Spatial limitations of fast temporal segmentation are best modeled by V1 receptive fields

Patrick T. Goodbourn  
Melbourne School of Psychological Sciences,  
University of Melbourne, Victoria, Australia  
School of Psychology, University of Sydney,  
New South Wales, Australia  

Jason D. Forte  
Melbourne School of Psychological Sciences,  
University of Melbourne, Victoria, Australia

The fine temporal structure of events influences the spatial grouping and segmentation of visual-scene elements. Although adjacent regions flickering asynchronously at high temporal frequencies appear identical, the visual system signals a boundary between them. These “phantom contours” disappear when the gap between regions exceeds a critical value ($g_{\text{max}}$). We used $g_{\text{max}}$ as an index of neuronal receptive-field size to compare with known receptive-field data from along the visual pathway and thus infer the location of the mechanism responsible for fast temporal segmentation. Observers viewed a circular stimulus reversing in luminance contrast at 20 Hz for 500 ms. A gap of constant retinal eccentricity segmented each stimulus quadrant; on each trial, participants identified a target quadrant containing counterphasing inner and outer segments. Through varying the gap width, $g_{\text{max}}$ was determined at a range of retinal eccentricities. We found that $g_{\text{max}}$ increased from 0.3° to 0.8° for eccentricities from 2° to 12°. These values correspond to receptive-field diameters of neurons in primary visual cortex that have been reported in single-cell and fMRI studies and are consistent with the spatial limitations of motion detection. In a further experiment, we found that modulation sensitivity depended critically on the length of the contour and could be predicted by a simple model of spatial summation in early cortical neurons. The results suggest that temporal segmentation is achieved by neurons at the earliest cortical stages of visual processing, most likely in primary visual cortex.

**Introduction**

The human visual system exploits a variety of cues to perform rapid grouping and segregation of scene elements. While one might usually think of static cues, such as differences in color or luminance across space, the temporal structure of visual input can also contribute to visual organization. In a natural scene, all of the elements comprising a moving object typically move together. Accordingly, simultaneous changes in visual features are likely to promote the grouping of scene elements on the basis that they are likely to be part of the same object, while asynchronous changes can induce segmentation.

Rapid luminance modulation can provide a strong cue for grouping and segmentation, even when elements are changing too quickly to discern individual light and dark phases (Forte, Hogben, & Ross, 1999; Goren & Flanagan, 2008; Livingstone & Hubel, 1987; Parton, Donner, Donnelly, & Usher, 2006; Quaid & Flanagan, 2005; Ramachandran, 1992; Ramachandran & Rogers-Ramachandran, 1991; Rogers-Ramachandran & Ramachandran, 1998; Sekuler & Bennett, 2001; Usher & Donnelly, 1998). Initially demonstrated by Livingstone and Hubel (1987), the effect was further investigated by Ramachandran and Rogers-Ramachandran (1991), who devised a stimulus containing two fields of spots modulating in luminance from black to white in counterphase. When modulation exceeded 7 Hz, the rate of flicker made it impossible to compare the temporal phase of any two spots. Nevertheless, a clear boundary was apparent between the regions. Because the stimulus induced the percept of a contour or texture border within an otherwise undifferentiated field, the authors referred to the phenomenon as a phantom contour.

Rogers-Ramachandran and Ramachandran (1998) noted that spatial separation of the modulating regions prevented the appearance of phantom contours. Subsequent studies have also supported the notion that
fast temporal-phase discrimination is a spatially limited process. Forte et al. (1999) demonstrated that the perception of phantom contours is solely dependent upon modulation within a narrow spatial window. They found that contrast sensitivity for a phantom-contour stimulus improved when the number of rows of spots on each side of the contour increased from one to two but did not improve further with the addition of more rows. Furthermore, sensitivity was notably impaired when the counterphasings regions were separated by 0.4° and was eliminated altogether at separations of 0.6° and greater. Similarly, Victor and Conte (2002) found that fast temporal-phase discrimination failed when centrally presented elements were separated by 0.125° or more. These results indicate that the mechanism underlying fast segmentation is sensitive only to modulation within a critical window surrounding the counterphasings border.

In contrast, Goren and Flanagan (2008) have suggested that the perception of form defined by phantom contours is primarily mediated by the surface properties of a stimulus. They asked participants to detect annular targets defined by spots flickering in counterphase to the background. They found no effect on contrast sensitivity of varying the total length of contour while the area within the figure was held constant. However, they reported a small improvement in sensitivity when the figure area was increased while the contour length was held constant (see also Quaid & Flanagan, 2005). These findings led the authors to conclude that phantom-contour perception is more dependent on the surface area of a stimulus than on the contour itself.

Here, we show that sensitivity to counterphasings modulation depends critically on two properties of the contour: the separation of the modulating regions and the length of the contour. Furthermore, we show that the effects of these properties on sensitivity depend on retinal eccentricity in a manner consistent with a detection mechanism situated early in the visual-processing hierarchy. We also discuss how the results reported by Goren and Flanagan (2008) are consistent with a receptive-field model of fast temporal-phase segmentation, whereby a border between modulating regions is detected within a limited spatial window.

Rationale

Forte et al. (1999) noted that the spatial limitations of fast temporal segmentation were consistent with a neural mechanism that sums modulation energy within a limited receptive field (RF). Consider two adjacent, flickering elements in a visual display. For the response of a neuron to depend on whether the elements flicker in phase or in counterphase, that neuron must be capable of detecting modulation on both sides of the border between elements. Accordingly, fast temporal segmentation should only be possible when the separation between counterphasings elements is smaller than the RF size of neurons capable of signaling the modulation. This hypothesis accounts for the dramatic effect of element separation on temporal-phase sensitivity: Beyond a given separation, no neuron has an RF sufficiently large to encompass both elements. It also accounts for the lack of effect of additional modulation at a distance from the counterphasings border: The response of any neuron will only be affected by modulation within the spatial limits of its RF.

