Spatiotemporal mechanisms for simple image feature perception in normal and amblyopic vision

Shuang Song
Vision Science, University of California, Berkeley, CA, USA

Dennis M. Levi
Vision Science, School of Optometry, and Helen Wills Neuroscience Institute, University of California, Berkeley, CA, USA

We used the method of image classification to investigate the spatiotemporal mechanisms for simple feature perception in normal and amblyopic vision. In the first experiment, we estimated the spatiotemporal mechanism for detecting a luminance increment of a bright bar embedded in spatiotemporal noise for normal and amblyopic observers. The normal template is characterized by a temporal summation zone surrounded by symmetric spatial inhibition zones and followed by a temporal inhibition zone. The abnormal amblyopic template lacks inhibition but shows normal temporal summation. Neither blurring the stimuli in space nor varying the target signal-to-noise ratios caused any significant change in the normal or amblyopic template. However, decreasing the fundamental frequency of the stimuli restored the normal template in the amblyopic eye. Furthermore, the normal periphery demonstrates a spatiotemporal template similar to that of the amblyopic eye. The second experiment mapped the spatiotemporal interaction dynamics of flankers in an orientation discrimination task. For normal peripheral vision, assimilation of the perceived target orientation into the flunker orientation (i.e., crowding) occurs at flanker locations adjacent to the target in space and around or before the target presentation in time. Cueing the target may influence the spatiotemporal interaction map, but the strongest crowding never coincides with the target presentation. The amblyopic spatiotemporal interaction map is similar to that of the normal periphery, except that in general the spatiotemporal interaction is more widely distributed around and after the target presentation. Distant flankers induced “anti-crowding” (repulsion) of perceived orientation in both normal and amblyopic vision.

Keywords: crowding, feature detection, amblyopia, classification image


Introduction


The temporal properties of amblyopic vision have not received nearly as much attention as the spatial properties. Among the few studies on amblyopic temporal visual deficits, an abnormal spatiotemporal modulation function with reduced contrast sensitivity more markedly at low temporal frequencies and at medium to high spatial frequencies has been found (Bradley & Freeman, 1985; Levi & Harwerth, 1977; Manny & Levi, 1982a, 1982b). Interestingly, this defect of sensitivity at low temporal frequencies was much milder or absent for apparent contrast tasks using suprathreshold gratings (Manny & Levi, 1982a, 1982b), suggesting that the mechanism for detecting flicker at near-threshold contrast may differ from that at suprathreshold contrast. Moreover, at low spatial frequencies, little difference in temporal sensitivity has been noted between amblyopic and normal eyes (Bradley & Freeman, 1985; Manny & Levi, 1982a, 1982b). These studies suggest that the amblyopic spatiotemporal mechanism for detecting a simple image feature is abnormal—the spatial mechanism may be coarser with decreased resolution, and given the extant data, we can at least expect that the temporal mechanism differs from that of the normal at high spatial frequencies. On the other hand, the amblyopic spatiotemporal template for identifying a crowded feature would be similar to that of the normal peripheral vision if both share the same mechanism for spatial interactions.

In the present study, we used the method of image classification, which is essentially a correlation analysis of the observer’s decision with the random noise present in the stimuli. The purpose of this method is to estimate the detection template (i.e., strategy or mechanism) for
performing a specific task for an observer (Ahumada, 1996; Ahumada & Lovell, 1971; Beard & Ahumada, 1998; Gold, Murray, Bennett, & Sekuler, 2000; Levi & Klein, 2002; Nandy & Tjan, 2007; Neri & Heeger, 2002; Neri, Parker, & Blakemore, 1999; Watson & Rosenholtz, 1997). The classification image technique is closely related to the reverse correlation technique—a tool for neural receptive field analysis that was developed in the 1980s and 1990s (Jones & Palmer, 1987; Ohzawa, DeAngelis, & Freeman, 1993; Ringach, Hawken, & Shapley, 1997), and they have both been mathematically justified (Ahumada, 2002; Bussgang, 1952; de Boer, 1967; de Boer & Kuyper, 1968; Neri & Levi, 2006; Ringach & Shapley, 2004). Classification image methods have also been applied in amblyopic vision studies. Levi et al. found that in amblyopia the templates for target detection and position discrimination were shifted to lower spatial frequencies, especially the latter that was substantially impaired at higher spatial frequencies. These abnormal templates as well as the high internal noise together account for the loss of performance for amblyopic vision (Levi & Klein, 2002, 2003; Levi, Klein, & Chen, 2008). Interestingly, there is evidence that the abnormal template for position discrimination can at least partly be corrected after training through perceptual learning (Li, Klein, & Levi, 2008).

In the present study, we applied spatiotemporal noise (either luminance or orientation noise) to the corresponding visual feature and estimated the two-dimensional spatiotemporal classification images for both normal and amblyopic observers.

### Experiment 1: The spatiotemporal mechanism for luminance discrimination

In the first experiment, we use a variant of the method described by Neri and Heeger (2002) to estimate the spatiotemporal mechanism for luminance discrimination. Their study directly demonstrated this spatiotemporal mechanism for feature detection for the first time, and the resulting classification image resembled the spatiotemporal receptive field of a V1 neuron to some degree (DeAngelis, Ohzawa, & Freeman, 1993; Neri & Levi, 2006).

The spatial profile of the classification image has a “Mexican hat” shape typically seen in both the neural receptive field and the behavioral “perceptive field” (DeAngelis et al., 1993; Hubel & Wiesel, 1959). However, the temporal profile of the classification image is rather widely distributed, and the often observed temporal inhibition (or transience) in detecting asynchronous light stimuli was absent (Kelly, 1971a, 1971b; Rashbass, 1970; Saarela & Herzog, 2008; Watson & Nachmias, 1977). A probable explanation for the purely sustained temporal profile of the classification image is the high temporal uncertainty associated with the task. High uncertainty is not only inefficient in estimating the template but can also severely diminish the value of the result. It has been shown that for classification images, a relatively strong signal may reduce perceptual uncertainty and provide the task-relevant template (the “signal clamped method”—Tjan & Nandy, 2006). In the present study, we repeated Neri and Heeger’s main experiments but reduce the temporal uncertainty by using a discrimination (instead of detection) task, in which the target signal was an increment on a relatively high pedestal that can be reliably localized temporally and spatially.

### Methods

#### Observers

Five observers with normal vision and six with amblyopia (2 purely anisometropic and 4 with strabismus) participated in our study. All participants had been exposed to hours of visual psychophysical experiments and were naive as to the purpose of the study, except for the author. None of the amblyopic observers has eccentric fixation. One amblyope, JS, had received intensive training with her amblyopic eye through perceptual learning, and her visual acuity as well as other visual functions of the amblyopic eye had improved substantially by the time she entered the current study. The detailed characteristics of each observer are listed in Table 1.

#### Stimuli

Stimuli were generated by Matlab 7.01 with Psychtoolbox 2.0 and presented on a flat CRT monitor (39.6 × 29.7 cm, 1600 × 1400 pixels, 100 Hz, NEC). At 50% brightness and 100% contrast, the luminance gradient ranges from 0.32 to 117.7 cd/m². The monitor was calibrated to generate a linear luminance function of pixel value. There were six conditions in this experiment.

*Condition 1 (basic noise–signal).* The background luminance was 58 cd/m². The basic stimulus element was a 1.1 × 0.8 deg square composed of 11 vertical bars in juxtaposition. Each bar measured 0.1 × 0.8 deg and was randomly assigned one of 11 luminance noise increments (from −6.9 cd/m² to 6.9 cd/m² with step size of 1.38 cd/m²). A complete stimulus consisted of a 0.42-s sequence of 21 such elements (i.e., frames), each of which lasted 20 ms. The middle (6th) bar in the middle (11th) frame of the sequence was the target, and the rest were noise (i.e., distracters). The target luminance on each trial was randomly selected from one of two values, 94.7 cd/m² (high) and 86.4 cd/m² (low), and was noise free. It was separately determined for each amblyopic observer according to his/her visual function (see Preliminary experiments). Two red bars, 1/2 the length and 1/3 the width of the target bar, abutting the middle bar of each
frame from above and below, served as the spatial cue for the target and were present throughout a trial. These noise–signal sessions served as the basic condition for both normal and amblyopic observers (Figure 1).

