Psychological Review
2008, Vol. 115, No. 1, 83–100

0033-295X/08/$12.00 DOI: 10.1037/0033-295X.115.1.83

Modeling Spatial and Temporal Aspects of Visual Backward Masking

Frouke Hermens, Gediminas Luksys, Wulfram Gerstner, and Michael H. Herzog
École Polytechnique Fédérale de Lausanne (EPFL)

Udo Ernst
Universität Bremen

Visual backward masking is a versatile tool for understanding principles and limitations of visual information processing in the human brain. However, the mechanisms underlying masking are still poorly understood. In the current contribution, the authors show that a structurally simple mathematical model can explain many spatial and temporal effects in visual masking, such as spatial layout effects on pattern masking and B-type masking. Specifically, the authors show that lateral excitation and inhibition on different length scales, in combination with the typical time scales, are capable of producing a rich, dynamic behavior that explains this multitude of masking phenomena in a single, biophysically motivated model.

Keywords: visual backward masking, temporal vision, spatial vision, modeling

Perception is not immediate. The brain processes visual information from a scene over a considerable time before a conscious percept is formed. A remarkable demonstration of this time-consuming processing comes from visual masking, in which performance on a target can be affected by a mask trailing the target by several hundred milliseconds (for a review, see Breitmeyer & Öğmen, 2006). Because of these long lasting effects, visual masking is an important technique in studying the dynamics of perception.

In addition, masking is a widely used tool in many other research areas. Bachmann (1994) estimated that in 14% of the articles in vision research and psychology, either masking by itself is investigated or masking is used as a tool to curtail the processing time of the target. A recent survey yielded a very similar estimate (Enns & Di Lollo, 2000). In spite of the large numbers of studies on visual masking, the underlying mechanisms are still only poorly understood, possibly due to the complex timing issues involved.

Visual masking can be classified along various directions. First, in forward masking, the mask precedes the target, whereas in backward masking, the mask follows the target. Second, a mask that spatially overlaps with the target is called a pattern mask, whereas a mask that does not overlap with the target is termed a metacontrast mask. Third, in A-type masking, masking is strongest when the target and the mask are presented simultaneously and becomes weaker once the mask is delayed with respect to the target. In B-type masking, the strongest masking occurs for intermediate stimulus onset asynchronies (SOAs). This means that in A-type masking, the masking curve, plotting performance as a function of SOA, is a monotonic function, whereas in B-type masking, the masking curve shows a dip at intermediate SOAs and is therefore nonmonotonic.

Visual masking research shows an interesting dichotomy. When masking itself is under investigation, B-type masking and metacontrast masks are of primary interest both in experimental and in modeling work. To account for the complex and nonlinear aspects of visual masking, researchers proposed several models. In one class of models, two processing pathways are assumed, one faster than the other. The mask signals in the faster pathway catch up with the target signals in the slower pathway, and thus, the mask signals have their strongest impact at a nonzero SOA (e.g., Bachmann, 1994; Breitmeyer & Ganz, 1976; Öğmen, 1993). Other types of models make use of recurrent connections to explain the complex characteristics of temporal masking (Di Lollo, Enns, & Rensink, 2000).

If masking is used as a tool, A-type masking is often implicitly assumed to prevail, and pattern masks are used. Surprisingly, in this research area, there are hardly any quantitative models to explain how the pattern mask interacts with target processing.

The main focus in all kinds of temporal masking is, as the name suggests, on the temporal aspects. Investigations of spatial aspects were restricted to basic dimensions, such as the spatial distance between the target and the mask or the similarity between the target and the mask (e.g., Parlee, 1969; Sekuler, 1965; Wehrhahn, Li, & Westheimer, 1996). There are only a few studies in which more complex spatial aspects were investigated (e.g., Werner, 1935; Williams & Weisstein, 1981, 1984), whereas this topic is more intensively studied in simultaneous masking. However, in simultaneous masking, as the word suggests, the temporal dimension is lacking. It was only recently that the effects of the spatial layout of pattern and metacontrast masks were investigated systematically in temporal masking. In these studies, temporal and

Frouke Hermens and Michael H. Herzog, Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland; Gediminas Luksys and Wulfram Gerstner, Laboratory of Computational Neuroscience, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL); Udo Ernst, Institut für Theoretische Physik, Universität Bremen, Bremen, Germany.

This work was supported by the Swiss National Science Foundation. Udo Ernst was supported by E.U. Grants BIND MECT-CT-20095-024831 and BACS FP6-IST-027 140. We thank Bruce Bridgeman, Greg Francis, and Haluk Öğmen for their useful comments on earlier versions of the article.

The Matlab code used for the simulations is available from our website at http://lpsy.epfl.ch/Masking_Model/

Correspondence concerning this article should be addressed to Frouke Hermens, Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Station 15, CH-1015, Lausanne, Switzerland. E-mail: frouke.hermens@gmail.com

spatial aspects were combined (Cho & Francis, 2005; Duangudom, Francis, & Herzog, 2007; Francis & Cho, 2005) or complex spatial aspects, such as the overall structure of the mask (Herzog & Fahle, 2002; Herzog & Koch, 2001), were varied.

In some of these more recent investigations, a vernier target was followed by a grating consisting of aligned verniers (*shine-through effect*; Herzog, Fahle, & Koch, 2001; Herzog & Koch, 2001). Surprisingly, gratings with 25 aligned verniers yielded weaker masking than did smaller gratings with, for example, 5 elements (see also Sturr, Frumkes, & Veneruso, 1965; Wehrhahn et al., 1996; Westheimer, 1967). However, it is not the sheer number of elements of the mask that determines the masking strength, but the overall spatial layout of the mask. For example, removing 2 lines from the weak 25-element grating, thereby creating an irregularity by means of gaps, makes vernier offset discrimination as difficult as with the 5-element grating (Herzog et al., 2001). In general, any kind of mask irregularity near the vernier position deteriorates performance (Hermens & Herzog, 2007; Herzog & Fahle, 2002). Moreover, masking functions can be changed from A-type to B-type and vice versa just by changing the spatial layout of the mask (Duangudom et al., 2007).

On the basis of these findings, it was proposed that purely temporal as well as purely spatial approaches to masking and to vision in general are both insufficient. Models have to combine both spatial and temporal processing (see also Herzog, 2007). In explaining temporal and spatial aspects of masking, different description levels may be used. For example, up to now, we described masking effects on a stimulus description level (removing elements) and in terms of perceptual organization (creating gaps; irregularities). In addition, explanations based on grouping mechanisms (Herzog, Dependahl, Schmonsees, & Fahle, 2004; Herzog & Fahle, 2002) were proposed.

However, masking effects may also be explained on other description levels, for example, on a neural description level. We recently proposed that particularly the strong masking with the small (5-element) grating and the gap (25-element grating with 2 elements removed) grating can be explained by dynamic lateral inhibition. For this explanation (Hermens & Ernst, 2007; Herzog, Ernst, Etzold, & Eurich, 2003), we used a simple one-dimensional neural network model inspired by the work of Wilson and Cowan (1973). Here, we extend this structurally simple one-dimensional model to a two-dimensional model to explain how even more complex arrangements of the spatial layout of a mask determine its masking strength. In addition, we show that our extended model is also capable of explaining many important temporal effects in masking, such as B-type masking.

The model that we propose is possibly one of the structurally simplest models that can explain such a large range of experimental findings on masking. Given the complexity of the human brain, it is not entirely surprising that a simple two-layer network cannot explain the entire range of masking phenomena. Because they are informative for model development, we also show the shortcomings of the model.

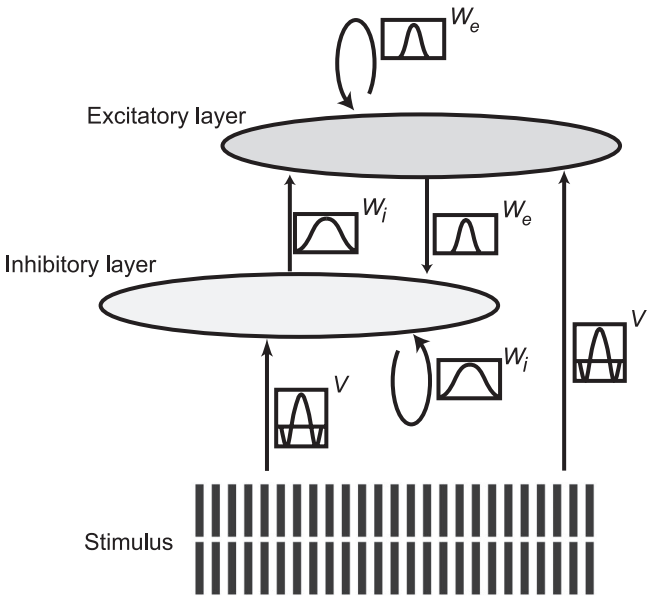

Figure 1. Setup of the model. The stimulus is fed through a filter (V) into an inhibitory and an excitatory layer. Interactions of neural activity between and within layers are mediated by kernels W_i and W_e.

Model

Structure and Dynamics of the Model

We performed simulations with a two-dimensional version of a neural network model comprising an excitatory and an inhibitory layer of mutually interconnected neurons. The structure of the model is illustrated in Figure 1. The input I from a time-varying stimulus S, presented in the visual field, feeds into both an excitatory and an inhibitory layer via a Mexican-hat input kernel V, acting effectively as a low-pass filter. Neurons in both layers mutually interact with excitatory couplings W_e and inhibitory couplings W_i. The strength of both inhibitory and excitatory interactions decays with the distance between neural populations, following a Gaussian shape. The length scales are larger for inhibitory weights, which establish a typical Mexican-hat shape of the combined couplings W_e–W_i, serving the process of edge enhancement (Grossberg, 1982; Wilson & Cowan, 1973). The excitatory and inhibitory layers have the same size as the input map.

The dynamics of the model are described by a set of partial differential equations for the population activities A_e and A_i. These are simplified versions of the equations originally introduced by Wilson and Cowan (1973):

$$\tau_e \frac{\partial A_e(\mathbf{x},t)}{\partial t} = -A_e(\mathbf{x},t) + h_e\{w_{ee}(A_e * W_e)(\mathbf{x},t) + + w_{ie}(A_i * W_i)(\mathbf{x},t) + I(\mathbf{x},t)\} \quad (1)$$

and

$$\tau_i \frac{\partial A_i(\mathbf{x},t)}{\partial t} = -A_i(\mathbf{x},t) + h_i\{w_{ei}(A_e * W_e)(\mathbf{x},t) + + w_{ii}(A_i * W_i)(\mathbf{x},t) + I(\mathbf{x},t)\}. \quad (2)$$

The parameters τ_e and τ_i denote time constants; w_{ee}, w_{ei}, w_{ie}, and w_{ii} are coupling strengths; $\mathbf{x}$ denotes a two-dimensional position vector in one of the neural layers; and t parameterizes time. h_e and h_i are defined as

$$h_{e,i}(x) = \begin{cases} s_{e,i} \cdot x & \text{for } x > 0 \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

with neuronal gain constants s_e and s_i.

The coupling structures are defined as

$$W_{e,i}(\mathbf{x} - \mathbf{x}') = \frac{1}{2\pi\sigma_{e,i}^2} \exp\left(-\frac{|\mathbf{x} - \mathbf{x}'|^2}{2\sigma_{e,i}^2}\right), \quad (4)$$

for excitatory and inhibitory interactions, respectively, whereas $\sigma_{e,i}$ represents the widths of the interaction kernels. The convolution, denoted by a star (*), is a shorthand notation for

$$w_{ee}(A_e * W_e)(\mathbf{x},t) = w_{ee}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A_e(\mathbf{x}',t)W_e(\mathbf{x} - \mathbf{x}')d\mathbf{x}'. \quad (5)$$

This equation operates as a filter weighting the activation across the layer.