According to this hypothesis, the size of the critical gap at which segmentation fails could indicate the location within the visual pathway of the neurons that mediate the signal. Single-cell electrophysiological studies in macaques have revealed that RFs centered at a given spatial location typically increase in size at each successive stage of the visual pathway. Further, within a given stage or brain region, the mean RF size is typically smallest at the fovea and increases approximately linearly with retinal eccentricity (Burkhalter & Van Essen, 1986; Croner & Kaplan, 1995; Derrington & Lennie, 1984; Dow, Snyder, Vautin, & Bauer, 1981; Felleman & Van Essen, 1987;Gattass, Gross, & Sandell, 1981; Gattass, Sousa, & Gross, 1988;Hubel & Wiesel, 1974; Nakamura & Colby, 2000). Forte et al. (1999) reasoned that the critical gap would most likely reflect the extent of the largest RFs responding to the modulation, those situated at the most eccentric point of the counterphasings border (around 5° in their stimulus). However, while their 0.4° gap is consistent with RF sizes in V1 at 5° eccentricity (Dow et al., 1981; Hubel & Wiesel, 1974), retinal eccentricity was neither controlled nor systematically varied in their experiments. Consequently, the influence of retinal eccentricity on the spatial parameters of fast temporal segmentation remains unclear.

The current study thus aimed to investigate the relationship between retinal eccentricity and the spatial limitations of fast temporal segmentation. Experiments 1 and 2 measured, at a range of eccentricities, the maximum separation permitting the detection of fast counterphasings modulation. These data were compared to RF size measurements reported in single-cell physiological studies to infer the location in the visual system of the underlying mechanism. A close match with the RF profile of a given brain area would provide strong evidence for the hypothesis that temporal segmentation is mediated by a neuronal mechanism located in that brain region. Conversely, a lack of variation with eccentricity in the spatial parameters of temporal segmentation, or variation inconsistent with RF profiles, would suggest that temporal segmentation and perception of phantom contours are mediated by
mechanisms that are not limited by early visual neurons. Our results indicate that the spatial limitations of fast temporal segmentation are consistent with RF sizes in V1.

Experiment 3 measured, at a range of eccentricities, sensitivity to a counterphasing edge as a function of edge length. If temporal segmentation is achieved by early visual neurons, we would expect sensitivity at first to increase rapidly with length, owing to physiological summation within individual RFs, then to increase more modestly with longer edges, owing to probability summation between spatially distributed independent units (Anderson & Burr, 1987, 1991; Graham, 1989; Howell & Hess, 1978; Robson & Graham, 1981). Our results followed this pattern, and by fitting a simple, physiologically plausible model to the data, we obtained further, independent RF size estimates consistent with single-cell measurements in V1. Together, the findings of the three experiments suggest that temporal segmentation is achieved by neurons at the earliest stages of cortical visual processing.

General methods

Participants

Observers were six males and two females aged from 24 to 38 years, including the authors and six colleagues who were unaware of the particular aims of the study. Five observers (AM, JF, JP, OC, and PG) participated in Experiments 1 and 2, and four observers (JH, LJ, PG, and XV) in Experiment 3. All had normal visual acuity and were experienced in psychophysical observation.

Apparatus

Stimuli were generated on a Power Mac G4 computer (Apple Inc., Cupertino, CA) using Matlab software (MathWorks, Natick, MA) with PsychToolbox–2 routines (Brainard, 1997; Pelli, 1997) and displayed via a BITSS++ Digital Video Processor (Cambridge Research Systems, Rochester, UK) on a Trinitron Multiscan G520 monitor (Sony Corporation, Tokyo, Japan) operating at a spatial resolution of 1024 × 768 pixels and a refresh rate of 120 Hz. The screen of the monitor subtended 56.0° × 43.7°. The luminance behavior of the monitor was linearized for each of the three phosphors independently using a CS-100A colorimeter (Konica Minolta Inc., Tokyo, Japan) and gamma corrections to the lookup table. Across the full luminance range, gamma correction was effective in achieving a linear relationship between requested and displayed luminance values ($r > 0.99$).

Psychophysical detection thresholds can be affected by luminance edges close to the target (McCann & Hall, 1980; Novak & Sperling, 1963). Thus to maximize the effective area of the uniform background, a 660 × 590 mm (83.4° × 77.1°) piece of white card was placed immediately in front of the monitor and illuminated to match the average spatiotemporal luminance of the display ($L_0 = 55.0$ cd/m²). Stimuli were viewed through a centered circular aperture in the card subtending 40.6°. Observers used a chin rest to maintain a viewing distance of 380 mm from the face of the monitor.

Stimuli

Figure 1 depicts a single frame of a stimulus used in the experiments. It consisted of a luminance-modulated segmented circle, subtending 38.4°, on a static gray surround of luminance $L_0$. The circle was partitioned into 70° temporal (T), superior (S), nasal (N), and inferior (I) quadrants separated by a 20° sector gap of fixed luminance $L_0$. Each quadrant was further divided...
The luminance of all other segments was given by

\[ L_{\text{other}} = 2L_0 - L_{\text{target}}. \]  

(3)

In previous studies, phantom-contour stimuli have most often been constructed using fields of spots (Goren & Flanagan, 2008; Quaid & Flanagan, 2005; Ramachandran & Rogers-Ramachandran, 1991; Rogers-Ramachandran & Ramachandran, 1998; Sperling, Lu, Manis, & Seidenberg, 2003) or Gaussian blobs (Forte et al., 1999). However, for our purposes, solid segments were preferable because they allowed us to make much finer manipulations of the separation between regions. Forte et al. (1999) demonstrated that Gaussian blobs and solid elements produce comparable results.

Our experiments measured either threshold contrast (Experiments 1 and 3) or the proportion of correct responses at near-threshold contrast (Experiment 2). As in the present study, most previous studies of phantom contours have used near-threshold stimuli. We chose to present our stimuli at low contrasts primarily because simple linear models are most appropriate for predicting near-threshold performance (Graham, 2011). Nonetheless, the separation between regions that is required to prevent fast temporal segmentation appears to be similar for high-contrast stimuli (less than 0.7° in Ramachandran & Rogers-Ramachandran, 1991) and low-contrast stimuli (about 0.4° in Forte et al., 1999).

