APA Type
1(1)

Spatial grouping determines temporal integration

Frouke Hermens, Frank Scharnowski & Michael H. Herzog

Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL)

To make sense out of a continuously changing visual world, features have to be integrated across space and time. Despite more than a century of research, the mechanisms of features integration are still a matter of debate. To examine how temporal and spatial integration interact, we measured the amount of temporal fusion (a measure of temporal integration) for different spatial layouts. We found that *spatial* grouping by proximity and similarity can completely block *temporal* integration. Computer simulations with a simple neural network capture these findings very well, suggesting that the proposed spatial grouping operations may occur already at an early stage of visual information processing.

Keywords: Temporal integration, Feature fusion, Spatial grouping, Gestalt laws, Modeling

When going to the cinema, movies are presented at a rate of 25 frames per second. Nevertheless, the sequences of single images of a film appear to us as temporarily coherent. To achieve coherence, first, objects in each frame have to be identified by spatial grouping of their elements. Second, for a continuous percept across time, features and objects have to be integrated across the individual frames. One could argue that we do not perceive the individual frames in a movie because the visual system is too sluggish to keep up with the rapid presentation. However, investigations have demonstrated that this is not the reason why information across frames is combined. In fact, the human brain is able to detect very fast temporal changes (Exner, 1875; Fahle, 1993; Sweet, 1953) (for a review, see Blake & Lee, 2005). This suggests that individual frames are integrated only if their elements can be grouped across frames. As soon as the single frames in a film are spatially unrelated, the movie fragments into a series of discontinuous individual frames (as in certain RSVPs, e.g., Potter & Fox, 2004).

Usually, spatial grouping and temporal integration are considered separately. However, this approach fails short as the example with the movies shows, indicating that spatial and temporal integration processes must be investigated jointly. For this purpose, we investigated how spatial grouping, namely grouping by proximity and similarity, interact with temporal integration. To measure the amount of temporal integration, we made use of a feature fusion paradigm. In feature fusion, stimuli are presented in rapid succession leading to the fusion of the features of the stimuli. If, for example, a red disc is immediately followed by a green disc, the colors fuse and the two discs are perceived as one yellow disc (Efron, 1967, 1973; Yund, Morgan, & Efron, 1983). Similarly, when a vernier is immediately followed by its cor-

responding anti-vernier, i.e. a vernier with opposite offset direction, observers cannot resolve the two verniers individually. Only one fused vernier is perceived with an almost aligned offset (Figure 1A; Herzog, Leseman, & Eurich, 2006; Herzog, Parish, Koch, & Fahle, 2003; Scharnowski, Hermens, Kammer, Öğmen, & Herzog, 2007). In both color and vernier fusion, the features from the two frames are combined in the fused percept: the *perceived* yellow disc is a combination of the red and green disc. Similarly, the vernier and the anti-vernier offsets almost cancel each other out and, hence, an almost aligned vernier is perceived on average. Interestingly, there is a slightly higher weight given to the second stimulus. If a red disc is followed by a green disc, the fused disc appears yellow with a slight greenish tone, whereas a green disc followed by a red disc looks slightly reddish. Likewise, integration is slightly more influenced by the anti-vernier on average (Figure 1A).

To investigate how temporal fusion is affected by spatial grouping, we determined how fusion between a vernier and an anti-vernier is affected by the presence of other elements in the display. In particular, we presented the anti-vernier either in isolation or in the context of an array of anti-verniers, as illustrated in Figure 1. We hypothesize that two effects can take place in when the anti-vernier is embedded in an array of anti-verniers (Figure 1B). According to the first hypothesis, the spatial grouping of the anti-verniers does not influence feature fusion. In this case, participants more often report an offset direction corresponding to that of the anti-vernier, just as when only the vernier and the anti-vernier are presented (Figure 1B, left). According to the second hypothesis, the anti-vernier is spatially grouped with the surrounding anti-verniers and temporal fusion is hindered. Earlier studies (e.g. Herzog, Fahle, & Koch, 2001) have shown that a vernier followed by a grating of several aligned verniers is well visible. This effect is known as the shine-through effect, because the vernier seems to shine through the successive grating (Herzog & Fahle, 2002). We expect for the second hypothesis that participants more often select the offset of the vernier, just as they would when performing the shine-through task

(Figure 1B, right).

In Experiment 1, we will show that the second hypothesis is true: spatial grouping dominates temporal fusion. If the anti-vernier is embedded in an array of anti-verniers, participants no longer fuse the vernier with the anti-vernier. This indicates that temporal integration cannot be considered without considering the spatial surround: Temporal and spatial aspects should be studied together (see also Herzog, 2007) In the experiments that follow, we will build on this result. First, we will show that the increased dominance of the vernier is not simply the consequence of the anti-vernier masking the vernier, but rather involves the combination of the features of the vernier and the anti-vernier as in feature fusion (Experiment 2). Second, we show that when spatial grouping of the anti-verniers is broken by increasing the spacing between the anti-verniers, temporal fusion re-occurs (Experiment 3). This shows that the trade-off between spatial grouping and temporal fusion can be varied by varying the spatial distance between the elements in the scene. The trade-off between temporal and spatial integration can also be varied by other factors, such as the similarity of the elements in the scene and their temporal proximity, as demonstrated in Experiment 4. Finally, we will show that interactions between temporal and spatial integration is not limited to the successive presentation of two stimuli, but extends to longer sequences (Experiments 5 and 6).

Our findings reveal a complex interplay between spatial grouping and temporal fusion, which, at first, seems difficult to explain. However, we will show that a simple neural network model with dynamical lateral inhibition and excitation can well explain our data. This is because the dynamics of the model suppresses the inside of regular structures. An anti-vernier embedded in anti-verniers forms a regular structure. This results in inhibition of the neural activation corresponding to the anti-vernier, which makes that the anti-vernier no longer has an effect on the preceding vernier. By increasing the distance between the anti-vernier and the surrounding anti-verniers, the inhibition of the anti-vernier is reduced, reducing the dominance of the preceding vernier. Similarly, by varying the temporal proximity and item similarity, the regularity of the structure in which the anti-vernier is embedded is varied. The model therefore suggests that the data can be understood from regularity processing. Moreover, it demonstrates that seemingly complex Gestalt effects, such as grouping by proximity and similarity, can be understood from basic low-level neural interactions, as long as neural processing is allowed to take some time.

In summary, measuring the extent to which features fuse across successively presented stimuli as a function of the spatial layout of the display offers a method to study quantitatively how spatial and temporal integration interact. Simulations with a neural network model suggest that these seemingly complex interactions can be explained by low-level neural interactions.
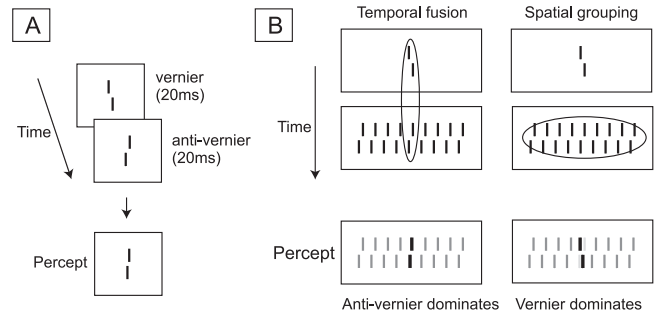

Figure 1. A. Feature fusion. A vernier is immediately followed by its anti-vernier, of which the offset is opposite to that of the vernier (in the experiments, the vernier offset direction was randomly chosen; if the vernier was offset to the left, the anti-vernier was offset to the right and vice versa). For short durations and small offsets, the vernier and anti-vernier are not perceived individually. Instead, participants perceive only one single fused vernier. When the offset sizes and durations of the vernier and anti-vernier are equal, the perceived offset of the fused vernier is dominated by the trailing anti-vernier. In that case, *on average*, participants report the anti-vernier offset more often than the vernier-offset. B. If the vernier is followed by an array of anti-verniers, two scenarios are possible. Either the vernier temporally fuses with the central anti-vernier (left), resulting in a dominance of the anti-vernier, as it is the case without the surround (see A). Alternatively, the anti-verniers spatially group without temporal fusion (right) resulting in a dominance of the vernier.

General Materials and Methods

Participants

The first two authors and university students (age ranging from 21 to 31 years) took part in the experiments. The observers were informed about the general purpose of the experiments. Before the experiments, they all signed an informed consent form. Observers were told that they could quit the experiment at any time. Only participants with normal or corrected-to-normal vision were included in the study. The visual acuity of each participant was tested by means of the Freiburg visual acuity test (Bach, 1996). To participate, observers had to score at least 1.0 (corresponding to 20/20) for at least one eye in this test. Participants other than the authors either received course credits or were paid for their participation.

Apparatus

Stimuli were presented on an X-Y display (HP 1334A or Tektronix 608 with P11 phosphor) controlled by a PC (Pentium 4 or Power Macintosh) via fast 16 bit D/A converters (dot pitch of 200μm, 1 MHz pixel rate). The refresh rate of the screen was 200 Hz. The luminance of the stimuli was set to approximately 80 cd/m^2 as determined with a Minolta LS-100 luminance meter. A dim background light illuminated the room at about 0.5 lux.

Stimuli

In all experiments, a vertical vernier served as the target. The vernier consisted of two vertical lines, each measuring 10' (arc minutes) in height separated by a 1' vertical gap (illustrated in Figure 1A), that were slightly offset horizontally. The horizontal offset direction (left/right) of the vernier was randomly selected on each trial. A mask consisting of one or more verniers was presented immediately after the vernier target. Some of the masks contained anti-verniers, which had the same parameters as the vernier, but an opposite offset direction. This means that if the offset direction of the target vernier was randomly chosen to be to the left on a particular trial, the anti-vernier in the mask was made offset to the right, and vice versa. The offset structure of the mask was varied across the experiments and will be discussed in the methods sections of each experiment. The target vernier and the mask were both presented for 20ms. Before the presentation of the stimuli, a fixation screen was shown for 1s, consisting of a fixation cross in the center and four lines in each of the corners.

