Local Interactions in Neural Networks Explain Global Effects in Gestalt Processing and Masking

Michael H. Herzog
michael.herzog@uni-bremen.de
Human Neurobiology, University of Bremen, D-28211 Bremen, Germany

Udo A. Ernst
udo@neuro.uni-bremen.de

Axel Etzold
aetzold@physik.uni-bremen.de

Christian W. Eurich
eurich@physik.uni-bremen.de
Institute of Theoretical Neurophysics, University of Bremen, D-28359 Bremen, Germany

One of the fundamental and puzzling questions in vision research is how objects are segmented from their backgrounds and how object formation evolves in time. The recently discovered shine-through effect allows one to study object segmentation and object formation of a masked target depending on the spatiotemporal Gestalt of the masking stimulus (Herzog & Koch, 2001). In the shine-through effect, a vernier (two abutting lines) precedes a grating for a very short time. For small gratings, the vernier remains invisible while it regains visibility as a shine-through element for extended and homogeneous gratings. However, even subtle deviations from the homogeneity of the grating diminish or even abolish shine-through. At first glance, these results suggest that explanations of these effects have to rely on high-level Gestalt terminology such as homogeneity rather than on low-level properties such as luminance (Herzog, Fahle, & Koch, 2001). Here, we show that a simple neural network model of the Wilson-Cowan type qualitatively and quantitatively explains the basic effects in the shine-through paradigm, although the model does not contain any explicit, global Gestalt processing. Visibility of the target vernier corresponds to transient activation of neural populations resulting from the dynamics of local lateral interactions of excitatory and inhibitory layers of neural populations.

1 Introduction

How objects emerge from their backgrounds as independent entities is one of the tenacious problems in computer vision as well as in the cognitive

and neurosciences. Two fundamental questions arise when studying the emergence of an object: How is a particular object spatially segmented from other parts of an image, and how does object formation evolve in time? Psychophysical investigations of the segmentation of objects concentrate mainly on static stimuli, neglecting the dynamics of the processing (e.g., Julesz, 1981; Beck, 1983; Nothdurft, 1994; Bergen & Adelson, 1988; Sagi, 1995). Studies of the time course of object formation usually use masked stimuli, often ignoring the global aspects of the mask important for object segmentation (e.g., Breitmeyer, 1984).

The shine-through paradigm combines both approaches by backward-masking a target with spatiotemporally varying gratings. The strength of masking is determined by both the exact timing of the vernier and the global Gestalt of the grating rather than by its local properties (Herzog & Koch, 2001; Herzog, Fahle, & Koch, 2001).

In detail, shine-through occurs only for extended and homogeneous gratings (see Figure 1a). For small or spatially inhomogeneous gratings, visibility strongly diminishes (see Figures 1b and 1c; Herzog, Fahle, et al., 2001). Detailed investigations reveal that the temporal homogeneity of the grating

![Figure 1: Spatial conditions. (a) A vernier precedes a grating comprising 25 elements (only 19 are shown in this and the following illustrations). In this condition, the vernier emerges as a transient flash superimposed on the grating appearing wider, brighter, and even longer than the vernier really is, called the shine-through element. Performance is determined by the threshold for which subjects reach 75% correct responses in discriminating the spatial offset direction of the vernier (right versus left). Standard errors of thresholds are shown in parentheses. (b) If the grating contains five elements, the vernier remains largely invisible. Thresholds increase dramatically—that is, performance deteriorates compared to condition a. (c) Inserting gaps in the grating diminishes the shine-through effect. As with the homogeneous grating with 5 elements, thresholds are strongly degraded compared to condition a.](http://www.mitpressjournals.org/doi/pdf/10.1162/089976603322297304)
is also a prerequisite for shine-through to occur (Herzog, Koch, & Fahle, 2001a). These subtle dependencies on the spatial and temporal layout of the grating suggest that the visual system is on the brink of its spatiotemporal processing characteristics. Hence, the shine-through effect allows one to investigate the time course of object formation and segmentation depending on global spatial aspects.

The shine-through element appears perceptually as a flash superimposed on the grating, suggesting that shine-through emerges in a dynamical fashion. When studying the underlying neuronal mechanisms, it is therefore reasonable to consider transients of the neural dynamics rather than steady states or other attractor properties of networks. We show that a commonly used Wilson-Cowan type of model (Ben-Yishai, Bar-Or, & Sompolinsky, 1995; Li & Dayan, 1999; Li, 1999b, 2001, 2002; Ernst, Pawelzik, Sahar-Pikelny, & Tsodyks, 2001; see Ernst & Eurich, 2002, for a review), capable of reproducing the basic results in the shine-through paradigm, indeed expresses shine-through as a transient in its activation dynamics, as suggested by physiological and psychophysical investigations of masked stimuli (Macknik & Livingstone, 1998).

The Wilson-Cowan model employs only one excitatory and one inhibitory layer of neural populations with lateral connections. Inputs to the network consist of the time-dependent stimulus intensities along the horizontal stimulus direction. The two-dimensional geometry of the vernier and the gratings is neglected since fundamental principles of information processing can be studied more exactly with a minimal model. Although in this model no Gestalt processing is explicitly implemented, still, the subtle dependencies of vernier visibility on the spatiotemporal layout of the masking grating can be explained in a unifying framework.

2 Materials and Methods

2.1 Psychophysical Experiments. In the psychophysical experiments, a vernier, i.e. two abutting lines horizontally displaced by an offset $d$, was displayed on an analog monitor for a short time (Herzog et al., 2001a; Herzog, Fahle, et al., 2001). The vernier was subsequently masked by one of a set of grating stimuli. Depending on the type of grating, visibility of the vernier varies strongly. Subjects were asked to indicate the direction of the offset of the vernier. In the psychophysical experiments, we determined the spatial displacement $d$ of the vernier segments for which an observer reached 75% correct responses. This value $d$ is called the threshold.

