Wiskott, L. and Berkes, P. (2003). Slow feature analysis yields a rich repertoire of complex-cells properties. Cognitive Sciences EPrint Archive (CogPrint) 2785, http://cogprints.ecs.soton.ac.uk/archive/00002785/

Slow feature analysis yields a rich repertoire of complex cell properties

Pietro Berkes and Laurenz Wiskott
Institute for Theoretical Biology
Humboldt University Berlin
Invalidenstraße 43, D-10115 Berlin, Germany
{p.berkes,l.wiskott}@biologie.hu-berlin.de
http://itb.biologie.hu-berlin.de/{~berkes,~wiskott}

Abstract

In this study, we investigate temporal slowness as a learning principle for receptive fields using slow feature analysis, a new algorithm to determine functions that extract slowly varying signals from the input data. We find that the learned functions trained on image sequences develop many properties found also experimentally in complex cells of primary visual cortex, such as direction selectivity, non-orthogonal inhibition, end-inhibition and side-inhibition. Our results demonstrate that a single unsupervised learning principle can account for such a rich repertoire of receptive field properties.

Key words: Complex cells, slow feature analysis, temporal slowness, model, spatio-temporal, receptive fields

1 Introduction

Primary visual cortex (V1) is the first cortical area dedicated to visual processing. This area has been intensively studied neurophysiologically since the seminal work by Hubel and Wiesel (1962), who also introduced the standard classification of neurons in V1 in two main groups: *simple* and *complex cells*. These neurons are conceived as edge or line detectors: simple cells respond to bars having a specific orientation and position in the visual field; complex cells also respond to oriented bars but are insensitive to their exact position.

Idealized simple and complex cells are well described by Gabor wavelets, which have the shape of sine gratings with a Gaussian envelope function (Figure 2 A). A single Gabor wavelet used as a linear filter is similar to a simple cell, since the response depends on the exact alignment of a stimulus edge on an excitatory (positive) subfield of the wavelet. Taking the square sum of the responses of two Gabor wavelets with identical envelope function, frequency, and orientation but with a 90° phase difference, one obtains a response similar to that of a complex cell, insensitive to the exact location of the bar (following a rule similar to the relation $\sin(x)^2 + \cos(x)^2 = 1$) while still being sensitive to its orientation. We will refer to these models as the *classical models* of simple and complex cells.

This idealized picture, however, is clearly not complete. Complex cells in V1 show a much richer repertoire of receptive field properties than can be explained with this simple model. For example, they show end-inhibition, direction selectivity, and sharpened or broadened tuning to orientation or frequency.

In this work we investigate *temporal slowness* as a possible computational principle for the emergence of complex cell receptive fields[1]. The slowness principle is based on the observation that the environment, sensory signals, and internal representations of the environment vary on different time scales. The environment,

[1] An early description of the principle was given by Hinton (1989) (p. 208) and early models based on temporal slowness were presented by Földiák (1991) and Mitchison (1991). Successive studies applied this principle to the extraction of disparity from stereograms (Stone, 1996) or from artificially generated simple cell outputs (Wiskott and Sejnowski, 2002), and to blind source separation (Stone, 2001). Other studies modeling complex and simple cells are discussed in the Conclusion section.

e.g. the objects we see around us, changes usually on a relatively slow time scale. Sensory signals on the other hand, such as the responses of single receptors in the retina, have a faster time scale, because even a small eye movement or shift of a textured object may lead to a rapid variation of the light intensity received by a receptor neuron. The internal representation of the environment, finally, should vary on a time scale similar to that of the environment itself, i.e. on a slow time scale. If we succeed in extracting slowly varying features from the quickly varying sensory signal, then the features are likely to reflect the properties of the environment and are in addition invariant or at least robust to frequent transformations of the sensory input, such as visual translation, rotation, or zoom. Our working hypothesis is that the cortex organizes according to this principle in order to build a consistent internal representation of the environment. In order to verify this hypothesis we consider a general space of non-linear input-output functions, determine those functions that extract the slowest features in response to natural image sequences, and compare their behavior to that of complex cells in V1 described in the literature.

The following section introduces the slow feature analysis algorithm. Section 3 presents the input data set and the methods used to analyze the results. Section 4 describes the simulation results and compares the learned functions with neurons described in the physiological literature.

2 Slow Feature Analysis

2.1 Problem statement

The problem of extracting slow features from time sequences can be formally stated as follows (Wiskott and Sejnowski, 2002): given a multidimensional input signal $\mathbf{x}(t) = (x_1(t), \ldots, x_N(t))^T$, $t \in [t_0, t_1]$, find a set of real-valued functions $g_1(\mathbf{x}), \ldots, g_M(\mathbf{x})$ so that for the output signals $y_j(t) := g_j(\mathbf{x}(t))$

$$\Delta(y_j) := \langle \dot{y}_j^2 \rangle_t \quad \text{is minimal} \tag{1}$$

under the constraints

$$\begin{aligned} \langle y_j \rangle_t &= 0 \quad \text{(zero mean)}\,, && (2)\\ \langle y_j^2 \rangle_t &= 1 \quad \text{(unit variance)}\,, && (3)\\ \forall i < j, \quad \langle y_i y_j \rangle_t &= 0 \quad \text{(decorrelation)}\,, && (4) \end{aligned}$$

with $\langle . \rangle_t$ and $\dot{y}$ indicating time averaging and the time derivative of y, respectively. The zero-mean constraint (2) is present for convenience only, so that eqs. (3) and (4) take a simple form. Constraint (3) means that each signal should carry some information and avoids the trivial solution $g_j(\mathbf{x}) = 0$. Alternatively, one could drop this constraint and divide the right side of (1) by the variance $\langle y_j^2 \rangle_t$. Constraint (4) forces different signals to code for different aspects of the input.

We solve this optimization problem with slow feature analysis (SFA) (Wiskott, 1998; Wiskott and Sejnowski, 2002), an unsupervised algorithm based on an eigenvector approach. Slow feature analysis permits to consider general finite dimensional function spaces and is efficient enough to do simulations of reasonable scale in terms of size and dimensionality of the input signals. The following two sections sketch the mathematical background and the definition of the algorithm.

2.2 The linear case

Consider first the linear case $g_j(\mathbf{x}) = \mathbf{w}_j^T \mathbf{x}$, for some input $\mathbf{x}$ and *weight vectors* $\mathbf{w}_j$. In the following we assume $\mathbf{x}$ to have zero mean (i.e. $\langle \mathbf{x} \rangle_t = \mathbf{0}$) for simplicity and without loss of generality.