Design

Each condition was tested twice in separate blocks ('blocked design') each consisting of either 40 or 80 trials. For the main conditions, we used 2x80 trials per condition, whereas for the parametric variations (varying the offset size, SOA, or duration), we used 2x40 trials per condition only. Within each experiment, the order of conditions was randomized across participants. After all conditions were presented once, the conditions were repeated in reverse order to counteract effects of practice and fatigue in the averaged data. The results of the two runs were collapsed into one mean.

Each new experiment contained at least one condition from an earlier experiment. This repeated condition provided a measure of the stability of the results across participants, because not all participants took part in all experiments.

Procedure

Participants observed the stimuli from a distance of 2 meters. A sequence of a target vernier and one or several masks were presented. When only the vernier and anti-vernier were presented, the task of the observers was to report the offset direction of the *fused* vernier by pressing the corresponding of two push buttons (participants were asked to report the position of the lower segment relative to the upper element). When the vernier was followed by a grating, observers were asked to report the offset of the central vernier element (this element was clearly visible because luminance summation of the vernier and central anti-vernier. Because of this, the center element appeared brighter than the surrounding elements). In both cases, we analyzed how often, observers responded in accordance with the vernier offset ("vernier dominance"). A vernier dominance larger than 50% indicates vernier dominance, a performance below 50% anti-vernier dominance, and a performance of 50% that both the vernier and the anti-vernier contribute equally strong to performance

on average. However, also in this latter case a very small vernier offset is perceived on each individual trial, but on average offsets balance each other out to 50%.

Before the actual experiment, participants received a few practice trials with auditory feedback. In these trials, the vernier was presented alone or it was followed by a grating consisting of 25 aligned verniers ('V' or 'V-25N', respectively). Once the participant performed well in the practice conditions, the condition in which a vernier was immediately followed by an anti-vernier ('V-AV') was presented to determine the individual offset size for each observer, which was used in the remainder of the experiment. For this, several offset sizes of the vernier and the anti-vernier were tested, and the offset size that yielded around 30% vernier dominance was chosen. This means that for the sequence of a vernier and an anti-vernier ('V-AV'), even though participants see only one fused vernier on each trial, they all chose the vernier offset direction in about 30% of the trials, and the offset direction of the anti-vernier in the remaining 70% of the trials. For some participants, we had to select a slightly smaller offset size to avoid motion percepts. For most participants, this procedure resulted in a vernier offset size of 40", whereas for a few participants, we used an offset of 30" or 50".

Modeling

To investigate the mechanisms underlying our experimental findings, we compared our experimental data with predictions from a two dimensional extension (Hermens, Luksys, Gerstner, Herzog, & Ernst, 2008) of a model introduced by Herzog and colleagues (2003), which was derived from a system initially proposed by Wilson and Cowan (1973).

Figure 2 shows the general setup of the model. The sequence of a target and a mask enters the excitatory and the inhibitory layer via a Mexican hat type input kernel (V). This input kernel highlights the edges of a homogeneous structure, while suppressing the inner structure. A similar operation is performed by the excitatory and inhibitory kernel, which further highlights the edges thereby inhibiting the interior of the homogeneous structure.

The neurons within the excitatory layer excite each other and neurons in the inhibitory layer. The amount of excitation that a neuron receives, depends on the mutual distance to the neuron that is exciting it. Neurons nearby excite each other more strongly than neurons at larger distances, mediated by the excitatory kernel W_e. Similarly, neurons in the inhibitory layer inhibit each other, but also neurons in the excitatory layer. The amount of inhibition again depends on the distance between the neurons as indicated by the inhibitory weighting kernel W_i. Because the inhibitory kernel is larger than the excitatory one, the combination of the two serves as an edge detector ($W_e + W_i$ gives a Mexican hat function). This edge detector is very sensitive to small inhomogeneities in a structure, as will be illustrated in the discussion of the results of Experiment 1.

More details about the model can be found in the appendix (also see Hermens et al., 2008; Herzog, Ernst, et al., 2003).
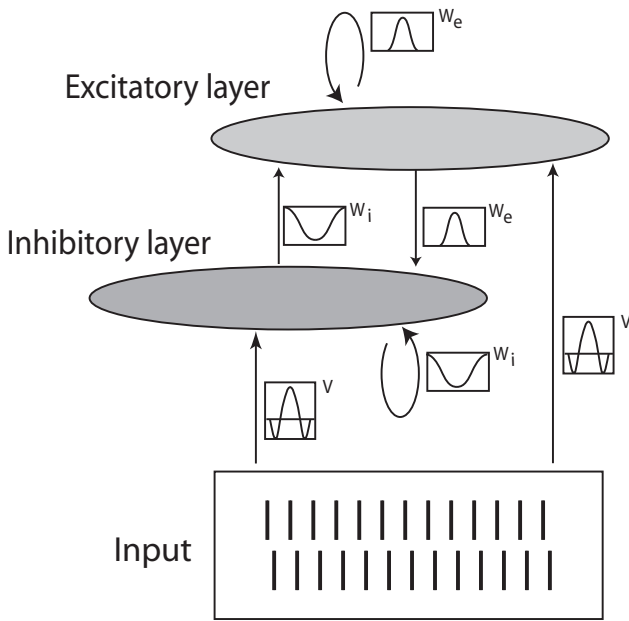
Figure 2. General setup of the neural network model. The input (e.g., a sequence of an image of a vernier and an anti-vernier) enters into the excitatory and the inhibitory layer via a Mexican hat type kernel (V). The activation of the layers is updated via kernels W_i and W_e. At the same time, the activation within both layers undergoes a passive decay (not depicted).

Experiment 1

If a vernier is immediately followed by its anti-vernier, and if their offsets and durations are sufficiently short, only one vernier is perceived, and the individual offsets are fused (see Figure 1 and Herzog et al., 2006; Herzog, Parish, et al., 2003). In Experiment 1, we show that spatial grouping can prevent the fusion of the vernier and the anti-vernier.

Methods

Nine observers took part in the experiment. The sequences of stimuli are illustrated at the bottom of Figure 3. In the condition labeled 'V', only the vernier was presented, whereas in the remaining conditions, the vernier target was followed by either its anti-vernier ('V-AV') or a grating. This grating either consisted of 25 anti-verniers ('V-25AV'), 25 aligned verniers ('V-25N'), 24 aligned verniers surrounding a central anti-vernier ('V-AV24N'), or 24 anti-verniers surrounding an aligned vernier ('V-N24AV'). Note that we use 'V' for a vernier, 'AV' for an anti-vernier, and 'N' for an aligned (i.e. Non-offset) vernier. Each condition was tested blockwise in two blocks of 80 trials.

Results and discussion

Figure 3 shows the results of Experiment 1. Participants could well discriminate the offset direction of a single vernier ('V'), indicated by a vernier dominance close to 100%. When the vernier was followed by its anti-vernier ('V-AV'), their offsets fused and the trailing anti-vernier dominated the percept (vernier dominance drops to 30%, indicating that participants more often report the offset of the anti-vernier; see also Herzog et al., 2006; Herzog, Parish, et al., 2003). This agrees with the common finding that backward masking is stronger than forward masking (Alpern, 1953; Bachmann & Allik, 1976; Breitmeyer & Öğmen, 2006; Growney, Weisstein, & Cox, 1977; Öğmen, Breitmeyer, & Melvin, 2003). In terms of neural processing, one can understand the dominance of the anti-vernier by assuming that neural signals decay over time. At the time the participant responds, the signal of the preceding vernier has decayed more than that of the trailing anti-vernier, causing the anti-vernier to dominate the percept.

When the anti-vernier of the 'V-AV' condition is surrounded by 12 anti-verniers ('V-25AV' condition) on each side, the vernier strongly dominates (performance is over 90%, indicating participants selected the offset of the vernier on almost each trial). Hence, which stimulus (the vernier or the anti-vernier) dominates, reverses due to the contextual elements. Perceptually, in the 'V-25AV' condition, the vernier appears to be superimposed on the anti-vernier grating and its offset can easily be discriminated. This agrees with the percept in the shine-through effect, in which a vernier followed by a grating of 25 aligned verniers seems to be superimposed on the grating (e.g. Herzog et al., 2001). We suggest that grouping of the central anti-vernier with the surrounding anti-verniers prevents feature fusion across time occurring for the 'V-AV' sequence. A comparison with the 'V-25N' condition, in which 25 aligned verniers follow the target vernier, shows that the embedded anti-vernier only weakly fuses with the preceding vernier. Vernier dominance is almost as high in the 'V-25AV' as in the 'V-25N' condition, which indicates that the offset of the anti-verniers in the '25AV' grating hardly affect the perceived offset of the central vernier. Although there is a significant difference in vernier dominance between the 'V-25AV' and 'V-25N' condition ($p < 0.001$), this difference is negligible when compared to that of the 'V-25AV' and the 'V-AV' condition.

One might argue that spatial low-pass filtering of the array of anti-verniers prevented the temporal fusion of the central anti-vernier with the preceding vernier. Such a low-pass filter might blur the grating so strongly that the fine structure inside the grating is erased. To exclude this possibility, we included a condition in which the vernier was followed by its anti-vernier embedded in 24 aligned verniers ('V-AV24N'). If blurring takes place for larger structures, this would also blur the signal of the anti-vernier in the context of the aligned verniers. As performance in the 'V-AV24N' condition is significantly below that of the 'V-25AV' condition ($t(8) = 8.19$, $p < 0.001$), it is unlikely that the high performance for the 'V-25AV' condition was due to blurring. Instead, it suggests that the visual system is very sensitive to small deviations in a homogeneous structure. In terms of grouping operations, the reduced vernier dominance in the 'V-AV24N' condition suggests that the central anti-vernier

is no longer completely grouped with its surround. Instead, at least a partial fusion with the preceding vernier seems to occur.

