

Conditions of Activity Bubble Uniqueness in Dynamic Neural Fields

Inna Mikhailova, Christian Goerick

2005

Preprint:

This is an accepted article published in Biological Cybernetics. The final authenticated version is available online at: [https://doi.org/\[DOI not available\]](https://doi.org/[DOI not available])

Conditions of activity bubble uniqueness in dynamic neural fields

Inna Mikhailova, Christian Goerick

Honda Research Institute Europe GmbH, Carl-Legien Str. 30, D-63073 Offenbach/Main, Germany

Received: 24 May 2004 / Accepted: 24 November 2004 / Published online: ■

Abstract. Dynamic neural fields (DNFs) offer a rich spectrum of dynamic properties like hysteresis, spatiotemporal information integration, and coexistence of multiple attractors. These properties make DNFs more and more popular in implementations of sensorimotor loops for autonomous systems. Applications often imply that DNFs should have only one compact region of firing neurons (activity bubble), whereas the rest of the field should not fire (e.g., if the field represents motor commands). In this article we prove the conditions of activity bubble uniqueness in the case of locally symmetric input bubbles. The qualitative condition on inhomogeneous inputs used in earlier work on DNFs is transferred to a quantitative condition of a balance between the internal dynamics and the input. The mathematical analysis is carried out for the two-dimensional case with methods that can be extended to more than two dimensions. The article concludes with an example of how our theoretical results facilitate the practical use of DNFs.

planning in robot navigation tasks (Giese 1999), for attention control in the case of single object tracking (Backer et al. 2001), and for dynamic link matching (Konen et al. 1994). We use the methods developed in the fundamental works on DNFs (Amari 1977; Taylor 1999) to answer the single most important question for such applications: How does one choose the DNF parameters for the given inhomogeneous input so that the stationary solution of the DNF equation has a single region of active neurons?

This question can be formulated in a more general way: How does one preserve the balance between input- and self-driven dynamics? Indeed, on the one hand the impact of the input should be sufficiently large to generate an excitation in the DNF. On the other hand, the competition between excited regions provided by internal dynamics should take care that only one excited region is left after convergence to a stable attractor. In earlier theoretical works (Amari 1977; Taylor 1999) an input “with small inhomogeneity” is considered as one that does not disturb the internal dynamics. We would like to give a quantitative estimation for this qualitative definition.


A closely related problem is discussed in Hahnloser and Seung (2001) for the case of recurrent linear threshold (LT) networks. The authors provide a stability analysis and give a general characterization of neuron sets allowed to be stably coactivated at the fixed point. The description of these sets depends solely on the connection weights matrix and not on the input. For DNFs it is essentially different. Due to the piecewise linearity of LT networks, the local linearization is constant, whereas this is not the case for DNFs with a threshold transfer function (see also (Wersing et al. 2001)). Our results will highlight this by showing the dependency of the stability on the actual profile of the input.

As our work is based on the DNF theory of Amari (1977) and Taylor (1999), we first restate their results (Sect. 2.1). Then we make an analysis of the field dynamics for the case of locally radial-symmetric input (Sect. 2.2). In Sect. 3 we show how the theory can be used in practice. Finally, we discuss the results in Sect. 4.

1 Introduction

In recent years dynamic neural fields (DNFs) have been successfully applied to robot control (Backer et al. 2001; Engels and Schöner 1995) as well as to modeling and understanding the brain (Schneider and Erlhagen 2001; Trappenberg et al. 2001; Giese 1999). In robotics the parameters of DNFs are either determined by evolutionary and learning algorithms or chosen heuristically. In brain modeling the parameters are mostly chosen according to the available neurophysiological experimental data. The influence of the parameter choice on the field dynamics is only roughly estimated. However, the application's semantics often put restrictions on a DNF's behavior. For example, it can be required that the activity bubble in the DNF be unique. This is the case for path

Correspondence to: I. Mikhailova
(e-mail: inna.mikhailova@honda-ri.de,
Tel: +49-69-89011752)

	4	2	2	0	0	5	3	7	B	Despatch: 22/1/2005	Journal: B C Y B	No. of Pages: 10
	Journal No.	Article No.								Author's Disk received <input checked="" type="checkbox"/>	Used <input checked="" type="checkbox"/>	Corrupted <input type="checkbox"/>

2 Analysis of activity bubble dynamics

2.1 Dynamic neural field

The DNF was proposed by Amari (1977) in the late 1970s to model pattern formation in the cortex. He described the neural tissue not as a set of discrete neurons, but as a continuous medium. Let us denote by $u(\mathbf{x}, t) \in \mathbf{R}$, $\mathbf{x} \in \mathbf{R}^2$, $t \in \mathbf{R}$ the state of the DNF. It satisfies the equation

$$\begin{cases} \tau \frac{\partial u(\mathbf{x}, t)}{\partial t} = -u(\mathbf{x}, t) + \int w(\mathbf{x}, \mathbf{x}') f(u(\mathbf{x}', t)) d\mathbf{x}' + h + s(\mathbf{x}, t) \\ u(\mathbf{x}, t_0) = u_0(\mathbf{x}), \end{cases} \quad (1)$$

where $\tau > 0$ is a relaxation constant, h is a rest level, $w(\mathbf{x}, \mathbf{x}')$ are connection weights, and $s(\mathbf{x}, t)$ is an external input. The function $f(u)$ is an activation function that maps u to the interval $[0, 1]$ and is monotonically nondecreasing: $f: u \in \mathbf{R} \rightarrow [0, 1]$.

Without the integral term, (1) would describe a simple exponential decay driven by the input $s(\mathbf{x})$. The integral term in (1) introduces interaction within regions of positive activity $f(u) > 0$. We call such regions **excited regions**. Reducing the field dynamics to the dynamics of the boundaries of these excited regions was the method proposed by Amari (1977) in his pioneer work on DNFs. Amari studied the DNF equation in the one-dimensional case. Taylor (1999) generalized the results of Amari to the two-dimensional case. In this section we restate some important results.

The method of excited regions was developed under the following assumptions:

1. The activation function is a step function with zero threshold:

$$f(u) = \begin{cases} 1, & \text{if } u > 0, \\ 0, & \text{else.} \end{cases}$$

This assumption is made for simplicity. The dynamics of the field would be qualitatively the same for other types of monotonically nondecreasing activation functions with saturation (Amari 1977).

2. Connections between neurons depend only on the Euclidean distance between the neurons:

$$w(\mathbf{x}, \mathbf{x}') = w(\mathbf{x}', \mathbf{x}) = W(\|\mathbf{x} - \mathbf{x}'\|).$$

Let us consider first the case of a homogeneous input only: $s(\mathbf{x}, t) \equiv 0$. It was shown in Taylor (1999) that the only possible shapes of the excited regions in this case are disks, concentric rings, or the whole field. We are especially interested in the case of disks. The radius R of the excited disk

$$D_R = \{\mathbf{x} \in \mathbf{R}^2 \mid f(u(\mathbf{x})) > 0\} = \{\mathbf{x} \in \mathbf{R}^2 \mid \|\mathbf{x} - \mathbf{x}_0\| < R\}$$

is governed by the equation (Taylor 1999)

$$\frac{dR}{dt} = -(G(R) + h)/(\tau v), \quad (2)$$

where $G(R)$ is a **neural interaction force**:

$$G(R) = \int_{D_R} W(\|\mathbf{x} - \mathbf{x}'\|) d\mathbf{x}', \quad \mathbf{x} \in \partial D_R \quad (3)$$

and v is the gradient of the neural field u normal to the boundary of the excited disk ∂D_R and is thus negative.

