

Visual channels, Hebbian assemblies and the effect of Hebb's rule

U. Mortensen, C. Nachtigall

Received: 10 May 1999 / Accepted in revised form: 22 October 1999

Abstract. A model of sensory learning is proposed that is based upon Hebb's rule, where Hebb's rule has been generalised by introducing a stabilising function representing some feedback process within or at the adapting (cortical) neuron, preventing synaptic weights from increasing without limit. It will be shown that neurons adapting according to this stabilised Hebb rule will turn into a matched filter for that part of the stimulus pattern that covers the receptive field of a neuron. It follows that the presentation of a stimulus pattern may imply the formation of a set of neurons with overlapping receptive fields, where each neuron has adapted to a certain part of the stimulus. Making simplifying assumptions about the detection process, the model will be illustrated, fitting it to data from Meinhardt and Mortensen [Meinhardt G, Mortensen U (1998) *Biol Cybern* 79:413–425] which are not compatible with the classical matched filter model introduced by Hauske et al. [Hauske G, Wolf W, Lupp U (1976) *Biol Cybern* 22:181–188].

1 Introduction

For simple visual patterns such as bars, edges, circular discs, etc. it is often assumed that (1) stimulus processing is bottom-up, and (2) the stimulus patterns are represented by the activity of certain "channels" whose characteristics do not change during the course of an experiment. Further, it is postulated that a given pattern is either detected by a single, approximately linear channel, or that detection is the result of probability summation or nonlinear pooling of activities among different channels. The channels are assumed to be spatial frequency channels (Campbell and Robson 1968), or having receptive field structures defined by Hermite functions (Koenderink and van Doorn 1990; Yang and Reeves 1995), or difference of Gaussian (DOG) functions (Wil-

son and Bergen 1979). Sachs et al. (1971) and Graham (1977) argued that detection is by probability summation among spatial frequency channels (for a review, see Graham 1989). Graham (1977) adopted Quick's (1974) model of detection, which may, quite generally, be called a p -norm model. Such a model had already been employed by Mostafavi and Sakrison (1976) to discuss the role of spatial frequencies in detection tasks. Wilson and Bergen (1979) employed the model assuming DOG functions to describe the basic channels; Daugman (1984) introduced Gabor functions and du Buf (1992, 1993, 1994) employed quadrature pairs of Gabor functions to represent the basic channels of the visual system.

Evidence that the assumption of a hardwired system providing a representation of the sensory input may not be quite adequate comes from the work of Gilbert and Wiesel (1992), Pettet and Gilbert (1992), Darian-Smith and Gilbert (1994, 1995). Das and Gilbert (1995) showed that the size of receptive fields of V1 neurons may change depending upon experimental conditions, although the experiments of these authors differ from those usually performed in visual psychophysics (creation of artificial scotoma). Caelli and Oğuztöreli (1987) summarise and provide arguments and psychophysical data supporting the conceptualisation of the visual system as an adaptive system. Poggio et al. (1992), Beard et al. (1995) and Kirkwood et al. (1996) provided data from perceptual learning experiments indicating that the visual system adapts during experimental sessions. Karni and Sagi (1991) also argued, referring to psychophysical data, that adaptational processes take place already at the level of V1, a view shared by Christ et al. (1997), Ahissar et al. (1998), however, provided evidence that perceptual learning does not manifest itself at V1 but at higher levels of the visual system.

In order to model the process of learning during an experiment, we will assume that learning can be modelled invoking Hebb's rule (Hebb 1949), according to which the strength h_{ij} of the synaptic connection between two neurons N_i and N_j , i.e. between presynaptic and postsynaptic activities, is increased proportional to the correlation between the activity of N_i and N_j ; rele-

vant physiological data have been provided, e.g. by Bonhoeffer et al. (1989). We suggest a modification of the original Hebb rule by introducing a stabilising term preventing the synaptic weights from growing indefinitely. It will be shown that a neuron adapting to a stimulus according to this modified Hebb rule turns into a filter that is matched to that part of the stimulus pattern that covers the receptive field of the neuron. A similar result was already derived by Oja (1982), who assumed that unlimited growth of the synaptic weights is prevented by normalisation of the weights. Oja's approach requires additional and not very plausible, in any case very restrictive assumptions about the way synaptic weights may change and about the learning rate. No such assumptions are necessary in our model.

Since a stimulus will activate more than a single neuron, it follows that there will exist a set of cortical neurons whose receptive fields are partially or completely "covered" by inputs generated by the stimulus pattern. This set may form a type of cell assembly (CA) for the presented stimulus pattern; CAs and not single neurons are often considered as the basic unit of stimulus detection (e.g. Braitenberg 1978). We will illustrate our result (convergence of the receptive field of an adapting neuron towards one representing a matched filter for a stimulus or part of a stimulus) with respect to some psychophysical detection data employing an approximation to the activation of the CA.

That a given stimulus pattern may be detected by a filter matched to the stimulus and that such a filter may be the result of some adaptive process was first suggested by Hauske et al. (1976, 1978). Their approach has been criticised since the data from their experiments can be explained in terms of probability summation among, for example, spatial frequency channels (Graham 1977, 1989). Meinhardt and Mortensen (1998) presented some data supporting the hypothesis of detection by matched filters, and some data which cannot be explained by this model. We have used these data in order to illustrate (not test) our model. Detection data may be discussed in terms of probability summation among some suitably chosen channels or some sort of nonlinear pooling of the activity of such channels, for instance by adopting a p -norm model ("Quicks' rule"; Quick 1974). In particular, one may choose as channels neurons that have adapted to the stimulus or to parts of the stimulus according to our modified Hebb rule. It turns out that, for the considered stimuli, the case $p \rightarrow \infty$, i.e. p is large, can be assumed, which is equivalent to the case where the stimuli are detected by the maximally activated neuron.

A more explicit modelling of the interaction of neurons that have adapted to different parts of a stimulus is the focus of our future work.

2 Hebbian learning

2.1 Matched filters

The receptive field of a neuron may be represented by a weighting function \mathbf{h} of a spatial coordinate x ; the

implicit assumption of linearity serves as a first approximation. We restrict ourselves to the one-dimensional case; the generalisation to the two-dimensional case is straightforward, but complicates the notation without clarifying the notions involved.

The temporal characteristics of the neuron will be neglected; only the temporal course of the adaptation or learning process will be modelled.

The stimulus is defined by a luminance distribution $L(x) = L_0[1 + ms(x)]$, where x is the retinal coordinate, L_0 is the mean luminance, $m > 0$ denotes contrast (usually Maxwell contrast) and $s(x)$ is a function defining the (spatial) form of the stimulus; ms defines the modulation of L_0 , and in detection experiments it is this modulation that has to be detected. We thus concentrate on the response $v(x)$ of the visual system to the modulation generated by $ms(x)$. Assuming linearity, one has $v(x) = m \int_{-\infty}^{\infty} H(\omega)S(\omega)e^{-j\omega x}/2\pi d\omega$, where H is the Fourier transform of the spatial impulse response h , and S is the Fourier transform of s ; as usual, $\omega = 2\pi f$, f a spatial frequency, $j = \sqrt{-1}$. Given the stimulus pattern ms , one may ask for which weighting function h the response v will be maximal at some $x = x_0$. Assuming detection in white noise, it may be shown that the optimal filter is given by

$$H(\omega) = S^*(\omega) \exp(-j\omega x_0) , \quad (1)$$

i.e. the response is maximal at x_0 if H is defined by the complex conjugate S^* of s (Peterson et al. 1954; Whalen 1971)¹. If H satisfies (1), the filter represented by H is said to be matched to the input s . Equivalently, the condition for H expressed in (1) may be written with respect to the impulse response as

$$h(x) = s(x_0 - x) . \quad (2)$$

For later reference, we note the response $v_{\max} = v(x_0)$ of a matched filter at x_0 , the position of the maximal response. Since $H(\omega) = S^*(\omega) \exp(-j\omega x_0)$, one has

$$v_{\max} = v(x_0) = \frac{m}{2\pi} \int_{-\infty}^{\infty} |S(\omega)|^2 d\omega > 0 . \quad (3)$$

Without loss of generality, we will assume $s = \text{constant}$ in time and $\|\mathbf{s}\|^2 = 1$, and $x_0 = 0$; these conditions are introduced to simplify the notation. This remark applies to the assumption $s = \text{constant}$ as well; the effect of s varying over time will be left to future work.

The integral representation of the output v is a generalisation or idealisation in so far as it approximates a sum. One may, therefore, discretise \mathbf{h} into components $h_i = h(x_i)$, $s_{-i} = s(-x_i)$ and write \mathbf{h} and \mathbf{s} as a vectors² $\mathbf{h} = (h_{-n}, h_{-(n-1)}, \dots, h_{-1}, h_n)'$, $\mathbf{s} = (s_{-n}, \dots, s_n)'$, $-n \leq i \leq n$. Then, the response $v(x_0)$ to ms is given by

$$v = m \sum_{i=-n}^n h_{-i} s_i . \quad (4)$$

¹ Not by $mS^*(\omega) \exp(-j\omega x_0)$; application of Schwartz' inequality to the expression for $v(x)$ in order to arrive at the system function that yields a maximal response for the stimulus ms (e.g. Papoulis 1988) shows that m cancels out

² As usual, we denote by ' a transposed vector or matrix

In the following, we will consider the response v only at x_0 with $x_0 = 0$ and identify $v(x_0)$ with the neuron's response.

