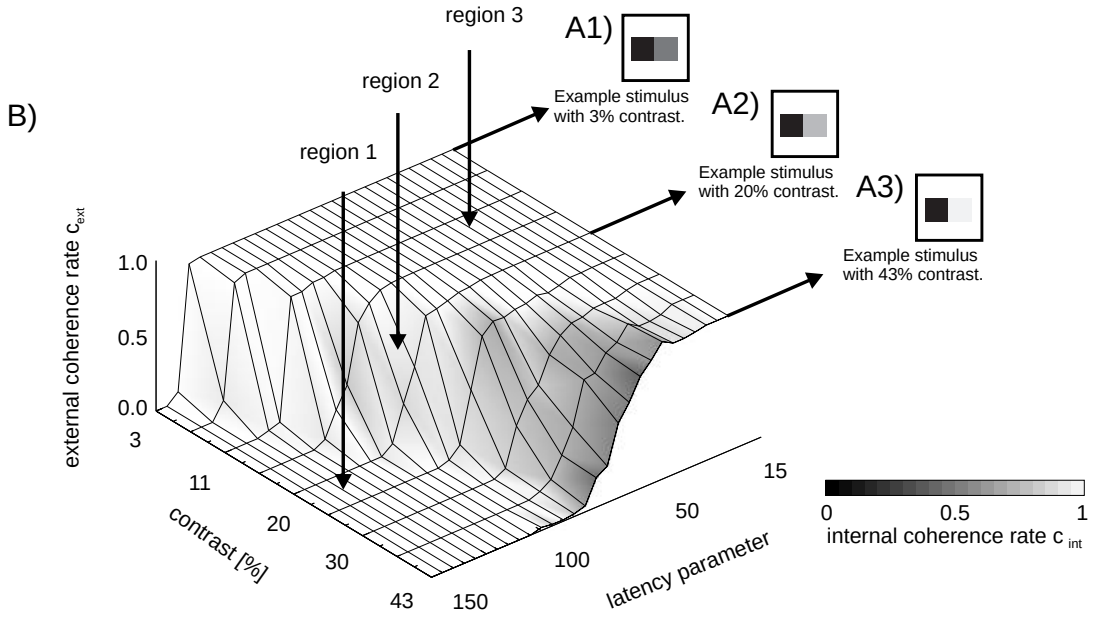

Figure 4: A) The stimulus consists of two adjacent squares. The contrast between the squares is varied from 3% up to 43%. B) The external (surface) and internal (gray level distribution) coherence rate c_{ext} of two cell assemblies representing two adjacent objects as a function of contrast and latency parameter.

with:

I_0	:	luminance without noise,
$rand(noise_{max})$	:	positive random number smaller than $noise_{max}$,
$noise_{max}$	:	maximal noise in % of I_0.

Internal and external coherence rates are computed and averaged as before (cf. Fig. 4B). In Fig. 5B four different areas can be distinguished. The first region is characterized by a low internal coherence rate and a low coherence rate between the cell assemblies. This is the case at very long latencies (latency parameter = 140) and very noisy stimuli (50% noise). If the latencies are too large, a synchronization is difficult to realize, because noisy image parts are artificially separated over long distances. However, such long latencies are physiologically unrealistic.