However, we need to exclude another possibility. It might be argued that rather than fusion between the central anti-vernier of the 'AV24N' with the preceding vernier, strong masking by the 'AV24N' grating caused the low performance in this condition. Possibly, the AV24N is a very strong backward mask, because of the local inhomogeneities close to the vernier position caused by the difference in offset direction between the central anti-vernier and the surround. It has been shown that inhomogeneities in the mask, such as gaps, strongly increase masking, so it might well be that this is also the case in the 'V-AV24N' condition (Herzog et al., 2001). Two results makes such a backward masking explanation is unlikely. First, no strong decrement in performance is found in the condition in which the target vernier is followed by an aligned vernier surrounded by anti-verniers ('V-N24AV'). In this condition, a similar inhomogeneity is present in the center. Nevertheless, hardly any masking is obtained, and masking is virtually the same as in the 'V-25N' condition ($t(8) = 1.48$, $p = 0.18$, two-tailed). Second, we will show in Experiment 2 that increasing the offset of the central anti-vernier increases its effect on performance. The larger the offset of the anti-vernier, the more often participants choose the offset of the anti-vernier, which indicates that fusion rather than masking takes place.
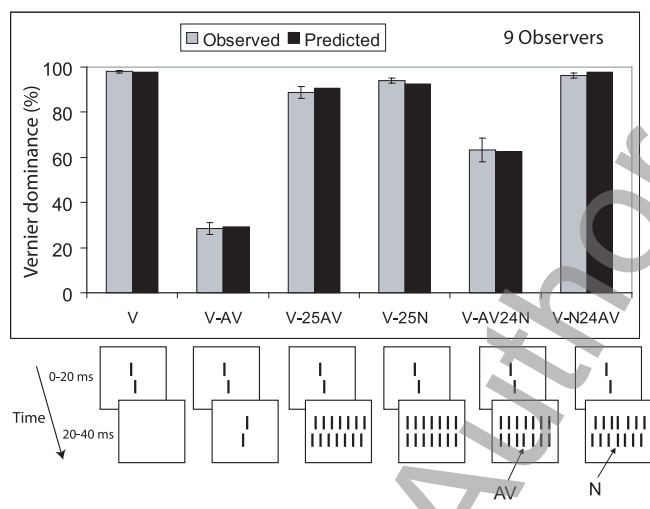

Figure 3. Mean percentage of trials in which participants reported an offset direction in accordance with the vernier ('vernier dominance') in Experiment 1 (gray bars), together with predictions by a neural network model (black bars). The sequences of stimuli are illustrated below the data plot. Note that for the purpose of illustration, only the seven central elements of the 25 elements of each grating are shown. Error bars show the standard error of the mean.

Finally, one may argue that participants might have used a cognitive strategy to reach high vernier performance in the 'V-25AV' condition. In such a strategy, participants look at the edges of the grating instead of looking at the center to determine the offset of the vernier (assuming they discov-

ered that the vernier offset was always opposite to that in the grating in the 'V-25AV' condition). To exclude this strategy, we asked five participants of Experiment 1 to perform an additional control condition. In this condition, the offset direction of the grating element was chosen at random on each trial (either the mask consisted of 25 verniers or 25 anti-vernier). Vernier dominance in this condition was at 91%. If participants would have followed such a cognitive strategy, vernier dominance would have been at 50% (average of 0% for the '25V' and 100% for the '25AV').

Model.

Quantitative predictions of vernier dominance were obtained by determining how well the activation in the excitatory layer of the model corresponded to the vernier or to the anti-vernier (for details, see the appendix). The black bars in Figure 3 show that the predicted vernier dominance closely matches the experimentally obtained one, for each of the conditions (a correlation between observed and predicted means across participants equal to 0.999). Note that for this excellent data fit, we used only two free parameters (see appendix). The remaining model parameters correspond to those used before (Hermens et al., 2008; Herzog, Ernst, et al., 2003).

Figure 4 illustrates how the neural network model explains three important conditions in Experiment 1. The activation of the excitatory layer is shown at different moments in the simulation, for three different sequences consisting of a target vernier and a mask. The masks in these sequences are the single anti-vernier ('V-AV'), the 25 anti-vernier grating ('V-25AV'), and the grating with the anti-vernier surrounded by 24 aligned verniers ('V-AV24N'). For the purpose of illustration, only the center 100 by 140 cells of the excitatory layer are displayed, which is the location where the vernier and the central anti-vernier are presented. The activation is shown immediately after vernier onset (3ms), just after mask onset (26ms), and just after mask offset (47ms). The activation after the presentation of the vernier is the same for the three masks ('3ms'). Just after the mask appears ('26ms'), the vernier offset direction, represented by the cells in the center, starts to change from the vernier offset to the anti-vernier offset for the 'V-AV' and the 'V-AV24N' conditions. However, no change in offset direction occurs for the 'V-25AV' mask. After mask termination ('47ms'), the neurons of the excitatory layer represent the offset of the anti-vernier for the 'V-AV' and 'V-AV24N' conditions. In the 'V-25AV' condition, an offset corresponding to that of the vernier is found.

The model predictions can be understood as follows. In the 'V-AV' condition, the anti-vernier dominates the final network activation, because, at the end of the simulation, the activation for the preceding vernier has decayed more than that of the trailing anti-venier (see also Scharnowski, Hermens, & Herzog, 2007). In the 'V-25AV' condition, the fine structure within the grating is suppressed by the combined effects of lateral inhibition and excitation, and only the edge elements show strong activation. Because the inner structure of the '25AV' grating is inhibited, there is no longer an

anti-vernier representation to fuse with the vernier representation. In the 'V-AV24N' condition, the model highlights the inhomogeneity caused by the difference between the center anti-vernier offset and the two neighboring aligned verniers. The central anti-vernier representation is therefore no longer inhibited and hence it fuses with the vernier.
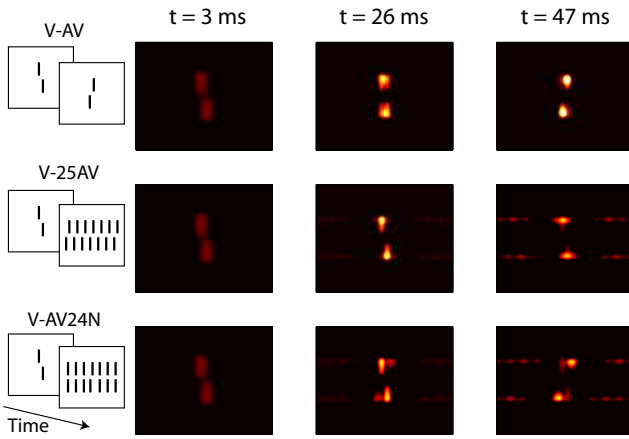

Figure 4. Activation of the center 100 by 140 neurons of the excitatory layer of the neural network model. The activation of the neurons is shown at different time points (t = 3ms, t = 26ms, t = 47ms; shown in the columns), for three different sequences of vernier target and mask (illustrated on the left).

Experiment 2

In Experiment 2 , we investigate the role of the central anti-vernier in the 'AV24N' grating in more detail. We determine whether this central anti-vernier fuses with the preceding vernier or whether it unspecifically masks the vernier.

Methods

Eight observers took part in Experiment 2. The top of Figure 5 shows the stimulus sequences. Either a vernier was followed by a single anti-vernier embedded in 24 neutral elements ('V-AV24N'), or it was followed by five anti-verniers embedded in 20 neutral elements ('V-5AV20N').

The offset size of the single anti-vernier or the five central anti-verniers was varied from 40, through 60, 80, to 120 arc seconds. The vernier offset was set to 40 arc seconds for all participants.

Each condition was tested blockwise in two blocks of 40 trials each.

Results and discussion

Figure 5 shows the results of Experiment 2. The dominance of the vernier decreases as the anti-vernier offset in the 'AV24N' grating increases. At large anti-vernier offsets, vernier dominance is below 50%, which indicates that for these offsets, the anti-vernier dominates performance. No

such decrease in vernier dominance is found for five central anti-verniers.

It might be argued that for small offsets of the anti-vernier, the vernier is well visible, which is reflected in a in high vernier dominance. The larger the offset of the central anti-vernier, the stronger it masks the vernier, until at some point, performance drops below 50%. This is when participants switch from reporting the vernier to reporting the offset of the anti-vernier more often. This is because the larger the offset of the anti-vernier, the easier it is to report its offset, resulting in high anti-vernier dominance. Such an explanation, however, becomes more difficult to hold when the condition with five central anti-verniers is taken into account. It is not clear why the mask should become stronger when there is just a single anti-vernier with an increasing offset, but no such increase of masking strength is taking place for five central anti-verniers.

In terms of fusion, the results can be understood as follows. The larger the offset of the single central anti-vernier, the more it will contribute to the perceived offset of the fused vernier and the lower vernier dominance is. The five central anti-verniers, however, group, thereby preventing fusion of the vernier and the anti-vernier. It is for this reason that vernier dominance remains high.

From the above, it is clear that the results can be explained both in terms of masking and in terms of fusion. We want to argue that the two explanations do not exclude each other, but that feature integration can be viewed as a form of specific masking. We will return to this issue in the general discussion.

From the model, the lack of fusion between the five offset anti-verniers and the vernier can be understood from the lateral inhibition of elements within a homogeneous substructure. The five central elements all have the same spatial offset, and this exerts a strong inhibition of information within the block of five anti-verniers. Only the edges of these five elements "survive". As in the case with the 25 anti-vernier grating, there is no activation to "overwrite" the vernier representation at the center. Vernier dominance occurs. Because we fixed the offset size of the target in the model, we were not able to produce any quantitative predictions for Experiment 2. However, for equal offset sizes of the vernier and the anti-verniers, the model accurately predicts the vernier dominance in both conditions.