To simplify the following derivations, we introduce the following notation. h_i defines the weights characterising the influence of position i of the receptive field of a neuron to the position $i=0$. Consider the index transformation $i \rightarrow i + n + 1$; then, $1 \leq i \leq 2n + 1$. The new index is renamed to i and $2n + 1$ is labelled by n , so that for the new index $1 \leq i \leq n$. The index transformation implies that h_{n-i} has to be replaced by h_i .

2.2 Hebb's rule and convergence towards a matched filter

According to Hebb's rule, the change of a synaptic weight is proportional to the correlation of output and input. The correlation is proportional to the product $s_i h_i$. Since the change may be expressed as a derivative, one has, for an arbitrary spatial pattern ms ,

$$dh_i(t)/dt = \lambda v(t)ms_i, \quad \lambda > 0, \\ 1 \leq i \leq n, \quad (\text{Hebb's rule}) \quad (5)$$

where λ represents the learning rate. We have written $v(t)$ instead of v in order to express the dependency of v upon t as a result of the adaptation of \mathbf{h} ; $v(t) = \sum_i h_i(t)s_i$, and $h_i(t)$ indicates the change of the h_i with adaptation and not the temporal course of the spatial impulse response. The formalisation (5) implies, however, that the solution for the weights h_i is not bounded. To see this, consider the simple case $n = 1$. Then, $dh(t)/dt = \lambda v(t)ms = \lambda m^2 s^2 h(t)$, and for $ms > 0$, $h(t) = \lambda_0 \exp(\lambda m^2 s^2 t)$ is the general solution with starting value λ_0 . Obviously, \mathbf{h} diverges for all $\lambda_0 \neq 0$, $\lambda > 0$.

In order to avoid this instability, Hebb's rule as formalised in (5) has to be generalised such as to encompass mechanisms that force the synaptic weights to saturate. Nachtigall (1991) introduced a stabilising function $F(v)$ and added the term $-F(v)h_i(t)$ to (5), thus assuming some sort of feedback mechanism within the neuron achieving the required saturation. F is postulated to have the following properties.

Definition 1. Let F be a continuous, real-valued function having the following properties; there exists a $v^* > 0$ such that

$$F(v) < F(0), \quad v \in (0, v^*) \quad (6)$$

$$F(v) > F(0), \quad v \notin (0, v^*) \quad (7)$$

Then F will be called a stabilising function.

Note that

$$F(v^*) = F(0), \quad (8)$$

since from (6), $F(v) < F(0)$ for $v \in (0, v^*)$, and from (7), $F(0) < F(v)$ for $v \notin (0, v^*)$. The continuity of F implies $F(v) \uparrow F(0)$ for $v \uparrow v^*$, and $F(v) \downarrow F(0)$ for $v \downarrow v^*$, and thus $F(v^*) = F(0)$.

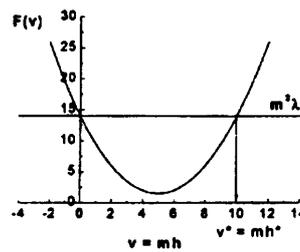


Fig. 1. An example of the stabilising function F introduced in definition 1; the definition of F implies that the differential equation (10), i.e. the stabilised Hebb rule, has only a single stable fixed point

A more explicit definition of F would require a more detailed characterisation of the adapting neuron; we will not go into such detail here, since the experiments to be discussed later in this paper do not allow a fine-grained consideration of neurophysiological mechanisms. An example of F is shown in Fig. 1.

Providing (5) with the additive term $-F(v)h_i(t)$ leads to

$$dh_i(t)/dt = \lambda v(t)ms_i(t) - F(v)h_i(t). \quad i = 1, \dots, n. \quad (9)$$

Equation (9) may be written as a vector equation; dropping the explicit reference to t one has, with $v = m \sum_i h_i s_i = m \langle \mathbf{s}, \mathbf{h} \rangle$,

$$d\mathbf{h}/dt = \lambda v m \mathbf{s} - F(v) \mathbf{h}. \quad (10)$$

This equation will be referred to as the stabilised Hebb rule³, in the following. We begin with a general statement about the equilibrium vectors \mathbf{h}_0 of the system (10), i.e. the vectors satisfying $d\mathbf{h}_0/dt = 0$.

Lemma. Let $\mathbf{C}_s = \mathbf{s}\mathbf{s}'$; \mathbf{C}_s is an $n \times n$ symmetric matrix: the element in the i th row and j th column of \mathbf{C}_s equals the product $s_i s_j$ of the i th and the j th component of the vector \mathbf{s} . Let \mathbf{w}_i be the eigenvectors of \mathbf{C}_s and let μ_i be the corresponding eigenvalues; suppose further that $\|\mathbf{s}\|^2 = 1$. The eigenvector \mathbf{w}_1 and the eigenvalues μ_i are known to satisfy the following equations:

$$\mathbf{w}_1 = \langle \mathbf{s}, \mathbf{w}_1 \rangle \mathbf{s} \quad (11)$$

$$\mu_1 = \langle \mathbf{s}, \mathbf{s} \rangle = 1, \quad \mu_i = 0 \quad \text{for } i \neq 1. \quad (12)$$

Proof. A proof of (12) and (11) may be found in Appendix A1.

Theorem 1. Suppose $\|\mathbf{s}\|^2 = 1$, let $\mathbf{h}_0 \neq 0$ be an equilibrium vector for the system (10), i.e. $d\mathbf{h}_0/dt = 0$, and let $v_0 = m \langle \mathbf{s}, \mathbf{h}_0 \rangle$. Then

$$\mathbf{h}_0 = \pm \langle \mathbf{s}, \mathbf{h}_0 \rangle \mathbf{s} \quad (13)$$

$$F(v_0) = m^2 \lambda. \quad (14)$$

³ This expression was suggested by an anonymous reviewer of this paper

Proof. From (10), one has $d\mathbf{h}/dt = \lambda m\mathbf{s}v - F(v)\mathbf{h} = m^2\lambda\mathbf{s}\mathbf{s}'\mathbf{h} - F(v)\mathbf{h} = m^2\lambda\mathbf{C}_s\mathbf{h} - F(v)\mathbf{h}$, i.e. with I the $n \times n$ unity matrix,

$$d\mathbf{h}/dt = (m^2\lambda\mathbf{C}_s - F(v)I)\mathbf{h} . \quad (15)$$

If $\mathbf{h} = \mathbf{h}_0 \neq 0$ is an equilibrium vector, it follows from (15) that

$$\mathbf{C}_s\mathbf{h}_0 = [F(v_0)/(m^2\lambda)]\mathbf{h}_0 ,$$

i.e. \mathbf{h}_0 is an eigenvector of \mathbf{C}_s and $\mu_1 = F(v_0)/m^2\lambda$ is the corresponding eigenvalue. Therefore, (11) implies (13), and (12) implies (14).

Note that (13) implies that, for $\mathbf{h}_0 = +\langle s, \mathbf{h}_0 \rangle s$, the equilibrium vector represents a matched filter for \mathbf{s} (the proportionality constant $\langle s, \mathbf{h}_0 \rangle$ will be discussed later) and has the Fourier transform $\langle s, \mathbf{h}_0 \rangle S^*$.

An equilibrium vector \mathbf{h}_0 may be either stable or not; if \mathbf{h}_0 is not stable, the slightest disturbance of the system such that $\mathbf{h} = \mathbf{h}_0 + \epsilon$, where ϵ is a vector representing the disturbance, causes the system to move away from \mathbf{h}_0 . A stable equilibrium in the following means, in particular, an asymptotically stable equilibrium vector, i.e. a vector \mathbf{h}_0 satisfying the condition that every vector \mathbf{h} within a certain neighbourhood N of \mathbf{h}_0 approaches \mathbf{h}_0 as $t \rightarrow \infty$. A concise discussion of different types of stability of a system may be found, e.g. in Arrowsmith and Place (1982).

If \mathbf{h}_0 is an equilibrium vector, then $-\mathbf{h}_0$ is an equilibrium vector as well. Let us assume that $-\mathbf{h}_0$ is a stable equilibrium vector. Then, the neuron's system function will correspond to $-\langle s, \mathbf{h}_0 \rangle S^*$, yielding the response $-v_{\max}$. This will mean that, for a "bright" stimulus, a filter could develop that signals the stimulus as a "dark" one, and vice versa, which does not seem to correspond to empirical findings. It will now be shown that the conditions on F as specified in definition 1 are sufficient for the existence of a single, stable equilibrium vector of (10).

Theorem 2. Let the stimulus pattern be given by $m\mathbf{s}$ with $m > 0$ and $\|\mathbf{s}\| = 1$. Suppose (10) holds with F satisfying the conditions given in (1). Then, there exists a real number $r > 0$ such that $\mathbf{h} \rightarrow \mathbf{h}^* = r\mathbf{s}$, and \mathbf{h}^* is the single stable equilibrium of the system characterised by (10).

Proof. Consider the special case, $\mathbf{s} = (1, 0, \dots, 0)'$; since now

$$v = m\langle \mathbf{s}, \mathbf{h} \rangle = mh_1 , \quad (16)$$

(10) leads to the system of equations

$$\frac{d}{dt}h_i = \begin{cases} \lambda m^2 h_1 - F(mh_1)h_1, & i = 1 \\ -F(mh_1)h_i, & i > 1 \end{cases} . \quad (17)$$

In this case, the equation for $i = 1$ is decoupled from the remaining ones. The implication is that, if a stationary solution of (17) for $i = 1$ exists, then the solutions for $i > 1$ approach zero exponentially.

The proof is based on the application of a coordinate transformation; the variables (i.e. the vectors) \mathbf{s} and \mathbf{h} in (10) will be transformed, and this transformation implies a transformation of the equation. The transformation will be chosen such that a system of equations corresponding to (17) results. Then, the solution to the transformed equation is found, which is easy since there is now just a single variable; the inverse transformation then yields a solution to the original equation (10). A justification of this procedure may be found, for instance, in Arnold (1980), and Arrowsmith and Place (1990) (chapters on diffeomorphisms).