2.1.1 Spatial Conditions. Figure 1 shows the stimuli used in the psychophysical experiments. Gratings were regularly spaced, comprising either 25 elements, 5 elements, or 25 elements containing gaps. The spacing between the grating elements was $200''$ (arc sec). (For a detailed description
of the methods and materials used in the psychophysical experiments, see Herzog, Fahle, et al., 2001.\(^1\)

Shine-through occurs only for extended and homogeneous gratings (see Figure 1a). For small gratings, the vernier is completely masked and therefore almost invisible (see Figure 1b). When gaps are inserted to both sides of the central part of the grating, the vernier is strongly masked (gap grating; see Figure 1c). In general, if an extended but inhomogeneous grating follows the vernier, visibility strongly diminishes (Herzog, Fahle, et al., 2001).

2.1.2 Temporal Conditions. In the conditions investigating the timing of object formation, we varied the timing of some of the mask parts. The timing is characterized by a parameter \(\tau\); for a detailed description of the timing in the four stimulus configurations, see the caption of Figure 2. The different stimuli employed can be summarized as follows:

- In the condition shown in Figure 2a, a 25-element grating with gaps followed immediately after the vernier. At the locations of the gaps, two grating elements were added before \((\tau < 0)\), simultaneously \((\tau = 0)\) or after \((\tau > 0)\) the other grating elements.

- In the condition displayed in Figure 2b, a 5-element grating followed the vernier. Ten peripheral context elements on each side of the 5-element grating were added before \((\tau < 0)\), simultaneously \((\tau = 0)\), or after \((\tau > 0)\) the 5-element grating.

- In the condition depicted in Figure 2c, a homogeneous 25-element grating followed the vernier. At time \(\tau > 0\), two elements were removed, leaving a 25-element grating with gaps (for details, see Herzog et al., 2001a).

- In the condition shown in Figure 2d, a homogeneous 25-element grating followed the vernier. At time \(\tau > 0\), the 2*10 peripheral elements were removed, leaving only a 5-element grating (for details, see Herzog et al., 2001a).

In all conditions, performance deteriorates the more the grating elements deviate from being presented simultaneously at \(\tau = 0\). In the conditions shown in Figures 2c and 2d, performance deteriorates the stronger the earlier some of the grating elements are removed from the homogeneous 25-element mask. (The empirical results are presented together with the modeling results in Figure 6.)

2.1.3 Sensitive Measure. Vernier acuity, instead of vernier detection itself, was used as the sensitive measure to avoid floor and ceiling effects.

\(^1\) For an animation of the stimuli, see http://www-neuro.uni-bremen.de/~vernier/vernier_english/vindex.html.
Local Interactions Explain Global Effects

Figure 2: Temporal conditions. In all four conditions (a–d), the masks consisted of two different spatial parts that add up to the standard, homogeneous 25-element grating. The first part of a mask was displayed immediately after the vernier offset and always lasted for 300 ms. The timing of the second part of the mask was variable and is characterized by a time parameter $\tau$ whose usage differs among the conditions. $\tau$ is defined relative to the temporal offset of the vernier. (a) The first part of the mask was a 25-element grating with gaps. The second part consisted of the elements filling the gaps, which were switched on at time $\tau$. For $\tau > 0$, the gap elements completed the gap grating to a 25-element grating after a delay $\tau$ (upper panel). For $\tau < 0$, the gap elements were already switched on at time $\tau$ prior to vernier offset (lower panel). (b) The first part of the mask was a 5-element grating. The second part consisted of 2*10 peripheral elements, which were switched on at time $\tau$. For $\tau > 0$, the peripheral elements completed the 5-element grating to a 25-element grating after a delay $\tau$ (upper panel). For $\tau < 0$, the peripheral elements were switched on at time $\tau$ prior to vernier offset (lower panel). (c) The first part of the mask was a 25-element grating with gaps. The second part consisted of the elements filling the gaps, presented immediately after vernier offset and removed at time $\tau$. (d) The first part of the mask was a 5-element grating. The second part consisted of the 2*10 peripheral elements completing a 5-element grating to a 25-element grating, presented immediately after vernier offset and removed at time $\tau$. The expressions 20 ms, $\tau$, and 300 ms-$\tau$ on the left denote the duration of each mask part.
Floor and ceiling effects would occur for 2AFC detection tasks since performance varies over a large range from no to clear visibility of the vernier in the various conditions. In this sense, vernier acuity is used only as a carrier. However, since performance is determined mainly by the visibility of the vernier (Herzog, Koch, & Fahle, 2001b), we focus only on the strength of the visibility of the vernier in the simulations. Even if we ignore the spatial offset of the vernier in the modeling, we will use the term vernier for the preceding target element in the model.

2.2 Structure of the Model. A neural network model was employed to account for the visibility of the shine-through element in the various conditions shown in Figures 1 and 2. The model consists of two one-dimensional continuous layers, one of which is excitatory and the other one inhibitory. The spatial dimension of the layers is described by the variable $x$, which represents the horizontal dimension of the stimulus (see Figure 3). The mathematical description assumes a large population of neurons at each position $x$ in both layers, justified by the huge number of neurons found in even small patches of cortical tissue. In the following, variables and functions of the model for excitatory and inhibitory populations are designated by a subscript $\{. . .\}_{\text{e}, \text{i}}$ respectively.