With this notation the equation for the radius of the excited disk at the fixed point becomes

$$G(R) + h = 0. \quad (4)$$

How many solutions this equation has and which of them correspond to stable stationary solutions of the DNF equation depends on the function $G(R)$ and thus on the connection weights $W(\|\mathbf{x} - \mathbf{x}'\|)$. A weight function of Mexican Hat form is often used:

$$W(r) = E \exp(-r^2/(2\sigma_E^2)) - I \exp(-r^2/(2\sigma_I^2)), \quad (5)$$

where $E > 0$, $\sigma_E^2 > 0$ are the amplitude and variance of the excitatory Gaussian and $I > 0$, $\sigma_I^2 > 0$ are the amplitude and variance of the inhibitory Gaussian. The inhibitory Gaussian is normally set to be wider than the excitatory Gaussian ($\sigma_I^2 > \sigma_E^2$).

Figure 1 shows this function. From Fig. 2, which shows the corresponding integral $G(R)$, one can see that for the considered type of excitatory-inhibitory connections four situations are possible:

1. h is positive and there is no solution to (4) ($-h$ line below $G(R)$ line, $h_3 < h$).
2. h is positive and there is one solution to (4) ($-h$ line intersects $G(R)$ line, $h_2 < h < h_3$).
3. h is negative and there are two solutions to (4) ($-h$ line intersects $G(R)$ line twice, $h_1 < h < h_2$).
4. h is negative and there is no solution to (4) ($-h$ line above $G(R)$ line, $h < h_1$).

Here $-h_3 = \min_{R>0} G(R)$, $h_2 = 0$, and $-h_1 = \max_{R>0} G(R)$. In the first case the field gets excited everywhere because h is positive. In the second case the excited region of the

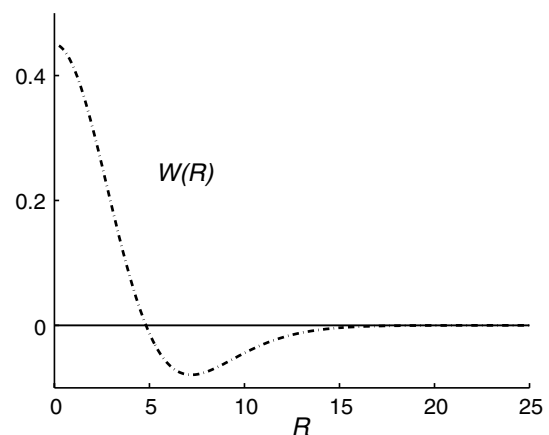


Fig. 1. Excitatory-inhibitory connections. Function $W(R)$ illustrates the typical shape of excitatory ($W > 0$) and inhibitory ($W < 0$) connection weights. The corresponding parameters for (5) are: $E = 0.8$, $I = 0.35$, $\sigma_E = 3.0$, $\sigma_I = 5.0$

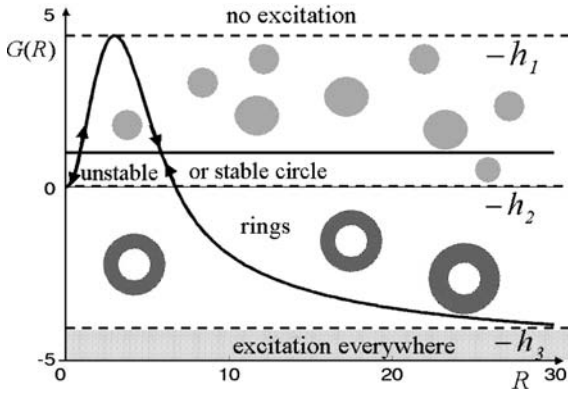


Fig. 2. Classification of excited region's behavior depending on rest level h . The excitation spreads across the complete field if $-h$ is less than the minimum of $G(R)$ and disappears if $-h$ is greater than the maximum of $G(R)$. The behavior in the middle region $0 < -h < \max(G(R))$ depends on the initial radius of the excited region. If the initial radius is too small, the excited region is unstable and disappears

field has the form of rings, centered around the initial excitation (Taylor 1999). In the last case the excitation disappears completely. The third case is the most interesting one. Depending on the initial state of the field, we can get no excitation, excitation everywhere, or excited disks. The radius of the disks converges to one of two possible values. Which of these solutions is a stable one? As v is negative, the solution is a stable attractor if the derivative of the function $G(R)$ is negative at the fixed point.¹ The resulting classification of field attractors for homogeneous inputs is illustrated in Fig. 2.

In Amari (1977) and Taylor (1999) a small stationary inhomogeneous input is also considered. It is proved that an excited region moves in the direction of an increasing input. It does not necessarily find the maximum, as it stops as soon as the input values are the same on the boundary of the excited region.

If the field has more than one excited region, they can be treated separately. For the sake of completeness we repeat here the analysis made in Amari (1977). Let us choose one region and call it region \mathcal{A} . We separate the total interaction [integral term in (1)] into two parts: one coming from the chosen excited region and another coming from all other excited regions (for example, a region we call \mathcal{B}). The second part can be seen as an external input $s(\mathbf{x})$:

$$s(\mathbf{x}) = \int_{\mathcal{B}} W(\|\mathbf{x} - \mathbf{x}'\|) d\mathbf{x}'.$$

Figure 3 shows the typical shape of this input as a function of the distance R between the center $\mathbf{x}_0^{\mathcal{B}}$ of region \mathcal{B} and the considered location x : $S(R) = S(\|\mathbf{x} - \mathbf{x}_0^{\mathcal{B}}\|) = s(\mathbf{x})$. Based on this figure we can distinguish three different regions:

1. $R \in (0, R_E)$: The input provided by interaction with region \mathcal{B} increases as the distance R to the center of the region \mathcal{B} shrinks.

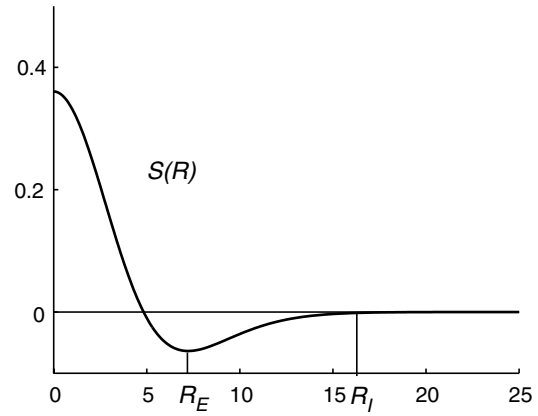


Fig. 3. Influence of the excited region as a function of the distance between the considered location and the center of the excited region. The curve is calculated for the connections presented in Fig. 1 and the excited region of the radius $R = 5$

2. $R \in (R_E, R_I)$: The input provided by interaction increases with the distance R to the center of region \mathcal{B} .
3. $R \in (R_I, \infty)$: The input provided by interaction stays constant.