Since the required transformation turns out to be linear, let us briefly recall some basic results concerning these transformations (for example, Rao 1975 or Fischer 1997). There exists an orthogonal matrix T such that $|T| = 1$, $T'T = T^{-1}T = I$, I the identity matrix, and $\tilde{\mathbf{s}} = T\mathbf{s} = (1, 0, \dots, 0)'$; one has to supplement the vector \mathbf{s} with some orthonormal vectors $\mathbf{w}_2, \dots, \mathbf{w}_n$ to form an orthonormal basis of \mathbb{R}^n ; then $T = (\mathbf{s}, \mathbf{w}_2, \dots, \mathbf{w}_n)$ is an orthonormal matrix and $T\mathbf{s} = (1, 0, \dots, 0)'$, provided $\|\mathbf{s}\| = 1$; of course, $\mathbf{s} = T^{-1}(1, 0, \dots, 0)'$. The scalar product $\langle \mathbf{s}, \mathbf{h} \rangle$ is invariant under the transformation T ; let $\tilde{\mathbf{h}} = T\mathbf{h}$, then $\langle \mathbf{s}, \mathbf{h} \rangle = \mathbf{s}'\mathbf{h} = \langle \tilde{\mathbf{s}}, \tilde{\mathbf{h}} \rangle$, since $\langle T\mathbf{s}, T\mathbf{h} \rangle = (T\mathbf{s})'(T\mathbf{h}) = \mathbf{s}'T'T\mathbf{h} = \mathbf{s}'\mathbf{h}$. Note that because $\tilde{\mathbf{s}} = T\mathbf{s} = (1, 0, \dots, 0)'$, one has, in particular, $\langle \tilde{\mathbf{s}}, \tilde{\mathbf{h}} \rangle = \tilde{h}_1$, and v is again given by (16).

Now, \mathbf{h} is the solution of (10) defining the stabilised Hebb rule $d\mathbf{h}/dt = \lambda v m \mathbf{s} - F(v)\mathbf{h}$. Applying the transformation T to this equation means applying T to \mathbf{h} and \mathbf{s} ; with $\tilde{\mathbf{h}} = T\mathbf{h}$ and $\tilde{\mathbf{s}} = T\mathbf{s}$ one has

$$\begin{aligned} \frac{d}{dt}T\mathbf{h} &= \frac{d}{dt}\tilde{\mathbf{h}} = \lambda \langle mT\mathbf{s}, T\mathbf{h} \rangle mT\mathbf{s} - F(\langle mT\mathbf{s}, T\mathbf{h} \rangle)T\mathbf{h} \\ &= \lambda m^2 \langle \tilde{\mathbf{s}}, \tilde{\mathbf{h}} \rangle \tilde{\mathbf{s}} - F(m\langle \tilde{\mathbf{s}}, \tilde{\mathbf{h}} \rangle) \tilde{\mathbf{h}} \\ &= \lambda v m \tilde{\mathbf{s}} - F(v)\tilde{\mathbf{h}} , \end{aligned} \quad (18)$$

since because of the invariance of the scalar product $\langle \mathbf{s}, \mathbf{h} \rangle$ under the transformation T , $m\langle \tilde{\mathbf{s}}, \tilde{\mathbf{h}} \rangle = v$.

Equation (18) is obviously of the same general form as the stabilised Hebb rule (10). Applying the inverse transformation T^{-1} to $\tilde{\mathbf{h}}$ yields again \mathbf{h} , i.e. $T^{-1}\tilde{\mathbf{h}} = \mathbf{h}$, and applying T^{-1} to (18) yields again (10) in the same way the application of T to (10) lead to (18), i.e. $\mathbf{h} = T^{-1}\tilde{\mathbf{h}}$ is a solution of (10).

Equation (18) is a vector equation; the individual components of the vector $\tilde{\mathbf{h}}$ obey the equations

$$\frac{d}{dt}\tilde{h}_i(t) = \begin{cases} m^2\lambda\tilde{h}_1(t) - F(v)\tilde{h}_1(t), & i = 1 \\ -F(v)\tilde{h}_i(t), & 2 \leq i \leq n \end{cases} \quad (19)$$

corresponding to (17), with v as defined in (16).

Consider first the equation for $i = 1$:

$$d\tilde{h}_1/dt = [m^2\lambda - F(v)]\tilde{h}_1 . \quad (20)$$

$\tilde{h}_1 = 0$ is an equilibrium point; it will be shown that it is unstable. Let $r > 0$ be such that, with $v^* = mr$, $F(v^*) = \lambda m^2$; it will be shown that r is a stable fixed point of (20).

Suppose that $\tilde{h}_1 < r$. If, in particular, $\tilde{h}_1 < 0$, $v = m\tilde{h}_1 < 0$ and definition 1, (7) implies $m^2\lambda - F(v) \leq 0$, so that $d\tilde{h}_1/dt > 0$ and \tilde{h}_1 increases. Now let $0 \leq \tilde{h}_1 < r$. Then $m^2\lambda - F(v) > 0$, and again \tilde{h}_1 increases; this implies that $\tilde{h}_1 = 0$ is an unstable fixed point. Let now $\tilde{h}_1 > r$. Definition 1 and (7) imply that the right-hand side of (20) is negative, and \tilde{h}_1 decreases. Since F is continuous, it follows that $v = m\tilde{h}_1\tilde{s}_1 \rightarrow v^* = mr\tilde{s}_1$, i.e. $\tilde{h}_1 \rightarrow r$ for $t \rightarrow \infty$.

For $i > 1$, if $F(v) \rightarrow m^2\lambda$ for $v \rightarrow v^*$, (19) can be approximated by

$$\frac{d}{dt}\tilde{h}_i(t) = -m^2\lambda\tilde{h}_i(t), \quad 2 \leq i \leq n, \quad (21)$$

the solutions of which satisfy

$$\tilde{h}_i(t) = \tilde{h}_i(0) \exp(-m^2\lambda t) \rightarrow 0. \quad (22)$$

Then

$$\tilde{\mathbf{h}} \rightarrow \tilde{\mathbf{h}}^* = \begin{pmatrix} r \\ 0 \\ \vdots \\ 0 \end{pmatrix} = r \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (23)$$

With respect to the original coordinates or vector \mathbf{h} , it follows that $\mathbf{h} \rightarrow \mathbf{h}^* = T^{-1}\tilde{\mathbf{h}}^* = rT^{-1}(1, 0, \dots, 0)' = r\mathbf{s}$, i.e. the equilibrium vector \mathbf{h}^* is proportional to \mathbf{s} . Substituting r , $v^* = mr$ for v into (20) shows that r is a solution, so $T^{-1}\tilde{\mathbf{h}}_1 = \mathbf{h}^*$ is a solution of (10); the solution is stable.

From (15), $\mathbf{h} = \mathbf{0}$ is an equilibrium vector as well. Although this equilibrium is unstable [recall that $\tilde{h}_1 = 0$ is an unstable fixed point for (20)], the system (10) may get stuck here unless some disturbance acts upon \mathbf{h} . There will always be some random fluctuations within the system (see, for example, Gardiner 1990) and eventual convergence towards v^* will occur; otherwise the response of the neuron would fluctuate around zero and the stimulus pattern would not be detected. A detailed discussion of the matter would require modelling in terms of stochastic differential equations, which will, however, not be pursued here; suffice it to say that underlying our approach is the assumption that the noise is "small" and running on a time scale much smaller than the temporal processes considered here. Psychophysical support for this assumption is provided in Mortensen and Suhl (1991).

To summarise, it was shown that Hebb's rule, as defined in (10) together with the definition of F in definition 1, implies convergence of \mathbf{h} towards the single stable equilibrium vector $\mathbf{h}^* = r\mathbf{s}$. From (13), we conclude $r = \langle \mathbf{s}, \mathbf{h}_0 \rangle$, and of course $\mathbf{h}^* = \mathbf{h}_0$. Since we can always arrange for $\|\mathbf{s}\|^2 = 1$, the value of $r = \langle \mathbf{s}, \mathbf{h}_0 \rangle = \langle \mathbf{s}, \mathbf{h}^* \rangle$ depends only upon the unit of measurement with respect to which the components of \mathbf{h}^* are evaluated. We may thus say that \mathbf{h} converges towards a matched filter, at least for the range of \mathbf{s} covering the receptive field of the neuron.

In relation to other work, the assumption that a neuron is equipped with adjustable synapses implies that the neuron is conceived as an adaptive filter of the

closed-loop type; a general discussion of such filters may be found, e.g. in Haykin (1984).

A complete account of work on the application of Hebb's rule is beyond the scope of this paper. Suffice it to say that models may be found in Amari (1977), or Kohonen (1989), where references for earlier models based on Hebb's rule may also be found. Bienenstock, et al. (1982) in particular provided a model for stimulus selectivity in the primary cortex based upon Hebb's rule. Herzog and Fahle (1998) proposed a model of perceptual learning in particular for hyperacuity tasks. Gerster et al. (1993) discussed Hebbian learning with respect to detailed reference to the spiking of neurons; this spike response model (SRM) was implemented by Ritz et al. (1994) in a model of feature linking and pattern separation; we will further comment upon this work in Sect. 5.