The dynamics of the system are given by partial differential equations often used in visual processing (Wilson & Cowan, 1973). The equations describe the firing rates $A_{\text{e}, \text{i}}(x, t)$ of the populations at position $x$ and time $t$:

$$
\tau_{\text{e}} \frac{\partial A_{\text{e}}(x, t)}{\partial t} = -A_{\text{e}}(x, t) + h_{\text{e}}\{w_{\text{ee}}(A_{\text{e}} \ast W_{\text{e}})(x, t) + w_{\text{ie}}(A_{\text{i}} \ast W_{\text{i}})(x, t) + I(x, t)\} \tag{2.1}
$$

$$
\tau_{\text{i}} \frac{\partial A_{\text{i}}(x, t)}{\partial t} = -A_{\text{i}}(x, t) + h_{\text{i}}\{w_{\text{ei}}(A_{\text{e}} \ast W_{\text{e}})(x, t) + w_{\text{ii}}(A_{\text{i}} \ast W_{\text{i}})(x, t) + I(x, t)\}. \tag{2.2}
$$

$\tau_{\text{e}, \text{i}}$ are the membrane time constants of the excitatory and the inhibitory populations, respectively. The terms $-A_{\text{e}}(x, t)$ and $-A_{\text{i}}(x, t)$ in equations 2.1 and 2.2, respectively, are decay terms yielding an exponential decay of network activity in the case of absent external input. $h_{\text{e}, \text{i}}$ denote nonlinear transfer functions, which map the total synaptic input onto the firing rate of the excitatory or inhibitory population at position $x$ and time $t$, respectively. We assume the simplest case: the transfer functions are piecewise linear, approximating the response behavior of typical cortical neurons (see Figure 3 inset; Yu & Lewis, 1989; Eurich, Pawelzik, Ernst, Cowan, & Milton, 1999). Note that we omitted a saturation regime for the firing frequencies, thereby reducing the number of system parameters: neurons in vivo hardly oper-
Local Interactions Explain Global Effects

Figure 3: Overview of the network model. Excitatory and inhibitory neural layers are indicated by elongated filled and open ellipses, respectively. Interactions within a layer (lateral interactions) and between layers (mutual interactions) and the external, afferent input that feeds in both layers are symbolized by arrows. Small inset figures show the kernel functions $W_e(\Delta x)$, $W_i(\Delta x)$, and $V(\Delta x)$. (Inset at top right) Piecewise linear transfer functions mapping the total synaptic inputs $J_e$ and $J_i$ for excitatory and inhibitory populations onto firing rates $h_e$ and $h_i$, respectively. Parameters $\theta_e$ and $\theta_i$ are the activation thresholds, and $s_e$, $s_i$ are the slopes of the transfer functions. The values of the various parameters are given in Table 1.

The neural populations have both lateral and mutual connections. For the excitatory neurons, interactions are characterized by a kernel function $W_e(x - x')$, which specifies the connectivity as a function of distance $x - x'$ between populations at positions $x$ and $x'$ for the lateral connections within the excitatory layer and for the excitatory input to the inhibitory population. Likewise, $W_i(x - x')$ is the interaction kernel of the inhibitory populations for the lateral connections within the inhibitory layer and for the inhibitory
input to the excitatory layer. Interaction kernels are described by normalized gaussians, as shown in the insets of Figure 3,

\[ W_{e,i}(x - x') = \frac{1}{\sqrt{2\pi \sigma_{e,i}^2}} \exp \left( -\frac{(x - x')^2}{2\sigma_{e,i}^2} \right), \tag{2.3} \]

with widths \( \sigma_{e,i} \), which introduce a specific length scale for the couplings. This simplifying choice of only two kernel functions was motivated by the observation that cortical interactions may have identical dependence on cortical distance irrespective of, for example, the type of the target neuron (Peters & Yilmaz, 1993; White, 1989). The change of the synaptic input to a population at position \( x \) at time \( t \) due to an interaction is given by a convolution (denoted by a star) of the kernel function with the actual firing frequencies of the respective populations; the resulting value is multiplied by synaptic weights \( w_{ee} \geq 0, w_{ei} \geq 0, w_{ii} \leq 0, \) and \( w_{ie} \leq 0 \) in the case of an excitatory \( \rightarrow \) excitatory, excitatory \( \rightarrow \) inhibitory, inhibitory \( \rightarrow \) inhibitory, and inhibitory \( \rightarrow \) excitatory interaction, respectively. For example, the excitatory \( \rightarrow \) excitatory interaction is given by

\[ w_{ee} (A_e \star W_e) (x, t) = w_{ee} \int_{-\infty}^{\infty} A_e(x', t) W_e(x - x') dx'. \tag{2.4} \]

In addition to the input from lateral and mutual connections, both excitatory and inhibitory populations at position \( x \) receive afferent input \( I(x, t) \) given by the convolution

\[ I(x, t) = (S \star V)(x, t) = \int_{-\infty}^{\infty} S(x', t) V(x - x') dx'. \tag{2.5} \]

\( S(x', t) \) denotes the spatiotemporal input corresponding to the different stimulus conditions described in section 2.1. \( S(x', t) \) takes the value 1 if it is part of the vernier or a grating element and 0 otherwise. For example, \( S(0, t) \) describes the luminance of the vernier presented at \( x = 0 \) for 20 ms followed by the central element of the 25-element grating, which is presented for 300 ms. Since we deal with the horizontal spatial coordinate only, the spatial offset of the vernier is not considered. \( V(x - x') \) in equation 2.5 is a filter function quantifying the influence of the input \( S(x', t) \) on a neuron at position \( x \). The filter is modeled as a Mexican hat function (i.e., a difference of gaussians) whose integral vanishes,

\[ V(x - x') = \frac{1}{\sqrt{2\pi \sigma_E^2}} \exp \left( -\frac{(x - x')^2}{2\sigma_E^2} \right) - \frac{1}{\sqrt{2\pi \sigma_I^2}} \exp \left( -\frac{(x - x')^2}{2\sigma_I^2} \right). \tag{2.6} \]
The parameters $\sigma_E$ and $\sigma_I$ denote the widths of the excitatory and the inhibitory parts of the kernel function, respectively (see the corresponding inset of Figure 3 for a sketch of the filter function).

