Spatial structure of contextual modulation

I. Mareschal

School of Psychology & Australian Centre of Excellence in Vision Science, The University of Sydney, Sydney, New South Wales, Australia

School of Biological and Chemical Sciences, Psychology, Queen Mary University of London, London, UK

C. W. G. Clifford

School of Psychology & Australian Centre of Excellence in Vision Science, The University of Sydney, Sydney, New South Wales, Australia

Contextual effects are ubiquitous in vision and provide a means for detectors with localized receptive fields to encode global properties of a stimulus. Although the nature of the neural connections is complex, the majority of evidence supports the Gestalt idea of collinearity; interactions are greatest when the target and surround orientations are spatially aligned to form a contour. Here we create a novel stimulus that simultaneously probes all areas around a detector to determine which spatial positions influence perception in human observers. We find that the surrounding spatial areas that contribute most to contextual effects for our perception of orientation and motion are not confined to a specific location. Rather our results reveal that human perception displays some interobserver variability in the weighting of detector interactions that is largely independent of collinear structure. We propose that these more extensive surround stimuli reveal how complex visual structure may modulate performance in a manner that is not easily predictable using more conventional stimuli.

Introduction

The world around us appears coherent even though detectors in our visual system only convey information about small localized areas of space known as their receptive fields (Enroth-Cugell & Robson, 1966; Hubel & Wiesel, 1968). Despite this piecemeal representation of the visual scene, the human brain reconstructs a faithful and coherent representation of the complex visual images it receives. A clue to understanding how this is achieved is the discovery of contextual effects whereby neighboring detectors interact to signal the global properties of a stimulus (Gilbert & Wiesel, 1990; Levitt & Lund, 1997; Nelson & Frost, 1978; Nothdurft, Gallant, & van Essen, 1999; Sillito, Grieve, Jones, Cudeiro, & Davis, 1995; Walker, Ohzawa, & Freeman, 1999). In human psychophysics the majority of evidence supports the importance of collinearity; interactions are greatest when the target and surround orientations are spatially aligned, evidenced in a variety of phenomena such as contour integration (Field, Hayes, & Hess, 1993; Kovacs & Julesz, 1993), crowding (Mareschal, Morgan, & Solomon, 2010; Toet & Levi, 1992), contrast facilitation (Polat & Sagi, 1993), and orientation illusions (Kapadia, Westheimer, & Gilbert, 2000; Schwartz, Sejnowski, & Dayan, 2006). Indeed, in their pivotal contour paper, Field et al. (1993) proposed the concept of a local association field, whereby detectors tuned to similar orientations and aligned “end to end” (rather than side to side) had enhanced interactions to signal continuity. In their model, integration can occur between filters that are not perfectly aligned, as long as their relative orientations do not exceed 50°–60°. Recently, there has been a burgeoning interest in linking the tuning properties of detectors to the statistics of natural images (Felsen & Dan, 2005; Rust & Movshon, 2005). Given that natural images are rich in highly structured statistical properties that define edges, the selectivity of contextual effects for collinearity might provide a neural substrate optimal to process natural scenes (Schwartz et al., 2006; Sigman, Cecchi, Gilbert, & Magnasco, 2001). There are two caveats, however, to the above findings; the first is that most of the stimuli used to test this idea have strong second-order orientation cues that bias results (Morgan & Baldassi, 1997). The second is that the presence of additional surround stimuli along the sides of a target can modulate the strength of collinear alignment in both contrast facilitation (Solomon &
Morgan, 2000) and contour detection (Dakin & Baruch, 2009), suggesting that conventional stimulus configurations do not capture the complex layout of target–surround interactions.

Physiological experiments have revealed that both classical and non-classical receptive fields can have nonhomogenous layouts (e.g., Nishimoto, Ishida, & Ohzawa, 2006; Walker et al., 1999). Using a noise stimulus larger than the neuron’s receptive field, Nishimoto et al. (2006) measured properties of subfields within the receptive field. Here we were motivated by a similar idea to examine contextual interactions underlying human perception. Specifically, we assessed which surround areas (spatial hotspots) influence our perception of a target when the entire surround is stimulated.

Materials and methods

Apparatus and observers

A Dell Optiplex computer running Matlab (MathWorks Ltd.) and incorporating elements of the PsychToolbox (Brainard, 1997) was used for stimulus generation and recording subjects’ responses. Stimuli were displayed on a calibrated Diamond Digital monitor (1024 × 768 pixels; 85 Hz) driven by the computer’s Radeon graphics card and viewed binocularly. At the viewing distance of 57 cm, 1 pixel subtended 2.1 arcmin. The authors and six naïve observers served as subjects. All experiments adhered to the Declaration of Helsinki guidelines.