Oja (1982) assumed \mathbf{s} and \mathbf{h} to be contaminated with noise. Secondly, he started with a time discrete version of Hebb's rule, namely (again in our notation) $h_i(t+1) = h_i(t) + \lambda v s_i$. Saturation of h_i is achieved by normalisation, i.e. Oja generalised Hebb's rule to $h_i(t+1) = [h_i(t) + \lambda v s_i] / (\|h_i(t) + \lambda v s_i\|)$. Assuming that $\lambda \rightarrow 0$ as $1/t$, he arrived at the differential equation

$$\frac{d}{dt}\mathbf{h}(t) = C\mathbf{h}(t) - (\mathbf{h}'(t)C\mathbf{h}(t))\mathbf{h}(t), \quad (24)$$

with $C = E(\mathbf{s}\mathbf{s}')$, the matrix of expected values of $s_i s_j$ implying convergence $\mathbf{h}(t) \rightarrow \mathbf{w}$ ($-\mathbf{w}$), and \mathbf{w} ($-\mathbf{w}$) are asymptotically stable equilibrium points of (24). Note that the structure of this equation is different from that of our (10). Obviously, our approach is more direct and does not require restrictive assumptions about λ .

Sanger (1989, 1990) generalised Oja's (1982) model to the case of multiple inputs.

3 Some models of detection

The main aim of this paper is the presentation of the results of the last section. In order to illustrate the results of the last section, we (1) present a simple model of detection containing the original matched filter model of Hauske et al. (1976) as a special case, and (2) fit the model to some experimental data that cannot be explained by the model of Hauske et al. (1976). The discussion of the model with respect to the data is not meant to be a general test of the model's validity but serves illustrative purposes; a general test would require more data, in particular from experiments with two-dimensional stimuli, along with tests of alternative models. Such tests are, however, beyond the scope of this paper.

3.1 The classical matched-filter model

The detecting neurons will respond to a filtered version of the stimulus pattern; even if the pattern is detected by a receptor, the pattern processed by that neuron will be one that has been filtered by the lens of the eye. Most likely, however, the site of the detecting neuron will be

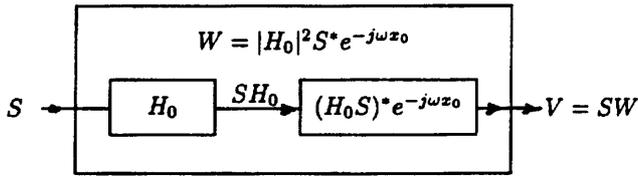


Fig. 2. The structure of a matched channel as introduced in Meinhardt et al. (1998). The matched channel is the cascade of a pre-filter with system function H_0 and a matched filter with system function $(H_0 S)^* \exp(-j\omega x_0)$

cortical. It follows that these neurons respond to a filtered version of the stimulus pattern. Let S be the Fourier transform of the stimulus pattern, and let H_0 be the system function of that part of the visual system preceding the “detector”, i.e. the neuron or system of neurons actually generating the detection response; this part will be called the pre-filter. The input to the detector has Fourier transform $G(\omega) = S(\omega)H_0(\omega)$, $\omega = 2\pi f$, f a spatial frequency. A filter matched to the response g of this filter, with maximal output at x_0 , has the system function $G^*(\omega) \exp(-j\omega x_0) = H_0(\omega)S(\omega) \exp(-j\omega x_0)$. Thus, a stimulus is first filtered by H_0 , then by $G^*(\omega) \exp(-j\omega x_0)$ (Fig. 2).

The overall system function is then given by

$$W_s(\omega) = |H_0(\omega)|^2 S^*(\omega) e^{-j\omega x_0}, \quad j = \sqrt{-1}. \quad (25)$$

It is essentially this model that was proposed by Hauske et al. (1976).

The system defined by (25) was called a matched channel in Meinhardt et al. (1998) because W_s does not only represent the matched filter but the cascade consisting of the pre-filter and the matched filter. We write W_s in order to indicate that the channel is matched to the pattern s . In the following, we will speak of the MC model. It follows from (25) that the system function of the detecting channel is proportional to the complex conjugate spectrum of the stimulus pattern. Hauske et al. (1976) explicitly assumed $x_0 = 0$; however, in the test of the model performed by Meinhardt et al. (1998) no specific assumption about the value of x_0 had to be made.

3.2 A cell-assembly model

In this section, we present a model of detection based on the assumption that there exists a set of neurons that have already adapted to a stimulus pattern in the sense of the theorem presented in the preceding section.

In a detection experiment, the stimuli are presented as brief temporal – usually rectangular – pulses; in our experiments, the stimulus duration is 500 ms. If the value of λ is sufficiently large, learning may take place within a single trial. Data from EEG experiments suggest that learning may require not more than 90–100 ms (Skrandies, personal communication). However, so far we are not able to arrive at an independent estimation of λ .

3.2.1 Assumptions

Let $N(x_i)$ denote a neuron positioned at the retinal coordinate x_i , $1 \leq i \leq k$, k the number of adapting neurons involved in the detection process; x_i may be thought of as corresponding to the centre of the neuron’s receptive field. The neuron is characterised by some spatial impulse function h_{x_i} .

1. The pre-filter: The pre-filter H_0 has a system function defined by the DOG

$$H_0(\omega) = A_1 \exp\left[-\frac{(\omega\sigma_1)^2}{2}\right] - A_2 \exp\left[-\frac{(\omega\sigma_2)^2}{2}\right], \quad (26)$$

where $\omega = 2\pi f$; f is a spatial frequency and A_i and σ_i , $i = 1, 2$ are free parameters to be estimated from the data.

2. Cell assembly:

- A. Let $M_s = \{N(x_i) | x_i \in I_s \subset \mathbb{R}\}$ be the set of neurons that have adapted to the response g of the pre-filter as generated by the stimulus s . The receptive field of a neuron $N(x_i) \in M_s$ at position x_i can be represented by a weighting function h_{x_i} which may assume non-zero values only upon some interval (a_{x_i}, b_{x_i}) , $x \in (a_{x_i}, b_{x_i})$; $h_{x_i}(\xi) \equiv 0$ for $\xi \notin (a_{x_i}, b_{x_i})$. The size of the receptive field is given by $\rho = b_{x_i} - a_{x_i}$, i.e. ρ is a constant for all $N(x_i) \in M_s$, and $a_{x_i} = x_i - \rho/2$, $b_{x_i} = x_i + \rho/2$.
- B. Suppose the neuron $N(x_i)$ has adapted to g as defined on (a_{x_i}, b_{x_i}) in the sense of Theorem 2. Then, $h(x_i - \xi) \propto g(\xi)$ for $\xi \in (a_{x_i}, b_{x_i})$.
3. Detection: Let $v_i = v(x_i) \geq 0$ be the response of the neuron $N(x_i) \in M_s$. The stimulus pattern is detected if

$$\sum_{i=1}^k v_i^p = d, \quad p > 0, \quad (27)$$

and $d = \text{constant}$.

The set M_s will be referred to as a cell assembly, and the model specified by assumptions 1 and 2 will be called the cell-assembly model, CA model.

3.2.2 Comments

$\rho = \text{constant}$: The assumption that the size ρ of the receptive fields of the neurons belonging to an assembly is constant may appear as being very restrictive; it serves as a first (but working!) approximation to the certainly more complex true state of affairs. Note that the receptive field of a neuron collects inputs only from the pre-filter, not from other neurons belonging to the set M_s ; this will be further commented upon in Sect. 5.

DOG postulate: The DOG postulate may be considered standard (e.g. Kulikowski and King-Smith 1973; Hines 1976; Wilson 1978; Wilson and Bergen 1979; see Wilson et al. 1990 for an overview of models on early visual processing). The inverse (Fourier) transform of (26) yields the line spread function (LSF), representing the x -profile of a point spread function (PSF) (assuming

radial symmetry of the PSF). The estimated parameters of the LSF depend, via the adopted model of the early processing stages, upon the method of determining the LSF or MTF (since the model implies the method); one may assume a single-channel model such as that of Kulikowski et al. (1973) or Hines (1976), or one may adopt a multiple-channel model plus probability summation among channels (e.g. Sachs et al. 1972; Graham 1977; Wilson et al. 1979)⁴.

A detailed discussion of these matters with respect to assumption 1 within the framework of the CA model is, however, beyond the scope of this paper and not necessary since our aim is to illustrate the hypothesis of detection by neurons that have adapted to the stimulus or to parts of the stimulus; for this purpose, only global properties of the pre-filter have to be assumed; in particular, the parameters of the pre-filter should be identical (up to estimation errors), since only cortical cells are meant to adapt, not the pre-filter.

The learning parameter λ : The only parameter characterising the adoption of the neurons to the stimulus, i.e. of the cell assembly, is the learning rate λ , and this parameter will not be estimated from the data; it is assumed that the neurons have already adapted. This is equivalent to postulating that adaptation is fast, i.e. that λ is large and that the precise value of λ is not required; we return to this assumption in Sect. 5. So far, estimating the parameters A_i and σ_i , $i = 1, 2$ of the pre-filter and adopting the assumptions 1–3 [in particular, (28)] represents indeed a fitting of the CA model because detection is assumed to depend upon the response of the cells with impulse response $h(x - \xi) \propto g(\xi)$, $g(\xi)$ the response of the pre-filter at ξ . This also means that the properties of the detecting cells are – except for the receptive field size ρ – completely determined by the particular stimulus employed in the experiment and by the parameters of the pre-filter.