**Experiment 1**

The present study aimed to compare the spatial limitations of temporal segmentation at different locations across the visual field. However, sensitivity to counterphasing edges may vary with eccentricity (Goren & Flanagan, 2008; Quaid & Flanagan, 2005; Rogers-Ramachandran & Ramachandran, 1998) and with visual quadrant. Accordingly, the aim of Experiment 1 was to determine, individually for each observer, sensitivity to modulating edges at different visual-field locations. The data enabled the salience of stimuli to be equated across the visual field in Experiment 2. This ensured that the observed limitations of temporal segmentation accurately reflected the spatial range of underlying mechanisms and were not confounded by retinotopic variations in absolute sensitivity.

**Methods**

We determined modulation sensitivity by measuring the minimum contrast required for the discrimination...
of counterphasing modulation at six eccentricities \((E = 2^\circ, 4^\circ, 6^\circ, 8^\circ, 10^\circ, \text{ and } 12^\circ)\) in each of the temporal (T), superior (S), nasal (N), and inferior (I) visual-field quadrants, with no gap between inner and outer segments. For each trial, presentation contrast was set using a ZEST adaptive staircase (King-Smith, Grigsby, Vingrys, Benes, & Supowit, 1994; Watson & Pelli, 1983) with a 72% correct threshold criterion. On completion of a staircase, contrast threshold was calculated as the mean of the posterior probability density function (pdf).

Each block comprised 24 interleaved staircases (4 visual-field quadrants \(\times\) 6 eccentricities). Observers completed six blocks, each lasting approximately 12 to 14 min. A preliminary calibration block containing 40 trials per staircase was used to obtain initial threshold estimates, with the mean and standard deviation of the prior pdf set to \(-1\) and 2 log units of contrast, respectively. The estimates were subsequently employed in five experimental blocks, each with 30 trials per staircase, by setting the mean of the prior pdf at the initial estimate and the standard deviation at 1.5 log units. A final threshold estimate for each observer in each of the 24 conditions was calculated as the mean of the thresholds derived in the five experimental blocks. Modulation sensitivity was defined as the inverse of threshold contrast.

**Results**

Data were analyzed separately for each observer. The results in Figure 2 show that with no gap between modulating regions, observers can detect phase differences at all eccentricities and in all quadrants tested. Sensitivity increased with eccentricity, reaching a plateau at \(8^\circ\) or \(10^\circ\) for most observers in the S, N, and I quadrants. Sensitivity in the T quadrant followed a similar pattern but declined markedly between \(10^\circ\) and \(12^\circ\), which corresponds to the typical location of the foveal border of the physiological blind spot (Østerberg, 1935). At each eccentricity, sensitivity also varied between visual quadrants, though patterns were inconsistent between observers.

**Discussion**

The results of the first experiment show that sensitivity to temporal-phase differences in our partic-
ular stimulus varies with eccentricity and visual-field location. The general increase in sensitivity with eccentricity may be due to the corresponding increase in the absolute length of the border between inner and outer segments, which may affect the number of cells responding to an edge. However, temporal contrast sensitivity functions are known to change substantially across the retina even when stimuli are scaled to account for variations in receptor density (Tyler, 1987). Thus our findings might be influenced by differences across the visual field in any number of factors, including underlying physiology, decision variables, or attentional resources. In any case, the results of Experiment 1 allowed us to adjust the contrast of stimuli in order to standardize baseline performance across all conditions in Experiment 2.

### Experiment 2

Experiment 1 showed that sensitivity to a counterphasing edge varied according to the retinal location at which it was presented. The aim of Experiment 2 was to determine the gap between segments at which sensitivity to their counterphasing luminance modulation fell to half of that observed with no gap, an index we call $g_{\text{max}}$. This was measured at a range of retinal eccentricities while controlling for the variation in sensitivity across the visual field that was assessed in Experiment 1.

### Methods

The method of constant stimuli was used to determine, at each eccentricity, the width of the gap between counterphasing segments at which sensitivity fell to half of that observed with no gap, an index we call $g_{\text{max}}$. The method of constant stimuli was used to determine the gap between segments at which sensitivity to their counterphasing luminance modulation fell to half of that observed with no gap, an index we call $g_{\text{max}}$. This was measured at a range of retinal eccentricities while controlling for the variation in sensitivity across the visual field that was assessed in Experiment 1.

To ensure comparable baseline performance in each condition, stimuli were presented for each observer at a contrast derived from the threshold obtained for the corresponding visual-field location in Experiment 1. Contrast thresholds for intermediate eccentricities ($E = 2^\circ$ to $12^\circ$ in increments of $1^\circ$), 11 gap widths scaled with eccentricity ($g = 0^\circ$ to $0.2E^\circ$ in increments of $0.02E^\circ$), and four target quadrants (I, S, T, and N).

To ensure comparable baseline performance in each condition, stimuli were presented for each observer at a contrast derived from the threshold obtained for the corresponding visual-field location in Experiment 1. Contrast thresholds for intermediate eccentricities ($E = 2^\circ$ to $12^\circ$ in increments of $1^\circ$), 11 gap widths scaled with eccentricity ($g = 0^\circ$ to $0.2E^\circ$ in increments of $0.02E^\circ$), and four target quadrants (I, S, T, and N).

