An image-computable model of attention and texture segmentation

Michael Jigo[1*], David J. Heeger[1,2] & Marisa Carrasco[1,2]

[1] Center for Neural Science and [2] Department of Psychology, New York University

[*] corresponding author

Key words:

computational model, texture segmentation, visual field, voluntary attention, involuntary attention, normalization, spatial resolution

Corresponding author:

Michael Jigo
Center for Neural Science
New York University
6 Washington Place, Room 974
New York, NY 1003
Email: michael.jigo@nyu.edu

Author contributions

M.J., D.J.H. and M.C. conceived the project; M.J. implemented the model with input from D.J.H. and M.C.; M.C. provided all the data to be modelled; M.J. wrote the manuscript with guidance and supervision from M.C. All three authors edited the manuscript.

Acknowledgements

We thank Jonathan Winawer, Michael Landy, Hsin-Hung Li and other members of the Carrasco lab for their helpful comments and discussion.

ABSTRACT

Attention can facilitate or impair texture segmentation, altering whether objects are isolated from their surroundings in visual scenes. We simultaneously explain several empirical phenomena of texture segmentation and its attentional modulation with a single image-computable model. At the model's core, segmentation relies on the interaction between sensory processing and attention, with different operating regimes for involuntary and voluntary attention systems. Model comparisons were used to identify computations critical for texture segmentation and attentional modulation. The model reproduced (i) the central performance drop, which is the parafoveal advantage for segmentation over the fovea, (ii) the peripheral improvements and central impairments induced by involuntary attention and (iii) the uniform improvements across eccentricity by voluntary attention. The proposed model reveals distinct functional roles for involuntary and voluntary attention and provides a generalizable quantitative framework for predicting the perceptual impact of attention across the visual field.

INTRODUCTION

Texture segmentation is crucial for detecting objects across the visual field (Bergen and Adelson, 1988; Landy and Graham, 2004; Landy, 2013; Victor et al., 2017). This process isolates regions of local structure from their surroundings. Textures comprise information that span multiple spatial scales (spatial frequency), each conveying different levels of detail (Landy and Graham, 2004; Landy, 2013; Victor et al., 2017). The visual system's spatial resolution (acuity) constrains the visibility of detail and limits texture segmentation (Kehrer, 1989; Morikawa, 2000). Spatial resolution is high at the fovea and declines with eccentricity, which blurs the neural representation of the periphery (Graham, 1989; DeValois and DeValois, 1990). Covert attention—the prioritization of a spatial location without concurrent eye movements—modifies spatial resolution (Carrasco, 2011; Anton-Erxleben and Carrasco, 2013), shaping the level of detail that can be resolved. As a result, the success of segmentation hinges on the eccentricity of a texture target (the to-be-segmented region) (Kehrer, 1989; Morikawa, 2000) and on the attentional locus (Carrasco, 2011; Anton-Erxleben and Carrasco, 2013).

Conventional models treat texture segmentation as a feedforward process that filters an image into its elementary features (e.g., spatial frequency (SF) and orientation), transforms them to reflect the local structure (e.g., a region of similarly oriented bars irrespective of their individual positions), then pools across space to emphasize texture-defined contours (Bergen and Adelson, 1988; Landy and Graham, 2004; Landy, 2013). Few of these models account for variations in spatial resolution across eccentricity (Kehrer, 1997; Kehrer and Meinecke, 2003) and all postulate that texture segmentation is a 'preattentive' (Victor et al., 2017) operation whose underlying neural processing is impervious to attention (Bergen and Adelson, 1988; Kehrer, 1997; Kehrer and Meinecke, 2003; Landy and Graham, 2004; Landy, 2013).

However, attention alters texture segmentation (Carrasco, 2011; Anton-Erxleben and Carrasco, 2013). Attending adjacent textures can facilitate or impair segmentation, depending on the type of attention deployed and the texture eccentricity. Involuntary spatial attention—the transient, stimulus-driven form of selection (Carrasco, 2011)—improves segmentation for peripheral stimuli but impairs it near the fovea, a consequence of an automatic attentional enhancement of spatial resolution (Yeshurun and Carrasco, 1998, 2000; Talgar and Carrasco, 2002; Carrasco et al., 2006; Yeshurun and Carrasco, 2008; Jigo and Carrasco, 2018). Voluntary spatial attention—the sustained, goal-driven form of selection (Carrasco, 2011)—improves segmentation at all attended locations by optimizing spatial resolution to distinguish the textural boundary (Yeshurun et al.,

2008a; Barbot and Carrasco, 2017; Jigo and Carrasco, 2018). Owing to these distinct attentional effects, texture segmentation provides a formidable illustration of the stark differences between involuntary and voluntary attention.

Two main hypotheses have been proposed to explain how attention alters spatial resolution. Psychophysical studies ascribe attentional effects to modulations of SF sensitivity (Carrasco et al., 2006; Barbot and Carrasco, 2017). Neurophysiological (Moran and Desimone, 1985; Womelsdorf et al., 2006; Womelsdorf et al., 2008) and neuroimaging (Fischer and Whitney, 2009; Klein et al., 2014) studies bolster the idea that attention modifies spatial profiles of neural receptive fields (Anton-Erxleben and Carrasco, 2013). Both hypotheses qualitatively predict attentional effects but do not specify the neural computations underlying the dissociable effects of involuntary and voluntary attention.

Here, we propose a computational model in which feedforward processing and attentional gain operate across eccentricity and SF and contribute to texture segmentation performance. (i) Feedforward processing. We augment a conventional model of texture processing (Bergen and Adelson, 1988; Landy and Graham, 2004; Landy, 2013). Our model varies with eccentricity and includes contextual modulation within local regions in the stimulus via normalization (Heeger, 1992). The defining characteristic of normalization is that the response of an individual neuron is (divisively) suppressed by the summed activity of neighboring neurons responsive to different positions and features in the scene. This canonical neural computation is widespread across basic sensory and higher-order perceptual and cognitive processing (Carandini and Heeger, 2011). (ii) Attentional gain. We model attention as multiplicative gains (attentional gain factors (Reynolds and Heeger, 2009)) that vary with eccentricity and SF. Attention shifts sensitivity toward fine or coarse details depending on the range of SFs enhanced.

Our model is image-computable, which allowed us to reproduce behavior directly from grayscale texture images from several experiments (Yeshurun and Carrasco, 1998; Talgar and Carrasco, 2002; Carrasco et al., 2006; Yeshurun and Carrasco, 2008; Yeshurun et al., 2008a; Barbot and Carrasco, 2017). The model explained three signatures of texture segmentation within a single computational framework (**Figure 1**): (i) the central performance drop (CPD) (Kehrer, 1989; Yeshurun and Carrasco, 1998; Morikawa, 2000; Yeshurun and Carrasco, 2000; Talgar and Carrasco, 2002; Carrasco et al., 2006; Yeshurun and Carrasco, 2008; Yeshurun et al., 2008a; Barbot and Carrasco, 2017; Jigo and Carrasco, 2018) (**Figure 1a**), i.e., the parafoveal advantage of segmentation over the fovea, (ii) the improvements in the periphery and impairments at central

locations induced by involuntary attention (Yeshurun and Carrasco, 1998, 2000; Talgar and Carrasco, 2002; Carrasco et al., 2006; Yeshurun and Carrasco, 2008; Yeshurun et al., 2008a; Jigo and Carrasco, 2018) (**Figure 1b**) and (iii) the equivalent improvements across eccentricity by voluntary attention (Yeshurun et al., 2008a; Barbot and Carrasco, 2017; Jigo and Carrasco, 2018) (**Figure 1c**). Whereas our analyses are focused on texture segmentation, our model is general and can be applied to any psychophysical discrimination or detection task with static, grayscale images. For example, the model can predict contrast sensitivity functions across eccentricity (**Supplementary Figure 1**).

To preview our results, model comparisons revealed that normalization is necessary to elicit the CPD (Morikawa, 2000) and that the effects of involuntary and voluntary attention are generated by distinct profiles of gain enhancement across SF (Jigo and Carrasco, 2020). Impairments by involuntary attention are explained by a preferential high-SF enhancement that shifts visual sensitivity toward details too fine to distinguish the texture at central locations. The transition from impairments to improvements in the periphery results from involuntary attentional gain gradually shifting to lower SFs that are more amenable to distinguish the texture target. Improvements by voluntary attention result from a uniform enhancement of SFs that encompass the target, optimizing visual sensitivity for the texture across eccentricity.

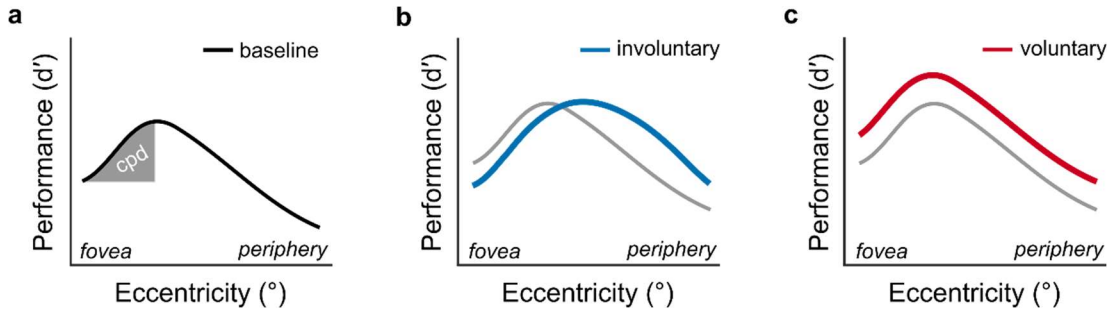


Figure 1. Signatures of texture segmentation.
(**a**) Central performance drop. Shaded region depicts the magnitude of the central performance drop.
(**b**) Involuntary attention modulation. Involuntary attention improves segmentation performance in the periphery and impairs it near the fovea.
(**c**) Voluntary attention modulation. Voluntary attention improves segmentation performance across eccentricity.

RESULTS

Image-computable model of attention and texture segmentation

The model is motivated by established principles of neural computation (Heeger, 1992; Carandini and Heeger, 2011), human pattern (Graham, 1989; DeValois and DeValois, 1990) and texture vision (Bergen and Adelson, 1988; Landy and Graham, 2004; Landy, 2013) and attentional modulation (Reynolds and Heeger, 2009). The model extends the Reynolds-Heeger normalization model of attention (NMA) (Reynolds and Heeger, 2009) to capture effects of attention on contrast and texture sensitivity across SF and eccentricity. The model emulates the decline in contrast sensitivity and peak SF preference with eccentricity, generates texture selectivity via spatial summation of normalized inputs and incorporates distinct modes through which involuntary and voluntary attentional gain modulate SF sensitivity. The model is composed of four components: stimulus drive, attentional gain, suppressive drive and spatial summation (**Figure 2a**). Following NMA, attention adjusts the gain on the stimulus drive before normalization. For a full description of the model, see **Online Methods**.

Stimulus drive. We simulate bottom-up responses of a large number of linear receptive fields, each jointly tuned to spatial position, SF and orientation. Images are processed through a filter bank (Simoncelli et al., 1992) covering the visual field at several SFs and orientations using bandwidths compatible with neurophysiological (DeValois and DeValois, 1990) and psychophysical (Graham, 1989) measurements. Filter outputs are combined across quadrature phase (Adelson and Bergen, 1985), yielding contrast energy images corresponding to different SFs and orientations. These outputs simulate the responses of complex cells in primary visual cortex (Adelson and Bergen, 1985; DeValois and DeValois, 1990). The gain on individual receptive fields varies as a function of SF and eccentricity preference (**Figure 2a, green**). Following the behavior of individual neurons (DeValois and DeValois, 1990) and human pattern vision (Graham, 1989), gain modulation is narrowly tuned to high SFs near the fovea and progressively shifts to low SFs in the periphery. Consequently, the stimulus drive reflects local spectral energy within each patch in the image, filtered through feature-selective receptive fields that vary with eccentricity.

Attentional gain. Attention is implemented as a gain control mechanism that multiplicatively scales the gain on the stimulus drive (Reynolds and Heeger, 2009). The magnitude of attentional gain is largest at the cued location (**Figure 2a, orange ring**) and varies with the eccentricity and SF preference of each receptive field. Motivated by findings of psychophysical experiments that manipulated voluntary and involuntary attention (Jigo and Carrasco, 2020), two candidate profiles

are assessed—narrow and broad. The narrow profile selectively enhances a small range of SFs at each eccentricity (**Figure 2a, blue**); the broad profile uniformly enhances SFs (**Figure 2a, red**).

Suppressive drive. Suppression operates via divisive normalization, a canonical computation in sensory processing and cognition (Heeger, 1992; Carandini and Heeger, 2011). Normalized responses are proportional to the attention-scaled stimulus drive divided by a normalization pool plus a constant σ^2 that increases with eccentricity. This constant adjusts the model's overall sensitivity to contrast (i.e., contrast gain; **Figure 2a, black**). The normalization pool consists of the attention-scaled stimulus drive across nearby spatial locations (surround suppression (Cavanaugh et al., 2002)), uniformly across orientation (cross-orientation suppression (Brouwer and Heeger, 2011)) and across preferred and neighboring SFs (cross-frequency suppression (Petrov et al., 2005)) of individual receptive fields.

Spatial summation. Normalized responses are weighted and summed across space within each SF and orientation filter. Spatial summation accentuates texture-defined contours within the image. The size of pooling regions scale with the SF preference of each receptive field (Landy and Graham, 2004; Landy, 2013) (**Figure 2a, purple**); larger for low than for high SFs. This implements an inverse relation between the integration area of individual receptive fields and their SF tuning.

Target discriminability. We used the model to generate measures of discriminability (d′) in a texture segmentation task (**Figure 2b**). Two texture images were input to the model and their respective population responses were computed. One texture contained a target patch of orientation that differed from its surround (target-present) and the other consisted of uniform orientation throughout (target-absent). The difference between their respective population responses was computed and the length of this vector (i.e., Euclidian norm) indexed stimulus discriminability. This measure is proportional to behavioral performance, assuming normally-distributed noise is added after normalization (**Online methods, Decision mechanism**).