The second region is characterized by a good internal synchronization and low

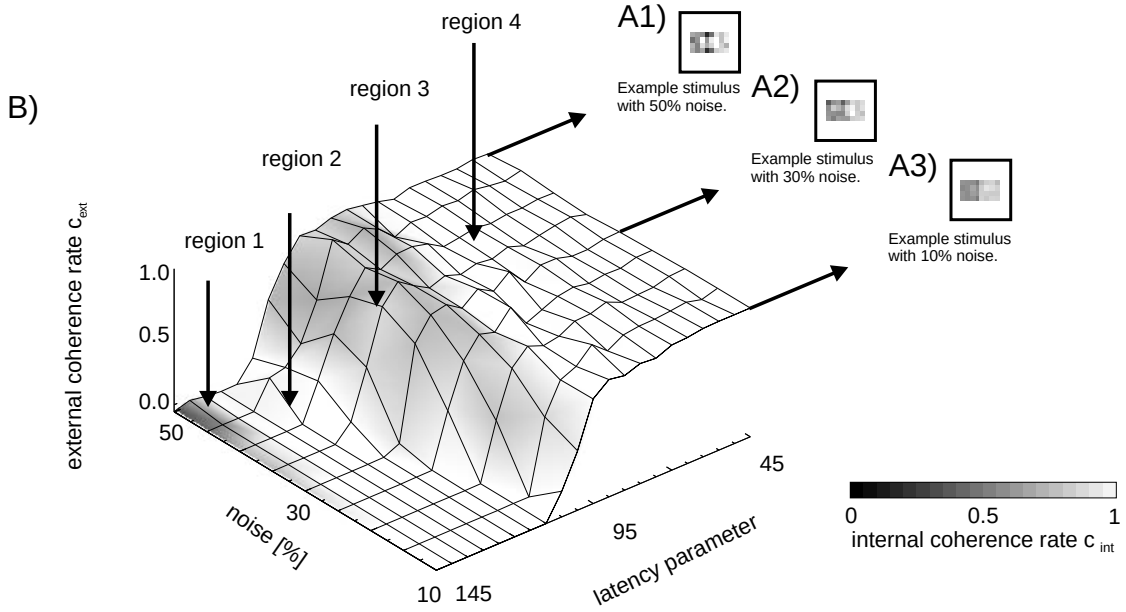

Figure 5: A) The stimulus which was given to the network consists of two adjacent squares with a contrast of 20% between the squares. The noise-induced variation of luminance is calculated according to equation 7. B) Coherence rate c_{int} within a cell assembly shown as surface and the coherence rate c_{ext} between two cell assemblies coded as grey level distribution on the surface. In this representation the latency parameter of the network response and the noise of the stimulus is varied.

coherence rate between the cell assemblies. This behavior occurs at an intermediate range of latencies (latency parameter > 120 at 50% noise and latency parameter $>$ 110 at 10% noise). In this parameter range the network groups the activity of the cell assemblies internally and cell assemblies coding different objects are separated from each other (low coherence rate between cell assemblies).

In region three the internal coherence rate is low, while the external coherence rate is at an intermediate level. Cell activity interferes with each other and the cell assemblies are not able to group their activity internally over the whole spike train. The cells switch between the states "one object" and "two objects". In such a region the system cannot make a clear decision, but this is only a small range.

The last region shows high internal and external coherence rate. The cells are in-

ternally grouped together and the two cell assemblies are synchronized. This indicates a decision : one object is recognized.

4.3 Shaded surfaces

In a natural environment it is quite common that the luminance changes continuously along the object surface. An important feature of a contrast based system must be the tolerance against such shaded surfaces. The connectivity pattern of the network, as mentioned above, leads to a synchronization and grouping of activity of cells representing similar contrast and spatial neighborhood. Shaded surfaces are characterized by smooth contrast changes within the boundary of the object. As a consequence of our model, shaded surfaces with a small shading gradient are much stronger grouped together than surfaces with a steep gradient.

In a series of simulations the response of the network has been calculated as a function of the grey level gradient and latency. The internal (Eq. 5) and external (Eq. 6) coherence rate are calculated and shown in Fig. 6B.

The stimulus is shown in Fig. 6A1 - A3 and consists of two squares with a contrast of 25% between them which are connected by a grey level ramp. The contrast of the squares is fixed, while gradient of the grey level ramp is varied from 2.7 $pixel^{-1}$ to 10 $pixel^{-1}$.

In Fig. 6B three different areas can be distinguished. The first region is characterized by a high internal synchronization but a low coherence rate between the cell assemblies. Here the gradient of the shading is so steep that the activity is grouped into two different cell assemblies. This is the case for shadings with a large gradient for latency parameters greater than 100 and for smaller gradients at long latencies (latency parameter = 125).

In region two internal coherence rate is low, while the external coherence rate is at an intermediate level. Cell assemblies are not able to group their activity internally