### Results

We initially analyzed data separately for each observer. Data were collapsed across visual-field locations, resulting in 100 trials per gap width at each eccentricity. A measure of sensitivity ($d'$) was derived from the percentage of correct responses (Hacker & Ratcliff, 1979). At each eccentricity, we used a least-squares algorithm to model sensitivity as a function of gap width using the complementary Gaussian error function

$$d' = \text{erfc}(g) = x \left[ 1 - \text{erf} \left( \frac{g}{\beta} \right) \right],$$

where $x$ and $\beta$ are the amplitude and standard-deviation parameters, respectively, and $\text{erf}(x)$ is the Gaussian error function

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-t^2} dt.$$

There was a good fit between the curve and the data for each participant at each eccentricity, with $r^2$ values above 0.80 in all but one case (mean $r^2 = 0.95$, $SD = 0.04$). Figure 3 shows, for a single typical observer, sensitivity as a function of gap width at each eccentricity tested. Sensitivity declined with increasing gap width at all eccentricities. The slope of the function was shallower for lower eccentricities and steeper for higher eccentricities.

We next examined $g_{\text{max}}$ as a function of eccentricity. For each eccentricity, $g_{\text{max}}$ was calculated as the point at which the sensitivity curve fell to one-half of its maximum. The value was chosen because curves were relatively steep at this point; thus it was likely to be less volatile than lower values. As shown in Figure 4, values of $g_{\text{max}}$ increased approximately monotonically with eccentricity, ranging from about 0.3° to 0.8° for eccentricities from 2° to 12°. Results showed little variation between observers and contrast conditions. The values obtained are consistent—both in absolute...
terms and in terms of relative variation with eccentricity—with RF diameters of neurons derived from single-cell recordings in V1 of awake, behaving macaques (Dow et al., 1981; Hubel & Wiesel, 1974).

We then derived a linear function for $g_{\text{max}}$ as a function of retinal eccentricity. We pooled $g_{\text{max}}$ values across all observers and contrast conditions and fitted a regression line using a least-squares procedure. The data were well described ($r^2 = 0.56$) by the linear function

$$g_{\text{max}}(E) = 0.046E + 0.350,$$

where $E$ is retinal eccentricity in degrees. We then assessed, by visual inspection, the correspondence between the $g_{\text{max}}$ fit line and functions describing RF diameters of neurons at different stages of the visual hierarchy. Dow et al. (1981) provide a function representing variation with eccentricity in the mean RF diameter of macaque V1 cells; we also derived functions from published data to describe the typical diameter of cell RFs in other cortical and subcortical visual areas.

As can be seen in Figure 5, the $g_{\text{max}}$ data and regression line correspond closely to the V1 function.
By comparison, they are inconsistent—both in slope and in magnitude—with the next closest functions describing RF diameters for lateral geniculate magnocell (LGN M-cell) centers and V2 cells. Thus our data are consistent with the proposal that $g_{\text{max}}$ is limited by RF sizes in V1. We also note that the minimum separation between the outer segments of adjacent quadrants in our stimulus was considerably greater than $g_{\text{max}}$ at all eccentricities tested. This suggests that sensitivity was mediated solely by the boundaries between inner and outer segments and was not influenced by the boundaries between quadrants.

**Discussion**

The results of Experiment 2 show that the spatial limitations of fast temporal segmentation are consistent with the spatial limitations of typical neurons in V1. This suggests that temporal-phase segmentation and the perception of phantom contours may be mediated by early cortical visual neurons. Indeed, it has already been suggested that spatiotemporal filtering by neurons in the early stages of visual processing can account for segmentation in other forms of temporally structured display. For example, Lee and Blake (1999) found that in a dynamic texture, a target region in which all elements reversed their motion in synchrony appeared distinct from a background in which reversals were uncorrelated between elements. In the apparent absence of cues from luminance, contrast, and coherent motion, this suggested that observers could rely on “temporal microstructure” to perform spatial grouping and segregation. Adelson and Farid (1999), however, showed that the temporal filtering properties of early vision would reveal static form in such displays, at least in their original form (see also Farid & Adelson, 2001).

If temporal segmentation in phantom-contour stimuli is also achieved by spatiotemporal filtering, how might a single cell extract the modulation energy falling within its RF? A counterphasing edge may be characterized in two dimensions: one dimension of space perpendicular to the edge ($x$), and one dimension of time ($t$). The spatiotemporal ($x-t$) profile comprises a step function in the $x$ dimension multiplied by a sinusoid in the $t$ dimension. The RF of an ideal detector...
will have a sensitivity profile that matches the stimulus (Watson, Barlow, & Robson, 1983). As noted by Forte et al. (1999), this ideal profile is well approximated by linear spatiotemporal separable RFs such as those initially mapped in cat striate cortex (DeAngelis, Ohzawa, & Freeman, 1993a, 1993b, 1995). Such RFs are also present in V1 of macaques: Pack, Conway, Born, and Livingstone (2006, figure 12a), for example, present an RF in which the spatial (x) profile reverses in polarity over approximately 25 ms. This interval is ideal for the detection of modulation at 20 Hz, the rate used in the current study.

It would thus appear that basic neural mechanisms found in primary visual cortex are suited to the detection of counterphasing luminance modulation. Given the results of Experiment 2, it is reasonable to suggest that these units might underlie fast temporal segmentation and the perception of phantom contours. If this is the case, it should be feasible to obtain independent psychophysical estimates of RF size using spatial-summation techniques (Anderson & Burr, 1987, 1991; Gorea, 1985; Howell & Hess, 1978). We did so in Experiment 3, providing further support for our hypothesis that temporal segmentation is mediated by an early cortical mechanism.

**Experiment 3**

The results of Experiments 1 and 2 indicate that neurons in V1 may underlie fast temporal segmentation. As we have noted, linear x–t separable RFs suited to the detection of counterphasing modulation have been mapped in cat and macaque V1. If temporal segmentation is mediated in the first instance by approximately linear and independent units, we can make predictions about the summation of modulation energy across space. As the length of a counterphasing edge is increased from zero, we initially expect spatial summation to proceed in a linear manner as a result of physiological summation (sometimes called synaptic or total summation) within the RF of individual units. This would be reflected psychophysically by a rapid increase in modulation sensitivity (or, conversely, a rapid decrease in contrast threshold) with increasing contour length. As contour length is increased beyond the spatial extent of single RFs, we expect probability summation across space. That is, the probability of detecting the contour increases with contour length, owing to an increase in the number of independent mechanisms capable of responding. This would be reflected psychophysically by a more gradual increase in contrast sensitivity (or a gradual decrease in contrast threshold) with increasing contour length.