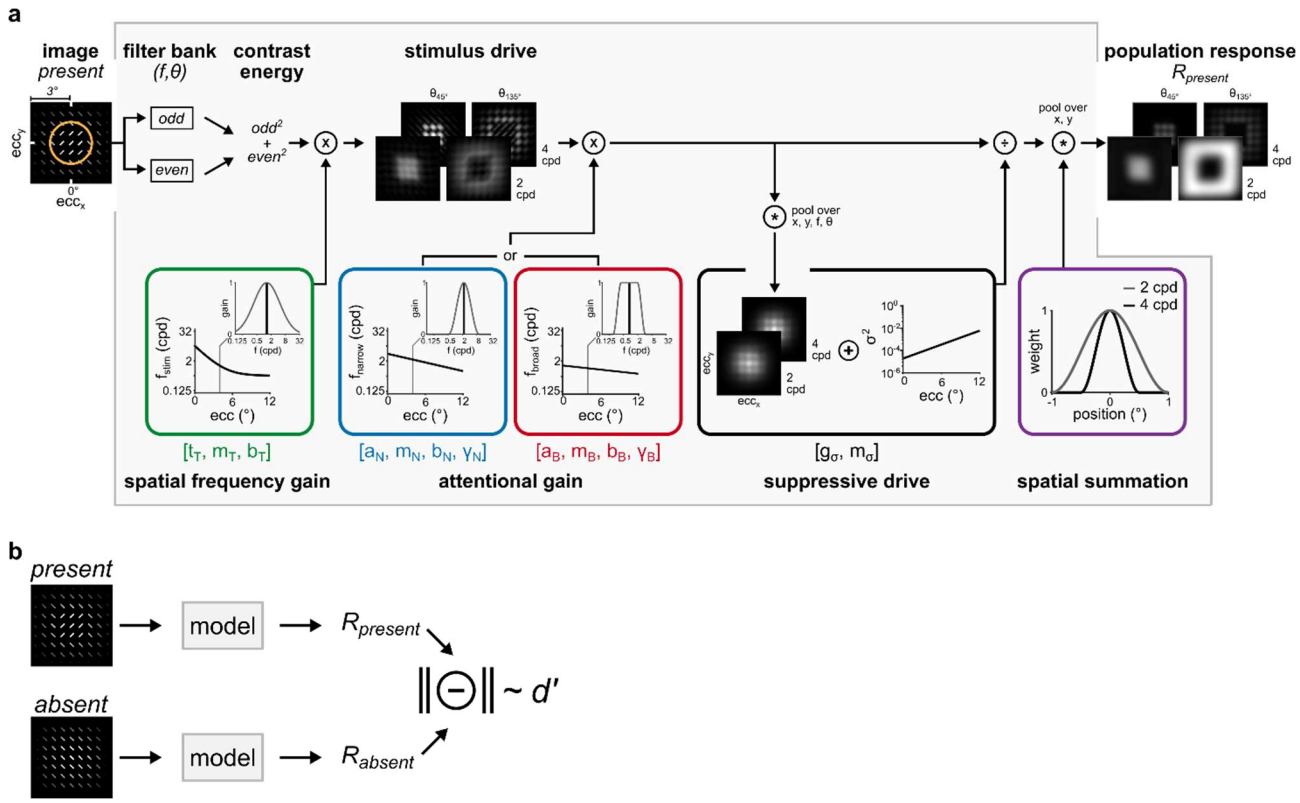


Figure 2. Model structure and target discriminability.
(**a**) Model structure. An image is filtered through a bank of linear receptive fields, each varying in their preferred position, SF and orientation. Filter responses are squared and summed across quadrature-phase pairs (odd, even), yielding contrast energy outputs. *Spatial frequency gain* scales contrast energy across spatial frequency and eccentricity (green box). The solid black line depicts the center frequency of the tuning function (f_{stim}); the full tuning function at a single eccentricity is shown in the inset. The *stimulus drive* characterizes the contrast energy at each pixel in the image, filtered through feature-selective and eccentricity-dependent receptive fields. A subset of receptive fields tuned to two SFs and orientations are displayed. *Attentional gain* multiplicatively scales the stimulus drive at a circumscribed region within the image (orange circle in left panel) and varies across spatial frequency and eccentricity. The center SF of attentional gain varies with eccentricity (solid black lines in blue and red boxes) and is distributed according to distinct profiles. The narrow profile (blue box) selectively enhances a given frequency (f_{narrow}), as depicted in the inset. The broad profile (red box) is centered on a given frequency (f_{broad}) and uniformly enhances a range of SFs (inset). The *suppressive drive* is computed through pooling the attention-scaled stimulus drive across a local neighborhood of positions, SFs and uniformly across orientation. Overall suppression magnitude is adjusted by σ^2, which increases with eccentricity. *Spatial summation* follows normalization (purple box) and generates the *population response*. Pooling area varies inversely with SF tuning. Parameters that were fit to behavior are displayed within square brackets for each model component.
(**b**) Target discriminability. Population responses are computed for a texture image containing a patch (*present*) and one without (*absent*). The vector magnitude of their difference produces a metric proportional to d′, assuming independent and identically distributed Gaussian output noise.

Texture stimuli, behavioral protocol and optimization strategy

Stimuli. Model parameters were constrained by data from ten published psychophysical experiments. Involuntary attention was manipulated in six experiments (Yeshurun and Carrasco, 1998; Talgar and Carrasco, 2002; Carrasco et al., 2006; Yeshurun and Carrasco, 2008; Yeshurun et al., 2008a) (**Figure 3a**) and voluntary attention in four experiments (Yeshurun et al., 2008a; Barbot and Carrasco, 2017) (**Figure 3b**). In each experiment, observers distinguished a patch of one orientation embedded within a background of differing orientation at several possible eccentricities. Grayscale images of the experiments' stimuli were input to the model.

Behavioral protocol. Performance was typically measured with a two-interval forced choice protocol (**Figure 3c**). Observers maintained fixation at the display's center while viewing two intervals of texture stimuli, one of which randomly contained a target texture. Involuntary or voluntary attention was manipulated with different pre-cues at their optimal timing. Involuntary attention was manipulated with brief peripheral pre-cues displayed before both intervals, but near the upcoming target location in the interval containing the target (Yeshurun and Carrasco, 1998, 2000; Talgar and Carrasco, 2002; Carrasco et al., 2006; Yeshurun and Carrasco, 2008; Yeshurun et al., 2008a; Jigo and Carrasco, 2018). Voluntary attention was manipulated with symbolic pre-cues that appeared near fixation and indicated the target location in the target-present interval (Yeshurun et al., 2008a; Barbot and Carrasco, 2017). The effects of each attention type were determined relative to a neutral condition, in which attention was distributed across all possible target locations. Behavioral performance displayed the three signatures of texture segmentation. First, the CPD emerged in the neutral condition (**Figure 1a**). Second, peripheral pre-cues improved performance in the periphery and impaired it at central locations (**Figure 1b**). Third, central, symbolic pre-cues improved performance at all eccentricities (**Figure 1c**).

Optimization. To identify the computations that underlie each signature of texture segmentation, the model was separately fit to three subsets of behavioral data. First, the CPD was isolated from attentional effects by fitting to the neutral condition from all ten experiments. Second, involuntary attentional effects were assessed by fitting to neutral and peripheral cueing conditions from the six involuntary attention experiments. Third, voluntary attentional effects were assessed by fitting to neutral and central cueing conditions from the four voluntary attention experiments (**Online methods, Optimization strategy**).

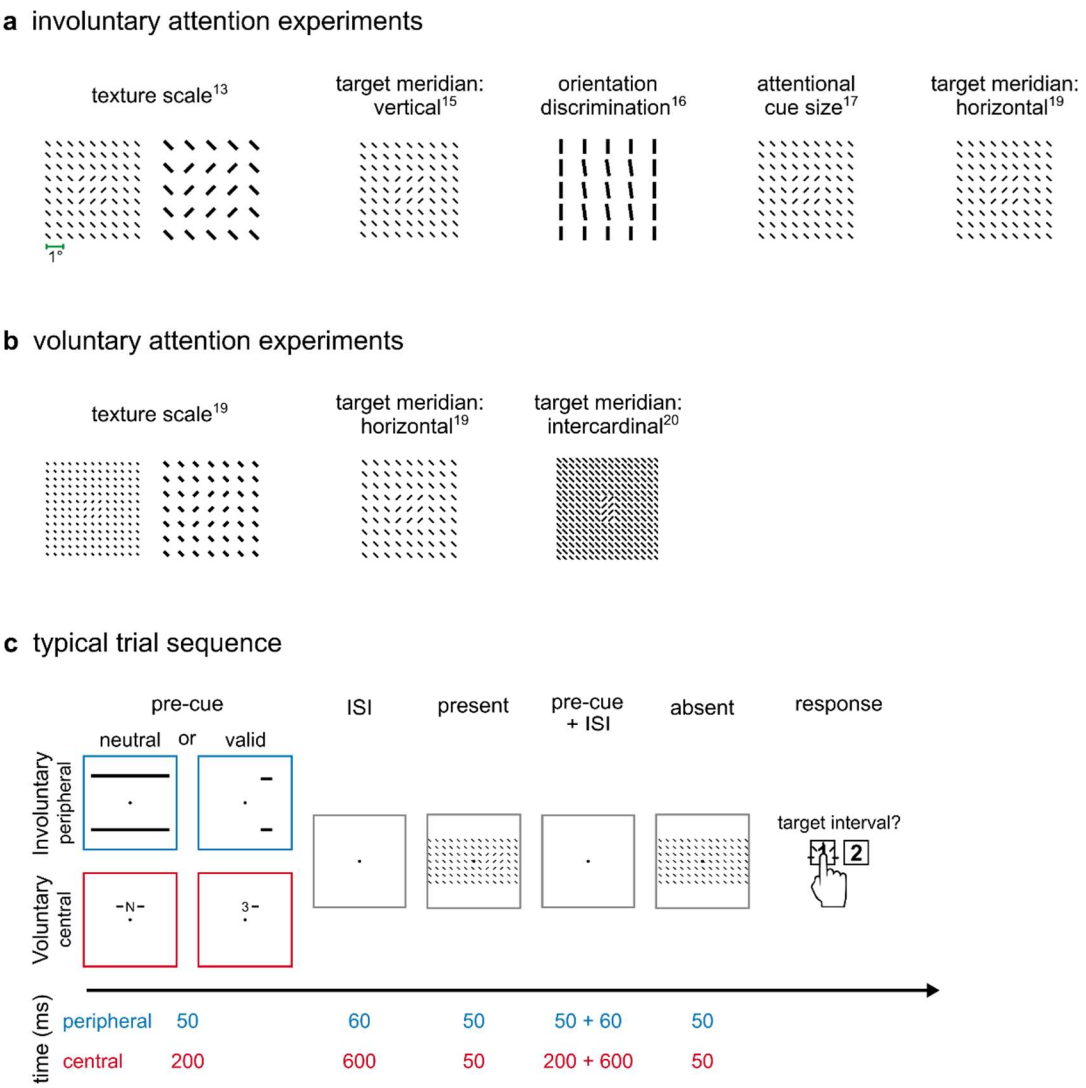


Figure 3. Texture stimuli and a typical behavioral protocol.
Target-present texture stimuli, displayed at their respective spatial scales, used in experiments in which (**a**) involuntary attention and (**b**) voluntary attention were manipulated. Scale bar (green) applies to all texture stimuli. Target-absent stimuli were homogenous textures. Each texture is labeled with the experimental manipulation conducted in the corresponding experiment; numbers indicate the study's citation.
(**c**) Two-interval forced choice protocol typically used to assess texture segmentation performance across eccentricity. Representative timing information for peripheral (blue) and central (red) pre-cues, and their corresponding inter-stimulus intervals (ISI) are displayed. Peripheral and central pre-cues manipulated involuntary and voluntary attention, respectively. Neutral pre-cues equally distributed attention to all possible target locations. Valid peripheral pre-cues were displayed near the upcoming target location whereas valid central pre-cues symbolically indicated the upcoming target location. In the displayed example, the number "3" and the adjacent line indicate that the target would appear at a peripheral eccentricity in the right visual hemifield.

Contextual modulation and spatial summation mediate the central performance drop

To identify the computations mediating the CPD, the model was simultaneously fit to group-average performance in each experiment's neutral condition (103 data points). 15 model parameters constrained performance (**Online methods, Table 2**). To account for differences in contrast sensitivity due to variable display properties among experiments (e.g., mean luminance), foveal contrast gain (g_σ; **Figure 2a**) was independently determined for each of ten experiments (10 parameters). Foveal spatial frequency tuning (t_T) was determined by two parameters–one shared among involuntary attention studies and another among voluntary attention studies. The remaining three parameters–SF bandwidth (b_T), slopes of contrast (m_σ) and SF tuning decline (m_T)–were shared among all experiments. Attentional gain was not included for these fits.

The model reproduced the CPD and its dependence on texture scale (**Figure 4**). With no change in parameters, the model predicted contrast sensitivity functions reasonably well (**Supplementary Figure 1**). For a fine-scale texture—characterized by narrow, densely spaced lines—performance peaked within the parafovea (4 deg) and declined toward the fovea and periphery (**Figure 4a**). Model fits to a fine-scale texture revealed that differences between target-present and target-absent stimuli were largest within the 2 cpd filter (**Figure 4a, middle**). This filter best differentiated the target patch from a homogenous texture; we denote its center SF as f_{fine}. A coarser texture was best distinguished by lower SFs (1 cpd, f_{coarse}), which exaggerated the CPD, moving peak performance to a farther eccentricity (~6 deg; **Figure 4b**). The CPD was well-fit by the model for all experiments (**Figure 4c**) and explained 77% of the variance (95% bootstrapped CI [70 80]), with the best-fitting regression line falling close to the unity line.

Previous models have qualitatively matched the CPD through spatial summation (Kehrer, 1997; Kehrer and Meinecke, 2003), but ignored the contributions of contextual modulation via normalization. To assess whether each operation is required to quantitatively reproduce behavior, the full model (described above) and two variants were compared. One variant lacked contextual modulation (**Figure 4d**, –context), wherein each receptive field was normalized by its response alone. Consequently, modulation due to spatial positions and featural attributes outside a receptive field's tuning was removed. Another variant lacked spatial summation (**Figure 4d**, –sum), wherein the population response comprised only normalized outputs. Without spatial summation, regions of similar orientation (e.g., target patch) were not pooled, attenuating the response to texture. Each model was fit to behavioral performance in the neutral condition across all experiments and compared using Akaike Information Criterion (Akaike, 1974) (AIC).

Each model variant could not capture performance. In each variant, the magnitude of the CPD was attenuated or eliminated (**Supplementary Figure 2a**). Model performance was measured relative to the full model, which yielded a ΔAIC score; positive values represent poorer performance than the full model. Without contextual modulation, model performance dropped by an average ΔAIC score of 47 (*SE*=5.1), and without spatial summation model performance fell by an average score of 77 (*SE*=2.3). Thus, contextual modulation and spatial summation were required to reliably generate the CPD.

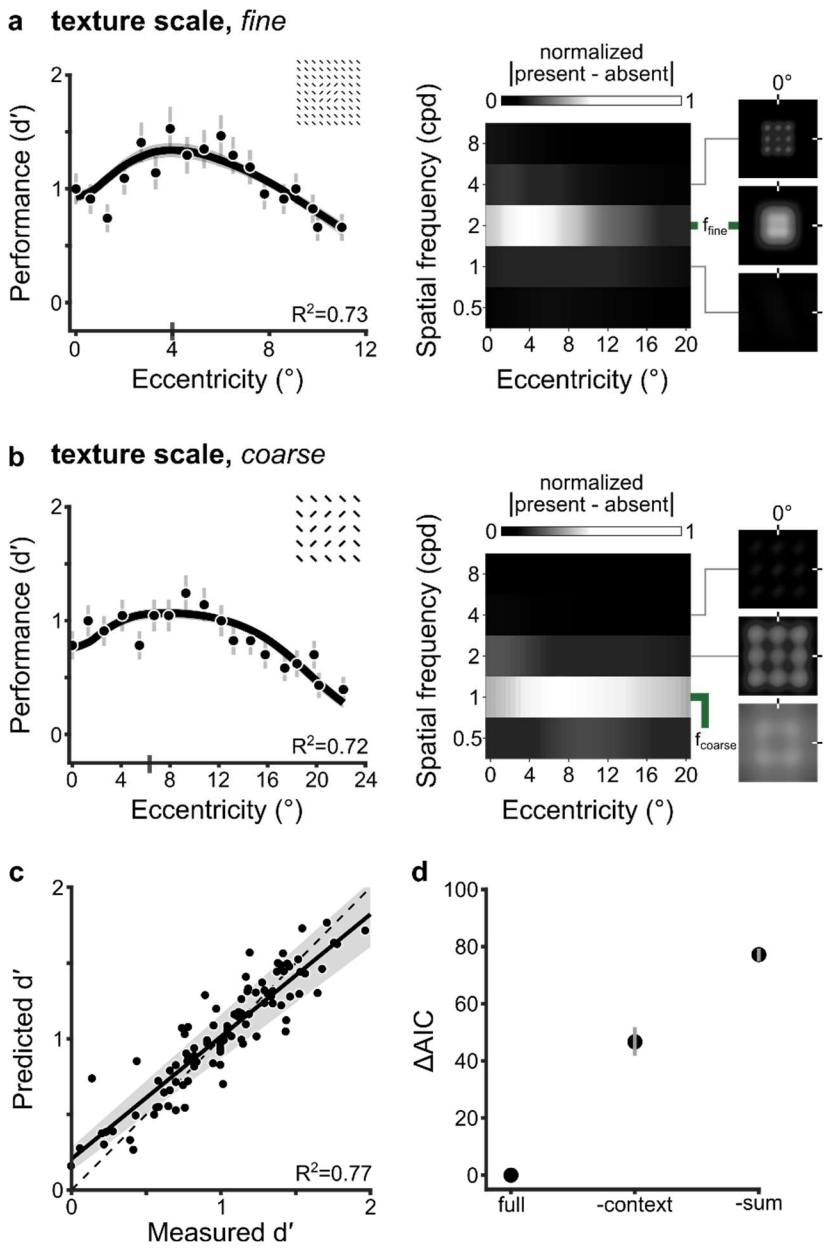


Figure 4. Model fits to the central performance drop.
(**a**) *Left*. Fit to Experiment 1 in Yeshurun & Carrasco (1998) (Yeshurun and Carrasco, 1998) which used a fine-scale texture. Black dots and error bars depict group-average performance and ±1 SEM. The black line indicates the median and the shaded region indicates the 68% confidence interval of the bootstrapped distribution of model fits. The model was fit simultaneously to the neutral condition of all ten experiments. The vertical gray bar on the x-axis indicates the eccentricity at which performance peaked. The inset in the top-right depicts the experiment's texture stimulus.
Middle. The matrix depicts the absolute value of the difference between target-present and target-absent population responses, normalized by the maximum across eccentricity. Differences were averaged across orientation and space. The spatial frequency filter that contained the largest difference between the fine-scale textures is denoted by f_{fine}. Note that target discriminability was determined using the vector length of differences (**Figure 2b**); the absolute difference displayed in the matrix is used to visualize the spatial frequencies that drove discriminability.
Right. This column depicts the spatial distribution of the absolute value of differences between target-present and target-absent population responses. Each panel depicts a subset of receptive fields when the texture was centered on the fovea (indicated by tick marks). Receptive fields were tuned to one of three spatial frequencies (4, 2, 1 cpd) and an orientation of 30°.
(**b**) Fit to Experiment 2 in Yeshurun & Carrasco (1998)(Yeshurun and Carrasco, 1998), which used a coarse-scale texture. This fit was determined simultaneously with the fit shown in *a*. Visualization follows the conventions in *a*. Note that eccentricity (x-axis) is twice that of *a*. f_{coarse} denotes the spatial frequency filter that best distinguished the coarse-scale target.
(**c**) Goodness-of-fit for simultaneous fit to performance in all neutral conditions. Each dot corresponds to the measured (x-axis) and predicted (y-axis) behavioral performance at a given eccentricity. The solid line and shaded area depict the best-fitting regression line and its 95% confidence interval, respectively. The dashed line indicates the unity line y=x.
(**d**) Model comparisons using AIC. ΔAIC is the AIC of each model minus the AIC of the full model. '-context' denotes the model variant without contextual modulation and '-sum', without spatial summation. The dots represent average ΔAIC and error bars denote 68% confidence interval of the bootstrap distribution.