Experiment 3

In Experiments 1 and 2, we found that a spatially grouped central anti-vernier could no longer fuse with the preceding vernier. From the Gestalt principle of proximity (for a recent review, see Palmer, Brooks, & Nelson, 2003), we expect that such spatial grouping of the anti-vernier elements only occurs for small distances between the anti-verniers. In Experiment 3, we increased the spacing between the anti-verniers to investigate the strength of grouping by spatial proximity with respect to fusion over time.
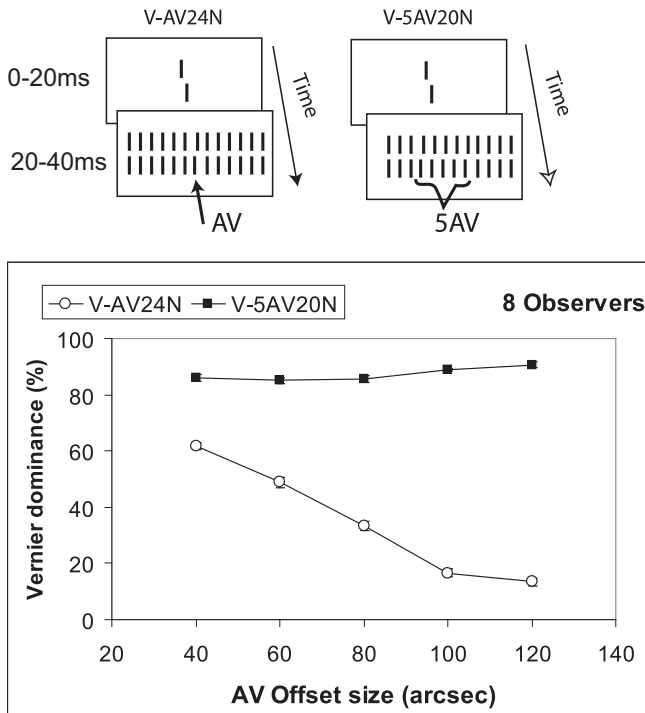
Figure 5. Vernier dominance as a function of anti-vernier offset size in Experiment 2. Above the data plot, the two stimulus sequences are illustrated. For the purpose of illustration, only the central 13 elements out of the 25 of each grating are shown. In most conditions, the symbols were larger than the error bars denoting the standard error of the mean.

Methods

Five participants took part in the experiment. They were presented with sequences of a vernier followed by a grating of 25 anti-verniers ('V-25AV'). The spacing between the elements of the '25AV' grating was varied from 200" (the same distance as in Experiment 1) to 1000". We included the 'V-AV' condition as a baseline to determine at which spacing spatial grouping is no longer preventing fusion over time. Each condition was presented blockwise in two blocks of 80 trials.

Results and discussion

Figure 6 shows the results of Experiment 3. The curve suggests that the transition from spatial grouping to temporal fusion is gradual. The larger the spacing between the elements of the '25AV' grating, the smaller the vernier dominance, which implies an increasing dominance of the anti-vernier. For a spacing of 600" or more, vernier dominance is almost the same as in the 'V-AV' condition. This suggests that from this spacing on, the anti-verniers are no longer spatially grouped, which allows fusion between the vernier and the central anti-vernier to take place. This result is in good agreement with the results of Westheimer and Hauske (1975), who showed the interference of lateral flanks

on vernier acuity is limited to a distance of about 600" (see also Malania, Herzog, & Westheimer, 2007).
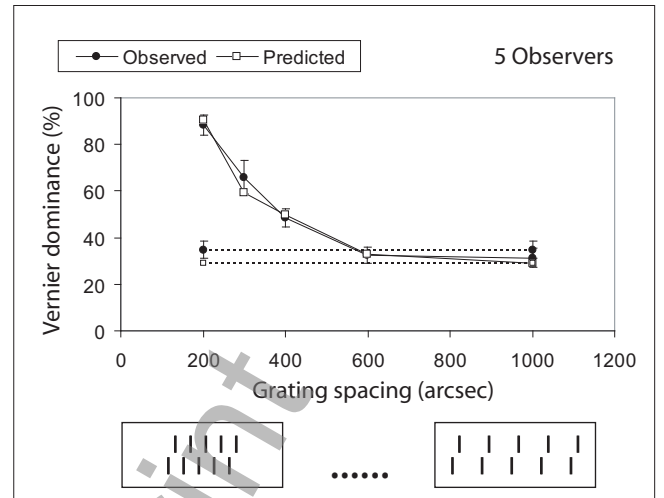

Figure 6. Vernier dominance as a function of the spacing between the anti-verniers in Experiment 3 (filled circles), together with the model predictions (open squares). The dotted lines show the performance for a sequence of vernier and anti-vernier only. The symbols at the smallest and the largest spacing are included to show the corresponding error bars, denoting the standard error of the mean.

The model predictions (Figure 6, filled circles) closely match the experimental findings (Figure 6, open squares)squares), which is reflected in a correlation between the predicted and observed means of 0.99. For this excellent fit, we used the same model parameters as in Experiment 1. On a conceptual level, the model prediction can be understood from the weaker interactions of neurons at a longer distance. If the distance between the mask's elements increases, the neurons that are sensitive to these elements become more remote from each other. At a larger distance, neurons interact less, which therefore results in less inhibition of the elements within the grating. Because the elements within the grating are less inhibited, they will affect the vernier activation more, and vernier dominance is predicted to decrease.

Experiment 4

Besides spatial proximity, item similarity (e.g., similarity in color, size, or orientation) is a well-known factor in Gestalt grouping (e.g. Palmer et al., 2003). Less well investigated is the effect of temporal proximity, i.e. the time between the presentation of the stimuli (however, see Oyama & Yamada, 1978). In Experiment 4, we use the feature fusion paradigm to study the relative importance of temporal proximity and item similarity in grouping. We presented a sequence of a vernier and an anti-vernier. In addition, we presented 24 spatially surrounding grating elements (Figure 7A). The temporal proximity of these elements with the vernier and the anti-vernier was varied by presenting the surround at different stimulus onset asynchronies (SOAs). The item similarity of

the surround was varied by using three different offset directions in the surround. Either the surround's elements had the same offset as the vernier, or the same as the anti-vernier or they were aligned (i.e., had no offset). We expect that if temporal proximity plays a role, the surrounding elements group the vernier at short SOAs and to the anti-vernier at longer SOAs. If item similarity is important, different surrounds will affect the fusion of the vernier and the anti-vernier differently.

Methods

Seven observers took part in the experiment. The stimulus sequences are illustrated in Figure 7A. We presented a sequence of a vernier and an anti-vernier, together with 24 surrounding verniers with an offset equal to that of the vernier ('24V'), the anti-vernier ('24AV'), or without an offset ('24N'). The surround was presented at different stimulus onset asynchronies (SOAs) with respect to the onset of the vernier, ranging from 0 to 20ms. For an SOA of 0ms, the 24 flankers were presented together with the vernier, whereas for an SOA of 20ms, the 24 flankers coincided with the anti-vernier. The experiment also included a baseline condition in which no surround was presented ('V-AV'). Each combination of surround offset and SOA was tested blockwise in two blocks of 40 trials each.

Results and discussion

The results show that both temporal proximity and item similarity strongly affect feature fusion (Figure 7B). Temporal proximity affects feature fusion, because at shorter SOAs the anti-vernier dominates (i.e., the earlier vernier is grouped with the flankers) and at longer SOAs the vernier does (i.e., the later anti-vernier is grouped with the flankers). Item similarity is important, because the curves for the three different offset directions in the surround are separate. Beside highly significant main effects of temporal proximity and item similarity ($F(2,12) = 25.1$; $p < 0.001$; $h_p^2 = 0.81$; and $F(4,24) = 81.8$; $p < 0.001$; $h_p^2 = 0.93$, respectively), there is a significant interaction between the two factors ($F = 9.43$; $p < 0.001$; $h_p^2 = 0.61$).

To understand how spatial grouping can explain the findings, consider the effects of the three surrounds at an SOA of zero. At this SOA, vernier dominance is highest for the anti-vernier surround, less for the non-offset vernier surround, and even less for the vernier offset surround. This can be understood by realizing that elements with the same offset direction group. The anti-vernier surround groups with the anti-vernier, which makes the effect of the anti-vernier on the percept small, and therefore vernier dominance is relatively high. The vernier surround groups with the vernier, and therefore the influence of the vernier on the percept is small, which results in a very low vernier dominance. The grouping by offset direction is very strong: Even if the surround is presented simultaneously with the vernier (SOA=0ms) the anti-vernier surround partially groups with the anti-vernier that follows, resulting in a vernier dominance as high as 40%.

A. Stimulus sequences
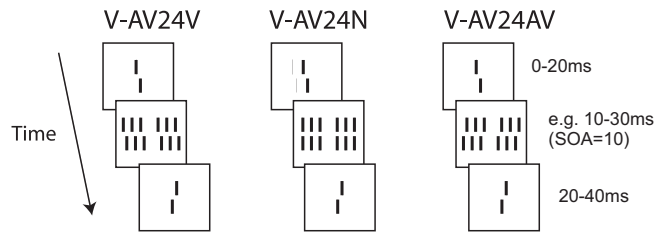
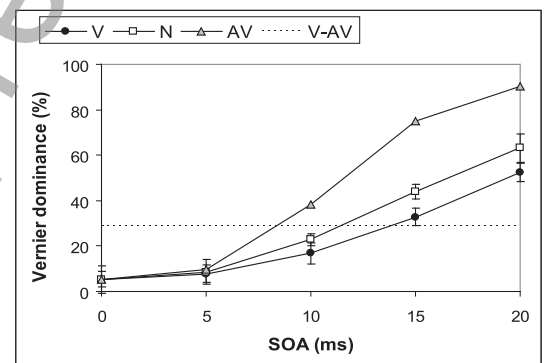### B. Experimental observations
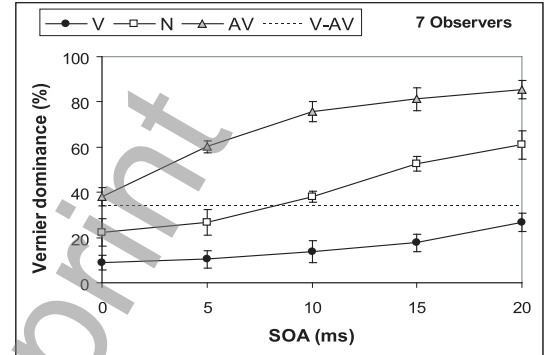
C. Model predictions

Figure 7. A. Illustration of the sequences of vernier, anti-vernier and surround used in Experiment 4. The anti-vernier was presented immediately after the vernier, whereas the surround was presented at different SOAs with respect to the onset of the vernier, ranging from 0ms to 20ms. At an SOA of 0ms, the surround was presented simultaneously with the vernier, at an SOA of 20ms, simultaneously with the anti-vernier; this condition is the 'V-25AV' condition used in the previous experiments. For the purpose of illustration, only the central 6 elements of the grating are shown, illustrating a surround presented from 10 to 30ms. B. Vernier dominance as a function of the SOA between vernier and surround. Error bars denote the standard error of the mean. C. Predictions of the neural network model. The symbols in the legend indicate the type of surround elements (e.g., 'V' for 24 verniers in the surround).