Detection process: Assumption 3 has become standard in visual psychophysics and may come across as a simplification; one may argue that a more realistic description of the detection process could be given in terms of von der Malsburg's (1983) correlation hypothesis or by the models of Aertsen and Gerstein (1991), and Aertsen et al. (1993). After all, Braitenberg (1978) discussed a feedback mechanism leading to the “ignition” of M_s , so that the complete assembly M_s would represent the stimulus pattern. However, apart from some unresolved questions concerning the conceptualisation of the interneural connections (see Sect. 5), any model taking interneural connections into account will

most likely imply the need for estimating further free parameters. Therefore, (27) was adopted as some summarising description of the neuronal interactions. Note that if one wants to interpret (27) as reflecting the effect of probability summation (Quick 1974), the neurons cannot be considered as being pairwise stochastically independent, even when no interactions among them are assumed, since many of them will adapt to overlapping parts of the stimulus, although the theory of asymptotically independent random variables (see Leadbetter et al. 1983) may lead to the same limiting distribution defining the psychometric function, i.e. $P(m) = 1 - \exp(-\sum_i v_i^p)$, that is assumed for independent neurons.

The expression (27) allows for a special case. It is well known that for “large” values of p (formally: $p \rightarrow \infty$, i.e. $p > 4$ or $p > 6$; values of p in this range were found by Meinhardt 1999) one has $\sum_i v_i^p \approx v_{\max}^p$, where $v_{\max} = \max(v_1, \dots, v_k)$. This means that the stimulus is detected if

$$v_{\max} \approx c, \quad c = d^{1/p}, \quad (28)$$

i.e. the stimulus is detected predominantly by the maximally activated neuron. With respect to the difficulties encountered when trying to estimate the parameters of the pre-filter (see Sect. 4.3 and Sect. 4.4); one may attempt to assume the validity of (28), since then the free parameter p does not have to be explicitly determined. Indeed, adopting (28) turns out to be compatible with the data.

MC versus CA model: Finally, let us point out the difference between the CA and the MF, i.e. the original matched filter model of Hauske et al. (1976).

1. MC model. In the model of Hauske et al. (1976), the existence of a single channel matched to the complete stimulus pattern is postulated.
2. CA model. Suppose there exists a neuron adapting to the presented stimulus pattern (or a pre-filtered version of it) in a way described by the stabilised Hebb rule (10), having a certain, fixed receptive field size, and suppose further that the activation of the neuron by the stimulus pattern is a necessary condition for the detection of the stimulus. The size ρ of the receptive field of the adapting neuron may be smaller than the spatial extend of the stimulus pattern. Recall that, according to assumption 2a, the receptive field of a neuron N_x , i.e. a neuron at position x , can assume nonzero values only within the interval (a_{x_i}, b_{x_i}) . Clearly, the neuron can only adapt to that part of the response $g(x)$ of the pre-filter to the stimulus for which $x \in (a_{x_i}, b_{x_i})$; let us call this stimulus part the stimulus truncated with respect to (a_{x_i}, b_{x_i}) . The neuron may be considered as a system (“channel”) characterised by a system function which is, according to Theorem 2, proportional to the complex conjugate of the Fourier transform of the stimulus truncated with respect to (a_{x_i}, b_{x_i}) . Since the Fourier transforms of the stimulus truncated with respect to some interval (a_{x_i}, b_{x_i}) will differ from that of the complete stimulus, one is led to predictions of stimulus thresholds that differ from those of the

⁴ Suppose, for instance, that the visual system contains neurons that adapt to the lines employed in an experiment performed to estimate the LSF of the system, and that the pre-filter is characterised by a DOG function with parameters A_i and σ_i . Suppose further that the LSF is determined as in Hines (1976), who adopted a single-channel model without adapting neurons. Then, the empirically determined LSF turns out to be wider than the “true” LSF actually specifying the pre-filter

MC model. The predictions of the CA model will be identical to those of the MC model if the size of the receptive field is large enough to cover the complete stimulus.

Neighbouring neurons will have overlapping receptive fields, and consequently the parts of the stimulus to which they adapt also overlap.

The difference between the MC model and the CA model may, of course, be more pronounced when stimulus detection depends upon the activity of all neurons in the assembly, i.e. if (27) with p “small” holds.

4 Example

While the model of detection presented in the previous section is a first and very simple illustration of the way in which the results of Sect. 2 may be employed to formulate a model of detection, the purpose of this section is to illustrate this model by showing that it is possible to “explain”, using this model, detection data that are not compatible with the model of Hauske et al. (1976).

4.1 Experiment

We consider data from Meinhardt et al. (1998), where the matched-filter model introduced by Hauske et al. (1976) was discussed. The data were collected in superposition experiments, the stimulus pattern is presented superimposed upon a sinusoidal grating being in either sine or cosine phase with respect to $x = 0$. The contrast m of the background grating is chosen to be subthreshold, and for a given value of m , the corresponding value $\phi(m)$ of the contrast of the stimulus pattern is determined. ϕ as a function of m is known as the contrast interrelationship function (CIF).

Let $m = 0$ and $m_{0s} = \phi(0)$; in this case, the stimulus pattern is presented without a background pattern, and m_{0s} is the threshold contrast of the stimulus pattern. The function $\phi(m)/m_{0s}$ is called the normalised CIF. In general, CIFs and, consequently, the normalised CIFs, turn out to be nonlinear (Logvinenko 1993; Meinhardt 1999); however, for sufficiently small values of m , the CIF ϕ and, consequently, the normalised CIF ϕ/m_{0s} may be approximated by a linear function of m . Let w_b be the unit response (i.e. the response to a signal with contrast equal to 1) of the detecting channel (neuron) to the background grating s_b ; then, w_b is proportional to either the real or the imaginary part of the system function of the detecting channel (see Hauske et al. 1976 or Meinhardt et al. 1998 for a derivation of this result). The slope of the normalised CIF is equal to w_b/d [see (A2) in Appendix 2]. Meinhardt et al. (1998) defined

$$\Phi(\omega) \stackrel{\text{def}}{=} \frac{w_b(\omega)}{d}, \quad (29)$$

as the sensitivity of the channel to a grating $\sin(\omega x)$ or $\cos(\omega x)$; $\omega = 2\pi f$, f the spatial frequency of the background grating. Given that detection is by a single

channel and that assumption (28) holds, the sensitivity Φ as a function of ω (or f) is proportional to the system function of the detecting channel and therefore allows for a direct test of the hypotheses concerning the detecting channel, e.g. of the hypothesis that this channel is matched to the response of the pre-filter to the pattern s . The prediction of the sensitivity $\Phi = w_b/d$ by the model is computed according to (A4) in Appendix 2. On the other hand, Φ can be estimated from the data by the quotient $[\hat{m}_{0s} - \hat{\phi}(m; \omega)]/(\hat{m}_{0s}m)$ [see (A3) in Appendix 2], where \hat{m}_{0s} is an estimate of the threshold contrast m_{0s} of the stimulus pattern when presented without a background grating; a discussion of the above interpretation of Φ and of further details concerning its estimation may be found in Meinhardt et al. (1998).

4.2 Stimulus patterns

We consider data for two patterns: $s_{\text{even}}(x) = 1$ for $x \in [-.5, .5]$ and $s_{\text{even}}(x) = 0$ for $x \notin [-.5, .5]$, and $s_{\text{odd}} = -1$ for $x \in [-.5, 0]$, $s_{\text{odd}} = +1$ for $x \in (0, .5]$ and $s_{\text{odd}} \equiv 0$ for $x \notin [-.5, .5]$, (see inset in Fig. 3). We will refer to s_{even} also as “bar-even” and to s_{odd} as “bar-odd”. Note that the spectrum of s_{even} is real while that of s_{odd} is imaginary. In order to test the matched filter as well as the matched CA hypothesis, it is sufficient to determine the sensitivities with respect to either a subthreshold cosine or a sine grating.

4.3 Parameter estimation

We assume that the value of p in (27) is “large” so that (28) holds. Then, there are altogether five free parameters to be estimated from the data: A_i and σ_i , $i = 1, 2$, and ρ . In order to arrive at the predictions for the CA model, the activity $v(x)$ of a neuron from M_s at x has to be calculated. Let g be the response of the pre-filter. From A3, one has

$$v(x) = \int_{x-\rho/2}^{x+\rho/2} h(x-\xi)g(\xi)d\xi = \int_{x-\rho/2}^{x+\rho/2} g^2(\xi)d\xi, \quad (30)$$

and g is computed according to assumption 2. Let Q be the sum of the squares of the differences $\Phi(f_i) - \hat{\Phi}(f_i)$, where $\hat{\Phi}(f_i)$ is the empirically estimated sensitivity for the spatial frequency f_i of the background grating (A3) and $\Phi(f_i)$ is the corresponding sensitivity as predicted by the model (A4), i.e. let

$$Q = \sum_i [\Phi(f_i) - \hat{\Phi}(f_i)]^2. \quad (31)$$

The parameters have been estimated such as to minimise Q as a function of the parameters A , B , s_1 and s_2 . The attempt to use the Levenberg-Marquardt (LM) algorithm (Press et al. 1996) as implemented in the Mathematica (version 2.24) statistics add-on package failed, however. It turned out that Q is a rapidly fluctuating function of the parameters; the local minima of Q at which the algorithm stops depend heavily upon the

