Evaluating spatiotemporal interactions between shapes

Michael Slugocki
Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Ontario, Canada

Catherine Q. Duong
Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Ontario, Canada

Allison B. Sekuler
Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Ontario, Canada

Patrick J. Bennett
Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Ontario, Canada

Spatiotemporal interactions between stimuli can alter the perceived curvature along the outline of a shape (Habak, Wilkinson, Zakher, & Wilson, 2004; Habak, Wilkinson, & Wilson, 2006). To better understand these interactions, we used a forward and backward masking paradigm with radial frequency (RF) contours while measuring RF detection thresholds. In Experiment 1, we presented a mask alongside a target contour and altered the stimulus onset asynchrony between this target–mask pair and a temporal mask. We found that a temporal mask increased thresholds when it preceded the target–mask stimulus by 130–180 ms but decreased thresholds when it followed the target–stimulus mask by 180 ms. Furthermore, Experiment 2 demonstrated that the effects of temporal and spatial masks are approximately additive. We discuss these findings in relation to theories of transient and sustained channels in vision.

Introduction

It is generally believed that local features encoded in early visual cortical areas (i.e., V1, V2) are integrated in extrastriate areas to form increasingly complex visual representations (Van Essen, Anderson, & Felleman, 1992; Kourtzi, Tolias, Altmann, Augath, & Logothetis, 2003; Ostwald, Lam, Li, & Kourtzi, 2008; Wilson & Wilkinson, 2015). The last two decades has seen advancement in understanding how midlevel visual areas combine low-level information to form representations of extended curves and simple shapes, but the majority of this work has used static contours (see Loffler, 2008, 2015 for reviews). Given that neurons throughout the visual pathway integrate information across space and time (Breitmeyer & Ganz, 1977; Lamme and Roelfsema, 2000; Hess, Hayes, & Field, 2003; Tanskanen, Saarinen, Parkkonen, & Hari, 2008), it is important to understand how midlevel representations may be altered by spatiotemporal interactions arising between shapes.

Radial frequency (RF) contours have been used by many visual researchers to probe midlevel representations of curvature along closed contours (Wilkinson, Wilson, & Habak, 1998; Jeffrey, Wang, & Birch, 2002; Loffler, Wilson, & Wilkinson, 2003; Habak, Wilkinson, Zakher, & Wilson, 2004; Habak, Wilkinson, & Wilson, 2006; Bell & Badcock, 2009; Bell, Wilkinson, Wilson, Loffler, & Badcock, 2009; Schmidtmann, Kennedy, Ordach, & Loffler, 2012). These stimuli are useful because they provide an easy way to manipulate features, such as curvature and angular frequency of curvature extrema, that drive population responses in area V4 of macaques (Gallant, Connor, Rakshit, Lewis, & Van Essen, 1996; Pasupathy & Connor, 1999, 2001, 2002). Compound forms of these stimuli also can be used to represent outlines of more complex shapes and objects and, thus, are useful in deconstructing complex forms into simpler components that are easier to study (Wilkinson et al., 1998; Wilson, Wilkinson, Lin, &
Several studies have demonstrated that spatial interactions between adjacent shapes can increase RF detection thresholds (Habak et al., 2004; Bell, Badcock, Wilson, & Wilkinson, 2007; Habak, Wilkinson, & Wilson, 2009). In the one study that examined temporal interactions between shapes, Habak et al. (2006) found that RF detection thresholds along RF contours were elevated significantly by the onset of a mask presented approximately 80–110 ms after the onset of the target and that shapes presented after the first backward mask did not increase the magnitude of masking. However, it remains unclear how spatial and temporal masking interact. Given that visual scenes typically consist of multiple moving objects, it is important to understand how spatial positioning of objects might contribute to disruptions in processing of curvature information over time. Therefore, the aim of the current study is to examine how spatial interactions might modulate dynamic processes involved in the perception of closed-contour shapes.

In Experiment 1, we use a spatiotemporal masking paradigm similar to Habak et al. (2006) while measuring RF detection thresholds. A spatial mask always appeared alongside targets, and a temporal mask was presented at one of 11 stimulus onset asynchronies (SOAs) relative to target–mask pairs. In Experiment 2, we examined possible spatiotemporal interactions by measuring temporal masking with and without a spatial mask. Results across both experiments show that the spatial mask has an additive masking effect beyond that evoked by a temporal mask alone. However, the additive effect between spatial and temporal masks is observed only when the temporal mask precedes the spatial mask. We discuss these findings in relation to theories of sustained and transient channels of vision.

Methods

Participants

Two of the authors (MS and CQD) and one naïve, experienced psychophysical observer (VAL) participated in Experiment 1. The mean age of the observers was 24.3 years (age range: 22–26), and all observers had normal or corrected-to-normal visual acuity. Experimental protocols were approved by the McMaster University research ethics board, and consent of the participant was collected prior to the start of the experiment.

Apparatus and stimuli

Stimuli were generated in MATLAB 7.10.0 (Math-Works, Natick, MA) on an iMac 3GHz Quad-core Intel Xeon computer and were displayed using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The visual display was a NEC Multisync FE700+ graphics monitor with a pixel resolution of 1,024 × 768 (86 pixels/°) and a refresh rate of 60 Hz. The display had a mean luminance of 35.2 cd/m² and was the only light source in the room. Stimuli were viewed binocularly at a distance of 131 cm, which was maintained through the use of a chin rest. From this distance, a single pixel subtended 41.5 arcsec.