2.3 Model Parameters and Implementation

2.3.1 Length, Time, and Intensity. The spatial coordinate $x$ of the neural layers is expressed in units of horizontal visual angle. For example, the spacing of the grating elements in the model is identified with $\Delta = 200''$ (arc sec), as in the experimental setup. The timescale was fixed by the presentation time of the vernier, which was chosen to be 20 ms. Input intensities and neural activities have arbitrary units. An intensity scale is introduced, as described above, by assigning the value $S(x', t) = 1$ if $x'$ is part of the vernier or a mask element and 0 otherwise.

2.3.2 Model Parameters. Model parameters were adjusted according to the following methods: (1) In order to yield a simple model whose parameter ranges can be explored in detail, the number of free variables was reduced by exploiting symmetries in the network, obeying anatomical and physiological facts. (2) Relationships between parameters were established using anatomical and physiological constraints. (3) Two experimental conditions were used to adjust the remaining parameters: the vernier followed by a grating of 25 (see Figure 1a) or 5 elements (see Figure 1b). Parameters as used are listed in Table 1.

1. The following symmetries are assumed in the network: Both interaction kernels $W_e(\Delta x)$ and $W_i(\Delta x)$ given by equation 2.3 are supposed to be independent of the target population. Therefore, only two parameters are associated with the interaction kernels of the lateral connections and the connections between the layers: the length scales $\sigma_e$ and $\sigma_i$. The inter-

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Meaning</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_e$</td>
<td>Membrane time constant of excitatory neurons</td>
<td>16 ms</td>
</tr>
<tr>
<td>$\tau_i$</td>
<td>Membrane time constant of inhibitory neurons</td>
<td>4 ms</td>
</tr>
<tr>
<td>$s_e$</td>
<td>Slope of excitatory neuron transfer function</td>
<td>3</td>
</tr>
<tr>
<td>$s_i$</td>
<td>Slope of inhibitory neuron transfer function</td>
<td>5.4</td>
</tr>
<tr>
<td>$\theta_e, \theta_i$</td>
<td>Activation thresholds of excitatory/inhibitory populations</td>
<td>0</td>
</tr>
<tr>
<td>$\sigma_e$</td>
<td>Length scale of excitatory interaction kernel</td>
<td>150''</td>
</tr>
<tr>
<td>$\sigma_i$</td>
<td>Length scale of inhibitory interaction kernel</td>
<td>500''</td>
</tr>
<tr>
<td>$w_{ee}, w_{ei}, w_{ii}$</td>
<td>Interaction weights $e \rightarrow e, e \rightarrow i, i \rightarrow i$</td>
<td>1, 1, -1</td>
</tr>
<tr>
<td>$\sigma_E, \sigma_I$</td>
<td>Length scale of excitatory part of afferent kernel function</td>
<td>100'', 200''</td>
</tr>
<tr>
<td>$A_0$</td>
<td>Activity threshold level</td>
<td>0.08</td>
</tr>
</tbody>
</table>
action weights of the excitatory population, $w_{ee}$ and $w_{ei}$, are assumed to be identical. The same holds for the interaction weights of the inhibitory population, $w_{ie}$ and $w_{ii}$. An inspection of the dynamical equations, 2.1 and 2.2, and the transfer functions shown in Figure 3 shows that only the products $w_{ee}s_e$ (identical with $w_{ei}s_i$) and $w_{ie}s_i$ (identical with $w_{ii}s_i$) enter the dynamics. Therefore, we can set $w_{ee} = w_{ei} = 1$ and $w_{ie} = w_{ii} = -1$, and use $s_e$ and $s_i$ as nonredundant parameters.

The neural activation functions are supposed to have identical activation thresholds. The simplest choice is $\theta_e = \theta_i = 0$: the input signal is rectified having linear output for positive membrane potentials.

The kernel function $V(\Delta x)$ for the external input given by equation 2.6 was chosen to have a zero integral. For spatially uniform stimuli, this input filter yields a vanishing net excitation. Therefore, the important parameters of this filter are the length scales $\sigma_E$ and $\sigma_I$ only.

2. We employed the following constraints on the parameters, thereby reducing the relevant parameter space further. From cortical physiology, it is well known that inhibitory neurons can fire at higher rates than excitatory neurons (McCormick et al., 1985). This is reflected in the condition $s_i > s_e$ for the slopes of the neural activation functions. Furthermore, cortical models (Ben-Yishai et al., 1995; Ernst et al., 2001) suggest that typical dynamical properties of visual cortical neurons, such as orientation and direction tuning, may occur in the regime where intracortical interactions dominate over the influence of the efferent inputs (the so-called marginally stable regime; cf. Ben-Yishai et al., 1995; Ernst et al., 2001). In our model, this implies $s_{ie,ii} > 1$ for the slopes of the transfer functions. Finally, for the external input, the relationship $\sigma_E < \sigma_I$ was used to obtain a proper Mexican hat type of interaction, which resembles on-off receptive field properties of efferent lateral geniculate nucleus (LGN) neurons.