**Condition 2 (basic noise–noise).** The luminance of the target bar was replaced by a fixed luminance pedestal (i.e., 86.4 cd/m²) plus a noise luminance increment randomly selected in the same way as for any other distracters from the 11 possible values. In this case, the signal was “imaginary” but the target could still be spatially and temporally localized because of its relatively high base luminance. The rest of a noise–noise session was the same as the noise–signal sessions described above. This condition was also part of the basic setup for other conditions.

**Condition 3.** For two normal observers, SS and SK, blurred stimuli were generated by spatially convolving each original frame with a two-dimensional disk filter with the radius equal to half of the bar width. The luminance levels of different components of the stimuli were varied in order to maintain the performance.

**Condition 4.** For one normal observer SK, target signal-to-noise ratios of the stimuli were decreased by increasing the noise luminance of the distracters. The step size of the luminance increment of the distracters was elevated from 1.38 cd/m² to 2.3 cd/m². Here, the target signal-to-noise ratios are defined to consist of both the ratio of the target luminance pedestal to the noise luminance step size and the ratio of the target luminance increment to the noise luminance step size.

**Condition 5.** The target signal-to-noise ratios of the stimuli was increased by either increasing both the target luminance pedestal and the target luminance increment, or reducing the noise (i.e., decreasing the absolute values of the noise luminance) or both. When presented in the normal periphery, the target luminance values were 108.6 cd/m² (high) and 78.3 cd/m² (low), and the range of noise luminance increments was reduced from [−6.9, 6.9] cd/m²

### Table 1. Observers’ characteristics.

<table>
<thead>
<tr>
<th>Observer</th>
<th>Age (years)</th>
<th>Strabismus (at 6 m)</th>
<th>Eye</th>
<th>Refractive error (diopters, D)</th>
<th>Line letter VA (single letter VA)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Normal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS</td>
<td>30</td>
<td>None</td>
<td>R</td>
<td>−0.25</td>
<td>20/12.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L</td>
<td>−1.00 × 0.25</td>
<td>20/12.5</td>
</tr>
<tr>
<td>SK</td>
<td>19</td>
<td>None</td>
<td>R</td>
<td>−4.75/−1.00 × 165</td>
<td>20/16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L</td>
<td>−3.00</td>
<td>20/16</td>
</tr>
<tr>
<td>CN</td>
<td>23</td>
<td>None</td>
<td>R</td>
<td>−0.50</td>
<td>20/12.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L</td>
<td>−0.50</td>
<td>20/16</td>
</tr>
<tr>
<td>EP</td>
<td>25</td>
<td>None</td>
<td>R</td>
<td>+0.75</td>
<td>20/16</td>
</tr>
<tr>
<td>MN</td>
<td>36</td>
<td>None</td>
<td>R</td>
<td>−1.25/−1.50 × 106</td>
<td>20/12.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L</td>
<td>−1.75/−1.00 × 73</td>
<td>20/12.5</td>
</tr>
<tr>
<td><strong>Strabismic (including strabismic and anisometric) amblyopic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BK</td>
<td>62</td>
<td>L ExoT 8°</td>
<td>R (NAE)</td>
<td>−1.50/−2.5 × 105</td>
<td>20/12.5°</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R hyperT 6°</td>
<td>L (AE)</td>
<td>−3.00/−0.25 × 135</td>
<td>20/25° (20/30° 2)</td>
</tr>
<tr>
<td>BN</td>
<td>22</td>
<td>L EsoT 3°</td>
<td>R (NAE)</td>
<td>+5.50/−2.25 × 5</td>
<td>20/16°</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L (AE)</td>
<td>+5.50/−1.50 × 175</td>
<td>20/50° (20/25° 2)</td>
</tr>
<tr>
<td>GW</td>
<td>59</td>
<td>R EsoT 4°</td>
<td>R (AE)</td>
<td>pl</td>
<td>20/100° (20/32° 1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L (NAE)</td>
<td>+0.50/−0.75 × 180</td>
<td>20/16°</td>
</tr>
<tr>
<td>GJ</td>
<td>24</td>
<td>R EsoT 4°</td>
<td>R (AE)</td>
<td>+3.50/−1.00 × 97</td>
<td>20/40° (20/25° 1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L (NAE)</td>
<td>pl</td>
<td>20/16°</td>
</tr>
<tr>
<td><strong>Anisometric amblyopic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GD</td>
<td>45</td>
<td>None</td>
<td>R</td>
<td>+0.25/−0.50 × 90</td>
<td>20/12.5°</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L</td>
<td>+3.75/−1.00 × 30</td>
<td>20/50° (20/30 2)</td>
</tr>
<tr>
<td>JS</td>
<td>27</td>
<td>None</td>
<td>R (AE)</td>
<td>+1.00</td>
<td>Pre-training: 20/25° (20/25° 2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L (NAE)</td>
<td>+0.25</td>
<td>Post-training: 20/16° (20/16° 1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20/12.5°</td>
</tr>
</tbody>
</table>

Figure 1. Illustration of the stimuli and task for Experiment 1.
to \([-4.6, 4.6]\) cd/m\(^2\) with the decreased step size of 0.92 cd/m\(^2\). The background luminance was lowered to 46.0 cd/m\(^2\). For three amblyopic observers, BK, GW, and GD, the amount of modification varied slightly depending on individual performance.

**Condition 6.** In addition to the increased target signal-to-noise ratios either for amblyopic eyes in the fovea or for normal eyes in the periphery, the bar width was increased from 0.1 deg to 0.3 deg for both the bright target and the noise, hence the fundamental frequency of the stimulus was decreased from 5 cpd to 1.67 cpd.

**Procedures**

All experiments were done monocularly in a room with dim overhead light (10 cd/m\(^2\)). The viewing distance was 1 m. The stimuli were presented to the fovea of all normal and amblyopic observers. For some normal observers, the task was also performed at 2.5 deg in the inferior visual field, in which case a fixation cross (20 × 20 min arc) was presented to the fovea.

**Preliminary experiments.** The luminance increment (i.e., noise step size) was set at the threshold for detecting a single bar on a blank screen. More specifically, a 40-trial session was completed by each observer. In every single interval trial, either a target bar or a blank screen was presented for 20 ms, and the observer had to respond whether a bar was presented or not. The luminance of the target bar was determined by Quest with the criterion set at 75%.

The lower target luminance in Condition 1 was set to an easily visible level while the rest of the stimulus square was filled with the luminance of the brightest noise for normal observers. In a preliminary session of 40 trials, a single stimulus frame uniformly filled with the brightest noise luminance, i.e., the sum of the background luminance and the largest noise increment (among the 11 noise increments), was presented in both intervals of each 2-interval forced choice (2-IFC) trial. In one of the two intervals, a brighter target bar was presented in the middle of the square, and its luminance is determined by Quest with a 95% criterion to ensure that it was easily detected. The observer had to respond as to which interval contains the target bar. The mean threshold luminance for all normal observers or the threshold luminance of individual amblyopic observer was taken as the low target luminance for the basic experimental conditions. The higher target luminance was set at around 75% threshold of discriminating a bright bar from the less bright bar of the fixed lower luminance described above, for the basic experimental conditions. A similar 2-IFC session of 40 trials was completed by every observer. A bar at the target location with the fixed lower luminance was presented in one interval, and one with a higher luminance was presented in the other interval. The observer had to respond by indicating which interval contained the brighter bar. The higher luminance is a variable and was determined by Quest with a 75% criterion for each trial. The mean threshold for all normal observers or the threshold luminance for individual amblyopic observer was taken as the high target luminance for the basic experimental conditions. Note that the luminance levels of the stimuli determined by methods described above may not generate the same performance with the presence of spatiotemporal noise as in the experimental sessions.

