Functional MRI, ERP, and psychophysical measures show that contextual effects are orientation tuned and suppressive

Sung Jun Joo
Institute for Learning & Brain Sciences, University of Washington, Seattle, WA, USA

The response of V1 neurons to a stimulus placed inside the classical receptive field can be modulated by stimuli presented outside the classical receptive field. However, the specific nature of these contextual modulations is unknown. Both enhancement and suppression have been observed as well as variability across measurement methodologies. To assess whether the contextual effect is facilitative or suppressive, we measured neural responses to an oriented Gabor stimulus (“target”) in three conditions: in isolation, with two Gabor flankers that were the same orientation as the target, and with two flankers that were orthogonal to the target orientation. We show that the target-related fMRI response, event-related potential amplitude, and the amount of contrast adaptation are all lower when the flankers were the same orientation compared to both the isolated and orthogonal conditions. There was no evidence of response enhancement. These results all point to an orientation-tuned suppressive effect of contextual stimuli measured in the periphery that is well explained by models incorporating divisive or subtractive inhibition.

Introduction

The receptive field (RF) of a neuron in primary visual cortex (V1) is defined as the region of retinotopic visual space that elicits an increase in spike rate in response to the presentation of a stimulus. While stimulation outside the RF alone does not increase spike rate, stimuli placed outside the RF (“flankers”) can modulate the response to a stimulus placed in the RF (“target”). Though these contextual effects are well known, there is little consensus on their specific nature, functional role, or underlying mechanisms. Most previous contextual modulation experiments can be segregated into two broad categories: (a) those reporting enhancement—often referred to as “collinear facilitation” and (b) those reporting suppression—often referred to as “surround suppression.”

Collinear facilitation experiments have demonstrated increased behavioral sensitivity and increased spike rate when a target is flanked by elements that have orientations that align with the target orientation (Kapadia, Ito, Gilbert, & Westheimer, 1995; Kapadia, Westheimer, & Gilbert, 2000; Polat, Mizobe, Pettet, Kasamatsu, & Norcia, 1998; Polat & Sagi, 1993, 1994). On the contrary, surround suppression experiments have shown decreased perceived contrast and suppressed spike rate in the same stimulus orientation relationship between a target and surrounding flankers (Cavanaugh, Bair, & Movshon, 2002a, 2002b; Sceniak, Hawken, & Shapley, 2001; Zenger-Landolt & Heeger, 2003). Given these discrepancies in the literature in which different sets of stimuli were used and inconsistencies across different measurement methodologies (e.g., Kinoshita, Gilbert, & Das, 2009), it is difficult to draw definite conclusions about the nature of contextual effects. In the present study we attempted to resolve these discrepancies by using similar stimulus properties and configurations across multiple methodologies.

Recently, theories of mental disorder such as autism spectrum disorder and psychosis have suggested the excitation/inhibitory imbalance or reduced surround suppression as underlying neural mechanisms (Flevaris & Murray, 2015b; Foss-Feig, Tadin, Schauder, & Cascio, 2013; Rosenberg, Patterson, & Angelaki, 2015; Yoon et al., 2010; Yoon et al., 2009). Results from these studies have shown lack of surround suppression effects in fMRI responses and better behavioral performance on the same stimulus configuration as surround suppression experiments. Normalization models (Carandini & Heeger, 2012; Heeger, 1992; Reynolds & Heeger, 2009) that describe the suppressive effects from the surround can successfully capture these results. However, these results can also be interpreted as facilitation instead of reduced suppression, and thus, it is important to study the nature of these contextual modulations to better understand the neural bases of these mental disorders.

Here we used stimuli that are often used in collinear facilitation experiments (isolated Gabor patches) and relatively large target–flanker distances such that flankers are likely located outside of the RF of V1 neurons. We observed no evidence of collinear facilitation but show a remarkably consistent pattern of results across three different methodologies—fMRI, event-related potential (ERP), and psychophysics—all demonstrate significant suppression when the flanker orientation matches the target orientation. These results are well explained by orientation-tuned normalization models.