We can write equation (1) as

$$\Delta(y_j) := \langle \dot{y_j}^2 \rangle_t = \langle (\mathbf{w}_j^T \dot{\mathbf{x}})^2 \rangle_t = \mathbf{w}_j^T \langle \dot{\mathbf{x}} \dot{\mathbf{x}}^T \rangle_t \mathbf{w}_j \,. \tag{5}$$

If we integrate Constraint (3) in the objective function (5), as suggested in the previous section, we obtain:

$$\Delta(y_j) = \frac{\mathbf{w}_j^T \langle \dot{\mathbf{x}} \dot{\mathbf{x}}^T \rangle_t \mathbf{w}_j}{\mathbf{w}_j^T \langle \mathbf{x} \mathbf{x}^T \rangle_t \mathbf{w}_j} =: \frac{\mathbf{w}_j^T \mathbf{A} \mathbf{w}_j}{\mathbf{w}_j^T \mathbf{B} \mathbf{w}_j} \,. \tag{6}$$

It is known from linear algebra that the weight vectors $\mathbf{w}_j$ that minimize this equation correspond to the eigenvectors of the generalized eigenvalue problem

$$\mathbf{AW} = \Lambda \mathbf{BW} \,, \tag{7}$$

where $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_N)$ is the matrix of the generalized eigenvectors and Λ is the diagonal matrix of the generalized eigenvalues $\lambda_1, \ldots, \lambda_N$ (see e.g. Gantmacher, 1959, chap. 10.7, Theorem 8, 10 and 11, pp.310–320). In particular the vectors $\mathbf{w}_j$ can be chosen such that $\mathbf{w}_i^T \mathbf{B} \mathbf{w}_j = \delta_{ij}$, which implies that Constraints (3) and (4) are fulfilled:

$$\langle y_j^2 \rangle_t = \langle (\mathbf{w}_j^T \mathbf{x})^2 \rangle_t = \mathbf{w}_j^T \langle \mathbf{x}\mathbf{x}^T \rangle_t \mathbf{w}_j = \mathbf{w}_j^T \mathbf{B} \mathbf{w}_j = 1\,, \tag{8}$$

$$\langle y_i y_j \rangle_t = \langle (\mathbf{w}_i^T \mathbf{x})(\mathbf{w}_j^T \mathbf{x}) \rangle_t = \mathbf{w}_i^T \langle \mathbf{x}\mathbf{x}^T \rangle_t \mathbf{w}_j = \mathbf{w}_i^T \mathbf{B} \mathbf{w}_j = 0\,. \tag{9}$$

Note that, by equation (6), $\lambda_j = \Delta(y_j)$, so that by sorting the eigenvectors by increasing eigenvalues we induce an order where the slowest signals have lowest indices (i.e. $\Delta(y_1) \leq \Delta(y_2) \leq \ldots$).

2.3 The general case

In the more general case of a finite dimensional function space $\mathcal{F}$, we consider a basis $h_1, \ldots, h_M$ of $\mathcal{F}$. For example, in the standard case where $\mathcal{F}$ is the space of all polynomials of degree n, the basis will include all monomials up to order n.

Defining the *expanded input* $\mathbf{h}(\mathbf{x}) := (h_1(\mathbf{x}), \ldots, h_M(\mathbf{x}))$, every function $g \in \mathcal{F}$ can be expressed as $g(\mathbf{x}) = \sum_{k=1}^{M} w_k h_k(\mathbf{x}) = \mathbf{w}^T \mathbf{h}(\mathbf{x})$. This leads us back to the linear case if we assume that $\mathbf{h}(\mathbf{x})$ has zero mean (again, without loss of generality), which can be easily obtained in practice by subtracting the mean over time from the expanded input signal (this reduces the dimensionality of the problem by one).

We can now formulate the Slow Feature Analysis (SFA) algorithm (Wiskott and Sejnowski, 2002):

Non-linear expansion: expand the input data and subtract the mean over time to obtain the expanded signal $\mathbf{z} := \mathbf{h}(\mathbf{x}) = (h_1(\mathbf{x}), \ldots, h_M(\mathbf{x}))$.

Slow Feature Extraction: solve the generalized eigenvalue problem

$$\mathbf{A}\mathbf{W} = \Lambda \mathbf{B} \mathbf{W}\,, \tag{10}$$

$$\text{with} \quad \mathbf{A} := \langle \dot{\mathbf{z}} \dot{\mathbf{z}}^T \rangle_t \tag{11}$$

$$\text{and} \quad \mathbf{B} := \langle \mathbf{z}\mathbf{z}^T \rangle_t\,. \tag{12}$$

The N eigenvectors $\mathbf{w}_1, \ldots, \mathbf{w}_N$ corresponding to the lowest generalized eigenvalues $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_N$ define the non-linear input-output functions $g_1(\mathbf{x}), \ldots, g_N(\mathbf{x})$:

$$g_j(\mathbf{x}) := \mathbf{w}_j^T \mathbf{h}(\mathbf{x}) \tag{13}$$

which satisfy constraints (2-4) and minimize (1).

In the simulations presented here, the derivative of $\mathbf{z}(t)$ is computed by the linear approximation $\dot{\mathbf{z}}(t) = \mathbf{z}(t + \mathrm{d}t) - \mathbf{z}(t)$. Other simulations performed with cubic interpolation show equivalent results. The function space $\mathcal{F}$ is chosen to be the set of all polynomials of order 2, mainly because of limited computational resources. This function space, however, is much larger than the ones spanned by the neural networks used in standard studies in the field (cf. Conclusions). Those models usually rely either on linear networks, which span the space of all linear functions, or on networks with one layer of a small number (typically 2 or 4) of linear units followed by a quadratic non-linearity, which span a small subset of the space of polynomials of order 2.