**Tasks.** At the beginning of each session, target bars with high and low luminances were presented side by side and the observer was instructed to memorize the luminance difference. Then, there was a 15-trial training session with feedback. In each training trial, either a bright or a less bright target bar, but no noise, was presented, and the observer was asked to respond whether it was a bright or less bright bar by pressing the appropriate key. Audio feedback was given.

The short training session was followed immediately by an experimental session. In each trial, the observer had to respond whether the target bar was bright or less bright by pressing a key (yes/no). No feedback was given. We used this discrimination task instead of a detection task in an attempt to avoid high temporal and spatial uncertainty, especially the former, which is often associated with such experiments and can severely contaminate the results. Each session had 200 trials and only one of the five conditions described above. A total of at least 2000 trials were collected for each condition for each eye.

**Data analysis.** We used reverse correlation to compute the classification images. The stimulus sequence of each trial is characterized by a 21 (time) × 11 (space) noise matrix and a 1–0 (1 for high and 0 for low) target luminance. The noise at a given time and space has a uniform distribution over all eleven luminance noise levels. The response for each trial is either 1 (i.e., bright bar) or 0 (i.e., less bright bar). All noise matrices and responses are grouped and averaged into 4 categories (Table 2).

The classification image is computed by

\[
\mu = \mu_{[1,1]} + \mu_{[0,1]} - \mu_{[1,0]} - \mu_{[0,0]},
\]

<table>
<thead>
<tr>
<th>Response</th>
<th>Stimulus</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hit</td>
</tr>
<tr>
<td></td>
<td>S1R1</td>
</tr>
<tr>
<td></td>
<td>False alarm</td>
</tr>
<tr>
<td></td>
<td>S0R1</td>
</tr>
<tr>
<td>(\mu_{[1,1]})</td>
<td>(\mu_{[0,1]})</td>
</tr>
<tr>
<td>(\sigma_{[1,1]})</td>
<td>(\sigma_{[0,1]})</td>
</tr>
<tr>
<td>0</td>
<td>Miss</td>
</tr>
<tr>
<td></td>
<td>S1R0</td>
</tr>
<tr>
<td></td>
<td>Correct rejection</td>
</tr>
<tr>
<td></td>
<td>S0R0</td>
</tr>
<tr>
<td>(\mu_{[1,0]})</td>
<td>(\mu_{[0,0]})</td>
</tr>
<tr>
<td>(\sigma_{[1,0]})</td>
<td>(\sigma_{[0,0]})</td>
</tr>
</tbody>
</table>

Table 2. Categories of stimuli and responses in computing classification images.
where $\mu(\cdot, \cdot)$ is the average $21 \times 11$ noise matrix of a given stimulus and response category. Each element in the classification image $\mu$ roughly represents the correlation of the noise at this spatial temporal location and the perception that the target is bright (it is slightly different from the standard definition of correlation in that the four categories are equally weighted regardless of the different number of samples in each group). Positive and negative correlations imply facilitation and inhibition, respectively.

The statistical significance of each correlation element is represented by the $z$-score:

$$z = \frac{\bar{x}}{\sigma},$$

where $\bar{x}$ is the estimated correlation at a given spatial temporal location, i.e., an element of the matrix $\mu$; the standard deviation $\sigma$ of the correlation is estimated using Bootstrapping, i.e., from each of the four groups, the noise matrices are sampled with replacement, averaged, and then combined to obtain one sample of the classification image, and the standard deviations of the classification image are computed from 3,000 such samples.

The $z$-scores above the Bonferroni corrected ($p$-value $< 0.05$) threshold of statistical significance (i.e., $|z| > 3.7$) are shown as red (positive) and blue (negative). This is a conservative estimate of significance (Chauvin, Worsley, Schyns, Arguin, & Gosselin, 2005). The False Discovery Rate method (FDR, Benjamini & Hochberg, 1995), which controls the expected proportion of true null hypotheses rejected out of the total number of null hypotheses rejected, instead of the traditional familywise error rate gave identical results.

**Results**

For the basic condition (Condition 1), the task performance for the normal and non-amblyopic eyes was above chance, whereas for about half of the amblyopic eyes it was near the chance level (50%). Increasing the target signal-to-noise ratios improved the performance in both the normal periphery on average and in the amblyopic fovea, so did decreasing the fundamental spatial frequency of the stimuli. For all but the first (basic) condition, the performance levels of all observers are comparable.

All normal observers show similar templates for detecting a luminance increment on a bright bar embedded in spatiotemporal noise, which is more clearly revealed by the average of all normal classification images (Figure 2, Table 4). Individual data are shown in the Supplementary material (Supplementary Figure S1). First, a temporal integrative zone is present at the target location from nearly 40 ms before the target onset until 30–40 ms after the target offset. Around the same period of temporal summation, two spatial inhibition zones are symmetrically distributed 0.1–0.15 deg from the target. At the target location, almost immediately following the temporal summation is the temporal inhibition zone that peaks at about 40 ms and extends to about 100 ms after the target presentation. The spatial correlation profile at the time of the target presentation (Figure 2, lower right graph) resembles a “Mexican hat” shape typical for a neural receptive field at various stages of visual processing (e.g., retinal ganglion cells, LGN, and V1 neurons). The temporal correlation profile at the target location (Figure 2, upper right graph) is consistent with the paired-pulse or flicker detection time course reported by others (Kelly, 1971a, 1971b; Rashbass, 1970; Watson & Nachmias, 1977). These results suggest that the summation-only temporal profile in the previous study may have been affected by the high temporal uncertainty (Neri & Heeger, 2002).

Another observation of Figure 2 is that the facilitative (red) and suppressive (blue) locations, including those with smaller absolute correlation values, in the classification image show an alternating pattern that is unlikely to be random, which suggests that part of the spatiotemporal template may be a product of the independent temporal and spatial functions, although obviously this is not true for the predominant mechanism or template.

On average, the classification images of the non-amblyopic eyes are normal or nearly so; however, the classification images of amblyopic eyes show a striking absence of inhibition (Table 4). The averaged classification images are shown in Figure 3. (The individual classification images are shown in Supplementary Figure S2.) Both the temporal inhibition that normally peaks about 40 ms after the target onset and the spatial inhibition adjacent to the target location are absent for all five amblyopic eyes. On the other hand, despite the reduction in inhibition, the temporal summation for amblyopic eyes does not seem to be much different from that of the normal.

In an attempt to understand what has caused the anomaly in the spatiotemporal template of the amblyopic eye, we presented modified stimuli to the normal fovea and hoped to reproduce the amblyopic template. First, the stimuli were blurred in space by a disk filter (see Methods section). The blurred stimuli had little effect on task performance (Table 3); neither did they result in a template that mimicked the amblyopic eye (Figure 4, Table 4; the individual classification images are shown in Supplementary Figure S3). All components, especially the inhibition zones, are present in the classification image, which distinguishes it from that of the amblyopic eye. Compared to those of the control condition (Figure 4, black dashed curves), classification images for blurred stimuli are scaled with larger spatial summation zones (Figure 4, red solid curves). This is presumably due to positive correlations between spatially adjacent pixels that result from the filtering process. Clearly, it is not simply the image blur that caused the anomaly in the amblyopic spatiotemporal template.
Second, the target signal-to-noise ratios of the stimuli were decreased for normal observer SK. The rationale is that perhaps the amblyopic eye could not detect any target and the resulting template is not relevant to this task, since after all, their performance (correct response rate in Table 3) is near the chance level (50%). Reducing target signal-to-noise ratios did make the task much more difficult and SK’s correct response rate decreased to 58%. However, SK’s classification image does not resemble that of the amblyopic eye (Figure 5). Another piece of evidence that the target signal strength is not a significant factor in determining the template comes from the amblyopic eyes’ null responses to stimuli with increased target signal-to-noise ratios (Figure 6, Table 4). Enhancing the visibility of the target did almost nothing for the amblyopic observers (Supplementary Figure S4). On the other hand, we know that these amblyopic observers did detect the target from their substantially improved performance (Table 3). Hence, the low visibility of the target is not likely to be responsible for the abnormal amblyopic template either.