**Materials and methods**

All of the experiments described below examine the neural response to a single parafoveal Gabor stimulus (target) flanked by two Gabors (flankers)—one positioned above and the other positioned below the target (dimensions specified below). Across all methodologies we compared three conditions: the target presented alone (single), the target with flankers that matched the target orientation (same), and the target with flankers that were orthogonal to the target orientation (orthogonal). There were two target orientation conditions (vertical and horizontal; Figure 1). The same condition with a vertical target specifically examines possible collinear facilitation effects.

First, we conducted a human psychophysical contrast adaptation experiment (as described in more detail below) using a vertical target embedded in either the collinear (vertical) flankers or orthogonally oriented (horizontal) flankers. The hypothesis based on collinear facilitation effects predicts that the strength of adaptation to the vertical target would be higher when the target is flanked by vertical flankers compared with when it is surrounded by horizontal flankers. We tested this hypothesis across multiple target contrast levels (6.25%, 12.5%, 25%, 50%, and 100%) while maintaining the flanker contrast (50%) and other stimulus properties (i.e., center-to-center distance, target eccentricity) identical across different target contrast.

Given our stimulus configuration, there was a surprisingly similar result across multiple target contrast levels: The mount of adaptation was lower for the collinear configuration than the orthogonal configuration (Figure 2). To generalize this finding, we sought to measure contextual effects on both vertical and horizontal targets using multiple methodologies. Importantly, we kept the stimulus properties as similar as possible across different methodologies.

We slightly modified the stimulus properties for each modality to get better measurements. For the psychophysical contrast adaptation experiment, we set the target contrast to 25% because the effect (the difference in threshold ratio between the collinear and orthogonal configuration) was the largest. For fMRI and ERP experiments, we set the target contrast to 50% because we wanted to maximize the evoked response to the briefly displayed target stimuli. In the psychophysical contrast adaptation experiment, the target eccentricity was horizontally displaced 6° from the fixation point. However, in fMRI and ERP experiments we moved the target location (displaced 3° horizontally and 3° vertically from the fixation point) to clearly localize visual areas (V2 and V3) in the fMRI experiment, and to avoid ERP signals being cancelled out due to opposite sign of evoked responses along the calcarine sulcus. These subtle differences would not lessen our interpretation of the results.

**Functional magnetic resonance imaging**

Six observers including the author participated. Each observer gave written and informed consent in accord with the human subjects Institutional Review Board at the University of Washington, in adherence to the Declaration of Helsinki. All Gabor patches had a standard deviation of 0.72°, a spatial frequency of 2 cycles per degree (cpd), and were 50% contrast. The fixation point was placed 3° below the center of the display. The target stimuli were displayed in the upper quadrant of both visual fields 3° horizontally and 3° vertically from the fixation point. The center-to-center distance between the stimuli comprising a pattern was 3°.

One experimental session consisted of two localizer scans and eight stimulus scans. A localizer scan (252 s) consisted of three conditions each presented in 12 s blocks: (a) blank fixation (“B”), (b) stimulus block with checkerboards in the location of the targets (“T”), (c)
stimulus block with checkerboards in the location of the flankers (“F”). The fixation condition and stimulus condition were alternated during the scan (B-T-B-F-B, etc.). The localizer scan ended with blank fixation block. The checkerboards were counterphased flickered (6 Hz) and windowed using the same Gaussian envelope used for Gabor stimuli in the main experiment.