Although at long SOAs the model successfully predicts the effects of the spatial offset of the surround, the model has problems at shorter SOAs (Figure 7C). Despite this discrep-

ancy, the overall correlation for Experiment 4 between predicted and observed means is still at 0.74. The problems at short SOAs probably reflect a floor effect. The model clearly underestimates the vernier dominance at these SOAs.

Experiment 5

In the previous experiment, only two elements (a vernier followed by a mask) were presented in sequence. Here, we show that the findings for sequences of two elements also hold for sequences of more elements. We will show that the exact duration and order of elements in the sequence determines the perceived offset of the fused vernier. This means that although participants cannot report the individual elements in a sequence, the visual system accurately registers each element in a sequence. This suggest that although the information in the sequence does not seem to reach awareness, it is still represented in the brain (see also Scharnowski, Hermens, & Herzog, 2007).

Methods

Six participants took part in the experiment. They were presented in different blocks with five sequences (Figure 8): (1) a vernier followed by an anti-vernier ('V-AV'), (2) a vernier followed by 25 anti-verniers ('V-25AV'), (3) a vernier followed by 25 aligned verniers ('V-25N'), (4) a vernier followed by 25 anti-verniers, and an anti-vernier ('V-25AV-AV'), and (5) a vernier followed by 25 anti-verniers, an anti-vernier, and another 25 anti-verniers ('V-25AV-AV-25AV'). Each element was presented for 20ms, thus, sequences lasted for 40ms, 60ms, or 80ms. Each sequence was tested blockwise in two blocks of 80 trials each.

Results and discussion

The first three conditions of Experiment 5 replicate the results of Experiment 1. If a vernier is followed by an anti-vernier of equal duration and with an equally large but opposite offset ('V-AV'), the anti-vernier dominates the participants' choices. If the vernier is followed by 25 anti-verniers ('V-25AV'), vernier dominance is equal to the condition in which 25 aligned verniers ('V-25N') follow the vernier ($t(5) = 1.14$, $p = 0.30$), indicating that the spatial grouping of the central anti-vernier with the surround prevents fusion with the preceding vernier.

The next two sequences show that the spatial grouping of the anti-vernier with the surrounding 24 anti-verniers also prevents temporal fusion in longer sequences. If the vernier and the 25 anti-verniers are followed by another anti-vernier ('V-25AV-AV'), performance is very close to that in the 'V-AV' condition, indicating that the 25 anti-verniers only slightly affects the fusion of the vernier and the anti-vernier. Similarly, adding 25 anti-verniers at the end of the sequence only slightly chances vernier dominance.

In the 'V-25AV-AV-25AV' condition, the vernier is followed by three anti-verniers (the single anti-vernier and two central anti-verniers of the gratings). Still, anti-vernier dominance hardly increases with respect to the 'V-AV' condi-
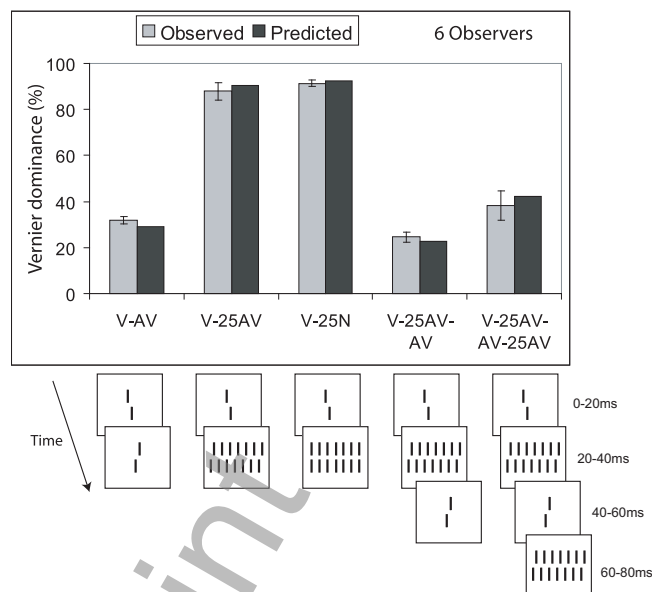

Figure 8. Vernier dominance for five different sequences in Experiment 5 (gray bars) together with the model predictions (black bars). Below the plot, the stimulus sequences are illustrated. For the purpose of illustration, only the central seven elements of each grating are shown.

tion. This indicates again that grating elements are not fused with single elements. Moreover, the single anti-vernier is fused with the target vernier even though it is masked by a forward and a backward mask (both '25AV' gratings), suggesting the visual system can sort out very briefly presented stimuli, pointing at a high temporal resolution. The grating does slightly mask the vernier, as shown by the increased anti-vernier dominance in the 'V-25AV-AV' condition with respect to the 'V-25AV-AV-25AV' condition. Still, its effect seems to be rather weak. This suggests that spatial grouping prevents the temporal fusion of elements inside gratings both for preceding and successive elements.

The results support our earlier conclusion (Scharnowski, Hermens, & Herzog, 2007) that feature fusion is not the result of a temporal blurring of the stimuli. Although participants cannot report the offsets of the individual verniers in a sequence, the perceived offset of the fused vernier very precisely reflects the duration and offset size of the verniers presented in the sequence.

A very good match between the experimental data and the model predictions is obtained ($r = 0.997$). As before, the weak effect of a '25AV' mask can be understood from the lateral inhibition and excitation that takes place between the mask elements. These lateral effects inhibit the activation corresponding to elements of the inner structure of the mask. At the center, this weak activation hardly affects that of the vernier, and therefore the only effect of the mask duration is an increase of the time until the presentation of the second vernier. The model therefore suggests that the only reason for the (small) difference between the 'V-AV' and the 'V-

25AV-AV' sequence is an increase in the time between the vernier and the anti-vernier.

Experiment 6

In Experiment 5, we found that 25 anti-verniers in a sequence of a vernier and an anti-vernier only had a minor effect on vernier dominance. We therefore argue that the grating acts as a 'neutral' element in the sequence, which means that increasing the duration of the 25 anti-verniers should not affect vernier dominance within a reasonable time range. We tested this prediction in Experiment 6 by varying the duration of the 25 anti-verniers from 0 to 70ms, and by comparing the resulting vernier dominance with that for a single anti-vernier of equal duration.

Methods

Six participants took part in the experiment. Sequences of three elements were used, as illustrated in Figure 9A. The first element in the sequence was a vernier, and the last was an anti-vernier, which were both presented for 20ms. Between the vernier and the anti-vernier, either 25 anti-verniers ('25AV') or a single anti-vernier ('AV') were presented. The duration of the 25 anti-verniers and the single anti-vernier was varied from 0, through 10, 30, 50, to 70ms. Each combination of central element type and duration was presented blockwise in two blocks of each 40 trials.

Results and discussion

Figure 9 shows that vernier dominance is essentially independent of the duration of the 25 anti-verniers for the range of durations tested ('V-25AV-AV'; $F(4,24) = 0.56$, $p = 0.69$, $h_p^2 = 0.085$). In contrast, strong effects are found, when the anti-vernier is presented without the context ('V-AV-AV'). Here, vernier dominance strongly decreases as the duration of the intermediate anti-vernier increases ($F(4,24) = 5.61$, $p = 0.0025$, $h_p^2 = 0.48$), which suggests that the anti-vernier without context is fused with the preceding vernier and anti-vernier. This is not a surprising finding, because in this condition the vernier is followed by two identical anti-verniers, which can be considered to be just one anti-vernier lasting 20ms plus the variable duration of the first anti-vernier.