Stimulus and procedure

We created a novel surround that consisted of irregularly shaped abutting windows that contained one of two opposite stimuli such that globally the stimulus averaged to an uninformative mean for the task but that, locally, sections were dominated by one of the two stimuli. Borders between the windows were used to create a new mask on each trial and were defined by the zero crossings of low-pass filtered white Gaussian noise (Figure 1b). Their shapes provided no cue to orientation thus bypassing problems associated with second-order information. We hypothesized that if certain areas of the surround (spatial hotspots) exert greater influence on the target, then the observer’s perception of the target will be determined by how the local stimulus statistics map onto these areas (Figure 1c).

Orientation

The stimuli consisted of a concentric annular surround (outer diameter 8.9°) separated by a 2 pixel wide gap from the circular target (diameter 2.2°) that contained a 1.84 c/° grating that, in separate runs, was either horizontal or vertical. The surround consisted of irregularly shaped abutting windows whose shapes were defined by a bipolar mask (Figure 1b). Borders between the windows were defined by the zero crossings of low-pass filtered white Gaussian noise that was recreated on each trial (using one of two filter cutoffs; 8 c/image and 6 c/image). Observer IM tested an additional two filter cutoffs of 10 c/image and 4 c/image.). The polarity value of a window in the mask determined the orientation of the 1.84 c/° grating contained within it; in white sections (pixel value = +1) the grating was tilted +15° from the target, and in black sections (pixel value = −1) the grating was tilted at −15° from the target (Figure 1c). Both orientations covered the same total area of the surround. All gratings were 30% contrast and their spatial phases were independently varied on each trial. Observers fixated a central square (1 s) that was extinguished during stimulus presentation (150 ms) then reported whether the target appeared tilted clockwise (CW) or counterclockwise (CCW) of vertical or horizontal using the keys “a” and “l,” respectively. Observers completed a minimum of 2000 trials per condition in blocks of 100 trials, collected over several weeks. A control condition was performed using surround orientations (+60° and −60°) that do not elicit a tilt illusion (e.g., Westheimer, 1990).

Lightness

The central target was an elongated vertical bar set to the mean grey (0.84° × 0.28°) within a square stimulus that subtended 4.4°. The surround consisted of windows (as above, with filter cutoff of 12 c/image) onto one of two polarity-reversed, vertical, square-wave gratings whose spatial frequency matched the width of the target bar (1.78 c/°). The stimulus was presented in two spatial locations, to the right of center and in opposite polarity to the left of center (inner edges separated by 4.2°). The observer was instructed to scan (free view) the stimulus for 1 s and then reported with a key press which side appeared lighter, completing 500 trials in blocks of 100 trials.

Hotspots maps were constructed from the bipolar masks in the following manner: Masks on trials where the observer judged the left target as lighter were summed together and masks on trials where the observer judged the right target as lighter were sign inverted and then added to the summed left lighter masks.
Motion

The central target (diameter 2.8°) contained isotropic fractal noise scaled to cover the full range of lightness values that drifted upwards (4.0°/sec). The surround (outer diameter 15.4°) contained windows (filter cutoff 16 c/image) onto different sets of isotropic fractal noise that drifted ± 36° from the direction of the target (3.6°/sec). Observers fixated a central square (1 s) that was extinguished during stimulus presentation (300 ms) then reported whether the direction of the center was CW or CCW of upwards motion, completing 2000 trials in blocks of 100.

White’s predictions

The spatial resolution of our technique is inevitably limited by the granularity of the mask. Using a square target rather than an elongated bar fails to elicit a compelling illusion, suggesting that the illusion may arise in large part due to the elongated nature of the stimulus. In order to determine whether this was the case, we created a series of predictive hotspots maps that we cross correlated with each observers’ data. To generate the predicted data for the White’s illusion experiment, one thousand random bipolar (black/white) masks were generated. Those in which the central pixel was black were negated and then summed with those in which the central pixel was already white to produce a summed image of one thousand masks each centered on a white pixel. The resulting image was taken as an estimate of the effective point spread function of our classification images arising from spatial correlations within the mask.

Results

Figure 2 shows the surround areas of influence (hotspots maps) in the tilt illusion stimulus when the observers’ task was to report whether the central target appeared tilted CW or CCW of vertical (Figure 2a) or horizontal (Figure 2b). Masks used to create the tilt illusion stimulus were stored on each trial into the CW or CCW category, based on the observer’s response to the target. At the end of all trials the masks in the CCW category were summed together and subtracted from
the sum of the masks in the CW category to build the hotspots maps.