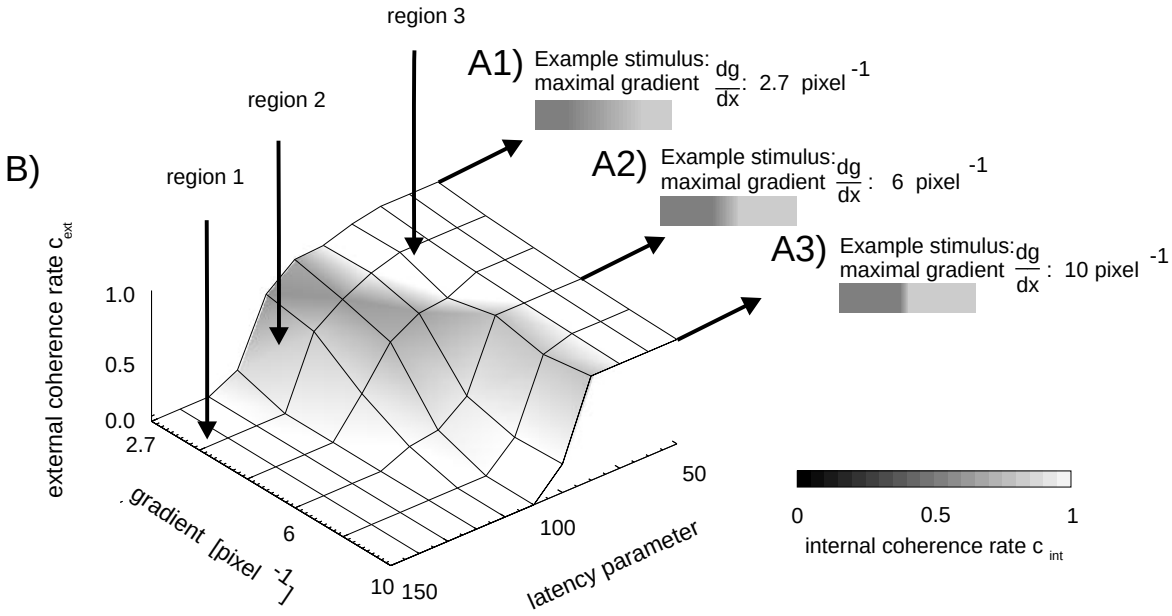

Figure 6: A) The Stimulus consists of two adjacent squares with a smooth transition. The contrast between the left and right edge is 25%. B) Coherence rate c_{ext} between two cell assemblies shown as grid and the coherence rate c_{int} within the cell assemblies coded as grey level distribution at the grid, as a function of latency and the gradient of the transition between the square.

and the cells are unpredictably switching between different phase relations.

The last region (3) shows a high internal and external coherence rate. The cells are internally grouped together and the two cell assemblies are synchronized, thus, coding a single object.

4.4 Influence of adjacent objects

The simulations above have been performed for two adjacent objects which are represented by two cell assemblies of the same size in feature space (Fig. 4B, 5B,6B). This means that only one contact surface exists. In Fig. 7B the influence of many contact surfaces is shown. The stimulus consists of a high contrast square in the middle of the image and several rectangles with a lower contrast which are placed around the high contrast square (Fig. 7A1-A4).

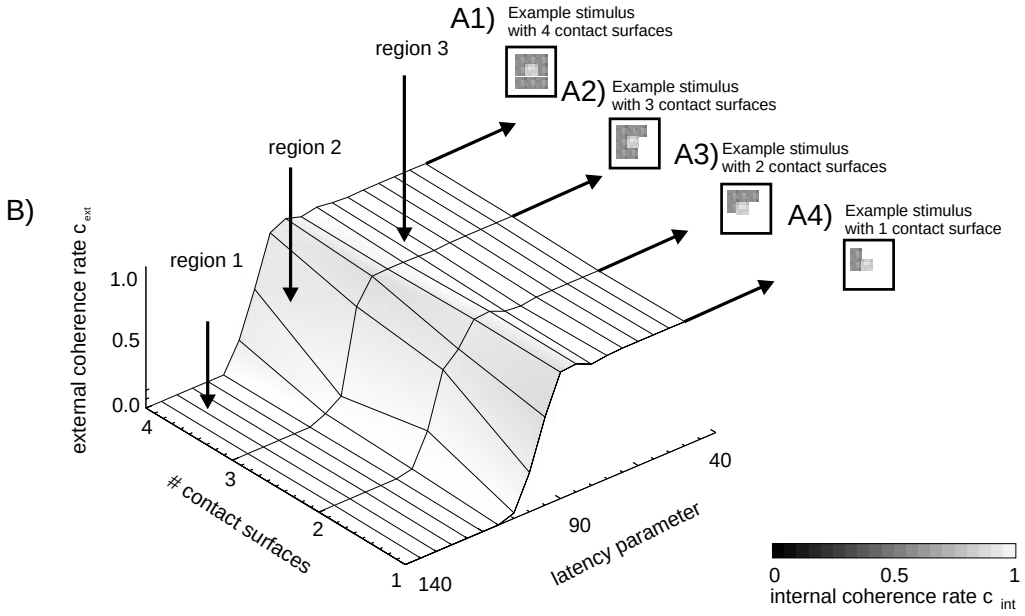

Figure 7: A) The Stimulus consists of two objects. One bright square in the middle of the image and one surrounding object with one up to four contact surfaces to the square in the middle. The contrast between the objects is 20%. B) Coherence rate (c_{ext} as grid and c_{int} as grey level distribution at the surface) of cell assembly 1 (high contrast) and cell assembly 2 (low contrast) as a function of number of contact surfaces and latency.

As above, we can distinguish three different areas in Fig. 7B. The first region is characterized by a high internal synchronization and low coherence rate between the cell assemblies. This occurs at long latencies (latency parameter greater than 105 with only one contact surface and at latency parameter greater than 110 with four contact surfaces). For four contact surfaces this region is only a little bit shifted to higher latencies as compared to the network response for one contact surface. This indicates that many contact surfaces will have only little effect on the network response.