When, instead of the 25 anti-verniers, a blank screen is presented between the vernier and the anti-vernier, a different effect is obtained. In a separate experiment with the participants of Experiments 5, we presented a vernier, followed by a blank screen for a variable duration between 10 and 50 ms, followed by an anti-vernier. When the blank screen lasted for longer than 10ms, participants reported seeing apparent motion and flickering of the display. This resulted in several participants asking which vernier they had to report on, meaning that the individual verniers became visible as single entities. These observations suggest that the intermediate element, which can be an array of 25 anti-verniers or a single anti-vernier, is needed for the fusion of the individual elements in the stream. This relates to the frames in a movie,

A. Stimulus sequences
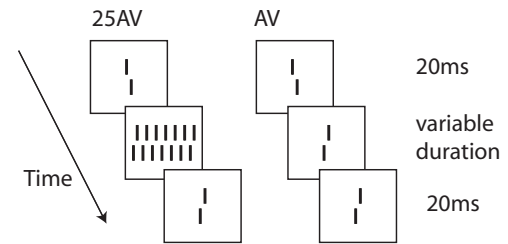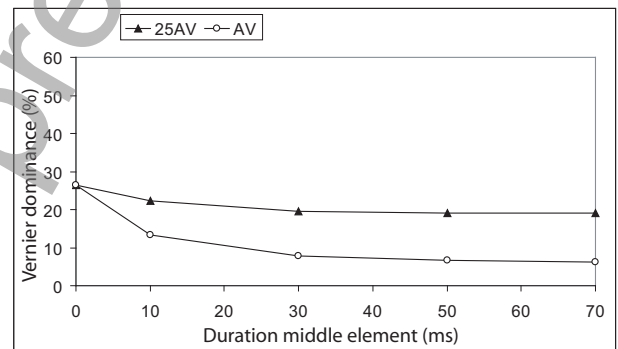

B. Experimental observations
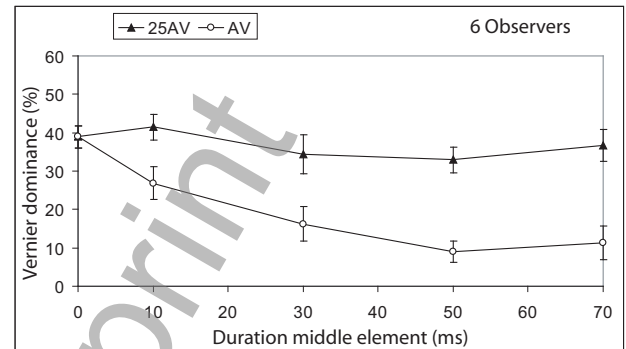
C. Model predictions

Figure 9. A. Illustration of the stimulus sequences used in Experiment 6. Either a vernier was followed by 25 anti-verniers of a variable duration and an anti-vernier (condition '25AV'), or a vernier was followed by a single anti-vernier of a variable duration and an anti-vernier (condition 'AV'). For the purpose of illustration, only the central seven elements of the grating are shown for the 25AV mask. B. Vernier dominance as a function of the duration of the central element. Error bars denote the standard error of the mean. C. Predictions of the neural network model.

which need to be presented in close succession to create a continuous percept.

The model predictions (Figure 9C) match the experimental observations (Figure 9B) well, with a correlation between the predicted and the observed means equal to 0.95. Due to the decay, the model predicts that an increase in the duration of the 25 anti-verniers results in a slight decrease of the vernier dominance. This is caused by the increased

amount of time between the vernier and the anti-vernier, during which the vernier decays. On the other hand, a strong decrease in vernier dominance is predicted as the duration of an intermediate *anti-vernier* increases, showing that in this condition not only neural decay plays a role, but that also the offset of the intermediate element is taken into account.

General discussion

Not only in the cinema, but in most daily life situations, our visual system has to integrate features across space and time to establish which elements in a visual scene belong to which object. Despite extensive research, it is still unclear how the visual system carries out this task. To understand the interactions between spatial and temporal integration, we have investigated how spatial grouping and temporal feature fusion influence each other.

The feature fusion paradigm that we used has an important advantage over previously used techniques, namely: The individual features are no longer consciously accessible after they have been fused. This becomes evident when participants are asked to indicate whether the vernier in a vernier-anti-vernier sequence is offset to the right or to the left. Participants are at chance in this task (Scharnowski et al., submitted). This means that the feature fusion paradigm probes into automatic processes instead of reflecting cognitive strategies.

Previously, we have argued that feature fusion is not simply an indication of the sluggishness of the visual system (Scharnowski, Hermens, & Herzog, 2007). The reason for this argument was that each vernier and anti-vernier in the sequence has an effect on the perceived offset of the fused vernier. For example, if the sequence contains a vernier with a large offset lasting for 40ms, the perceived offset of the fused vernier will be more in the direction of the vernier than if the same vernier in the sequence would be replaced by one with a smaller offset lasting only 20ms. This means that even though the offsets of the individual verniers in the sequence are no longer consciously accessible, the visual system does still register each of the offsets and durations. It is just that after fusion, this individual information is no longer available. Our current data support this interpretation. In Experiment 2, we find that increasing the offset of the anti-vernier results in a stronger dominance of the anti-vernier, but only if the anti-vernier is not grouped with its surround. In Experiment 5, each vernier and anti-vernier in the sequence affects vernier dominance, again, as long as it is not grouped with its surround. Experiment 6 shows that increasing the duration of the intermediate anti-vernier in a sequence, increases the dominance of the anti-vernier, as long as it is not spatially grouped.

Besides providing evidence for a high *temporal* resolution, our data also suggests that the visual system operates at a high *spatial* resolution. When only the vernier-25-anti-vernier condition is considered, one could argue that the central anti-vernier has no effect due to the fine spacing (200 arc seconds) between the anti-verniers. The consequence would be that the visual system cannot resolve the offset,

because, for example, blurring of the images makes the inner structure of the anti-vernier grating inaccessible. Such blurring would make the grating mask appear like a homogeneous light field. However, several findings argue against such an interpretation. Most importantly, when the contextual verniers are aligned instead of anti-offset, vernier dominance strongly decreases (Experiments 1 and 4). The offset differences between the aligned and anti-offset verniers are only 40 arc seconds, i.e. well below the spacing of the grating of 200 arc seconds (Experiment 1). Moreover, earlier findings (e.g. Herzog, Harms, et al., 2003) show that light masks are usually much weaker masks than grating masks, indicating that some structure of the grating is still preserved. These findings indicate that the visual system actually is very sensitive to the fine structure of the gratings.

Taken together, the fusion paradigm does not reflect temporal or spatial blurring, but instead offers insights into *active* integration processes performed by the human visual that are beyond cognitive and conscious control.

Spatial grouping

In order to detect objects in a visual scene, the elements in a display have to be grouped into objects. For example, in order to perceive a grating, lines need to be grouped. With the feature fusion paradigm, it is possible to measure the strength of this spatial grouping as a function of, for example, the distance between the lines. This is by the assumption that if spatial grouping becomes weaker, temporal fusion becomes gradually stronger (Experiment 3). We found grouping by proximity to act for spacings between anti-verniers up to 600". Similarly, the strength of the spatial grouping can be varied by changing the offset direction of the individual elements. By presenting the anti-vernier in a surround of anti-verniers, grouping is strong, as indicated by weak temporal fusion. If the anti-vernier is presented in the context of aligned verniers, spatial grouping is much weaker, which is reflected in stronger fusion (Experiment 4).

It is very likely that the above grouping operations occur without cognitive control. In an additional experiment, reported elsewhere (Hermens, Scharnowski, & Herzog, 2007), participants performed two tasks. In the first task, a setup very similar to that in Experiments 1 and 2 was used. Participants were presented with a vernier target followed by either a single anti-vernier or five anti-verniers embedded in a grating of aligned verniers (Figure 5). The task was to report the offset of the preceding vernier. As in Experiments 1 and 2, vernier dominance was high with five anti-verniers, but low with a single anti-vernier, indicating that in order for fusion to take place between the vernier and the anti-vernier, there has to be just one anti-vernier in the grating. In the second task, the preceding vernier was removed from the sequence, and now the task of the participants was to report the number of anti-verniers in the grating (one or five). Performance on this task varied greatly across participants. Whereas some participants reported the correct number of offset elements in 80% of the trials, others were at chance. More importantly, performance on this task did not correlate with the effects

the gratings had on feature fusion. Participants who could not determine the number of anti-verniers in the grating still showed a large effect of the number of anti-verniers in the grating on the offset discrimination of the preceding vernier. This suggests that conscious access to the structure of the grating is not necessary for fusion to take place.

Most text books typically discuss each of the Gestalt laws in isolation. This makes it appear that the Gestalt grouping laws operate relatively independently. For example, illustrations are provided in which elements either intuitively group by color, orientation, or proximity (e.g. Palmer et al., 2003). There are only a few studies that compare the relative strengths of Gestalt cues directly. For example, Quinlan and Wilton (1998) showed an interaction between the principles of proximity and similarity (see also Han, 2004; Han, Ding, & Song, 2002) and Van Lier and Wagemans (1997) between the factors regularity and proximity. Similarly, interactions were found between the principles of temporal and spatial proximity (Gepshtein & Kubovy, 2007; Oyama & Yamada, 1978). Typically, participants in these experiments are asked to decide between one grouping versus another (e.g., horizontal versus vertical). Decisions are based on the percepts and the corresponding reports. We think that the feature fusion paradigm offers an interesting alternative approach by using an objective binary task. Our experiment demonstrate that with the fusion paradigm, the relative strengths of the various grouping factors, such as spatial and temporal proximity (Experiments 3 and 4) and similarity (Experiment 4) can be measured without the need to explicitly ask observers for the grouping strength (see also Razpurker-Apfeld & Kimchi, 2007).

Experiment 4 demonstrated an interaction between temporal proximity and item similarity. Such interactions between temporal and spatial aspects are also found in paradigms, such as visual masking (Duangudom, Francis, & Herzog, 2007), motion perception in Ternus-Pickler displays (Kramer & Yantis, 1997), and in the detection of spatial structure mediated by purely temporal cues (Lee & Blake, 1999), suggesting that temporal and spatial aspects should not be studied in separation (Herzog, 2007).

Temporal integration

An often used task to study temporal integration is the form-part integration task (Dixon & Di Lollo, 1994). In this task, an image is broken up into two frames which are presented in sequence. Only if the two frames can be integrated, the image can be identified. For example, an array of dots is presented in two frames, with half of the dots presented in one frame and half of the other half minus one dot in the other frame. The task of the participant is to locate the missing dot. Performance is measured as a function of the time between the two frames and the duration of the two frames.

Although vernier fusion resembles the classical paradigms of temporal integration, there are some important differences. First, in temporal integration, a blank screen can be presented between the two frames and integration still takes place. This is not the case for the fusion of two verniers.

As soon as a brief interstimulus interval is used, the percept changes. Instead of a single fused vernier, a vernier that seems to rotate or, at longer interstimulus intervals, two separate verniers are perceived. Temporal integration with an interstimulus interval could only be obtained by presenting the anti-vernier at a different spatial position (Scharnowski, Hermens, Kammer, et al., 2007). Second, in temporal integration, a superposition of the two frames is perceived. In vernier fusion, however, a single vernier is perceived which is not the result of superposition of the two images. If superposition of the two verniers would occur, two aligned verniers were perceived rather than one single vernier.