3 Methods

3.1 Input data

Our data source consisted of 36 gray-valued natural images extracted from the Natural Stimuli Collection of H. van Hateren[2] and preprocessed as suggested in (van Hateren and van der Schaaf, 1998) (block averaging and taking the logarithm of the pixel intensities). We constructed image sequences by moving a window over the images by translation, rotation, and zoom and subsequently rescaling the frames to a standard size of 16×16 pixels. The transformations were performed simultaneously, so that each frame differed from the

[2]The Natural Stimuli Collection of H. van Hateren is available online at http://hlab.phys.rug.nl/archive.html

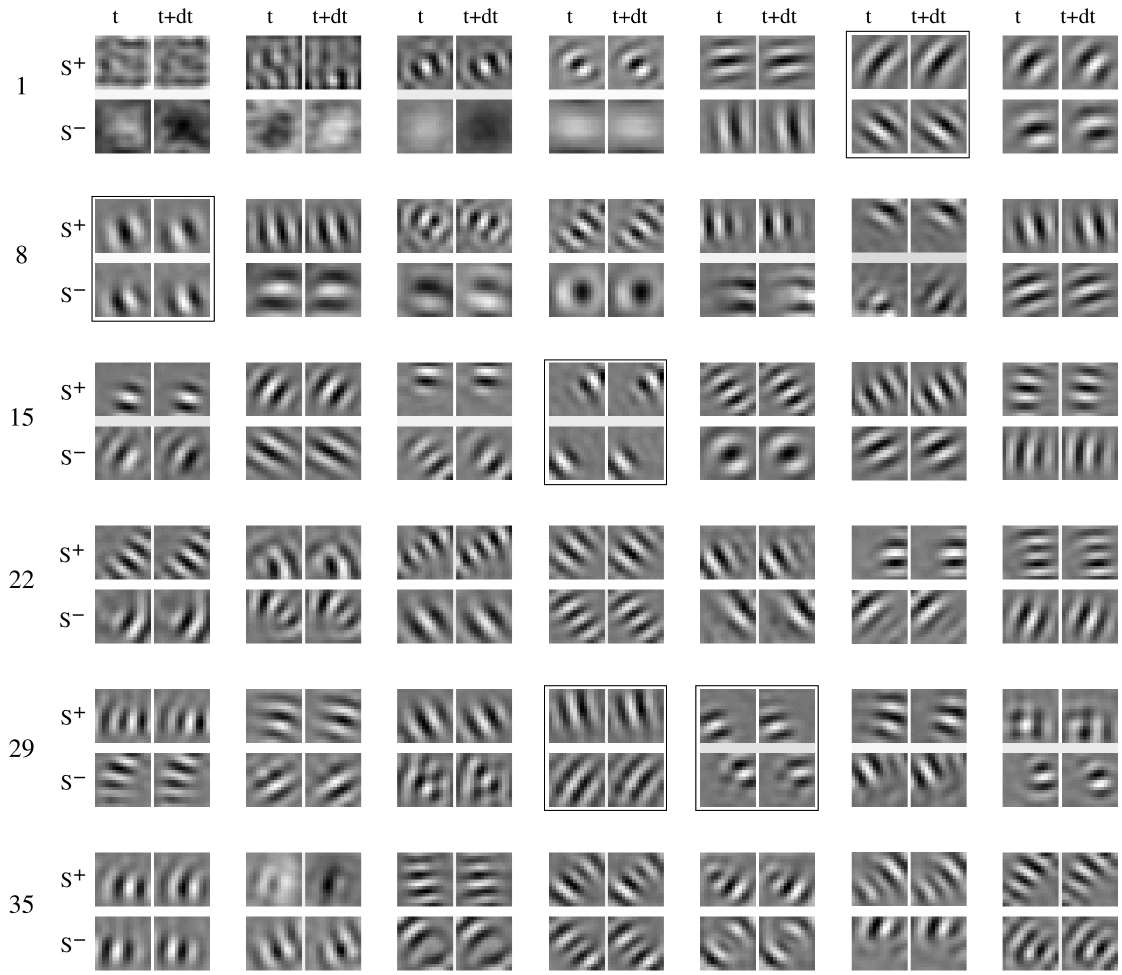

Figure 1 Optimal stimuli, overview. Optimal excitatory stimuli (S^+) and optimal inhibitory stimuli (S^-) of the first 41 units of the simulation described in the text. For most units S^+ and S^- look like Gabor wavelets in agreement with physiological data. The optimal stimuli give information about the preferred frequency and orientation of a unit. A comparison between the patches at time t and at time $t + \mathrm{d}t$ hints at the temporal structure of the receptive field, e.g. its preferred speed. The units surrounded by a black frame are magnified and described in Figure 2.

previous one by position, orientation, and scale. We produced in this way 250,000 frames per simulation. To include temporal information, the input vectors to SFA were formed by the pixel intensities of two consecutive frames at times t and $t + \mathrm{d}t$.

A run with these settings requires the computation of large covariance matrices having in the order of $\mathcal{O}(N^4)$ elements, where N is the input dimension. We perform a standard preprocessing step using PCA to reduce the dimensionality of the input vectors from $512 = 16 \times 16 \times 2$ dimensions to 100.

3.2 Analysis of results

SFA learns a set of second degree polynomials that applied to our input visual stimuli have the most slowly varying output. We refer to the i-th polynomial as the *i-th unit*. The units are ordered by slowness (the first unit being the slowest) and their outputs are mutually uncorrelated. As the sign of a unit's response is arbitrary in the optimization problem, we choose it in this paper such that the strongest response to an input vector with norm r is positive (i.e. such that the magnitude of the response to S^+ is greater than that to S^-; see below).

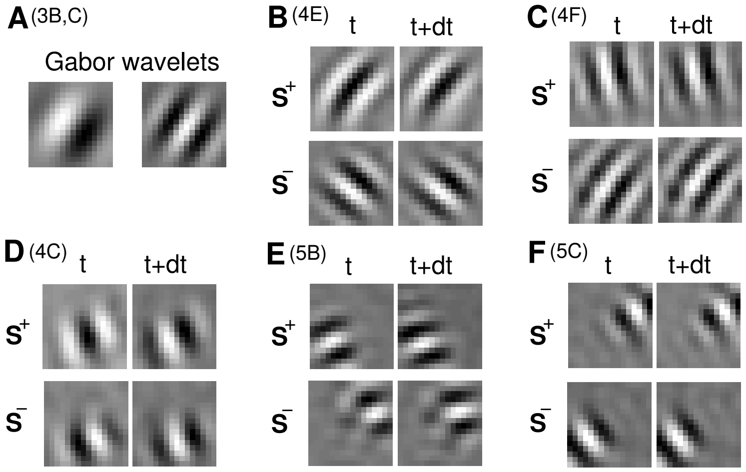

Figure 2 Gabor wavelets and selected optimal stimuli. **(A)** Two Gabor wavelets with frequencies of 2.5 and 4.5 cycles/degree of arc and equal orientation. These wavelets were used to produce the images of Figures 3 B and C according to the classical simple and complex cell models. **(B–F)** Optimal excitatory stimuli (S^+) and optimal inhibitory stimuli (S^-) for some of the units shown in Figure 1, obtained from the simulation described in the text. The units are chosen to illustrate typical behaviors, namely orthogonal inhibition (B), non-orthogonal inhibition (C), direction selectivity (D), end-inhibition (E), and side-inhibition (F). Response images corresponding to B–F are shown in Figures 4 E,F,C and Figures 5 B,C, respectively, as indicated in brackets.