In Experiment 3, we varied the length of the modulating border and measured the minimum contrast required to discriminate counterphasing modulation at a range of eccentricities. Contrast thresholds that decrease as contour length increases—at first rapidly, and subsequently more gradually—would provide evidence for our proposal that fast temporal segmentation and phantom-contour perception are mediated by linear neurons early in the visual pathway. Conversely, contrast thresholds not conforming to this pattern would challenge the proposal. We then employed a simple computational model based on known physiology, which we fitted to the data to estimate the spatial extent of the RFs detecting counterphasing modulation. If estimated RF sizes increase with eccentricity in a manner consistent with early cortical visual areas, the proposal is further supported; conversely, RF size estimates that do not increase with eccentricity or that increase in a manner inconsistent with early visual areas would challenge the proposal. The results, both psychophysical and computational, support our hypothesis that rapid temporal segmentation is achieved by neurons in primary visual cortex.

**Methods**

We determined contrast thresholds for the discrimination of counterphasing modulation at six retinal
eccentricities ($E = 2^\circ, 4^\circ, 6^\circ, 8^\circ, 10^\circ$, and $12^\circ$) for 12 different lengths of contour scaled by eccentricity ($l = 0^\circ, 10^{-1.2}E^\circ, 10^{-1.35}E^\circ, 10^{-1.2}E^\circ, 10^{-1.05}E^\circ, 10^{-0.9}E^\circ, 10^{-0.75}E^\circ, 10^{-0.6}E^\circ, 10^{-0.45}E^\circ, 10^{-0.3}E^\circ, 10^{-0.15}E^\circ$, and $E^\circ$). Each stimulus quadrant comprised a $70^\circ$ sector (as in Experiments 1 and 2), with contour length manipulated by obscuring the ends of each quadrant border with two wide arcs of luminance $L_0$ (see Figure 6). This arrangement allowed contour length to vary while keeping the total amount of modulation in the stimulus relatively constant.

Contrast thresholds were measured using the same staircase procedure as in Experiment 1. Each block comprised 12 interleaved 30-trial staircases, one for each contour length, at a single retinal eccentricity. Observers first completed one preliminary calibration block, containing 40 trials per staircase, for each of the six eccentricities. The threshold estimates derived from these blocks were then employed in five experimental blocks per eccentricity (30 in total). Block order was counterbalanced using a Latin-square design. In all, each observer completed 2,880 calibration trials and 10,800 experimental trials. Final estimates of threshold for each of the 72 combinations of contour length and eccentricity were calculated as the mean of the thresholds derived in the five experimental blocks.

Model

While we give a full exposition of the model in the Appendix, we provide here a basic account of its operation. The key element is a single neuron, or unit, which responds to counterphasing modulation within a limited spatial window (RF). The unit has a spatiotemporal RF profile, or weighting function, that determines its response to a stimulus: The better the stimulus matches its weighting function, the greater its response. The spatial weighting function is that of a typical edge detector, with an excitatory region on one side and an inhibitory region on the other. The unit will give the greatest response to a luminance edge aligned such that the brighter region of the stimulus coincides with the excitatory region of the RF and the darker region of the stimulus coincides with the inhibitory region of the RF. The spatial weighting function reverses in polarity over 25 ms; thus, the greatest response will be to a luminance edge that reverses in polarity at the same rate.

The model mimics the structure of the visual system by employing many of these units, with their RF centers forming a grid across the visual field. The RF of any given unit will cover some part of the stimulus. The probability of a single unit detecting the target contour is a simple function of three factors: the contrast of the stimulus, the absolute sensitivity of the unit, and the match between the unit’s weighting function and the part of the stimulus that falls within its RF. The probability that the observer detects the target contour is the probability that at least one unit detects the contour; and the probability that the observer gives a correct response is the probability that the observer detects the target, with an adjustment for guessing. The observer’s contrast threshold is the stimulus contrast at which the probability of a correct response is equal to the threshold criterion.

The model is applied independently to the data from each observer at each eccentricity. For simplicity, we assume that at a given eccentricity, every unit has the same RF size, the same absolute sensitivity, and the same spatiotemporal weighting function. Only two parameters are allowed to vary: The first controls the size of the RF ($\sigma$), and the second represents absolute sensitivity of a unit ($A$). The fitting procedure finds the values of $\sigma$ and $A$ that best account for the data describing contrast threshold as a function of contour length. Because stimuli were presented with a fixed duration (500 ms), we do not here attempt to estimate

![Figure 6. Single frame of the stimulus used in Experiment 3. The stimulus was divided into four quadrants, each containing an inner and outer segment abutting at a boundary of constant retinal eccentricity. On each trial, the contrast of the stimulus reversed at 20 Hz for 500 ms, with the outer segment of a target quadrant modulated 180° out of temporal phase with other elements. The length of the contour ($l$) varied between trials, and the retinal eccentricity of the contour ($E$) varied between blocks. The width of the masking arcs ($s$) at each eccentricity was sufficient to prevent temporal-phase discrimination across the gap.](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933542/ on 11/09/2018)
the temporal extent of the weighting function. Instead, we assume that each unit integrates across the full presentation time. This assumption, however, only impacts the absolute-sensitivity parameter of the model, not the size parameter that is the focus of the study. The same applies to our choice of stimulus duration: Changing the presentation time could affect estimates of absolute sensitivity, but not of size, for the units in the model.

**Results**

We initially analyzed the data separately for each observer. The results presented in Figure 7 show that as contour length increased, contrast thresholds decreased, at first rapidly, and subsequently more gradually. This was true for all observers at each eccentricity tested. In every case, discrimination of counterphasing modulation where $l = 0^\circ$ was impossible, demonstrating that the width of the masking arcs in the stimulus was sufficiently large to prevent discrimination of phase across the gap. (Because contrast threshold is undefined, these data points have been omitted from Figure 7). Absolute thresholds decreased with eccentricity, although this may be a result of scaling the contour lengths tested.