We already know that an excited region will move in the direction of an increasing input. It means that in the first case region \mathcal{A} will move toward the center of region \mathcal{B} . Since the action of region \mathcal{A} on \mathcal{B} is reciprocal, the regions will attract each other. Arguing in a similar way we conclude that in the second case the regions will repel each other. At a distance greater than R_I , the regions can coexist.

How can we ensure that only one excited region exists? Intuitively it is clear that every excited region [where $u(\mathbf{x})$ is positive] should prevent all remaining parts of the field from becoming positive. This is the case if the area with positive states sends a negative input to all distant areas.

Various proposals on how to achieve such a behavior can be found in the literature. Amari and Arbib (1977) describe a model consisting of a one-dimensional inhibitory DNF $v(x, t)$ and a two-dimensional excitatory DNF $u(x, y, t)$:

$$\begin{cases} \tau_u \frac{\partial u(x, y, t)}{\partial t} = -u(x, y, t) + h_u + s(x, y, t) \\ \quad + \int w_u(x - x', y - y') f(u(x', y', t)) dx' dy' \\ \quad - \int w_v(x - x') g(v(x', t)) dx' \\ \tau_v \frac{\partial v(x, t)}{\partial t} = -v(x, t) + \int f(u(x, y', t)) dy' + h_v \\ u(x, y, t_0) = u_0(x, y) \\ v(x, t_0) = v_0(x) \end{cases}, \quad (6)$$

where $w_u(x, y)$ is a weighting function of excitatory connections within the excitatory field and $w_v(x)$ is a weighting function of connections from the inhibitory field to the excitatory field. Function $g(x)$ is a linear-threshold function:

$$g(u) = \begin{cases} u, & \text{if } u > 0, \\ 0, & \text{else.} \end{cases}$$

For the rest we use the same notations as for system (1): function $f(u)$ is an activation function, τ_u, τ_v are

¹ It is easy to show (Taylor 1999) that if $W(R) \in C^1(0, \infty)$, then also $G(R) \in C^1(0, \infty)$.

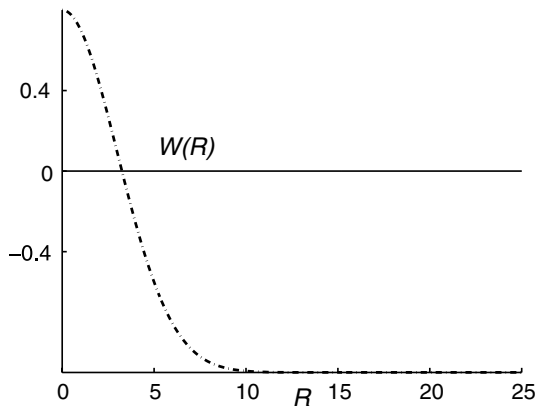


Fig. 4. Example of global inhibition connections

relaxation constants, h_u, h_v are rest levels, and $s(x, y, t)$ is an external input.

Another possibility to introduce inhibition is to use one DNF with connections $w(x, y)$ of the global inhibition type: positive at a short distance and negative at a far distance, as shown in Fig. 4.

In the limit $\tau_v \ll \tau_u, h_v \approx 0$, system (6) and the DNF with global inhibition connections have the same cooperation-competition behavior of excited regions.

Can the uniqueness of the excited region be proved mathematically? In Amari and Arbib (1977) it was done after the connections from the inhibitory field to the excitatory one were simplified to a homogeneous inhibition ($w_v(x) \equiv w_v$) and system (6) was reduced to the system of two one-dimensional DNFs. Another simplification was done by assuming that the input consisted of identical bubbles.

In Konen et al. (1994) a model with global inhibition connections was analyzed. The authors presented a proof for the one-dimensional case in the absence of an inhomogeneous input. The proof uses a Lyapunov functional and gets very complex in the higher-dimensional case.

This concludes our review of previous work. In the next section we present our proof of uniqueness conditions, which is valid for high-dimensional cases and input consisting of different (in radius and shape) symmetric local bubbles.

2.2 Uniqueness of activity bubble

Let us consider a connection function of a global inhibition type:

$$W(\|\mathbf{x} - \mathbf{x}'\|) = W_E(\|\mathbf{x} - \mathbf{x}'\|) - I,$$

where the function W_E has a finite support ($W_E(R) \equiv 0$, for $R \geq R_{\max}$) and I is the global inhibition level ($I > 0$). It is often assumed that the weighting function W_E is non-negative so that we can call W_E the excitatory part of the connections. This condition ($W_E(R) \geq 0$) is not necessary for the following analysis.

For this type of connection the input provided from an excited disk D_R to its own boundary becomes:

$$\begin{aligned} G(R) &= \int_{D_R} W_E(\|\mathbf{x} - \mathbf{x}'\|) d\mathbf{x}' - I\pi R^2 \\ &= G_E(R) - I\pi R^2, \quad \mathbf{x} \in \partial D_R. \end{aligned} \quad (7)$$

It was shown in Taylor (1999) that $G_E(R) \in C^1(0, \infty)$ if $W_E(R) \in C^1(0, \infty)$.

Assume now that the excitation in the field is produced by n symmetric nonoverlapping stationary input bubbles. This assumption makes sense because in brain modeling the input to a DNF is usually either an output of another DNF or a data representation in the form of activity bubbles. The profile of each bubble is locally described by the function $S_i(R)$, $R = \|\mathbf{x} - \mathbf{x}_i\|$, where \mathbf{x}_i is the center of the i th bubble and $S_i(R)$ has a finite support: $S_i(R) \equiv 0$, for $R \geq R_i^{\max}$. We assume that the input bubbles are not overlapping. With these notations the inhomogeneous stationary input to the DNF is $s(\mathbf{x}) = \sum_{i=1}^n S_i(\|\mathbf{x} - \mathbf{x}_i\|)$. For this particular input it is possible to reduce the field dynamics to the dynamics of the boundaries of excited disks in a way similar to the one proposed in Amari (1977) and Taylor (1999) for the homogeneous input.