3. The parameters that remained to be adjusted using the above restrictions are $\tau_{ie,i}, \sigma_{ie,i}, s_{ie,i}, \sigma_E, \sigma_I$. Parameter values were selected by hand using the experimental results of the 25 and 5 element grating conditions, which yield good or no visibility of the shine-through element, respectively (see Figure 1).

The range of excitation in the network was set to a value of $\sigma_e = 150''$, which is below the spacing $\Delta = 200''$ of the grating elements. In contrast, the range of inhibition was set to a larger value of $\sigma_i = 500'' = 2.5 \Delta$. This choice leads to maximal inhibition of the vernier exactly when the masking grating consists of 5 elements. To amplify the effect of the inhibition in this case, $\tau_i$ was chosen to be smaller than $\tau_e$. For the external input, we choose $\sigma_E = 100''$ and $\sigma_I = 200''$ to yield enhanced activity at the edges of the masking gratings.

2.3.3 Model Implementation. The differential equations, 2.1 and 2.2, were numerically integrated using an embedded fifth-order Runge-Kutta scheme with Cash-Karp parameters (Press, Teukolsky, Vetterling, & Flannery, 1990).
We employed open boundary conditions but made sure that the grating elements were presented far away from the boundary such that its influence on the dynamics was negligible.

All results described below are obtained with the set of parameter values listed in Table 1.

2.3.4 Calibration Between Experimental and Modeling Results. After having adjusted the parameters of the model, activities of the neural populations of the model had to be related to the observed psychophysical thresholds \( d \). For this purpose, we first reduced the data by condensing the spatiotemporal activation pattern \( \phi_e(x,t) \) into one or more order parameters from which we assumed it will be possible to predict the thresholds. We chose two order parameters: the time interval \( T \) that the activity \( \phi_e(0,t) \) remains above an activity level \( A_0 \), and the maximum activation level \( A_m \) of \( \phi_e(0,t) \) during this period \( T \). This choice reflects our assumption that the longer and stronger a neuronal activation, the lower the threshold of its detection is. \( A_0 \) can be interpreted as a noise level above which neural activity clearly indicates the presence of a stimulus and becomes psychophysically relevant. Here, we chose \( A_0 = 0.08 \) (see Table 1). Using \( T \) and \( A_m \), we could check if a calibration using one or a combination of these parameters is possible and if this calibration can be used to predict the results in various other experimental conditions (for the details of this calibration procedure, see the appendix).

We calibrated our model by comparing psychophysical and simulation results for the stimulus condition shown in Figure 2a. Experimentally, thresholds \( d(\tau) \) were measured as a function of the timing parameter \( \tau \); in the model, \( T(\tau) \) and \( A_m(\tau) \) are obtained from the simulations. A combination of the functions \( T(\tau) \) and \( A_m(\tau) \) with \( d(\tau) \) then yields the calibration function (surface) \( d_{cal}(T, A_m) \). This function assigns each possible model output \( T, A_m \) a theoretical threshold \( d_{cal} \). The model can be evaluated quantitatively by computing the theoretical threshold \( d_{cal} \) and comparing this estimation with the empirical threshold \( d \) for all remaining temporal conditions shown in Figures 2b through 2d.

3 Results

3.1 Spatial Aspects. The modeling results for the spatial conditions (see Figure 1) are summarized in Figure 4. In all three conditions, the vernier presentation leads to an increased activation of the center column at \( x = 0 \) during the first 20 ms. After this period, the vernier activity decays, while additional patches of activity emerge corresponding to the edges of the masking gratings. In the 25-element condition (see Figure 4a), two patches appear remote from the center; in the 5-element condition (see Figure 4b), two patches appear in close vicinity of the center, and in the gap condition (see Figure 4c), six patches appear, four of which are located close to the
Figure 4: Spatiotemporal activation levels $A_e(x, t)$ of the excitatory layer for the (a) 25-element, (b) 5-element, and (c) gap grating conditions. Activation levels are expressed on a gray scale (see the right scale). Time $t$ proceeds on the vertical axis. The position $x$ within the excitatory neuronal layer is depicted at the horizontal axis. (d) Activation $A_e(0, t)$ of the center neural population at position $x = 0$, being directly stimulated by the preceding vernier and the following central element of the various gratings. The solid line shows $A_e(0, t)$ for the 25-element, the dashed-dotted line for the 5-element, and the dashed line for the gap grating condition. While in the 25-element grating condition, $A_e(0, t)$ remains above $A_0 = 0.08$ for a long period $T$, the inhibitory interactions in the 5-element and the gap grating conditions drive $A_e(0, t)$ very quickly below $A_0$. Note that the visibility is not reflected in some steady state to which the system converges, but occurs as a transient activity in the system dynamics. The maximum activity is denoted with $A_m$.

center. As soon as activity emerges near the center column (see Figures 4b and 4c), the decay of the vernier activity is accelerated through the inhibitory lateral interactions. Figure 4d shows this suppressive effect in comparison to the decay of activity in the 25-element grating condition. In the 5-element grating and gap grating conditions, the time interval $T$ of the vernier activity above the level $A_0$ is strongly reduced, providing an explanation of the
Local Interactions Explain Global Effects

...deteriorated visibility in the psychophysical experiments. It is important to note that the final steady state of activity at the position of the vernier element is the same in all three conditions. However, the transients in the activation dynamics are very different.