The input to both populations is computed from

$$I(\mathbf{x},t) = (S * V)(\mathbf{x},t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(\mathbf{x}',t)V(\mathbf{x} - \mathbf{x}')d\mathbf{x}', \quad (6)$$

with the input kernel defined as

$$\mathrm{V}(\mathbf{x} - \mathbf{x}') = \frac{1}{2\pi\sigma_E^2}\exp\left(-\frac{|\mathbf{x} - \mathbf{x}'|^2}{2\sigma_E^2}\right) - \frac{1}{2\pi\sigma_I^2}\exp\left(-\frac{|\mathbf{x} - \mathbf{x}'|^2}{2\sigma_I^2}\right), \quad (7)$$

with σ_E and σ_I representing the widths of the input kernel. The values of the different constants in the model are listed in the Appendix table.

Linking Model Output to Perception

A linking hypothesis relates one or more model values to the experimentally obtained values. For example, the activity of a set of neurons in the model is mapped onto the percentage of correct responses or a threshold value. In our particular implementation of the model, we linked the activation in the excitatory layer corresponding to the target position to a measure of performance (such as a threshold or the percentage of correct responses). In detail, we determined the activation of the excitatory neurons 80 ms after target onset.[1] From the activation in the excitatory layer, we took only the activation of cells at the exact position of the vernier target. The activation at those positions was then summed. This procedure is summarized by the following equation, computing the target-related activation T:

$$T = \int_{\mathbf{x}} A_e(\mathbf{x},r_0) \cdot S_T(\mathbf{x})d\mathbf{x}, \quad (8)$$

where S_T is the representation of the target vernier, r_0 is the read-out time of the excitatory activity (80 ms after target onset in our implementation of the model), $\mathbf{x}$ is the spatial coordinate, and A_e denotes the activation in the excitatory layer.

For experiments related to the shine-through effect, we converted the target-related activation T to a predicted threshold value that we then compared with the corresponding psychophysical discrimination threshold. This psychophysical threshold is defined as the offset (in arc seconds, i.e., a measure of angular distance) between the upper and the lower segment of the vernier, for which 75% correct responses are obtained. For the conversion, we used a sigmoid function that takes into account the fact that thresholds for a masked vernier were generally not below 15 arc seconds and that performance was truncated to a maximal threshold of 350 arc seconds in the psychophysical experiments,

$$\hat{\Theta} = 15 + \frac{335}{1 + \exp(-a \cdot (T_{25\ grating} - T) + s)}. \quad (9)$$

The free parameters a and s were determined once by means of a least-squares fit of the data of one selected experiment (training data set; for details of this experiment, see the *Varying mask width* subsection). For predicting the results of all other experiments (test data set), this function was held fixed. To obtain positive numbers for a and s, we subtracted T for each mask from T for the 25-element mask, which served as a baseline, because it typically yielded the weakest masking. To allow for an evaluation of the model fit, we always plotted the model predictions and the experimental thresholds in one graph. For an optimal fit, we expected the predicted thresholds to be within the confidence interval of the experimental thresholds.

We used the simplest linking hypothesis possible, namely, comparing the network activation directly with the target template. We also considered more complex linking hypotheses. Most yield comparable results (see Appendix for more discussion).

Results

The Results section is divided into two parts; in the first part, we consider spatial aspects of masking, whereas in the second one, we investigate temporal aspects of masking. We start by describing a particular set of experiments on the shine-through effect. One of these experiments also served to determine the threshold predictor's parameters a and s.

Spatial Aspects

Basic model behavior and the shine-through effect. If a vernier target, displayed for about 20 ms, is followed immediately by a 5-element grating mask (illustrated to the left of Figure 2), the vernier is rendered invisible. However, much weaker masking occurs for a mask consisting of 25 aligned verniers. In this condition, participants report seeing the vernier target shining through the mask (Herzog et al., 2001). This effect might be related to the sheer number of mask elements: The larger the mask size, the weaker the masking. However, if only 2 elements are removed from the central part of the mask (gap grating; bottom left illustration of Figure 2), masking is as strong with a 5-element mask.

In recent publications (Herzog et al., 2004; Herzog & Fahle, 2002), we argued that these results can be explained in terms of perceptual organization by the Gestalt cue of proximity. Inserting gaps into the 25-element grating changes the grouping of the mask

[1] The exact value of this moment in time proved not to be critical: Other values ranging from 60 ms to 120 ms yielded very similar results.

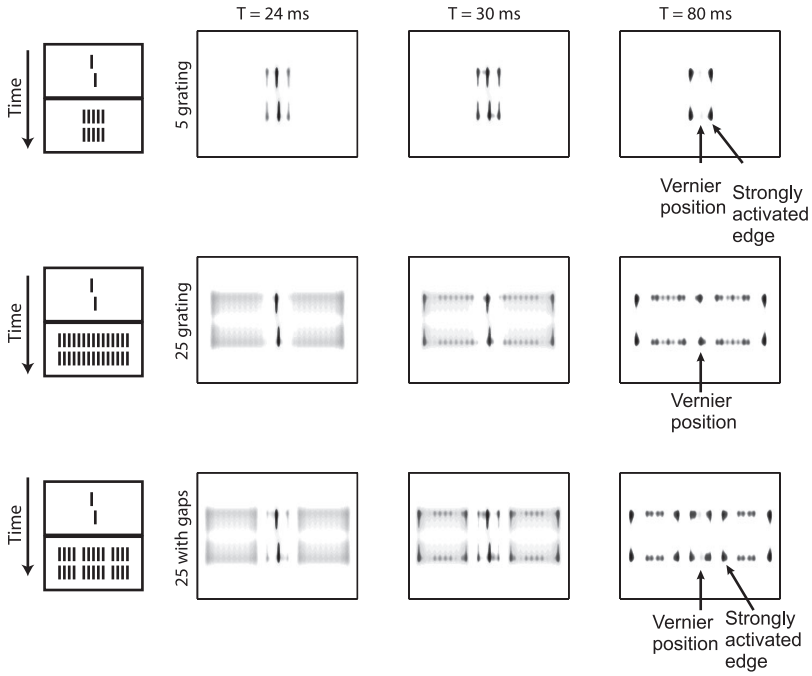

Figure 2. Activation $A_e(\mathbf{x}, t_0)$ of the excitatory layer at three points in time: $t_0 = 24$ ms, $t_0 = 30$ ms, and $t_0 = 80$ ms after vernier onset. Three masks were used for the simulations (see the illustrations on the left): a grating consisting of 5 elements (5 grating), a grating consisting of 25 elements (25 grating), and a gap grating, that is, a grating of 25 elements from which 2 elements were removed (25 with gaps). Arrows indicate the positions of the vernier target and the strongly activated edges. Arrows next to Time indicate the direction of time. For graphic purposes, only the center 15 elements of the grating are shown in the illustrations. Movies of the different simulations, showing the time course of activation, can be found at http://froukehe.googlepages.com/simulations.

elements: Three smaller gratings are perceived, not one large grating with gaps. Because the central grating of the three gratings consists of 5 elements, it masks the vernier as strongly as the 5-element grating presented alone. Hence, it seems that the grouping of the grating elements determines the strength of masking. These explanations, in terms of perceptual organization, leave open explanations on other description levels, for example, on a neural description level.

Often grouping operations are thought to be accomplished by higher cortical mechanisms dedicated to grouping (for a review, see Palmer, Brooks, & Nelson, 2003). We show that our model demonstrates that some of these grouping aspects can be entirely explained by low-level neural interactions. This is an interesting finding, because the model was not explicitly developed to explain grouping. It is important to note that propositions at the various description levels are independent, just as, for example, micro theories and macro theories are independent from each other in the sense that either one can be true or false.

To understand the mechanisms that may underlie the suggested grouping in an intuitive way, we compare the activation of the excitatory layer at three different points in time for three different masks (Figure 2). In the simulations, we used a vernier target consisting of two elements each of 600 arc seconds (10 arc minutes) in height and 20 arc seconds in width. The elements were separated by a vertical gap of 60 arc seconds (1 arc minute) and a horizontal displacement of 40 arc seconds. The grating elements were similar to those in the vernier target (with a height of 600 arc seconds, a width of 20 arc seconds, and a vertical gap of 60 arc seconds). The only difference was that the segments of each grating element were aligned (no vernier offset).

Figure 2 shows that the Mexican-hat input filter and the two interaction kernels strengthen the edges of the grating mask, thereby suppressing the individual elements inside the mask. The highlighted edges of the 5-element grating dynamically suppress the target vernier signal. However, no such suppression occurs for the 25-element grating, because the edges of this grating are too remote from the target. When 2 elements from such a large mask are left out (25 with gaps), the elements making up the gaps are strongly activated, suppressing the vernier signals as efficiently as the edges of the 5-element grating.

This intuitive understanding of the basic mechanisms in the model can be formalized in mathematical terms by means of a stability analysis, which links critical sizes of the mask to the typical length scales of the intracortical interactions, revealing which spatial modes become enhanced or suppressed by recurrent activity (see the Appendix). Our simulation results agree well with those of Herzog, Ernst, et al. (2003), in which a one-dimensional version of the model was used and which indicated that the generalization of the model to two dimensions left its properties intact.

Varying mask width: Calibrating the linking hypothesis. If the number of elements in the grating mask is systematically varied from 3 to 25 (Herzog, Harms, et al., 2003), while keeping the spacing between the elements constant at 200 arc seconds (3.3 arc

minutes), strongest masking is found for a mask consisting of 5 elements (open squares in Figure 3). Our neural network model is qualitatively capable of explaining this finding. We now use the empirical discrimination thresholds to calibrate the free parameters of our linking hypothesis, obtaining $a = 0.4419$ and $s = 1.7547$. A very good fit is obtained for this calibration data set (Figure 3), showing an accurate match between the predicted thresholds (filled circles) with the experimentally observed ones (measured in seconds of arc, open squares).

Gaps in the grating. The simulations with the gap grating as a mask (Figure 2) suggested that the model is very sensitive to irregularities in the mask's structure. To quantify this observation, the size of the gap was increased from 200 arc seconds (normal spacing) to 600 arc seconds (3 times the spacing; Herzog et al., 2001). Figure 4A shows the predictions by the model (solid circles) together with the experimentally observed thresholds (open squares). The increase of thresholds with gap size was correctly predicted by the model, although the linear shape of the experimentally obtained curve could not be reproduced.

To improve our understanding of how a gap in a grating is processed, lines of various lengths were inserted in the gaps, thereby morphing the gap grating into a grating in which all elements have the same length. Figure 4B shows how the length of the gap element affects the predicted masking strength (solid circles), together with the experimental observations (open squares, Herzog et al., 2001). The model predictions match the experimental observations very well. Data and predictions indicate that a small element inside the gap yields almost as strong a masking as a full gap grating. Only when the size of the element increases beyond 300 arc seconds (5 arc minutes) do thresholds decrease and approach the level of the standard 25-element grating.

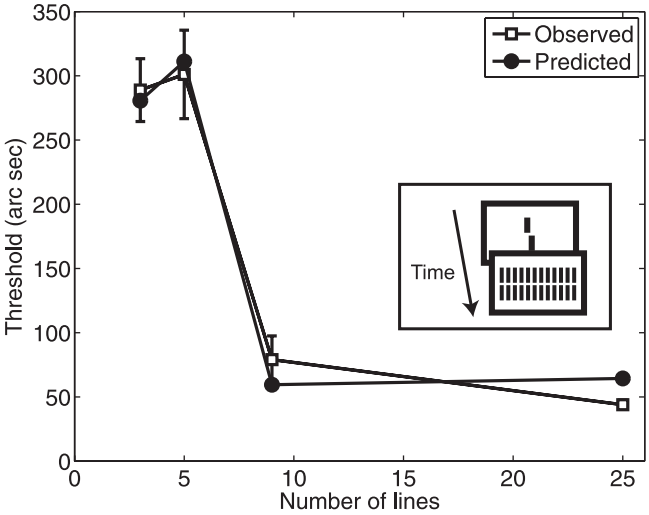

Figure 3. Vernier target offset discrimination thresholds (open squares; data replotted from "Extending the Shine-Through Effect to Classical Making Paradigms," by M. H. Herzog, M. Harms, U. Ernst, C. Eurich, S. Mahmud, & M. Fahle, 2003, *Vision Research, 43,* Figure 7, p. 2664. Copyright 2003 by Elsevier) displayed as a function of the number of aligned verniers in the successively presented mask (an example of the stimulus sequence is shown in the inset). In comparison, the predicted threshold (filled circles) is shown. Error bars represent the standard error of the mean across participants for the experimentally observed thresholds.

