Chapter 6

A Population Coding Model of Visual Crowding
Abstract

An object in the peripheral visual field is more difficult to recognize when surrounded by other objects. This phenomenon is called crowding. Crowding places a fundamental constraint on human vision that limits performance on numerous tasks. While there is a growing consensus that crowding results from spatial feature integration, the underlying mechanism and functional origin of the effect remain questioned. Here, we present a quantitative and physiologically plausible model for spatial integration of orientation signals, based on the principles of population coding. Using simulations, we demonstrate that this model coherently accounts for several fundamental properties of crowding, including critical spacing, compulsory averaging, and a foveal-peripheral anisotropy. Moreover, we show that the model predicts increased saliency of correlated visual stimuli such as contours. Altogether, these results suggest that crowding is a by-product of a general integration mechanism aimed at improving signal quality in early vision.
6.1 Introduction

The important role of spacing for object recognition was first described in the 1920s, when Korte discovered that a letter is harder to recognize when it is flanked by other letters [Korte, 1923]. This phenomenon is now popularly known as crowding [Stuart & Burian, 1962] and has since been studied extensively (for two recent reviews, see [Levi, 2008; Pelli & Tillman, 2008]).

The strength of the crowding effect depends on the spacing between objects (Figure 6.1). The largest spacing at which there is a measurable effect is commonly referred to as the critical spacing. An important and often replicated finding is that the critical spacing for object recognition is proportional to the viewing eccentricity [Bouma, 1970]. Moreover, critical spacing is found to be highly invariant to a great variety of stimulus manipulations, such as contrast and size [Pelli et al., 2004; Strasburger et al., 1991; Tripathy & Cavanagh, 2002]. Critical spacing is the most extensively studied crowding property and, because of its robustness, now sometimes considered the defining property of crowding [Pelli & Tillman, 2008].

Crowding is a general phenomenon in vision. It is not confined to letter and shape recognition, but affects a broad range of stimuli and tasks, including the identification of orientation [Andriessen & Bouma, 1976; Parkes et al., 2001; Wilkinson et al., 1997] object size, hue and saturation of colors [van den Berg et al., 2007], recognition of faces [Louie et al., 2007; Martelli et al., 2005], reading [Pelli et al., 2007], and visual search [Motter & Simoni, 2007; Vlaskamp & Hooge, 2005; Wertheim et al., 2006]. Altogether, crowding emerges as a fundamental limiting factor in vision, making the question about its neural basis and functional origin rather pressing.

Several theories have been proposed to explain the crowding effect. Currently, there is a growing consensus that crowding results from feature integration over an area that is larger than the target object [Levi, 2008]. However, there is a marked controversy about both the underlying mechanism and the functional origin of the effect. Some authors assert the existence of bottom-up hardwired integration fields (e.g., [Pelli & Tillman, 2008]), while others claim that feature integration arises from limitations related to the spatial resolution of attention (e.g., [Cheung et al., 2006; He et al., 1996]). Postulated functions of feature integration include texture perception [Parkes et al., 2001], contour integration [May & Hess, 2007], and object recognition [Pelli & Tillman, 2008]. In the absence of formal models, however, it is neither clear whether these theories can also quantitatively account for the mysteries of crowding [Levi, 2008] nor how plausible they are from a biological perspective.

Here, we present a quantitative model for spatial integration of orientation signals. Our model is based on the principles of population coding [Pouget et al., 2000], which is an approach that mathematically formalizes the idea that information is encoded in the brain by populations of cells, rather than by single cells. A population code is defined as the simultaneous activity of an ensemble of neu-
rons that encodes a probability distribution over a single variable, such as the orientation of a visual edge. An appealing property of this approach is that it allows for studying neural computation in a mathematical and biologically plausible way. Stimuli are first encoded as neural signals, typically using tuning functions and a noise model. Next, these signals are manipulated (e.g., integrated, as in our case). Finally, a decoder can be derived in order to transform the signals back to stimulus space and assess what the effects of these computations are for perception.