A stimulus scan (372 s) consisted of four conditions each presented in 12 s blocks: (a) blank fixation (B), (b) single, either vertical or horizontal (V or H), (c) same, in which flankers had the same orientation as the target (VVV and HHH), and (d) orthogonal, in which flankers had an orientation that was orthogonal to the target orientation (HVH and VHV). There were 10 repetitions per each condition in a stimulus scan. The fixation condition was inserted between the three stimulus conditions (B-single-B-same-B-orthogonal-B, etc.). The stimulus scan ended with blank fixation block. The checkerboards were counterphased flickered at 0.5 Hz to preclude fading effects and the target Gabors were flashed on (250 ms) and off (250 ms). The phase of target Gabor was always in-phase with the flanker Gabors. An attention-demanding contrast decrement task on the fixation mark was used throughout each stimulus scan to equate attentional state across the conditions. We found no significant differences between conditions in either accuracy (hit rates: 54.3%, 54.9%, and 55.9% for single, same, and orthogonal condition, respectively; a repeated measures analysis of variance [ANOVA]: \( F[2, 10] = 2.004, p = 0.185 \)) or reaction times (483 ms, 432 ms, and 439 ms for single, same, and orthogonal condition, respectively; \( F[2, 10] = 0.535, p = 0.601 \)).

Functional MRI data were acquired using a Philips Achieva 3T scanner using an eight-channel head coil and an echo-planar imaging sequence (repetition time, 2 s; flip angle, 70°; 31 axial slices of 3.5-mm thickness (no gap) and 3.44 × 3.44-mm resolution, field of view, 220 mm). Each scanning session began with a T1-weighted structural scan with 1- × 1- × 1-mm resolution, which was used for visualization of retinotopic visual areas. Visual cortical V1 was localized using standard retinotopic mapping and cortical-flattening techniques using BrainVoyager QX. Regions of interest (ROIs), corresponding to the retinotopic location of the target and flankers within V1, were determined using the localizer scan described above. Voxels within V1 with a significantly larger response to the target compared to the flanker locations were included in further analyses.

Time courses for each of the stimulus scans were extracted and averaged across voxels within V1. For each scan, the signal intensity in each condition at each
of eight time points time-locked to the onset of the stimuli was averaged, beginning 4 s before the onset and ending 10 s after the onset. The two time points at 2 and 4 s before trial onset served as baseline. The averaged time courses were converted to percent signal change by subtracting the corresponding baseline measurement and then dividing by that value. The resulting time course for each condition was then averaged across scans. The signal intensity that averaged from 4 to 10 s poststimulus presentation was used as the measured response for each condition. This single number was then averaged across observers and a repeated-measures ANOVA and planned-comparison paired $t$ tests were used for statistical evaluation.

**Event-related potentials**

Eighteen observers including the author participated. Except for the first author, all were naive observers. All naive observers were paid $20/hr for participation. Again, each observer gave written and informed consent in accord with the human subjects Institutional Review Board at the University of Washington, in adherence to the Declaration of Helsinki. The stimuli were generated and controlled by Presentation (Neurobehavioral Systems, Inc., Berkeley, CA) software on a PC, and they were displayed on a 21-in. CRT monitor (60-Hz refresh rate). The viewing distance was approximately 70 cm. The spatial stimulus arrangement was identical to the fMRI experiment. On a given trial, flankers appeared before the target onset and remained throughout the trial. After a random duration chosen from a uniform distribution between 1 and 2 s, targets were briefly flashed for 100 ms. After the target offset, the flankers remained in the display for 500 ms. The intertrial interval was 2 s. Observers were asked to maintain fixation and to limit eye blinks to the intertrial interval.

One experimental block consisted of 12 trials [2 orientation (vertical and horizontal) $\times$ 3 stimulus conditions (single, same, and orthogonal) $\times$ 2 repetitions]. Trials were randomized within a block. Observers completed 22–51 blocks (264–612 trials). Observers initiated each block after a 5-s break by pressing a designated key in the button box. The first block served as practice.

EEG waveforms were recorded using BioSemi active Ag-AgCl electrodes from 64 sites. The signals were referenced to the left mastoid during online acquisition and re-referenced to the average of right and left mastoids offline. Vertical electrooculography (EOG) was measured using an electrode placed below the left eye and horizontal EOG was measured using an electrode placed at the outer canthus of the right eye. The signals were digitized at a sampling rate of 256 Hz.