Narrow high-spatial frequency enhancement mediates involuntary attention effects

To identify the computations mediating involuntary attentional effects, the model was simultaneously fit to group-average data from neutral and peripheral cueing conditions across six experiments (146 data points). Involuntary attention was modeled as a narrow gain profile (**Figure 2, blue**), motivated by psychophysical measurements (Jigo and Carrasco, 2020). 14 free parameters constrained model behavior (**Online methods, Table 3**). Model parameters that determined neutral cueing performance—foveal contrast gain (g_σ), SF tuning (t_T), bandwidth (b_T) and the slope of contrast (m_σ) and SF decline (m_T)—were configured identically as described above. Attentional gain was controlled by four parameters that were shared among experiments–foveal spatial frequency preference (a_N), slope of shift to lower frequencies with eccentricity (m_N), bandwidth (b_N) and amplitude (γ_N). Consequently, attention operated identically on each texture stimulus, with the spatial spread of attention fixed across experiments (**Online methods, Table 1**).

The model reproduced the central impairments and peripheral improvements and their variation with texture scale. For a fine-scale texture, the narrow gain profile across SF yielded improvements within the parafovea (4-12°), impairments across a small range of central eccentricities (0-2°) and

shifted peak performance toward the periphery (~6°; **Figure 5a**). For the coarser texture, the same attentional gain profile generated improvements in the periphery (8-22°), impairments within the parafovea (0-8°) and shifted peak performance farther toward the periphery (~15°; **Figure 5b**).

The transition from impairments to improvements was generated by a gradual shift of attentional gain to lower SFs with eccentricity (**Figure 5c**), in line with empirical measurements (Jigo and Carrasco, 2020). At the fovea, attentional gain was centered on a SF (4 cpd) higher than those distinguishing the fine- (2 cpd, f_{fine}) or coarse-scale (1 cpd, f_{coarse}) textures. As a result, the population response was shifted away from the target and impaired performance. With increasing eccentricity, attentional gain progressively overlapped the SFs of fine- and coarse-scale targets, improving performance. Attention enhanced the fine-scale target SF within the parafovea (4-12°) then enhanced the coarse-scale target at farther eccentricities (8-22°). Overall, across both neutral and peripheral cueing conditions in six experiments, the model explained 77% of the variance (95% bootstrapped CI [49 82]; **Figure 5d**).

Across all six experiments, impairments at central locations were generated through attentional gain on SFs higher than the target (**Figure 5e**). Consequently, the overlap between fine- (f_{fine}) or coarse-scale (f_{coarse}) target SFs and the SF tuning of attentional gain was minimal at the fovea and peaked in the periphery (**Figure 5f**). This mismatch between the SF tuning of attention (f_{narrow}) and the target is suggested to be driven by involuntary attention operating above intrinsic SF preferences at each eccentricity (Jigo and Carrasco, 2020). We corroborated this relation. We compared f_{narrow} to the model's baseline SF tuning, indexed by the stimulus drive (f_{stim}, **Figure 2a**). Consistent with empirical measurements, we found that the narrow attentional gain profile preferred SFs higher than baseline (**Supplementary Figure 3**).

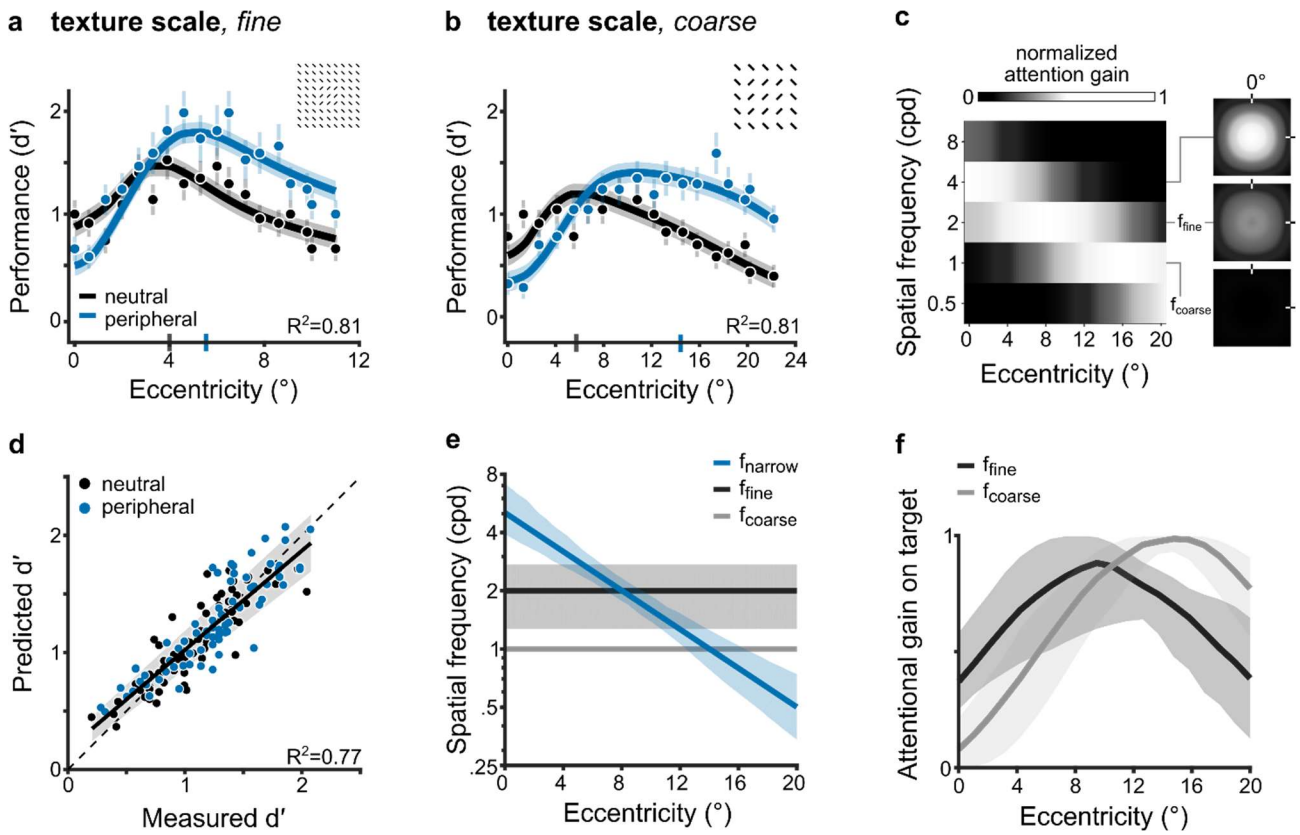


Figure 5. Model fits to involuntary attention experiments.
(**a**) Fit to Experiment 1 in Yeshurun & Carrasco (1998)(Yeshurun and Carrasco, 1998), which used a fine-scale texture. The dots depict the group-average and error bars denote ±1 SEM. Solid lines indicate the median and shaded regions indicate the 68% confidence interval of the bootstrapped distribution of model fits. The vertical blue bar on the x-axis indicates the eccentricity at which performance peaked with valid peripheral cues. The model was fit simultaneously to all six involuntary attention experiments.
(**b**) Fit to Experiment 2 in Yeshurun & Carrasco (1998)(Yeshurun and Carrasco, 1998), which used a coarse texture scale. This fit was determined simultaneously with that shown in *a*. Visualization follows the conventions in *a*.
(**c**) Best-fitting narrow gain profile.
Left. The matrix depicts attentional gain across eccentricity, normalized by the maximum and averaged across space and orientation. The spatial frequencies that best distinguished fine- (f_{fine}) and coarse-scale targets (f_{coarse}) in *a* and *b* are displayed.
Right. The spatial distribution of attentional gain on a subset of receptive fields when the texture was centered on the fovea (indicated by tick marks). Receptive fields were tuned to one of three spatial frequencies (1, 2, 4 cpd) and an orientation of 30°.
(**d**) Goodness-of-fit for simultaneous fits to performance on all neutral and peripheral cueing conditions. Plotted as in **Figure 4c.**
(**e**) Spatial frequency preference of the narrow attentional gain profile (f_{narrow}) and the spatial frequency that best distinguished target-present from target-absent stimuli for fine- (f_{fine}) and coarse-scale textures (f_{coarse}). The solid lines indicate the median and the shaded areas denote the 68% confidence interval of bootstrapped distribution. Shaded area for f_{coase} overlaps the solid line.
(**f**) Normalized magnitude of attentional gain on the fine- and coarse-scale target spatial frequency across eccentricity (median and 68% confidence interval of bootstrapped distribution).

Broad spatial frequency enhancement mediates voluntary attention effects

The model was simultaneously fit to group-average data from neutral and central cueing conditions across four experiments (60 data points). Voluntary attention was modeled as a broad gain profile (**Figure 2a, red**), motivated by psychophysical measurements (Jigo and Carrasco, 2020). 12 free parameters constrained model behavior (**Online methods, Table 4**). Four parameters determined attentional gain and were shared among experiments–foveal SF preference (a_B), slope of shift to lower SFs with eccentricity (m_B), bandwidth (b_B) and amplitude (γ_B).

The model reproduced improvements across eccentricity for both fine- (**Figure 6a**) and coarse-scale textures (**Figure 6b**). To generate these improvements, attentional gain encompassed the target SF for each texture scale (**Figure 6c**). Across all four experiments, the model explained 89% of the variance (95% bootstrapped CI [67 92]; **Figure 6d**).

Across all four experiments, broad attentional gain was centered near the target SF across eccentricity (f_{broad} in **Figure 6e**), unlike the narrow gain profile that modulated higher SFs at central locations (**Figure 5e**). Although the center SF of attention declined with eccentricity, the modulation profile's plateau ensured that it overlapped both fine- and coarse-scale target SFs across eccentricity (**Figure 6f**). Psychophysical measurements of attentional effects on contrast sensitivity (Jigo and Carrasco, 2020) suggest that the SF range enhanced by voluntary attention is centered near those intrinsically preferred by an observer at each eccentricity. However, our model fits to texture segmentation experiments revealed that attentional gain enhanced lower SFs than baseline tuning (f_{stim}) at central locations (**Supplementary Figure 4**).

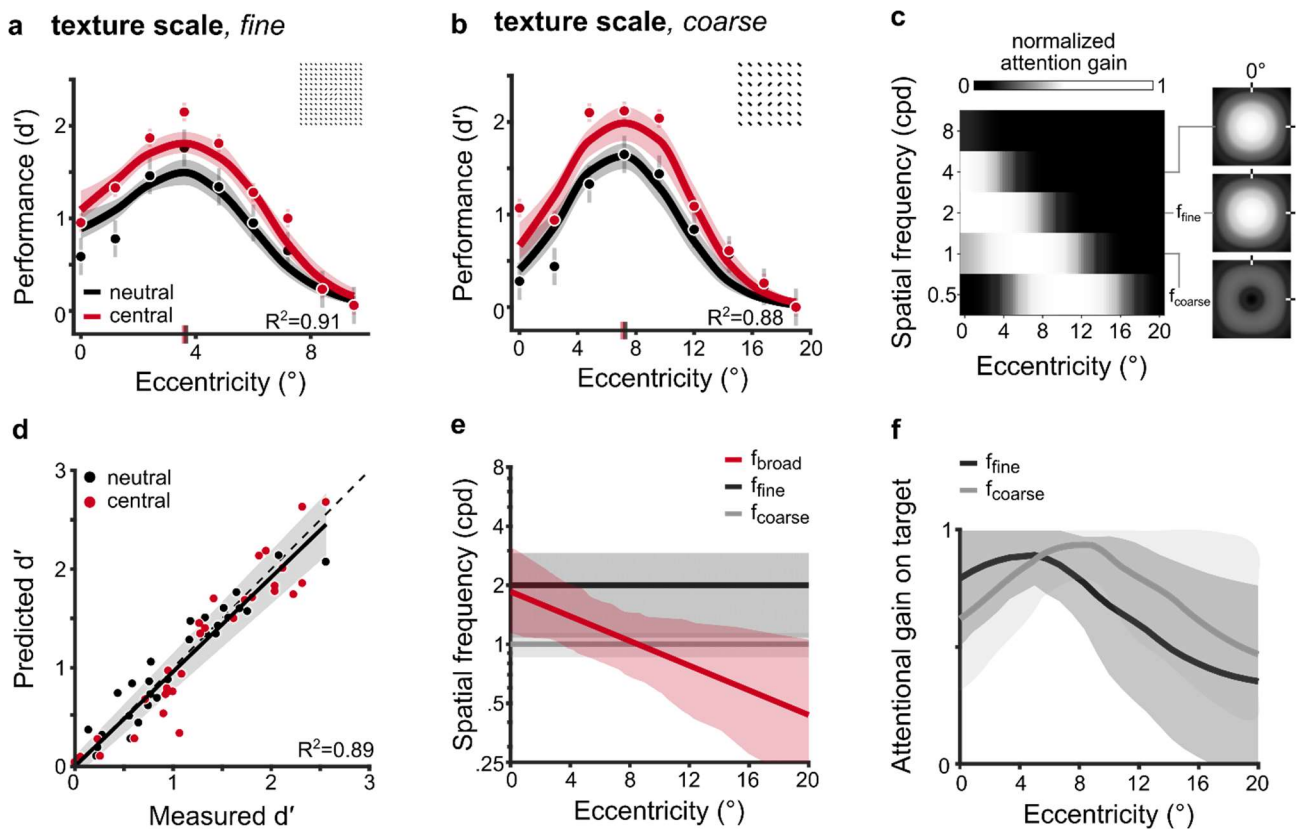


Figure 6. Model fits to voluntary attention experiments.
(**a**) Fit to Experiment 3 in Yeshurun, Montagna & Carrasco (2008)(Yeshurun et al., 2008a), which used a fine-scale texture. The dots depict group-average performance and error bars denote ±1 SEM. Solid lines indicate the median and shaded regions indicate 68% confidence intervals of the bootstrapped distribution of model fits. The model was fit simultaneously to all four voluntary attention experiments.
(**b**) Fit to Experiment 4 in Yeshurun, Montagna & Carrasco (2008)(Yeshurun et al., 2008a), which used a coarse-scale texture. This fit was determined simultaneously with that shown in *a*. Visualization follows the conventions in *a*.
(**c**) Best-fitting broad gain profile. Plotted as in **Figure 5c**.
(**d**) Goodness-of-fit for simultaneous fits to performance on all neutral and central cueing conditions. Each dot corresponds to the measured (x-axis) and predicted (y-axis) behavioral performance at a given eccentricity. Plotted as in **Figure 4c**.
(**e**) Average spatial frequency preference of the broad attentional gain profile (f_{broad}) and the spatial frequency that best distinguished target-present from target-absent stimuli for fine- (f_{fine}) and coarse-scale textures (f_{coarse}). Solid lines indicate the median and the shaded areas denote 68% confidence intervals of bootstrapped distributions.
(**f**) Normalized magnitude of attentional gain on the fine- and coarse-scale target spatial frequency across eccentricity (median and 68% confidence interval of bootstrapped distribution).

Different gain profiles mediate involuntary and voluntary attention effects

To provide statistical evidence that different gain profiles—narrow or broad— generate the effects of involuntary and voluntary attention, each profile was separately fit to involuntary and voluntary attention experiments. ΔAIC (AIC_{broad}-AIC_{narrow}) was used for model comparison (**Figure 7**); positive values represent an advantage of the narrow profile.