Figure 6.1: Top: An example demonstrating crowding. The two Bs are at equal distance from the fixation cross. On the left, where the center-to-center spacing between the letters is approximately one half of the eccentricity of the central letter, the B can easily be recognized when fixating the cross. Letter spacing on the right is much smaller, and the B appears to be jumbled with its neighbors. Bottom, left: Human data from a typical crowding experiment. Crowding diminishes as target-flanker spacing is increased, up to a certain critical spacing after which flankers have no effect. Bottom, right: Findings from psychophysical studies show that critical spacing is a linear function of target eccentricity. Data from [van den Berg et al., 2007].

Motivated by findings from physiological [Kapadia et al., 1995; Sceniak et al., 1999] and theoretical [Ma et al., 2006] studies, we model feature integration as a (weighted) summation of population codes. Using simulations, we demonstrate that our model reproduces fundamental crowding behavior, including aspects of critical spacing [Pelli et al., 2004, 2007], compulsory averaging of crowded orientation signals [Parkes et al., 2001], and an asymmetry between the effects of foveally and peripherally placed flankers [Bouma, 1973; Petrov et al., 2007]. These results provide strong quantitative support to the notion that hard-wired feature integration underlies crowding. Finally, we show that the model predicts enhancement of signals that encode visual contours, which could facilitate subsequent contour
6.2 A POPULATION CODING MODEL OF VISUAL CROWDING

detection and segmentation. Altogether, these results suggest that crowding is a by-product of a general integration mechanism aimed at improving signal quality in early vision.

6.2 Results

6.2.1 Model

Several different population coding schemes have been proposed in the literature [Ma & Pouget, 2009]. Although they differ in their details, the general idea behind all of them is that variables are encoded in the brain by entire populations of cells. Our model is based on the distributional population coding (DPC) scheme that was proposed by Zemel et al. (1998). In this scheme, a population code explicitly encodes a probability distribution over the stimulus domain. In this section we will only provide a general overview of our model. Mathematical details can be found in the Methods section.

The input to the model consists of a set of stimuli, each one defined by a location, orientation, contrast, and size (Figure 6.2a). These parameters define the probability distributions (Figure 6.2b) that are subsequently used as inputs to the DPC encoder that computes a population code representation for each of the stimuli (Figure 6.2c). In the second layer, stimulus representations from the first layer are spatially integrated, in the form of weighted summations of cell responses (Figure 6.2d). The integration weights depend on the cortical distance in primary visual cortex between the locations of the integration cell and the cells encoding the input stimuli (for details of the weight function and mapping of visual field to cortical locations, see Methods). This function can be interpreted as defining a cortical integration field.

Several of the simulation experiments that we conducted required that a response be generated (e.g., when simulating psychophysical experiments involving target tilt estimation). In those simulations, a maximum-likelihood decoder was used to decode the post-integration population code associated with the target position back to a stimulus distribution (Figure 6.2e). The number of components of the returned mixture model was interpreted as the number of distinct orientations perceived at the location associated with the decoded population code, the mixing proportions as the amounts of evidence for the presence of an orientation, the means as estimates of these orientations, and the standard deviations as the amounts of uncertainty about these estimates.