In region two the internal coherence rate is low and external coherence rate is at an intermediate level, and as in the examples above (Fig. 4B, 5B, 6B) no grouping is achieved. The last region shows a high internal and external coherence rate. The cells are internally grouped together and the two cell assemblies are synchronized to

represent one object.

The effect of many contact surfaces is characterized by the tendency to bind the cell assemblies a little bit earlier together compared to only one contact surface. This effect, however, is small related to other effects (Fig. 4B, 5B, 6B).

The simulations (Fig. 4, 5, 6 and 7) show that the network can simultaneously treat different combinations of individual stimulus parameters. A significant overlap of valid decision surfaces exists for all four kinds of stimulus parameters used in these figures in large ranges. This is most easily demonstrated in an example of only two stimulus combinations, e.g., Fig. 4 (contrast of 15% - 25%) and Fig. 5 (noise 10% - 30%). For a combination of these parameters (noise 10%-30% and contrast 15%-25%) the network will still bind the two cell assemblies if the latency parameter is lower than 85 and for a latency parameter greater than 115 it will separate them. This example can be extended to more than two stimulus parameters.

5 Discussion

In this study we combined several physiologically inspired aspects of real neural networks (i.e., oscillation, synchronization) together with a latency mechanism to a network that simultaneously achieves efficient object binding and image segmentation. The following discussion shall be devoted to two questions: 1) What are the advantages and the inherent performance limitations of the network? and (2) Is this mechanism physiologically feasible?

5.1 Advantages and limitations

The central advantage of our network relies on the fact that the most general feature of each image - it's contrast distribution - is directly used to control the synchronization process. Features with a high contrast are thereby favored and reach the last network

layer faster than objects that do not "jump to the eye". This mechanism thereby mimics human perception (Burgi and Pun 1991) but more importantly it efficiently limits the information flow which needs to be evaluated in the network at any point in space (viz. time). In other approaches a small focus of attention is used (Milanese 1993, Niebur *et al.* 1993, Olshausen *et al.* 1993, Fellenz 1994, Niebur and Koch 1996, for a recent review see Posner and Dehaene, 1994 and Desimone and Duncan 1995) to also impose spatial restrictions on the information flow. In most cases the focus of attention is shifted through the network either arbitrarily or by means of a precomputed saliency map. The term "saliency", however, is not well-defined and what shall be regarded as salient is mostly relying on the taste of the network-designer. In our design we circumvent the problem of how to define and shift a focus of attention and instead rely entirely on the inherent saliency differences of features which have a different contrast. As a result, our network is rather simple and almost no restrictions are imposed on the design as long as it contains a delay-line and is able to synchronize its units.

In our model synchronization occurs only between spatially adjacent units over increasing lateral distances. This apparent disadvantage, could be easily overcome by a more complicated lateral connectivity which can lead to a synchronization within almost any kind of desired geometry. Another restriction of the network is even more prominent: In most cases objects with a graded shading of a high contrast will be cut by the network into two or more pieces traveling separately through the layers. There seems to be no imminent way of how to overcome this restriction without changing the network architecture to a large degree (e.g., by introducing feedback loops and/or higher order feature detector neurons).

In general our network will fail as soon as higher level "cognitive decisions" have to be made. As a solution to this problem, in the brain of vertebrates feedback-loops, higher order feature detector neurons and more complex receptive fields are introduced along the hierarchy of the visual pathway in order to finally lead to our

own advanced abilities for image analysis. Our current knowledge about the network of the visual pathway, however, is still so strongly limited that it remains unknown of how to assemble these aspects in order to create a generic image analysis network (for physiologically oriented models see e.g., Wörgötter and Koch 1991, Somers *et al.* 1995). Therefore, so far all artificial synchronization-models introduce a certain degree of arbitrariness in the design as soon as higher level image segmentation and binding is to be achieved (Sporns *et al.* 1989, Tononi *et al.* 1992).

The two restrictions of our network discussed above (spatial neighborhood restriction and graded shading restriction) can probably be lifted by rather simple additions to the design. The goal of this study, however, was not to try to achieve complex network performance by an arbitrary connection structure but rather to promote a novel idea of how to improve image segmentation by a very simple, generic and physiologically plausible (see below) mechanism. It should be noted that even some of the existing connections in the network (e.g., feedback excitation, feedback inhibition) could be removed without altering its basic behavior. These connections have been introduced to make the system more robust against noise in the stimuli and they are not responsible for generating latencies or leading to synchronization.