For involuntary attention, the narrow profile outperformed the broad profile (ΔAIC=8±4), which could not capture impairments at central locations (**Supplementary Figure 2b**). For voluntary attention, the broad profile outperformed the narrow profile (ΔAIC=-30±4), which could not capture improvements at central locations (**Supplementary Figure 2b**). These model comparisons corroborated psychophysical measurements(Jigo and Carrasco, 2020): involuntary and voluntary attention effects are best explained by different attentional gain profiles across SF.

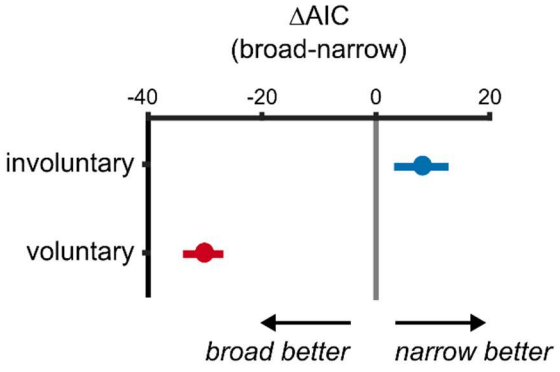


Figure 7. Model comparisons using ΔAIC. Dots represent median and error bars denote 68% confidence interval of the bootstrap distribution.

A parsimonious explanation for several experimental manipulations

Figure 8 depicts behavioral data for a variety of experimental manipulations. Whereas we focus on the impact of texture scale in **Figure 5 and 6**, the model is general. It simultaneously accounted for multiple target locations (vertical, **Figure 8a**; horizontal, **Figure 8c-e**; and intercardinal meridians, **Figure 8f**), behavioral tasks (orientation discrimination, **Figure 8b**; **Online methods, Stimulus generation**) and attentional manipulations (cue size, **Figure 8c**). Although the model was fit using texture images with fixed positions and orientations (**Figure 3**), it behaved similarly for textures with randomly jittered elements (**Supplementary Figure 5**). Therefore, the proposed model provides a parsimonious explanation for and a quantitative match to human texture segmentation performance (**Figure 8**) and contrast sensitivity (**Supplementary Figure 1**).

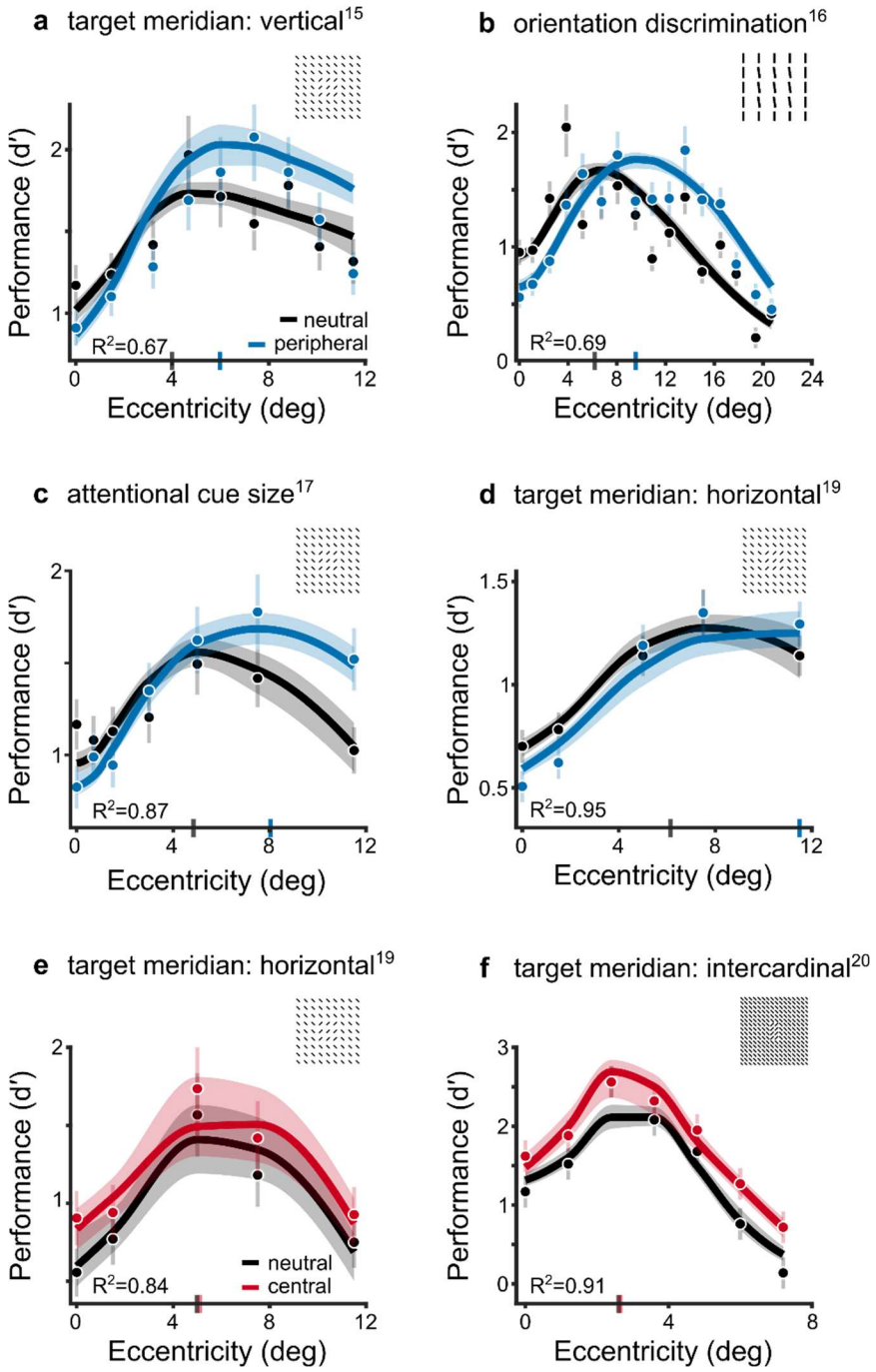


Figure 8. (**a-d**) Fits to involuntary attention experiments that used a narrow gain profile and were fit simultaneously with data displayed in **Figure 5**. (**e-f**) Fits to voluntary attention experiments that used a broad gain profile and were fit simultaneously with data displayed in **Figure 6**. Dots depict group-average and error bars indicate ±1 SEM. Solid lines indicate the median and shaded regions indicate 68% confidence intervals of the bootstrapped distribution of model fits. Labels above each panel match the labels in **Figure 3**, which describe the experimental manipulation used in the corresponding study; numbers denote the study's citation.

DISCUSSION

An image-computable model of sensory and attentional processing

The proposed model explains texture segmentation via normalization, spatial summation and attention. It links three signatures of texture segmentation to distinct cortical computations. (i) Divisive normalization and spatial summation mediate the central performance drop (Kehrer, 1989; Yeshurun and Carrasco, 1998; Morikawa, 2000; Yeshurun and Carrasco, 2000; Talgar and Carrasco, 2002; Carrasco et al., 2006; Yeshurun and Carrasco, 2008; Yeshurun et al., 2008a; Barbot and Carrasco, 2017; Jigo and Carrasco, 2018). (ii) Narrow high-SF attentional gain drives involuntary attention effects (Yeshurun and Carrasco, 1998, 2000; Talgar and Carrasco, 2002; Carrasco et al., 2006; Yeshurun and Carrasco, 2008; Yeshurun et al., 2008a; Jigo and Carrasco, 2018). (iii) Broad attentional gain modulation across spatial frequency drives voluntary attention effects (Yeshurun et al., 2008a; Barbot and Carrasco, 2017; Jigo and Carrasco, 2018).

The fact that our model operates on arbitrary grayscale images and can be applied to any psychophysical detection or discrimination task across the visual field, enabled us to predict human contrast sensitivity across SF and eccentricity. Importantly, our model allows for predictions for how involuntary or voluntary attention would adjust performance and ultimately visual perception across the visual field.

Extension to a normalization model of attention

Our model adopted the same algorithm specified by the Reynolds-Heeger normalization model of attention (NMA) (Reynolds and Heeger, 2009)—attentional gain modulates the stimulus drive before normalization—but we required distinct profiles across SF to capture involuntary and voluntary attention effects. The NMA has reconciled the effects of each attention type on physiological measures of contrast sensitivity (Reynolds and Heeger, 2009) and its predictions have been empirically confirmed with psychophysical experiments (Herrmann et al., 2010). The NMA incorporates a variable attention field size to account for attentional effects. However, manipulating the size of the attention field during texture segmentation did not elicit shifts between typical effects of involuntary and voluntary attention (Yeshurun and Carrasco, 2008). By modeling an extensive set of texture segmentation experiments, we reveal a computational dissociation between these types of attention.

Our current implementation extends NMA (Reynolds and Heeger, 2009) to capture differential effects of voluntary and involuntary attention on contrast and texture sensitivity across SF and eccentricity. (i) Multiplicative gains (**Figure 2a, green**) implement neural (DeValois and DeValois, 1990) and psychophysical (Graham, 1989) bandpass SF selectivity. (ii) Spatial summation (**Figure 2a, purple**) emphasizes textural contours (Bergen and Adelson, 1988; Landy and Graham, 2004; Landy, 2013). (iii) Different attentional gain profiles (Jigo and Carrasco, 2020) scale responses to SF prior to normalization (Reynolds and Heeger, 2009), which adjusts the balance between fine and coarse-scale perception.

Divisive normalization is necessary to explain the central performance drop

Contextual modulation through normalization has been proposed as a critical component for models of texture perception (Landy and Graham, 2004; Landy, 2013), but has not been implemented to explain the CPD. Experimental studies demonstrate that the CPD is attenuated when the pool of SFs contributing to normalization is restricted**.** When distracting high SFs are filtered out of the texture, the CPD is eliminated (Morikawa, 2000). When visual sensitivity to high SFs is reduced, via selective adaptation, the CPD is diminished (Carrasco et al., 2006; Barbot and Carrasco, 2017).

The present model demonstrates that the CPD depends on contextual modulation, whereas previous models (Kehrer, 1997; Kehrer and Meinecke, 2003) relate the phenomenon solely to an increase in the pooling area of receptive fields with eccentricity. Our model directly linked the summation area of receptive fields to their SF tuning. Consequently, the dominant summation area increased with eccentricity as SF preferences lowered, consistent with an increase in receptive field size with eccentricity (Anton-Erxleben and Carrasco, 2013). Crucially, without contextual modulation, the model did not reliably capture performance.

Involuntary and voluntary attention differentially modulate perception

Texture segmentation studies have provided clear evidence for a dissociation between involuntary and voluntary attention (Yeshurun and Carrasco, 1998, 2000; Talgar and Carrasco, 2002; Carrasco et al., 2006; Yeshurun and Carrasco, 2008; Yeshurun et al., 2008a; Barbot and Carrasco, 2017; Jigo and Carrasco, 2018) (reviews (Carrasco, 2011; Anton-Erxleben and Carrasco, 2013)). Our model demonstrates that their differential perceptual effects are driven by distinct operating regimes across SF. Importantly, the different attentional gain profiles were empirically motivated by independent psychophysical measurements of contrast sensitivity with grating stimuli (Jigo and Carrasco, 2020), bolstering the model's plausibility.

Involuntary and voluntary attention differentially modulate contrast sensitivity across eccentricity (Jigo and Carrasco, 2020) and differentially adjust spatial resolution (Carrasco et al., 2006; Barbot and Carrasco, 2017) (reviews (Carrasco, 2011; Anton-Erxleben and Carrasco, 2013)). Although both types of attention similarly improve overall sensitivity (Pestilli et al., 2009) (review (Carrasco, 2011)), involuntary attention preferentially enhances a narrow range of SFs whereas voluntary attention uniformly enhances SF. Through these distinct modulation profiles, we explicitly link changes in SF sensitivity to modifications of spatial resolution and provide a quantitative explanation for how attention alters spatial resolution during texture segmentation tasks (Yeshurun and Carrasco, 1998, 2000; Talgar and Carrasco, 2002; Carrasco et al., 2006; Yeshurun and Carrasco, 2008; Yeshurun et al., 2008a; Barbot and Carrasco, 2017; Jigo and Carrasco, 2018) (reviews (Carrasco, 2011; Anton-Erxleben and Carrasco, 2013)).

Preferential high-SF enhancement generates an increase in spatial resolution, characteristic of involuntary attention (Yeshurun and Carrasco, 1998, 2000; Talgar and Carrasco, 2002; Carrasco et al., 2006; Yeshurun and Carrasco, 2008; Yeshurun et al., 2008a; Anton-Erxleben and Carrasco, 2013; Jigo and Carrasco, 2018). The model provides converging evidence that the SF tuning of involuntary attention is linked to, and operates above, intrinsic (i.e., baseline) SF tuning across eccentricity (Jigo and Carrasco, 2020) (**Supplementary Figure 3**). The fact that our model was constrained by texture stimuli yet produced a preference for higher SFs similar to measurements with gratings, highlights the hallmark inflexibility of involuntary attention (Yeshurun and Carrasco, 1998, 2000; Talgar and Carrasco, 2002; Carrasco et al., 2006; Yeshurun and Carrasco, 2008; Yeshurun et al., 2008a; Giordano et al., 2009; Jigo and Carrasco, 2018) (review (Carrasco, 2011)). Thus, in addition to involuntary attention effects being invariant to cue predictability (Giordano et al., 2009) and sometimes detrimental to perception (Yeshurun and Carrasco, 1998, 2000; Talgar and Carrasco, 2002; Carrasco et al., 2006; Yeshurun and Carrasco, 2008; Yeshurun et al., 2008a; Jigo and Carrasco, 2018), our modeling suggests that involuntary attentional modulation is also invariant to the type of stimulus being attended.

Voluntary attention, in contrast, optimizes spatial resolution for fine or coarse-scale textures (Yeshurun et al., 2008a; Anton-Erxleben and Carrasco, 2013; Barbot and Carrasco, 2017; Jigo and Carrasco, 2018) through a broad SF gain modulation. Owing to its plateau, the broad profile overlapped and equivalently enhanced the SFs of both fine- and coarse-scale targets. The model revealed that attentional gain during texture segmentation was centered on SFs lower than the baseline tuning preferences, particularly near the fovea (**Supplementary Figure 4**). This differs

from measurements of voluntary attentional effects made with gratings, for which the attentional gain modulates SFs lower and higher than the baseline tuning preference at each eccentricity (Jigo and Carrasco, 2020). This finding also seemingly conflicts with the explanation that voluntary attention suppresses sensitivity to high SFs at central locations instead (Barbot and Carrasco, 2017). However, suppressed sensitivity to high SFs at central locations would decrease cross-frequency suppression (Petrov et al., 2005) and result in an effective dominance of lower SFs. In any case, the fact that the model revealed attentional gain for textures that differed from those measured with gratings provides converging evidence regarding the flexibility of the voluntary attention system. Thus, in addition to voluntary attention effects conforming to cue predictability (Giordano et al., 2009) and adapting to task demands (Yeshurun et al., 2008a; Giordano et al., 2009; Barbot and Carrasco, 2017; Jigo and Carrasco, 2018), our model suggests that the center frequency of voluntary attentional modulation may vary with the attended stimulus.

Neural mechanisms

Our model simulates bottom-up responses of phase-invariant complex cells in V1 (Adelson and Bergen, 1985). Contrast sensitivity of the model's receptive fields decline and their tuning shifts toward coarser spatial scales with eccentricity, following the spatial characteristics of the cone mosaic, retinal and V1 cells (Graham, 1989; DeValois and DeValois, 1990). The model generates texture selectivity through pooling bottom-up inputs, following the notion that texture selectivity emerges gradually along the visual hierarchy (El-Shamayleh and Movshon, 2011). Consistent with this idea, evidence from human functional magnetic resonance imaging (fMRI) exhibit a steady increase in selectivity for texture-defined patterns beyond V1 (Larsson et al., 2006; Hallum et al., 2011).