6.2.2 Critical regions for crowding

A well-established behavioral finding in human observers is that identification thresholds for a crowded target decrease as a function of target-flanker spacing until a certain critical spacing is reached. Beyond this critical spacing flankers
A graphical illustration of our model. (a) In this example, the input consists of three oriented bars (the colors are only for visualization purposes and not part of the input to the model); (b) probability distributions are defined for the input stimuli; these distributions capture the stimulus uncertainty caused by neural noise in processing stages prior to the first layer of the model; (c) In the first layer, a neural representation is computed for each of these distributions. (d) In the second layer, the stimulus representation at each location is integrated with the representations of stimuli at neighboring locations. Integration is implemented as a weighted summation, such that nearby stimuli receive higher weights than stimuli that are far away; (e) The resulting population codes are decoded to a mixture of normal distributions, with each component representing a perceived orientation at the respective location; (f) Due to integration, the resulting percept of closely spaced stimuli will be crowded.
no longer have an effect (see, for example, the results shown in Figure 6.1). In our model, the integration fields are implemented as weight functions of stimulus spacing in cortex. Consequently, flanker stimuli affect the identification of a target only when positioned within a certain distance from the target, yielding a critical region for target identification. To examine whether our model can quantitatively account for critical regions found for human subjects, we performed a simulation that mimicked the psychophysical experiment by Pelli et al. (2007), who estimated critical regions for letter identification at several positions in the visual field.

Critical regions predicted by our model were estimated as follows. For each target position, identification thresholds were determined for a range of target-flanker spacings (see Figures 6.3a and 6.3b; we refer to Methods for details about the procedure that was used to estimate identification thresholds). A clipped line was fit to the resulting data, providing an estimate of the critical spacing (Figure 6.3c). By varying the positions of the flankers, we estimated critical spacing in several directions around the target. Combining these spacings gives an estimate of the critical region around a given target location (Figure 6.3d). We estimated model parameter values that result in a good model fit to one of the critical regions measured by Pelli et al. Subsequently, we repeated the experiment for the other target locations using the same parameter values, and found that the model accurately predicts all reported human critical regions (Figure 6.3d). These results thus provide quantitative evidence for the suggestion that the behavioral crowding regions found in humans can be explained as the result of fixed-sized, hard-wired integration fields in visual cortex.

6.2.3 Effect of stimulus manipulations on critical spacing

The critical spacing for crowding is known to scale with eccentricity and is consistently found to be in the range 0.3-0.6 times the target eccentricity [Pelli et al., 2004]. Moreover, it is found to be largely invariant under changes to the physical properties of the stimulus, such as the size, contrast, and number of flankers [Pelli et al., 2004] and the scaling of stimuli (i.e., changing the size of both the target and flankers) [Pelli et al., 2004; Strasburger et al., 1991; Tripathy & Cavanagh, 2002]. To further verify our model, we conducted another series of simulation experiments, in which we manipulated several stimulus properties. We found that the results are compatible with findings in human subjects: critical spacing predicted by our model scales linearly with target eccentricity and is hardly affected by stimulus manipulations (Figure 6.4).

6.2.4 Compulsory averaging of crowded orientation signals

Human observers are able to report the mean orientation of a set of crowded stimuli, but not the orientations of the individual stimuli [Parkes et al., 2001]. This peculiar crowding property is generally referred to as compulsory averaging. In the experiment of Parkes et al., observers reported the tilt direction of a variable
Figure 6.3: Comparison of crowding regions reported for humans with crowding regions estimated by our model. (a) The input stimulus on each trial consisted of a 10 tilted target stimulus and two 30 tilted flankers placed on opposite sides of the target. If the sign of the post-integration stimulus representation associated with the target position was the same as the sign of the input target, then performance on that trial was considered correct. (b) Performance was estimated for a range of target contrasts, yielding a curve that is very similar to psychometric curves typically found with human experiments (compare, for example, with data shown in Figure 6.1). Based on these curves, contrast thresholds were estimated that produce 75% correct performance. (c) Contrast thresholds decrease as target-flanker spacing is increased. The smallest spacing at which the flankers do not have an effect is defined as the critical spacing. (d) Critical spacings were estimated in several directions around the target, at five different target positions. These simulation data accurately reproduce the critical regions measured psychophysically in humans. Human data from [Pelli et al., 2007].
Figure 6.4: Simulation results showing the effect of several stimulus manipulations on estimated critical spacing. The shaded areas represent the range of critical spacings that are typically reported in the literature (0.3-0.6 times target eccentricity). Standard errors are smaller than the marker size. (a) Critical spacing scales linearly with target eccentricity. (b-f) Critical spacing is only weakly affected by various stimulus manipulations. The eccentricity of the target was 6 degrees in these experiments.
number of equally tilted targets positioned among horizontal flankers. Parkes et al. found that a relatively simple pooling model could account for human data when the total number of stimuli is kept constant. However, when targets are presented without flankers, identification thresholds dropped significantly slower as a function of the number of targets than predicted by their model (Figure 6.5b). They postulated a late noise factor to explain the discrepancy between data and model.