For three of the four observers (GV excepted) the hotspots maps for the vertical and horizontal targets appear similar across the two conditions (Figures 2a & b), highlighting a lack of dependency on the stimulus configuration. This suggests no greater weighting of interactions at the ends of the stimulus, contrary to what would be predicted from Kapadia et al. (2000) who report that collinear interactions (flankers aligned end to end) produced greater repulsion than side to side (which sometimes produced weak attraction) using a three bar stimulus. This is readily seen in the vertical/

horizontal difference maps (Row C) that contain no structure. Observer GV displayed a difference between the vertical and horizontal target but in a direction opposite to the collinearity prediction. The control experiment (Row D) where observers did not experience a tilt illusion failed to produce any structure in the hotspots maps.

To aid visualization of the hotspots maps, we also show significant pixels maps (not corrected for multiple comparisons), calculated relative to the control condition. For each observer, we calculated the mean and standard deviation of the hotspots maps for their control condition and pixel values in the horizontal,
vertical, control, and horizontal–vertical conditions that were significantly different ($p = 0.001$) were set to white.

Hotspots maps summed across three observers (bottom row) show uniform structure around the target. In order to examine whether the structure around the target was isotropic, we summed pixel values within small subregions of the hotspots maps. If the maps are isotropic, the summed pixel values should be similar across wedge positions. Two different wedge radii are used, overlaid on the summed hotspots maps using a 30° bandwidth wedge. For the two sizes tested there is no difference in summed pixel values within the subregions for the summed horizontal or vertical maps; the plots are circular (circular variances ranged from 0.84 to 0.92).

The finding that the hotspots maps reveal some differences between observers is consistent with earlier reports of interobserver variability in classification images (Gold, Murray, Bennett, & Sekuler, 2000; Rajashekar, Bovik, & Cormak, 2006). To determine how much variation there was between observers we performed the following analysis: We randomly selected half of an observer’s data (10 runs) to build a subset map and used the other half (10 runs) to build a second subset map. We then cross correlated the first subset map with the second one for the same observer and with the subset maps of different observers. When subset maps were tested within the same observer, the average correlation (across conditions and observers) was $0.36 \pm 0.12$; when they were tested across observers, the average correlation was reduced to $0.24 \pm 0.08$. It appears that the observer’s decision process throughout the trials was fairly stable and that there is some overlap between different observers’ hotspots maps. This reflects the fact that all observers showed maximum influence for areas immediately adjacent to the target.

The lack of structure in the control condition hotspots maps when observers never experienced a tilt illusion (Row D) indicates that they are not simply reporting the orientation of parts of the surround during the orientation task. When mask size was varied, the larger mask led to a coarser sampling of the surround space, but the hotspots maps were defined by structure in similar sections (IM, Rows 1 & 4). To examine in detail how the ratio of mask to target sizes may alter the hotspots maps, observer IM performed the experiment for horizontal and vertical targets using two additional target and mask sizes (Figure 3).

This observer displays similar structure in the hotspots maps located mainly along the upper section of the surround across a range of target/mask sizes and orientations. When the masks are very large, the coarse sampling of the surround limits the effectiveness of the technique. Similarly for very small mask sizes, reducing the windows on the surround will start to introduce noise at a fine spatial scale. The significant pixels maps here were produced by calculating the mean and standard deviations of the distribution of pixel values for each hotspot map. Pixels greater than $\pm 2$ standard deviations from the mean were set to white, while those that did not exceed this criterion were set to black.

Figure 4A (left panels) shows the hotspots maps using White’s illusion, a phenomenon where identical grey patches placed on the black and white bars of a square wave grating appear to differ in lightness (White, 1979; see Figure 1e). Masks that led the observer to respond “left lighter” are summed and added to the polarity inverted masks that led the observer to respond “right lighter.” The corresponding significant pixels for the summed masks are shown on
the right. White’s illusion is reported to be due to assimilation from the flanks or contrast from the ends. For all observers, the hotspots maps reveal structure localized along the flanks of the stimulus that could arise due to the elongated nature of the stimulus. In order to examine this we created a series of combinatorial predictive hotspots maps that reflected a weighted sum of side flank assimilation and end flank contrast. This was done by making two predictive maps, one for side flank assimilation only and one for end contrast only by convolving the estimated point spread function of the filter used to create the mask with one of two configurations: assimilation only along the sides (a) and contrast only along the ends (c), resulting in predictive hotspots map for assimilation only (d) and contrast only (f). The relative contributions of the assimilation and contrast maps were varied to construct a set of combinatorial predictive maps (equal contribution shown in [e]). (C) Combinatorial predictive maps were correlated with each observer’s hotspots maps data to determine which ratio of assimilation to contrast produced the greatest correlation, zero is pure contrast and one is pure assimilation (g). For all observers, the addition of contrast influences did not increase the correlations obtained from assimilation only.