Since the input vectors are pairs of image patches, the functions g_j can be interpreted as non-linear spatio-temporal receptive fields of neurons in V1 and tested with input stimuli much like in neurophysiological experiments. The units can have a spontaneous firing rate, i.e. a non-zero response to a blank visual input. As in physiological experiments we interpret an output lower than the spontaneous one as active inhibition.

In order to analyze the units, we first compute for each of them the optimal excitatory stimulus S^+ and the optimal inhibitory stimulus S^-, i.e. the input that elicits the strongest and the weakest output from the unit, respectively, given a constant norm r of the input vector (Figure 1, Figure 2 B-F). This is in analogy to the physiological practice of characterizing a neuron by the stimulus to which the neuron responds best. We choose r to be the mean norm of the training input, since we want S^+ and S^- to be representative of the typical input. The optimal stimuli give us information for example about the preferred frequency and orientation of the unit. In addition, a comparison between the optimal stimulus at time t and at time $t + \mathrm{d}t$ hints at the temporal structure of the receptive field. However, since the units are non-linear the optimal stimuli give only a partial view of their behavior.

To gain further insight into the response properties we use an appropriate pair of *test images* (one at time t and one at time $t + \mathrm{d}t$) and compute for each unit the corresponding *response image*[3]. The response image is computed by cutting at each point of the test images a 16×16 window, using it as the input to the unit and plotting its output at the corresponding point. This leads to response images like those shown in Figure 4 B-F and Figure 5 B,C. An illustrative example for the classical simple and complex cells models on a static test image is shown in Figure 3.

To study the response to a range of frequencies and orientations, we use a test image that consists of a circular pattern of sine waves with frequency increasing from the circumference to the center; we let the ring patterns move outward at the preferred speed of the unit (Figure 4 A). These images contain information not only about the whole range of frequencies and orientations to which the unit responds or by which it is inhibited, but also about the sensitivity of the unit to the phase of the grating. If a unit is sensitive, the response image will show oscillations in the radial direction, while if there are no oscillations the unit is phase-invariant. Moreover, since at two opposite points of a ring the grating is moving in opposite directions,

[3] A different technique, that consists in computing the invariances of the optimal stimuli, i.e. the directions in which a variation of S^+ or S^- has the least effect on the output of the unit, is described in (Berkes and Wiskott, 2002) or online at http://itb.biologie.hu-berlin.de/~berkes/ICANN02/results.shtml.

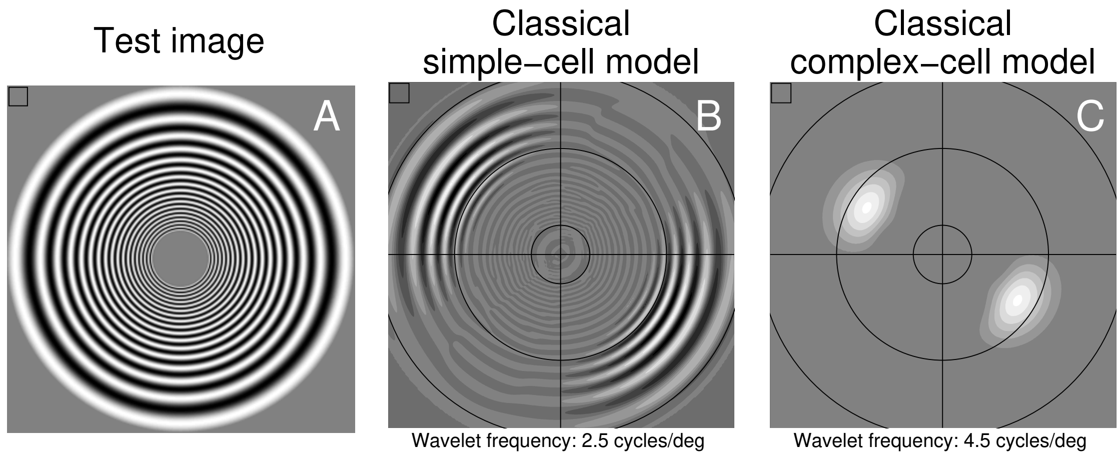

Figure 3 Circular test- and response images for the classical simple and complex cells models. This figure illustrates the technique of response images. The images are created by applying the classical models to all locations of the test image (we used the two wavelets shown in Figure 2 A for the simple and the complex cell model, respectively). The output is normalized such that white indicates the maximal response and dark gray values indicate inhibition. The square in the upper left corner indicates the size of the windows used as input to the unit. The three circular guidelines in the response images coincide with the internal, the middle, and the external circumference of the test image. **(A)** Test image used to investigate the response of a unit to a range of frequencies and orientations. **(B,C)** Response images for the classical models of simple and complex cells when applied to the test image shown in A. Simple cells respond to oriented bars and are sensitive to the exact position of edges in their receptive field, which causes the oscillations in B. Complex cells also respond to oriented edges but are insensitive to their position, and thus there are no oscillations in C. In contrast with our model, there is no inhibitory region in C and therefore no possibility to modify the orientation- or frequency-tuning of the complex cell. *A color version of this figure can be found at the end of the paper.*

different responses indicate selectivity to the direction of motion.

4 Results

For each simulation we extracted the 100 slowest units and analyzed their behavior. We now describe units computed in a single simulation and make a comparison with corresponding cells reported in the literature[4]. Although we concentrate for illustrative purposes on response images, all observations have been confirmed by independent measures, such as linear sinus gratings.

Gabor-like optimal stimuli and phase invariance: The optimal stimuli of almost all units look like Gabor wavelets, in agreement with physiological data. This means that the units respond best to edge-like stimuli. Figure 1 shows an overview of the optimal stimuli of the first 41 units of the simulation. The response of all these units is largely invariant to phase shift, as illustrated by the lack of oscillations in the response images (Figure 4), and would thus be classified as complex cells in a physiological experiment. 98 units out of the 100 studied in each simulation had Gabor-like S^+ and phase-shift invariance, the missing two cells are described under the paragraph *Tonic cells*.