Radial frequency contours were generated by sinusoidally modulating the radius of a circle in polar coordinates according to the equation:

\[ r(\theta) = \bar{r}(1 + A \sin(\omega \theta + \phi)), \]

where \( \theta \) is the angle in radians, \( \bar{r} \) is the mean radius of the contour, \( A \) is the amplitude of modulation expressed as a proportion of the radius of the circle, \( \omega \) is the radial frequency in cycles per circumference (cy/2\( \pi \)), and \( \phi \) is angular phase (Wilkinson et al., 1998). The modulation amplitude, \( A \), determines the magnitude of curvature: values could range between zero and one to prohibit overlap between the RF contour and the polar center.\(^1\)

The cross-sectional luminance profile was defined by a fourth derivative of a Gaussian (see Wilkinson et al., 1998) with a peak spatial frequency of 8 cy/° and a luminance contrast of 99%. Across all experimental conditions, target contours had a radial frequency of five and a mean radius of 1.14°. Baseline thresholds were measured in displays that contained only the target contour (i.e., no spatial or temporal mask) at three phases (45°, 135°, and 225°). Modulation amplitudes of RF spatial and temporal masks were set to 15 times the baseline detection thresholds because masks of higher modulation amplitude elicit stronger masking effects relative to masks appearing near threshold (Habak et al., 2004; Habak et al., 2006). The mean target–mask distance was 0.60°.

The phase of spatial and temporal masks was always 0°, whereas the target–mask phase could be 0°, 90°, or 180°. Varying the target–mask phase served two purposes. First, varying phase introduces spatial uncertainty about the location of deformed segments along target shapes. Spatial uncertainty promotes stronger summation of curvature signals (Green, Dickinson, & Badcock, 2018a, 2018b, 2018c), presumably because the uncertainty forces observers to monitor multiple spatial locations along a contour onscreen in order to detect modulations in curvature (Green et al., 2018a). Thus, masking effects that are a consequence of combining curvature signals along a contour should be amplified by varying phase across trials. Second, previous research has shown that masking strength is approximately linearly related to the phase alignment between two shapes (Habak et al., 2006).
Therefore, we can examine the effect of phase on masking strength across SOA conditions, albeit at a group level using amalgamated data (see Results sections for more detail).

To evaluate how dynamic interactions affect spatial interactions between shapes, one spatial mask was always presented alongside a target contour, and a temporal mask was presented at one of 11 SOAs relative to the target–mask pair (±280 ms, ±230 ms, ±180 ms, ±130 ms, ±80 ms, and 0 ms). At an SOA of 0 ms, the contrasts of the spatial and temporal masks were summed and averaged. Because the contrast of each mask was set to the same value of 99%, the effect of performing this averaging procedure is equivalent to using only a single spatial mask appearing alongside the RF target.

**Psychophysical procedure**

Detection thresholds were measured using the method of constant stimuli and a two-interval, forced-choice paradigm. Prior to the start of the experiment, observers underwent a 60-s light adaptation period, during which the observer fixated the center of the display, followed by practice trials with auditory feedback to ensure they were familiar with the stimuli and task. Observers initiated each trial by pressing the space bar on a computer keyboard. Upon initiating a trial, a fixation dot flickered on-screen for 50 ms and was followed, after a delay of 200 ms, by the presentation of two stimulus intervals that were separated by an interstimulus interval of 700 ms (Figure 1A). In backward masking conditions, each interval began with the presentation of a target stimulus that was followed by the presentation of a temporal mask. In forward masking conditions, each interval began with the presentation of a temporal mask followed by a target stimulus. In one interval, the target stimulus was a contour deformed according to Equation 1, and in the other interval, the stimulus was a comparison contour that was a circle. Observers were asked to identify which of the two intervals contained a deformed target contour by pressing one of two keys on the keyboard. A spatial mask contour appeared concurrently with the target and comparison contours. The duration of each temporal mask and each target/comparison stimulus was 30 ms. The SOA between the temporal mask and target stimulus, which was one of the 11 SOAs listed above, remained the same within a block of trials. Across trials and intervals, stimuli were spatially jittered 0.17° in a random direction from the center of the screen. Figure 1 illustrates a typical sequence of events for both forward and backward masking conditions.

A single experimental session contained seven different radial modulation amplitudes that were shown 30 times in random order for a total of 210 trials per block. The three target–mask relative phase combinations (0°, 90°, and 180°) were randomly interleaved within a given block. Experimental sessions consisted of either five or six blocks, and a minimum of eight experimental sessions were completed to ensure two thresholds were recorded at each SOA condition. Each session took approximately 1.5 h to complete.

**Data analysis**

The statistical computing software R was used to perform all analyses reported within this paper (R Core Team, 2017). Data for each block were fit using maximum likelihood estimation with a psychometric function defined as

\[
\psi(x; \alpha, \beta, \gamma, \lambda) = \gamma + (1 - \gamma - \lambda)F_W(x; \alpha, \beta),
\]

where \(x\) is RF modulation amplitude, \(\lambda\) is the lapse rate, \(\gamma\) the guess rate, and \(F_W(x; \alpha, \beta)\) is a Weibull function (Weibull, 1951) defined as

\[
F_W = 1 - \exp\left(-\left(\frac{x}{\alpha}\right)^\beta\right).
\]

Threshold was defined as the RF modulation amplitude yielding 75% detection accuracy.