As an additional problem it should be realized that the network architecture cannot directly be transfered onto the brain, because a strict layering over more than 3 stages is not existing there. Abeles (1993) has, however, introduced the idea of a synfire chain which essentially represents information processing by subsequently excited groups of neurons. This idea supports a more serially arranged processing structure as demanded by our model.

The latencies in our network are defined by different contrasts. It would, however, also be possible to employ more complicated aspects of an image or image sequence (e.g., the distributions of orientation, depth, motion, etc.) in order to define the latencies. In fact, Niebur and Koch (1996) proposed to use a combination of different image features with a predominance of motion in order to create a saliency map which

defines how the focus of attention is shifted across the image in their model. This could very well be regarded as a clever pipelining of several spotlights which would be processed subsequently.

5.2 Physiological background

The model we presented in this study is far from any biological realism but rather it is meant to demonstrate the applicability of a combination of several physiologically inspired mechanisms. There are nevertheless several aspects which bear a certain degree of similarity to the visual system.

From psychophysical experiments is has long been known that latencies can have a distinct influence on visual perception. Most famous is probably the Pulfrich effect (1922, see Enright 1985) where luminance differences between both eyes result in a depth percept of a stimulus which is actually moving in an equidistant plane in front of the observer. More recently evidence was provided that the Pulfrich effect is actually reflected at the level of single cell behavior in the visual cortex (Carney *et al.* 1989). Visual latencies do not only depend on the contrast but also on color and spatial frequency of the stimulus. In addition, they are depending on the state of light adaptation. All these effects do not interfere with the general idea our model, because for us contrast only represents the simplest way of creating a functional dependence for the latencies which could, however, also be replaced by more complicated combined dependencies.

While the physiological findings point to the general importance of latencies for visual perception, it is still unclear if a delay-line/synchronization model such as ours could reflect reality. As a basic requirement our model demands that synchronous oscillations directly follow the onset of a visual stimulus. Whittaker and Siegfried (1983) in a study of visually evoked potentials provide evidence for such a stimulus locked oscillation onset.

Over the last years there is a heated discussion going on about the physiological

relevance of models which use synchronization for feature binding (see Singer and Gray 1995 for a recent review). It does not seem to be appropriate to enter this discussion here because in the context of this study the major focus lies on the latency mechanism. In fact, Burgi and Pun (1994) have demonstrated that a delay mechanism can be used to improve image analysis without employing synchronization in an oscillator network. The difference between the approach of Burgi and Pun and our model lies mainly in the fact that they used spatio-temporal filters and a convolution algorithm for image analysis whereas our approach is an artificial neural network.

There are, however, several questions associated with the latency mechanism, five of which shall be discussed in the following.

(1) Are the latencies in the visual system long enough ?

Several studies have shown that the visual delays in V1 (or area 17, cat) of the cortex lie approximately within the range of 30 to 130 ms depending on the contrast[2]. Given an oscillation period of between 15 and 30 ms the latency range would cover about 3-6 oscillation cycles which could be enough to subserve our purposes.

(2) Are the latencies narrowly distributed ?

The answer to this question is plainly: No. Statistical fluctuations already at the level of the retina will influence the individual latencies of cortical cells. The X- and Y-subsystems have different propagation delays (Bolz *et al.* 1982, Sestokas *et al.* 1987). The time to response of an individual cell depends on its size and the distribution of the input synapses. These and other effects lead to a significant broadening of the latency distribution at the level of the cortex (Dinse and Krüger 1994). The strong noise robustness of our network (see Fig. 5), however, points to the fact that a rather large degree of broadening in the latency distribution is easily tolerated, so that this problem seems to be of minor relevance.

(3) What happens if objects have only a slight contrast difference ? After all human are able to distinguish between those, too.

[2]This range was estimated from several studies (Levick 1973, Bolz *et al.* 1982, Sestokas *et al.* 1987) and own observations (Wörgötter, Funke and Eysel, unpublished observations).

The major problem of a reliable latency estimate is the question of what to actually measure: The first, second, third, etc. spike ? The time until a significant proportion of the response has built up ? Or other measures. Most of these measures suffer from the effect that they are only accurate if there is no spontaneous activity. In in some cases even more restrictions apply. Thus, latency estimation in real neurons is a non-trivial problem and the above given values can only be regarded as estimates

A slight contrast difference will certainly not lead to truly visible latency differences, it could however be sufficient to induce small initial temporal differences (phase differences) which could also facilitate a mutually undisturbed synchronization. It is conceivable that networks with feedback loops could amplify the small initial differences and result in a good separation of the objects.

(4) What happens when the network changes to a new scene ? For example a bright object in the new scene could "overtake" the dark objects from the old scene which are still processed in the network and this would cause a disturbance of the synchronization process.