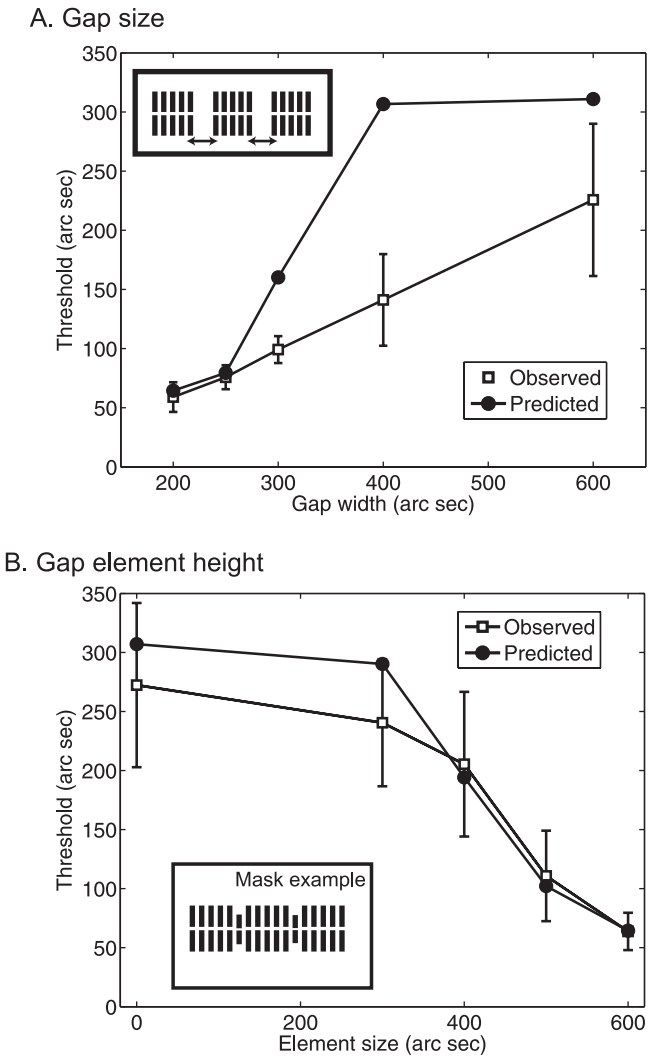

Figure 4. Model predictions (solid circles) shown together with experimental data (open squares; replotted from "Spatial Aspects of Object Formation Revealed by a New Illusion, Shine-Through," by M. H. Herzog, M. Fahle, & C. Koch, 2001, *Vision Research, 41,* Figures 5 & 7, pp. 2331–2332. Copyright 2001 by Elsevier) for masks with gaps (gap grating). A: Shows the dependence of thresholds on the gap width between the mask elements (see inset). B: Shows the dependence of thresholds on the size of the elements at the location of the gaps between the mask elements (see inset). The error bars represent the standard error of the mean across participants.

This data set demonstrates the need for a two-dimensional model. Without a complete representation of the structure of the mask, it is impossible to show that even small irregularities in the grating strongly suppress the vernier-related activity in good accordance with the experimental findings.

Contextual modulation. In the previous paragraphs, we showed that nonhomogeneities in the mask, such as gaps or changes in the length of elements, can strongly affect masking strength, pointing to complex spatial processing. Herzog, Schmonsees, and Fahle (2003) showed that also contextual lines outside the mask could strongly modulate masking strength (Figure 5A,

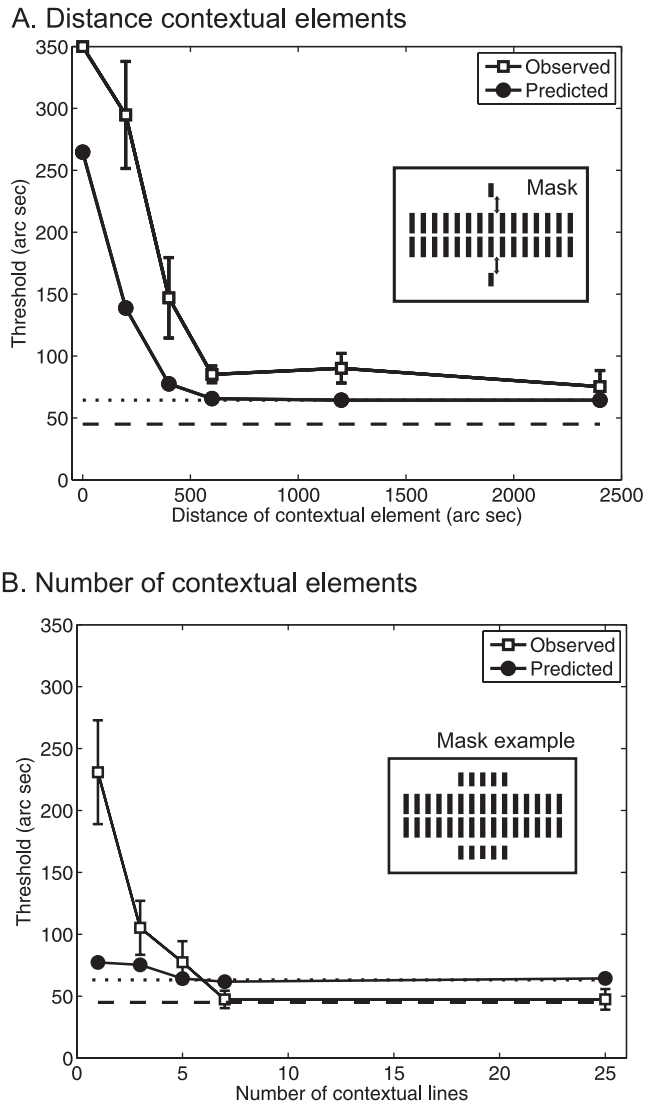

Figure 5. Model predictions (solid circles) shown together with experimental data (open squares; replotted from "Collinear Contextual Suppression," by M. H. Herzog, U. Schmonsees, & M. Fahle, 2003, *Vision Research, 43,* Figures 2 & 4, pp. 2917–2918. Copyright 2003 by Elsevier) for masks in which single contextual lines were added above and below a standard 25-element mask. A: Shows the dependence of thresholds on the contextual element's distances (see inset). B: Shows the dependence of thresholds on the number of contextual elements (for a stimulus example with 5 elements, see inset). The dashed and the dotted lines show, respectively, observed and predicted performance for a 25-element grating, which served as the baseline. Arrows indicate that the distance between the contextual elements and the grating was varied. Error bars represent the standard error of the mean across participants.

open squares). The spatial range across which these contextual elements have an effect is surprisingly large. Even at a distance of 2,400 arc seconds (40 arc minutes) to the vernier (4 times the height of a vernier segment), the contextual elements still modulate performance.

The model correctly predicts that contextual lines increase the masking strength (collinear inhibition; Figure 5A, solid circles). It also correctly predicts that increasing the distance between the contextual elements and the grating makes masking weaker. However, the contextual effects at very long spatial distances could not be reproduced (the curve reaches the baseline, indicated by the dotted line, for a separation around 500 arc seconds), which probably reflects the limited spatial range of the kernels in the model. This issue might be resolved by the use of a bank of kernels of various widths. However, we chose not to do so to keep the model as simple as possible.

Besides the distance of the contextual element, the number of contextual lines strongly affects the masking strength (Herzog, Schmonsees, & Fahle, 2003). Surprisingly, even though the overall intensity of the mask (Luminance $\times$ Surface $\times$ Duration) increased due to the additional elements, less masking was obtained when the number of contextual lines was increased (Figure 5B, open squares). In this respect, the contextual elements have a similar effect on the masking strength as the elements of the grating itself.

The model correctly predicts that adding contextual elements weakens masking (Figure 5B, solid circles). However, the decrement of the masking strength is clearly underestimated. Moreover, an asymptotic masking strength stronger than that of the 25-element grating is predicted, which clearly contradicts the experimental findings. This might again reflect a too limited range of the model kernels, which we discuss in more detail in the general discussion.

Forward and backward masking sequences. The number of elements in the mask also strongly affects masking strength in sequences comprising a vernier target and two masks. In the first experiment, the vernier was followed by a briefly presented grating of a variable number of elements, which was then followed by a second grating of 25 elements (backward masking sequence; Figure 6A, illustration). A strong effect of the number of lines in the first grating on the vernier offset discrimination was found, with strongest masking at a grating size of 5 elements (Figure 6A, open squares).

A temporal variation of this sequence shows an entirely different pattern. If the briefly presented grating serves as a forward mask to the vernier target, which is then backwardly masked by a 25-element grating for 280 ms (forward masking sequence; Figure 6B, illustration), strong masking is found for almost all grating sizes (Figure 6B, open squares).

Although accurate model predictions are found for the backward masking sequence (Figure 6A, solid circles), the model completely fails to predict the data of the forward sequence (Figure 6B, solid circles). In contrast to the experimental observation, the model predicts strong facilitation with a forward mask with thresholds around that of a target without a mask. This incorrect prediction can be understood from the summation of the activation corresponding to the target and the mask. The energy in the forward mask adds to the target-related energy instead of suppressing it. This problem is likely to appear in all models that do not separate the energy of the target and the mask into different channels. We elaborate on this issue in the general discussion.

A closer analysis of the kernel widths of the model shows why the model could explain the finding that a backward mask of 3 elements masked the vernier less than did a mask of 5 elements. In Figure 7, the model kernel widths are plotted, centered at the vernier position and at the mask edges, superimposed on the

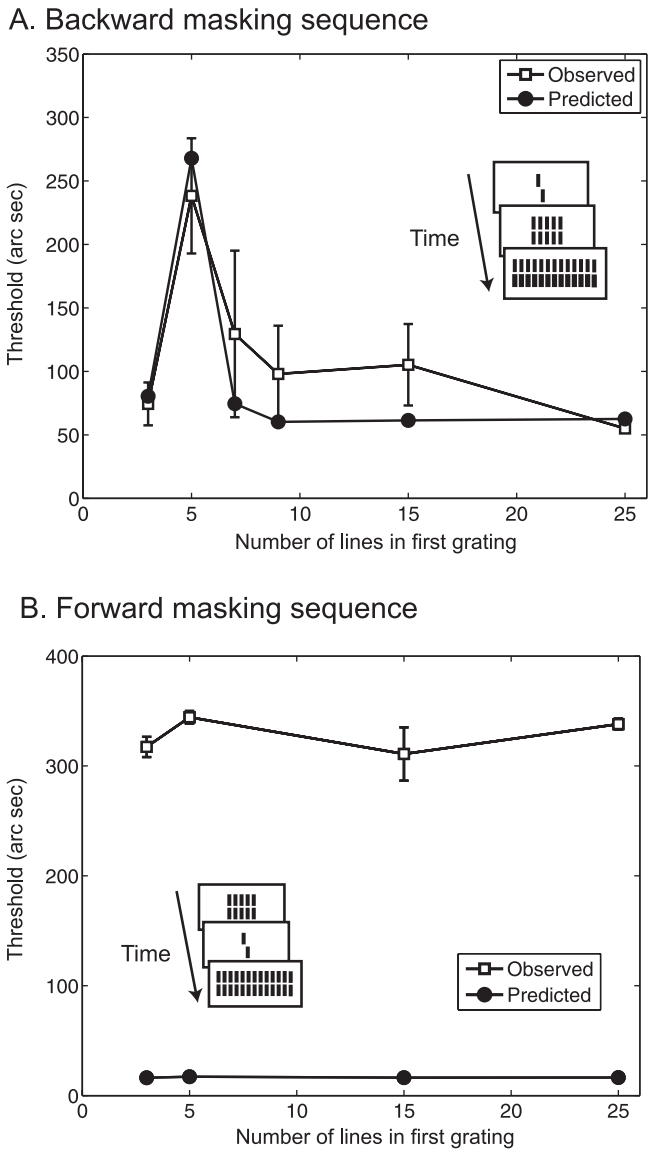

Figure 6. Model predictions (solid circles) are shown together with experimental data replotted from "Grouping in the Shine-Through Effect," by M. H. Herzog, U. Schmonsees, J. M. Boesenberg, T. Mertins, & M. Fahle, 2007 (Copyright 2007 by M. H. Herzog, U. Schmonsees, J. M. Boesenberg, T. Mertins, & M. Fahle), for backward and forward masking sequences (see insets). A: Shows dependence of thresholds on the number of lines in the first mask. The sequence consisted of a target vernier followed by a grating of a variable number of lines, followed by a 25-element mask. B: Shows dependence of thresholds on the number of lines in the forward mask. The sequence consisted of a grating of a variable number of lines, followed by the target vernier, followed by a 25-element grating. Error bars represent the standard error of the mean across participants.

activation of the excitatory layer at 33 ms after vernier onset. The smaller discs (solid lines) show the size (2 times the standard deviation) of the excitatory kernels, whereas larger discs (dotted lines) represent the size of the inhibitory kernels. For a 5-element grating, only the inhibitory kernels at the edges overlap with the target vernier, which explains why this mask yields the strongest masking. With 3 elements in the grating, the edges of the mask both excite and inhibit the target vernier, which results in a higher overall activation than when only inhibition takes place. For a 25-element grating, the edges of the grating are so remote that they hardly affect the vernier target, which is therefore clearly visible.