Another possibility to which the abnormal amblyopic template may be attributed is that the spatiotemporal mechanisms in the normal and the amblyopic eyes may be shifted on the scale of spatial frequency (Levi, Waugh, & Beard, 1994). More specifically, the template may vary with spatial frequency of the stimuli, and the relationship between them may differ for the normal and the amblyopic eyes. Therefore, while 5 cycles/deg (i.e., the fundamental spatial frequency of the stimuli) probably elicits a template corresponding to mid-high spatial frequencies for the normal eye, it may elicit a template at the high end of the spectrum for the amblyopic eye. Presumably, the template used by the normal eyes is more efficient than that used by the amblyopic eye.

To test the hypothesis that amblyopic template is shifted in spatial scale, the fundamental spatial frequency of the stimuli was decreased from 5 cpd to 1.67 cpd by widening...
Discrimination task performance (percentage of correct responses) in all but the second conditions for the tested eyes of normal observers and both amblyopic (AE) and non-amblyopic (NAE) eyes of amblyopic observers. Condition 2 is excluded because the target position is filled by noise only. The asterisk denotes 2.5 deg in the parafovea.

### Table 3

<table>
<thead>
<tr>
<th>Condition</th>
<th>Eye</th>
<th>1</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS</td>
<td></td>
<td>68.3%</td>
<td>70.4%</td>
<td>—</td>
<td>73.7%*</td>
<td>75.2%*</td>
</tr>
<tr>
<td>SK</td>
<td></td>
<td>66.8%</td>
<td>69.4%</td>
<td>58.6%</td>
<td>68.1%*</td>
<td>73.5%*</td>
</tr>
<tr>
<td>MN</td>
<td></td>
<td>73.6%</td>
<td>—</td>
<td>—</td>
<td>66.9%*</td>
<td>—</td>
</tr>
<tr>
<td>EP</td>
<td></td>
<td>69.2%</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>CN</td>
<td></td>
<td>70.2%</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>BK</td>
<td>NAE</td>
<td>63.0%</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>AE</td>
<td>63.1%</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>GW</td>
<td>NAE</td>
<td>64.8%</td>
<td>—</td>
<td>—</td>
<td>68.9%</td>
<td>71.1%</td>
</tr>
<tr>
<td></td>
<td>AE</td>
<td>55.7%</td>
<td>—</td>
<td>—</td>
<td>67.1%</td>
<td>67.1%</td>
</tr>
<tr>
<td>BN</td>
<td>NAE</td>
<td>63.5%</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>AE</td>
<td>61.3%</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>JS</td>
<td>NAE</td>
<td>67.4%</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>AE</td>
<td>64.9%</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>GJ</td>
<td>NAE</td>
<td>63.0%</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>AE</td>
<td>53.2%</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>GD</td>
<td>NAE</td>
<td>65.8%</td>
<td>—</td>
<td>—</td>
<td>77.1%</td>
<td>75.5%</td>
</tr>
<tr>
<td></td>
<td>AE</td>
<td>58.3%</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Figure 3. Average classification images (left and middle) and the profiles (right) of perceiving a bright bar for amblyopic observers in both amblyopic (AE) and non-amblyopic (NAE) eyes. The correlation value is represented by the gray level of the grids in the background of the classification images. Statistically significant correlations after Bonferroni correction (i.e., $|z| > 3.7$) are overlapped by the positive (red) and negative (blue) $z$-score images that are interpolated using the bicubic spline method. Correlation profiles across the spatial and temporal target locations for both amblyopic (red solid) and non-amblyopic (blue solid) eyes are smoothed using the cubic spline method. The average normal correlation profile (black dashed) is superimposed. The positively correlated central areas (red) of the templates in both amblyopic and non-amblyopic eyes of all observers are not much different from those in the normal eyes. The negatively correlated areas are, on average, normal in non-amblyopic eyes but are missing in amblyopic eyes.
each bar in the stimuli from 0.1 deg to 0.3 deg and presented to two amblyopic eyes. Finally, normal classification images are seen in both amblyopic eyes in response to the modified stimuli (Figure 7, Table 4; individual classification images are shown in Supplementary Figure S5). The substantial change in the amblyopic template cannot be attributed to the task performance, which, on average, is not significantly different from the classifica-

tion images for the blurred stimuli show different spatial scales, but both positively and negatively correlated components are complete.

**Figure 4.** Average classification images (left) and the profiles (right, solid red) for detecting a luminance increment on a spatially blurred bright bar in similar noise stimuli for normal observers. The profiles for the control stimuli (dashed black, from the solid black curves in Figure 2) are superimposed. The classification images for the blurred stimuli show different spatial scales, but both positively and negatively correlated components are complete.

**Figure 5.** Classification images (left) and the profiles (right, solid red) of perceiving a bright bar embedded in distracters with enhanced luminance noise for a normal observer. The profiles for the control stimuli (dashed black, from the solid black curves in Figure 2) are superimposed. The classification image for the noisy stimuli lacks the temporal inhibitory component.
previous condition (i.e., increased target signal-to-noise ratios). Interestingly, the spatial correlation profiles in Figure 7 show that the spatial extents and shapes of the templates depend on the relative distance of the stimulus elements (i.e., distance normalized by bar width) instead of the absolute distance in degree. This result is not inconsistent with the amblyopic spatial classification image for detection reported by Levi et al. (2008), which demonstrated a moderate shift of the amblyopic template toward lower spatial frequencies (i.e., wider and flatter spatial template). The target used in Levi and Klein’s study was a fixed discrete frequency pattern (DFP) that had a negative–positive–negative luminance profile around the target, and they did not examine the effect of signal spatial frequency like the present study does by varying the stimulus spatial frequency. Our result suggests that the spatiotemporal mechanism of feature detection may depend on spatial scale.

As noted previously, we did not succeed in an attempt to produce the abnormal amblyopic template in normal foveal vision by either spatially blurring the stimuli or decreasing the target signal-to-noise ratios of the stimuli. Because of the close connection between amblyopic foveal vision and normal peripheral vision, we also estimated classification images from the periphery (2.5 deg eccentricity) of two normal observers using stimuli with increased target signal-to-noise ratios such that the performance levels were maintained above chance (Table 3). As Figure 8 and Table 4 show, like those of the amblyopic eye, the inhibition is missing in the normal peripheral classification image (individual classification images are shown in Supplementary Figure S6).

The subsequent question is whether a normal foveal template can be generated in the periphery by scaled stimuli. Similarly, the bar width was increased from 0.1 deg to 0.3 deg to decrease the fundamental spatial frequency of the stimuli from 5 cpd to 1.67 cpd, and the stimuli were presented to the periphery for two normal observers. The resulting classification images show partial restoration in spatial inhibition, but the difference is not statistically significant (Figure 9, Table 4; individual classification images are shown in Supplementary Figure S7). Again, the spatial profile of the template depends on the relative distance but not the absolute distance in degrees. Further investigation is required to confirm the possible common spatial mechanisms in simple feature perception shared by normal peripheral vision and the amblyopic foveal vision.