Our model suggests the following explanation for the compulsory averaging phenomenon. When two features are highly similar, their population code representations have a high degree of overlap and will merge when summed. Consequently, the resulting post-integration code will be interpreted as representing a single feature with a value somewhere in between the values of the input stimuli (Figure 6.5a). To examine whether our model can also quantitatively account for compulsory averaging, we conducted a simulation experiment with conditions and stimuli similar to those used in the psychophysical experiment performed by Parkes et al. (2001). The results show that our model produces accurate fits to the psychophysical data for both the condition with and without flankers (Figure 6.5b).

An important difference between our model and the pooling model proposed by Parkes et al. is that the latter integrates all stimuli with equal weight, while integration in our model is weighted by object spacing. To verify the relevance of this aspect in explaining why the models make different predictions, we reran the simulations with varying stimulus spacing. We found that, first, when spacing is set to 0.5, our model provides a good fit to the human data that were found with the same object spacing. Second, when spacing is set to 0, the predictions of our model are the same as the predictions of the pooling model proposed by Parkes et al. Finally, when spacing is set to a value close to or larger than the critical spacing (which was about 1.25 deg in this simulation), then predicted identification thresholds are independent of the number of targets, as one would expect.

We found that when we set all integration weights in our model to one (implying an object spacing of zero), the identification thresholds predicted by our model are similar to those predicted by the pooling model of Parkes et al. Additionally, the predictions of the models increasingly diverge when object spacing is increased. These results confirm that object-spacing related weighting of integration is an essential difference between the models. Moreover, they challenge the need for the late noise factor proposed by Parkes et al. to explain their results.

6.2.5 Peripheral flankers cause stronger crowding than foveal flankers

Several studies [Bouma, 1970; Petrov et al., 2007] have found that, with equal target-flanker spacing, flankers positioned at the peripheral side of a target cause stronger crowding effects than flankers positioned at the foveal side. As has been noted previously [Motter & Simoni, 2007], this asymmetry follows directly from
Figure 6.5: (a) Simulation results illustrating a possible explanation of the compulsory averaging effect of crowded orientation signals. Top row: example input stimuli, consisting of a vertical target flanked by two equally tilted flankers. Second row: single trial examples of population codes representing the post-integration stimulus at the target position. Third row: distributions of the orientations encoded at the target locations after integration (1000 trials). Bottom row: corresponding distributions of the number of perceived stimuli at the target position. When target and flanker tilt are nearly identical, their population code representations merge into a single hill of activity when integrated. The resulting code is decoded to a single orientation, with a value intermediate between the values of the input stimuli. This effect diminishes when the difference between target and flanker tilt is increased. (b) Model fit to human psychophysical data. Top: Example stimuli of the experiment described in [Parkes et al., 2001]. The task was to report the tilt direction of a variable number of equally tilted targets positioned within a set of horizontal flankers. Bottom: Identification thresholds predicted by our model are very close to those found for human subjects. Human data from [Parkes et al., 2001], subject LP.
the way that the visual field is mapped onto the cortex. With increasing eccentricity, the representation of visual space becomes more and more compressed. Consequently, for equal target-flanker spacing in visual space, the cortical distance between the representation of a target and a foveal flanker is larger than that between a target and a peripheral flanker. Assuming that cortical integration fields are isotropic, peripheral flankers will, therefore, contribute more to the integrated target signal than foveal flankers.