The simplest solution which would avoid such a problem would be to introduce a reset mechanism which erases the network activity entirely before a new scene can enter. In the visual system the saccadic eye movements could indeed reflect such a reset mechanism in particular in view of the saccadic suppression which is in most cases associated with saccades in particular in the magnocellular pathway (Burr *et al.* 1994).

(5) What happens if a bright object enters the scene ?

If this happens the network performance could be reduced, but only in the unlikely case that the activity of the bright object would get close enough to other active neurons so as to disturb their internal synchronization. A moving stimulus represents a very salient feature (Niebur and Koch 1996) and it is conceivable that such cases are strongly governed by attentional mechanisms which are not included in our model.

5.3 Conclusion

The intention of this study was not to design a biologically realistic model but rather to outline an artificial neural network which represents an abstraction from the real visual pathway. It was attempted to demonstrate the applicability and the limitations of the approach and in order to achieve this we have restricted ourselves to a very simple architecture which focussed only on the algorithmic principles which were to be tested. The enhancement of assembly formation achieved as the result of the latency mechanism can be used as a first but highly efficient step in image analysis problems. It is very likely that these originally formed cell assemblies provide not more than the basis for an immediately following regrouping by other mechanisms to analyze more complicated visual problems. While we did not attempt this, it is clear that the model provides no restrictions for a possible rearrangement of the activity which could for instance be achieved by feedback from higher visual areas (Sporns *et al.* 1989, Tononi *et al.* 1992).

We were able to show that visual latencies could be used in a rather robust way to enhance image segmentation. If future work should show that the proposed mechanisms to utilize latencies does not reflect the computational reality in the visual pathway, then the intriguing question remains how the brain would analyze a visual scene which is indeed stretched in time by the unequivocally existing latencies ?

6 Appendix

6.1 A) The synchronization process

The process of synchronization of two neurons A and B (Fig. 8A) with excitatory connections is shown in Fig. 8B.

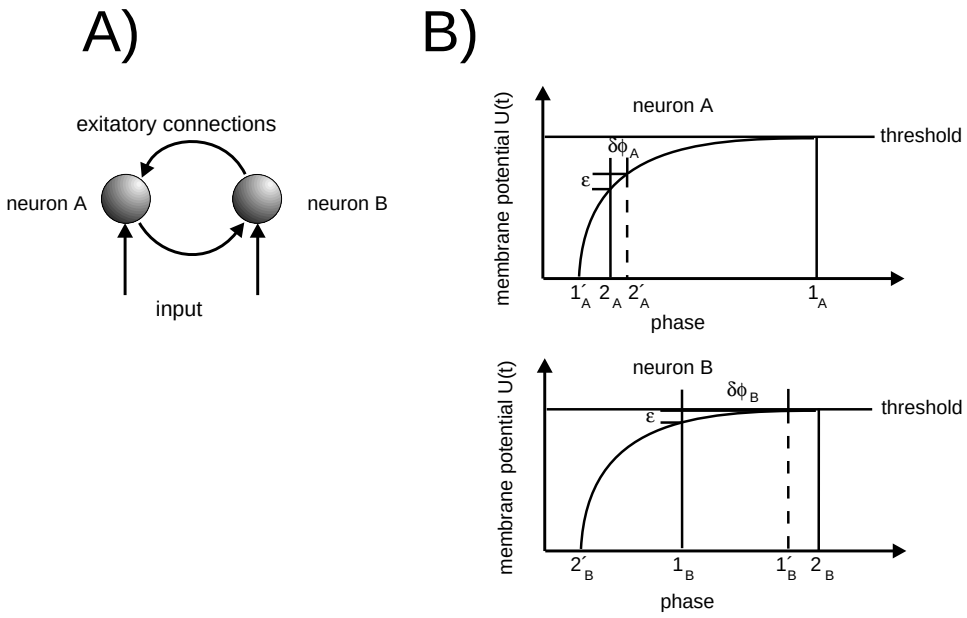

Figure 8: A) Diagram of two excitatorily connected neurons A and B. B) Membrane potential of neurons A and B demonstrating the process of synchronization of the two connected cells.

Both cells are driven by the same stimulus and the initial phase Φ_0 between the oscillators is 180 degree (distance between 1_A,1_B).

If neuron A fires at time (1_A) the membrane potential of neuron B (1_B) is increased by an amount ϵ which in this example induces a shift in phase of about $\delta\Phi_B = 77°$ ($1'_B$). Because of the small gradient of U(t) at 1_B the small shift by ϵ leads to a large shift of the relative phase between neuron A and B (Fig. 8B).