Thresholds were analyzed using a mixed linear model estimated with the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). Degrees of freedom for the mixed model were approximated using the Kenward–Roger method (Kenward & Roger, 1997) as this method rescales the \(F\) ratios in addition to adjusting the degrees of freedom to better approximate \(F\) distributions for mixed linear models (Judd, Westfall, & Kenny, 2012). For brevity, we report only the \(F\) tests from the linear mixed-effects regression analyses (i.e., analysis of variance of type III sums of squares with Kenward–Roger approximation for degrees of freedom). Post hoc comparisons were performed using paired, two-tailed \(t\) tests with \(p\) values adjusted with the Holm–Bonferroni method and familywise \(x\) equal to 0.05 (Holm, 1979) unless otherwise stated.

**Results**

Figure 2 plots detection thresholds for each observer and group-averaged thresholds as a function of SOA between a temporal mask and target–mask pair. In general, RF detection thresholds were elevated relative to baseline measures across all SOAs tested, and forward masks elicited greater masking than backward masks (see Figure 3). For forward masks, a clear peak in masking occurred at SOAs of either 130 or 180 ms in each observer. No consistent peak was found in the backward masking conditions; however, a consistent improve-
A dip in the masking function was found at 180 ms. Results for Experiment 1 are summarized in Table 1, in which peak masking in the backward and forward masking conditions is listed for each observer.

These observations were confirmed statistically through the use of a mixed linear model. Because the zero SOA condition did not contain a temporal mask, it was omitted from these analyses. The mixed model was fitted with two fixed effects (Mask Type and SOA) and two random effects (Observer and Session). The analysis revealed a significant interaction between Mask Type and SOA, $F(4, 59.03) = 8.40, p < 0.0001$.

To further investigate the cause of the interaction, two separate mixed models, one for each masking condition (Backward and Forward) were performed. Each model was fitted with a fixed effect of SOA and two random effects (Observer and Session). For backward masks, the main effect of SOA was significant, $F(4, 26.42) = 3.32, p = 0.025$. Likewise, for forward masks, there was also significant main effect of SOA, $F(4, 26.24) = 6.38, p = 0.001$. The significant main effects of SOA were analyzed with follow-up pairwise comparisons. For backward masking conditions, post hoc comparisons revealed differences between thresh-

Figure 1. (A) A typical sequence of events for both forward and backward masking conditions. Temporal offset between the presentation of a mask and a target–mask pair was varied across 11 SOAs ($\pm 280, \pm 230, \pm 180, \pm 130, \pm 80, 0$ ms). (B) Example of RF contours used in Experiment 1. Target and mask contours are shown at 10% modulation amplitude for illustrative purposes.
olds obtained at an SOA of 180 ms and thresholds at SOAs of 80 ms and 280 ms SOA conditions ($p_{Holm} < 0.05$). For forward masks, post hoc comparisons revealed significant differences between thresholds obtained at an SOA of −80 ms and thresholds at SOAs of −130 ms and −180 ms SOA conditions ($p_{Holm} < 0.01$).

To examine in greater detail how masking strength changes as a function of SOA between forward and backward masks, difference scores were computed at corresponding SOAs between the two conditions as seen in Figure 4. At short and long SOAs, the strength of masking between conditions remained similar as evidenced by difference scores near zero. However, the magnitude of masking differed at intermediate SOAs as forward masks exerted greater elevations in threshold compared to backward masks. These data suggest that a temporal mask affects spatial interactions between adjacent shapes primarily at SOAs of ≈180 ms. It should be noted that masking increased as SOA increased from −80 to −180 ms but decreased as SOA increased from 80 to 180 ms (see Figure 2), which serves to further exaggerate the difference between thresholds at SOAs of ±180 ms. However, this result only reinforces our main finding that temporal masks interact with target shapes at both positive and negative SOAs at approximately 180 ms, albeit in potentially different ways. Pairwise $t$ tests were used to determine whether thresholds at SOAs of ±180 ms differ from the zero SOA condition in which only a spatial mask is present: A significant difference was found between thresholds for the zero and −180 ms, $t(7) = 3.20$, $p = 0.015$, and 180 ms, $t(7) = −3.24$, $p = 0.014$, SOA conditions.

<table>
<thead>
<tr>
<th>Observers</th>
<th>Peak negative SOA</th>
<th>Peak positive SOA</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS</td>
<td>180 ms</td>
<td>280 ms</td>
</tr>
<tr>
<td>CQD</td>
<td>180 ms</td>
<td>280 ms</td>
</tr>
<tr>
<td>VAL</td>
<td>130 ms</td>
<td>80 ms</td>
</tr>
<tr>
<td>Average</td>
<td>163.33 ms</td>
<td>213.33 ms</td>
</tr>
</tbody>
</table>

Table 1. SOAs for forward and backward masks that led to peak elevations in detection thresholds.

Figure 2. Results from Experiment 1. RF detection thresholds are plotted as a function of the SOA between a mask and target–mask pair. The dashed horizontal lines represent baseline thresholds measured in the absence of spatial and temporal masks. Note that only a spatial mask was present in the 0 SOA condition. Errors bars represent ±1 SEM.