Next, we developed a simple, physiologically plausible computational model of the manner in which independent units (corresponding to $x-t$ separable simple cells) might mediate fast temporal segmentation (see the Appendix for details). We fitted the model to the data describing contrast threshold as a function of contour length to derive estimates of RF size for each observer at each eccentricity. As can be seen in Figure 8, RF size estimates derived from the model increase approximately monotonically with eccentricity. Pooling across all observers, the linear function best fitting the data ($r^2 = 0.59$) is

$$\text{size}_{R}\text{F}(E) = 0.088E + 0.159,$$

(7)

where $E$ is the retinal eccentricity in degrees. Visual comparison of our data to functions describing RF diameters of neurons in subcortical and cortical visual areas shows that they are closest to the V1 function at all eccentricities. As in Experiment 2, these data are consistent with the notion that temporal segmentation is performed by neurons in V1.
Discussion

Experiment 3 aimed to provide independent estimates of RF size for neurons mediating fast temporal segmentation using a spatial-summation technique. The results point to a V1 locus, supporting the findings of Experiment 2. The estimates of RF size derived from Experiment 3 were generally larger than those from Experiment 2; this is in line with reports that RF length typically exceeds RF width in simple cells (Dow et al., 1981; Jones & Palmer, 1987; Ringach, 2002).

As contour length increased, contrast thresholds declined rapidly for very short contours, then declined much more gradually. This may explain why Goren and Flanagan (2008) failed to observe any effect of contour length on sensitivity: The smallest contour tested in their experiments was 29.3°, well beyond the length at which sensitivity begins to plateau at any eccentricity.

General discussion

This study investigated the spatial limitations of temporal segmentation at high frequencies of luminance modulation, with the aim of identifying the level in the visual hierarchy at which the neural signal for phantom contours originates. With increasing spatial separation, sensitivity to counterphasing regions was reduced; this effect was more gradual when the boundary between regions was farther from the fovea. A threshold index ($g_{\text{max}}$), defined as the gap width at which sensitivity fell to half of that observed with no gap between regions, increased monotonically with eccentricity. The values of $g_{\text{max}}$ corresponded closely to typical receptive-field diameters of macaque V1 neurons reported in the literature. This finding is consistent with the proposal that temporal-phase segmentation is moderated by neurons that detect modulation energy within their receptive field. According to this model, $g_{\text{max}}$ indicates the separation at which the counterphasing regions fall beyond the spatial range of neurons capable of signaling the modulation.

In a further experiment, we found an increase in sensitivity with contour length. The pattern was suggestive of physiological summation within, and probability summation between, neurons sensitive to contrast modulation at the counterphasing edge. A simple, physiologically plausible model fitted to the data predicted RF sizes close to those measured in V1 of macaques. The correspondence between the present data and single-cell measurements in V1 thus suggests that these neurons play a critical and limiting role in temporal segmentation and the perception of phantom contours.

Comparison with previous findings

Previous psychophysical studies have found fast temporal segmentation to be spatially limited (Forte et al., 1999; Ramachandran, 1992; Ramachandran & Rogers-Ramachandran, 1991; Rogers-Ramachandran & Ramachandran, 1998; Victor & Conte, 2002). The values of $g_{\text{max}}$ reported here accord with the spatial limitations reported by the two groups that have explicitly investigated counterphasing luminance modulation. Although they did not systematically vary the separation between regions, Rogers-Ramachandran and Ramachandran (1998) observed that a gap of 0.75° prevented the appearance of phantom contours. Their stimulus was presented centrally and subtended 13.4° × 15.4°; assuming a neuronal segmentation mechanism, the largest receptive fields responding to the edge would be those nearest the periphery of the stimulus at eccentricities of 6.7° and 7.7°. From the function derived in the present study, the corresponding values of $g_{\text{max}}$ are 0.66° and 0.70°, respectively. While the separation of 0.75° is beyond $g_{\text{max}}$, some residual sensitivity might nevertheless be expected. However, their display comprised modulating fields of spots, to which observers are somewhat less sensitive than to the solid-element stimuli used in our study (Forte et al., 1999).
Our findings are also consistent with the spatial limitations reported by Forte et al. (1999), who parametrically varied the separation of counterphasing solid quadrants and found that fast temporal segmentation failed at gaps greater than 0.4°. Their stimulus subtended 10.5° × 10.5°, with the outermost edges located at an eccentricity of 5.25°. The critical value of 0.4° is smaller than our corresponding \( g_{\text{max}} \) of 0.59°, though it is close to the range of \( g_{\text{max}} \) values for individual observers in our study (from 0.46° to 0.71° at 5° eccentricity). For the Forte et al. stimulus, typical RF sizes varied across the length of the contour. As the gap between quadrants approaches a critical width, sensitivity is mediated only by those neurons with the largest RFs, which are situated near the outermost edge of the contour. This decreases the effective length of the contour and thus reduces sensitivity. In comparison, for the stimulus used in the current study, typical RF size is constant across the length of the contour. Thus even as the gap between segments approaches a critical width, neurons with RFs situated along the full length of the contour will continue to mediate sensitivity. This difference could easily account for the small discrepancy between their findings and those of the present study.

Goren and Flanagan (2008) proposed that surface properties were more important than the contour itself in determining sensitivity to phantom contours. Their proposal was made on the basis of three experiments, which required the detection of annular targets flickered in counterphase to the background. In the first, no effect of ring thickness was found for annuli with a fixed outer diameter. The total amount of contour (outer plus inner) in these stimuli was always greater than 31°. According to our results and model, no readily observable effect would be predicted within this range.

In the second experiment, no effect of contour length was found for rings of a constant area. As in the first experiment, the contour lengths tested (all greater than 29°) were within the plateau region of the sensitivity curve measured in our Experiment 3. Thus, while it is true that the effect of contour length is negligible for long contours, our results show that the effect is striking for smaller contours.