Active inhibition: In V1 the tuning to orientation and frequency is shaped by active inhibition and other non-linear interactions (Sillito, 1975; De Valois et al., 1982) and is not simply a consequence of the shape of S^+ like in the classical model of complex cells.

[4] Further examples will be made available via the author's homepages.

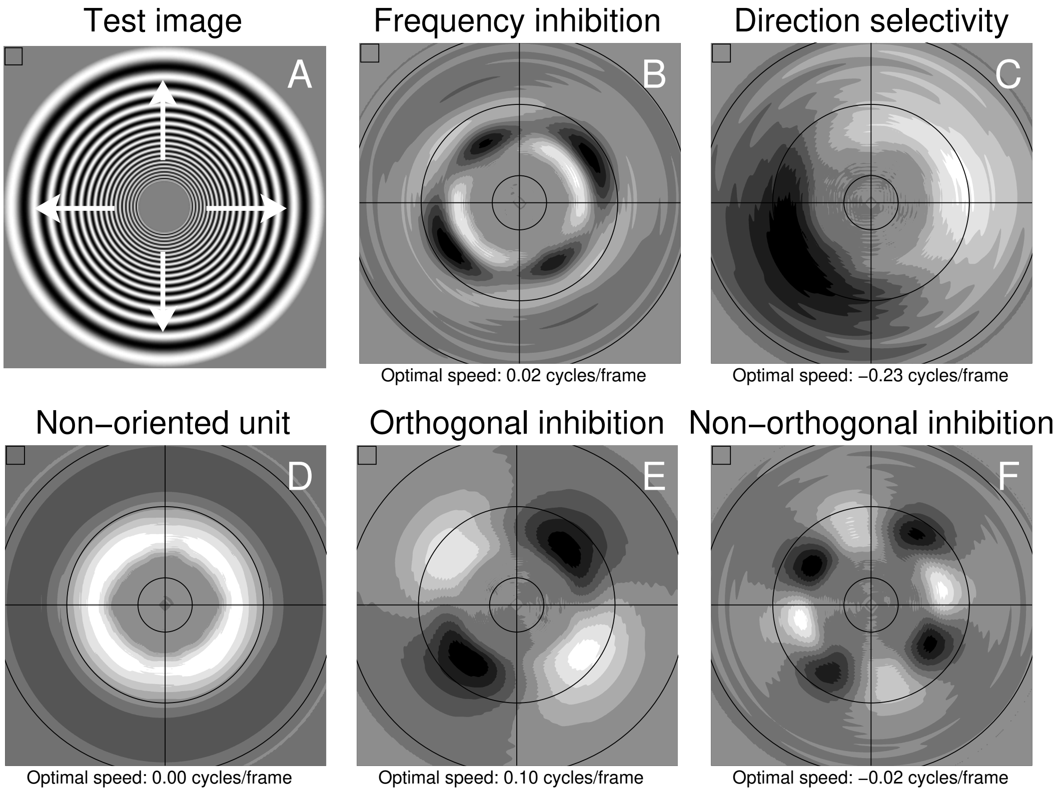

Figure 4 Circular test- and response images for some learned units. **(A)** Test image at time t used to investigate the response of a unit to a range of frequencies and orientations. A second test image at time $t + \mathrm{d}t$ is also used similar to the first one but with the rings shifted outwards by the preferred speed of the unit. **(B–F)** Response images of representative units. The response images are computed and normalized as described in Figure 3. The labels over the images indicate the illustrated behavior. **(B)** The strong inhibitory flanks at an orientation equal to the preferred one but at a slightly lower frequency than the optimal makes this unit very frequency-selective. **(C)** Different responses at opposite points of the rings, where the gratings move in two opposite directions, indicate direction selectivity. **(D)** This unit is phase-shift invariant (since there are no oscillations in the response image) and selective to relatively high frequencies, but it responds rather uniformly to all orientations. **(E)** This unit show inhibition at an orientation orthogonal to the preferred one. **(F)** In this unit the orientation of maximal inhibition is near to that of maximal excitation. This causes a second peak of activity to appear at the orientation orthogonal to the optimal one (secondary response lobe). The optimal stimuli of the unit shown in C,E,F are shown in Figure 2 D,B,C, respectively. *A color version of this figure can be found at the end of the paper.*

In our model inhibition is present in most units and typically makes an important contribution to the output. As a consequence, S^- is usually well-structured and has the form of a Gabor wavelet as well (Figure 1, Figure 2 B-F). Its orientation (which corresponds to the orientation of maximal inhibition) is often not orthogonal to that of S^+ (Figure 2 C), like also found in V1 (De Valois et al., 1982). This results in sharpened orientation tuning. When the orientation of S^- is very close to that of S^+, it is possible to observe a second peak of activity at an orientation orthogonal to the optimal one, known in the experimental literature as *secondary response lobe* (De Valois et al., 1982) (Figure 4 F).

On the other hand, we also find units that would be classified as complex by their response to edges and their phase-invariance but which are not selective for a particular orientation (Figure 4 D). Similar cells are present in V1 and known as *non-oriented cells* (De Valois et al., 1982). Analogous considerations also hold for the frequency domain: here again, the tuning varies from a sustained response to a