Temporal Aspects

Masking curves: Shape of the curve. Previous modeling of visual masking strongly focused on explaining the various shapes of the masking function. In A-type masking, performance increases monotonically with an increasing SOA between the target and the mask, as is intuitively expected: The later the mask is presented, the weaker is its effect. In B-type masking, strongest masking occurs at a nonzero SOA. This type of masking is also referred to as U-shaped masking, which reflects the shape of the curve when percentage correct is plotted against SOA. Several studies have found, if the spatial layout of the mask is kept constant, A-type masking with high mask-target intensity ratios, and B-type masking occurs with low mask-target intensity ratios (Breitmeyer & Öğmen, 2006). Moreover, B-type masking is typically observed with masks that do not overlap spatially with the target (metacontrast masks).

For our simulations of the effect of SOA on performance, we used a spatially nonoverlapping metacontrast mask. As the target, we used a filled square, and for the mask, we used a square outline (Figure 8). The target measured 400 arc seconds (6.7 arc min-

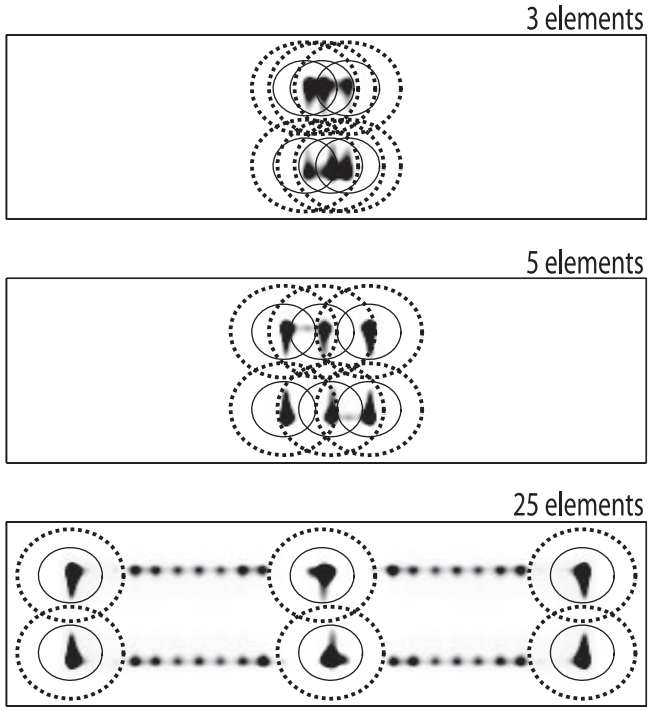

Figure 7. Illustration of the kernel size centered around the vernier position and the grating's edges. The smaller discs (solid lines) represent the excitatory kernels, whereas the larger discs (dotted lines) show the inhibitory kernels. Individual subplots show the activation after 33 ms for the 3-, 5-, and 25-element grating, respectively.

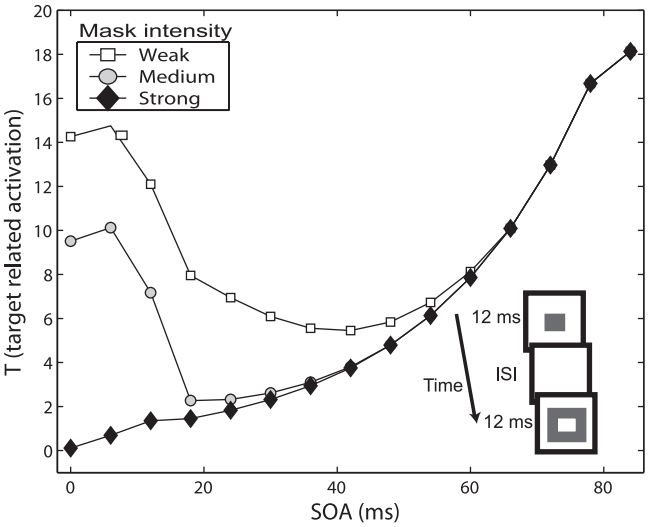

Figure 8. Target-related excitatory activation (T) as a function of the stimulus onset asynchrony (SOA) between target and mask for three different mask intensities. The inset illustrates the stimulus sequence. ISI = interstimulus interval.

utes) × 400 arc seconds, whereas the mask outline surrounded the center of the display at 1,440 arc seconds (24 arc minutes) and was 80 arc seconds (1.3 arc minutes) in width. The target and the mask were both presented for 12 ms. We varied the SOA between the target and the mask from 0 ms to 84 ms. Mask intensities, $I(\mathbf{x},t)$ in the model, of 0.7 (weak), 1.1 (intermediate), and 2.5 (strong) were used. The intensity of the target was kept constant at 0.5. As a measure of performance, we report T (Equation 8), which is monotonically related to the percentage of correct responses.

In accordance with previous experimental findings and modeling results (Bridgeman, 2001; Francis, 1997), the SOA yielding strongest masking is predicted to shift to lower values when the mask intensity is increased (Figure 8). For high intensity masks, strongest masking is found for an SOA of 0 ms, which means that the curve has become monotonic; that is, A-type masking occurs.

For SOAs shorter than 12 ms, in which the target and mask overlap in time, a small nonlinearity can be observed. This is probably the result of the gradual build-up of activation in the model, which contrasts with the immediate build-up of activation in other models (e.g., Anbar & Anbar, 1982).

Wilson and Cowan (1973) reported that they could not obtain a U-shaped masking curve with their model. At this point, it is unclear what difference between their model and ours caused this discrepancy. For our model, we used only slightly modified equations as well as a different linking hypothesis. An analysis in which we used a linking hypothesis that integrates the activity in the excitatory layer over time, which better matches the linking hypothesis used by Wilson and Cowan, also yielded B-type masking, suggesting that the discrepancy in results might have resulted from the difference in equations.

Masking curves: Spatial separation. The spatial distance between the target and a metacontrast mask has been explored in many experiments (see Francis, 2003, for an overview) and with several models (e.g., Bridgeman, 2001; Francis, 2003). Generally, the effect of a metacontrast mask weakens with an increase in the spatial distance to the target. However, how the SOA with strongest masking changes with increasing target–mask distance is not well established (for an overview of results, see Francis, 2003). Model predictions of the effect of the spatial separation also vary. The boundary contour system model (Francis, 1997) predicts only a very weak effect of the spatial separation on the SOA with maximum masking, whereas the efficient masking model (Francis, 2003) predicts a strong shift of the optimal SOA. Here, we present the predictions of our model and compare them with previous model predictions and experimental data.

To investigate the effects of target–mask distance on the predicted masking curve, we performed simulations with a 400 arc seconds (6.7 arc minutes) by 400 arc seconds square target and two flanking bars, each 400 arc seconds in height and 200 arc seconds in width. The distance between target and mask was varied in steps from 200 to 1,800 arc seconds. The input strength, $I(\mathbf{x},t)$, was set to 0.5 for the target and 0.7 for the mask. The target and the mask were both presented for 12 ms.

Figure 9 shows that the model correctly predicts that masking is weaker for larger spatial separations between the target and the mask. Moreover, in agreement with the psychophysical results of Growney (1978), strongest masking (indicated by a larger symbol in the curve) is predicted to shift to longer SOAs with an increasing separation between target and mask. Our model predictions also agree with those of the efficient masking model (Francis, 2003).

Masking curves: Effects of practice. Hogben and Di Lollo (1984) showed that the strength of a metacontrast mask weakens with practice (see also Ventura, 1980). At the same time, the SOA yielding strongest masking remains constant. A possible effect of practice may be an increased attention to the target. This increased attention might be modeled by assuming that the intensity of the target in the neural representation increases. Such an increase of target intensity, however, cannot explain why the SOA with strongest masking remains constant. This is because, as Francis (2003) pointed out, most of the existing models, all sharing a similar underlying mechanism, predict a shift of the bottom of the masking curve when the relative intensity of the target is increased. This also holds for our model (Figure 8).

Instead, we suggest that the effect of practice can be modeled by decreasing the read-out time. If participants are well trained at the task, they might be able to read out the activation in the excitatory layer at an earlier point in time. Such a change in the read-out time would be similar to Bridgeman's (1971) assumption that the effect of task difficulty can be modeled by varying the integration time.

To investigate whether read-out time could explain the upward or downward shift of the masking curve, without changing the SOA when maximal masking occurs, we performed simulations with the stimuli from Figure 8. The input strength was set to 0.5 for the target and to 0.7 for the mask.

Figure 10 shows three masking curves for three different read-out times. In agreement with the data of Hogben and Di Lollo (1984), masking becomes weaker with practice, and the SOA with strongest masking does not change. Model predictions deviate from the experimentally obtained results at two points: First, at short and long SOAs, the experimentally obtained curves converge, whereas the simulated curves remain at an almost constant distance. This discrepancy is probably the result from not scaling the target-related activation to percentages, for which a sigmoid

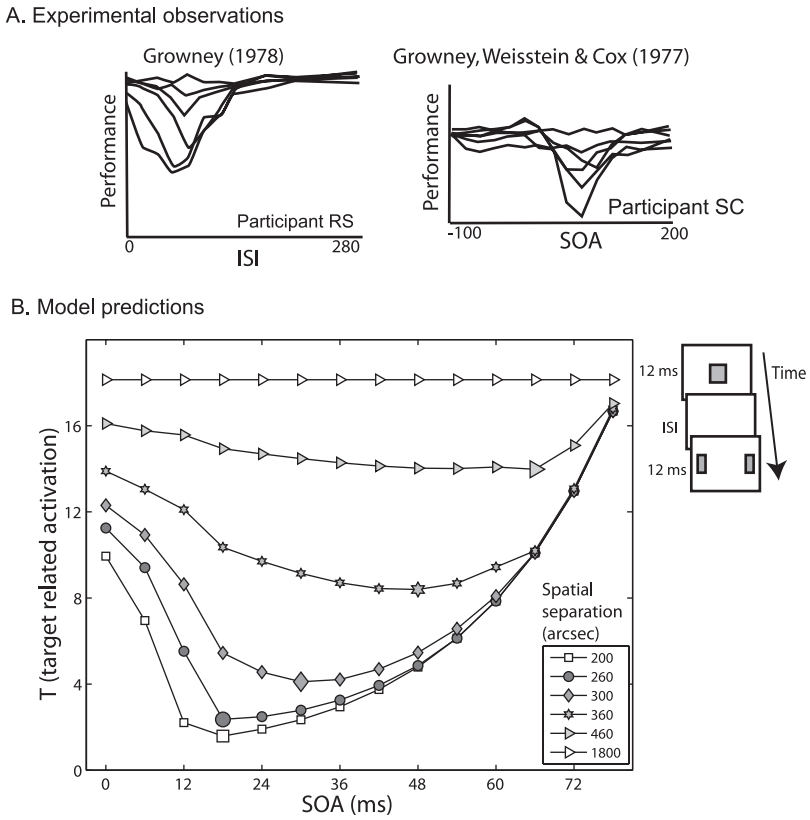

Figure 9. A: Two examples of masking curves as a function of spatial separation. The left set of curves is adapted from "Metacontrast as a Function of the Spatial Frequency Composition of the Target and Mask," by R. Growney, 1978, *Vision Research, 18*, Figure 1, p. 1119. Copyright 1978 by Elsevier. The right set of curves is adapted from "Metacontrast as a Function of Spatial Separation With Narrow Line Targets and Masks," by R. Growney, N. Weisstein, and S. Cox, 1977, *Vision Research, 17,* Figure 1, p. 1207. Copyright 1977 by Elsevier. B: Simulation results of the effect of an increasing spatial separation between target and mask (see legend for size of separation). The stimulus onset asynchrony (SOA) with strongest masking is indicated by larger symbols. The RS and SC abbreviations indicate names of participants in the Growney (1978) and Growney et al. (1977) experiments. ISI = interstimulus interval.

linking function could be used. Such a linking function brings the curves together for near-perfect performance. Second, the value of the SOA with strongest masking is not reproduced. Experimentally, masking was strongest at an SOA of 80 ms, whereas in the model, masking was strongest at an SOA of 40 ms. Many factors could be underlying this discrepancy, such as the particular read-out time used in the model, the exact shape of the stimuli, or the response required from the participant. Because our aim here was to investigate whether we could reproduce the pattern of results rather than to provide an exact fit of the experimental data, we leave this issue to future work.