EEG epochs started 100 ms before the target onset and lasted 400 ms after the target onset. Each waveform was baseline corrected to the average voltage of the interval $-100$ ms to 0 ms before the target onset and low-pass filtered at 40 Hz to remove high-frequency noise. Trials with waveforms that had a larger than 50 $\mu$V peak-to-peak vertical EOG amplitude and that exceeded $\pm$50 $\mu$V on other electrodes were excluded as these were trials deemed to be contaminated with eye blinks or other sources of noise. Data from five observers were discarded due to excessive artifact rejection (>40% of trials). The resulting waveforms were averaged across conditions individually for statistical analyses and then averaged across observers for visualization.

P1 amplitude, the first positive peak at $\sim$150 ms in our data, on six electrodes (Oz, O1, O2, POz, PO3, and PO4) was measured by averaging the ERP amplitudes during the time window of 130 to 170 ms. These electrodes were centered over the maximum of the P1 component as determined through visual inspection of the scalp topography. These individual amplitudes were averaged across six electrodes to represent P1 amplitude.

**Human psychophysical contrast adaptation**

Seven observers participated in each target orientation (vertical and horizontal) version including the author. Observers participated in two sessions (left and right visual field). The experiments were conducted in a dark room and head position was stabilized using a chin rest. Stimuli were presented on a 19-in. linearized CRT monitor with vertical refresh rate of 60 Hz. The distance between observers and the monitor was 50 cm. We used a video attenuator device (Video Switcher, Xiangrui Li, Los Angeles, CA) to generate 10-bit grayscale luminance values (Li, Lu, Xu, Jin, & Zhou, 2003). We used the MATLAB Psychtoolbox (Brainard, 1997; Pelli, 1997) on a PC to create stimuli, control stimulus presentation, and record responses.

Stimuli consisted of Gabor patches ($\sigma = 0.72^\circ$, spatial frequency = 2 c/pd, sinusoidal counterphase-flickering at 2 Hz). A fixation point (0.48$^\circ$ in diameter) was displayed at the center of the display. The distance between the fixation point and the center of the target location was 6$^\circ$. To measure the contrast detection threshold for a target, we used two randomly interleaved independent QUEST (Watson & Pelli, 1983) staircases. The detection task was a two-interval forced choice (2IFC) task where observers indicated which interval had the target. Each interval (200 ms) was indicated by a high-pitched tone and there was a gray blank (300 ms) between intervals. Auditory feedback was given for an incorrect response. There were 41
trials (20 trials per each staircase) in a session. The response of the first trial was discarded. The last contrast values of two staircases were averaged to estimate an observer’s contrast detection threshold for 82% performance. If the two staircases did not converge observers were asked to rerun the session.

Adapting stimuli consisted of a target and flankers. The center-to-center distance between stimuli was 3°. Observers were initially adapted for 30 s, followed by the first 2IFC task trial. A 5-s top-up adaptation period was inserted between subsequent trials to maintain stable adaptation. A 500-ms gray blank was inserted before each trial began. During this blank period, a black line was displayed next to the fixation point to indicate the beginning of a trial. The target location was always marked during both the adaptation and task periods to remove effects of location uncertainty on the detection task (Petrov, Verghese, & McKee, 2006). To equate the attentional state across conditions observers performed a contrast decrement task on the fixation mark during adaptation periods (Bi, Cai, Zhou, & Fang, 2009). The contrast decrement (10%) was displayed for 150 ms and the onset of the contrast decrement was selected randomly from a uniform distribution between 1.5 and 2 s. To quantify the amount of adaptation we defined the threshold ratio between detection threshold before and after adaptation (threshold_after / threshold_before). We measured the amount of adaptation with two different target orientations (vertical, horizontal) embedded in three conditions: single, same, and orthogonal. The target and the flanker contrasts were set to 50% and 25%, respectively.