Figure 3. Average elevation in detection thresholds relative to baseline for both forward and backward masking conditions, in which thresholds were collapsed across SOAs. For all three observers, strength of masking is greater in the forward relative to the backward masking condition as shown by larger elevations in detection thresholds compared to baseline performance. Error bars represent ±1 SEM.

Figure 2. Results from Experiment 1. RF detection thresholds are plotted as a function of the SOA between a mask and target–mask pair. The dashed horizontal lines represent baseline thresholds measured in the absence of spatial and temporal masks. Note that only a spatial mask was present in the 0 SOA condition. Errors bars represent ±1 SEM.

Figure 3. Average elevation in detection thresholds relative to baseline for both forward and backward masking conditions, in which thresholds were collapsed across SOAs. For all three observers, strength of masking is greater in the forward relative to the backward masking condition as shown by larger elevations in detection thresholds compared to baseline performance. Error bars represent ±1 SEM.
To visualize the effect of phase at each SOA, data for each observer were collapsed across sessions prior to fitting a psychometric function to ensure reliable estimates of threshold were obtained. Data were then averaged across observers at each SOA for each target–mask phase combination. As seen in Figure 5, at all SOAs, the largest elevations in threshold occurred when target and masks were phase aligned (i.e., 0°), and essentially no masking occurred when the target–mask phase was 180°. Also, the difference between forward and backward masking conditions was much greater in phase-aligned conditions.

Experiment 1 demonstrates that a temporal mask can modulate the effect of a spatial mask on RF detection thresholds. Compared to thresholds obtained with only a spatial mask, thresholds decrease when a temporal mask is presented approximately 180 ms after the target–mask pair and increase if the temporal mask precedes the target–mask pair by 180 ms. Consistent with previous masking studies, the largest elevations in threshold are observed when the target and spatial mask are phase aligned (i.e., 0°). To further investigate how the presence of a spatial mask affects temporal masking, in Experiment 2, we measured the magnitude of forward masking with and without spatial masks.

**Experiment 2**

**Methods**

**Participants**

Six young adults participated in Experiment 2 ($M = 24.50$ years; $SD = 3.45$, range: 20–26). Three observers (CQD, MS, VAL) had participated in Experiment 1, and three (EAM, LUC, CIV) were new, experienced psychophysical observers. All participants, excluding the main author (MS), were naïve with regards to the experimental hypotheses and had normal or corrected-to-normal visual acuity. Experimental protocols were approved by the McMaster University research ethics board, and consent of the participant was collected prior to the start of the experiment.

**Stimuli, apparatus, and procedure**

The stimuli, apparatus, and psychophysical procedure were the same as those used in Experiment 1 except for the following differences. First, all observers except CIV were tested with only a subset of forward (and zero) SOA conditions (0, −80 ms, −130 ms, −180 ms) in two masking conditions (i.e., spatial mask present vs. absent), for a total of eight experimental conditions (4 SOAs × 2 mask combinations). In conditions that used only a temporal mask, the zero SOA condition was identical to the baseline condition in which no mask appeared alongside the target contour. The sequence of events on each trial for spatial mask present and absent conditions is illustrated in Figure 6.

A single experimental session contained seven different radial modulation amplitudes that were shown 30 times in random order for a total of 210 trials per block. Experimental sessions consisted of eight blocks of trials, one block for each experimental condition, with block sequence randomized within a session. A minimum of two experimental sessions were completed to ensure at least two thresholds were recorded in each condition. Each session took approximately 2.5 h to complete.

After completing the main experiment, we measured thresholds in three observers (MS, VAL, and CIV)
using SOAs of ±80 ms, in which only a temporal mask was present. Each observer ran in a minimum of two sessions per condition, yielding two thresholds per observer per condition. Each session took approximately 1 h to complete.

**Results**

Results for Experiment 2 are shown in Figure 7. Masking functions varied across observers, especially in conditions in which a spatial mask was present. For most observers, elevations in detection thresholds were largest at the largest SOAs tested, in which the onset of the temporal mask far preceded that of target shapes. Furthermore, masking strength was greatest when a spatial mask was present compared to thresholds observed when absent, and the effect of the spatial mask was, on average, similar across SOAs.

The data in Figure 7 were analyzed with a linear mixed-effects model, with Spatial Mask (Presence vs. Absence) and SOA as two fixed effects and two random effects (Observer and Session). The ANOVA revealed a significant main effect of Spatial Mask, $F(1, 52.35) = 23.94, p < 0.0001$, and SOA, $F(2, 49.11) = 9.41, p < 0.001$. The interaction between Spatial Mask Presence and SOA was not significant, $F(2, 49.11) = 1.52, p = 0.23$.

Post hoc comparisons using the Holm–Bonferroni procedure were performed between SOAs collapsed across masking conditions. Comparisons between thresholds measured with SOAs of 130 and 180 ms and other SOAs tested were significant ($p_{Holm} < 0.01$ in all cases), but the comparison between 130 and 180 ms were not significant ($p_{Holm} = 0.31$).