If the bubbles are close to each other, then excited regions will merge to one region due to cooperation (Fig. 3). Let us suppose that the input bubbles are far from each other. Then the input of the i th excited disk to the boundary of all other excited disks is $-I\pi R_i^2$. Using (2) we obtain the system of equations that describes the excited disks' dynamics

$$\begin{aligned} \dot{R}_1 &= -(G_1(R_1) + S_1(R_1) + h - I\pi R_2^2 \cdots - I\pi R_n^2) / (\tau v_1), \\ &\vdots \\ \dot{R}_n &= -(G_n(R_n) + S_n(R_n) + h - I\pi R_1^2 \cdots - I\pi R_{n-1}^2) / (\tau v_n), \end{aligned} \quad (8)$$

with initial conditions $R_i(0) = 0$, $i = 1, \dots, n$. At the fixed point the radii R_i of excited disks satisfy the system of equations

$$G_E(R_i) + S_i(R_i) = I\pi \sum_{j=1}^n R_j^2 - h, \quad i = 1, \dots, n, \quad (9)$$

where we replaced $G_i(R_i)$ with (7). In the appendix we give the proof of the following theorem:

Theorem 1 *Let $G_E(R)$ and $S_i(R)$, $i = 1, \dots, n$ have continuous derivatives for $R \in (0, \infty)$. Let $\mathbf{R} = (R_1, \dots, R_n)$ be a fixed point with multiple excitation ($n > 1$ and $R_i > 0$). If the inequality*

$$G'_E(R_i) + S'_i(R_i) < 0 \quad (10)$$

holds for all radii R_i , then this multiple excitation is stable. If the inequality

$$G'_E(R_i) + S'_i(R_i) > 0 \quad (11)$$

holds at least for two radii R_k, R_l , then this multiple excitation is unstable.

Conditions (10) and (11) can be seen as conditions of a balance between the influence of the input (expressed by the slope of the bubble profile S'_i) and the internal dynamics (measured by the growth of positive neural interaction G'_E). Condition (11) rewritten as

$$-S'_i(R_i) < G'_E(R_i) \quad (12)$$

gives a quantitative bound on how inhomogeneous the input is allowed to be if we want to preserve the uniqueness of activity bubbles in DNFs. This quantitative bound can be better handled in practice than the descriptions of small inhomogeneities of the form “ $\|s(\mathbf{x})\| \ll -h$ ” or “ $s(\mathbf{x}) = \varepsilon \tilde{s}(\mathbf{x})$ with small ε ,” which have been used until now (Konen et al. 1994; Amari 1977).

In the case of an arbitrary input we cannot make an exact theoretical analysis. Still, we can argue in a similar way if the input can be approximated with symmetric bubbles.

The extension to higher dimensions is straightforward. Instead of an excited disk, we can consider an excited n -dimensional sphere. Then we just need to exchange the integral over the excited region in the definition of the neural interaction $G(R)$ by a higher-dimensional integral.

In the next section we show how the developed theory can be used in practice.

3 Using a DNF as a selection operator

3.1 Appropriate parameter choice

In order to illustrate the practical use of Theorem 1, we consider using a DNF as a selection operator. At the input the DNF gets two identical Gaussian bubbles:

$$s(\mathbf{x}) = S(\|\mathbf{x} - \mathbf{x}_1\|) + S(\|\mathbf{x} - \mathbf{x}_2\|),$$

with $S(R) = A \exp(-R^2/(2\sigma^2))$. The task of the DNF is to select one of these input bubbles. This means that at the fixed point we expect only one activity bubble left at the location of one of the input bubbles. In order to break the symmetry of the input, a small white Gaussian noise is added to the input. If the condition of the theorem is fulfilled, the fixed point with multiple activity bubbles is unstable and a small noise is sufficient to make the DNF converge to the fixed point with only one activity bubble.

Besides the uniqueness of the activity bubble we impose the following restrictions:

- In the absence of an input there should be no activity bubbles in the field.
- If an input is larger than a fixed activation threshold S_{\min} , then the field should get activated.

These requirements imply (Fig. 2) that

$$\max_R G(R) < -h \leq S_{\min}.$$

Thus we will get activity in the DNF only if the amplitude A of input bubbles is larger than $\max_R G(R)$. On the other hand, the amplitude of input bubbles is bounded by the uniqueness condition $-S'(R) < G'_E(R)$. Together these two restrictions define the appropriate range of the DNF parameters for the given range of the input.

Let us fix the variance σ^2 of the input bubble at 9.0 and let the amplitude vary from 0.0 to 1.5. We set the activation threshold in our example to $S_{\min} = 0.7$ and choose the smallest possible rest level $-h = S_{\min} = 0.7$

To keep the calculations simple, we choose constant connection weights in the DNF:

$$W(R) = \begin{cases} E, & \text{if } R < R_{\max}, \\ -I, & \text{else.} \end{cases}$$

In this simple case the integral $G(R)$ can be calculated analytically:

$$\begin{aligned} G(R) &= G_E(R) - I\pi R^2 \\ G_E(R) &= (E + I)R^2 [\pi + \alpha(R) \cos(\alpha(R)) - \sin(\alpha(R))] \\ G'_E(R) &= (E + I)2R [\pi - \alpha(R) - \sin(\alpha(R))], \end{aligned} \quad (13)$$

where α is defined by

$$\alpha(R) = \begin{cases} 2 \arccos(R_{\max}/(2R)), & \text{if } R > R_{\max}/2 \\ 0, & \text{else.} \end{cases}$$

The price for simplicity is the discontinuity of the derivative of $G_E(R)$ at the point $R = R_{\max}/2$. In practice we could smooth the connection weights and thus the considered derivative without affecting much the resulting field behavior. The following theoretical considerations are carried out outside of the neighborhood of the discontinuity.

We fix the radius of the excitatory connections at $R_{\max} = 5$ in order not to blur the input bubbles too much. Taking into account the condition on the rest level h ($\max_R G(R) < -h$), we choose $E = 0.025$ and $I = 0.03$. Now we must concern ourselves with the uniqueness condition. Figure 5 shows the gradient of the curves $G_E(R)$ and $S(R)$. From this figure we see that theoretically two-bubble solutions could be stable for the chosen parameters.

We test now the impact of changing the amplitude A in the simulation. As long as the amplitude is less than 0.7 we

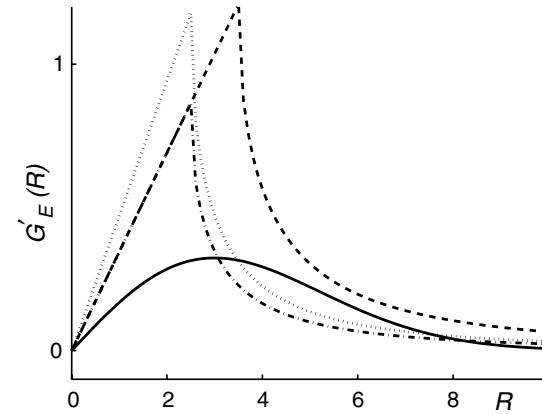


Fig. 5. Parameter tuning of the DNF for restoring the uniqueness condition (Fig. 6 a–c). The *solid line* represents the slope of the cross section of the input bubble $-S'(R)$ for amplitude $A = 1.6$ and variance $\sigma^2 = 9$. Other *lines* show the growth of a positive interaction force $G'_E(R)$ for different parameter choices. The uniqueness condition for activity bubble $-S'(R) < G'_E(R)$ implies that all these *lines* should lie above the *solid line*. 1 *Dot-dashed line* $E = 0.025$, $I = 0.03$, $R_{\max} = 5$. The uniqueness condition is not fulfilled for a wide range of R . 2 *Dotted line* $E = 0.025$, $I = 0.05$, $R_{\max} = 5$. The uniqueness condition still fails on a large interval. However, due to the increase of inhibition, the fixed points of system (8) do not fall into this interval. Thus the uniqueness of an activity bubble is provided with this choice of parameters (Fig. 6 a–c, case 3). 3 *Dashed line* $E = 0.025$, $I = 0.03$, $R_{\max} = 7$. The uniqueness condition is fulfilled for all fixed points (Fig. 6 a–c, case 4)