The effect of changing the read-out time in our model is different from that of changing the integration time in the lateral inhibition model (Bridgeman, 1978). A change in the integration time in the lateral inhibition model results in a shift of the SOA yielding the strongest masking, whereas no such shift was found when decreasing the read-out time in our model. This suggests that changes in read-out time can explain the effects of practice, whereas changes in the integration times in the lateral inhibition model cannot.

Common onset masking. In the common onset masking paradigm (Di Lollo et al., 2000), the target and the mask are presented on the screen together. After some time, the target is removed from the display, whereas the mask remains on for a variable duration. Performance is plotted as a function of mask duration and the resulting curves typically show a decrease in performance with increasing mask duration (Di Lollo et al., 2000; Di Lollo, Von Mühlenen, Enns, & Bridgeman, 2004). Moreover, masking strength is strongly affected by the number of distractor elements in the display. This effect is usually explained in terms of attention. A single target followed by the mask (attention is focused) results in only weak masking. However, increasing the number of possible targets in the display (attention is distributed) strongly impairs performance (Figure 11C). Common onset masking under these conditions of divided attention is also known as object substitution masking, which refers to the mechanism assumed to underlie this type of masking (Enns, 2004).

Common onset masking can be modeled by many of the existing feed-forward and lateral inhibition models of masking (Bischof & Di Lollo, 1995; Francis, 1997; Francis & Hermens, 2002) and by

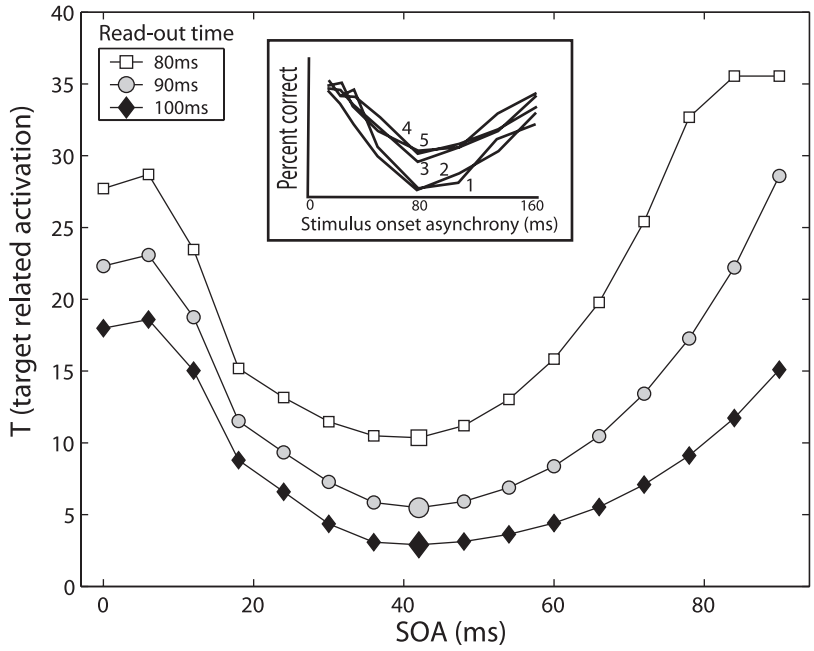

Figure 10. The effect of a decrease of the read-out time on the masking curve, simulating the effect of practice. The small inset shows a representation of the experimental results from "Practice Reduces Suppression in Metacontrast and in Apparent Motion," by J. H. Hogben & V. Di Lollo, 1984, *Perception & Psychophysics, 35*, p. 44. Copyright 1984 by The Psychonomic Society. Numbers 1–5 indicate the different repeated blocks of the experiment.

a recurrent model (Di Lollo et al., 2000). The effects of distributed attention have been modeled in two ways: Either the time to find the target in the search display is assumed to increase with the number of elements (Di Lollo et al., 2000) or the strength of the mask is assumed to vary with the number of distractors (Francis & Hermens, 2002). Here, we show the simulation results for our model in which we varied the intensity of the mask, which provided a slightly better match with the experimental data than a variation in the read-out time (comparable with the search time in the model by Di Lollo et al., 2000). Simulation results for both methods to model the effects of attention are shown on our website at http://lpsy.epfl.ch/Masking_Model/.

A target Landolt C (radius of 200 arc seconds) was presented to the model among three distractor Landolt Cs (Figure 11B; Enns & Di Lollo, 2000). The mask consisted of four dots positioned at the corners of an imaginary square (720 arc seconds × 720 arc seconds) around each Landolt C. To model the effects of attention, we used mask intensities of 0.2 (weak), 0.5 (intermediate), and 0.8 (strong). The intensity of the target was set to 1.0. The target was presented for 12 ms, and mask duration was varied in steps of 6 ms from 0 ms to 54 ms.

Both common onset masking and effect of distributed attention could be explained well by our model (Figure 11C). Predicted masking strength is higher with both increased durations of the mask and higher intensities of the mask (modeling the effects of distributed attention).

General Discussion

We showed that a simple neural network model could explain a broad spectrum of masking phenomena, including both spatial and temporal aspects. It is interesting to note that the model does not rely on multiple channels (Breitmeyer & Öğmen, 2000; Weisstein, 1968) or feedback connections between different layers (Di Lollo et al., 2000) and comprises only two layers in contrast to, for example, the boundary contour system model that comprises six layers (Francis, 1997). Given this simplicity, the range of experimental results that can be explained is remarkable.

Spatial Aspects

Traditionally, masking research has focused on temporal aspects. However, more recently, it was also shown that the spatial layout of a mask strongly affects the masking strength of both pattern (Herzog et al., 2004; Herzog & Fahle, 2002; Herzog, Harms, et al., 2003; Herzog, Schmonsees, & Fahle, 2003) and metacontrast masks (Duangudom et al., 2007; Williams & Weisstein, 1981, 1984). These spatial aspects turned out to be more important than traditional measures, such as the overall mask and the target energy (Luminance × Duration × Surface) ratio, once thought to be the most crucial component to explain the masking strength (e.g., Breitmeyer & Öğmen, 2006). The importance of the spatial layout becomes clear in Figure 3, in which a grating consisting of 5 elements (having a low overall mask energy) masks a vernier target more strongly than a similar grating with 25 elements (having a high overall energy, see also Herzog, Harms, et al., 2003).

More important, it is not the sheer number of elements in the mask that determines the masking strength but rather its complex spatial layout. If only 2 lines are left out of the 25-element mask, thereby creating two gaps, performance drops to a level comparable with that of the 5-element grating. On a description level of perceptual organization, the results can be explained by assuming

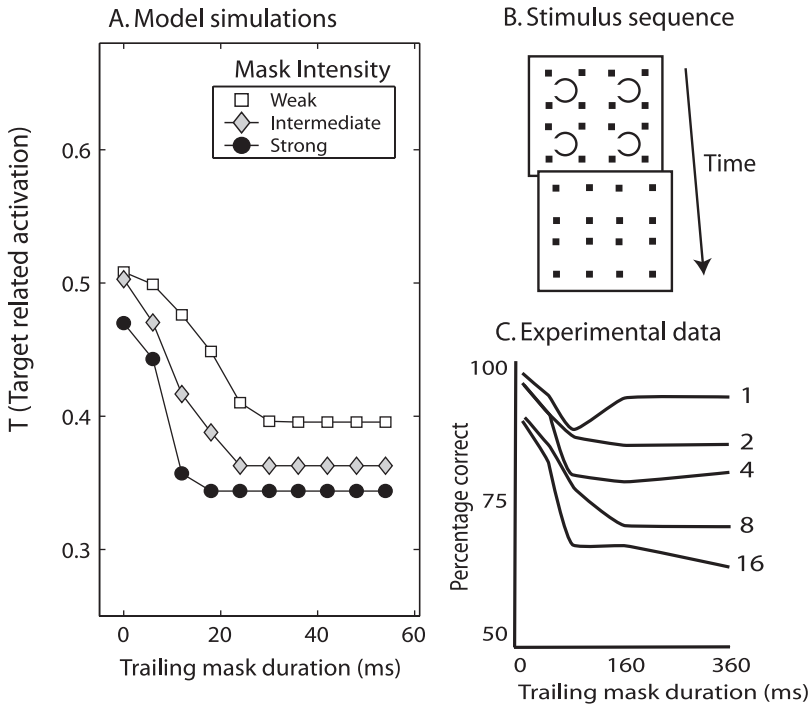

Figure 11. A: Simulation results of common onset masking shown as a function of trailing mask duration for different mask intensities (see legend and main text). B: Illustration of the sequence of target and mask used for the simulations. C: Representation of the experimental data (Di Lollo et al., 2000) with the numbers near the graphs indicating the number of possible targets.

that the gaps group the center 5 elements into one grating and the other elements into two peripheral gratings (due to grouping by proximity). In terms of neural processing, our model simulations show that a first step of this grouping may be explained by a neural highlighting of irregularities, such as gaps in the grating.

We therefore believe that our results bridge the usually separate research areas of temporal and spatial vision by showing how spatial vision emerges over time, which suggests that masking cannot be explained solely by temporal explanations. Spatial vision needs to incorporate temporal aspects, and temporal vision needs to incorporate spatial ones (Herzog, 2007). Our model suggests that although complex grouping mechanisms seem to be at work, the experimental findings can be explained without explicitly implementing such grouping mechanisms and instead can be explained by relying on basic neural lateral interactions only.

To proceed to more complex spatial processing, our simulations show that contextual elements may affect target processing, following the same principles as described for irregularities in the masking gratings (Figure 5; Herzog, Schmonsees, & Fahle, 2003). Single collinear lines deteriorate performance because their neural activity interferes with that of the vernier. Contextual gratings, which contain multiple contextual lines, have less impact because their inner elements are inhibited before they can interfere with the vernier. This again indicates that regularity plays an important role in masking and contextual modulation. Our model suggests that the inner parts of regular structures are strongly inhibited. It is interesting to note that there are experimental data providing evidence for such inhibition: A vernier offset inside a grating is much more difficult to discriminate than one at the edges of the grating (Malania, Herzog, & Westheimer, 2007; Sharikadze, Fahle, & Herzog, 2005).

These findings are not only of theoretical interest but can also aid in constructing masks of varying strengths. For example, as a rule of thumb, to create a strong mask one should include irregular elements such as gaps (Herzog et al., 2001) or contextual elements (Herzog, Schmonsees, & Fahle, 2003) near the target.