After firing the membrane potential of neuron A is reset to zero ($1'_A$). At time 2_B neuron B reaches the threshold θ and also fires. The membrane potential of neuron

A (2_A) is increased by ϵ ($2'_A$). Although the membrane potential is again shifted by ϵ the effect of the phase shift $\delta\Phi_A = 13°$ is smaller than at time step 1, because the gradient of the membrane potential $U(t)$ at time 2_A is much steeper than at time 1_B.

The resulting phase $\Phi = 180° - \delta\Phi_{eff}$, after both neurons have fired, is decreased by $\delta\Phi_{eff} = \delta\Phi_B - \delta\Phi_A = 77° - 13° = 64°$. This procedure is repeated several times which yields synchronously firing cells already after 3 firing cycles. Note, the results which are presented in this paper are only little affected by the special choice of neural dynamics as long as the dynamic leads to a synchronization between the neurons.

6.2 B) Connectivity parameters

Appendix B shows the parameters used in the simulation shown in Fig. 3. In some of the other simulations small modification of these parameter tables have been made.

The following tables show the connection strengths between neurons. A common multiplication factor is given with which all values have to be multiplied. The leftmost column is the source cell column. All connections are made from those source cells to the target cells with the connection strength as given at the location of the target cell in the table. All tables are symmetric with respect to the source cell column and only one side is plotted. Layers (rows) which do not contain connections are not plotted. The actual direction (lateral, feedback, feedforward) of a connection type is described in the header of each table and the reader is refered to Fig. 1. Inhibitory connections take negative values.

Connection type 1) All basic excitatory feedforward connections from layer l to layer l+1 have a weight of 0.55λ,

with:

$\lambda =$ (200-latency parameter)/200; in layer 1-3,

$\lambda =$ 1; in layer 4-9.

In layer 1-3 λ is varied between 0.25 (latency parameter=150) and 0.9 (latency parameter=15) (Figs. 4-7).

Connection type 2) excitatory and inhibitory connections within a layer:

common multiplication factor: 0.04

Example: Source cell S_9 connects to B_9 with weight 1.0 times 0.04.

source cell	A	B	C	D	E	F	G	layers
S_9	0	1.00	0.80	0.50	-0.25	-0.20	-0.20	9
S_8	0	0.83	0.60	0.41	-0.21	-0.17	0	8
S_7	0	0.66	0.53	0.33	-0.17	0	0	7
S_6	0	0.50	0.40	-0.12	0	0	0	6
S_5	0	0.30	-0.02	0	0	0	0	5
S_4	0	0.20	0	0	0	0	0	4

Connection type 3) excitatory connections from layer l to layer l-1:

common multiplication factor: 0.10

Example: Source cell S_9 connects to B_8 with weight 0.83 times 0.10.

source cell	A	B	C	D	E	F	G	layers
S_9	0	0	0	0	0	0	0	9
S_8	0	0.83	0.41	0	0	0	0	8
S_7	0	0.66	0.33	0	0	0	0	7
S_6	0	0.50	0.25	0	0	0	0	6
S_5	0	0.30	0.16	0	0	0	0	5
S_4	0	0.20	0.10	0	0	0	0	4

7 Acknowledgements

The authors are grateful to E. Nelle for critical comments on the manuscript. F.W. acknowledges the support of the Deutsche Forschungsgemeinschaft WO-388/4-2.

8 References

Abeles, M., Prut, Y., Bergman, H., Vaadia, E. and Aertsen, A. 1993 Integration, synchronizity and periodicity. In: *Brain Theory. Spatio-temporal aspects of brain function* (A. Aertsen, ed.) Elsevier Science Publ. B.V. 149-181.

Bolz, J., Rosner, G. and Wässle, H. 1982. Response latency of brisk-sustained (X) and brisk-transient (Y) cells in the cat retina. H. *J. Physiol*, **328**, 171-190.

Burgi, P. Y. and Pun, T. 1991. Figure-ground separation: evidence for asynchronous processing in visual perception? *Perception* **20**, 69.

Burgi, P. Y. and Pun, T. 1994. Asynchrony in image analysis: using the luminance-to-response-latency relationship to improve segmentation. *J. Opt. Soc. Amer. A* **11**/6, 1720-1726.

Burr, D. C., Morrone, M. C. and Ross, J. 1994. Selective suppression of the magno-cellular visual pathway during saccadic eye movements. *Nature*, **371**, 511-513.

Carney, T., Paradiso, M. A. and Freeman, R. D. 1989. A physiological correlate of the Pulfrich effect in cortical neurons of the cat. *Vision Res.* **29**, 155-165.

Desimone, R. and Duncan, J. 1995. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193-222.