Normalization model

We implemented a modified version of a recent normalization model (Reynolds & Heeger, 2009). We used the following parameters that approximately matched the spatial configuration of the above experiments. The width of stimuli (σ) was 15 and the center-to-center distance was 60. Note that the unit of these spatial parameters is arbitrary. The orientation tuning width was set to 10°. We used 100% contrast stimuli. Different contrast levels did not change the pattern of results. For excitatory drive (E) of hypothetical neurons we fixed the spatial extent of the RF as 5, and the orientation tuning width as 40°. For inhibitory drive (I) we fixed the spatial pooling width as 50, and the orientation pooling width as 5°. We used the following equation to calculate the response of a neuron (R):

\[ R = \frac{E}{I + k} \]

where we set the value of k as 0.01. The pattern of response did not change across these spatial and orientation parameters.

Results

Human fMRI

The fMRI response in V1 was measured in the retinotopic location of the central target, which was localized in a separate scan that alternated between flickering checkerboards in the target location and flanker location (Figure 3A). The responses of the target ROI were well localized—the checkerboards in the flanker location did not evoke an fMRI response significantly above the response to a blank screen fixation period (Figure 3B; t[5] = 0.89, p = 0.41).

Figure 3C and D shows fMRI responses in the V1 ROIs to vertical (Figure 3C) and horizontal (Figure 3D) targets across flanker conditions. A repeated measures ANOVA revealed that there was no significant main effect of target orientation, \( F(1, 5) = 0.19, p = 0.681 \), or an interaction between target orientation and flanker condition, \( F(2, 10) = 1.158, p = 0.353 \). However, there was a significant effect of flanker condition, \( F(2, 10) = 9.432, p = 0.005 \), indicating contextual effects on the target response.

For both horizontal and vertical targets there was a significantly smaller response in the same compared to the orthogonal condition: \( F(1, 5) = 17.813, p = 0.008 \), for the vertical target and \( F(1, 5) = 14.234, p = 0.013 \) for the horizontal target. The fact that we observed suppression in the vertical target condition in the same versus orthogonal condition is not consistent with collinear facilitation: If facilitation were occurring, we would expect the opposite pattern of results (greater response to same vs. orthogonal). Furthermore, there was significant suppression in the same condition compared to the single condition with a horizontal target, \( F(1, 5) = 6.723, p = 0.049 \). The response to the target in the orthogonal condition was not different from the response to the target in the single condition for both target orientations (ps > 0.05). Together, these results suggest that suppression is greatest when flankers share the orientation of the target stimulus, which is consistent with previous electrophysiological findings demonstrating orientation-tuned surround suppression.

Human ERP

We also performed an ERP experiment using similar stimulus configurations to examine the timing of the effect. The ERP response was measured to the onset of the target in the three different flanker arrangements. Figure 4A shows the timeline for each trial. Importantly, the flankers were presented first for a random duration between 1 and 2 s, followed by the target for...
Again, the flankers preceding the target was done to “configure” the surround suppression mechanism before target onset. The flankers remained for an additional 500 ms so that their offset would not contaminate the ERP response to the target.