Table 4 lists p-values of various hypothesis tests of different experimental conditions or among subjects. Since the shape of the template can differ from person to person, any tests that require the absolute location information of the classification image would be inappropriate. Since we are mostly interested in the inhibition, we tested the spatial and temporal inhibitions separately. A minimum correlation value within a region of interest was selected to represent either spatial or temporal inhibition, and comparison between experimental conditions or subjects was conducted using these representative values.
Table 4. *P*-values for comparisons between experimental conditions and subjects. Statistically significant *p*-values are indicated by the bold and italic font. The regions of interest are labeled for each classification image: (1) space: the areas adjacent to the target in space on both sides (−60 to 60 ms, −0.5 to 0.5 deg, excluding the target location—the origin); (2) time: the frames following the target offset (0–200 ms, 0 deg, excluding the target location—the origin). The minimum correlation value in each region of interest represents spatial or temporal inhibition, which is used to compute the *F*-statistics for comparison between test conditions or subjects. This method eliminates possible individual differences in the shape of template. (a) Control conditions. (b) Stimulus manipulation in the normal eye. (c) Normal periphery. (d) Stimulus manipulation in the amblyopic eye.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Spatial inhibition (<em>p</em>-value)</th>
<th>Temporal inhibition (<em>p</em>-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Among normal eyes (foveal)</td>
<td>0.2516</td>
<td>0.5654</td>
</tr>
<tr>
<td>Among NAEs</td>
<td>0.0033</td>
<td>0.2475</td>
</tr>
<tr>
<td>Among AE</td>
<td>0.4054</td>
<td>0.3615</td>
</tr>
<tr>
<td>AE vs. NAE</td>
<td>0.0062</td>
<td>0.0010</td>
</tr>
<tr>
<td>(b) Normal blurred vs. normal control</td>
<td>0.0751</td>
<td>0.3615</td>
</tr>
<tr>
<td>Normal high noise vs. normal control</td>
<td>0.0665</td>
<td>0.0016</td>
</tr>
<tr>
<td>(c) Normal 2.5 deg vs. AE</td>
<td>0.3269</td>
<td>0.1257</td>
</tr>
<tr>
<td>Normal 2.5 deg low SF vs. normal 2.5 deg</td>
<td>0.0725</td>
<td>0.9629</td>
</tr>
<tr>
<td>(d) AE high S/N ratio vs. AE control</td>
<td>0.8851</td>
<td>0.3851</td>
</tr>
<tr>
<td>AE low SF vs. AE control</td>
<td>&lt;0.0001</td>
<td>0.0023</td>
</tr>
</tbody>
</table>

Spatial interactions among cluttered visual inputs affect both sensitivity and perceptual bias (e.g., apparent tilt) simultaneously. Crowding is normally described as the decreased sensitivity to a target among nearby similar distracters (Levi, 2008; Pelli, Palomares, & Majaj, 2004). What is missing in this description is that the crowded perception is not necessarily featureless; rather, crowding is often associated with a perceptual bias determined by the cluttered features. Parkes, Lund, Angelucci, Solomon, and Morgan (2001) reported that statistical attributes of a crowded feature (e.g., orientation) are spatially pooled instead of being lost and that the accessible average feature covertly serves as a perceptual pedestal (i.e., bias), which can either elevate or lower the feature detection threshold under certain conditions, although later studies argued that crowding cannot be reduced to mere feature averaging (Balas, Nakano, & Rosenholtz, 2009; Petrov & Popple, 2007).

In peripheral vision, flankers can produce very striking perceptual biases or illusions. Thus, the perceived tilt of a line may be either undergo assimilation, where it appears to be attracted toward slightly differently tilted objects, or repulsion, the exaggeration of the difference between the target and flankers when their orientations are grossly different (Solomon, Felisberti, & Morgan, 2004). These studies demonstrate the importance of not only the sensitivity loss but also the possible perceptual changes in understanding spatial interactions. While the spatial properties of crowding have been widely studied, the temporal properties of the spatial interactions have rarely been investigated.

Westheimer and Hauske (1975) reported that flankers with positive stimulus onset asynchronies (SOAs; i.e., flanker onset after target onset) have stronger interfering effects than those of simultaneously presented flankers on foveal Vernier acuity. Similar results have been found by others (Huckauf & Heller, 2004). Abnormal spatial interactions (crowding) are a well-known feature of strabismic amblyopia. In the present study, we systematically investigated the spatiotemporal interactions between flankers and the target, specifically on feature (i.e., orientation) perception, using the method of classification images.

**Methods**

**Observers**

Two observers with normal vision and two with strabismic amblyopia participated. All participants had substantial experience with visual psychophysical experiments and...
Figure 7. Classification images (left) and peak correlation profiles (middle and right, solid red) of perceiving a bright bar embedded in similar distracters with decreased fundamental spatial frequency, for amblyopic eyes. The bar width is increased to 0.3 deg. Profiles of control stimuli for the same amblyopic eye (solid blue) as well as the normal average profiles (dashed black) are superimposed. For correlation vs. distance profiles (lower middle), the distance is in degree. For correlation vs. normalized distance profiles (lower right), the distance in degree is normalized by the number of stimulus elements (i.e., number of bars). Classification images for both observers are corrected and appear normal with complete positively and negatively correlated zones. The correlation profile extent in space depends on the relative distance but not the absolute distance in degree, for both observers (lower middle and right).

Figure 8. Classification images (left) and peak correlation profiles (right, solid red) at 2.5 deg eccentricity for normal observers. Profiles in the fovea (dashed black) are superimposed. Classification images in the parafovea of normal eyes are noisy and lack regular anti-correlated zones seen in Figure 2.
were naïve as to the purpose of the study except the author. None of the amblyopic observers has eccentric fixation. The detailed characteristics of each observer are listed in Table 5.

**Stimuli**

Stimuli were generated by Matlab 7.01 with Psychtoolbox 2.0 and presented on a flat CRT monitor (30.1 cm², 1600 × 1400 pixels, 100 Hz, Sony). At 41% brightness and 98% contrast, the luminance gradient ranges from 3.8 to 133.2. The monitor was calibrated to generate luminance that is a linear function of pixel value.

The background luminance of the stimuli was 28.4 cd/m².

A stimulus sequence for each trial was composed of 9 frames. For normal observers, in every frame, there were either 11 or 9 possible positions evenly distributed along a horizontal line; each of the 5 or 4 positions on

![Correlation peak profile](image)

Figure 9. Classification images (left) and peak correlation profiles (middle and right, solid red) of perceiving a bright bar embedded in similar distracters with decreased fundamental spatial frequency at 2.5 deg eccentricity for normal observers. The bar width is increased to 0.3 deg. For correlation vs. distance profiles (lower middle), the distance is in degree. For correlation vs. normalized distance profiles (lower right), the distance in degree is normalized by the number of stimulus elements (i.e., number of bars). Compared to Figure 8, lowering the fundamental spatial frequency of stimuli partially, but not completely, restored the missing (spatially) anti-correlated zones of the classification images. The correlation profile extent in space depends on stimuli but not absolute distance in these conditions for both observers (lower middle and right).

Table 5.