Related paradigms

Several paradigms are related to feature fusion in that they probe into similar processes. We here discuss some of them.

Forward and backward masking. Feature fusion can be viewed as a special case of forward and backward masking (for a monograph on masking, see Breitmeyer & Öğmen, 2006). The vernier acts as a forward mask to the anti-vernier and the anti-vernier as a backward mask to the vernier. Due to this mutual masking, performance is much lower than when a single unmasked vernier is presented (Figure 3, condition 'V').

Masking effects are often divided in two major types, for which the terms integration and interruption masking are mostly used. Integration masking is thought to occur with pattern masks which spatially overlap with the target. Target performance is assumed to decrease because of unspecific masking effects such as contrast reduction. The mask and the target are treated as a double exposure and for the amount of masking the mutual luminances of target and mask are of more importance than the time course of the stimulus sequence (Eriksen, 1966). In contrast, interruption masking is often attributed to the use of metacontrast masks, which do not spatially overlap with the target. These masks are believed to impair performance on the target because they interact with its dynamical processing.

We here propose that within integration masking, there are two different aspects. The first aspect, as mentioned above, is of an unspecific nature, possibly caused by low level effects of the mask, such a reduction in target contrast. It is the unspecific aspect of integration masking that impairs performance on a vernier target if it is followed by a grating of 25 aligned verniers (Experiment 1 and 5). The second aspect of integration masking comes with a specific integration of features, such as the integration of the offsets of a vernier and an anti-vernier (Figure 1). This type of masking is specific, because it depends on the exact features of the stimuli involved, such as the offset sizes in the example of a sequence of vernier and anti-vernier. It is this aspect of integration masking that causes the decrease of vernier dominance as the offset of the anti-vernier increases (Experiment 2). This aspect of masking is clearly different from the unspecific one, because the contrast of the vernier is not affected by an increase in the offset size of the anti-vernier. If the offset di-

rection of the anti-vernier would be the same as that of the preceding vernier, i.e. it is actually a vernier, completely changes the pattern of results. Moreover, these specific effects of integration masking depend in subtle ways on the spatial layout of the mask. In all pattern masking paradigms, both types of masking can occur at the same time, often making it difficult to tell which mechanism is at work.

In the above, we used the term mutual masking. However, studies of mutual masking are often slightly different from the paradigm that we used. Commonly mutual masking refers to the situation in which two stimuli are presented in close succession and participants are asked to report the identity of both of them (Bachmann & Allik, 1976; Bachmann & Sikka, 2005). If the stimuli are presented in isolation, performance is highest on the second object. This aspect of mutual masking agrees with our findings in feature fusion, in which the latter of the two objects dominates the percept. However, feature fusion is different from mutual masking in that participants only perceive one fused stimulus instead of two successive stimuli. In feature fusion, it therefore is impossible to identify the individual offsets of the two stimuli (Scharnowski et al., submitted), whereas in mutual masking this is still possible to some degree. Both in mutual masking and in feature fusion, the dominance of the last stimulus can be changed into a dominance of the first stimulus by other elements in the sequence. In mutual masking, if the same two stimuli are presented in a sequence of stimuli, the dominance of the two stimuli reverses. In a sequence, performance is highest on the first of the two stimuli (Bachmann & Sikka, 2005). In feature fusion, whether dominance reverses is very much dependent on the idenitity of the other elements in the sequence. When the vernier and anti-vernier are presented in a stream of aligned verniers, the anti-vernier remains to dominate performance (Otto, Ögmen, & Herzog, 2006). However, if a vernier and anti-vernier are followed by a grating, dominance does reverse, and performance is dominated by the vernier (Herzog et al., 2006; Herzog, Scharnowski, & Hermens, 2007). In summary, although mutual masking and feature fusion are related, the two paradigms differ in one important aspect: In mutual masking the task is to identify both stimuli, in feature version only one decision has to be made.

Our present data clearly show how strongly masking can be affected contextual elements (see also Herzog, Schmonsees, & Fahle, 2003). If instead of the single anti-vernier, an array of anti-verniers follows the vernier, masking is strongly diminished even though the anti-vernier grating has a much higher energy (Experiment 1). This contrasts to typical masking findings in which masks of higher intensity and luminance are stronger masks (Breitmeyer & Öğmen, 2006).

The prominent role of the spatial layout in pattern masking was shown in a series of experiments related to the so-called shine-through effect. In shine-through, a vernier is followed by gratings of aligned verniers for 300ms. Even though such a grating is a high energy mask (it contains many elements and it is presented for a long time), masking was relatively weak (Herzog & Koch, 2001). Masking is strongly increased by small spatio-temporal irregularities in the grating (Herzog

& Fahle, 2002). Similar strong effects of the spatial layout of the mask were also obtained for metacontrast masks presented for only 20ms (Duangudom et al., 2007).

Feature inheritance. Feature fusion is also related to the feature inheritance phenomenon. In feature inheritance, a vernier is followed by a five element grating presented for 300ms. Because of the long duration of the small grating, in combination with the brief presentation of the vernier, the vernier is completely masked (for the 20ms grating duration, we used here, the vernier would clearly be visible). Still, observers perceive all gratings elements to be offset in the direction of the vernier offset (Sharikadze, Fahle, & Herzog, 2005; Hamker, 2007; Herzog & Koch, 2001). Attention is important for feature inheritance to occur. The offset of the target is only seen in attended items (Sharikadze et al., 2005). Similar attention effects are found for streams of items (Otto et al., 2006) and in feature fusion (Scharnowski, Hermens, Kammer, et al., 2007).

Modeling

We could show that a simple neural network with lateral inhibition and excitation explains the trade-off between spatial grouping and temporal fusion very well. Due to the lateral interactions, neural activity corresponding to the inner elements of a homogeneous grating is inhibited over time. This means that the activation corresponding to an anti-vernier embedded in an array of anti-verniers is suppressed (see Figure 4). As there is no neural activation corresponding to the anti-vernier, there is no activation to fuse with the vernier. The vernier dominates performance. The situation changes if the anti-vernier is embedded in an array of aligned verniers. Due to the inhomogeneity caused by the small offset difference between the central anti-vernier and the surrounding aligned verniers, the anti-vernier is no longer filtered out. The corresponding activation combines with the vernier related activation. The model thereby provides a mechanism by which the proposed grouping mechanisms may be implemented.

Whereas these considerations are highly speculative, there is evidence that the visibility of the individual lines of a grating is strongly diminished. If participants have to report the offset of a vernier embedded in a grating, performance is strongly deteriorated compared to when the vernier is presented alone or at the edges of the grating (Malania et al., 2007; Sharikadze et al., 2005) (in these experiments all elements are displayed simultaneously). This suggests that for a vernier presented inside a grating, grouping of the vernier with its surrounding elements impairs offset discrimination (which we attribute to inhibition). This interpretation is supported by experiments related to the spatial resolution of attention. Counting the number of seemingly clearly visible lines in a grating is a difficult task (see also Intriligator & Cavanagh, 2001). This finding suggests that seeing each individual line in a grating might be an illusion. The explicit perception of the individual lines may be suppressed in favor

of seeing the grating as a whole: It is like seeing the forest for the trees.

It is important that the model type, we used, was designed to explain cortical processing in general (Wilson & Cowan, 1973), i.e. the model was not designed for the current experiments. Quite to the contrary, the model was used in previous studies and could explain a large range of visual masking data, including spatial aspects of masking, such as the number of elements in a grating mask, and temporal aspects, such as A-type masking for high intensity masks and B-type masking for masks of lower intensity (Hermens et al., 2008).

Several models exist to explain simultaneous or temporal masking effects. However, most of these models explain only one aspect. On the one hand, there are models that aim to describe the spatial aspects of masking. These are often models of simultaneous masking (e.g. Watson, Solomon, & Watson, 1997; Yu, Klein, & Levi, 2003), in which the target and the mask are presented simultaneously and the spatial layout of the target and the mask is varied. Because temporal aspects are not varied in simultaneous masking (the target is always presented together with the mask), the models typically lack the temporal component, which makes that they cannot be used to explain the current data. On the other hand, there are models that focus on temporal aspects. Typically, the target and the mask are represented as an event, for example by activating a single neuron (Anbar & Anbar, 1982; Di Lollo, Enns, & Rensink, 2000; Francis, 2003; Francis & Cho, 2005; Weisstein, 1968). A few models exist that can both explain temporal and spatial aspects of masking (Bridgeman, 1971; Bugmann & Taylor, 2005; Francis, 1997; Öğmen, 1993). However, these models tend to be very complex, which makes it difficult to understand the exact mechanisms by which they explain the data. An exception is the model by Bridgeman (1971). However, in its current form the offset of the elements in the grating cannot be represented. A model that is similar to our model, although more complex, is the model by Zhaoping (Zhaoping, 1999, 2000, 2003), and it is likely that this model also explains our data well.

The good fit of our model to the current data demonstrates that low-level neural interactions suffice to explain the seemingly complex interactions between spatial grouping and temporal fusion. Hence, complex grouping interactions may well be implemented by simple neural circuits, in accordance with physiological findings (Yasuko Sugase, Yamane, & Ueno, 1999), and contrary to assumptions that grouping happens after basic visual processing is accomplished (e.g. Palmer et al., 2003).

References

Alpern, M. (1953). Metacontrast. *Journal of the Optical Society of America*, *43*, 648-657.

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

Bach, M. (1996). The "Freiburg visual acuity test". Automatic measurement of visual acuity. *Optometry and Vision Science*, *73*, 49–53.

Bachmann, T., & Allik, J. (1976). Intergration and interruption in the masking of form by form. *Perception*, *5*(1), 79–97.

Bachmann, T., & Sikka, P. (2005). Perception of successive targets presented in invariant-item streams. *Acta Psychologica*, *120*, 19–34.

Blake, R., & Lee, S.-H. (2005). The role of temporal structure in human vision. *Behavioral Cognitive Neuroscience Review*, *4*(1), 21–42.

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.

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

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*(4), 481–507.