The first prominent ERP component that was observed in response to the target was a positive deflection that peaked at approximately 150 ms (P1). Previous research has suggested that the P1 originates from early extrastriate visual areas (Clark, Fan, & Hillyard, 1995; Di Russo, Martínez, Sereno, Pitzalis, & Hillyard, 2002). The C1 component—which is believed to originate in V1 (Clark et al., 1995; Di Russo, Martínez, Sereno, Pitzalis, & Hillyard, 2002)—was not observed in our data. The scalp distribution of the P1 was confined to the occipital pole (Figure 4B). To characterize the magnitude of the P1, signals from the six electrodes around the occipital pole (Oz, O1, O2, POz, PO3, and PO4) were averaged. We calculated the magnitude of the P1 as the average amplitude between 130 and 170 ms. As with the fMRI experiment, a repeated measures ANOVA showed no significant effect of orientation, $F(1, 12) = 0.995, p = 0.338$, or interaction between target orientation and flanker condition, $F(2, 24) = 0.354, p = 0.705$. Figure 3C and D shows the average waveforms and P1 amplitudes for each target orientation in each flanker conditions. Planned contrasts showed that, consistent with the fMRI experiment, the P1 amplitude was smaller for the same condition compared to orthogonal ($F[1, 12] = 5.607, p = 0.036$ for vertical target; $F[1,12] = 12.633, p = 0.004$ for horizontal target) and single conditions ($F[1, 12] = 11.911, p = 0.005$ for vertical target; $F[1, 12] = 23.160, p < 0.001$ for horizontal target). These results indicate that when the flanker orientation matches the...
Figure 4. The procedure and results in the ERP experiment. (A) An example trial. On a given trial, flankers appeared before target onset. The initial flanker duration was randomly chosen from a uniform distribution between 1 and 2 s. The targets were flashed for 100 ms. After target offset, the flankers remained in the display for 500 ms to reduce the influence of flanker offset to the target response. (B) The scalp topography at 150 ms after target onset. The color bar represents visually evoked potential. We included ERPs from the six electrodes (Oz, O1, O2, POz, PO3, and PO4) near the occipital pole for data analysis. (C) The averaged ERP waveforms for the vertical (left panel) and horizontal (right panel) target across observers for the three stimulus conditions: single (blue), same (green), and orthogonal (red). For each observer, the P1 amplitudes of each electrode were measured by averaging the amplitude within the temporal window of 130–170 ms after the target onset (shaded region). P1 amplitudes of each electrode were then averaged and used for the statistical analysis. (D) The P1 amplitude for the vertical (left panel) and horizontal (right panel) target. *p < 0.05; **p < 0.01; ***p < 0.001. Error bars are the SEM across observers.
target orientation, there is neural suppression at the initial processing of the target.

**Human psychophysical contrast adaptation**

Across two separate methodologies (human fMRI and ERP) we observed smaller neural responses to the target when it matched the orientation of the flankers. An important question is whether this neural modulation has direct perceptual consequences. Specifically, we measured psychophysical contrast detection thresholds for a target after adaptation to the different flanker configurations—either when the flanker orientation matched the target orientation or when the flanker orientation was orthogonal to the target orientation. Based on the results from fMRI and ERP experiments, we had a strong prediction that the adaptation strength would be higher in the orthogonal condition than the same condition.

Observers were initially adapted to a stimulus for 30 s before performing a 2IFC detection task. A 5-s top-up adaptation period was inserted between trials to maintain stable adaptation (Figure 5A). The target location was always marked during both the adaptation and task periods to remove any effects of location uncertainty on the detection task (Petrov et al., 2006). To equate the attentional state across conditions observers performed a contrast decrement detection task on the fixation mark during adaptation periods (Bi et al., 2009).

To quantify adaptation strength, we calculated the ratio of each observer’s contrast detection threshold for a target before and after adaptation. Our assumption is that more adaptation—as indexed by an increase in postadaptation detection thresholds—reflects stronger neural activity in response to the adapting stimulus (Blake, Tadin, Sobel, Raisian, & Chong, 2006; Blakemore & Campbell, 1969; Boynton & Finney, 2003; Carandini, Movshon, & Ferster, 1998; Dragoi, Sharma, & Sur, 2000; Engel, 2005; Fang, Murray, Kersten, & He, 2005; Kohn & Movshon, 2003; Larsson, Landy, & Heeger, 2006; Priebe, Churchland, & Lisberger, 2002).