<table>
<thead>
<tr>
<th>Observer</th>
<th>Age (years)</th>
<th>Strabismus (at 6 m)</th>
<th>Eye</th>
<th>Refractive error (diopters, D)</th>
<th>Line letter VA (single letter VA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS</td>
<td>29</td>
<td>None</td>
<td>R</td>
<td>−0.25</td>
<td>20/12.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L</td>
<td>pl/−0.25 × 119</td>
<td>20/12.5</td>
</tr>
<tr>
<td>LL</td>
<td>26</td>
<td>None</td>
<td>R</td>
<td>−1.50</td>
<td>20/12.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L</td>
<td>−0.50</td>
<td>20/12.5</td>
</tr>
<tr>
<td>Strabismic amblyopic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GJ</td>
<td>24</td>
<td>R EsoT 4–5λ</td>
<td>R (AE)</td>
<td>+3.50/−1.00 × 97</td>
<td>20/40⁻² (20/25⁺¹)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L (NAE)</td>
<td>pl</td>
<td>20/16⁻¹</td>
</tr>
<tr>
<td>AP</td>
<td>22</td>
<td>L EsoT 4–5λ and HyperT 2–3λ</td>
<td>R (NAE)</td>
<td>−1.25/−0.50 × 175</td>
<td>20/16⁻²</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L (AE)</td>
<td>−0.50/−0.25 × 60</td>
<td>20/50⁺¹ (20/40⁻²)</td>
</tr>
</tbody>
</table>

Table 5. Observers’ characteristics.
one side was independently filled with a high contrast (90%) Gabor patch with a probability of 0.5 and these Gabor patches were positioned symmetrically on both sides; therefore, the significant number of positions (excluding the target position) was 5 or 4 instead of 10 or 8 due to symmetry. Here, we assumed symmetry in the noise effect on perception because the distance from each stimulus patch to the fovea is presumably the same for both normal (in the lower visual field) and amblyopic (at the fixation) eyes. The middle (6th or 5th) position was occupied by the target Gabor patch in the middle (5th) frame and blank in all other frames. For amblyopic observers, there were only 4 possible positions on each side of the target. Each Gabor patch was tilted by a small angle either clockwise (i.e., left end higher, $\alpha$) or counterclockwise (i.e., right end higher, $\pi - \alpha$) and the magnitude of the tilt (around 6 deg) was determined for each individual (discussed in the Procedures section). Again, the tilt of the Gabor patch was symmetric around the target position, i.e., if the Gabor patch at position 1 was tilted 6 deg clockwise, then the one at position 11 was also tilted 6 deg clockwise.

The target in the middle position of the middle frame was cued in time and space. A short red line segment below the target, a red solid circle around the target, or a red dotted circle around the target was used as the cue that appeared at the onset of the target and disappeared at the offset of the target; hence, the cue was synchronized with the target in time. Since there was no significant difference between the results of the solid circle cue and those of the dotted circle cue, they are combined as one group in the analysis. The duration of each frame was 40 ms, and the entire sequence was 0.36 s (Figure 10).

A Gabor patch is mathematically defined as a two-dimensional sinusoidal wave multiplied by a Gaussian envelop:

$$G'(x, y) = -\lambda \cos\{\nu(x\sin\alpha + y\cos\alpha)\}e^{\frac{-x^2+y^2}{2\sigma^2}},$$  

where $\alpha$ is the tilted angle of the Gabor patch, $\nu$ is the frequency of the sinusoid, $\sigma$ is the standard deviation of either dimension of the Gaussian function, and $\lambda$ is a scaling constant. Here, $\lambda$ is selected such that $G'(\cdot)$ is within the interval $[-1, 1]$, with the minimum and maximum being $-1$ and $1$, respectively.

The contrast of a Gabor patch is defined as the Weber contrast:

$$C = \frac{L_{\text{max}} - L_{\text{background}}}{L_{\text{background}}}, \quad L_{\text{background}} \in (0, L_{\text{max}}].$$  

Therefore, the Gabor patch with contrast $C$ is

$$G(x, y) = L_{\text{background}}[1 + CG'(x, y)],$$  

where $C \in [0, 1]$.

**Procedures**

All experiments were done monocularly in a room with dim overhead light (10 cd/m²). The stimuli were presented at fixation for amblyopic observers or at either 4 deg or 5 deg in the inferior visual field of normal observers. A fixation cross (20 × 20 min) was presented to normal

![Figure 10. Illustration of the stimuli and task for Experiment 2.](https://jov.arvojournals.org/pdfaccess.ashx?url=data/journals/jov/933482/)
observers. The distance from the stimuli and the tested eye was 100 cm for normal observers or 300 cm for amblyopic observers.

**Preliminary experiments.** For each observer, the spatial frequency of the Gabor patches was chosen to be highly visible, i.e., half of the wavelength of a Gabor patch is between 1.5 and 2 multiples of the minimum angle of resolution (MAR) estimated by a letter chart. The Gaussian envelope standard deviation was 0.09–0.16 deg, resulting in 2–3 visible cycles in a Gabor patch. The tilt angle was set at the detection threshold absent of noise. In a 40-trial preliminary session, the stimulus, a single Gabor patch in the target position and no distracter, was tilted by a small angle (clockwise or counterclockwise), and the observer responded whether the target was tilted or not (i.e., yes/no). The tilt direction (i.e., clockwise or counterclockwise) was randomly selected for each trial, and the angle was determined by Quest with the criterion of 75%. The threshold angle $\alpha$ was taken as the amount of tilt of the Gabor patches for the observer. Table 6 summarizes the stimulus characteristics for each observer.

**Tasks.** In each trial, a complete stimulus sequence was presented, and the observer had to answer whether the target Gabor patch was tilted clockwise or counterclockwise by pressing a key. No feedback was given. After every five trials, there was a training trial in which only the target and the cue but no distracter was presented to the fovea of the observer, and the response to the training trial was not counted. Each block had 30 trials (not including training trials), and the number of blocks in each session was between 2 and 4. A total of 750–1700 trials were collected for each condition for each eye.

**Data analyses**

The reverse correlation method described in Experiment 1 was used to compute the classification images for the current experiment. Using the notation from Table 2, we can code “clockwise tilt” as 1 and “anti-clockwise tilt” as 0, and apply them to both the stimuli and the responses. The noise matrix is $9 \times 5$ (space) or $9 \times 4$ for normal observers and $9 \times 4$ for amblyopic observers. The spatial positions are labeled mono-sided 1 to 5 from near to far relative to the target, due to the symmetry of distracters around the target. The noise at a given time and space has the following distribution:

$$N = \begin{cases} 
0 \text{ blank} & \text{with probability } \frac{1}{2} \\
-1 \text{ clockwise} & \text{with probability } \frac{1}{4} \\
1 \text{ counterclockwise} & \text{with probability } \frac{1}{4}
\end{cases}$$

**Results**

Without distracters, the correct response rate is around 75% for each observer. With closely placed distracters, the performance is much worse for both normal peripheral and amblyopic central vision (Table 7), suggesting a crowding effect.

The main significant interaction between the distracters and the target is orientation assimilation (red areas), i.e., an observer is likely to perceive the target orientation as similar to that of the flankers, consistent with previous studies of normal peripheral vision (Parkes et al., 2001; Solomon et al., 2004). This assimilation effect is strongest at locations immediately adjacent to the target for every observer, as shown by the narrow vertical red zones on both sides of the target in Figure 11. The assimilation effect extends further in space for flankers presented simultaneously with the target and is limited in space for flankers with non-zero asynchronies. Interestingly, at locations beyond the assimilation area, the target–flanker interaction becomes repulsion (blue), i.e., an observer is likely to perceive the orientation of the target as opposite to that of the flankers. The repulsion effect in orientation perception as a function in space was recently reported in peripheral vision by Mareschal, Morgan, and Solomon (2010). We refer to it as “anti-crowding”; along with assimilation, the spatial interaction shows a “Mexican hat” profile (discussed below).

It seems unexpected that the assimilation zone for flankers with non-zero asynchronies is so narrow that the
The adjacent repulsion zone is still within the critical distance for crowding (i.e., $0.5E = 2.5\text{ deg}$ at $5\text{ deg}$ eccentricity or $2\text{ deg}$ at $4\text{ deg}$). However, for flankers presented non-simultaneously with the target, the crowding effect on the spatially closest flankers was substantially relieved since the target position was not occupied. “Anti-crowding” occurs immediately adjacent to assimilation in space, as shown by the blue areas in the $z$-score image and the dark gray squares of the classification images, although not all meet the statistically significant criterion (i.e., $|z| > 3$).