Temporal Aspects

Whereas we primarily focused on modeling spatial processing, we found that the model could also explain the most prominent temporal aspects of masking, such as B-type masking and common onset masking. Although it has been shown that these aspects can be explained with fairly simple models (Bridgeman, 1978; Francis & Cho, 2005; Francis & Hermens, 2002), often more complex mechanisms were assumed to be necessary, such as dual-channel interactions (Bachmann, 1994; Breitmeyer & Ganz, 1976; Öğmen, 1993) or recurrent processing (Di Lollo et al., 2000; Enns & Di Lollo, 2000). Our results show once more that many temporal aspects of masking can be explained with a simple model structure.

An analysis by Francis (2000) revealed that although models of masking differ in their structure (Anbar & Anbar, 1982; Bridgeman, 1971; Francis, 1997), they in fact rely on a common mechanism that he termed *mask blocking*. In mask blocking, B-type masking occurs when the target elicits a relatively strong response with respect to the mask. Such a strong target response prevents the mask signals from influencing the target signals for a simultaneous presentation of the target and mask. For the mask to affect

the target, the strong signals of the target have to decay. Hence, the mask can only be effective after a delay between the target and the mask. Because at even later presentations of the mask, the mask will come in too late to have an effect, such a scheme results in strongest masking for an intermediate SOA, that is, a B-type mask. It is difficult to prove that our model is a mask blocking system; however, it reveals similar characteristics.

Alternative Approaches and Models

Fourier analysis. Our experiments demonstrate that minor changes in the spatial arrangement of the masks can lead to strong performance differences when these changes modify the perceptual organization of the mask. Particularly, the regularity of the mask seems to be a determining factor for the shine-through effect. Our model provides an explanation as to why a regular mask yields weak masking.

An alternative approach to explain the effects of mask regularity may be based on a Fourier analysis of the stimuli. The rationale is that a regular mask, such as the 25-element grating, activates a smaller range of frequency detectors, concentrated at lower frequencies, than does an irregular mask, such as the gap grating. For this reason, irregular masks that activate a larger range of frequency detectors, comprising also higher frequencies, are more likely to inhibit those detectors that are important for target processing (Weisstein, Harris, Berbaum, Tangney, & Williams, 1977), which could result in an increase of offset discrimination thresholds. It should be mentioned that although such a Fourier approach could possibly explain spatial aspects, it does not address the temporal issues involved in masking.

To test such a Fourier approach, we converted several of our masks related to the shine-through effect to Fourier space. We then quantified the difference between the pairs of masks in Fourier space by applying an appropriate distance measure.

In particular, if a mask i was given by a two-dimensional, real-valued function, $M_i(\mathbf{x})$ of position x, the Fourier transform $F[\ldots]$ gave a complex function, $\tilde{M}_i(k) = F[M_i(\mathbf{x})]$, of frequency **k.** The difference $\Delta\phi_{ij}$ between two Fourier-transformed masks i and j was quantified as an integral over their absolute values, normalized by the overall luminance of the original stimulus. The corresponding equations read

$$\Delta\phi_{ij} = \int_{\mathbf{k}} d\mathbf{k}\, D_{ij}(\mathbf{k})$$

$$D_{ij}(\mathbf{k}) = \left| \frac{|\tilde{M}_i(\mathbf{k})|}{\int_{\mathbf{x}'} d\mathbf{x}' M_i(\mathbf{x}')} - \frac{|\tilde{M}_j(\mathbf{k})|}{\int_{\mathbf{x}'} d\mathbf{x}' M_j(\mathbf{x}')} \right| \tag{10}$$

This procedure did not show obvious relations between the differences $\Delta\phi_{ij}$ of any two masks considered in Fourier space and the differences in the thresholds (θ). This is exemplified in Figure 12 for four selected masks, suggesting that simply comparing the Fourier spectra of the masks cannot explain the experimental findings (see also Hermens & Herzog, 2007).

Alternative models of masking. Several models have been proposed to explain visual masking. In Table 1, we provide a possibly incomplete overview. One important characteristic to distinguish the various models is the number of spatial dimensions in which the input can be coded. For example, the model by Weisstein (1968) represents the target as input to one target detector neuron, which means that zero spatial dimensions of the stimuli are coded.

Bridgeman's (1978) model allows for a vector input that describes the stimulus in one dimension. The model, in principle, can be extended to handle two-dimensional input, although it is not clear at this point exactly how this should be done. The weighting of the inhibition between the neurons depends on the distance to the receptor neuron. In two dimensions, one has to decide how this distance is computed (city block, Euclidean, etc.). Moreover, we recently showed that Bridgeman's (1978) model has difficulties explaining the effects of the number of elements of a grating mask (5 or 25) as well as the effect of two gaps in the 25-element mask (Figure 7 of Hermens & Ernst, 2007). The difference in simulation outcomes of the Bridgeman (1978) model and our model, which both apply lateral inhibition, is probably the result of the interaction kernels used. Our model applies both an excitatory kernel and an inhibitory kernel that interact by means of two separate layers, whereas the model by Bridgeman only uses lateral inhibition within one layer. It is probably the combination of the two kernels that allows for a precise explanation of the effects of the size of a grating and the effects of the gaps.

A few models, such as the models by Francis (1997) and by Öğmen (1993) allow for coding in two dimensions. However, these models are so complex that simplifications to the models' structure are often needed to be able to perform the simulations on a standard computer (see Francis, 1997). Some of these simplifications involve representing the input structure in a single dimension. However, because the models, in principle, can handle two-dimensional input, we classified them as two-dimensional models.

A model that can probably explain most of our current results is the model by Zhaoping (1999, 2000, 2003), which is structurally very similar to the model presented here. However, because of the collinear facilitation used in that model, it probably cannot explain the observed collinear suppression (Figure 5 in Herzog, Schmonsees, & Fahle, 2003) that is captured by our model. Moreover, the two models differ in their complexity. For example, Zhaoping's (1999, 2000, 2003) model uses different types of cells, such as orientation-sensitive cells, whereas the neurons in our model are not selective to orientation. We chose to build a model that is as simple as possible, to be able to clearly work out the important mathematical mechanisms (see Appendix) while keeping the number of free parameters to a minimum (avoiding overfitting of the data).

Spatial aspects explained by alternative models. One aspect of visual backward masking that has been modeled with several models is the effect of the spatial distance between the target and a metacontrast mask. This has been simulated with the lateral inhibition model (Bridgeman, 1978), the boundary contour model (Francis, 1997), the efficient masking model (Francis, 2003), and the dual channel model (Breitmeyer & Ganz, 1976).

At least two models have been applied to other spatial aspects as well. The lateral inhibition model (Bridgeman, 1978) was used to simulate the effects of the size of the mask and the difference between spatially overlapping and nonoverlapping masks, in addition to the spatial separation. The boundary contour system (Francis, 1997) was also shown to explain the influence of the distribution of the contour of the mask. Our current simulations

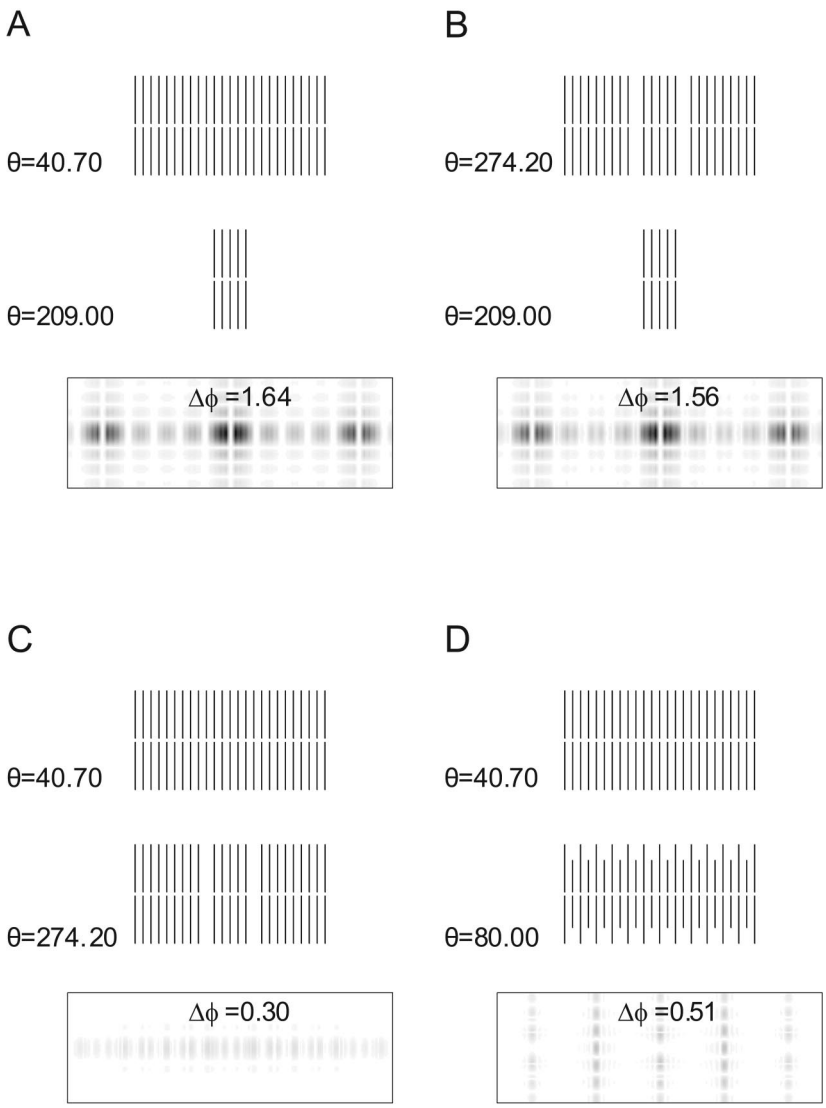

Figure 12. For four pairs of masks, the difference in Fourier power spectra ($\Delta\phi$) was computed and compared with the psychophysical thresholds (θ). Each subgraph depicts a mask *i* (top third of each panel) and a mask *j* (middle of each panel), together with the absolute difference in Fourier power (lower third of each panel). Note that the difference in power for masks pairs C and D is very small, resulting in only faint traces in the difference spectrum. In A–C, we compare the masks with 25 elements, 5 elements, and the 25 elements with gaps to each other. In D, we compare the 25-element mask with a mask comprising alternating short and long elements (Hermens & Herzog, 2007). In subgraphs A and D, the difference in thresholds seems to be related to $\Delta\phi_{ij}$ (i.e., $\Delta\phi_{ij}$ is large when the difference in thresholds is large and small when the difference in thresholds is small). Subgraphs B and C, however, clearly demonstrate the opposite relation.

add to this work by providing an explanation of a broad set of results involving both spatial and temporal aspects.

Limitations of the Current Approach

Energy summation. One important limitation of the current model concerns the summation of target-related and mask-related energy. Activation related to both stimuli enters the same layers, which results in a summation of the energy. As a consequence, the model incorrectly predicts that if the target is followed by a spatially overlapping regular mask, the target will be better visible than if presented without a mask.

This is an important problem that is likely to show up for all modeling of spatially overlapping pattern masks in which the target and the mask are processed in the same channel. It is therefore not surprising that hardly any modeling of pattern masking was performed up to now.

Although there are several possible ways to solve the energy summation problem, such as the use of reset signals (Francis, 1997; Öğmen, 1993) or gain control (Grossberg, 1982), they all

Table 1
Comparison of Different Models of Visual Masking by Number of Spatial Dimensions in Which Input Can Be Coded and Their Complexity

Model	Spatial dimensions	Complexity
Anbar & Anbar (1982)	0	Simple (lateral inhibition, stimulus decay)
Bachmann (1994)	0	Two channels
Breitmeyer & Ganz (1976)	0	Two channels
Di Lollo et al. (2000)	0	Recurrent connections
Francis & Cho (2005)	0	Simple (mask blocking)
Francis (2003)	0	Simple (efficient masking)
Weisstein (1968)	0	Two channels
Bridgeman (1978)	1	Simple (lateral inhibition)
Bugmann & Taylor (2005)	2	Five layers, complex, computationally demanding
Francis (1997)	2	Six layers, complex, computationally demanding
Öğmen (1993)	2	Two channels, complex, computationally demanding
Zhaoping (2003)	2	Two layers, orientation sensitive cells
Current model	2	Two layers, simple structure, computationally feasible

increase the complexity of the model. Because we tried to keep the model as simple as possible, we decided not to incorporate these mechanisms in the current version.