We found less adaptation in the same compared to the orthogonal condition, (paired two-tailed t test, \( t[6] = 6.145, p < 0.001 \) for vertical target; \( t[6] = 2.801, p = 0.031 \) for horizontal target; Figure 5B). Furthermore, there was less adaptation with the vertical target in the same condition compared to the single condition, \( t[6] = 2.998, p = 0.024 \). In an additional experiment we found that the horizontal and vertical flankers alone (no stimulus in the target position) did not produce any adaptation, indicating that differences in adaptation observed between the same and orthogonal condition were due to differences in the orientation relationship between the target and flankers and not due to differences in the flankers themselves (Figure 5C).

**Normalization model**

Our human fMRI, ERP, and psychophysical results are consistent with previous findings of a suppressed neural response to the target when the orientation of the flankers matches the target. One well-known and influential model of early visual cortical processing that can explain this general effect incorporates divisive normalization (Cavanaugh et al., 2002a; Heeger, 1992). We implemented a recent version of this model (Reynolds & Heeger, 2009). The model predicts the neural response to the target by dividing the response of a linear orientation-selective RF by the summed response of linear units with neighboring RFs. To account for our data all that is required is (a) a reasonable spatial pooling parameter—that is, we assume that the responses are being pooled across a distance that includes the flanker stimuli—and (b) that the pooling is orientation tuned—that there is more weight attached to neighboring units that share the orientation of the target than neurons that are tuned to orientations that do not match the target. Figure 6 shows an implementation that produces an output that closely corresponds to our actual measured data across all of our methodologies. This pattern is very robust across a range of spatial and orientation tuning parameters.

**Discussion**

Across multiple methodologies—fMRI, ERP, and psychophysics—we observed a reduced neural response when the orientation of a pair of flankers matched the target orientation. We have also shown that a well-established model of early visual cortical processing can explain the effect: orientation-tuned divisive normalization from neighboring neurons. In particular, we observed no evidence for response enhancement induced by the flankers—for example, similar to that described as “collinear facilitation.” The neural suppression in our data regardless of collinearity between the target and flankers is consistent with orientation-tuned surround suppression in early visual cortex (Allman, Miezin, & McGuinness, 1985; Blakemore & Tobin, 1972; Cavanaugh et al., 2002a; DeAngelis, Freeman, & Ohzawa, 1994; Maffei & Fiorentini, 1976; Sillito, Grieve, Jones, Cudeiro, & Davis, 1995).

In contrast, human neuroimaging studies have shown inconsistent results. Steady-state visual evoked potential (Polat & Norcia, 1996) and ERP (Khoe,
Freeman, Woldorff, & Mangun, 2004) studies that used small foveal or perifoveal (0.67° from fixation) stimuli have found enhanced responses to the target with collinear flankers, whereas a simultaneous EEG/magnetoencephalography (MEG) recording study has reported suppressed responses to the collinear configuration (Haynes et al., 2011). A majority of fMRI studies on surround suppression documented suppressed fMRI responses in early visual cortex for the collinear stimulus configuration (Nurminen, Kilpeläinen, Laurinen, & Vanni, 2009; Williams, Singh, & Smith, 2003; Zenger-Landolt & Heeger, 2003). What caused these inconsistent results? Can we predict contextual effects—whether facilitation or suppression—based on what we know from the literature?

First, contextual effects depend on the target eccentricity. Psychophysically, when a small foveal target is used to measure contextual effect, the contrast detection threshold for the target decreases when the target is surrounded by collinear flankers (Polat & Sagi, 1993, 1994). However, when measured in the periphery, this psychophysical contextual effect becomes suppressive (Petrov & McKee, 2006; Williams & Hess, 1998; Zenger-Landolt & Koch, 2001; but also see Lev & Polat, 2011). Second, contextual effects also depend on the contrast of the target (Levitt & Lund, 1997; Polat et
In particular, when a target was surrounded by high contrast flankers, there was collinear facilitation effect for the low contrast target, whereas suppression effect was found for the high contrast target in a cat’s V1 neurons (Polat et al., 1998). Similar results have been observed in human fMRI responses (Tajima et al., 2010). These findings are in concert with the fact that the summation field of V1 neurons increases as the contrast of grating stimulus decreases (Sceniak, Ringach, Hawken, & Shapley, 1999), making flankers occupy portions of the high-threshold spiking RF “fringe” regions (Cavanaugh et al., 2002a). Thus, the inconsistent results that have been documented in the literature—whether collinear surround facilitate or suppress the neural response—might be explained by the target eccentricity and how the RF of a neuron is organized (Michel, Chen, Geisler, & Seidemann, 2013). Based on these previous results, measuring contextual modulation while manipulating stimulus properties such as the size of each stimulus and the distance between the target and flankers may provide a fruitful avenue for future research to understand the complex nature of contextual effects.