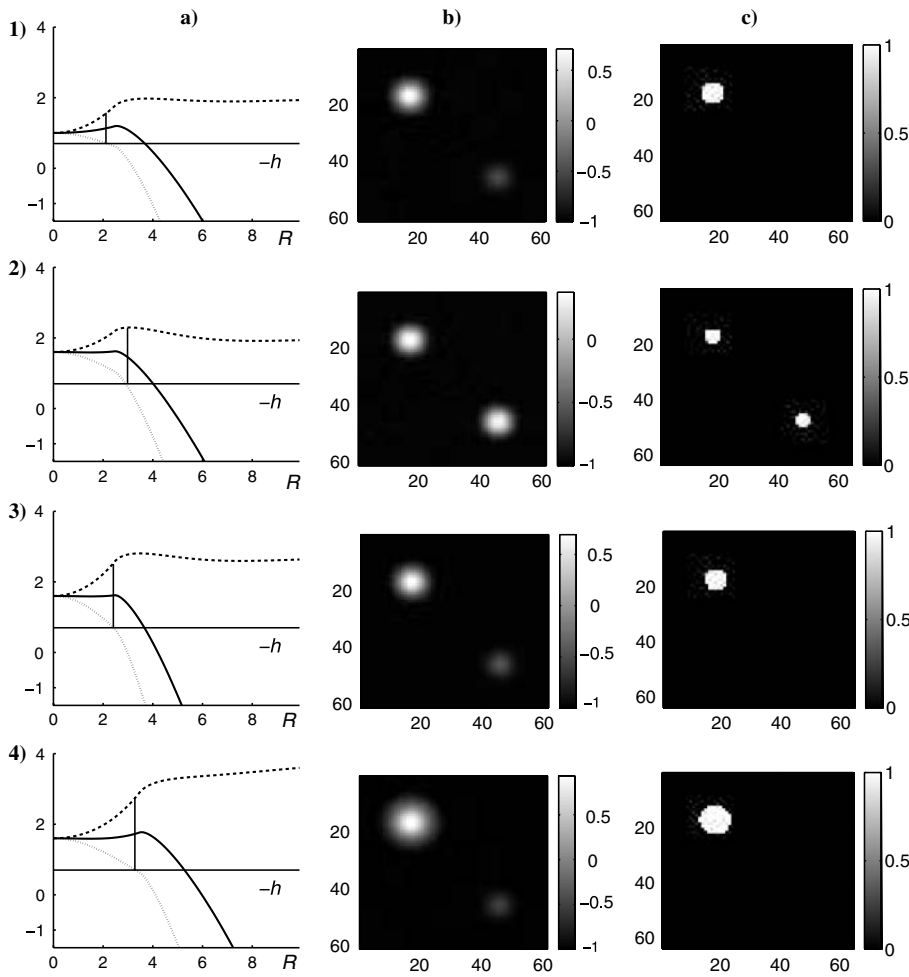


Fig. 6 a–c. Violation of the uniqueness condition. The input has two Gaussian bubbles with variance $\sigma^2 = 9$ and amplitude $S(0)$: 1) $S(0) = 1.0$; 2), 3), 4) $S(0) = 1.6$. With the growth of the amplitude the uniqueness condition becomes violated (case 2). In order to restore the uniqueness one can increase the inhibitory part of connections (case 3) or enlarge the radius of the excitatory connections (case 4). **a** Theoretical uniqueness condition. The dashed line is $G_E(R) + S(R)$. The solid line is $G(R) + S(R)$. The dotted line is $G_E(R) + S(R) - 2I\pi R^2$. The intersection of the dotted line with $-h$ line determines the radius R of the excited circle in the system of two circles. If at this radius the derivative of $G_E(R) + S(R)$ (dashed line) is not positive, then the system of two excited circles is stable. **b** The DNF at the fixed point. **c** Excited regions in the DNF at the fixed point

get no activity at all. In the range of the amplitude between $S(0) = 0.7$ and $S(0) = 1.2$ we have a single active region, and thus the DNF can be applied as a selection operator (Fig. 6 a–c, case 1). As soon as condition (11) of Theorem 1 is no longer fulfilled we get two active regions (Fig. 6 a–c, case 2). The intuitive remedy is to increase the inhibition I , which indeed helps (Fig. 6 a–c, case 3, $I = 0.05$). A less intuitive possibility of parameter tuning is to augment the radius R of excitatory connections (Fig. 6 a–c, case 4, $R = 7$). Figure 5 shows a comparison of the function $G'_E(R)$ for these different choices of the DNF parameters. This figure shows how, with the help of Theorem 1, we can estimate either *the range of input* where we have the uniqueness of an activity bubble for given weight connections or *the appropriate connection weights* for a given input.

3.2 Dynamical properties of selection

Until now we have considered a stationary input to a DNF. Next we show some important dynamical properties of DNFs that explain the advantage of using DNFs as a selection operator. Let us suppose that the input and the DNF are changing on different time scales, so that the DNF has enough time to converge to the fixed point every time the input changes. We do not conduct a full time anal-

ysis of bubble formation (Taylor 1999) but only follow the change of the excited region's location after convergence of the DNF to the fixed point.

Our test is schematically explained in Fig. 7. Without resetting the state of the neural field we supply five different inputs and let the DNF converge to the fixed point

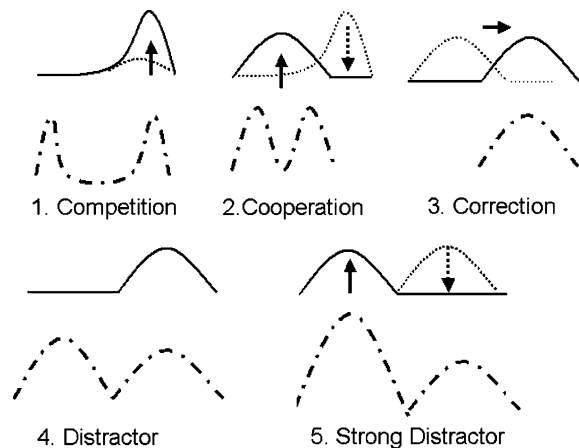


Fig. 7. Test of the DNF properties. The dash-dotted lines schematically represent the input. The solid lines represent the desired output of the neural field. The dotted lines show the previous stable state of the DNF, and the arrows indicate the direction of changes in the DNF