Figure 6.6: Simulation results illustrating the anisotropic effects of foveal vs peripheral flankers on target identification. (a) Stimuli consisting of a +/- 10° tilted target, flanked by either no flanker, a foveal flanker, or a peripheral flanker. (b) Both flankers elevate target tilt identification thresholds, but this effect is largest for peripheral flankers. We define threshold elevations \( TE_{\text{foveal}} \) and \( TE_{\text{peripheral}} \) as the 75%-correct target contrast found for the condition with a foveal and peripheral flanker, respectively, divided by the 75%-correct target contrast found for the condition without a flanker. (c) Predicted threshold elevations plotted as a function of target-flanker spacing. When target-flanker spacing is small or when it approaches the critical spacing, the effects of foveal and peripheral flankers are comparably strong. However, in the intermediate range, a peripheral flanker produces larger threshold elevations (i.e., stronger crowding) than a foveal flanker. (d) The same data as in (c), but now shown as a ratio (i.e., the values at black data points from panel (c) divided by those at the red data points).

We conducted a simulation experiment to verify whether our model replicates the foveal-peripheral anisotropy and to investigate how its predictions depend on target-flanker spacing. For several target-flanker spacings, we estimated 75%-correct target contrast thresholds for identifying the tilt of a target without a flanker, a target with a foveal flanker, and a target with a peripheral flanker.
6.2 A POPULATION CODING MODEL OF VISUAL CROWDING

(Figure 6.6a). The results show that while both the foveal and a peripheral flanker produce crowding (Figure 6.6b), the effect caused by a peripheral flanker is substantially larger than that caused by a foveal flanker (Figure 6.6c). Hence, our model exhibits a foveal-peripheral flanker anisotropy. Furthermore, the model predicts the anisotropy to be strongest at intermediate spacings while it predicts no anisotropy when target-flanker spacing is very small or approaches the critical spacing (Figure 6.6d). In these simulation data, the strongest anisotropy is found when target-flanker spacing is about 2 degrees (i.e., about 0.3 times the target eccentricity). At this spacing, threshold elevation caused by the peripheral flanker is predicted to be approximately 2.5 times that caused by the foveal flanker. This is comparable to the effect size measured for human observers [Petrov et al., 2007].

6.2.6 Spatial integration enhances signals from correlated stimuli

The results so far suggest that crowding is what happens when signals from closely-spaced, unrelated stimuli are integrated with each other. However, in normal viewing conditions, signals from closely-spaced stimuli are often correlated (e.g., neighboring line segments of an edge or smooth contour). It has been suggested that integration of such correlated (orientation) signals may underlie phenomena such as contour integration [Field et al., 1993; Gilbert et al., 1996; Kapadia et al., 1995; Schmidt et al., 1997].

![Figure 6.7: Simulation results showing how our model responds to visual contours. Left image: input stimulus, consisting of a set of oriented line segments comprising several contours within a noisy background. The + symbol indicates the center of the visual field and was not part of the stimulus. Central image: a visualization of the stimulus representation in the first layer of our model, which is a noisy version of the input. The contrast of the bars is set to the median of the contrasts in the right image. Right image: a visualization of the decoded stimulus representations after integration. At every original input location, the post-integration population code was decoded to a mixture of normal distributions. The contrast of each bar is proportional to the associated mixing proportion. Note the highlighting of the contours and the crowding effects in the periphery, which agrees well with the subjective experience when viewing the input stimulus.](image)
To investigate how our model responds to signals from correlated stimuli, we ran a simulation with an input stimulus consisting of a set of line segments comprising various contours within a noisy background (see Methods for details). The results are shown in Figure 6.7. Line segments that are part of a contour clearly stand out in the post-integration representation. This is because both stimulus density and orientation correlation are higher for contours than for the random background. This result provides a functional context for crowding and strongly supports an earlier suggested link between contour integration and crowding [May & Hess, 2007]. Note that in areas away from fixation, in the periphery of the visual field, the decoder often returned stimulus distributions that represent more than one orientation value. This indicates that the post-integration codes at those locations are ambiguous in terms of the encoded orientation. In other words, when stimulus spacing is small relative to eccentricity, stimuli become jumbled with their neighbors, just as observed in crowding.