Long-range contextual effects. Our model incorrectly predicts that contextual elements only have an effect when they are relatively close to the mask (Herzog, Schmonsees, & Fahle, 2003). To incorporate the observed long-distance effects, long-range neural connections could be introduced.

A related issue involves the scaling of the stimuli. If the target and the mask are simply increased in size, this hardly affects masking (Figure 4d, Herzog & Fahle, 2002), which is in contradiction to the model's predictions. The model fails at this point, because only one fixed set of kernel widths is used. A more complex version of the model could incorporate multiple sets of kernel widths or adaptive kernel widths.

Intersecting masking curves. Francis and Herzog (2004) have shown that models that rely on mask blocking predict that for all SOAs, B-type masking is weaker than A-type masking when the target and the task remain constant (whereas the mask changes across conditions). For example, vernier thresholds are predicted to be lower for B-type masking than for A-type masking for all SOAs. From this, it follows that A- and B-type masking curves must never intersect. At the same time, Francis and Herzog (2004) showed that this prediction can be violated experimentally: Masking curves can intersect. Up to this point, no simulation with our model has ever shown intersecting masking curves. It seems, therefore, that our model shares a shortcoming with all other quantitative models considered so far. Because our model incorporates spatial aspects, this indicates that adding spatial aspects to a temporal model alone is not sufficient to obtain intersecting masking curves (for a discussion, see Francis, 2007). Therefore, the solution might lie in postulating at least two separate processing mechanisms (Reeves, 1982).

Conclusion

We demonstrated that many effects of visual masking could be explained with a relatively simple neural network model. These results aid our understanding of visual information processing in general by showing why temporal aspects are important for spatial vision and spatial aspects for temporal vision—a relationship that seems to have been largely neglected in the past. We argue therefore that both aspects of vision should be considered in one model. Practically, our results may guide the construction of masks used as a tool in many research areas.

References

Anbar, S., & Anbar, D. (1982). Visual masking: A unified approach. *Perception, 11,* 427–439.

Bachmann, T. (1994). *Psychophysiology of visual masking: The fine structure of conscious experience.* Commack, NY: NOVA Science Publishers.

Bischof, W., & Di Lollo, V. (1995). On motion and metacontrast with simultaneous onset of the stimuli. *Journal of the Optical Society of America: A. Optics and Image Science, 12,* 1623–1636.

Breitmeyer, B. G., & Ganz, L. (1976). Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. *Psychological Review, 83,* 1–36.

Breitmeyer, B. G., & Öğmen, H. (2000). Recent models and findings in visual backward masking: A comparison, review, and update. *Perception & Psychophysics, 62*(8), 1572–1595.

Breitmeyer, B. G., & Öğmen, H. (2006). *Visual masking: Time slices through conscious and unconscious vision.* New York: Oxford University Press.

Bridgeman, B. (1971). Metacontrast and lateral inhibition. *Psychological Review, 78*(6), 528–539.

Bridgeman, B. (1978). Distributed sensory coding applied to simulations of iconic storage and metacontrast. *Bulletin of Mathematical Biology, 40,* 605–623.

Bridgeman, B. (2001). A comparison of two lateral inhibitory models of metacontrast. *Journal of Mathematical Psychology, 45,* 780–788.

Bugmann, G., & Taylor, J. G. (2005). A model of visual backward masking. *Biosystems, 79,* 151–158.

Cho, Y. S., & Francis, G. (2005). The highs and lows of temporal integration in backward masking [abstract]. *Journal of Vision, 5,* 763a.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General, 129,* 481–507.

Di Lollo, V., Von Mühlenen, A., Enns, J. T., & Bridgeman, B. (2004). Decoupling stimulus duration from brightness in metacontrast masking: Data and models. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 733–745.

Duangudom, V., Francis, G., & Herzog, M. H. (2007). What is the strength of a mask in visual metacontrast masking? *Journal of Vision, 7,* 1–10.

Enns, J. T. (2004). Object substitution and its relation to other forms of visual masking. *Vision Research, 44,* 1321–1331.

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences, 4,* 345–352.

Francis, G. (1997). Cortical dynamics of lateral inhibition: Metacontrast masking. *Psychological Review, 104,* 572–594.

Francis, G. (2000). Quantitative theories of metacontrast masking. *Psychological Review, 107,* 768–785.

Francis, G. (2003). Developing a new quantitative account of backward masking. *Cognitive Psychology, 46,* 198–226.

Francis, G. (2007). What should a quantitative model of masking look like and why would we want it? *Advances in Cognitive Psychology, 3,* 21–31.

Francis, G., & Cho, Y. (2005). Computational models of visual masking. In H. Öğmen & B. G. Breitmeyer (Eds.), *The first half second: The microgenesis and temporal dynamics of unconscious and conscious visual processes.* Boston, MA, The MIT Press.

Francis, G., & Hermens, F. (2002). Comment on: Competition for consciousness among visual events: The psychophysics of reentrant visual processes, by Di Lollo, Enns, and Rensink (2000). *Journal of Experimental Psychology: General, 131,* 590–593.

Francis, G., & Herzog, M. H. (2004). Testing quantitative models of backward masking. *Psychonomic Bulletin and Review, 11,* 104–111.

Grossberg, S. (1982). Why do cells compete? Some examples from visual perception. *The UMAP Journal, III,* 103–121.

Growney, R. (1978). Metacontrast as a function of the spatial frequency composition of the target and mask. *Vision Research, 18,* 1117–1123.

Growney, R., Weisstein, N., & Cox, S. (1977). Metacontrast as a function of spatial separation with narrow line targets and masks. *Vision Research, 17,* 1205–1210.

Hermens, F., & Ernst, U. (2007). Visual backward masking: Modeling spatial and temporal aspects. *Advances in Cognitive Psychology, 3,* 93–105.

Hermens, F., & Herzog, M. H. (2007). The effects of the global structure of the mask in visual backward masking. *Vision Research, 47,* 1790–1797.

Herzog, M. H. (2007). Spatial processing and visual backward masking. *Advances in Cognitive Psychology, 3,* 85–92.

Herzog, M. H., Dependahl, S., Schmonsees, U., & Fahle, M. (2004). Valences in contextual vision. *Vision Research, 44,* 3131–3143.

Herzog, M. H., Ernst, U., Etzold, A., & Eurich, C. (2003). Local interactions in neural networks explain global effects in the masking of visual stimuli. *Neural Computation, 15,* 2091–2113.

Herzog, M. H., & Fahle, M. (2002). Effects of grouping in contextual modulation. *Nature, 415,* 433–436.

Herzog, M. H., Fahle, M., & Koch, C. (2001). Spatial aspects of object formation revealed by a new illusion, shine-through. *Vision Research, 41,* 2325–2335.

Herzog, M. H., Harms, M., Ernst, U., Eurich, C., Mahmud, S., & Fahle, M. (2003). Extending the shine-through effect to classical masking paradigms. *Vision Research, 43,* 2659–2667.

Herzog, M. H., & Koch, C. (2001). Seeing properties of an invisible object: Feature inheritance and shine-through. *Proceedings of the National Academy for Science, USA, 98,* 4271–4275.

Herzog, M. H., Schmonsees, U., Boesenberg, J. M., Mertins, T., & Fahle, M. (2007). *Grouping in the shine-through effect.* Manuscript in preparation.

Herzog, M. H., Schmonsees, U., & Fahle, M. (2003). Collinear contextual suppression. *Vision Research, 43,* 2915–2925.

Hogben, J. H., & Di Lollo, V. (1984). Practice reduces suppression in metacontrast and in apparent motion. *Perception & Psychophysics, 35,* 441–445.

Malania, M., Herzog, M. H., & Westheimer, G. (2007). Grouping of contextual elements that affect vernier thresholds. *Journal of Vision, 7,* 1–7.

Öğmen, H. (1993). A neural theory of retino-cortical dynamics. *Neural Networks, 6,* 245–273.

Palmer, S. E., Brooks, J. L., & Nelson, R. (2003). When does grouping happen? *Acta Psychologia, 114,* 311–330.

Parlee, M. B. (1969). Visual backward masking of a single line by a single line. *Vision Research, 9,* 199–205.

Reeves, A. (1982). Metacontrast U-shaped functions derive from 2 monotonic processes. *Perception, 11,* 415–426.

Sekuler, R. W. (1965). Spatial and temporal determinants of visual backward masking. *Journal of Experimental Psychology, 70,* 401–406.

Sharikadze, M., Fahle, M., & Herzog, M. H. (2005). Attention and feature integration in the feature inheritance effect. *Vision Research, 45,* 2608–2619.

Sturr, J. F., Frumkes, T. E., & Veneruso, D. M. (1965). Spatial determinant of visual masking: Effect of mask size and retinal position. *Psychonomic Science, 3,* 327–328.

Ventura, J. (1980). Foveal metacontrast: I. Criterion content and practice effects. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 473–485.

Wehrhahn, C., Li, W., & Westheimer, G. (1996). Patterns that impair discrimination of line orientation in human vision. *Perception, 25,* 1053–1064.

Weisstein, N. (1968). A Rashevsky–Landahl neural net: Simulation of metacontrast. *Psychological Review, 75,* 494–521.

Weisstein, N., Harris, C. S., Berbaum, K., Tangney, J., & Williams, A. (1977). Contrast reduction by small localized stimuli: Extensive spatial spread of above-threshold orientation-selective masking. *Vision Research, 17,* 341–350.

Werner, H. (1935). Studies on contour: I. Qualitative analysis. *American Journal of Psychology, 47,* 40–64.

Westheimer, G. (1967). Spatial interaction in human cone vision. *Journal of Physiology, London, 190,* 139–154.

Williams, M. C., & Weisstein, N. (1981). Spatial frequency response and perceived depth in the time-course of object superiority. *Vision Research, 21,* 631–646.

Williams, M. C., & Weisstein, N. (1984). The effect of perceived depth and connectedness of metacontrast functions. *Vision Research, 24,* 1279–1288.

Wilson, H. R., & Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik, 13,* 55–80.

Zhaoping, L. (1999). Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proceedings of National Academy of Science, USA, 96,* 10530–10535.

Zhaoping, L. (2000). Pre-attentive segmentation in the primary visual cortex. *Spatial Vision, 13,* 25–50.

Zhaoping, L. (2003). V1 mechanisms and some figure-ground and border effects. *Journal of Physiology, Paris, 97,* 503–515.

(*Appendix follows*)

Appendix

Appendix Details of the Model, Its Parameters, and Its Dynamics

General Setup

For most of our simulations, we presented the stimuli on a visual field that was 6,000 arc seconds (100 arc minutes) in width and 2,800 arc seconds (46.7 arc minutes) in height (a larger field was used for simulating the effects of contextual elements). For the numerical approximation of integrals and for the representation of the stimulus, we used a spatial discretization of 20 arc seconds. In the input map, $S(\mathbf{x},t)$, the pixels belonging to the stimulus were coded as 1, whereas background pixels were set to 0 unless stated otherwise.

Table A1 gives an overview of the parameters used in the model. These parameters are the same as those used by Herzog, Ernst, et al. (2003) in their simulations of a one-dimensional version of the neural population model, except for a few minor changes. First, the interaction weights were reduced from 1.0 to 0.5. Second, for the numerical approximation of the solutions of the differential equations, a first-order Euler method was used, instead of the computationally more demanding fourth-order Runge-Kutta method used before. Third, the size of the time steps in the simulations was set to 2/3 ms, which allowed faster simulations. We realize that with this relatively large step size, a precise approximation of the solution of the differential equations cannot be obtained. However, because the general behavior of the model remained intact for these settings (see the *Basic model behavior and the shine-through effect* subsection), we decided to favor fast computations over a precise approximation of the original differential equations. The parameter values were kept constant across all simulations.