In the model, attention alters sensitivity to contrast and texture differently across SF. However, how the brain mediates these effects has not been extensively investigated for involuntary attention. Single-cell electrophysiology has demonstrated that involuntary attention modulates neural activity in V1 (Wang et al., 2015) and middle temporal area (Busse et al., 2008). Likewise, fMRI experiments have revealed modulations of neural activity in visuo-occipital areas and beyond (Liu et al., 2005; Dugué et al., 2018, 2020). Early visual cortex plays a key role in the generation of involuntary attention effects, as neurostimulation via transcranial magnetic stimulation eliminates the perceptual consequences of involuntary attention (Fernández and Carrasco, 2020). In contrast to involuntary attention, it is well-established that voluntary attention modulates cortical responses (Reynolds and Chelazzi, 2004; Womelsdorf et al., 2006; Womelsdorf et al., 2008; Fischer and Whitney, 2009; Carrasco, 2011; Anton-Erxleben and Carrasco, 2013; Klein et al., 2014; Dugué et

al., 2018, 2020). Our findings are consistent with human fMRI measurements demonstrating that involuntary and voluntary attention differentially modulate neural responses (Dugué et al., 2018, 2020). However, future measurements of attention should probe SF to assess the neural implementation of the distinct computations we describe here for each type of attention.

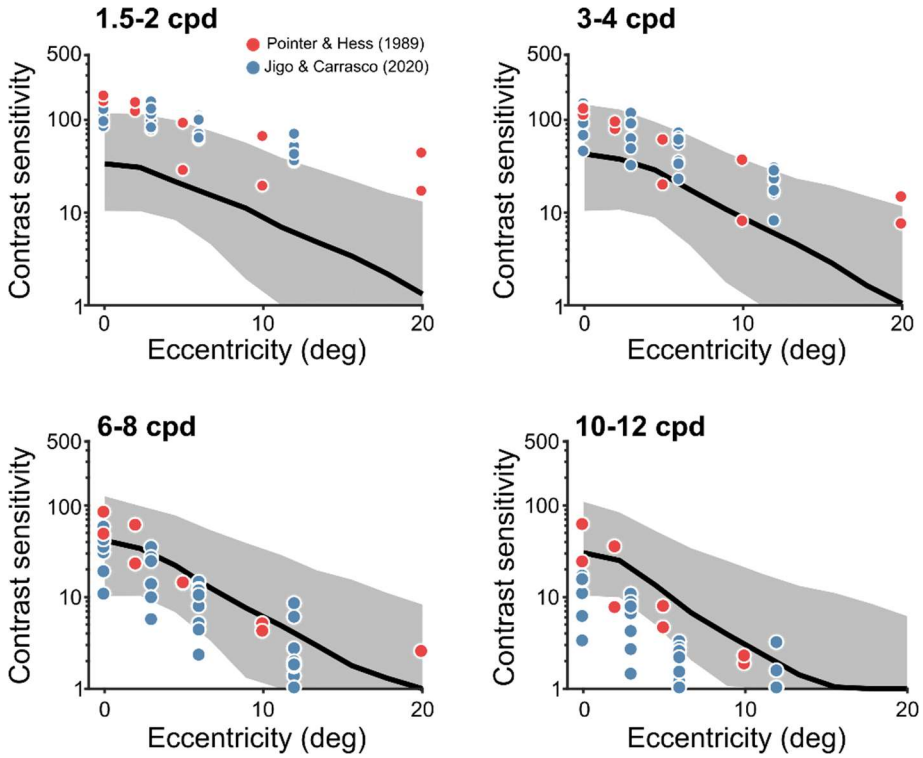
Relation to models of spatial resolution

Few computational models have implemented ways in which attention alters spatial resolution. Some have proposed that attention modifies how finely a spatial region is analyzed via an attention field that adjusts the spatial profile of receptive fields (Womelsdorf et al., 2008) or by attracting receptive fields toward and contracting them around the attended location (Baruch and Yeshurun, 2014). Other models suggest an attentional prioritization that selectively tunes responses for a given spatial location and attenuates responses to surrounding regions (Tsotsos et al., 1995). However, these models neither account for differences across eccentricity nor explain attentional shifts in visual sensitivity toward fine or coarse-scale details. Critically, these models do not distinguish between involuntary and voluntary attention effects. The present model advances our understanding of how involuntary and voluntary attentional selection differentially modulate perception across eccentricity and attributes these differences to operations on basic dimensions of early visual processing.

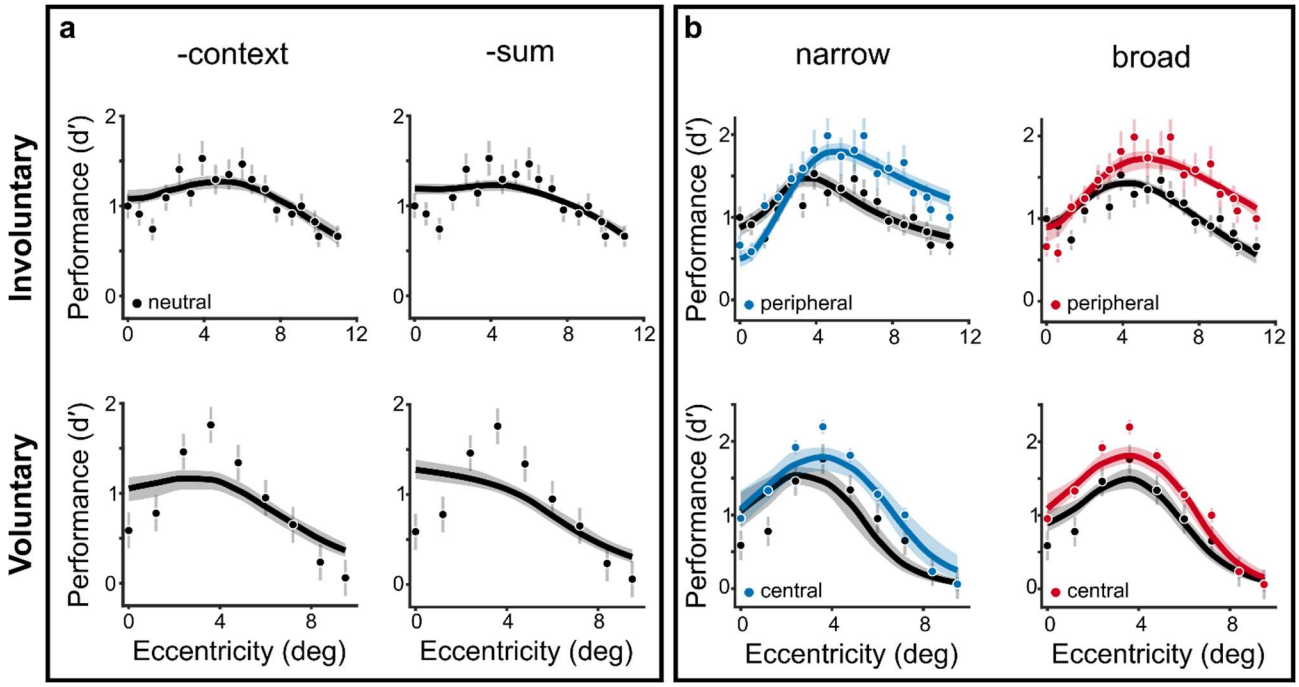
Conclusion

We quantitatively explain key signatures of texture segmentation (Kehrer, 1989; Yeshurun and Carrasco, 1998; Morikawa, 2000; Yeshurun and Carrasco, 2000; Talgar and Carrasco, 2002; Carrasco et al., 2006; Yeshurun and Carrasco, 2008; Yeshurun et al., 2008a; Barbot and Carrasco, 2017; Jigo and Carrasco, 2018) and characterize the contributions of attention to a process historically considered ‘preattentive’ (Victor et al., 2017). Moreover, we reveal the cortical computations that underlie both involuntary and voluntary attentional modulation of spatial resolution (Carrasco, 2011; Anton-Erxleben and Carrasco, 2013). Attention scales sensitivity to high and/or low SFs, adjusting the balance between fine and coarse-scale spatial resolution. Involuntary attention preferentially enhances fine details whereas voluntary attention uniformly enhances fine and coarse details to optimize task performance. Because the model distinguishes between involuntary and voluntary attention, varies with stimulus eccentricity, flexibly implements psychophysical tasks and operates on arbitrary grayscale images, it provides a general-purpose tool for assessing theories of vision and attention across the visual field.

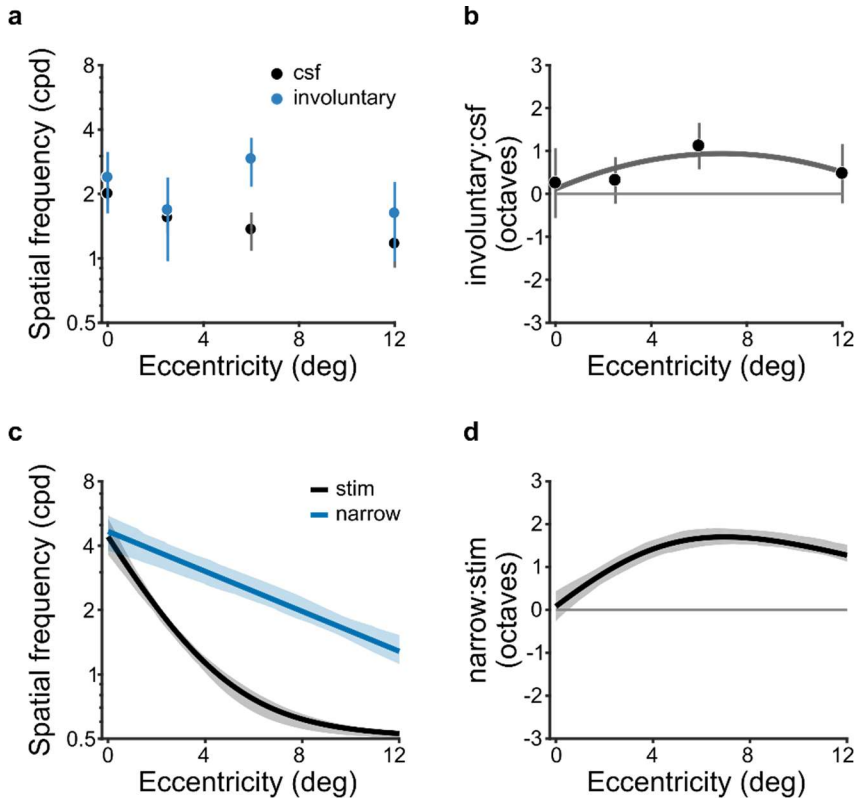
SUPPLEMENTARY FIGURES



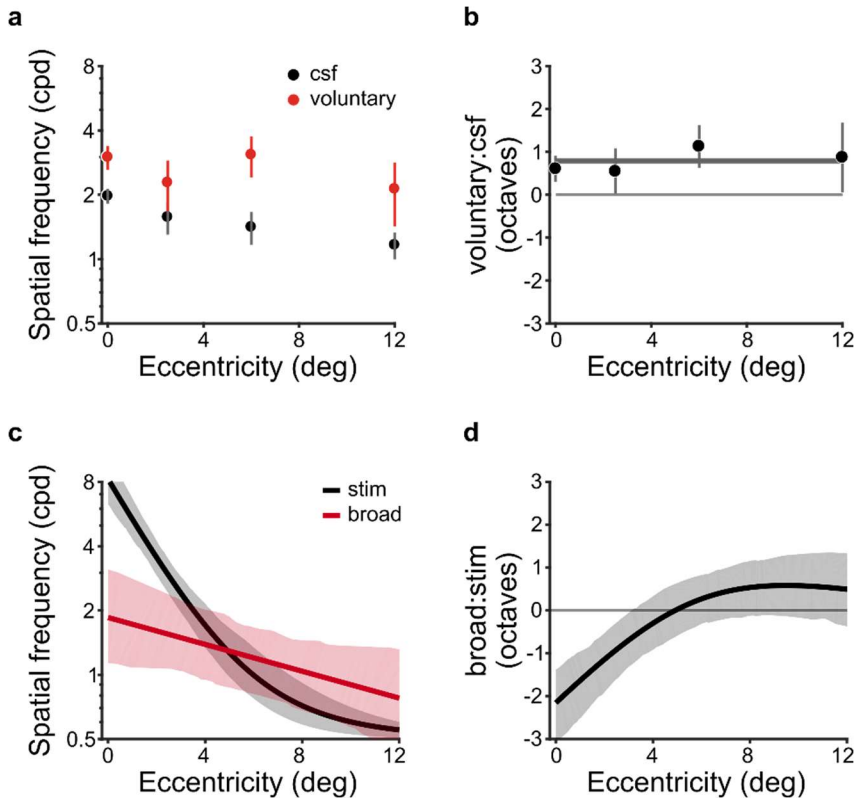
Supplementary Figure 1. Contrast sensitivity across eccentricity. The black line indicates the median and shaded regions denote 68% confidence intervals of the bootstrapped distribution of the model's contrast sensitivity. Using parameters fit to texture segmentation performance and grating images that matched each of the two experiments displayed above, the model predicted contrast sensitivity functions (see **Online methods, Predicted contrast sensitivity functions**). Model parameters were randomly selected, with replacement, from the bootstrapped distribution of parameters that simultaneously fit the central performance drop of all ten texture segmentation experiments. Dots indicate contrast sensitivity, measured psychophysically, for individual observers.



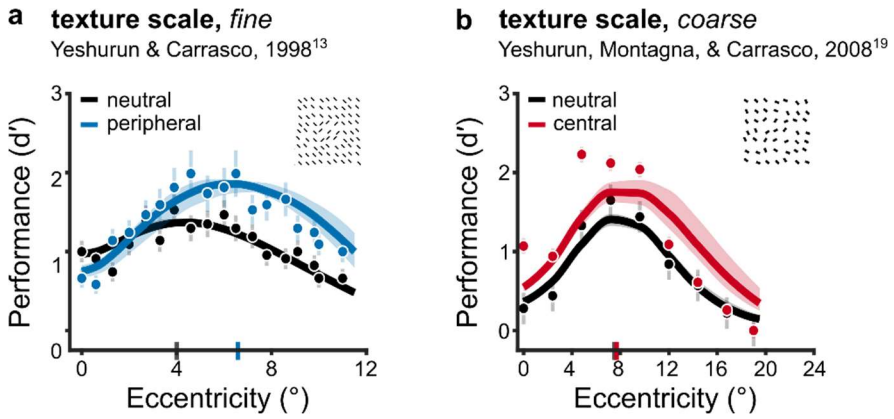
Supplementary Figure 2. Model variants fit to behavioral data from (*top row*) experiment 1 in Yeshurun & Carrasco, 1998(Yeshurun and Carrasco, 1998) and (*bottom row*) experiment 3 in Yeshurun, Montagna & Carrasco, 2008(Yeshurun et al., 2008a). Each column depicts different model variants. (**a**) The first two columns depict fits to the neutral condition. '-context' denotes the variant lacking contextual modulation and '-sum' lacks spatial summation. (**b**) Fits to attentional effects are color-coded by the attentional gain profile used to fit the data. Fits with the narrow profile are depicted in blue and fits with the broad profile are depicted in red. Dots and error bars denote group-average performance and ±1 SEM. The solid lines depict the median and shaded regions depict 68% confidence intervals of the bootstrapped distribution of model fits. Displayed are representative experimental datasets; fits were performed simultaneously with other experiments as described in **Online methods, Optimization strategy** and **Online methods, Alternative models**.



Supplementary Figure 3. Spatial frequency operating range of involuntary attention. (**a**) Peak spatial frequency of baseline contrast sensitivity (CSF) and involuntary attentional modulation from Jigo & Carrasco (2020)(Jigo and Carrasco, 2020). Estimates were based on human contrast sensitivity, measured psychophysically with narrowband gratings. (**b**) Ratio (in octaves) of attentional and baseline peak spatial frequency tuning across eccentricity. Positive values denote an attentional preference for spatial frequencies higher than baseline. The solid line depicts the best-fitting second-order polynomial (i.e., parabola). Polynomial order was determined using leave-one-subject-out cross-validation (**Online methods, Cross-validation procedure**). Dots in *a* and *b* depict group-average and error bars depict ±1 SEM. (**c**) Peak spatial frequency of the stimulus drive (stim) and the narrow attention gain profile (narrow). Estimates were derived from model fits to texture segmentation performance across all six involuntary attention experiments. (**d**) Ratio of the preferred spatial frequency for the stimulus drive and attentional gain. Solid lines indicate the median and shaded area denotes 68% confidence interval of bootstrapped distribution in *c* and *d*.