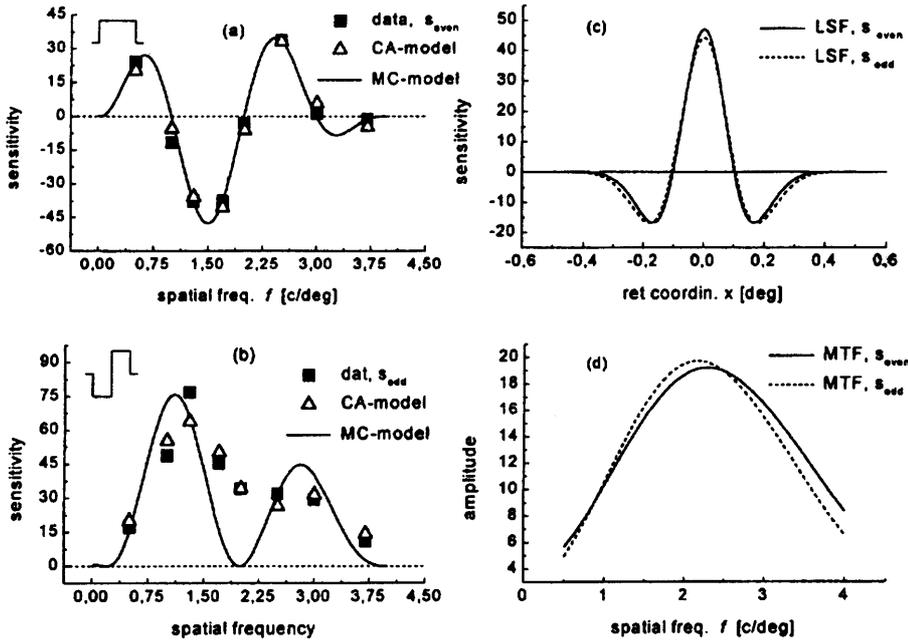


Fig. 3. Fit of the model of Hauske et al. (1976) and of the cell assembly (CA -) model **a** for s_{even} , **b** for s_{odd} , **c** the line spread functions (LSFs), estimated independently for s_{even} and s_{odd} , **d** corresponding MTFs. The data have been taken from Meinhardt et al. (1998)

starting values; minute changes of parameter values lead to rather drastic changes in the value of Q . Therefore, the estimation was carried out “by hand”; parameter values were varied depending upon an increase or decrease in the value of Q and separately for each pattern.

4.4 Results

The parameter estimates are summarised in Table 1; the great number of digits after the decimal point reflects the sensitivity of Q upon very small changes of the values.

The rather extreme sensitivity of the model with respect to minute differences of parameter values appears to be implied by the DOG postulate for the pre-filter with independent parameters A_i , σ_i , $i = 1, 2$; Meinhardt (1996) estimated the parameters of DOG functions when fitting a model of spatial probability summation to data from superposition experiments and reports the same difficulties that we encountered. One may therefore ask whether the estimates are meaningful. At present at least we obtain consistent estimates in the sense that basically the same pre-filter was found for different stimulus patterns. In future experiments, it will be attempted to estimate the LSF (or even better, the PSF) of the

pre-filter; not indirectly as was the case with the data discussed here, but directly in a way consistent with the CA model, i.e. assuming the existence of neurons that have adapted to thin lines according to the CA model (see the footnote concerning the estimation of LSFs in Sect. 3.2). Fitting the CA model to data from detection experiments with stimuli such as s_{even} , s_{odd} or two-dimensional stimuli employing such an estimate of the pre-filter would certainly represent more of a test of the model; as mentioned before, we only intend to give a simple illustration of the model here.

The empirical sensitivities for s_{even} and s_{odd} are presented in Fig. 3, together with the predictions for the MC-filter model and the CA model. Also, the LSFs and their Fourier transforms (the system functions) as estimated for each pattern are presented in Fig. 3.

The differences between corresponding parameter values for the pre-filter (i.e. the DOG functions) are sufficiently small to justify the assumption that the pre-filter is identical for the two patterns.

For the even-bar pattern, the MC and the CA model fit almost equally well. Note, however, that, in particular for the $f = 1$ c/deg background grating, the empirical sensitivity differs considerably from zero, while the MC model predicts a zero sensitivity. The CA model, on the other hand, can cope with this observation (there may exist a combination of parameter values for which the data are predicted even more perfectly!).

For the odd-bar pattern, the MC model is clearly inadequate, while the CA model agrees quite well with the data. The MC model predicts zero sensitivities, e.g. for the $f = 2$ c/deg background grating, and small sensitivities in the neighbourhood of this spatial frequency, while the data indicate positive sensitivities for these values of f .

Figure 4 shows the responses of the pre-filter to the two patterns, and the activation of the cells of the set M_s corresponding to each pattern. The curves for the CA

Table 1. Estimates of the pre-filter parameters A_i and σ_i , $i = 1, 2$, and ρ , the size of the receptive fields of the adapting neurons. x_{max} is the position of the neuron(s) with maximum response

Pre-filter parameter	s_{even}	s_{odd}
A_1	1500.5555	1500.3785
A_2	1496.76	1497.7255
σ_1	0.093614758	0.100283425
σ_2	0.095135322	0.101999505
ρ	1.176 deg	0.84665
x_{max}	0.102422	0.0

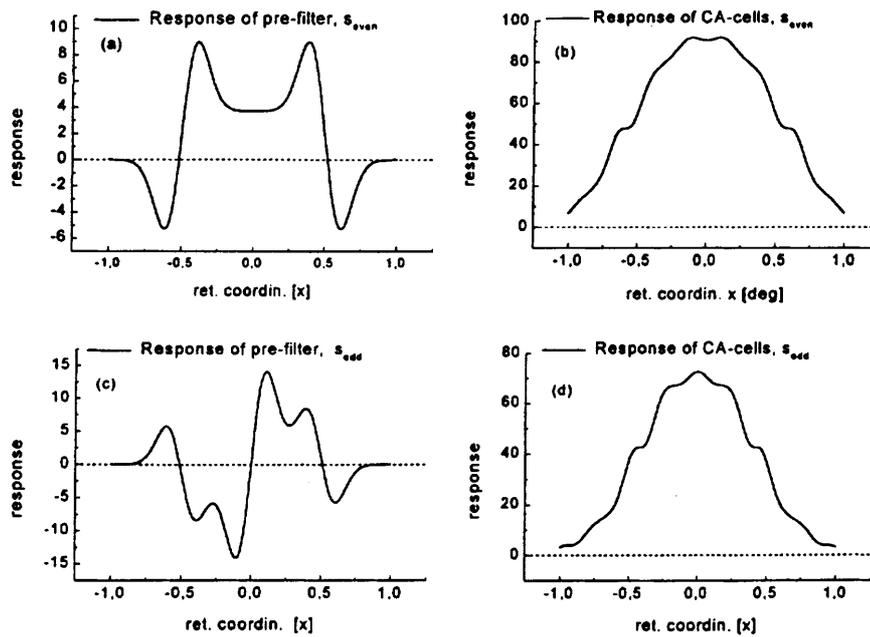


Fig. 4. Responses of the pre-filter (a to s_{even} , c to s_{odd}) and of the adapted neurons (b to s_{even} , d to s_{odd})

cells are, of course, symmetrical, since they depend upon the squared response of the pre-filter. For the even bar, there exist two maxima of equal size; it was assumed that the subject's response depends on the value of just one of the maxima. The position of the maxima differs from zero by 0.102422 (see Table 1), so for the even-bar pattern, the subject will not focus on the position $x = 0$. For the odd bar, there is just a single maximum at $x = 0$, so it is the odd pattern, but not the even pattern for which the assumption of Hauske et al. (1976) of detection at $x_0 = 0$ holds.

It remains to consider the estimates of the size ρ of the receptive field of the detecting neuron. For the even pattern, ρ is about 1.4 times as large as that for the odd pattern, although the size of the two patterns was identical; so far, the model itself does not allow an interpretation of the different values of ρ . Further experiments appear to be necessary to guide speculations about the relation between pattern and receptive field size.

5 Discussion

One may distinguish between two broad classes of models of early vision: one for models assuming processing by fixed channels, another for models assuming the operation of adaptive channels. Examples for the first class have been given in Sect. 1; a possible approach to the question how their general structure may be characterised can be found in Logvinenko (1996).

The CA model belongs to the second class. The finding that this model could be fitted to some data assuming detection by the maximally activated neuron may, however, not apply to other data from experiments thoroughly designed as a test of the model, employing also two-dimensional stimulus patterns; the special case “ p large”, equivalent to the assumption of

processing by a single neuron, may not apply in general. For instance, the correlational mechanisms postulated by Caelli and Moraglia (1985) for discrimination data, by Caelli and Oğuztörel (1987) for masking data and by Caelli et al. (1987) for pattern classification data are, according to these authors, indicative for the operation of adaptive and not of fixed mechanisms, and with respect to the size of the stimuli employed it is not plausible that these mechanism should depend on only a single neuron.

Stating that correlational mechanisms characterise processes such as discrimination, masking, detection and classification may, like postulating “nonlinear pooling”, be seen as a functional description of neuronal mechanisms, but not yet as a detailed characterisation of the specific neuronal interactions which they represent. Such a characterisation would be useful in order to predict, e.g. the way that the activity of channels is pooled, or how the adaptive, correlational mechanisms work in a given experiment, and so one is lead to look for possible models of interactive neuronal processes.

There exists a number of network models and we will briefly consider two of them in order to indicate (1) possible ways of approaching the question how networks of the visual system may be conceptualised, and (2) the difficulties one encounters when choosing a particular way. To begin with, let us note that the stabilised Hebb rule has been formalised as a differential equation, i.e. (10), without reference to the electrical and chemical mechanisms leading to the formation and timing of action potentials, so (10) describes the neuron in a rather global way; the input s is constant in time, and the response v , determined as the scalar product $m\langle \mathbf{s}, \mathbf{h} \rangle$, will change continuously in time as long as $d\mathbf{h}/dt \neq 0$. One way of interpreting this is to say that the response v represents the mean firing rate of the neuron, where the mean firing rate is defined as the reciprocal of the mean interspike interval.