As a test of the versatility of our technique, we applied it to direction repulsion (Marshak & Sekuler, 1979) often referred to as the motion analogue of the tilt illusion. Figure 5 plots hotspots maps using these stimuli that clearly display interobserver variability. As with the tilt illusion, there were localized areas in the surround that were dominating observers’ experience of direction repulsion, but these were not associated with any particular section of the stimulus. There also does not appear to be any obvious correlation between the areas driving perception for orientation and those underlying the motion repulsion effects in the same observers (compare data for IM and RM in Figures 2 & 5).

**Discussion**

Using our stimuli we find no evidence of any specific location within the surround that drives contextual effects in motion or orientation, but rather that all areas closest to the target influence perception. There was some variability between observers as to which sections most dominated perception; for example, in the tilt illusion experiment, observer IM appeared to be
mainly influenced by structure along the top of the target whereas this section appeared to be irrelevant for RM. We suggest that the use of stimuli tiling the full surround comprehensively probes the complex interactions driving our perception of a target by giving equal stimulation to all areas. Our technique relies on the premise that if the strength of the surround’s influence were spatially uniform, the two opposite stimuli activating it would null each other, and the observers’ responses would be random; no structure would emerge from the hotspots technique. If the surround’s strength is not spatially uniform, the observer’s perception of the target will be correlated with the statistics of the stimulus over the hotspot(s). Our result could be reconciled with the findings that noncollinear structure may null or modulate the collinear effects due to low-level interactions occurring along a different axis of the stimulus (Dakin & Baruch, 2009; Solomon & Morgan, 2000). It is likely that the use of simplified configurations in earlier experiments may have boosted the influence of parts of the surround that would otherwise be subject to inhibitory interactions.

Our method is akin to the classification image technique that is used to examine how observers make decisions about a visual task. The classification image method usually involves presenting a stimulus in white noise to determine how each pixel influenced performance (e.g., Abbey & Eckstein, 2002; Gold et al., 2000; Mareschal, Dakin, & Bex, 2006; Murray, 2011; Neri, Parker, & Blakemore, 1999; Solomon, 2002). A different method based on the rapid presentation of stimuli that act to temporally mask each other has also been successfully used to examine observers’ sensitivity to orientation (Mareschal & Clifford, 2012; Ringach, 1997; Wong, Roeber, & Freeman, 2010) and motion (Iyer, Freeman, McDonald, & Clifford, 2011). In our procedure, no noise (either spatial or temporal) is added onto the stimulus; however, the underlying idea is the same: By using masks that differ on every trial in how they reveal sections of two opposite stimuli, we can uncover which areas of the surround are most influential to performance.

The individual hotspots maps reveal an anisotropic layout of effects for orientation and motion. We propose that an observer’s perceptual decision about a target is determined by the interplay of many influences. This includes the neural architecture of interactions but also other factors, such as, for example, how the observer distributes their attention in space or what part of the target they fixate that can act to modulate the strength of connections. The implication from these results is that the processing of very basic features, such as bars and gratings, is not solely a low-level process determined by detectors in early cortical areas but rather is subject to flexible higher level input. This is consistent with a recent finding of higher-level feedback altering low-level detector characteristics (Neri, 2011).

Our technique is highly versatile; we have successfully applied it to map out contextual interactions in three realms of vision, demonstrating a lack of collinear superiority for motion and orientation. In the case of the lightness stimulus we examined the relative roles of side assimilation and end flank contrast in White’s illusion, with our data highlighting the dominance of assimilative effects. It is important to note, however,
that this structure arises largely due to the elongated nature of the stimulus. Comparison of the predictive hotspots maps driven solely by assimilation (Figure 4d) or by equal contribution of assimilation and contrast (Figure 4e) are very similar. Our data support previous reports that the illusion is dependent on both the geometrical configuration of the target as well as the luminance of the inducers (Spehar, Clifford, & Agostini, 2002).

Finally, these results have important implications for early vision models that feature excitatory and inhibitory connections between neurons based on their relative alignments, often used as a first step in extracting edges or contours (e.g., Field et al., 1993; Huang, Jiao, & Jia, 2008; Itti & Koch, 2001; Li, 2000). Our results reveal the need for modulatory factors, either acting directly on the low-level interactions or at a later stage, that would capture the observer variability by boosting the influence of sections of the stimulus while suppressing others.

**Keywords:** contextual effects, reverse correlation, visual illusions

**Acknowledgments**

This research was funded by the Australian Centre of Excellence in Vision Science. CC was supported by a Future Fellowship from the Australian Research Council.

Commercial relationships: none.

Corresponding author: Isabelle Mareschal.

Email: imareschal@gmail.com.

Address: School of Biological and Chemical Sciences, Psychology, Queen Mary University of London, London, UK.

**References**