Linking Hypothesis

Most models of masking (Bridgeman, 1978; Francis, 1997) integrate target-related evidence over some period of time. We chose, instead, to read out the target-related activity from the excitatory layer at one specific moment (80 ms after target onset), because this yielded slightly better model predictions (similar to the mechanism used by Di Lollo et al., 2000). This, however, does not mean that integration does not take place in our model. Because the neural activity is a function of the neural activity in the past, the model implicitly integrates activity over time. The main difference between a linking hypothesis that applies integration and one that does not is that the integrated activation is always a positive number, which increases over time, whereas the activity at a particular read-out time can approach zero if enough time is allowed to pass.

To determine whether the model predictions depend on whether an integration operation is used in the linking hypothesis, we performed simulations for both types of linking hypotheses. These simulations showed that predicted thresholds hardly depend on whether integration is used in the linking hypothesis. In general, we tested a variety of linking hypotheses that with a few exceptions yielded roughly comparable results.[2] An overview of these simulations and their results can be found at our website at http://lpsy.epfl.ch/Masking_Model/

Table A1
Parameters in the Model

Parameter	Description	Value
τ_e	Time constant excitatory neurons	16 ms
τ_i	Time constant inhibitory neurons	4 ms
s_e	Slope excitatory transfer function	3
s_i	Slope inhibitory transfer function	5.4
σ_e	Scale excitatory interaction kernel	150 arc seconds
σ_i	Scale inhibitory interaction kernel	250 arc seconds
w_{ee}, w_{ei}	Interaction weights $e\rightarrow e$, $e\rightarrow i$	0.5
w_{ii}, w_{ie}	Interaction weights $i\rightarrow i$, $i\rightarrow e$	−0.5
σ_E	Scale excitatory afferent kernel	100 arc seconds
σ_I	Scale inhibitory afferent kernel	200 arc seconds

Stability Analysis

Here, we show mathematically how our network model reproduces the experimentally observed efficacy of the different masks. For this purpose, we analyze the interplay between the typical length scales expressed in a specific mask with the length scales of both the feed-forward connections and the recurrent interactions in the model network. This analysis takes place in Fourier space, where the layout of a mask or the influence of the neuronal interactions are studied in terms of their modes k, each representing a certain length scale $x = 2\pi/k$ in retinal space.

Feed-Forward Input

The visual input is first passed through feed-forward filter kernel V, which resembles the on–off structuring of lateral geniculate nucleus receptive fields. According to the convolution theorem, in Fourier space this filter operation can be performed by a simple multiplication of the Fourier-transformed kernel with the Fourier-transformed visual input,

$$I(x,t) = (S * V)(x,t)$$
$$\rightarrow F[I](k,t) = 2\pi F[S](k,t)F[V](k,t). \quad \text{(A1)}$$

Note that the spatial coordinate undergoes a Fourier transformation, but not the temporal one. As an instructive example depicted in Figure A1 Panel A, we show the Fourier transform of a stimulus comprising an infinite number of aligned verniers with a periodic distance of $d = 200$ arc seconds. This stimulus S and its Fourier transform $F[S]$ are described by the following expressions:

[2] We thank Claus Bundesen for suggesting this extended analysis of possible linking hypotheses.

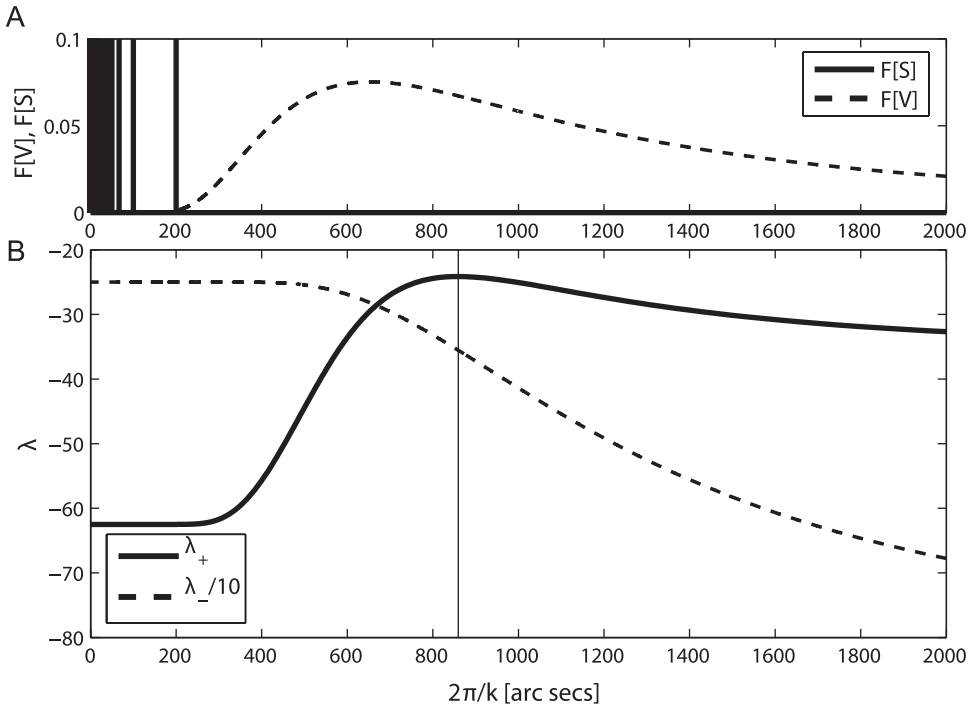

Figure A1. A: Fourier-transforms of an infinitely wide grating of aligned verniers $F[S]$ (solid lines) and of the feed-forward input kernel V, $F[V]$ (dashed lines), respectively. B: Characteristic exponents λ_+ (solid line) and λ_- (dashed line). The maximum of both exponents is marked with a vertical line. A and B are both shown in dependence of retinal space x, expressed in arc seconds.

$$S(x) = \sum_{j=-\infty}^{+\infty} \delta(x - jd)$$

$$F[S](k) = const. \sum_{j=-\infty}^{+\infty} \delta(k - 2\pi j/d). \qquad \text{(A2)}$$

That is, frequency components are at the basic frequency ($k = 2\pi/d$) and at higher harmonics but not at any other frequencies.

Note that in Figure A1 we plot the Fourier transform as function of $1/k$ rather than k. Thus, high spatial frequencies (periodicity with a short length scale) are to the left (close to zero) and low spatial frequencies (periodicity with a long length scale) are to the right. For the sake of comparison, we plot in the same graph the Fourier transform of the feed-forward kernel $F[V]$, using the parameters chosen for our simulations. A comparison of the two curves in Figure A1 with Equation A1 shows that the chosen feed-forward kernel actually suppresses all frequency components with a periodicity of less than $d = 200$ arc seconds (i.e., all higher frequencies). Therefore, for a stimulus that is composed of a regular arrangement of aligned verniers at $d = 200$ arc seconds, mainly its envelope is transferred by the feed-forward connections as an input $I(x, t)$ to the two neuronal layers, while the interior of regular structure is largely suppressed.

Recurrent Interactions

Once a filtered input, $I(x,t)$, is given to the network, activity is dynamically exchanged between the two neuronal layers. The nonlinearity of the system excludes an analytical solution of the response to arbitrary spatio-temporal stimuli as the ones used in the experiments. Instead, we use a linear stability analysis to investigate some of the basic properties of our neuronal network, which dominate the processing of complex, time-varying stimuli.

For this purpose, we analyze how the network reacts to arbitrary perturbations (i.e., feed-forward inputs), when it is in a spatially homogeneous steady state. Such a steady state is characterized by a spatially constant activation of all excitatory and inhibitory neuronal populations, $A_e(x, t) = A_{e0}$ and $A_i(x, t) = A_{i0}$, in response to a constant input $I(x, t) =$ constant. With the parameters chosen for our model, this steady state is stable; that is, any transient perturbation in the input I will lead to an exponentially decaying response of the network, which after some time settles back into its steady state. In the following, we quantify how fast this decay will be, given a perturbation with a typical length scale k.

(*Appendix continues*)

We first express A_e and A_i as the steady activations A_{e0} and A_{i0} plus some (small but arbitrary) perturbations δA_e and δA_i:

$$A_e(x, t) = A_{e0} + \delta A_e(x, t),$$

and

$$A_i(x, t) = A_{i0} + \delta A_i(x, t).$$

Inserting these expressions into the differential equations describing the model's dynamics (Equations 1 and 2) and linearizing the gain functions h_e and h_i yields

$$\tau_e \frac{\partial}{\partial t}\delta A_e(x,t) = -\delta A_e(x,t) + s_e\{w_{ee}(\delta A_e * W_e)(x,t) + w_{ie}(\delta A_i * W_i)(x,t)\}$$

$$\tau_i \frac{\partial}{\partial t}\delta A_i(x,t) = -\delta A_i(x,t) + s_i\{w_{ei}(\delta A_e * W_e)(x,t) + w_{ii}(\delta A_i * W_i)(x,t)\}. \quad \text{(A3)}$$

These equations are transformed into Fourier space with respect to x (we denote the Fourier transform with a bar placed over the corresponding variable), obtaining the characteristic equation

$$\frac{\partial}{\partial t}\tilde{\boldsymbol{\delta \mathbf{A}}} = L\tilde{\boldsymbol{\delta \mathbf{A}}}, \quad \text{(A4)}$$

where $\tilde{\boldsymbol{\delta \mathbf{A}}} = \{\delta\tilde{A}_e, \delta\tilde{A}i\}$ denotes the vector composed of the Fourier-transformed perturbations, and L describes the Jacobi-matrix of this system of partial differential equations with the corresponding coefficients

$$L = \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix} = \begin{bmatrix} \frac{1}{\tau_e}(2\pi s_e w_{ee}\tilde{W}_e(k) - 1) & \frac{1}{\tau_e}(2\pi s_e w_{ie}\tilde{W}_i(k)) \\ \frac{1}{\tau_i}(2\pi s_i w_{ei}\tilde{W}_e(k)) & \frac{1}{\tau_i}(2\pi s_i w_{ii}\tilde{W}_i(k) - 1) \end{bmatrix}. \quad \text{(A5)}$$

The eigenvalues of L are the characteristic exponents $\lambda_{\pm}(k)$ of our dynamic system that read

$$\lambda_{\pm}(k) = \pm\sqrt{\frac{(l_{11} + l_{22})^2}{4} - l_{21}l_{12} - l_{11}l_{22}} + \frac{l_{11} + l_{22}}{2}. \quad \text{(A6)}$$

These characteristic exponents determine the network behavior in the vicinity of its steady state: Positive exponents for a mode k would signal that perturbations with a length scale of $x = 2\pi/k$ grow exponentially with time, actually destabilizing the steady state. Negative exponents signal an exponential decay of the corresponding mode in a perturbation, whereas exponents with an imaginary part reveal that the system tends to develop oscillations.

In Figure A1 Panel B, we show the two characteristic exponents in dependence on the spatial scale of an arbitrary perturbation in the system. First, we observe that the exponents are both negative for all x, confirming that the system is inherently stable: In the regime in which our linear stability analysis is valid, recurrent excitation never leads to activity exponentially increasing over all bounds. Thus, the maximum exponent determines which perturbation decays slowest. Second, we realize that the exponent λ_- is, for $2\pi/k > 700$, significantly smaller than λ_+. While relaxing into its steady state, the system's dynamics will therefore be dominated by the exponent λ_+, which has a maximum at approximately $x \approx 860$ arc seconds.

In the final step of our analysis, we consider the transient presentation of a mask pattern as a spatio-temporal perturbation of our network. This means that features of a mask with a typical length scale of 860 arc seconds will dominate the network's activity. For example, the mask with 5 elements or the 25-element mask with the gaps, inserted at the element positions ± 3, have discontinuities at roughly this length scale. The outer elements of a 5-element grating, or the gaps, are separated by a distance of 800 arc seconds and therefore are dominant in the mask's representation within the network activation. Consequently, stimuli lying in between these dominant features, such as the remaining activity from the vernier, are rigorously suppressed. In this way, the stability analysis confirms the intuitive understanding of the network's behavior presented in the *Results* section.

Received May 8, 2007
Revision received November 1, 2007
Accepted November 2, 2007 ■