Dinse, H. R. and Krüger, K. 1994. The timing of processing along the visual pathway in the cat. *NeuroReport* **5**, 893-897.

Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M. and Reitböck, H. J. 1988. Coherent oscillations: a mechanism of feature linking in the visual cortex? *Biol. Cybern.* **60**, 121-130.

Eckhorn, R., Fiem, A., Bauer, R., Woelbern, T. and Kehr, H. 1993. High frequency (60-90Hz) oscillations in primary visual cortex of the awake monkey. *NeuroReport* **4**, 243-246.

Enright, J. T. 1985. On Pulfrich-illusion eye movements and accommodation vergence during visual pursuit. *Vision Res.* **25**, 1613-1622.

Ernst, U., Pawelzik, K. and Geisel, T. 1994. Multiple phase clustering of globally pulse coupled neurons with delay. In M. Marinaro & P.G.Morasso (eds.), *ICANN '94: Proceedings of the International Conference on Artificial Neural Networks, Volume I*, (Springer, London), p. 1063-1066.

Fellenz, A. W. 1994. A sequential model for attentive object selection. *Proc. of the IWK 94*, Illmenau, FRG, 109-116.

Gray, C.M., König, P., Engel, A.K. and Singer, W. 1989. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, **338**, 334-337.

Hopfield, J. J. 1995. Pattern recognition computation using action potential timing for stimulus representation. *Nature*, **376**, 33-36.

Ikeda, H and Wright, M.J. 1975. Retinotopic distribution, visual latency and orientation tuning of sustained and transient cortical neurons in area 17 of the cat. *Exp. Brain Res.*, **22**, 385-398.

Levick, W.R. 1973. Variation in the response latency of cat retinal ganglion cells. *Vision Res.*, **13**, 837-853.

v.d. Malsburg, C. 1981. The correlation theory of brain function. Int. report 81-2, Dept. of Neurobiol. Max-Planck-Institute for Biophysical Chemistry, Göttingen.

McClurkin, J. W., Optican, L. M., Richmond, B. J. and Gawne, T. J. 1991. Concurrent processing and complexity of temporally encoded neural messages in visual perception. *Science* **253**, 675-677.

Milanese, R. 1993. Detecting salient regions in an image: from biological evidence to computer implementation. Ph.D. thesis, Univ. of Genova.

Mirollo, R. E. and Strogatz, S. H., 1990. Synchronization of pulse-coupled biological oscillators. *Siam J. Appl. Math.*, **6**, 1645-1662.

Niebur, E. and Koch, C. 1996. Control of selective visual attention: modeling the "where" pathway. In *Proc. of NIPS*, Mozer, M. and Hasselmo, M. (eds.), Morgan Kaufmann, San Mateo, in press.

Niebur, E., Koch, C. and Rosin, C. 1993. An oscillation-based model for the neuronal basis of attention. *Vis. Res.* **33**, 2789-2802.

Olshausen, B. A., Anderson, C. H. and Van Essen, D. C. 1993. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *J. Neurosci.* **13**(11), 4700-4719.

Posner, M. I. and Dehaene, S. 1994. Attentional networks. *TINS* **17**/2.

Raiguel, S.E., Lagae, L. Gulyàs, B. and Orban, G. 1989. Response latencies of visual cells in macaque areas V1, V2 and V5. *Brain Res.* **493**, 155-159.

Sestokas, A.K., Lehmkuhle, S. and Kratz, K.E. 1987. Visual latency of ganglion X- and Y-cells: a comparison with geniculate X- and Y-cells. *Vision Res.*, **27**, 1399-1408.

Singer, W. and Gray, C. M. 1995. Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* **18**, 555-86.

Somers, D. C., Nelson, S. B. and Sur, M. 1995 An emergent model of orientation selectivity in cat visual cortex. *J. Neurosci.* **15**, 5448-5465.

Sporns, O., Gally, J. A., Reeke, Jr., G. N. and Edelman, G. M. 1989 Reentrant signaling among simulated neuronal groups leads to coherency in their oscillatory activity. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7265-7269.

Tononi, G., Sporns, O. and Edelman, G. 1992. Reentry and the problem of integrating multiple cortical areas: simulation of dynamic integration in the visual system. *Cerebral Cortex*, **2**, 310-335.

Whittaker, S.G. and Siegfried, J.B. 1983. Origin of wavelets in the visual evoked potential. *Electroenceph. and Clin. Neurophysiol.*, **55**, 91-101.

Wörgötter, F. and Koch, C. J. 1991. A detailed model of the primary visual pathway in the cat: comparison of afferent excitatory and intracortical inhibitory connection schemes for orientation selectivity. *Neurophysiol.* **11**, 1959-1979.