Dixon, P., & Di Lollo, V. (1994). Beyond visible persistence: an alternative account of temporal integration and segregation in visual processing. *Cognitive Psychology*, *26*, 33–63.

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

Efron, R. (1967). Duration of present. *Annals of the New York Academy of Sciences*, *138*(A2), 713–729.

Efron, R. (1973). Conservation of temporal information by perceptual systems. *Perception & Psychophysics*, *14*(3), 518–530.

Eriksen, C. W. (1966). Temporal luminance summation effects in backward and forward masking. *Perception & Psychophysics*, *1*, 87–92.

Exner, S. (1875). Experimentelle Untersuchungen der einfachsten psychischen Processe. III. Der persönlichen Gleichung zweiter Theil. *Pflugers Archiv*, *11*, 403–432.

Fahle, M. (1993). Figure-ground discrimination from temporal information. *Proceedings of the Royal Society London B*, *254*(1341), 199–203.

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

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

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.* The MIT Press.

Gepshtein, S., & Kubovy, M. (2007). The lawful perception of apparent motion. *Journal of Vision*, *7*(8), 1–9.

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.

Hamker, F. H. (2007). The mechanisms of feature inheritance as predicted by a systems-level model of visual attention and decision making. *Advances in Cognitive Psychology*, *3*, 111–123.

Han, S. (2004). Interactions between proximity and similarity grouping: an event-related brain potential study in humans. *Neuroscience Letters*, *367*(1), 40–43.

Han, S., Ding, Y., & Song, Y. (2002). Neural mechanisms of perceptual grouping in humans as revealed by high density event related potentials. *Neuroscience Letters*, *319*(1), 29–32.

Hermens, F., Luksys, G., Gerstner, W., Herzog, M. H., & Ernst, U. (2008). Modeling spatial and temporal aspects of visual backward masking. *Psychological Review*, *115*, 83–100.

Hermens, F., Scharnowski, F., & Herzog, M. H. (2007). Feature integration independent of awareness [abstract]. *Poster presented at the KNAW workshop 'Blinks of the Mind', Amsterdam (The Netherlands)*.

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

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 of Sciences of the USA*, *98*(7), 4271–4275.

Herzog, M. H., Leseman, E., & Eurich, C. W. (2006). Spatial interactions determine temporal feature integration as revealed by unmasking. *Advances in Cognitive Psychology*, *2*(1), 77–85.

Herzog, M. H., Parish, L., Koch, C., & Fahle, M. (2003). Fusion of competing features is not serial. *Vision Research*, *43*, 1951–1960.

Herzog, M. H., Scharnowski, F., & Hermens, F. (2007). Long lasting effects of unmasking in a feature fusion paradigm. *Psychological Research*(71), 653–658.

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

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, *43*(3), 171–216.

Kramer, P., & Yantis, S. (1997). Perceptual grouping in space and time: evidence from the ternus display. *Perception & Psychophysics*, *59*(1), 87–99.

Lee, S. H., & Blake, R. (1999). Detection of temporal structure depends on spatial structure. *Vision Research*, *39*(18), 3033-3048.

Malania, M., Herzog, M. H., & Westheimer, G. (2007). Grouping of contextual elements that affect vernier thresholds. *Journal of Vision*, *7*(2), 1–7.

Otto, T. U., Ögmen, H., & Herzog, M. H. (2006). The flight path of the phoenix the visible trace of invisible elements in human vision. *Journal of Vision*, *6*, 1079–1086.

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

Ögmen, H., Breitmeyer, B. G., & Melvin, R. (2003). The what and where in visual masking. *Vision Research*, *43*, 1337–1350.

Oyama, T., & Yamada, W. (1978). Perceptual grouping between successively presented stimuli and its relation to visual simultaneity and masking. *Psychological Research*, *40*(2), 101–112.

Palmer, S. E., Brooks, J. L., & Nelson, R. (2003). When does grouping happen? *Acta Psychologica*, *114*(3), 311-330.

Potter, M. C., & Fox, L. F. (2004). Perceiving and remembering multiple pictures in RSVP. *Journal of Vision*, *4*(8), 868.

Quinlan, P. T., & Wilton, R. N. (1998). Grouping by proximity or similarity? Competition between the Gestalt principles in vision. *Perception*, *27*, 417–430.

Razpurker-Apfeld, I., & Kimchi, R. (2007). The time course of perceptual grouping: the role of segregation and shape formation. *Perception & Psychophysics*, *69*(5), 732-743.

Scharnowski, F., Hermens, F., & Herzog, M. H. (2007). Bloch's law and the dynamics of feature fusion. *Vision Research*, *47*(18), 2444–2452.

Scharnowski, F., Hermens, F., Kammer, T., Ögmen, H., & Herzog, M. H. (2007). Feature integration reveals the temporal dynamics of retinotopic and non-retinotopic visual memory. *Journal of Cognitive Neuroscience*, *19*(4), 632-641.

Scharnowski, F., Rüter, J., Jolij, J., Hermens, F., Kammer, T., & Herzog, M. H. (submitted). Transcranial magnetic stimulation modulates feature integration preceding visual awareness.

Sharikadze, M., Fahle, M., & Herzog, M. H. (2005). Attention and feature integration in the feature inheritance effect. *Vision Research*, *45*(20), 2608-2619.

Sweet, A. L. (1953). Temporal discrimination by the human eye. *American Journal of Psychology*, *66*(2), 185–198.

Van Lier, R., & Wagemans, J. (1997). Perceptual grouping measured by color assimilation: Regularity versus proximity. *Acta Psychologica*, *97*, 37-70.

Watson, A., Solomon, J., & Watson, A. (1997). Model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America A*, *14*, 2379–2391.

Weisstein, N. (1968). A rashevsky-landahl neural net: simulation of metacontrast. *Psychological Review*, *75*(6), 494–521.

Westheimer, G., & Hauske, G. (1975). Temporal and spatial interference with vernier acuity. *Vision Research*, *15*, 1137–1141.

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

Yasuko Sugase, Y., Yamane, S., & Ueno, K., S.and Kawano. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, *400*, 869-873.

Yu, C., Klein, S. A., & Levi, D. M. (2003). Cross- and iso- oriented surrounds modulate the contrast response function: The effect of surround contrast. *Journal of Vision*, *3*(8), 527–540.

Yund, E. W., Morgan, H., & Efron, R. (1983). The micropattern effect and visible persistence. *Perception & Psychophysics*, *34*(3), 209–213.

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*(1), 25–50.

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

Appendix

Details of the model

Two equations describe the evolution of the activation of the excitatory and the inhibitory layer in the model:

$$\tau_e \frac{\partial A_e(x,t)}{\partial t} = -A_e(x,t) + h_e \left\{ w_{ee} \left(A_e \star W_e \right)(x,t) + \right.$$

$$+w_{ie}\,(A_i \star W_i)\,(x,t)+I(x,t)\} \tag{1}$$

$$\tau_i\frac{\partial A_i(x,t)}{\partial t} = -A_i(x,t)+h_i\{w_{ei}\,(A_e \star W_e)\,(x,t)+$$

$$+w_{ii}\,(A_i \star W_i)\,(x,t)+I(x,t)\}\,. \tag{2}$$

The equations state that the change in activation ($A_e(x,t)$ and $A_i(x,t)$) is determined by a set of factors. First, activation decays, which is indicated by the terms $-A_e(x,t)$ and $-A_i(x,t)$. Second, there is excitation by the neurons in the excitatory layer ($\{w_{ee}(A_e \star W_e)(x,t)\}$ and $\{w_{ei}(A_e \star W_e)(x,t)\}$). This excitation is stronger for neurons at nearby locations, as expressed by the Gaussian kernel W_e ($\star$ denotes the convolution with the kernel). Third, there is inhibition from the neurons in the inhibitory layer ($\{w_{ie}(A_i \star W_i)(x,t)\}$ and $\{w_{ii}(A_i \star W_i)(x,t)\}$), weighted by the kernel W_i. Finally, input ($I(x,t)$) into the layers drives the activation. The effect of the input and neighboring neurons is restricted to values above zero by the functions h_e and h_i. The time constants τ_i, τ_e and the coupling strengths w_{ii}, w_{ei}, w_{ie}, w_{ee} complete the system. For a detailed discussion of the equations and further details of the model, including the values of the constants in the model, we refer to earlier publications (Hermens et al., 2008; Herzog, Ernst, et al., 2003).

To link the activation in the network to the observed mean vernier dominance, we used the following function to determine the relative evidence for a left offset, E_L:

$$E_L = \sum_{x,y}A_e(x,y,t_r)\cdot M_L(x,y) - \sum_{x,y}A_e(x,y,t_r)\cdot M_R(x,y) \tag{3}$$

where A_e is the activation in the excitatory layer, t_r is the read-out time, M_L is the input map containing the left offset vernier, and M_R is the map with the right offset vernier. The function compares the evidence in the excitatory layer for a right offset to that for a left offset at a particular read-out time. This read-out time, we set to 20ms after the disappearance of the last stimulus in the sequence, although other values could be used as well. To compare the relative evidence for the vernier to the observed dominance, we converted E_L to a percentage ('P') by means of a logistic function, of which we fitted the slope (σ) and the center of the curve ('shift') to the data of the first experiment:

$$P = \frac{100-\text{lapse}}{1+\exp(\sigma\cdot E_L - \text{shift})} + \text{lapse} \tag{4}$$

Because participants make errors even when they can easily perform the task, we introduced a minimum and a maximum to the fitted logistic function. We assumed this minimum and maximum to be 5% ('lapse') and 95% ('100-lapse') respectively.

The shift parameter was introduced to be able to fit the parameters on the data of the first experiment only. The vernier dominances of this experiment were biased towards higher values, which led to a biased fit of the logistic function without the shift parameter. For the parameter fit, we used only the data of the first experiment to determine how well the parameter settings extrapolate to different experiments. The best fit of the logistic function on the data of Experiment 1 was obtained with $\sigma = 0.30$ and shift $= 1.1$.