In their third experiment, Goren and Flanagan (2008) found that for annuli of constant contour length, sensitivity increased with the area of the target. More specifically, sensitivity was lower for the stimulus with the smallest area than for stimuli with larger areas, although this effect was absent when stimuli were presented at fixation. The thickness of the annulus with the smallest area was 0.4°, with the next largest being 0.8°. Thus, the results are compatible with those of Forte et al. (1999), who reported an increase in sensitivity when the number of rows of spots comprising a border was increased from one to two (corresponding to an increase from about 0.33° to 0.66°), but no further effect of extra rows. In the context of our model, this represents the point at which the additional area fails to affect sensitivity because it falls outside the RF of any unit responding to the counterphasing border.

It has been suggested previously that phantom contours selectively stimulate the magnocellular visual pathway (Ramachandran & Rogers-Ramachandran, 1991; Rogers-Ramachandran & Ramachandran, 1998; Sperling et al., 2003). The proposal is controversial and has been challenged on a number of grounds (see Skottun & Skoyles, 2006). Our results suggest that phantom-contour perception is mediated by cortical mechanisms. However, while we may note that phantom contours are unlikely to be retinal in origin—and thus are not likely to be generated by magnocells—the present data do not allow us to determine whether or not the relevant cortical cells receive predominantly magnocellular input.

### Relation to \( d_{\text{max}} \) and local-motion detection

Our suggestion that temporal-phase segmentation could be achieved by linear RFs creates links to a body of literature concerned with a “short-range” process for the detection of motion identified by Braddick (1974; Braddick, Ruddock, Morgan, & Marr, 1980). When a subregion in a random array of dots is spontaneously displaced, it appears to be moving and is perceptually segregated from the remainder of the display. Like temporal-phase segmentation, detection of the displacement seems to be a local, spatially limited process. Observers are able to report the direction of apparent motion only if the displacement is below a spatial threshold known as \( d_{\text{max}} \).

In many models of local-motion detection (see, e.g., Adelson & Bergen, 1985), motion information is extracted from a stimulus by a cascade of spatiotemporal detectors. Detectors at the first stage of processing correspond exactly to the spatiotemporal separable detectors that we have proposed may be responsible for the detection of phantom contours. Temporal segmentation may thus be mediated by mechanisms implicated in local-motion processing, a possibility raised by Fahle (1993).

If these two mechanisms—detection of short-range motion and temporal-phase segmentation—have a common physiological substrate, one should expect the spatial limitations of local-motion detection (\( d_{\text{max}} \)) to exhibit a similar scaling with eccentricity as the spatial limitations of temporal-phase segmentation (\( g_{\text{max}} \)). As local-motion detectors receive inputs from a number of spatiotemporal separable subunits, their combination is
likely to yield an aggregate receptive-field size somewhat greater than that of the individual subunits. Accordingly, at any eccentricity, we expect \( d_{\text{max}} \) to exceed \( g_{\text{max}} \) but remain within the range of RF sizes associated with early cortical neurons. Indeed, this appears to be the case. Baker and Braddick (1985) measured critical displacement for short-range motion at retinal eccentricities from 0.4° to 10° and found \( d_{\text{max}} \) to scale linearly from 0.1° to 1.5° over this range. These values are similar to \( g_{\text{max}} \) at lower retinal eccentricities and up to 0.7° greater than \( g_{\text{max}} \) in the periphery. The difference in gradient may be due to the recruitment of more subunits or a greater degree of RF center scatter in the peripheral field compared to the central field.

Comparison of human and macaque RF properties

The current study used measurements of RF diameters in macaque LGN and visual cortex to infer the location of a neural mechanism for fast temporal segmentation. While electrophysiological investigations of receptive fields in human visual cortex are rare, experimental studies involving human patients who have had electrodes implanted for clinical purposes indicate a close correspondence between human and macaque visual physiology (Marg, Adams, & Rutkin, 1968; Wilson, Babb, Halgren, & Crandall, 1983; Yoshor, Bosking, Ghose, & Maunsell, 2007). Macaque single-cell data are also consistent with human receptive-field size estimates obtained from functional magnetic resonance imaging (fMRI; Dumoulin & Wandell, 2008; Kastner et al., 2001; Smith, Singh, Williams & Greenlee, 2001; Tootell et al., 1997). As the resolution of fMRI improves, increasingly refined estimates of human visual receptive-field sizes will become available. In fact, we believe that our stimulus could be useful for probing RF response properties in such studies. In any case, although there is good evidence of comparable RF sizes between human and macaque, the general interpretation of our results does not depend on exact correspondence: Fast temporal-phase segmentation operates within a restricted spatial window, the limits of which scale with eccentricity—and this is best explained by a neural mechanism limited by the spatial extent of its receptive field.

Conclusions

We have suggested that the spatial limits of fast temporal segmentation correspond to the spatial limits of receptive fields in early visual processing, and have proposed a physiologically plausible neural mechanism that accounts for our experimental data. This mechanism is also implicated in models of local-motion detection, a process that exhibits similar spatial limitations. Together, our psychophysical and computational results support our conclusion that fast temporal-phase segmentation is mediated by neurons in primary visual cortex.

Keywords: temporal segmentation, temporal-phase discrimination, phantom contours, flicker-defined form, receptive fields

Acknowledgments

We thank Philip Smith and Simon Cropper for helpful comments on an earlier version of the manuscript. This research was supported in part by an Australian Postgraduate Award to PTG. During preparation of the manuscript, PTG was supported by a grant from the John Templeton Foundation held by H. Price, A. Holcombe, K. Miller, and D. Rickles.

References


### Appendix

This section describes the model used to estimate RF size from the data of Experiment 3. It follows Anderson and Burr’s (1987, 1991) model of spatial summation in directionally selective mechanisms. The spatiotemporal profile for our model RFs was based on the zeroth-order Gaussian function

\[
s_0(x) = e^{-\frac{(x-x_0)^2}{2\sigma_x^2}},
\]

where \(x_0\) is the center and \(\sigma_x\) is the standard deviation. The spatial profile—depicted in Figure 9—was the first-order derivative to \(x\) of a two-dimensional Gaussian function.