Experiment 1 found that a temporal mask presented 80 ms after a target contour produced minimal masking when a spatial mask was presented simultaneously with the target. This result is surprising because Habak et al. (2006) found that a temporal mask at that SOA produced significant masking when a spatial mask was not present. To test whether removing the spatial mask would increase the effect of backward masks, we measured thresholds in three observers (MS, VAL, CIV) with temporal masks at SOAs of ±80 ms. We selected 80 ms because that SOA was near the peak of the masking function observed by Habak et al. (2006) for backward masks when no spatial mask was present. In our experiment, we found that backward masking was so great that some observers still did not reach ceiling...
performance at the highest amplitude modulations tested. Therefore, we fit psychometric functions to responses averaged across observers. The fitted psychometric functions, which can be seen in Figure 9, indicate that backward masking (SOA = 80 ms) was significantly greater than forward masking (SOA = -80 ms). That fact that significant masking occurred at an SOA of 80 ms is consistent with the results of Habak et al. (2006).

As in Experiment 1, we examined the effect of phase by estimating thresholds for each target–mask phase combination at each SOA. To increase the reliability of our threshold estimates, we collapsed data across test sessions before fitting psychometric functions to the data from each observer, and we then averaged thresholds across observers in each condition. Figure 8 shows that masking in the spatial mask present and absent conditions was greatest when the target–mask phase was 0° and that essentially no masking occurred when the phase was 180°. Also, the effect of SOA was much larger when the phase offset was 0°.

Figure 7. Results for Experiment 2 in which masking is evaluated at only negative and zero SOAs in the presence (dotted) and absence (solid) of spatial masks. Average RF detection threshold plotted as a function of SOA of a temporally offset mask. The dashed horizontal line represents the baseline detection thresholds with error bars representing ±1 SEM.

Figure 8. Average threshold for each target–mask phase combination plotted as a function of SOA for spatial mask present (right panel) and absent (left panel) conditions in Experiment 2. The dashed horizontal line represents the baseline detection threshold (averaged across observers). Errors bars represent ±1 SEM. In cases in which error bars are not visible, the standard error was smaller than the width of the symbols.

Figure 9. Psychometric functions fit to RF detection data averaged across three observers for conditions in which temporal masks (with no spatial mask) appeared at SOAs of ±80 ms. Error bars represent ±1 SEM.
Our results demonstrate that the presence of a spatial mask increases RF detection thresholds relative to thresholds obtained with a temporal mask and that the effect of the spatial mask did not (on average) differ significantly across SOAs. These results are consistent with the hypothesis that masking between shapes is the sum of two components: one due to the effect of a temporal mask and another due to the presence of a spatial mask. One important caveat is that our data suggest that potential interactions between spatial and temporal masking may differ significantly across observers (a point that we discuss further in the Section “Individual variability in the pattern of masking”). Also, as previously observed, elevations in threshold were largest for phase-aligned patterns with thresholds approaching baseline values for conditions with increasing degrees of target–mask phase offset.

Discussion

The current study investigated the effect of spatial masks on temporal interactions between shapes. Experiment 1 found that RF detection thresholds measured with forward, temporal masks elevated thresholds relative to thresholds obtained with a spatial mask alone. Conversely, thresholds measured with backward masks improved slightly relative to conditions obtained with a spatial mask alone. Experiment 2 found that masking produced by a mask that precedes the target and a mask presented concurrently with the target are (to a first approximation) additive. In other words, forward masking reflected approximately the additive effects of the temporal and spatial masks. Finally, in both experiments masking strength was strongly dependent on the phase alignment of patterns with masking decreasing with increasing phase offset between target and masking shapes. Overall, our results demonstrate that interactions between shapes can be explained by a simple additive model with static and dynamic components that are strongly modulated by phase.

Habak et al. (2006) argued that backward masking between shapes is driven by the first mask transient: When a stimulus sequence consisted of multiple masks presented after the target, only the first temporal mask produced significant masking. However, Habak et al. (2006) only examined the influence of a backward mask on spatial interactions between shapes across a limited range of SOAs (i.e., 80–110 ms). Experiment 1 demonstrated that backward masks presented 180 ms after the target–mask stimulus lower thresholds compared to conditions in which the temporal mask is absent. Furthermore, our results from Experiments 1 and 2 suggest that, for forward masking conditions, the first mask is not solely responsible for driving the magnitude of masking observed. Instead, spatial masks presented concurrently alongside a target contour can further elevate RF detection thresholds beyond the effect evoked by a temporal mask alone. Below, we discuss plausible origins for each shape-masking component in relation to theories of sustained and transient channels of vision.

Timescale of information processing along the visual hierarchy

The time needed for visual information to evoke neuronal responses (i.e., mean cortical latencies) differs dramatically across visual cortical areas (Nowak & Bullier, 1997; Schmolesky et al., 1998; Lamme & Roelfsema, 2000; Capalbo, Postma, & Goebel, 2008). For the purposes of our discussion, we are interested in the mean latency it takes for midlevel visual areas to process and encode representations of curvature along closed contours. An area of the visual cortex that has received much attention for encoding simple representations of shape is V4 as studies in macaques suggest that neurons within V4 produce population responses that represent curvatures at different polar angles relative to the center of a closed contour (Gallant et al., 1996; Pasupathy & Connor, 1999, 2001, 2002). This population code within V4 can also be used to reconstruct simple shapes (Pasupathy & Connor, 2001, 2002) and, therefore, can potentially serve as a foundation for building more complex representations of form at higher level cortical areas along the visual processing hierarchy (e.g., IT).