Another factor that might affect the contextual effect is attention. Collinear facilitation effects in the fovea depend on attention (Freeman, Sagi, & Driver, 2001): Only attended flankers facilitated the target detection threshold while unattended flankers did not. In the periphery, this attention-dependent facilitation effect was observed only for a subset of subjects (Shani & Sagi, 2005). In a recent fMRI study, Flevaris and Murray (2015a) demonstrated that attending to the central target led to the suppression effect while attending to the upper flanker led to the facilitative effect in V1. In our experiments, we avoided attention-driven contextual effects by drawing attention from stimuli using an attention-demanding fixation task.

Collinear facilitation through horizontal connections in V1 has been suggested as an underlying neural mechanism for basic visual tasks such as edge detection and contour integration (Kapadia et al., 1995; Li, 2002). Note that the contour stimuli in the previous experiments that showed facilitation effects were embedded in a background comprised of random orientation stimuli (Kapadia et al., 1995; Kourtzi, Tolias, Altmann, Augath, & Logothetis, 2003; Li, 2002). In these stimulus configurations, however, facilitation effects could be due to stimulus attribute such as saliency or pop-out (Kastner, Nothdurft, & Pigarev, 1997; Knierim & van Essen, 1992). Furthermore, our results and other single-cell recording studies suggest that the neural response to a stimulus within contour stimuli is more likely to be suppressed (Cavanaugh et al., 2002b). How the visual system utilizes suppressed neural response in early visual cortex to give rise to contour perception remains an open question.
It has been suggested that suppressive signals in orientation-tuned surround suppression originated from higher visual areas and communicated through feedback connections (Angelucci & Bressloff, 2006; Bair, Cavanaugh, & Movshon, 2003). However, the functional role of these feedback signals remains unknown. It could be one way of achieving normalization (Carandini & Heeger, 2012) that mechanistically inhibits the neural response due to the presence of contextual stimuli in an image. In contrast, it might show a different coding scheme in early visual cortex that is sensitive to high-level interpretation of an image. For example, we recently showed that the neural response in early visual cortex was modulated by grouping based on long-range orientation patterns (Joo, Boynton, & Murray, 2012) and perceived surface structures (Joo & Murray, 2014).

In summary, contextual effects in V1 have been received much attention because they may provide the link between neural activity and behavior at the first stage of visual processing in the cortex. However, it has been difficult to reconcile the opposite effects—facilitation or suppression—that have been observed in the previous experiments with different stimuli, and within or across different methodologies. Our results suggest that contextual effects in V1 are largely suppressive, at least for the geometry of stimulus we used in our experiments. More research should be conducted to understand how the visual system utilizes these suppressed responses for visual tasks.

**Keywords:** contextual modulation, fMRI, ERP, contrast adaptation, collinear facilitation, surround suppression

**Acknowledgements**

I thank Scott Murray, Elizabeth Huber, and Geoffrey Boynton for helpful comments and discussion on the manuscript.

Commercial relationships: none.
Corresponding author: Sung Jun Joo.
Email: sjjoo@uw.edu.
Address: Institute for Learning & Brain Sciences, University of Washington, Seattle, WA, USA.

**References**