The spatial profile reverses in polarity. We assume that for the ideal edge defined by \(H_{d}(x)\), the length of the contour in the \(y\) dimension, \(\phi(t)\) is the temporal envelope (Equation 2), and \(H_{d}(x)\) is the Heaviside step function

\[
H_d(x) = \begin{cases} 
0 & x < a \\
1 & x \geq a.
\end{cases}
\]

The temporal profile was a sinusoid windowed by a Gaussian. The full spatiotemporal profile of the RFs was thus given by

\[
S(x, y, t) = s_1(x') \cdot s_0(y) \cdot s_0(t) \cdot \sin(2\pi f_{i} t),
\]

where \(f_{i}\) is the temporal frequency at which the spatial profile reverses in polarity. We assume that for the ideal detector, \(f_{i}\) is matched to the stimulus.

The input stimulus was a vertical counterphasing edge defined by

\[
I(x, y, t) = [H_{d}(y) - H_{d}(y)] \cdot [H_0(-x) - H_0(x)] \cdot \phi(t) \cdot \sin(2\pi f_{i} t),
\]

where \(l\) is the length of the contour in the \(y\) dimension, \(\phi(t)\) is the temporal envelope (Equation 2), and \(H_{d}(x)\) is the Heaviside step function

\[
H_d(x) = \begin{cases} 
0 & x < a \\
1 & x \geq a.
\end{cases}
\]

We assumed that the response at \(t_0\) of an RF centered at \((x_0, y_0)\) was given by the correlative integral

\[
R(x, y, t) = \int I(x, y, t) \cdot S(x - x_0, y - y_0, t - t_0) \cdot dx \cdot dy \cdot dt.
\]

This was calculated by multiplying the RF and the stimulus in the frequency domain,

\[
\tilde{R}(f_x, f_y, f_z) = \tilde{I}(f_x, f_y, f_z) \cdot \tilde{S}(f_{-x}, f_{-y}, f_{-z}).
\]

The output of each unit was then half-wave rectified and integrated across time,

\[
O(x, y) = \int R_0[R(x, y, t)] \cdot dt.
\]

We used the Weibull psychometric function (Weibull, 1951) to relate the output of the \(i\)th unit to the probability \(P_i\) of that unit detecting the stimulus,

\[
P_i = 1 - e^{-[C_i A O(x, y)]^\beta},
\]

where \(C_i\) is the stimulus contrast, \(A\) is the absolute sensitivity of the unit, and \(\beta\) is the slope parameter of the psychometric function. Probability summation was performed across space according to the Quick pooling formula (Graham, 1989; Quick, 1974), such that the observer’s probability \(P\) of detecting the stimulus is

\[
P = 1 - e^{-\int [C_i A O(x, y)]^\beta \cdot dx \cdot dy}.
\]

For a given probability of detection, contrast threshold \(C_i\) is thus

\[
C_i = [\int A O(x, y)^\beta \cdot dx \cdot dy]^{-\frac{1}{\beta}}.
\]

Figure 9. Spatial sensitivity profile of the receptive field employed in the model. The profile was the first-order derivative to \(x\) of a two-dimensional Gaussian function.
By fitting this equation to the data from Experiment 3, we were able to find the RF size that best accounted for our observations. The two free parameters were the standard deviation of the two-dimensional spatial Gaussian ($\sigma_x$ and $\sigma_y$ in Equation 9 and Equation 10, employed here as a single parameter $\sigma$, where $\sigma = \sigma_x = \sigma_y$) and the absolute sensitivity of an individual unit ($A$ in Equations 15, 16, and 17). All units at a given eccentricity had the same values for each parameter. To estimate $\beta$, we fitted a Weibull function to the psychometric data from each staircase by a least-squares procedure. For each observer and eccentricity, the mean $\beta$ of the 60 (12 contour lengths × 5 staircases) functions was employed in the model (for JH, mean $\beta = 4.60$, $SD = 0.16$; for LJ, mean $\beta = 4.54$, $SD = 0.26$; for PG, mean $\beta = 4.61$, $SD = 0.10$; and for XV, mean $\beta = 4.60$, $SD = 0.21$).

For each observer individually, we determined the values of the two free parameters $\sigma$ and $A$ at each eccentricity that best fit the data describing contrast threshold as a function of contour length. We used a nonlinear least-squares procedure based on the Matlab Optimization Toolbox function lsqnonlin and performed each fit 25 times with randomized starting values. The algorithm successfully found reliable parameter estimates for 20 of the 24 functions; the mean 95% confidence interval on the estimate of $\sigma$ was 0.56° with a standard deviation of 0.38°. In four cases, 95% confidence intervals on the estimate of $\sigma$ exceeded 2°; these four data points (JH for $E = 6^\circ$ and 12°, LJ for $E = 12^\circ$, and XV for 4°) were excluded from further analyses.

For comparison with the $g_{\text{max}}$ method, the model was also fitted to the data from Experiment 2 (see Figure 3) according to the same procedure. We calculated the probability of a correct response as

$$P' = P + \frac{1 - P}{4},$$

where $P$ is the probability of detecting the stimulus according to Equation 16. The $\beta$ parameter was fixed at 5.20, which was the mean slope of all Weibull functions fitted to the data from each staircase in Experiment 1. Proportions correct were converted to $d'$, such that the fitting procedure optimized the parameters $\sigma$ and $A$ by minimizing the least-squares error of $d'$ values predicted by the model, compared to observed $d'$ values.

For the purposes of comparison with physiological RF size data, RF size was deemed to be $2\sigma$. This conversion was based on figure 3 of Dow et al. (1981), in which the width of a hand-mapped RF corresponds to two standard deviations of the Gaussian-like spatial function derived from computer mapping.