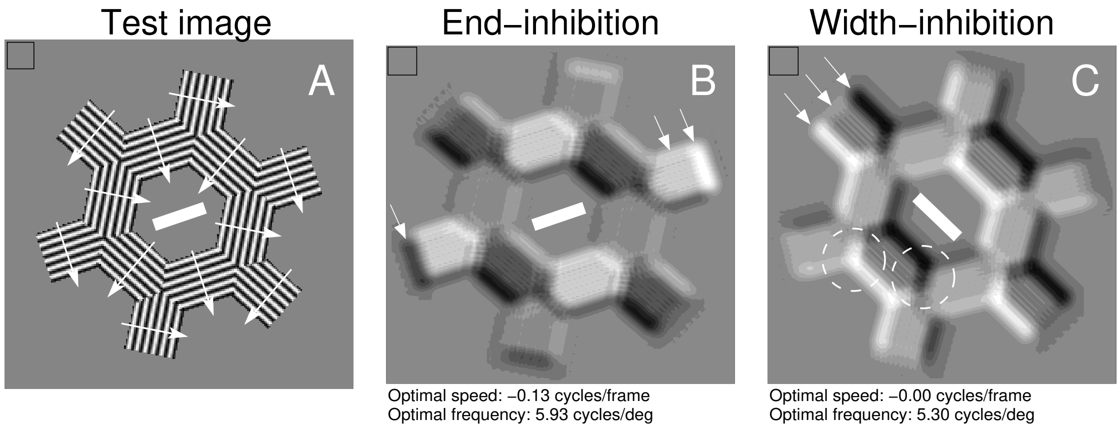

Figure 5 Hexagonal test- and response images for some learned units. (A) Test image used to investigate length- and width-inhibition. The hexagonal shape is oriented such that it is aligned to the preferred orientation of the considered unit, indicated by the central bar. The gratings are tuned to the preferred frequency and move with the preferred speed of the unit, in the direction indicated by the thin arrows. The response images are computed and normalized as described in Figure 3. **(B)** Response image of an end-inhibited unit. Going from right to left, the first arrow indicates the point of maximal response, when only the left part of the input window is filled. In the region indicated by the second arrow the grating fills the whole input window; the response has decreased, which corresponds to end-inhibition. The third arrow indicates the region where only the right part of the input window is filled, and the unit is inhibited. **(C)** Response image of a width-inhibited unit. The interpretation of the arrows going from left to right is similar to that in B, except that the grating extends laterally, which corresponds to width-inhibition. The two dotted circles surround two regions with opposite curvature. The unit responds strongly in one case and is inhibited in the other, which indicates curvature selectivity. The optimal stimuli of the unit shown in B,C are shown in Figure 2 E,F, respectively. *A color version of this figure can be found at the end of the paper.*

large range of frequencies (Figure 4 E) to a sharp (and sometimes even bimodal) tuning due to active inhibition (Figure 4 B).

End- and Side-Inhibition: Some of the complex cells in V1 are selective to the length (*end-inhibited* cells) or width of their input (*side-inhibited* cells). While in normal cells the extension of a grating at the preferred orientation and frequency produces an increase of the response up to a saturation level, in these cells the response drops if the grating extends beyond a certain limit (DeAngelis et al., 1994).

End- and side-inhibited units form approx. 1/6 of the units in our results. They can sometimes be identified by looking directly at the optimal stimuli: in these cases S^+ fills only one half of the window, while the missing half is covered by S^- with the same orientation and frequency (Figure 2 E,F). In this way, if we extend the stimulus into the missing half, the output receives an inhibitory contribution and drops.

To investigate this behavior in greater detail, we constructed hexagonal-shaped test images like the one shown in Figure 5 A. The hexagons are adapted to each unit to have its optimal orientation and frequency. If a unit is selective for the length or width of the input, one notices a higher response at the borders of the hexagon, where only part of the input window is filled, than in the inner, where the whole input window is filled. Figures 5 B,C show the response images of an end- and a side-inhibited unit.

A secondary characteristic of end-inhibited cells in V1 is that they are sometimes selective for different signs of curvature (Dobbins et al., 1987; Versavel et al., 1990). This can be observed in our simulation, for example in Figure 5 C: the dashed circles indicate two opposite curvatures. While in one the unit responds strongly, in the other one it is inhibited.

Direction selectivity: Complex cells in V1 are sensitive to the motion of the presented stimuli. Some of them respond to motion in both directions while others are direction selective (Hubel and Wiesel, 1962; Schiller et al., 1976; De Valois et al., 1982). Similarly, in our model some units are strongly selective for direction (Figure 4 C) while others are neutral (e.g. Figure 4 E). In the latter case the optimal speed may be non-zero but the response is nearly equal for both directions. In a few cases, the two wavelets in the optimal stimuli have a slightly different orientation or frequency at time t and at time $t + \mathrm{d}t$, which indicates a more complex behavior in time, such as selectivity for rotation or zoom.

Tonic cells: The first unit in every simulation codes for the mean pixel intensity, and is thus comparable to the *tonic cells* found in V1 (Schiller et al., 1976). Tonic cells respond to one sign of light but do not need a contour in order to respond. We find in addition one unit (the second one) that responds to the squared mean pixel intensity, and thus to both signs of light.

Response to complex stimuli: We also find some units that give a near-optimal output ($> 80\%$ of the optimum) in response to corners and T-shaped stimuli, related to the V1-cells described in (Shevelev, 1998). Other units respond to one sign of curvature only. These behaviors are often associated with length- or width-inhibition, as described above.

Relations between slowness and behavior: Although the precise shape and order of the units can vary in different simulations, there seem to be relations between the slowness of unit responses and their receptive field properties. The slowest units are usually less selective for orientation and frequency, have orthogonal inhibition and their preferred speed is near zero. Units with non-orthogonal inhibition, direction selectivity, and length- or width-inhibition predominate in a faster regime. Moreover, frequency tends to increase with decreasing slowness. We couldn't find any relation between preferred orientation and slowness. It would be interesting to see if similar relations can also be found experimentally.

5 Conclusions

Several theoretical studies have successfully reproduced the basic properties of simple cells (Olshausen and Field, 1996; Bell and Sejnowski, 1997; Hoyer and Hyvärinnen, 2000; Szatmáry and Lõrincz, 2001; Olshausen, 2002; Einhäuser et al., 2002; Hurri and Hyvärinnen, 2002) or complex cells (Hyvärinnen and Hoyer, 2000; Kayser et al., 2001) in models based on the computational principles *sparseness* (Olshausen and Field, 1996; Olshausen, 2002), *statistical independence* (Bell and Sejnowski, 1997; Hoyer and Hyvärinnen, 2000; Hyvärinnen and Hoyer, 2000; Szatmáry and Lõrincz, 2001), or *slowness* (Kayser et al., 2001; Einhäuser et al., 2002; Hurri and Hyvärinnen, 2002). Among the simple cells models, two included direction selectivity (Olshausen, 2002; Szatmáry and Lõrincz, 2001), two color-selectivity (Hoyer and Hyvärinnen, 2000; Einhäuser et al., 2002) and one disparity (Hoyer and Hyvärinnen, 2000). Each of these models focused on one particular aspect of the behavior of cells in V1. In particular, the complex cells resulting from the two cited models only reproduced phase shift invariance and learned units that were equivalent to the classical models. To the best of our knowledge, the model presented here is the first one that is able to reproduce complex cells that are in addition direction-selective, end-inhibited, width-inhibited, curvature selective etc.