Another seemingly surprising result is that the time when the target is on (i.e., time 0) does not coincide with the peak of the assimilation in any condition. This may be similarly explained by the flankers being subject to more crowding while the target was on (i.e., the target acts to crowd the flankers). This is partly consistent with the results of earlier studies (Huckauf & Heller, 2004; Westheimer & Hauske, 1975).

The shape of the crowding map depends on the cue type. The circle cue induces a bimodal profile in time that demonstrates two crowding peaks about $80\text{ ms}$ before and $40\text{ ms}$ after the target presentation. The line cue, on the other hand, induces predominant crowding before and around the target presentation and is strongest near the target.
beginning of the sequence. The observers’ performance associated with the circle cue is moderately better than that associated with the line cue (Table 7). However, every observer expressed that the task was difficult and the guessing rate was high.

The amblyopic spatiotemporal interaction maps also demonstrate strongest orientation assimilation at the location adjacent to the target, with anti-crowding present at farther locations. The greatest extent of orientation assimilation occurs simultaneously with or immediately before the target presentation for normal eyes (Figure 11; positive correlations are represented by light gray squares), whereas it occurs after the target presentation (i.e., while the target is off) for the amblyopic eyes (Figure 12). “Anti-crowding” seems to be weakened as the eccentricity increases in the normal eye as well as in the amblyopic fovea. However, in order to know whether or not these differences are significant or meaningful, further investigation is necessary. In general, the amblyopic crowding map is not substantially different from that of the normal periphery.

**Discussion**

**Temporal mechanisms for simple feature discrimination**

Using a discrimination task and the method of classification images, we were able to estimate the spatiotemporal templates for luminance discrimination of both normal and amblyopic observers under various conditions. Our results for the normal template differ from what Neri and Heeger (2002) reported in that we showed a temporal inhibition component that peaks around 40–60 ms after the target presentation. As explained previously, we think this difference is due to the elimination of the temporal uncertainty using a discrimination task.

Our results also differ from the classic paired-pulse, three-pulse, or flicker sensitivity time course in that the latter is symmetric around the target presentation (time 0), i.e., there are two temporal inhibition lobes before and after the target presentation (Bergen & Wilson, 1985; Rashbass, 1970; Watson & Nachmias, 1977). Note that in the three-pulse study by Bergen and Wilson the 1st and 3rd pulses always had the same luminance. Although we are not certain as to the reason for such differences, we think they are probably related to two factors. First, the stimulus target in the present study has suprathreshold luminance, whereas the classic paired-pulse experiments were designed to test the luminance threshold. It has been suggested that the mechanisms involved in suprathreshold and near-threshold stimuli may be very different (Manny & Levi, 1982a, 1982b). Second, the presence of spatiotemporal noise that shares similar features with the target in our study could result in unknown interactions that can affect the perception, especially since the classic paired-pulse experiments did not have a spatial variable at all. Further investigation is necessary to understand the mechanism underlying these differences.

The classic paired-pulse temporal sensitivity time course has been fairly well modeled by a two-stage
process by Rashbass (1970). The first (linear) stage is temporal filtering. A working model of the temporal impulse response function was proposed and developed by numerous people (Watson, 1986). The modeled function is composed of a transient and a sustained component. The transformed function (i.e., filter output of the stimulus input) is then processed at the second stage where the squared $L_2$-norm (i.e., the sum of the squares of the filter output signal) of the function is computed, which serves as the basis for the perceptual decision. This model correctly predicts that the paired-pulse sensitivity time course is determined by the autocorrelation function of the temporal impulse function that is always symmetric about the zero asynchrony (i.e., the target presentation time). As mentioned above, our result is not symmetric about the target presentation since the inhibition component is only present after the target offset, and therefore, it cannot be accounted for by Rashbass’ model. Neither can it be accounted for by a simpler model in which the second stage (i.e., summation of squares) is replaced by simply taking the maximum of the filtered signal to be the base for decision criterion, which predicts a horizontally flipped profile (Figure 13).

On the other hand, the asymmetric U shape of the temporal profile of the normal template for discriminating a bright bar, whose inhibition component peaks at 40–60 ms and extends to about 100 ms after the target presentation, somewhat resembles the time course of the metacontrast or structure backward masking function that typically has a U shape and the masking effect is strongest when the mask onset is 60–100 ms after the target onset (Breitmeyer & Ogmen, 2000). What distinguishes these two visual functions is that there is no summation in backward masking, i.e., the masking effect always interferes with the target detection. This difference could be due to two facts: first, the task in the present study is to discriminate a target at a suprathreshold level, whereas it is often detection of the stimulus near its threshold level in studies on backward masking; second, in our study the target and the distracter that may be considered as the “mask” are overlapping in space and uniform in structure (in terms of a single bar), whereas in metacontrast or structure backward masking, the mask does not overlap with the target or has structures different from that of the target. However, the second reason stated above might not be characteristic for backward masking, since in a recent study by Saarela and Herzog, the iso-oriented mask that spatially overlaps the target (Gabor patch) had strongest contrast masking effects at its offset and onset with no facilitation for any SOA. At the peaks of the masking effect, the detection threshold of the target even exceeds the mask contrast (Saarela & Herzog, 2008). One possible mechanism of backward masking is that the propagation of the target elicited neural activity is interrupted by the mask presentation. Along the same line, it is reasonable to hypothesize that the discrimination of a bright bar relies on the proper pattern of neural activity, instead of the maximum activity or its “energy” that can be defined to be the summed squares as in Rashbass’ model. A bright stimulus initiates a train of neural spikes whose rates form a pattern along time, which has been demonstrated to be similar to that of the temporal impulse function (Albrecht, Geisler, Frazor, & Crane, 2002). Neurons that receive these activities in the next stage may recognize specific patterns of the activities and respond according to the degree of matching. In the case of the present study, the temporal template profile has the typical shape of a temporal impulse response; therefore, the neural activities elicited by an optimal stimulus (predicted by the template) will have a pattern that

![Figure 13](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933482/ on 10/12/2018)

**Figure 13.** (a) The temporal impulse function—the working model. (b) Simulated from 10,000 trials, the temporal template profile by setting the decision strategy to “report the presence of a signal if the maximum of the temporally filtered visual input is greater than a threshold.” Red dots are simulated results, dashed curve is the smoothed profile. (c) Simulated from 10,000 trials, the temporal template profile using Rashbass’ model. The decision strategy is “report the presence of a signal if the sum of squares of the temporally filtered visual input is greater than a threshold.” In each of the three panels, the horizontal axis is the Stimulus Onset Asynchrony (SOA), and the y-axis is either (a) temporal response or (b, c) correlation. The dashed gray lines are at y = 0.
mimics the shape of the temporal impulse response, given that the stimuli have suprathreshold visibility. This pattern is likely to be recognized by the neurons of the next stage as coming from a strong visual stimulus. In addition, the polarity of the target determines that the profile is asymmetric.

The temporal impulse function has been modeled to have a sustained and a transient component (Watson, 1986). A pure-sustained temporal impulse function is non-negative, whereas a pure-transient temporal impulse function has zero area (i.e., positive areas = negative area). Likewise, anything in between has both sustained and transient components. It has been found that stimuli of low spatial frequency are associated with transient temporal response and the transience weighs less or even disappears for stimuli of high spatial frequency. In the present study, the control stimulus had a fundamental spatial frequency of 5 cpd, which was decreased to 1.67 cpd for the normal peripheral and amblyopic central vision testing. At the relatively high spatial frequency, the amblyopic template shows a lack of inhibition both spatially and temporally, with the temporal summation hardly affected. Interestingly, the amblyopic template was completely restored to normal using stimuli with low spatial frequency. This is consistent with previous studies that also found that at low spatial frequency the difference in temporal sensitivity in both amblyopic and normal eyes was substantially decreased or even became unnoticeable (Bradley & Freeman, 1985; Manny & Levi, 1982a, 1982b). These results suggest that the temporal template profile as a function of spatial frequency is shifted toward the lower spatial frequency for amblyopic vision. When the stimuli were presented in the normal periphery, a template similar to that of the amblyopic eye was reproduced. Decreasing the fundamental frequency of the stimuli partly restored the spatial but not temporal inhibition mechanism. Therefore, we hypothesize that the spatial properties of normal peripheral vision are similar to that of amblyopic centric vision, but their temporal properties may be quite different.