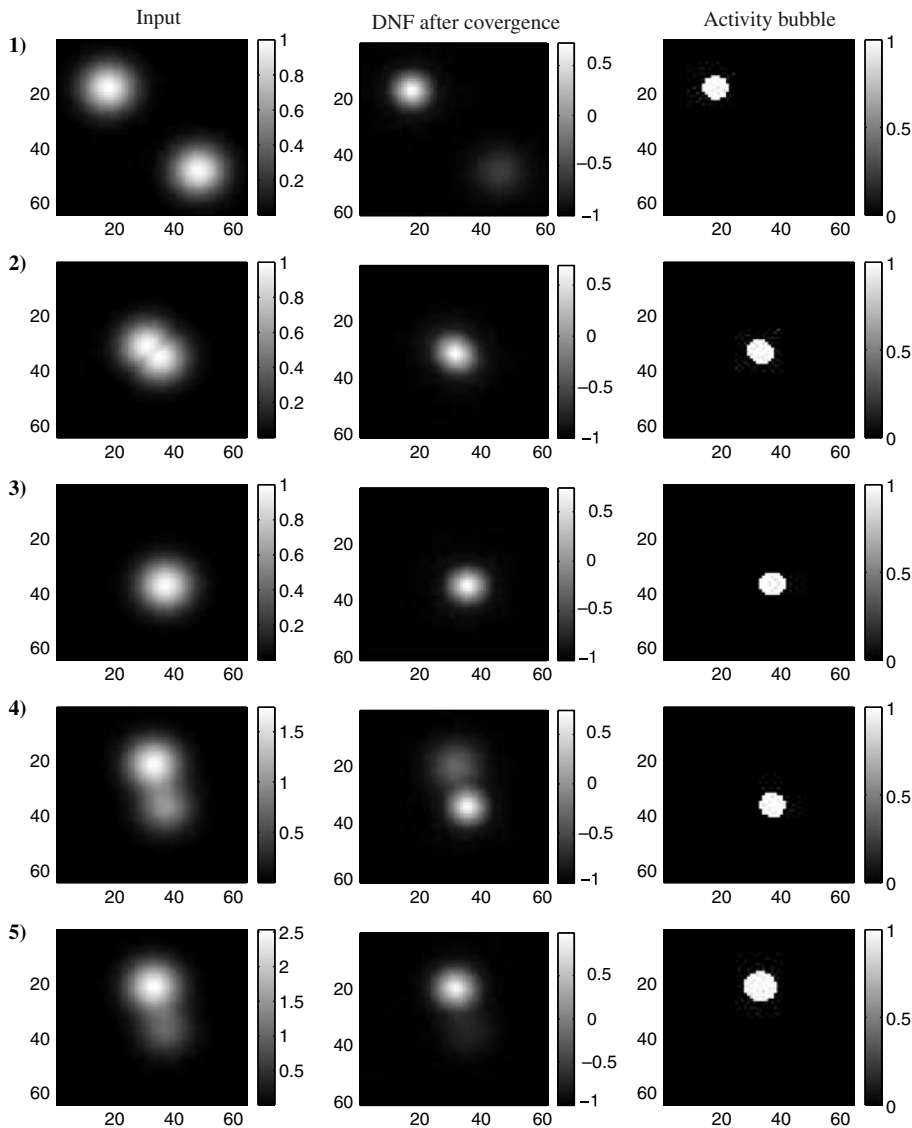


Fig. 8. Simulation results. 1 Competition. 2 Cooperation. 3 Correction. 4 Distractor. 5 Strong distractor. For more details see text

every time. Situation 1: We start with the DNF that has no active bubbles. However, one region is “preactivated.” This means that the value of the DNF in this region is slightly higher than in the rest of the field. As soon as two identical input bubbles are introduced, the DNF becomes active at the location of the bubble that is closer to the preactivated region. Thus in the “competition” of two input bubbles the preactivated location wins. Situation 2: The “cooperation” situation illustrates the ability of an excitation to merge to one region if maxima of an input are close to each other. Situation 3: The “correction” situation shows how an excited region follows the movement of a local input maximum. Situation 4: The excited region stays fixed despite changes in the input, which are far from the actual excitation location (“distractor” situation). Situation 5: If the distractor is strong enough, the excitation moves to a new maximum. Figure 8 presents the corresponding simulation results.

The test described above illustrates properties of DNF dynamics that are very useful for applications. These properties are:

- Conditioned selection (situation 1);
- Spatiotemporal integration (situations 1, 2, 3); and
- Hysteresis (situations 4 and 5).

These properties were used in the implementation of a gaze direction control for an active robot head. The robot has a neck with two degrees of freedom and a head with two cameras. From the visual input of the camera a saliency map of the viewed scene is generated (for more details see Itti et al. 1998). By moving the head the robot should direct its gaze toward the most salient object in the scene and fixate this object for some time. This task means that the robot must

- Correct the gaze direction if the movement of the head has not been precise,
- Focus on one object despite distractors, and
- React to the changes in the environment.

A simple maximum operator used for the choice of the most salient location would completely fail the first two requirements. Indeed, if the visual scene changes (due to head movement or due to distractors) the maximum of the

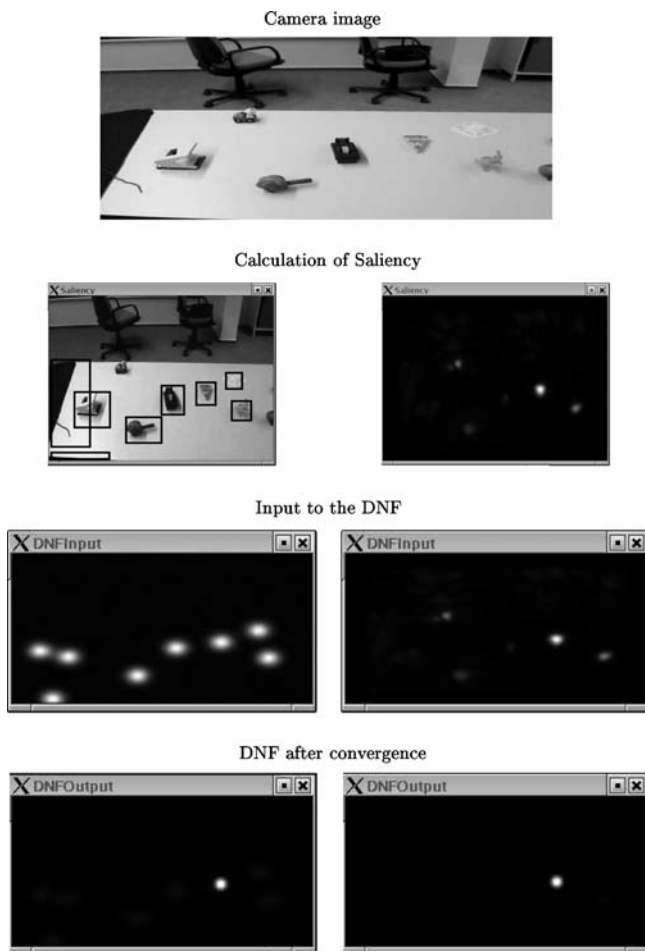


Fig. 9. Practical test for using a DNF as a selection operator. *Left side:* Saliency map is built with help of segmentation. *Right side:* Saliency uses complex features (colors, orientations). In both cases the DNF has only one region of active neurons after convergence. This region can be used to control the direction of the robot head

saliency can change its location from one salient object to another.