One important limitation of the previous models is that they assumed linear or non-linear but simple neural network architectures spanning a function space much smaller than the one we use. In particular, none of the non-linear models included inhibition, while many of the illustrated complex cell behaviors are impossible to obtain without it.

We would like to point out that the definitions of *slowness* in the models described in (Kayser et al., 2001; Einhäuser et al., 2002; Hurri and Hyvärinnen, 2002) and in the present work are to some extent incompatible with each other. In (Kayser et al., 2001), a decorrelation and a slowness term are optimized through gradient descent. The slowness term is defined by the *mean* of the Δ-values in eq. (1). The units found by this rule lie in the subspace of the most slowly varying functions, but they are unique only up to an orthogonal transformation, i.e. by mixing the results through a rotation one would find different but equally optimal units (Blaschke 2003, personal communication). In (Einhäuser et al., 2002; Hurri and Hyvärinnen, 2002), the temporal variation of the output is minimized after a rectifying non-linearity. When the non-linearity is symmetrical (for example when squaring or taking the absolute value of the output), solutions oscillating strongly around zero can be optimal, since the sign is canceled out by the rectification. Also in the non-symmetric case, the solutions found are different from the ones extracted by SFA. Further investigation is needed in order to compare the different definitions and to find a unifying framework.

In summary we have shown that slowness leads to a great variety of complex cell properties found also in physiological experiments. Our results demonstrate that such a rich repertoire of receptive field properties can be accounted for by a single unsupervised learning principle. It will be interesting to see to what extent theory and experiments agree also quantitatively, e.g. with respect to the shape of tuning curves, and whether the relations between the behavior of the neurons and the slowness of their output will be confirmed experimentally. Earlier modeling studies with SFA (Wiskott and Sejnowski, 2002) have shown that translation, scale, and other invariances can also be learned for whole objects in a hierarchical network. This suggests that slowness might be a rather general learning principle in the visual and possibly also other perceptual systems.

6 Acknowledgments

Supported by a grant from the Volkswagen Foundation. We thank Andreas Herz, Tim Gollisch, and Martin Stemmler for useful comments on the manuscript.

References

Bell, A. J., Sejnowski, T. J., 1997. The 'independent components' of natural scenes are edge filters. Vision Research 37 (23), 3327–3338. 9

Berkes, P., Wiskott, L., 2002. Applying slow feature analysis to image sequences yields a rich repertoire of complex cell properties. In: Dorronsoro, J. R. (Ed.), Artificial Neural Networks - ICANN 2002 Proceedings. Lecture Notes in Computer Science. Springer, pp. 81–86. 5

De Valois, R., Yund, E., Hepler, N., 1982. The orientation and direction selectivity of cells in macaque visual cortex. Vision Res. 22 (5), 531–44. 6, 7, 9

DeAngelis, G., Freeman, R., Ohzawa, I., 1994. Length and width tuning of neurons in the cat's primary visual cortex. Journal of Neurophysiology 71 (1), 347–74. 8

Dobbins, A., Zucker, S. W., Cynader, M. S., October 1987. Endstopped neurons in the visual cortex as a substrate for calculating curvature. Nature 329, 438–441. 8

Einhäuser, W., Kayser, C., Körding, K., König, P., 2002. Learning multiple feature representation from natural image sequences. In: Dorronsoro, J. R. (Ed.), Artificial Neural Networks - ICANN 2002 Proceedings. Lecture Notes in Computer Science. Springer, pp. 21–26. 9

Földiák, P., 1991. Learning invariance from transformation sequences. Neural Computation 3, 194–200. 1

Gantmacher, 1959. Matrix Theory. Vol. 1. AMS Chelsea Publishing. 3

Hinton, G., 1989. Connectionist learning procedures. Artificial Intelligence 40, 185–234. 1

Hoyer, P., Hyvärinnen, A., 2000. Independent component analysis applied to feature extraction from colour and stereo images. Network: Computation in Neural Systems 11 (3), 191–210. 9

Hubel, D., Wiesel, T., 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. Journal of Physiology 160, 106–154. 1, 9

Hurri, J., Hyvärinnen, A., 2002. Simple-cell-like receptive fields maximize temporal coherence in natural video. In press. 9

Hyvärinnen, A., Hoyer, P., 2000. Emergence of phase and shift invariant features by decomposition of natural images into independent features subspaces. Neural Computation 12 (7), 1705–1720. 9

Kayser, C., Einhäuser, W., Dümmer, O., König, P., Körding, K., 2001. Extracting slow subspaces from natural videos leads to complex cells. In: Artificial Neural Networks - ICANN 2001 Proceedings. Springer, pp. 1075–1080. 9

Mitchison, G., 1991. Removing time variation with the anti-Hebbian differential synapse. Neural Computation 3, 312–320. 1

Olshausen, B., 2002. Sparse codes and spikes. In: Rao, R. P. N., Olshausen, B. A., Lewicki, M. S. (Eds.), Probabilistic Models of the Brain: Perception and Neural Function. MIT Press. 9

Olshausen, B., Field, D., Jun 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381 (6583), 607–609. 9

Schiller, P., Finlay, B., Volman, S., 1976. Quantitative studies of single-cell properties in monkey striate cortex. I. Spatiotemporal organization of receptive fields. J. Neurophysiol. 39 (6), 1288–1319. 9

Shevelev, I. A., 1998. Second-order features extraction in the cat visual cortex: selective and invariant sensitivity of neurons to the shape and orientation of crosses and corners. BioSystems 48, 195–204. 9

Sillito, A., 1975. The contribution of inhibitory mechanisms to the receptive field properties of neurons in the striate cortex of the cat. J. Physiol. 250, 305–329. 6

Stone, J. V., 1996. Learning perceptually salient visual parameters using spatiotemporal smoothness constraints. Neural Computation 8, 1463–1492. 1

Stone, J. V., 2001. Blind source separation using temporal predictability. Neural Computation 13, 1559–1574. 1

Szatmáry, B., Lõrincz, A., 2001. Independent component analysis of temporal sequences subject to constraints by lateral geniculate nucleus inputs yields all three major cell types of the primary visual cortex. Journal of Computational Neuroscience 11, 241–248. 9

van Hateren, J., van der Schaaf, A., 1998. Independent component filters of natural images compared with simple cells in primary visual cortex. Proc. R. Soc. Lond. B 265, 359–366. 3