Supplementary Figure 4. Spatial frequency operating range of voluntary attention. (**a**) The peak spatial frequency of baseline contrast sensitivity (CSF) and the center frequency of broad voluntary attentional modulation from Jigo & Carrasco (2020)(Jigo and Carrasco, 2020). Estimates were based on human contrast sensitivity, measured psychophysically with narrowband gratings. (**b**) Ratio (in octaves) of attentional and baseline and spatial frequency preferences across eccentricity. Negative values denote an attentional preference for spatial frequencies lower than baseline. The solid line depicts the best-fitting zero-order polynomial (i.e., constant). Polynomial order was determined using leave-one-subject-out cross-validation (**Online methods, Cross-validation procedure**). Dots in *a* and *b* depict group-average and error bars depict ±1 SEM (**c**) The center spatial frequency of the stimulus drive (stim) and the broad attentional gain profile (broad). Estimates were derived from model fits to texture segmentation performance across all six voluntary attention experiments. (**d**) Ratio of the preferred spatial frequency for the stimulus drive and attentional gain. Solid lines indicate the median and shaded areas denote 68% confidence intervals of the bootstrapped distribution in *c* and *d*.



Supplementary Figure 5. Model fits to jittered texture stimuli.
(**a**) Predicted performance for experiment 1 in Yeshurun & Carrasco (1998)(Yeshurun and Carrasco, 1998) using texture stimuli generated with line elements spatially jittered within the stimulus parameters of the experiment (**Online methods, Stimulus generation**). An example jittered stimulus is shown in the top-right. Solid lines indicate the median and shaded regions depict 68% confidence intervals of the bootstrap distributions of model predictions. Gray and blue ticks on x-axis indicate peak of performance in the neutral and peripheral cueing condition, respectively. To generate the bootstrap distributions, model parameters were fixed to those that best-fit all involuntary experiments simultaneously then new jittered texture stimuli were input to the model on each iteration.
(**b**) Predicted performance for experiment 4 in Yeshurun, Montagna & Carrasco (2008)(Yeshurun et al., 2008a) using texture stimuli generated with line elements whose orientation and spatial location were randomly jittered within the parameters of the experiment (**Online methods, Stimulus generation**). Solid lines indicate the median and shaded regions depict 68% confidence intervals of bootstrap distributions of model predictions. Gray and red ticks on x-axis indicate peak of performance in the neutral and central cueing condition, respectively. To generate the bootstrap distribution, model parameters were fixed to those that best-fit all voluntary experiments simultaneously and new jittered texture stimuli were input to the model on each iteration.

METHODS

Model

Our model simulates the response of a collection of receptive fields (RFs) each narrowly tuned to spatial position (*x,y*), orientation (θ) and spatial frequency (*f*). Responses varied with eccentricity (α). The population response (*R*) was determined by the stimulus drive (*E*), attentional gain (*A*), suppressive drive (S and σ), and spatial summation (*F*), where * represents convolution:

$$R(f,\theta,x,y) = \frac{E(f,\theta,x,y)A(f,\alpha)}{\sigma^2(\alpha) + S(f,\theta,x,y)} * F(f) \quad (1)$$

All model parameters are given in **Table 1**.

Table 1. Model parameters. Lighter gray shading indicates that the parameter was fixed to a set value and not optimized during fitting.

Parameter	Description
Stimulus drive	
t_T	peak spatial frequency ($\log_2$-cpd)
t_{min}	minimum preferred spatial frequency (cpd); *fixed at 0.5*
m_T	spatial frequency tuning slope (octaves/°)
b_T	spatial frequency tuning bandwidth (octaves)
Contrast gain	
g_σ	gain at fovea
m_σ	slope along eccentricity
Normalization pool	
δ_f	Spatial frequency pool bandwidth (octaves); *fixed at 1*
δ_θ	Orientation pool bandwidth (°); *fixed at 180*
δ_{pos}	Spatial pool width (°); *fixed at* $\frac{2}{f}$

Spatial summation	
δ_{pos}	Summation pool width (°); *fixed at* $\frac{2}{f}$
Attentional gain profile	
a_N or a_B	attentional spatial frequency tuning at fovea ($\log_2$-cpd)
m_N or m_B	slope of frequency tuning along eccentricity (octaves/°)
γ_N or γ_B	amplitude
b_N or b_B	spatial frequency bandwidth (octaves)
b_{pos}	spatial spread (°); *fixed at FWHM of 4*

Stimulus drive

The stimulus drive characterizes responses of linear RFs in the absence of suppression, attention and spatial summation. Stimulus drive responses were determined using the steerable pyramid(Simoncelli et al., 1992). The pyramid decomposed the stimulus image into several spatial frequency and orientation subbands. Spatial frequency and orientation selectivity were defined by weighted sums of the image (i.e., linear filters). The weights were parameterized by raised-cosine functions that evenly tiled spatial frequencies, orientations and positions (eccentricities).

The number of spatial frequency and orientation subbands are parameters that can be flexibly chosen. We used a set of 30 subbands that comprised five spatial frequency bands and six orientation bands. The number of spatial frequency subbands is determined by the size of the stimulus image and the subband bandwidth. In our simulations, we used images that were 160 x 160 pixels (see **Stimulus generation**), and a full-width at half-maximum (FWHM) bandwidth of 1 octave, which allowed for five different spatial frequency subbands. The chosen bandwidth is comparable to empirical tuning curves measured in primate electrophysiological recordings(De Valois et al., 1982) and human psychophysical(Graham, 1989) measurements. The FWHM orientation bandwidth (60°) of each subband is comparable to empirical tuning curves measured in primate electrophysiological recordings(Ringach et al., 2002). Using narrower (30°) or wider (90°) bandwidths yielded similar results supporting the same conclusions.

For each spatial frequency and orientation subband, the pyramid includes RFs in quadrature phase. We computed a 'contrast energy' response(Adelson and Bergen, 1985), which is the sum of squared responses across phase. This response depends on the local spectral energy at each position in the image for a given spatial frequency and orientation subband and is fundamental to models of texture perception(Bergen and Adelson, 1988; Landy and Graham, 2004; Landy, 2013). The contrast energy response to spatial frequency (*f*) and orientation (θ) across space (*x*,y) is denoted as $C(f,\theta,x,y)$.

Spatial frequency gain. Human(Graham, 1989; Pointer and Hess, 1989; Jigo and Carrasco, 2020) and non-human primate(DeValois and DeValois, 1990) contrast sensitivity is narrowly tuned to spatial frequency across eccentricity. Spatial frequency tuning shifts from high to low frequencies with eccentricity. To model this behavior, contrast energy was multiplied point-by-point by a spatial frequency gain function, *T*, defined by a log-parabola(Watson and Solomon, 1997; Watson and Ahumada, 2005; Lesmes, 2010):

$$T(f,\alpha) = \exp(-[\frac{\log_2(\frac{f}{\lambda_T(\alpha)})}{b_T}]^2) \tag{2}$$

where α denotes the eccentricity of a RF and b_T determines the function's spatial frequency FWHM bandwidth. The preferred spatial frequency (λ_T) at a given eccentricity was given by:

$$\lambda_T(\alpha) = 2^{t_T - m_T\alpha} + t_{min} \tag{3}$$

Motivated by psychophysical measurements(Pointer and Hess, 1989), spatial frequency preferences converged onto a single value in the far periphery, t_{min}. As a result, the preferred spatial frequency at the fovea was given by $2^{t_T} + t_{min}$ and progressively shifted towards t_{min} at the rate m_T (octaves/degree). Whereas t_T was free to vary during simulations (see **Table 1-4**), t_{min} was fixed at 0.5 cpd because texture stimuli produced minimal contrast energy below that spatial frequency subband. Allowing t_{min} to vary yielded similar results supporting the same conclusions.

In sum, the stimulus drive (*E*) characterizes the contrast energy response of a collection of RFs whose spatial frequency preferences vary with eccentricity, computed as:

$$E(f,\theta,x,y) = T(f,\alpha)C(f,\theta,x,y) \tag{4}$$

Attentional gain

Attention was implemented as an attentional gain field, *A*, that multiplied the stimulus drive point-by-point as in the Reynolds-Heeger normalization model of attention(Reynolds and Heeger, 2009). Attentional gain was uniform across orientation. Across spatial frequency and spatial position, the distribution of gain was characterized by cosine window functions, *w*:

$$w(z;\mu,b) = \begin{cases} 0.5 + 0.5\cos(\frac{\pi[z-\mu]}{b}) & \\ 0 & \mu - b < z > \mu + b \end{cases} \quad (5)$$

where μ defined its center and *b* defined its FWHM. The units of μ and *z* depended on the dimension: for spatial frequency each variable was in units of $\log_2$-transformed cycles per degree, and for spatial position they were in units of degrees of visual angle. Thus, the window was defined on a logarithmic axis for spatial frequency but on a linear axis for position. Spatial frequency and spatial position functions were multiplied, point-by-point, to characterize the full distribution of attentional gain.

Spatial spread. Attentional gain was centered on the target location. In our simulations, the target was always placed along the horizontal meridian at eccentricity α_{targ} (see **Stimulus generation**). The spread of attention was determined by multiplying two separate cosine functions (*w*, equation 5): one that varied as a function of *x* and another as a function of *y*, each with an identical width b_{pos}. The width did not vary across eccentricity. This produced a two-dimensional spatial window A_{pos}:

$$A_{pos}(x,y) = w(x;\alpha_{targ},b_{pos})w(y;0,b_{pos}) \quad (6)$$

The precise spatial spread of attention is controversial(Eriksen and Hoffman, 1972; Tkacz-Domb and Yeshurun, 2018; Yeshurun, 2019) and can change based on task demands(LaBerge, 1983; Stigchel et al., 2009; Herrmann et al., 2010). Critically, the spread of attention has not been explicitly manipulated during texture segmentation tasks by varying the target's spatial uncertainty. Such a protocol has been used to test predictions of the NMA and, through functional magnetic resonance imaging, has been demonstrated to adjust the size of the attention field(Herrmann et al., 2010). Instead, a previous study(Yeshurun and Carrasco, 2008) measured involuntary attention effects while manipulating the size of a peripheral pre-cue. The authors found that involuntary attention altered performance as long as the cue was the same or smaller than the target size. In our simulations, the spread of attention was fixed at a FWHM of 4° (**Table 1**). This size was chosen

because it encompassed the largest target size used to constrain model parameters (**Table 5**). As a result, the spatial extent of attention was identical across eccentricity and experiments. Similar results were observed when the spread was fixed at 2° and 3°, or a free parameter fit to the data.

Spatial frequency profile. Motivated by empirical findings(Jigo and Carrasco, 2020), we implemented two attentional gain profiles: narrow and broad.

Narrow profile. In the narrow model (A_N), attentional gain was bandpass across spatial frequency. Attentional gain peaked at a given spatial frequency, λ_N, and fell gradually toward neighboring frequencies within its bandwidth, b_N. The window was characterized by a cosine function as defined in equation (5):

$$A_N(f, \alpha) = w(f; \lambda_N(\alpha), b_N) \tag{7}$$

The center frequency of attentional gain profiles (λ_N for narrow, λ_B for broad) varied with eccentricity:

$$\lambda_N(\alpha) = 2^{a_N - m_N \alpha} \tag{8}$$

where a_N (or a_B for the broad profile) defined the center frequency at the fovea, which gradually changed with eccentricity at the rate m_N (m_B for broad).

Broad profile. In the broad profile (A_B), attentional gain was broadband across spatial frequency. Attentional gain was uniformly large across a range of spatial frequencies, characterized by the sum of three overlapping cosine functions (w, equation 5):

$$A_B(f, \alpha) = w_1 + w_2 + w_3 \tag{9}$$

where $w_1 = w(f; \lambda_B(\alpha), b_B)$, $w_2 = w(f; \lambda_B(\alpha) - b_B, b_B)$, and $w_3 = w(f; \lambda_B(\alpha) + b_B, b_B)$. The bandwidth of each function was given by b_B. Relative to the center frequency, λ_B, the adjacent functions were centered $\pm b_B$ apart, which ensured that their sum yielded a plateau that spanned $\frac{2b_B}{3}$ octaves.

In sum, attentional gain multiplicatively scaled the stimulus drive uniformly across orientation, but differently across spatial frequency and eccentricity. Overall, attentional gain was given by:

$$A(f,\alpha) = \gamma_B A_{pos} A_B \tag{10}$$

where A_{pos} and A_B (or A_N) were four-dimensional matrices that characterized the distribution of attentional gain on spatial position, spatial frequency and orientation. The strength of attention was determined by γ_B (or γ_N). To simulate the neutral cueing condition, attentional gain was set to 1 across spatial frequency, orientation and spatial position, which removed modulations by attention.

Suppressive drive

The suppressive drive comprised contextual modulation, computed through pooling the attention-scaled stimulus drive(Reynolds and Heeger, 2009) across nearby spatial positions, all orientations and across neighboring spatial frequencies. This pooling procedure implemented lateral interactions between RFs that differed in their tuning profiles. Suppression magnitude was computed via convolution(Reynolds and Heeger, 2009). Convolution kernels were cosine window functions (*w*, equation 5), defined across spatial frequency, orientation and spatial position.

The bandwidth of the spatial frequency kernel, δ_f, equaled 1 octave:

$$K_f = \begin{cases} 1 & f_i - \delta_f \leq f_i \leq f_i + \delta_f \\ 0 & otherwise \end{cases} \tag{11}$$

where f_i denotes the center frequency of a steerable pyramid subband. This kernel summed the squared outputs of the steerable pyramid's subbands within and ±1 octave around each spatial frequency band. Given the parameters for the steerable pyramid's subband bandwidth (1 octave, see **Stimulus drive**), this sum yielded the response of a single large filter with a FWHM of 3 octaves. Consequently, each RF was suppressed by the response to spatial frequencies within and around its tuning bandwidth.

The bandwidth of the orientation kernel, δ_θ, equaled 180°, which encompassed all orientation subbands:

$$K_\theta = \begin{cases} 1 & \theta_i - \delta_\theta \leq \theta_i \leq \theta_i + \delta_\theta \\ 0 & otherwise \end{cases} \tag{12}$$

where θ_i denotes the center orientation of a steerable pyramid subband. This kernel summed the squared outputs of the steerable pyramid's subbands across orientation. This sum yielded the

response of a single large filter that spanned all orientations. As a result, each RF was suppressed by the contrast energy response to all orientations in the spatial and spatial frequency neighborhood of the RF.

Spatial position kernels were determined by multiplying two cosine windows defined in equation (5):

$$K_{pos}(x, y; f) = w(x; 0, \delta_{pos})w(y; 0, \delta_{pos}) \tag{13}$$

One window varied across x, another across y and their centers traversed across the image during convolution. The two-dimensional kernel summed to unity and therefore computed the average contrast energy within the pooled spatial area. The width of each kernel, δ_{pos}, equaled $\frac{2}{f}$ and was thus inversely proportional to subband spatial frequency *f*, which yielded two-dimensional spatial kernels, K_{pos}, which were broader for lower frequencies than for higher frequency subbands. Kernel widths did not vary with eccentricity.

Broad suppressive pools for responses to spatial frequency, orientation and spatial position are supported by physiological(Bonds, 1989; DeAngelis et al., 1992; Cavanaugh et al., 2002; Priebe and Ferster, 2006; Brouwer and Heeger, 2011) and psychophysical(Sagi and Hochstein, 1985; Meese and Hess, 2004; Petrov et al., 2005; Meese and Holmes, 2007) findings and models of visual processing(Heeger, 1992).

Overall, contextual modulation was characterized via separable convolution across the attention-scaled stimulus drive:

$$S(f, \theta, x, y) = K_f * (K_\theta * (K_{pos} * [E(f, \theta, x, y)A(f, \alpha)])) \tag{14}$$

where * denotes convolution of the suppression kernels, *K*. The magnitude of suppression was adjusted across eccentricity by adding σ^2, which controlled the level of contrast at which neural responses reached half-maximum and is referred to as contrast gain. Motivated by psychophysical measurements(Pointer and Hess, 1989; Jigo and Carrasco, 2020), contrast gain was implemented as an exponential function across eccentricity:

$$\sigma^2(\alpha) = 10^{2(g_\sigma - m_\sigma \alpha)} \tag{15}$$

where g_σ and m_σ are free parameters that determine contrast gain at the fovea and the rate at which it varies with eccentricity, respectively.

Spatial summation

Following divisive normalization, responses were weighted and summed across space and within each spatial frequency and orientation subband. Spatial summation was accomplished via convolution by two-dimensional cosine windows, *F*, computed using equation (13). Importantly, the width of each filter scaled with subband spatial frequency such that narrow (wide) regions of space were pooled for high (low) spatial frequencies(Landy and Graham, 2004). Filter widths did not vary with eccentricity. A spatial summation stage is common in models of texture perception(Landy and Graham, 2004; Landy, 2013). Motivated by empirical findings(Landy and Oruç, 2002; Westrick and Landy, 2013), pooling followed normalization and emphasized textural aspects of the stimulus image (e.g., a patch of similarly oriented lines).