An example for a model of the interactions among neurons with neuronal responses being explicitly defined as mean firing rates is that of Oğuztöreli (1979). In this model, the responses v_i , $1 \leq i \leq k$ of k neurons are specified by a nonlinear system of ordinary integro-differential difference equations. None of the parameters of the model (e.g. the coupling coefficients denoting the influence of the j th on the i th neuron, or the self-inhibition/excitation factors) is time dependent, i.e. the model may be said to be characterised by a rigid wiring. Still, Oğuztöreli and Caelli (1986) were able to show that the system is able, at least in principle, to explain response selectivity of retinal ganglion cells. Since some properties of Oğuztöreli's (1979) model have already been explored (Caelli et al. 1985; Oğuztöreli et al. 1986), it is tempting to try to implement the stabilised Hebb rule into this model in order to arrive at a characterisation of an assembly with adapting neurons. Note, however, that, in theorem 2, the receptive field of a neuron is just the set of all inputs defined as outputs of some pre-filter; there is no input to a neuron transmitted by a horizontal connection to another neuron, meaning that implicitly a "vertical" organisation of the cortex is assumed. As Gilbert et al. (1996) pointed out, the inputs from horizontal connections should, however, be considered as part of the receptive field of a neuron. Theorem 2 may still be valid for any member of a set of interconnected neurons; however, a detailed discussion of the convergence properties of adapting cells as members of such systems has to be left to future work⁵.

The assumption that the activity of an assembly can be adequately represented by the mean firing rates of its neurons may be challenged. Referring to the fact that, since the information capacity of a coding schemes is much higher when based on interspike intervals than on spiking rates, Gerstner et al. (1993) argued that it is the spike patterns and not the rate with which spikes are generated that should be taken to characterise the responses of neurons; the spatio-temporal spike pattern generated by a stimulus does not necessarily imply a change of the mean firing rate. The spike pattern may change within a few milliseconds, while the change of the mean firing rate may take much longer. Quoting Bialek et al. (1991), the authors point out that a fly could most likely not survive if information were encoded in mean firing rates. Gerstner et al. (1993) proposed what they called a spike response model (SRM) allowing for a time resolution of not more than 1 ms; learning is characterised by a Hebb rule [not containing a stabilising term as in (10), though] that is (1) formulated with explicit reference to the electrical and chemical processes at the synapses of a neuron and that (2) allows for the storage

(and retrieval) of different patterns. A very interesting aspect of the model refers to the speed with which neurons can adapt; the model of Gerstner et al. (1993) provides some (indirect) justification underlying our fitting the CA model to data, namely that adaptation to the stimulus pattern had already taken place or, equivalently, that adaptation is very fast compared to the duration of the stimulus presentation. Ritz et al. (1994) showed how the SRM model may be employed to discuss feature linking and pattern segmentation processes.

In order to arrive at a more realistic version of our CA model, one may try to adopt a SRM-type model with a stabilised Hebb rule; of course, the remarks concerning the convergence properties of neurons within a network made above with respect to Oğuztöreli's model carry over to SRM-type networks. On the other hand, the fact that we were able to fit our simple model to some data may mean that, indeed, for certain stimulus patterns some network of neurons mimics the behaviour of a single, adapted neuron of the type described in Sect. 3.2.

In any case, regardless of whether one adopts a model the type proposed by Oğuztöreli's (1979) or Gerstner et al. (1977), the question how a network model can be related to psychophysical data has no easy answer. For instance, the probability of detection depends upon the contrast m of a pattern, which suggests that the response of the visual system (usually) increases with m . To the extent that the value of m influences the firing rate of the neurons involved in the detection process, network models of the type proposed by Oğuztöreli (1979) may be taken as a promising point of departure, although one has then to face the criticism of Gerstner et al. (1993) of models based on mean firing rates (see also Gerstner et al. 1997). On the other hand, the relation between the value of m and the probability of triggering the spatio-temporal spike pattern representing a stimulus pattern has yet to be formulated. There are a number of open questions for future research.

Appendix 1: Eigenvectors and eigenvalues of C_s

Let \mathbf{w}_i , $1 \leq i \leq n$ be the eigenvectors of C_s and μ_i the corresponding eigenvalues. Suppose, in particular, that $\mu_1 \neq 0$. Then, $C_s \mathbf{w}_1 = \mathbf{s} \mathbf{s}' \mathbf{w}_1 = \mu_1 \mathbf{w}_1$. Multiplying from the left with \mathbf{s}' , $\mathbf{s}' \mathbf{s} \mathbf{s}' \mathbf{w}_1 = \mu_1 \mathbf{s}' \mathbf{w}_1$, from which $\mathbf{s}' \mathbf{s} = \|\mathbf{s}\|^2 = \mu_1$ follows. Further, $\mathbf{s} \mathbf{s}' \mathbf{w}_1 = \mathbf{s} (\mathbf{s}' \mathbf{w}_1) = \mu_1 \mathbf{w}_1$ so that

$$\mathbf{w}_1 = \frac{\langle \mathbf{w}_1, \mathbf{s} \rangle}{\|\mathbf{s}\|^2} \mathbf{s} .$$

Let $\mathbf{w}_i \neq 0$, $i \neq 1$ be some other eigenvector. Since C_s is symmetric, $\mathbf{w}_i \perp \mathbf{w}_1$, and therefore $\mathbf{w}_i \perp \mathbf{s}$. Then, $\mathbf{s} \mathbf{s}' \mathbf{w}_i = \mu_i \mathbf{w}_i = \mathbf{0}$, because $\mathbf{s}' \mathbf{w}_i = 0$; consequently, $\mu_i = 0$. See chapter 1 in Rao (1975) for a concise text on vector algebra.

Appendix 2: Superposition experiments

Consider some linear system with constant coefficients exposed to a stimulus pattern $s = \phi(m)s_t + ms_b$; here, s_t is some "test" pattern with contrast $\phi(m)$ and s_b is some "background" pattern with contrast m , in our context either a sine or a cosine grating. Let the unit responses, i.e. the responses to s_t and s_b , of the system be w_s

⁵ In the simplest case, the set of interconnected neurons can be considered as a channel with (1) the interconnections not part of the RF of the channel, (2) the RF of the channel defined as the union of the receptive fields of the individual neurons and (3) response as the linear superposition of the neurons' responses; theorem 2 would then imply that the network behaves like a matched filter. However, this model cannot be true in general because it would imply the validity of the matched filter model for any pattern

and w_b , so that the total response is given by $\phi(m)w_s + mw_b$. Suppose that the response is interpreted as “detection of s_i ” if

$$\begin{aligned} \phi(m)w_s + mw_b &= d, & m \neq 0, \\ m_0s w_s &= d, & m = 0. \end{aligned} \quad (A1)$$

Note that if the value of m is varied, ϕ has to be adjusted in a corresponding way such that (32) holds. We may assume that, for some small range of m values, ϕ can be approximated by a linear function (for a more detailed discussion, see Meinhardt et al. 1998). (32) implies $\phi(m)w_s/(m_0s w_s) + mw_b/(m_0s w_s) = 1$ and therefore

$$\frac{\phi(m)}{m_0s} = 1 - m \frac{w_b}{d}; \quad (A2)$$

this is the normalised CIF. This equation implies immediately

$$\frac{w_b}{d} = \frac{m_0s - \phi(m)}{m_0s m}, \quad (A3)$$

and so w_b can be estimated from the contrasts m , m_0s , and $\phi(m)$. Now, let w_s and w_b be the response of a neuron from C_s . The input is given by $g(x)$, the response of the pre-filter. If the input to the pre-filter is a sinusoidal grating $\sin(2\pi fx)$ or $\cos(2\pi fx)$, the response of the pre-filter is $|H(f)| \sin(2\pi fx)$. Let $g(x)$ be the response of the pre-filter to s_i ; since the impulse response of the neuron adapted to that part of g covering its receptive field is given by $h(x - \xi) = g(x)$ one has

$$\begin{aligned} w_b(x) &= |H(f)| \int_{x-\rho}^{x+\rho} h(x - \xi) \sin(2\pi f \xi) d\xi \\ &= |H(f)| \int_{x-\rho}^{x+\rho} g(x) \sin(2\pi f \xi) d\xi. \end{aligned} \quad (A4)$$

So, the sensitivity can be estimated from the data and be predicted if $H(f)$ is known.