6.2.7 Effect of model parameter settings on our main results

We reran the simulation experiment that was used to estimate critical spacings (see main text) with a range of different parameter settings. The results are shown in Figure 6.8). We observe that the critical spacing predictions are hardly affected by the parameter settings, indicating that critical spacing is a rather general property of the population code integration model that we presented. Note that the floor thresholds increase when we reduce the gain or the number of neurons in the model. This is to be expected, because lower gain or fewer neurons means that there is less information (spikes) in the population codes, which will increase stimulus uncertainty.

6.3 Discussion

We presented a model of spatial feature integration based on the principles of population coding. While there is a growing consensus for the theory that spatial feature integration is responsible for crowding [Levi, 2008], the model that we presented here is the first to quantitatively account for several fundamental properties of this phenomenon in a coherent and biologically plausible manner. Besides replicating the properties of the critical spacing of crowding, and the anisotropic crowding effects of foveal versus peripheral flankers, our model also replicates and explains compulsory averaging of crowded orientation signals. Furthermore, it provides an explanation of the functional origin of crowding, by suggesting that it is the by-product of a mechanism aimed at enhancing the saliency of ecologically relevant stimuli, such as visual contours.

The parameter settings (see Methods) in our model were fixed over the entire range of simulations that we performed, with one minor exception (Figure 6.3). We reran a number of simulations with different parameter values and found that
Figure 6.8: Results of a simulation that estimated critical spacing for a tilt identification task of a target located at 6 degrees of eccentricity. The stimuli and procedure were the same as for the simulations in the main experiment. These results show that critical spacing is hardly affected by the model parameters, which indicates that critical spacing is a general property of the type of model that we proposed.
this hardly affected our results (Figure 6.8). This suggests that crowding is an inherent property of signal integration by population code summation.

### 6.3.1 Comparison with other theories of crowding

These results also shed new light on other crowding theories. Some authors have proposed that crowding is, at least in part, the result of source confusion due to positional uncertainty [Popple & Levi, 2005; Strasburger, 2005]. We would like to note, however, that integrating signals over space necessarily increases positional uncertainty. Hence, we consider location uncertainty and, consequently, source confusion as a result of feature integration, rather than an additional factor in the explanation of crowding. Indeed, our results show clear evidence for source confusion, even though we did not explicitly incorporate positional uncertainty into our model (for an example, see Figure 6.5a).

Other authors argue that crowding is the result of attentional limitations [Chakravarthi & Cavanagh, 2007; He et al., 1996], although evidence for these theories is considered very slim [Levi, 2008]. While we deem it possible that attentional factors have modulatory effects on crowding, our present results show that the general properties of crowding can very well be accounted for without invoking attentional mechanisms.

A recent theory suggests that crowding is the breakdown of object recognition [Pelli & Tillman, 2008]. The reasoning is that spatial integration of object features (in the notion of binding) is required for object recognition, whereas crowding occurs when multiple objects fall within the same integration field. Our results show that the spatial signal integration underlying crowding also enhances responses for correlated signals, such as contours. This corroborates an earlier suggestion that the association fields that have been proposed to underlie contour integration [Field et al., 1993] may also cause crowding [May & Hess, 2007]. However, while the enhancement of responses to correlated signals will no doubt facilitate higher-order functions such as object recognition, in our view, the function of integration appears more elementary and general.