Instead of a maximum operator we use a DNF as a selection operator. We feed a saliency map into a DNF with global lateral inhibition. After convergence of the DNF to the stable state we can choose either the center of gravity of an excited region or the location of the maximum excitation for the control of the robot's gaze direction. If condition (11) of Theorem 1 is fulfilled, the DNF has only one active region. Therefore, the choice is unambiguous. The dynamical properties of the DNF described above provide both reaction and fixation.

Figure 9 presents the results for two versions of a saliency map. On the left side of the figure the saliency map has locally symmetric bubbles, as considered in Theorem 1. They are generated by putting identical bubbles in the regions where the objects are detected by segmentation (black rectangles). As the condition of Theorem 1 is fulfilled, we get only one excited region in the DNF after convergence. The right side of Fig. 9 shows a commonly used saliency map (Itti et al. 1998). In this case we also observe

the uniqueness of activity bubbles if the DNF parameters were chosen under consideration of Theorem 1.

Figure 10 compares two selection mechanisms: one using a DNF and another using resetting of selective neurons after every decision (Itti et al. 1998). Among the objects on the table are two ducks that produce very similar saliency blobs. What makes a gaze selection difficult is the fact that from the position of the camera focusing on the big duck the most salient point is located on the small duck, and vice versa. For this reason all algorithms that are input-driven only, like the one proposed in Itti et al. (1998), are unable to make a catchup saccade and fixate an object stably. Indeed one can see that the gaze is oscillating between the locations of the big and small duck. By contrast, the algorithm using a DNF is able to produce catchup saccades and fixations of objects.

4 Summary

In this paper we considered a DNF with global inhibition, which is frequently used in biologically motivated control applications. The stability analysis of DNF attractors led to a quantitative condition of activity bubble uniqueness in DNFs. This condition expresses a balance between the internal dynamics and the influence of the input. Already in the pioneer work on DNFs (Amari 1977) it was mentioned that "When the input is strong compared with the mutual excitation and inhibition, it will dominate the solution." We translated this statement into a more precise mathematical language for the case of high-dimensional DNFs with stationary locally symmetric input.

We validated our results in the simulations and applied them to the implementation of gaze direction selection for an active robot head. Due to the balance between the internal dynamics and the input's influence, the robot head was sufficiently self-driven to fixate objects and sufficiently input-driven to take changes in the scene into account.

Our findings are especially interesting for the applications using DNFs as a maximum selection operator. These are the applications where such properties of selection like hysteresis and spatiotemporal integration are desired [e.g., navigation (Giese 1999), attention (Backer et al. 2001)]. The uniqueness of an activity bubble is crucial for the unambiguity of the selection, whereas the sufficient impact of the input is needed in order to react to the strong changes in the input. The proposed condition helps to find the range of the input (resp. the appropriate DNF parameters) where both features are provided.

Due to the high nonlinearity of the DNF dynamics a theoretical analysis was conducted only for the case of stationary input with locally symmetric bubbles. However, the proposed condition clarifies the theoretical question of activity bubble uniqueness and eases the handling of DNFs in practice.

Appendix: Analysis of Jacobian eigenvalues

In order to simplify the calculations, let us use the notations $d_i = G'_E(R_i) + S'_i(R_i)$ and $a_i = -2\pi I R_i$. With these

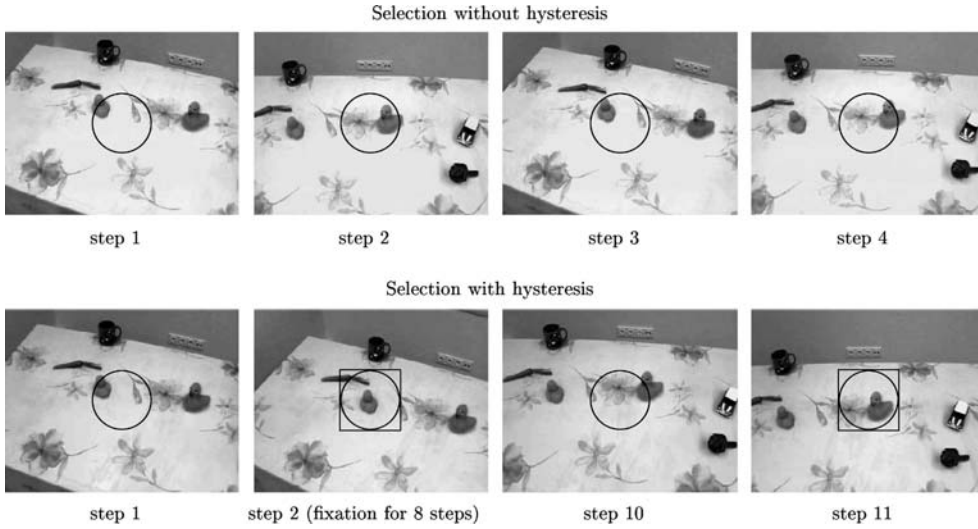


Fig. 10. Advantages of selection hysteresis. For visualization purposes the center of the camera image is marked by a *circle* and fixations are marked by a *square*. *Upper row:* Algorithm using resetting of neurons after every selection decision. The gaze direction is oscillating between the location of the small and the big duck without

fixating them. *Lower row:* Algorithm using a DNF. After a catchup saccade (steps 1–2) the gaze is fixating the small duck. After fixation time is over, the big duck is selected. Unperturbed by the fact that the maximal saliency is located at the small duck, a catchup saccade is done (steps 10–11)

notations the Jacobian of system (8) at the fixed point is given by the matrix

$$J = 1/\tau \begin{pmatrix} -(d_1 + a_1)/v_1 & -a_2/v_1 & \dots & -a_n/v_1 \\ -a_1/v_2 & \ddots & & \vdots \\ \vdots & & \ddots & -a_n/v_{n-1} \\ -a_1/v_n & \dots & & -(d_n + a_n)/v_n \end{pmatrix}.$$

The terms that include the derivative of $1/v_i$ vanish at the fixed point because they are multiplied by the terms that are zero at the fixed point.