References

- Abeles M, Prut Y, Bergman H, Vaadia E, Aertsen AMHJ (1993) Integration, synchronicity and plasticity. In: Aertsen AJHM (ed) Brain theory: spatio-temporal parts of brain function. Elsevier, Amsterdam, pp 205–244
- Aertsen A, Gerstein G (1991) Dynamical parts of neuronal cooperativity: fast stimulus-locked modulations of effective connectivity. In: Krüger J (ed) Neuronal cooperativity. Springer, Berlin Heidelberg New York, pp 52–67
- Aertsen A, Erb M, Palm G (1994) Dynamics of functional coupling in the cerebral cortex: an attempt at a model-based interpretation. *Physica D*:75, 103–128
- Ahissar M, Laiwand R, Kozminsky G, Hochstein S (1998) Learning pop-out detection: building representations for conflicting target-distractor relationships. *Vision Res* 38: 3095–3107
- Amari SI (1977) Neural theory of association and concept-formation. *Biol Cybern* 26: 175–185
- Arnol'd VI (1980) *Gewöhnliche Differentialgleichungen*. Springer, Berlin, Heidelberg, New York
- Arrowsmith DK, Place CM (1982) *Ordinary differential equations*. Chapman and Hall, London
- Arrowsmith DK, Place CM (1990) *An introduction to dynamical systems*. Cambridge University Press, Cambridge
- Beard BL, Levi DM, Reich LN (1995) Perceptual learning in parafoveal vision. *Vision Res* 35: 1679–1690
- Bialek W, Rieke F, Ruyter van Steveninck RR de, Warland D (1991) Reading a neural code. *Science* 252: 1854–1857
- Bienenstock EL, Cooper LN, Munro PW (1982) Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J Neurosci* 2: 32–48
- Bonhoeffer T, Staiger V, Aertsen A (1989) Synaptic plasticity in rat hippocampal slice cultures: local “Hebbian” conjunction in pre- and postsynaptic stimulation leads to distributed synaptic enhancement. *Proc Natl Acad Sci USA* 86: 8113–8117
- Braitenberg V (1978) Cell assemblies in the cerebral cortex. In: Heim R, Palm G (eds) *Theoretical approaches to complex systems*. Proceedings, Tübingen 1977. Springer, Berlin Heidelberg New York, pp 171–188
- Caelli T, Moraglia G (1985) On the detection of Gabor signals and discrimination of Gabor textures. *Vision Res* 25: 671–684
- Caelli T, Öguztöreli MN (1987) Some task and signal dependent rules for spatial vision. *Spat Vis* 2: 295–315
- Caelli T, Steil G, Öguztöreli NM (1985) Control properties of perceptual transient and sustained mechanisms. *Biol Cybern* 52: 177–186
- Caelli T, Rentschler I, Scheidler W (1987) Visual pattern recognition in humans. *Biol Cybern* 57: 233–240
- Campbell FW, Robson JG (1968) Application of Fourier analysis to the visibility of gratings. *J Physiol (Lond)* 197: 551–566
- Christ RE, Kapadia MK, Westheimer G, Gilbert CD (1997) Perceptual learning of spatial localization: specificity for orientation, position, and context. *J Neurophysiol* 78: 2889–2894
- Darian-Smith C, Gilbert CD (1994) Axonal sprouting accompanies functional reorganization in adult cat striate cortex. *Nature* 368: 737–740
- Darian-Smith C, Gilbert CD (1995) Topographic reorganization in the striate cortex of the adult cat and monkey is cortically mediated. *J Neurosci* 15: 1631–1647
- Das A, Gilbert CD (1995) Receptive field expansion in adult visual cortex is linked to dynamic changes in strength off cortical connectins. *J Neurophysiol* 74: 779–792
- Daugman JG (1984) Spatial visual channels in the Fourier plane. *Vision Res* 24: 891–910
- Du Buf JMH (1992) Modeling spatial vision at the threshold level. *Spat Vis* 6: 25–60
- Du Buf JMH (1993) Responses of simple cells: events, interferences, and ambiguities. *Biol Cybern* 68: 321–333
- Du Buf JMH (1994) Ramp edges, Mach bands, and the functional significance of the simple cell assembly. *Biol Cybern* 70: 449–461
- Fahle M, Edelman S (1993) Long-term learning in Vernier acuity: effects of stimulus orientation, range and of feedback. *Vision Res* 33: 397–412
- Fischer G (1997) *Lineare Algebra*. Vieweg, Braunschweig
- Gardiner CW (1990) *Handbook of stochastic methods for physics, chemistry and the natural sciences*. Springer, Berlin Heidelberg New York
- Gerstner W, Ritz R, van Hemmen JL (1993) Why spikes? Hebbian learning and retrieval of time-resolved excitation patterns. *Biol Cybern* 69: 503–515
- Gerstner W, Kreiter AK, Markram H, Herz AVM (1997) Neural codes: firing rates and beyond. *Proc Natl Acad Sci USA* 94: 12740–12741
- Gilbert CD, Wiesel TN (1992) Receptive field dynamics in adult primary visual cortex. *Nature* 356: 150–152
- Gilbert CD, Das A, Ito M, Kapadia M, Westheimer G (1996) Spatial integration and cortical dynamics. *Proc Natl Acad Sci USA* 93: 615–622
- Graham N (1977) Visual detection of aperiodic spatial stimuli by probability summation among narrowband channels. *Vision Res* 17: 637–652
- Graham N (1989) *Visual pattern analyzers*. Oxford University Press, Oxford
- Hauske G, Wolf W, Lupp U (1976) Matched filters in human vision. *Biol Cybern* 22: 181–188
- Hauske G, Lupp U, Wolf W (1978) Matched filters. A new concept in vision. *Soc Photogr Sci Eng* 20: 59–64
- Hebb DO (1949) *The organisation of behavior*. Wiley, New York
- Herzog MH, Fahle M (1998) Modeling perceptual learning: difficulties and how they can be overcome. *Biol Cybern* 78: 107–117

- Hines M (1976) Line spread variation near the fovea. *Vision Res* 14: 567–572
- Karni A, Sagi D (1991) Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proc Natl Acad Sci USA* 88: 4966–4970
- Kirkwood A, Rioult M, Bear MF (1996) Experience-dependent modification of synaptic plasticity in visual cortex. *Nature* 381: 526–528
- Koenderink JJ, van Doorn AJ (1990) Receptive field families. *Biol Cybern* 63: 291–297
- Kohonen T (1989) *Self-organization and associative memory*. Springer, Berlin Heidelberg New York
- Kulikowski JJ, King-Smith PE (1973) Spatial arrangement of line, edge and grating detectors revealed by subthreshold summation. *Vision Res* 13: 1455–1478
- Leadbetter MR, Lindgren G, Rootzén H (1983) *Extremes and related properties of random sequences and processes*. Springer, Berlin Heidelberg New York
- Logvinenko AD (1993) Lack of convexity of threshold curves for compound grating: implications for modelling visual pattern detection. *Biol Cybern* 70: 55–64
- Logvinenko AD (1996) Convexity of a set of subthreshold stimuli implies a peak detector. *Spat Vis* 2: 105–123
- Malsburg C von der (1983) The correlation theory of brain function. Internal report 81-2, July 1981, Department of Neurobiology, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Revised Printing 1983
- Meinhardt G (1996) Spatial summation in retinal inhomogeneous receptive fields. *Perception* 25 [Suppl]:79
- Meinhardt G (1999) Evidence for different nonlinear summation schemes for lines and gratings at threshold. *Biol Cybern* 81: 263–277
- Meinhardt G, Mortensen U (1998) Detection of aperiodic test patterns by pattern specific detectors revealed by subthreshold summation. *Biol Cybern* 79: 413–425
- Mortensen U, Suhl U (1991) An evaluation of sensory noise in the human visual system. *Biol Cybern* 66: 37–47
- Mostafavi H, Sakrinson DJ (1976) Structure and properties of a single channel in the human visual system. *Vision Res* 16: 857–958
- Nachtigall C (1991) Modellierung eines selbstorganisierten Matched-Filters durch ein System gewöhnlicher Differentialgleichungen. Diplomarbeit am Fachbereich Mathematik der Westf. Wilhelms-Universität, Münster
- Oğuztöreli MN (1979) Activity analysis of neural networks. *Biol Cybern* 34: 159–169
- Oğuztöreli MN, Caelli T (1986) The relationship between signal selectivity and the functional structure of complex retinal neural networks. *Biol Cybern* 53: 229–238
- Oğuztöreli MN, Steil G, Caelli TM (1986) Control mechanisms of a neural network. *Biol Cybern* 54: 21–28
- Oja E (1982) A simplified neuron model as a principal component analyzer. *J Math Biol* 15: 267–273
- Papoulis A (1988) *Signal analysis*. McGraw Hill, New York
- Peterson WW, Birsall TG, Fox WC (1954) The theory of signal detectability. *Trans IRE professional group in information theory*. PGIT 2: 171–212
- Pettet MW, Gilbert CD (1992) Dynamic changes in receptive-field size in cat primary cortex. *Proc Natl Acad Sci USA* 89: 8366–8370
- Poggio T, Fahle M, Edelman S (1992) Fast perceptual learning in visual hyperacuity. *Science* 256: 1018–1021
- Press WH, Teukolsky SA, Flamery BP, Vetterling WT (1996) *Numerical recipes in Pascal*. Oxford University Press, Oxford
- Quick RF (1974) A vector magnitude model of contrast detection. *Kybernetik* 16: 67–67
- Rao CR (1975) *Linear statistical inference and its applications*. Wiley, New York
- Ritz R, Gerstner W, Fuentes U, van Hemmen JL (1994) A biologically motivated and analytically soluble model of collective oscillations in the cortex. II. Application to binding and pattern segmentation. *Biol Cybern* 71: 349–358
- Sachs MB, Nachmias J, Robson JG (1971) Spatial frequency channels in human vision. *J Opt Soc Am* 61: 1176–1186
- Sanger TD (1989) Optimal unsupervised learning in a single-layer-linear feedforward neural network. *Neural Netw* 2: 459–473
- Sanger TD (1990) Analysis of the two-dimensional receptive fields learned by the generalized Hebbian algorithm in response to random input. *Biol Cybern* 63: 221–228
- Whalen AD (1971) *Detection of signals in noise*. Academic Press, New York
- Wilson HR (1978) Quantitative characterisation of two types of line-spread function near the fovea. *Vision Res* 18: 971–981
- Wilson HR, Bergen JR (1979) A four mechanism model of threshold spatial vision. *Vision Res* 19: 515–522
- Wilson HR, Levi D, Maffei L, Rovamo J, DeValois R (1990) The perception of form: retina to striate cortex. In: Spillmann L, Werner JS (1990) *Visual perception. The neurophysiological foundations*. Academic Press, San Diego, pp 231–272
- Yang J, Reeves A (1995) Bottom-up visual image processing probed with weighted Hermite polynomials. *Neural Netw* 8: 669–691
- Zeki S (1990) Parallelism and functional specialization in human visual cortex. *Cold Spring Harbor Symposia on Quantitative Biology*. Cold Spring Harbor Lab. Press, New York. LV 651–661