Decision mechanism

We used signal detection theory to relate population responses to behavioral performance (*d′*). Following an earlier image-computable model(Teo and Heeger, 1994), the available signal *s* for an observer was computed as the Euclidian norm of the difference between target-present ($\mathbf{r_t}$) and target-absent ($\mathbf{r_n}$) neural population responses: $s = ||\boldsymbol{r_t} - \boldsymbol{r_n}||$. The available signal simulates an ideal observer performing a 2IFC discrimination task wherein the response to each stimulus is compared without bias or noise. Performance on a discrimination task is proportional to the neural responses given the assumption of additive, independent and identically distributed (IID) noise. An alternative model with Poisson noise and a maximum-likelihood decision rule would yield the same linkage between neural response and behavioral performance(Jazayeri and Movshon, 2006; Pestilli et al., 2009). Based on signal detection theory, the signal and the magnitude of noise (σ_n) were used to compute behavioral performance $d' = \frac{s}{\sigma_n}$. In our simulations, $\sigma_n = \overline{r_{neutral}} \big/ \overline{s_{neutral}}$ where $\overline{r_{neutral}}$ denotes the observed neutral performance averaged across eccentricity and $\overline{s_{neutral}}$ denotes the average signal across eccentricity for the neutral condition. This ratio scaled the model's predicted behavioral performance to match the observed data.

Model fitting

Models were optimized by minimizing the residual sum of squared error (SSE) between model and behavioral *d′* using Bayesian adaptive direct search (BADS(Acerbi and Ma, 2017)). This

optimization procedure is well-suited for the number of parameters and cost function evaluation time of the model. Behavioral data were extracted from the original study publications using WebPlotDigitizer tool(Rohatgi, 2020). When applicable, performance data for a psychophysical experiment were converted from proportion correct, *p*, to *d′* with the assumption of no interval bias(Macmillan and Creelman, 2005): $d' = \sqrt{2}z(p)$ where *z* denotes the inverse normal distribution. Although performance on 2IFC tasks can exhibit biases between intervals(Yeshurun et al., 2008b), our conversion algorithm operated uniformly across eccentricity, which preserved the performance variation (i.e., the CPD) that is critical for the goals of this study.

Optimization strategy

Central performance drop

To address whether normalization is critical for eliciting the CPD, the model was first fit to performance on all neutral conditions across the ten texture segmentation experiments (103 data points). Peripheral and central cueing conditions (for involuntary and voluntary attentional conditions, respectively) were excluded to isolate the CPD. Fifteen free parameters fit all 103 data points (**Table 2**).

Ten separate free parameters independently controlled the minimum contrast gain at the fovea (g_σ; equation 15) for each of the 10 experiments. These parameters accounted for possible differences in overall contrast sensitivity that can arise from variable display properties (e.g., mean luminance) among experimental protocols (reviews(Kelly, 1977; Graham, 1989; DeValois and DeValois, 1990)).

Sensitivity to contrast and spatial frequency varies for stimuli placed at isoeccentric locations around the visual field; it is higher at the horizontal meridian and decreases gradually towards the vertical meridian(Carrasco et al., 2001; Cameron et al., 2002; Himmelberg et al., 2020; Barbot et al., 2021). Whereas 5 out of 6 involuntary attention experiments used texture targets placed on the horizontal meridian, 3 out of 4 voluntary attention experiments used targets presented along the diagonal meridians (**Table 5**). Because spatial frequency selectivity depends on stimulus polar angle, two parameters separately determined the highest preferred spatial frequency (t_T; equation 3)—one was shared among involuntary attention experiments and the other was shared among voluntary attention experiments. Alternatively, we could have fit separate parameters for horizontal, vertical and diagonal meridians. However, this approach would have added a third free parameter, reducing the parsimony of the model. The configuration we used (i.e., separate parameters for

involuntary and voluntary attention experiments; two parameters total) yielded reasonably good fits.

The remaining three parameters were shared among the neutral conditions for all experiments. Each controlled the bandwidth (b_T; equation 2) of the tuning function T, the gradual shift toward lower frequencies with eccentricity (m_T; equation 3) and the increase in contrast gain across eccentricity (m_σ; equation 15).

Table 2. Free parameters for the fits to the neutral condition of all ten psychophysical experiments. Shading denotes voluntary attention experiments. The mapping between the experiment numbers (1-10) and the respective references is given below the table; italicized text describe the manipulation conducted in a given experiment. Bold values indicate the median and values within square brackets denote the 95% CI of the bootstrapped distribution of fitted parameters. min. = minimum; SF = spatial frequency; bw = bandwidth.

<table>
<tr><th colspan="2" rowspan="3">parameter
description</th><th colspan="2">Contrast gain</th><th colspan="3">Stimulus drive</th></tr>
<tr><th>g_σ</th><th>m_σ</th><th>t_T</th><th>m_T</th><th>b_T</th></tr>
<tr><th>min.</th><th>slope</th><th>SF peak</th><th>SF slope</th><th>SF bw</th></tr>
<tr><td rowspan="10">Experiment</td><td>1</td><td>2.3
[2.2 2.5]</td><td rowspan="10">-0.09
[-0.09 -0.08]</td><td rowspan="6">1.9
[1.5 2.4]</td><td rowspan="10">-0.8
[-0.9 -0.6]</td><td rowspan="10">1.7
[1.6 2.0]</td></tr>
<tr><td>2</td><td>2.3
[2.1 2.5]</td></tr>
<tr><td>3</td><td>2.7
[2.5 2.8]</td></tr>
<tr><td>4</td><td>2.7
[2.6 2.8]</td></tr>
<tr><td>5</td><td>2.2
[2.0 2.5]</td></tr>
<tr><td>6</td><td>2.7
[2.4 2.8]</td></tr>
<tr><td>7</td><td>1.9
[1.5 2.3]</td><td rowspan="4">2.8
[2.4 3.1]</td></tr>
<tr><td>8</td><td>1.8
[1.7 2.0]</td></tr>
<tr><td>9</td><td>2.6
[2.5 2.8]</td></tr>
<tr><td>10</td><td>2.7
[2.6 2.8]</td></tr>
</table>

1. Yeshurun & Carrasco (1998)(Yeshurun and Carrasco, 1998). *Fine-scale texture; experiment 1.*
2. Yeshurun & Carrasco (1998)(Yeshurun and Carrasco, 1998). *Coarse-scale texture; experiment 2.*
3. Talgar & Carrasco (2002)(Talgar and Carrasco, 2002). *Target meridian: lower vertical.*
4. Carrasco, Loula & Ho (2006)(Carrasco et al., 2006). *Orientation discrimination: baseline adaptation.*
5. Yeshurun & Carrasco (2008)(Yeshurun and Carrasco, 2008). *Attentional cue size: cue size 1.*

6. Yeshurun, Montagna & Carrasco (2008)(Yeshurun et al., 2008a). *Target meridian: horizontal; experiment 2*.
7. Yeshurun, Montagna & Carrasco (2008) (Yeshurun et al., 2008a). *Target meridian: horizontal; experiment 1*
8. Yeshurun, Montagna & Carrasco (2008)(Yeshurun et al., 2008a). *Fine-scale texture; experiment 3.*
9. Yeshurun, Montagna & Carrasco (2008)(Yeshurun et al., 2008a). *Coarse-scale texture; experiment 4*.
10. Barbot & Carrasco (2017)(Barbot and Carrasco, 2017). *Target meridian: intercardinal; baseline adaptation.*

Attentional modulation

To assess whether narrow or broad attentional gain profiles generated the effects of attention, the model was fit separately to involuntary and voluntary attention experiments. Importantly, a pair of cueing conditions—neutral and valid—were tested in each experiment, as the neutral condition provided a baseline from which the effects of attention were measured. We preserved the link between the pair by jointly fitting the model to neutral and valid conditions of each experiment. We refer to the valid condition for involuntary attention as “peripheral” and as “central” for voluntary attention to match the different cues used in each set of experiments.

Involuntary attention. All six involuntary attention experiments were fit simultaneously (146 data points) with fourteen free parameters (**Table 3**). Minimum contrast gain at the fovea (g_σ; equation 15) was determined independently for each of six experiments, yielding six free parameters. The remaining eight parameters were shared among all experiments. Four determined the stimulus drive: its bandwidth (b_T; equation 2), the highest preferred spatial frequency at the fovea (t_T; equation 3), the shift to lower spatial frequencies with eccentricity (m_T; equation 3) and the slope of contrast gain across eccentricity (m_σ; equation 15). The remaining four controlled the narrow attentional gain profile, specifically its bandwidth (b_N; equation 7), center frequency (λ_N; equation 8), the shift to lower frequency preferences with eccentricity (m_N; equation 8), and its amplitude (γ_N; equation 10).

Table 3. Free parameters for the fits to the neutral and peripheral cueing conditions of the six involuntary attention experiments. Bold values indicate the median and values in square brackets depict the 95% CI of the bootstrapped distribution of fitted parameters. min. = minimum; SF = spatial frequency; bw = bandwidth; amp. = amplitude.

		Contrast gain		**Stimulus drive**			**Narrow gain profile**			
parameter		g_σ	m_σ	t_T	m_T	b_T	a_N	m_N	b_N	γ_N
description		*min.*	*slope*	*SF peak*	*SF slope*	*SF bw*	*SF peak*	*SF slope*	*SF bw*	*amp.*
Experiment	1	**1.9** [1.7 2.2]	**-0.05** [-0.07 -0.04]	**2.1** [1.8 2.8]	**-0.7** [-0.9 -0.5]	**1.6** [1.4 2.1]	**2.3** [1.8 3.0]	**-0.2** [-0.3 -0.1]	**2.9** [1.5 4.8]	**4.3** [3.0 7.8]
	2	**1.7** [1.5 2.1]								
	3	**2.6** [2.2 2.8]								
	4	**2.1** [2.8 2.5]								
	5	**1.5** [1.5 2.0]								
	6	**2.0** [1.7 2.7]								

Voluntary attention. All four voluntary attention experiments were fit simultaneously (60 data points) with 12 free parameters (**Table 4**). Minimum contrast gain at the fovea (g_σ; equation 15) was determined independently for each experiment, which yielded four free parameters. The remaining eight parameters were shared among experiments, as described above for involuntary attention.

Table 4. Free parameters for the fits to the neutral and central cueing conditions of the four voluntary attention experiments. Bold values indicate the median and values within square brackets depict the 95% CI of the bootstrapped distribution of fitted parameters. min. = minimum; SF = spatial frequency; bw = bandwidth; amp. = amplitude.

		Contrast gain		**Stimulus drive**			**Broad gain profile**			
parameter		g_σ	m_σ	t_T	m_T	b_T	a_B	m_B	b_B	γ_B
description		*min.*	*slope*	*SF peak*	*SF slope*	*SF bw*	*SF peak*	*SF slope*	*SF bw*	*amp.*
Experiment	7	**1.6** [1.5 2.1]	**-0.09** [-0.1 -0.06]	**3.4** [2.8 4.0]	**-0.8** [-1.0 -0.5]	**1.3** [1.1 1.6]	**1.66** [0.5 2.3]	**-0.2** [-0.3 -0.01]	**3.0** [1.4 4.4]	**3.1** [2.2 6.2]
	8	**1.7** [1.5 1.9]								
	9	**2.5** [2.2 2.8]								
	10	**1.7** [1.5 1.9]								

Alternative models

To assess whether contextual modulation, via normalization, and spatial summation are critical for generating the CPD, we implemented two model variants, one without contextual modulation ('-context') and another without spatial summation ('-sum'). Each model variant was fit separately to neutral performance data from all ten psychophysical experiments using the configurations described in **Optimization strategy, Central performance drop**.

In the '-context' model, each RF was suppressed by its own response, simulating an extremely narrow suppressive pool. Specifically, the extent of suppressive pools (δ_f, δ_θ, δ_{pos}; equations 11-13) were set to 0. As a result, the contributions of surround, cross-orientation and cross-frequency suppression were absent.

The '-sum' variant removed spatial summation (i.e., *F* in equation 1) from the model. As a result, normalized responses were not pooled across space and the response to regions of homogenous orientation were attenuated.

We additionally compared whether distinct attentional gain profiles—narrow or broad—were necessary to generate the effects of involuntary and voluntary attention. To this end, each profile was fit to involuntary and voluntary attention experiments using the configurations described in **Parameter configuration, Attentional modulation**.

Model comparisons

We compared models using the Akaike Information Criterion(Akaike, 1974) (AIC), computed with the assumption of a normal error distribution, which is asymptotically equivalent to model comparison via cross-validation(Burnham and Anderson, 2002). The difference in AIC values between model variants indexed model performance. '-context' and '-sum' models were compared to the full model; positive values reflect the main model outperforming the variants. Additionally, narrow and broad attentional gain profiles were compared; positive values reflect the narrow gain profile outperforming the broad profile.

Stimulus generation

Target-present and target-absent textures were re-created to match the stimulus parameters used in each psychophysical study (**Table 5**). For all stimuli, each pixel subtended 0.03125° (i.e., 32 pixels/°), roughly matching the spatial resolution of a 1280 × 960 monitor display placed 57 cm

away from the observer. Although these monitor display parameters were not identical to those used in each psychophysical study, they are typical of visual psychophysics experiments.

The full texture stimulus used in each experiment typically spanned the entire display. We generated 5°-wide square cutouts of the texture stimulus, centered on the target location. The target was displayed along the vertical meridian in Talgar & Carrasco (2002)(Talgar and Carrasco, 2002) to assess the impact of asymmetries in visual function between the cardinal meridians(Carrasco et al., 2001; Cameron et al., 2002; Himmelberg et al., 2020; Barbot et al., 2021). Additionally, targets were displayed at the diagonal (intercardinal) meridian in Barbot & Carrasco (2017)(Barbot and Carrasco, 2017) and in experiments 3 and 4 in Yeshurun, Montagna & Carrasco (2008)(Yeshurun et al., 2008a) to avoid the cardinal asymmetries. All other experiments presented targets along the horizontal meridian (**Table 5**). Because the model implemented visual sensitivity that varied with eccentricity, but was uniform at isoeccentric locations, all targets were assumed to be presented along the horizontal eccentricity for simplicity (as in equation 6).

Each texture array was composed of lines oriented 135°. For target-present textures, the target comprised a patch of lines that were oriented 45°. The exception was for the Carrasco, Loula and Ho (2006)(Carrasco et al., 2006) study, in which the texture array comprised vertical lines (0°) and the target patch contained lines tilted ±8° (**Figure 3**). In this study, observers' performed an orientation discrimination task wherein they reported the orientation of the target patch presented on each trial. To simulate orientation discrimination performance, the target-present and target-absent stimuli that were input to the model always contained a patch but their orientation differed.

To avoid edge artifacts, texture stimuli were windowed by the sum of three cosine window functions (as in equation 8) centered on the target that produced a uniform plateau covering the central 3.75 deg and fell off with cosine edges. Pixel intensities in each stimulus were constrained between 0 and 1.

To maximize model discriminability and minimize random fluctuations in model output, textures used to fit the model were generated without spatial or orientation jittering. In additional simulations, the stimuli of two representative experiments were jittered. The stimuli for experiment 1 of Yeshurun & Carrasco (1998)(Yeshurun and Carrasco, 1998) were spatially jittered (0.3 deg jitter), and the stimuli in experiment 4 of Yeshurun, Montagna & Carrasco (2008)(Yeshurun et al., 2008a), were jittered spatially (0.34 deg jitter) and in orientation (55° bandwidth). Parameters of the jitter were compatible with those specified in each study.

Table 5. Stimulus parameters for each texture segmentation experiment. Shaded rows denote voluntary attention experiments. Experiment numbers match those described in **Table 2**.

Experiment	**Line spacing**: horz × vert	**Line size**: width × height	**Target size**: width × height	**Target meridian:**
1	0.68° × 0.71°	0.1° × 0.4°	1.97° × 2.03°	horizontal
2	1.36° × 1.42°	0.2° × 0.8°	3.78° × 3.84°	horizontal
3	0.71° × 0.68°	0.1° × 0.4°	2.03° × 1.97°	vertical
4	1.43° × 1.37°	0.2° × 1°	4.09° × 3.97°	horizontal
5	0.68° × 0.71°	0.1° ×0.7°	2.28° × 2.34°	horizontal
6	0.68° × 0.71°	0.1° × 0.7°	2.28° × 2.34°	horizontal
7	*0.68° × 0.71°*	*0.1° × 0.7°*	*2.28° × 2.34°*	horizontal
8	*0.46° × 0.46°*	*0.1° × 0.2°*	*1.34° × 1.34°*	diagonal
9	*0.91° × 0.91°*	*0.2° × 0.4°*	*2.47° × 2.47°*	diagonal
10	*0.4° × 0.4°*	*0.1° × 0.24°*	*1.03° × 1.84°*	diagonal

Resampling procedures

We obtained confidence intervals on the parameter estimates and model predictions by bootstrapping the data and refitting the model 100 times per configuration (**Optimization strategy**) and for each model variant (**Alternative models**). Bootstrap samples were generated by drawing random samples from Gaussian distributions. Gaussians were centered on group-average performance at a given eccentricity and the standard error of the mean for each study defined the width of the distribution. The model was fit to each resampled dataset, which yielded 100 fits of resampled data. Confidence intervals on parameter estimates (**Tables 2-4**) were calculated from the bootstrapped estimates. Confidence intervals on model fits (**Figures 4-6, 8**) were calculated for each eccentricity using the bootstrapped model output.