Let us estimate the eigenvalues of the Jacobian J .

$$\det(J - \lambda \mathbf{E}) = \prod_{i=1}^n \left(-\frac{1}{\tau v_i} \right) \begin{vmatrix} d_1 + a_1 + \lambda \tau v_1 & a_2 & \dots & a_n \\ a_1 & \ddots & & \vdots \\ \vdots & & \ddots & a_n \\ a_1 & \dots & & d_n + a_n + \lambda \tau v_n \end{vmatrix}.$$

We set $b_i = d_i + \lambda \tau v_i$ and calculate the considered determinant by induction:

$$D_n = \begin{vmatrix} b_1 + a_1 & a_2 & \dots & a_n \\ a_1 & b_2 + a_2 & & \vdots \\ \vdots & & \ddots & a_n \\ a_1 & \dots & & b_n + a_n \\ b_1 & 0 & \dots & -b_{n-1} & 0 \\ 0 & b_2 & 0 & \vdots & \vdots \\ \vdots & \ddots & -b_{n-1} & 0 & \vdots \\ a_1 & a_2 & \dots & a_{n-1} + b_{n-1} & a_n \\ 0 & \dots & 0 & -b_{n-1} & b_n \end{vmatrix} =$$

$$= b_n \begin{vmatrix} b_1 & 0 & \dots & -b_{n-1} \\ 0 & b_2 & 0 & \vdots \\ \vdots & \ddots & -b_{n-1} & \\ a_1 & a_2 & \dots & a_{n-1} + b_{n-1} \end{vmatrix} - a_n \begin{vmatrix} b_1 & 0 & \dots & -b_{n-1} \\ 0 & b_2 & 0 & \vdots \\ \vdots & \ddots & -b_{n-1} & \\ 0 & \dots & 0 & -b_{n-1} \end{vmatrix} = b_n D_{n-1} + a_n \prod_{i=1}^{n-1} b_i.$$

If we now suppose that

$$D_{n-1} = \sum_{i=1}^{n-1} a_i \prod_{j=1, j \neq i}^{n-1} b_j + \prod_{i=1}^{n-1} b_i,$$

then we get

$$D_n = \sum_{i=1}^n a_i \prod_{j=1, j \neq i}^n b_j + \prod_{i=1}^n b_i.$$

We now use this result for estimating the eigenvalues of the Jacobian:

$$f(\lambda) = \det(J - \lambda \mathbf{E}) = \prod_{k=1}^n \left(-\frac{1}{\tau v_k} \right) * \left[\sum_{i=1}^n a_i \prod_{j=1, j \neq i}^n (d_j + \lambda \tau v_j) + \prod_{i=1}^n (d_i + \lambda \tau v_i) \right].$$

To shorten the last expression, we use the variables $\alpha_i = -a_i/(\tau v_i)$ and $\beta_i = -d_i/(\tau v_i)$. Then we get the following expression for the determinant $f(\lambda)$:

$$f(\lambda) = \sum_{i=1}^n \alpha_i \prod_{j=1, j \neq i}^n (\beta_j - \lambda) + \prod_{i=1}^n (\beta_i - \lambda). \quad (14)$$

Without loss of generality we can assume that the values β_i are sorted in ascending order: $\beta_i \leq \beta_{i+1}$, $i = 1, \dots, n$. We will show next that every interval $[\beta_i, \beta_{i+1}]$ contains a zero crossing of the function $f(\lambda)$ and thus an eigenvalue of the Jacobian.

At the point $\lambda = \beta_i$ all terms in the sum (14) vanish except for the i th term:

$$f(\beta_i) = \alpha_i \prod_{j=1, j \neq i}^n (\beta_j - \beta_i). \quad (15)$$

Similarly,

$$f(\beta_{i+1}) = \alpha_{i+1} \prod_{j=1, j \neq i+1}^n (\beta_j - \beta_{i+1}). \quad (16)$$

We recall that

$$\alpha_i = -\frac{a_i}{\tau v_i} = \frac{2\pi I R_i}{\tau v_i},$$

where v_i is the gradient of neural field u normal to the boundary of the i th excited circle and is thus negative for all i . All radii R_i are positive. Hence, α_i and α_{i+1} have the same sign. Now we compare the signs of the other multiplicands in products (15) and (16). As all β_i are ordered, the terms $\beta_j - \beta_i$ [(15)] and $\beta_j - \beta_{i+1}$ [(16)] have the same sign for all j except $j = i + 1$ [(15)] and $j = i$ (16). These terms are either both zero ($\beta_i = \beta_{i+1}$) or have the opposite signs. In the first case we have an explicit root of $f(\lambda) = 0$: $\lambda = \beta_i = \beta_{i+1}$. In the second case, the products in (15) and (16) have opposite signs. This means that the continuous function $f(\lambda)$ changes its sign in the interval $[\beta_i, \beta_{i+1}]$ and thus has a zero crossing in this interval.

We showed that every interval $[\beta_i, \beta_{i+1}]$ contains an eigenvalue of the Jacobian. These are $n - 1$ eigenvalues. The analysis of the sign of the function $f(\lambda)$ shows that the n th value belongs to the interval $(-\infty, \beta_1]$ and there are no eigenvalues in the interval (β_n, ∞) . In order to complete the analysis, we need to know whether β_i are positive or negative. We recall that

$$\beta_i = -\frac{d_i}{\tau v_i} = -\left(G'_E(R_i) + \frac{S'_i(R_i)}{\tau v_i}\right),$$

where v_i are negative for all i . Hence, if $G'_E(R_i) + S'_i(R_i)$ are negative for all radii, then all β_i are negative and all

eigenvalues are negative. If $G'_E(R_i) + S'_i(R_i) > 0$ holds at least for two radii, then we have two positive β_i . Therefore, the Jacobian has at least one positive eigenvalue. Theorem 1 is proved.

Acknowledgements. We would like to thank Edgar Körner for focusing our attention on the issues of this article and the entire Honda Research Institute Europe team, which is engaged in construction of the demonstration platform used in our experiments. Special thanks go to Heiko Wersing and Mark Dunn for careful proofreading and providing helpful comments on the manuscript. We are also grateful to Julian Eggert and Frank Joublin for stimulating discussions.

References

- Amari S (1977) Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol Cybern* 27:77–87
- Amari S, Arbib MA (1977) Competition and cooperation in neural nets. *Syst Neurosci* 119–165
- Backer G, Mertsching B, Bollmann M (2001) Data- and model-driven gaze control for an active-vision system. *IEEE Trans Pattern Anal Mach Intell* 23(12):1415–1429
- Engels C, Schöner G (1995) Dynamic fields endow behavior based robots with representation. *Robot Autonomous Syst* 14:55–77
- Giese M (1999) Dynamic neural field theory for motion perception. Kluwer, Dordrecht
- Hahnloser RHR, Seung HS (2001) Permitted and forbidden sets in symmetrical threshold linear networks. *Proc Nips* 13:217–223
- Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans Pattern Anal Mach Intell* 20(11):1254–1259
- Konen WK, Mauer T, von der Malsburg C (1994) A fast dynamic link matching algorithm for invariant pattern recognition. *Neural Netw* 7:1019–1030
- Schneider S, Erlhagen W (2001) A neural field model for saccade planning in the superior colliculus: speed-accuracy tradeoff in the double-target paradigm. In: 10th annual computational neuroscience meeting
- Taylor JG (1999) Neural ‘bubble’ dynamics in two dimensions: foundations. *Biol Cybern* 80:393–409
- Trappenberg TP, Dorris M, Klein RM, Munoz DP (2001) A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus. *J Cogn Neurosci* 13:256–271
- Wersing H, Beyn WJ, Ritter H (2001) Dynamical stability conditions for recurrent neural networks with unsaturating piecewise linear transfer functions. *Neural Comput* 13(8):1811–1825