Versavel, M., Orban, G. A., Lagae, L., 1990. Responses of visual cortical neurons to curved stimuli and chevrons. Vision Res. 30 (2), 235–248. 8

Wiskott, L., 1998. Learning invariance manifolds. In: Niklasson, L., Bodén, M., Ziemke, T. (Eds.), Proc. Intl. Conf. on Artificial Neural Networks, ICANN'98, Skövde. Perspectives in Neural Computing. Springer, pp. 555–560. 2

Wiskott, L., Sejnowski, T., 2002. Slow feature analysis: Unsupervised learning of invariances. Neural Computation 14 (4), 715–770. 1, 2, 3, 10

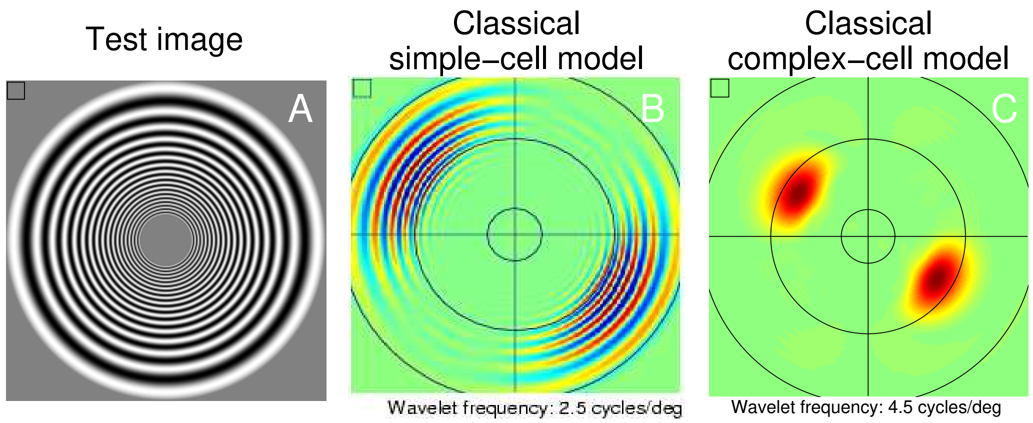

Figure 3 *(color version, see page 6 for figure caption).*

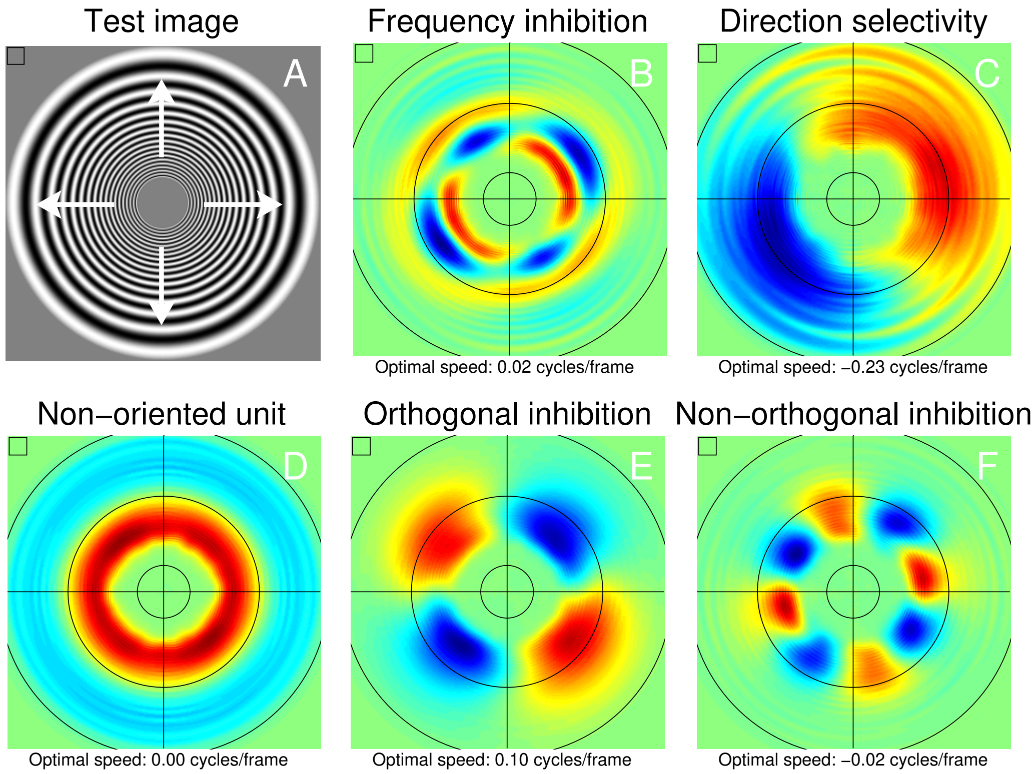

Figure 4 *(color version, see page 7 for figure caption).*

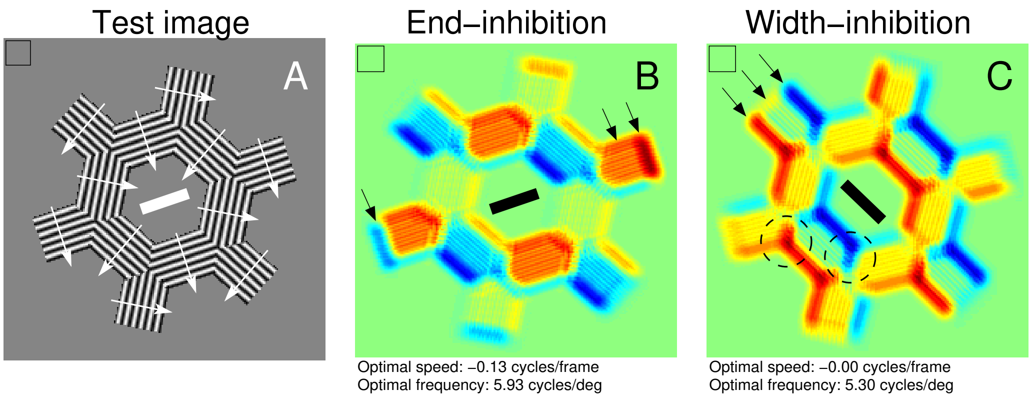

Figure 5 *(color version, see page 8 for figure caption).*