3.2 Temporal Aspects. To obtain quantitative comparisons, we related parameters derived from the model’s output to the threshold \( d \) in the psychophysical experiments. First, we found that a single-order parameter does not allow computing a unique threshold from the model output. This is shown in Figures 5a through 5c, where we used the duration \( T \) as the calibration parameter. Figure 5a shows the experimental thresholds \( d(\tau) \) for stimuli shown in Figure 2a. Figure 5b shows the corresponding results of the simulation for the order parameter \( T. d(\tau) \) and \( T(\tau) \) are combined via the reference \( \tau \) to yield a calibration map \( \hat{d}(T) \) shown in Figure 5c. The curve is composed of two branches: one for negative and one for positive \( \tau \) (see Figure 5c). These two branches do not allow a unique prediction of the psychophysical threshold \( d \) from the model output \( T \). Therefore, a second, independent neural property is necessary. We chose the maximum activation level \( A_m \) of the excitatory population \( A_e(0,t) \) during the period \( T \) as the second-order parameter, which disambiguates the data. The second parameter yields an additional function \( A_m(\tau) \). From the functions \( d(\tau), T(\tau), \) and \( A_m(\tau) \), a calibration curve (surface) \( d_{\text{est}}(T, A_m) \) is obtained via the reference \( \tau \). Figure 5d shows that this functional dependence is now indeed unambiguous (see the appendix for a detailed description of the calibration procedure).

Using this well-defined calibration curve, the predictive power of our model can be assessed by estimating the threshold \( d_{\text{est}} \) from the model output and comparing its value with the empirical data \( d \). Because we already used the data of the experimental condition shown in Figure 2a to the calibration, we can use only the remaining three temporal conditions in Figures 2b through 2d for this purpose. Figures 6b through 6d demonstrate that the modeling results closely approach the experimental curves. Therefore, the model is able to reproduce the various experimental phenomena quantitatively in a unifying framework.

4 Discussion

4.1 Shine-Through. The shine-through effect has been proven to be a very useful tool in the investigation of different areas such as feature binding (Herzog et al., 2001b; Herzog, Parish, Fahle, & Koch, in press), backward masking (Herzog & Koch, 2001), contextual modulation (Herzog & Fahle, 2002), the time course of information processing (Herzog et al., 2001a), and schizophrenia research (Herzog, Kopmann, & Brand, 2002). In particular, the sensitivity of the shine-through effect to subtle spatiotemporal changes of the masking grating makes shine-through a versatile tool. Moreover, the...
Figure 5: (a) Empirical results from Herzog et al. (2001a, Figure 2a). The abscissa denotes the onset time of the gap elements relative to the onset of the gap grating. \( \tau = 0 \) ms refers to the simultaneous presentation of all grating elements. The ordinate shows the vernier displacement \( d(\tau) \) necessary to yield 75\% correct responses. (b) Time interval \( T(\tau) \) of the activity \( A_e(0, t) \) being above the activity threshold level \( A_0 \), for the same stimuli as in a. (c) The resulting calibration map \( \hat{d}(T) \) has two branches, introducing ambiguities. (d) Taking \( T \) as well as the maximum activation level \( A_m \) during the period \( T \) yields a calibration map \( \hat{d} \) leading to a calibration surface resolving the ambiguities from c.
short presentation times of the preceding vernier allow one to test information processing on the brink of the temporal resolution of the human visual system.

4.2 High versus Low-Level Processing. The emergence of the shine-through element reveals complex spatial characteristics that cannot be explained with a phenomenological low-level terminology regarding features such as orientation or luminance. For example, the overall intensity is the same for a homogeneous grating with 25 elements as for a grating containing gaps since both gratings have an identical number of elements. Therefore, the Gestalt of the grating rather than its low-level features determines
whether the vernier appears as a shine-through element (for further results, see Herzog, Fahle, et al., 2001). While these descriptions, based on stimulus phenomenology, point to high-level processing, we could show here that, nevertheless, low-level models can explain the results in the shine-through effect without incorporating any explicit Gestalt processing as often assumed (e.g., Marr, 1982; Nakayama & He, 1995). It might be that Gestalt processing can occur with the circuits found already in the primary visual cortex V1 (see also Caputo, 1996; Wolfson & Landy, 1999; Li, 1999a, 2002; Sugita, 1999; Zhaoping, in press).

4.3 Considerations on the Model. The purpose of the current modeling study was to elucidate the basic neuronal mechanisms yielding the shine-through effect. Our strategy was to choose a structurally simple yet neurophysiologically plausible model, providing a framework to account for spatiotemporal signal processing on short timescales rather than providing a detailed description of the underlying cortical anatomy and physiology. The simplicity of our approach manifests itself in the existence of only a few model parameters. Thereby, we avoid the possibility that the good correspondence between our simulations and the empirical results occurs because of overfitting the data.

For these reasons, we restricted our model to a one-dimensional neural network of the Wilson-Cowan type (Wilson & Cowan, 1972, 1973) consisting of only one inhibitory and one excitatory layer of neuronal populations. We disregarded stimulus features such as a second spatial dimension or the stimulus orientation, since these features are not necessary to account for the visibility of the vernier. Versions of the Wilson-Cowan model class are ubiquitous in vision research and were successfully employed in different cases (see Ernst & Eurich, 2002, for an overview). Examples include texture segmentation (Li, 1999b), figure-ground-segmentation (Li, 2002), and contextual modulation (Bartsch, Stetter, & Obermayer, 2001). Moreover, a simplified version of this dynamics resembling a Hopfield type of model (Li, 1999a) accounts for the dynamics of orientation tuning in primary visual cortex (Ben-Yishai et al., 1995; Ernst et al., 2001). Clearly, this model is not explicitly designed to serve the stimuli of our study. Still, the simulations yield a good quantitative description of the empirical results.