Physiological studies suggest that the mean response latency of neurons within V4 is approximately 100 ms after the onset of a visual stimulus (Nowak & Bullier, 1997; Schmolesky et al., 1998; Lamme & Roelfsema, 2000; Capalbo et al., 2008). Given that the response latency of V4 neurons is approximately 56 ms (Lee, Willford, & Maunsell, 2007), then this latency roughly corresponds to the SOA at which peak masking (∼130–180 ms) was observed in Experiment 1. This correspondence between maximal masking and V4 mean response latencies is consistent with the idea that V4 is a candidate area for processing shapes and that masking between shapes may reflect interactions among V4 neurons. To begin our discussion on how temporal and spatial factors might manifest within V4, we must briefly review theories of sustained and transient channels in vision (Kulikowski & Tolhurst, 1973).

Sustained and transient channels

Kulikowski & Tolhurst (1973) first postulated that visual information is processed by sustained and
transient visual channels. Sustained channels partake in prolonged processing of visual information and exhibit a long response latency after brief stimulation (Breitmeyer & Ganz, 1977; Breitmeyer, 1992; Breitmeyer & Ogmen, 2000). These channels are hypothesized to play a critical role in processing object features as prolonged responses are associated with generation of featural codes (Breitmeyer & Ganz, 1977; Breitmeyer, 1992; Breitmeyer & Ogmen, 2000). In contrast, transient channels are characterized by brief, rapid activation that reorients attention toward different spatial locations and to moving stimuli (Breitmeyer & Ganz, 1977; Breitmeyer, 1992; Breitmeyer & Ogmen, 2000). Properties of sustained and transient channels are similar to neurons in, respectively, the parvocellular and magnocellular pathways in old and new world monkeys (Nealey & Maunsell, 1994; Kremers, 1998; Yabuta & Callaway, 1998; Ogmen, Breitmeyer, & Melvin, 2003), and those pathways have been considered by some as the neural correlates of sustained and transient channels in humans (Ogmen et al., 2003).

Although sustained and transient channels serve different functional roles in processing information, reciprocal inhibitory interactions between channels afford opportunities for interchannel cross talk to occur (Breitmeyer & Ganz, 1977; Breitmeyer, 1992; Breitmeyer & Ogmen, 2000; Ogmen et al., 2003). Interchannel inhibition is characterized by a sudden termination or degradation in sustained responses to a stimulus by the activity evoked from processing of a new stimulus onscreen (Breitmeyer & Ganz, 1977; Breitmeyer, 1992; Breitmeyer & Ogmen, 2000; Ogmen et al., 2003). Intrachannel interactions are thought to differ from interchannel inhibition that arises between stimuli that are close together in time because behavioral studies have found evidence for intrachannel inhibition even when stimuli are separated by long temporal gaps (Breitmeyer & Ganz, 1977; Breitmeyer, 1992; Breitmeyer & Ogmen, 2000; Ogmen et al., 2003). Although theories of intrachannel inhibition are based mostly upon behavioral evidence, such inhibitory interactions do have plausible neuro-physiological correlates (Kruse & Eckhorn, 1996; Yabuta & Callaway, 1998).

Disruption in sustained patterns of activation via intrachannel or interchannel inhibition produce characteristically different masking functions (Breitmeyer & Ganz, 1977; Breitmeyer, 1992; Breitmeyer & Ogmen, 2000; Ogmen et al., 2003). Type A masking functions, which are thought to reflect intrachannel inhibition, are obtained when temporal offsets between the presentation of a target and mask are small (Breitmeyer & Ganz, 1977; Ogmen et al., 2003). Type A masking functions describe patterns of masking whereby thresholds are most elevated at SOAs occurring proximal to the onset of a target stimulus. At longer intervals between target and mask onset, transient responses from the second stimulus are able to interfere with sustained responses evoked by the first stimulus at higher levels along the visual hierarchy (Ogmen et al., 2003). This interference results in interchannel inhibition and a so-called type B masking function, in which masking occurs only at longer temporal offsets between two stimuli (Breitmeyer & Ganz, 1977; Breitmeyer, 1992; Breitmeyer & Ogmen, 2000; Ogmen et al., 2003).

Type A functions do a better job of describing the interference observed in Experiment 1 with backward masks, whereas type B functions are better for describing the results obtained with forward mask. One possible reason why type B masking functions are observed for forward masking conditions is that transient responses evoked by a target–mask pair disrupts sustained processing of forward masks within midlevel visual areas (i.e., V4, LOC). A similar pattern of masking was observed by Habak et al. (2006), who showed that detection of curvature along a target shape was severely impaired when a mask was presented approximately 80–110 ms after onset of the target. Instead of a mask transient degrading the discriminability of a target, our experiments suggest that the transient response evoked by a target–mask pair interacts with sustained processing of a temporal mask (see Figure 10). Our results from Experiments 1 and 2 are consistent with the idea that interruptions in sustained processing produce masking at longer SOAs.