To generate bootstrap samples for simulations with jittered texture stimuli, the model was first fit to the data for each experiment using a non-jittered texture. Then, model parameters were fixed and jittered stimuli were input to the model. One hundred jittered stimuli were presented to the model,

which yielded 100 estimates of texture segmentation performance at each eccentricity. Confidence intervals on the model output (**Supplementary Figure 5**) were calculated for each eccentricity using the distribution of model outputs.

Parameter interpretation

Several parameters contributed to spatial resolution including the spatial frequency gain and bandwidth of both the stimulus drive (t_T, m_T, b_T) and attentional gain along eccentricity (a_N, m_N, b_N, a_B, m_B, b_B). We used the fitted parameters to calculate summary metrics describing the intrinsic spatial frequency tuning of the model, determined using the neutral condition, and separately, the center frequency of voluntary and involuntary attentional modulation produced by the model.

Predicted contrast sensitivity functions

Contrast sensitivity functions were generated from the model to assess their compatibility with two datasets of psychophysical measurements(Pointer and Hess, 1989; Jigo and Carrasco, 2020). These datasets were chosen as they measured contrast sensitivity across a wide range of spatial frequencies and eccentricities.

Two important differences exist between these two datasets. First, Jigo and Carrasco (2020)(Jigo and Carrasco, 2020) used a fixed grating size (4° wide, cosine window) for all spatial frequencies, whereas Pointer and Hess (1989)(Pointer and Hess, 1989) adjusted grating size so that 3.2 periods of each spatial frequency were contained in a Gaussian window. As a result, gratings of low spatial frequency were larger than gratings of high spatial frequency. Second, Jigo and Carrasco (2020)(Jigo and Carrasco, 2020) used an orientation discrimination task to measure contrast sensitivity, whereas Pointer and Hess (1989)(Pointer and Hess, 1989) used a detection task.

The flexibility of our model allowed us to simulate contrast sensitivity for these different experiments. To simulate the signal available to an observer in a detection task, the population response to a grating was compared to the population response to a uniform gray image. For the discrimination task, the difference in population response between orthogonal gratings determined the available signal.

For each grating spatial frequency, model population responses were evaluated for 0-20 degrees of eccentricity and at 9 levels of contrast that were interpolated to characterize the full contrast response function at a given eccentricity. Interpolation was used to reduce computational load;

similar contrast response functions were generated when the model was evaluated at finer contrast steps. The available signal was scaled by internal noise magnitude to yield model discriminability (**Decision mechanism**). Because internal noise varies with spatial frequency(Silvestre et al., 2018), the available signal across eccentricity was scaled so that the maximum *d′* for each spatial frequency at the fovea equaled 2. Contrast thresholds were then determined as the level of contrast needed to reach *d′*=1. The inverse of contrast threshold served as our index of contrast sensitivity.

To generate contrast sensitivity estimates, we randomly sampled (with replacement) from the full distribution of best-fitting parameters for the fits to the neutral condition across all ten texture segmentation experiments (**Table 2**). Parameters were resampled 100 times, resulting in 100 different contrast sensitivity estimates for each spatial frequency and eccentricity.

Cross-validation procedure

To characterize how the operating range of involuntary and voluntary attention varied with eccentricity, relative to baseline tuning preferences (**Supplementary Figures 3 and 4**), we fit polynomials to empirical measurements made by Jigo and Carrasco (2020)(Jigo and Carrasco, 2020). Leave-one-subject-out cross-validation was used to determine the best-fitting polynomial order. Specifically, the ratio, in octaves, between the peak spatial frequency of the neutral contrast sensitivity function and the preferred spatial frequency of attentional modulation were computed for individual observers. Eccentricities were aggregated between each of the two experiments conducted. The ratio for one observer was set aside, and the remaining were averaged across observers. Zero to second-order polynomials were fit to the group-average ratio across eccentricities. Cross-validation error was computed as the sum of squared residuals to the left-out data point. This process was iterated until each observation was left-out once, resulting in 19 total iterations. The best-fitting polynomial order was defined as one that produced the lowest median cross-validation error across all iterations.

REFERENCES

Acerbi L, Ma WJ (2017) Practical Bayesian optimization for model fitting with Bayesian adaptive direct search. Advances in neural information processing systems 30:1836-1846.

Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. Journal of the Optical Society of America A 2:284.

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Contr 19:716-723.

Anton-Erxleben K, Carrasco M (2013) Attentional enhancement of spatial resolution: linking behavioural and neurophysiological evidence. Nature Reviews Neuroscience 14:188-200.

Barbot A, Carrasco M (2017) Attention modifies spatial resolution according to task demands. Psychological science 28:285-296.

Barbot A, Xue S, Carrasco M (2021) Asymmetries in visual acuity around the visual field. Journal of Vision 21:2-2.

Baruch O, Yeshurun Y (2014) Attentional attraction of receptive fields can explain spatial and temporal effects of attention. Visual Cognition 22:704-736.

Bergen JR, Adelson EH (1988) Early vision and texture perception. Nature 333:363-364.

Bonds AB (1989) Role of Inhibition in the Specification of Orientation Selectivity of Cells in the Cat Striate Cortex. Visual Neuroscience 2:41-55.

Brouwer GJ, Heeger DJ (2011) Cross-orientation suppression in human visual cortex. Journal of Neurophysiology 106:2108-2119.

Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd ed Edition. New York: Springer.

Busse L, Katzner S, Treue S (2008) Temporal dynamics of neuronal modulation during exogenous and endogenous shifts of visual attention in macaque area MT. Proceedings of the National Academy of Sciences 105:16380-16385.

Cameron EL, Tai JC, Carrasco M (2002) Covert attention affects the psychometric function of contrast sensitivity. Vision Research 42:949-967.

Carandini M, Heeger DJ (2011) Normalization as a canonical neural computation. Nature Reviews Neuroscience.

Carrasco M (2011) Visual attention: The past 25 years. Vision Research 51:1484-1525.

Carrasco M, Talgar CP, Cameron EL (2001) Characterizing visual performance fields: Effects of transient covert attention, spatial frequency, eccentricity, task and set size. Spatial vision 15:61-75.

Carrasco M, Loula F, Ho Y-X (2006) How attention enhances spatial resolution: Evidence from selective adaptation to spatial frequency. Attention, Perception, & Psychophysics 68:1004-1012.

Cavanaugh JR, Bair W, Movshon JA (2002) Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. Journal of Neurophysiology 88:2530-2546.

De Valois RL, Albrecht DG, Thorell LG (1982) Spatial frequency selectivity of cells in macaque visual cortex. Vision Research 22:545-559.

DeAngelis GC, Robson JG, Ohzawa I, Freeman RD (1992) Organization of suppression in receptive fields of neurons in cat visual cortex. Journal of Neurophysiology 68:144-163.

DeValois RL, DeValois KK (1990) Spatial vision, 2nd ed Edition. New York: Oxford Univ. Press.

Dugué L, Merriam EP, Heeger DJ, Carrasco M (2018) Specific Visual Subregions of TPJ Mediate Reorienting of Spatial Attention. Cerebral Cortex 28:2375-2390.

Dugué L, Merriam EP, Heeger DJ, Carrasco M (2020) Differential impact of endogenous and exogenous attention on activity in human visual cortex. Scientific Reports 10:21274.

El-Shamayleh Y, Movshon JA (2011) Neuronal Responses to Texture-Defined Form in Macaque Visual Area V2. Journal of Neuroscience 31:8543-8555.

Eriksen CW, Hoffman JE (1972) Temporal and spatial characteristics of selective encoding from visual displays. Perception & Psychophysics 12:201-204.

Fernández A, Carrasco M (2020) Extinguishing Exogenous Attention via Transcranial Magnetic Stimulation. Current Biology.

Fischer J, Whitney D (2009) Attention narrows position tuning of population responses in V1. Current Biology 19:1356-1361.

Giordano AM, McElree B, Carrasco M (2009) On the automaticity and flexibility of covert attention: A speed-accuracy trade-off analysis. Journal of Vision 9:30-30.

Graham NVS (1989) Visual pattern analyzers. New York: Oxford University Press.

Hallum LE, Landy MS, Heeger DJ (2011) Human primary visual cortex (V1) is selective for second-order spatial frequency. Journal of Neurophysiology 105:2121-2131.

Heeger DJ (1992) Normalization of cell responses in cat striate cortex. Visual Neuroscience 9:181-197.

Herrmann K, Montaser-Kouhsari L, Carrasco M, Heeger DJ (2010) When size matters: attention affects performance by contrast or response gain. Nature Neuroscience 13:1554-1559.

Himmelberg MM, Winawer J, Carrasco M (2020) Stimulus-dependent contrast sensitivity asymmetries around the visual field. Journal of vision 20:18-18.

Jazayeri M, Movshon JA (2006) Optimal representation of sensory information by neural populations. Nature Neuroscience 9:690-696.

Jigo M, Carrasco M (2018) Attention alters spatial resolution by modulating second-order processing. Journal of Vision 18:2.

Jigo M, Carrasco M (2020) Differential impact of exogenous and endogenous attention on the contrast sensitivity function across eccentricity. Journal of Vision 20:11.

Kehrer L (1989) Central performance drop on perceptual segregation tasks. Spatial vision 4:45-62.

Kehrer L (1997) The central performance drop in texture segmentation: A simulation based on a spatial filter model. Biological Cybernetics 77:297-305.

Kehrer L, Meinecke C (2003) A space-variant filter model of texture segregation: Parameter adjustment guided by psychophysical data. Biological Cybernetics 88:183-200.

Kelly DH (1977) Visual Contrast Sensitivity. Optica Acta: International Journal of Optics 24:107-129.

Klein Barrie P, Harvey Ben M, Dumoulin Serge O (2014) Attraction of position preference by spatial attention throughout human visual cortex. Neuron 84:227-237.

LaBerge D (1983) Spatial Extent of Attention to Letters and Words. Journal of Experimental Psychology: Human Perception and Performance 9:371-379.

Landy MS (2013) Texture analysis and perception. In: The New Visual Neurosciences (Werner JS, Chalupa LM, eds), pp 639-652. Cambrige, MA: MIT Press.

Landy MS, Oruç İ (2002) Properties of second-order spatial frequency channels. Vision Research 42:2311-2329.

Landy MS, Graham N (2004) Visual Perception of Texture. In: The Visual Neurosciences, pp 1106-1118. Cambrige, MA: MIT Press.

Larsson J, Landy MS, Heeger DJ (2006) Orientation-Selective Adaptation to First- and Second-Order Patterns in Human Visual Cortex. Journal of Neurophysiology 95:862-881.

Lesmes LA (2010) Bayesian adaptive estimation of the contrast sensitivity function: The quick CSF method. Journal of Vision 10:1-21.

Liu T, Pestilli F, Carrasco M (2005) Transient attention enhances perceptual performance and fMRI response in human visual cortex. Neuron 45:469-477.

Macmillan NA, Creelman CD (2005) Detection theory: A user's guide. Mahwah, NJ: Lawrence Erlbaum Associates.

Meese TS, Hess RF (2004) Low spatial frequencies are suppressively masked across spatial scale, orientation, field position, and eye of origin. Journal of Vision 4:2.

Meese TS, Holmes DJ (2007) Spatial and temporal dependencies of cross-orientation suppression in human vision. Proc R Soc B 274:127-136.

Moran J, Desimone R (1985) Selective attention gates visual processing in the extrastriate cortex. Science 229:782-784.

Morikawa K (2000) Central performance drop in texture segmentation: the role of spatial and temporal factors. Vision Research 40:3517-3526.

Pestilli F, Ling S, Carrasco M (2009) A population-coding model of attention's influence on contrast response: Estimating neural effects from psychophysical data. Vision research 49:1144-1153.

Petrov Y, Carandini M, McKee S (2005) Two distinct mechanisms of suppression in human vision. Journal of Neuroscience 25:8704-8707.

Pointer JS, Hess RF (1989) The contrast sensitivity gradient across the human visual field: With emphasis on the low spatial frequency range. Vision research 29:1133-1151.

Priebe NJ, Ferster D (2006) Mechanisms underlying cross-orientation suppression in cat visual cortex. Nature Neuroscience 9:552-561.

Reynolds JH, Chelazzi L (2004) Attentional modulation of visual processing. Annu Rev Neurosci 27:611-647.

Reynolds JH, Heeger DJ (2009) The Normalization Model of Attention. Neuron 61:168-185.

Ringach DL, Shapley RM, Hawken MJ (2002) Orientation Selectivity in Macaque V1: Diversity and Laminar Dependence. The Journal of Neuroscience 22:5639-5651.

Rohatgi A (2020) Webplotdigitizer: Version 4.3.

Sagi D, Hochstein S (1985) Lateral inhibition between spatially adjacent spatial-frequency channels? Perception & Psychophysics 37:315-322.

Silvestre D, Arleo A, Allard R (2018) Internal noise sources limiting contrast sensitivity. Scientific Reports 8:2596.

Simoncelli EP, Freeman WT, Adelson EH, Heeger DJ (1992) Shiftable multiscale transforms. IEEE Transactions on Information Theory 38:587-607.

Stigchel SVd, Belopolsky AV, Peters JC, Wijnen JG, Meeter M, Theeuwes J (2009) The limits of top-down control of visual attention. Acta Psychologica 132:201-212.

Talgar CP, Carrasco M (2002) Vertical meridian asymmetry in spatial resolution: Visual and attentional factors. Psychonomic Bulletin & Review 9:714-722.

Teo PC, Heeger DJ (1994) Perceptual image distortion. In: 1st International Conference on Image Processing, pp 982-986: IEEE Comput. Soc. Press.

Tkacz-Domb S, Yeshurun Y (2018) The size of the attentional window when measured by the pupillary response to light. Scientific Reports 8:11878.

Tsotsos JK, Culhane SM, Wai WYK, Lai Y, Davis N, Nuflo F (1995) Modeling visual attention via selective tuning. Artificial intelligence 78:507-545.

Victor JD, Conte MM, Chubb CF (2017) Textures as probes of visual processing. Annual review of vision science 3:275-296.

Wang F, Chen M, Yan Y, Zhaoping L, Li W (2015) Modulation of Neuronal Responses by Exogenous Attention in Macaque Primary Visual Cortex. Journal of Neuroscience 35:13419-13429.

Watson AB, Solomon JA (1997) Model of visual contrast gain control and pattern masking. Journal of the Optical Society of America A 14:2379.

Watson AB, Ahumada AJ (2005) A standard model for foveal detection of spatial contrast. Journal of Vision 5:6.

Westrick ZM, Landy MS (2013) Pooling of first-order inputs in second-order vision. Vision Research 91:108-117.

Womelsdorf T, Anton-Erxleben K, Treue S (2008) Receptive Field Shift and Shrinkage in Macaque Middle Temporal Area through Attentional Gain Modulation. Journal of Neuroscience 28:8934-8944.

Womelsdorf T, Anton-Erxleben K, Pieper F, Treue S (2006) Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. Nature Neuroscience 9:1156-1160.

Yeshurun Y (2019) The spatial distribution of attention. Current Opinion in Psychology 29:76-81.

Yeshurun Y, Carrasco M (1998) Attention improves or impairs visual performance by enhancing spatial resolution. Nature 396:72.

Yeshurun Y, Carrasco M (2000) The locus of attentional effects in texture segmentation. Nature Neuroscience 3:622-627.

Yeshurun Y, Carrasco M (2008) The effects of transient attention on spatial resolution and the size of the attentional cue. Perception & Psychophysics 70:104-113.

Yeshurun Y, Montagna B, Carrasco M (2008a) On the flexibility of sustained attention and its effects on a texture segmentation task. Vision Research 48:80-95.

Yeshurun Y, Carrasco M, Maloney LT (2008b) Bias and sensitivity in two-interval forced choice procedures: Tests of the difference model. Vision Research 48:1837-1851.