4.4 Mechanisms for the Visibility of the Masked Vernier. Our simulations suggest that the activity of the excitatory neural population at the central position, \( A_e(0,t) \), determines the visibility of the shine-through element. Using both \( T \) and \( A_m \) as theoretical measures, we find a good agreement between the simulations and the empirical data. Therefore, we suggest that the persistence of neural activity (corresponding to \( T \)), together with the maximal activity (corresponding to \( A_m \)), are likely neural candidates underlying psychophysical performance and perception.
In the simulations, two mechanisms are responsible for the observed dynamics. First, the afferent Mexican hat coupling kernels yield enhanced neural activation at the edges or at other inhomogeneities of a grating. Inhomogeneities, such as gaps, are highlighted by enhanced activation if the length scale of the kernel is smaller than the length scale of the inhomogeneity. In this sense, the resulting activation can be interpreted as an edge or inhomogeneity detection. Second, the lateral inhibitory coupling kernels, having a length scale different from the afferent connections, yield a competition among activation patches, which is maximal in a distance of about two to three times the distance $\Delta$ between the grating elements. Therefore, competition and deterioration of visibility are maximal for the 5-element grating but vanish for a 25-element grating. In summary, both afferent and lateral couplings are necessary to explain Gestalt effects in the shine-through effect. Hence, our model extends model architectures based solely on lateral inhibition (e.g., Bridgeman, 1971). The paramount role of edges is supported by results of other psychophysical experiments and physiological recording studies (MacKay, 1973; McCarter & Roehrs, 1976; Sagi & Hochstein, 1985; Macknik, Martinez-Conde, & Haglund, 2000).

In our model, the neural activity of inner elements of the grating quickly approaches zero. However, not only the edges of the grating are subjectively perceived but the whole grating, including the inner grating elements. Our model can easily yield neural activity of these inner elements larger than zero by changing the afferent coupling kernel $V$ such that its integral is positive. We voluntarily chose a parameter regime in which activity approaches zero to highlight the mechanisms responsible for vernier visibility. Moreover, it may be that inner grating elements are indeed first eliminated and then filled in by subsequent processing stages in different areas of the brain.

The simulations also allow an interpretation of the psychophysical results in functional terms. In our simulations, neural activity enhances for neurons corresponding to edge elements, while activity corresponding to inner, homogeneous parts of the grating diminishes. This differential enhancement and weakening makes the grating salient against its background. This process can be understood as figure-ground segmentation (see Li, 1999b). In this sense, shine-through emerges as a by-product during the segmentation of visual scenes. Shine-through occurs only when edge elements, highlighted in the figure-ground segmentation process of the mask, are remote from the neural populations coding the vernier. Our results support models of figure-ground segmentation in which homogeneous objects are detected and coded by their borders, while the interior of the objects is neglected or filled in by a different processing system (e.g., Marr, 1982; Grossberg & Mingolla, 1985).

4.5 Dynamics. Subjectively, the shine-through element appears as a short flash superimposed on the grating. This perceptual appearance has its analogy in the simulations. Activity in the center of the neural layer rises
in a strong transient response, slowly decreasing after its peak time. This transient lasts much longer for a 25-element grating than for a 5-element or a gap grating. A closer inspection reveals that the ratio of the timescales of excitation and inhibition, $\tau_e$ and $\tau_i$, causes the different decays in addition to the inhibition mediated by the couplings kernels.

Therefore, our model suggests that information processing and object coding occur in transients rather than in steady states; that is, perceptual states do not correspond to attractors. This explanation is plausible considering the fact that we live in a dynamic, ever-changing environment. These results are, moreover, supported by physiological investigations. Macknik and Livingstone (1998) showed that “masking” the transient onset and offset response of neurons in primary visual cortex of macaque monkey reduces visibility. Martinez-Conde, Macknik, and Hubel (2000, 2002) suggested that burst transients correlate with increased visibility. Dynamic disambiguation of motion stimuli was found physiologically in the aperture problem (Pack & Born, 2001). In our simulations, the duration the transient $A_e(0, t)$ is above $A_0$ has been proved to be a good measure for vernier visibility. Transient neural activity has also been investigated recently in the context of cortical micro-circuits (Maass, Natschläger, & Markram, 2002; Jäger, 2001).

Another important aspect of the shine-through effect is the subtle dependency of the results on the specific timing of the grating elements. If the peripheral elements of the grating or the gap elements are delayed by only 10 or 20 ms, performance significantly deteriorates in the psychophysical experiments as well as in the simulations. Even if gap elements (see Figure 2a) or a 5-element grating (see Figure 2b) are displayed for only 10 ms before they are completed to the 25-element grating, performance deteriorates dramatically, since their edge elements have already inhibited activity corresponding to the vernier. In these conditions, remarkably, both 5-element and gap gratings yield enhanced activity at the same patches in the close neighborhood to the vernier. Thus, performance deteriorates regardless of whether only single elements or the whole peripheral context is delayed (see Figures 6a and 6b).

These results also impose restrictions on the coding of objects by temporal codes. Since the supposed binding of the grating to a homogeneous surface as well as the presentation time of the vernier, giving rise to a fast transient in the simulations, are in the range of only 10 ms, the underlying neural processes in the brain have to be initiated very quickly—in the range of a few spikes.