Spatial masks modulate interchannel inhibition

In our current study, we assume that temporal masks affect perception of target shapes in one of two ways: (a) Evoked transient responses interfere with sustained processing of similar information upstream, or (b) sustained processing of a mask is disrupted by downstream transients evoked from stimuli appearing later in sequence. Both of these are examples of interchannel inhibition. Although explanations based on ideas of interchannel inhibition may explain why peak forward masking is observed at longer SOAs, it does not explain why the effect of a spatial mask on target discriminability is approximately constant across SOAs tested as seen in Experiment 2. Transient responses evoked by temporal masks likely do not activate regions along the visual pathway at the same time as target contours unless there is enough time between stimulus presentations for the transient to affect a region where sustained processing of similar visual information is observed. Therefore, we suggest that intrachannel inhibition is needed to account for the fact that spatial masks further elevating detection
thresholds beyond that attributable to the presence of a temporal mask.

Spatial masks interfere with target encoding at different levels of the visual hierarchy over the same time course at which target shapes are processed. In primary visual cortex, interference between a spatial mask and target shape are hypothesized to arise via weak local inhibition between orientation selective filters (Poirier & Wilson, 2006, 2007). Inhibition between local filters that process contour orientation results in reduced neural responses, thus propagating weaker signals to areas further upstream that integrate such information to represent shape (Poirier & Wilson, 2007). At higher visual areas, such as V4, interference between spatially adjacent shapes is attributed to the improper summation of curvature signals along a target and mask contours at locations of peak curvature (Habak et al., 2004; Poirier & Wilson, 2006). Such theories are based on the response properties of V4 neurons that respond to curvature extrema relative to the center of a visual stimulus (Pasupathy & Connor, 2001, 2002). Although such encoding schemes confer benefits in representing shapes, such as scale invariance, if curvature extrema occur at the same polar angle relative to the center of a shape, the response of neurons may saturate, resulting in an impaired ability to encode changes in curvature. Consistent with this idea, Experiments 1 and 2 found that masking is greater in conditions in which curvature extrema are aligned compared to conditions in which curvature extrema are misaligned, a result that is consistent with the results obtained in previous visual masking studies (Habak et al., 2004; Habak et al., 2006). Taken together, studies of spatial masking between shapes suggest that interference likely arises from interactions arising within the same visual area and possibly along a similar time course.

Results from Experiment 2 are consistent with the idea that spatial masks are interfering with processing within the same visual areas in which target shapes are encoded. However, it should be emphasized that intrachannel inhibition arising from spatial interactions between shapes is likely not restricted to a specific cortical area or level of processing, but occurs at different tiers along the visual hierarchy. Furthermore, spatial masking, which we have argued arises from intrachannel inhibition, appears to have an additive effect on temporal masking. As demonstrated in Experiment 2, a spatial mask only scaled thresholds by a constant factor across all SOAs tested and, therefore, did not interact with components of masking that are attributed to temporal masks. This result suggests that the mechanism(s) responsible for dynamic interactions between shapes, such as intrachannel inhibition, is (are) likely separate from those mechanisms, such as intrachannel inhibition, that contribute to static spatial interactions between shapes. Thus, although both intrachannel and interchannel inhibition likely affect target visibility when both spatial and temporal masks are present, their effects are additive rather than multiplicative.

Alternatives to curvature-based interpretations

Although our discussion has focused on the role of curvature in discriminating shapes, our data are agnostic regarding the specific feature being used to discriminate two RF contours. Indeed, recent research has questioned whether changes in the rate of curvature modulation are the critical feature used by observers to discriminate different shapes (Dickinson, Harman, Tan, Almeida, & Badcock, 2012; Dickinson et al., 2013; Dickinson, Cribb, Riddell, & Badcock, 2015; Schmidtmann & Kingdom, 2017; Dickinson, Haley, Bowden, & Badcock, 2018). For example, Schmidtmann and Kingdom (2017) described a two-stage model in which the difference between maximum and minimum curvature is the primary computation used to code for shape, not curvature change per se. Such theories of shape encoding based on the periodicity of curvature extrema are consistent with the results from Experiments 1 and 2 as offsets in phase between two RF contours resulted in less masking.

Based on neuroimaging studies that suggest radial and concentric gratings evoke strongest responses in intermediate visual areas (Gallant et al., 1996; Wilkinson et al., 2000), we speculate many shape-specific
interactions originate in area V4. Nevertheless, it is plausible that disruptions in shape discrimination originate in other visual areas. Such a claim is supported by recent work by Salmela, Henriksson, and Vanni (2016) that investigated shape representations using fMRI and multivoxel pattern analysis. Despite the advantage conferred in using multivoxel analysis in decoding multidimensional patterns of activation relative to older methods (Norman, Polyn, Detre, & Haxby, 2006), Salmela et al. failed to find pattern selectivity for radial frequencies in area V4 although RF-specific patterns of activation were found in other intermediate cortical areas (e.g., V3d, IPS0). However, Salmela et al. did observe that RF contours elicited strong activation within area V4 and suggested that the neuroimaging device used (i.e., fMRI) may lack the resolution needed to discriminate between neural patterns generated by different RF contours (Salmela et al., 2016). Thus, area V4 still remains an important candidate area to consider in contributing to the ability to distinguish between shapes.

In identifying visual areas that contribute to accurate discrimination of shapes, consideration of the task used to probe these areas is paramount. Although the spatial summation of curvature signals has been well studied (Hess, Wang, & Dakin, 1999; Loffler et al., 2003; Schmidtmann et al., 2012; Green, Dickinson, & Badcock, 2017; Green et al., 2018a, 2018b, 2018c), far less attention has been devoted to understanding how curvature signals are processed over prolonged durations (Habak et al., 2006; Green et al., 2018b). This is problematic because much of our understanding of shape perception may be limited to feed-forward processes that occur following a brief stimulus presentation and fail to consider how shape perception is affected by recurrent cortical processes connections to shape processing that are known to that operate over longer durations (Lamme & Roelfsema, 2000). Therefore, additional research is needed to better understand the time course of recruitment of visual areas in passing diagnostic information regarding shape identity.