Both normal peripheral vision and amblyopic central vision are crowded in space and time. In contrast to the well-studied spatial characteristics of crowding, the temporal properties of crowding have rarely been investigated. Note that the temporal properties of crowding are not the same as “temporal crowding,” which often refers to the analogue of crowding in time, i.e., the target visibility is impaired by nearby temporal flankers that are presented at the target location (Pelli et al., 2004). The most relevant study on the spatiotemporal properties of crowding was done by Westheimer and Haukse (1975). They investigated the crowding effect on foveal Vernier acuity by varying the target–flanker separation as well as asynchrony. With various target–flanker spatial configurations and fixed target presentation duration (50 ms), the crowding effect appeared at zero asynchrony and increased with target–flanker asynchrony up to 150 ms (i.e., backward crowding). In a recent preliminary study on letter crowding by Chung, a similar time course of crowding was found (Chung, personal communication). We also showed that the strongest crowding effect did not occur during the target presentation in the present study. However, both backward crowding generated by post-target flankers and forward crowding generated by pre-target flankers exist in our results, especially in the normal periphery. It is not clear as to what caused this difference, but it may be related to the fact that the target was embedded in a relatively long sequence composed of 9 frames lasting 0.36 s in the present study, whereas the stimulus sequence in the previous studies had only two frames, one of which contained the target. The temporal order of the target and the flankers was unlikely to be mistaken by an observer for the 2-frame stimulus sequence but could be easily be confused for the 9-frame stimulus sequence. Further investigation is required to test this hypothesis.

**Spatial mechanisms for simple feature discrimination**

The results from Experiment 2 illustrate the effect of spatial interactions on target orientation perception as a function of space and time. The orientation assimilation effect (i.e., positive correlation) is distributed close to the target location within the critical space for the tested eccentricities, therefore representing crowding (Parkes et al., 2001; Pelli et al., 2004). This orientation assimilation effect of crowding is consistent with previous reports (Parkes et al., 2001; Solomon et al., 2004). Mounting evidence suggests that crowding emerges from feature integration over an undesirably large area following feature detection (Levi, 2008; Pelli et al., 2004), and that crowding is like texture perception (Balas et al., 2009; Levi & Carney, 2009; Parkes et al., 2001). Several explicit mechanisms have been proposed to explain crowding. Averaging features from the clutter explains certain orientation configurations but not others, suggesting averaging may be a special case of more general pooling mechanisms (Parkes et al., 2001). Solomon et al. developed a unified model that accounted for both the sensitivity and the perceptual bias for crowding (assimilation) and the tilt illusion (repulsion), which incorporates a local modulated response followed by an opponent mechanism for orientation identification (Solomon et al., 2004). A computational model that computes summary statistics over the crowding zone based on texture synthesis successfully predicted human performance under crowding condition for a variety of tasks (Balas et al., 2009). Since the flanking effect on target orientation threshold was not the focus of the present study, the estimated spatiotemporal map for perceiving a crowded orientation is not intended to be a complete picture of crowding. Amblyopic spatiotemporal interaction maps differ from those of the normal periphery in some aspects...
such as the relative weights of forward and backward crowding, although the difference is small and will need more evidence to confirm.

An important finding of Experiment 2 is “anti-crowding” (i.e., repulsion) at locations farther away from the target, demonstrating an attraction–repulsion structure similar to the facilitation–inhibition spatial template for discriminating a luminance increment of a bright bar in Experiment 1 and those for other visual tasks. In physiology, retinal ganglion cells and LGN neurons both have the radial symmetric on/off receptive field with a center–surround profile; so do some V1 simple cells in selective orientations. This center–surround organization for orientation perception was recently reported by Mareschal et al. (2010), who showed that in normal periphery, cortical distance determines whether flankers cause crowding (assimilation) or repulsion (the tilt illusion).

The center–surround spatial template was also found in Experiment 2 for orientation discrimination, which extends as far as 2.8 deg on each side of the target at 4 deg eccentricity. It seems that the center–surround structure, with the center effect much stronger than the surrounding subtractive effect as shown by the difference of their relative weights, is quite general for visual perceptual mechanisms associated with both lower and higher level visual processes (Badcock & Westheimer, 1985; Kienzle, Franz, Scholkopf, & Wichmann, 2009; Levi & Klein, 2003; Wilson, 1978). Moreover, since the significance of such a spatial profile with both positive and negative components is to detect the change in an image, e.g., a bright light spot in the darkness or a shift in position or phase in an otherwise uniform texture, it is reasonable to hypothesize that crowding, which represents averaging or pooling in some sense, occurs in the centroid in the feature identification process. The general center–surround mechanism is obviously limited by a certain spacing constraint (e.g., either effect disappears at a large enough spacing) but does not have a one-to-one relationship with the absolute distance; rather, it seems to be a function of the spatially ordered image feature. More specifically, when a feature is not crowded either because it is relatively isolated in space or very different from the surrounding features, it can occupy one of the off–on–off locations; whereas when some features are crowded, the pooled features as an ensemble are processed as one unit (on/off) for differentiation. This pooling–differentiation visual information processing probably exists at multiple levels.

Cueing a crowded target has no effect on identification sensitivity (Felisbert, Solomon, & Morgan, 2005). Similarly, we found that the ratio of correct responses with distracters was much lower than that without distracters even with spatiotemporal cueing in our experiment. However, our results show that a cue does affect perception under spatiotemporal interactions.

Summary and conclusion

The normal template for luminance discrimination has an inhibition–summation–inhibition structure in space and a summation–inhibition time course. The amblyopic spatiotemporal mechanism lacks inhibition, especially the temporal inhibition. The absence of temporal inhibition for moderate spatial frequency stimuli may provide a low-level explanation for the abnormal “attentional blink” in amblyopia (Popple & Levi, 2008). Neither increasing the signal-to-noise ratio of the stimuli for the amblyopic eye or decreasing the signal-to-noise ratio of the stimuli for the normal eye or blurring the stimuli spatiotemporally caused any significant change in the template. However, by decreasing the fundamental frequency of the stimuli the normal inhibition was successfully restored in the amblyopic eye, suggesting that the amblyopic mechanism for luminance discrimination is shifted toward lower spatial frequencies.

The normal peripheral spatiotemporal map for orientation perception is characterized by strong perceptual assimilation (i.e., crowding) near the target and relatively weak perceptual repulsion (i.e., “anti-crowding”) farther from the target. Simultaneous, forward, and backward assimilation effects (i.e., crowding) are all present but affected by different cues, and interestingly, the simultaneous assimilation effect is relatively less strong. We noted that the center–surround organization here may be a general spatial template for both lower level and higher level visual processing for a wide variety of visual tasks. The amblyopic spatiotemporal interaction map for orientation perception is similar to that of the normal periphery, except that in general the spatiotemporal interaction is more widely distributed around and after the target presentation, and the “anti-crowding” seems to be weakened.

Acknowledgments

We thank Susana Chung for her comments. This research was supported by National Institutes of Health Grant RO1EY-1728.

Commercial relationships: none.
Corresponding author: Shuang Song.
Email: songsh@berkeley.edu.
Address: Vision Science Program, University of California, Berkeley, CA, USA.

References