4.6 Limitations and Extensions of the Model. There are several limitations to our model resulting from the chosen minimalist approach. First, the model employs only one spatial dimension. Therefore, psychophysical effects resulting from the vertical spatial dimension of the stimulus are not captured by the current model version (Herzog et al., 2002). Second, in the shine-through effect, perception of a single observer can vary strongly from
trial to trial, even if the stimulus is the same. Perception can change from clear to no visibility, since the visual system operates on the brink of its temporal limits. To account for such effects, noise has to be added to the system, which we avoided for simplicity. Third, our model does not address questions arising from a possible involvement of the motion system contributing to masking. For example, in the Fröhlich effect, a line in motion can mask itself. Analogous effects may occur for a vernier masked by a small grating of lines.

The Wilson-Cowan model used is not restricted to grating masks. Shine-through-like effects (e.g., visibility increases with mask extension) occur also for other classical masking paradigms, such as masking by light, noise, pattern, or metacalorner. We could reproduce these results very well by employing the same Wilson-Cowan model used here (Herzog, Harms, Ernst, Eurich, Mahmud, & Fahle, 2002).

5 Conclusion

The shine-through paradigm reveals complex spatiotemporal effects pointing to high-order Gestalt processing. Here, we showed that these complex effects can be reproduced by a simple Wilson-Cowan type model commonly used in vision research. Because of its simplicity, the underlying neurophysiological mechanisms may occur already at the very early stages of visual processing—not incorporating any explicit high-order Gestalt processing. The model captures the experimental results not only qualitatively but also quantitatively remarkably well.

Appendix: Calibration Procedure

Before calibration, the parameters of the model have already been chosen appropriately. The goal of the calibration procedure is to establish a quantitative relationship between the model outputs $T$ and $A_m$ and the thresholds $d$ obtained in the psychophysical experiments. That is, we have to find a function $d_{est}(T, A_m)$ that maps $T$ and $A_m$ onto an estimation of the threshold, $d_{est}$. For this purpose, we use the model output and empirical results of the condition shown in Figure 2a. Once this relationship is fixed, $d_{est}$ can be used to predict the thresholds $d$ in the remaining conditions, Figures 2b through 2d. The difference between $d_{est}$ and $d$ is a measure of the performance of the model.

As a reference that allows us to relate pairs of $T$, $A_m$ to the experimental thresholds $d$, we use the time parameter $\tau$, which was varied in the experiment shown in Figure 2a (first temporal condition). The calibration is done in three steps:

1. Using the different time parameters $\tau$, from the first temporal condition as a reference frame, the empirical results $d(\tau)$ together with the
model outputs $T(\tau_i)$ and $A_m(\tau_i)$ define a discrete map $\hat{d}$,

$$\hat{d}: \{T(\tau_i), A_m(\tau_i)\} \mapsto d(\tau_i).$$  \hspace{1cm} (A.1)

2. In order to estimate the threshold for arbitrary pairs $\{T, A_m\}$, the calibration function $d_{est}$ is constructed by inter- and extrapolation from the map $\hat{d}$ known only at the discrete points $\{T(\tau_i), A_m(\tau_i)\}$. Because there are only a few points in the $(T, A_m)$-space, the choice of the extrapolation algorithm is crucial for avoiding pathological estimates. To define $d_{est}$, we used an $N$-nearest neighbor interpolation scheme defined as follows:

- Let $i(j), j = 1, \ldots, N$ denote the indices of the $N$ nearest neighbors $\{T(\tau_{i(j)}), A_m(\tau_{i(j)})\}$ to a test point $(T, A_m)$, according to the distance $D_i$ measured as

$$D_i := \sqrt{(T - T(\tau_i))^2 + c_0 (A_m - A_m(\tau_i))^2},$$  \hspace{1cm} (A.2)

with an appropriate scaling constant $c_0$ relating time intervals to activity levels, which was set in our study to $c_0 = 200$ with $N = 3$.

- The nearest-neighbor interpolate $d_{est}(T, A_m)$, which associates a higher weight to neighbors lying closer to the test points $(T, A_m)$, is given by the expression

$$d_{est}(T, A_m) = \frac{1}{N - 1} \sum_{j=1}^{N} \hat{d}(T(\tau_{i(j)}), A_m(\tau_{i(j)}))$$

$$\times \left(1 - \frac{D_{i(j)}}{\sum_{k=1}^{N} D_{i(k)}}\right).$$  \hspace{1cm} (A.3)

3. We now compute $d_{est}$ for the experimental conditions shown in Figures 2b, 2c, and 2d, and compare this prediction with the empirical threshold $d$ in order to access the performance of the model.

In general, the data have to be preprocessed before applying the first calibration step. Normally, the simulations allow sampling $T$ and $A_m$ for a large number $N_{sim}$ of times $\tau_{i,m}, i = 1, \ldots, N_{sim}$. In the experiments, $d$ has been sampled for fewer delays $\tau_{j,exp}, j = 1, \ldots, N_{exp}$. In order to have common reference frames $\tau_i$ for these variables, we interpolate the unknown thresholds $d(\tau_{i,m})$ for delays $\tau_{i,m} \neq \tau_{j,exp}$ $\forall j, i = 1, \ldots, N_{exp}$, from the known thresholds $d(\tau_{i,exp})$, using a piecewise cubic Hermite interpolation (Fritsch & Butland, 1984).
Acknowledgments

We thank Greg Francis and Li Zhaoping for useful comments. M.H., U.E., and C.E. were supported by the SFB 517 “Neurocognition” of the Deutsche Forschungsgemeinschaft (DFG). U.E. was also supported by the Volkswagen Stiftung, project 5425, “Neuronal Mechanisms of Gestalt Perception.”

References


Zhaoping, L. (in press). V1 mechanisms and some figure-ground and border effects. J. Physiol.

Received August 22, 2002; accepted March 4, 2003.