Constraints on models of shape perception

Our data suggest that the mechanisms responsible for discriminating shapes operate over a prolonged time course. In considering a purely feed-forward model of visual processing along the ventral stream, our results are consistent with the hypothesis that shape discrimination depends on computations performed in intermediate visual areas, such as V4, which tend to have moderate-to-long response latencies to visual stimuli (Lamme & Roelfsema, 2000). Our results also are consistent with reports returned by Habak et al. (2006), who demonstrated that backward masking was greatest at longer SOAs (approximately 80–110 ms).

Together, these results place important constraints on the time course over which shape processing unfolds. For example, our results pose a problem for theories that propose shape discrimination depends on analysis only of low-level orientation filters (Baldwin, Schmidtmann, Kingdom, & Hess, 2016) as activity in primary visual cortex typically decays rapidly after initial stimulus onset (Celebrini, Thorpe, Trotter, & Imbert, 1993). Evidence from single-cell recordings of neurons in macaque inferior temporal cortex also suggest that the spike train of these neurons fail to encode information specific to shape identity until after approximately 20 ms (Kovcs, Vogels, & Orban, 1995). If shape discrimination depended only on analysis of output from V1 orientation-selective neurons, the largest interactions between shapes should occur at short SOAs, a prediction that is inconsistent with our data. However, our results do not rule out the possibility that extrastriate areas are performing local computations over large spatial regions (Baldwin et al., 2016).

Individual variability in the pattern of masking

For Experiments 1 and 2, there was notable interobserver variability in the pattern of masking observed. Given the multitude of factors that influence masking, it is not surprising that these functions can differ between observers. Masking a stimulus introduces noise to the signal (i.e., target) either at early levels of visual processing, such as those arising from spatial masks, or along higher levels of visual processing typically evoked by temporal masks (Enns & Di Lollo, 2000). The addition of this sensory noise can make it difficult for a decision-making process to categorize a stimulus based on this noisy visual representation, and consequently, this uncertainty is reflected as additional noise in the behavioral responses of an observer. Furthermore, the level of attention directed by an observer to a stimulus can also impact levels of performance. The role of covert attention has been shown to affect the perception of a target in a variety of visual tasks (see Posner & Petersen, 1990, for a review). Therefore, additions of noise at sensory encoding, along with how these noisy representations affect higher cognitive processes (e.g., decision and/or attention), at least partly explain why masking functions are so variable across observers. However, more research is needed in understanding how uncertainty via noise, sensory or otherwise, can affect human perception of shape.
Conclusion

To summarize, the aim of the current study was to investigate the impact of spatial over dynamic interactions between shapes in multishape displays. Results from Experiment 1 demonstrate that backward and forward masks have very different effects on RF detection. Specifically, the effect of a forward mask depended significantly on SOA with peak masking occurring for SOAs near –180 ms, whereas backward masks had a smaller effect with thresholds improving at a SOA of 180 ms. Results from Experiment 2 demonstrate that the presence of a spatial mask serves to increase thresholds above those obtained with a backward mask alone and that the effect of the spatial mask was nearly constant across SOA. Across both experiments, masking strength was strongly modulated by phase alignment between target and masking patterns. Overall, these results suggest that spatial and temporal masks contribute to masking in RF detection tasks and that their effects are (to a first approximation) additive. Our results are consistent with theories of inhibition between transient-sustained channels of vision, in which both intrachannel and interchannel inhibition are considered in describing spatiotemporal interactions between multiple shapes. Future work should aim to further elucidate whether interactions between shapes attributed to static and dynamic components of shape processing are additive or whether the mechanisms governing shape interactions may operate under nonadditive routines.

Keywords: radial frequency, global form, shape perception, masking, contour perception, curvature processing

Acknowledgments

Supported by NSERC Discovery grants (ABS and PJB) and the Canada Research Chair Program (PJB).

Commercial relationships: none.
Corresponding author: Michael Slugocki.
Email: slugocm@mcmaster.ca.
Address: Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Ontario, Canada.

Footnote

1 It is important to note that the relation between contour curvature and modulation amplitude is not straightforward. For example, a common definition of curvature is the rate of change of orientation with respect to polar angle (Pasupathy & Connor, 2001, 2002; Poirier & Wilson, 2006; Dickinson, Bell, & Badcock, 2013), which is a continuous quantity that depends on several parameters in addition to modulation amplitude (e.g., RF). Nevertheless, increasing modulation amplitude results in an average increase in contour curvature when other RF parameters remain fixed. Therefore, for the purposes of reporting changes in curvature sensitivity with a single number, in the remainder of this article, we express curvature in terms of modulation amplitude with the understanding that this quantity does not represent curvature at specific polar angles along the RF contour.

References

Capalbo, M., Postma, E., & Goebel, R. (2008). Combining structural connectivity and response...
latencies to model the structure of the visual system. *PLoS Computational Biology*, 4(8), e1000159.