Finally, another theory suggests that crowding is texture perception when we do not wish it to occur [Parkes et al., 2001]. The motivation behind this proposal is the finding that observers cannot identify individual stimulus properties in a crowded display, but still have access to its average statistics (i.e., its texture properties). Our model is able to explain this finding (see Figure 6.5), and we agree that what occurs after pooling can be described as texture perception. However, in view of the plausible connection between spatial integration and contour integration, we hesitate to conclude that texture perception is the primary function of spatial integration. Moreover, if a functional link exists between spatial integration and texture perception, then we deem it just as likely that integration serves to compress visual information, in order to reduce energy requirements at higher levels of processing.
6.3.2 Target-flanker similarity and configural effects

Two crowding properties that our current model does not account for are the effects of target-flanker similarity and flanker configuration. The former refers to the finding that crowding is stronger for target-like flankers compared to dissimilar flankers [Andriessen & Bouma, 1976; Kooi et al., 1994; Levi et al., 2002]. The flanker configuration effect refers to the finding that crowding is partially released when surrounding flankers form a contour [Livne & Sagi, 2007; Saarela et al., 2009]. A rather natural extension to our model may allow it to account for these two effects as well. At present, the integration fields in our model represent exclusively long-range horizontal connections between cells (about 0.5-5mm). Alongside these long-range connections, however, many of the cells in primary visual cortex are known to have extensive short-range connections (<0.5mm) as well as feedback connections from higher-order brain areas [Stettler et al., 2002]. The short-range connections are thought to be involved in lateral inhibition mechanisms. Inhibition could reduce the integration of dissimilar pieces of information and thus be responsible for target-flanker similarity effects in crowding. Likewise, feedback connections might inhibit the integration of signals that are likely to represent different objects or perceptual groups and, therefore, be responsible for configuration influences on crowding.

6.3.3 Crowding in other feature domains

The model and simulations that were presented in this paper are limited to the orientation domain. However, crowding is a rather general phenomenon that affects a large number of tasks, including discrimination of letters and objects sizes, colors, and shapes. Since population coding is considered the general way by which variables are encoded in the brain [Pouget et al., 2000], crowding of other basic features such as size and color [van den Berg et al., 2007] can presumably be explained by a model that is largely analogous to the one presented here. Crowding of more complex structures (such as letters, object shapes, bodies, and faces) could follow both from crowding in their constituent features and from crowding within higher-order population codes that represent the structures themselves.

6.3.4 Predictions

Our model makes a number of predictions that can be tested experimentally. For example, the simulations related to the compulsory averaging effect provide precise predictions regarding the effect of object spacing on identification thresholds. Additionally, the model makes quantitative predictions regarding the effect of spacing on the foveal-peripheral flanker anisotropy of crowding.
6.3.5 Conclusion

In conclusion, our results lend strong support to the theory that the mechanism behind crowding is spatial feature integration. Additionally, they support the theory that crowding and contour integration are two sides of the same coin.

6.4 Methods

6.4.1 Stimulus encoding (first model layer)

Input stimuli are specified as 4-tuples $S = (\theta, \alpha, \vec{\lambda}, c)$, where $\theta$ is the orientation, $\alpha$ the size, $\vec{\lambda} = (\lambda_x, \lambda_y)$ the location, and $c$ the (relative) contrast of the stimulus. In the first layer of the model, the distributional population coding (DPC) scheme by Zemel et al. (1998) is used to compute internal population code representations for these stimuli. We first define probability distributions for the input stimuli, which capture the stimulus uncertainty caused by neural noise in processing stages prior to the first layer of the model. We relate the width $\sigma$ of these distributions to the eccentricity $\lambda = ||\vec{\lambda}||$, size $\alpha$, and contrast $c$ of a stimulus, in the following way. From [Romano & Virsu, 1979] we know that the cortical magnification factor $M(\lambda)$ - which describes the differential change in cortical position with respect to retinal eccentricity $\lambda$ - can be approximated as follows:

$$M(\lambda) \approx \frac{8}{0.4\lambda + 1} \quad (6.1)$$

From this, we infer that the cortical size $A$ of a stimulus that subtends $\alpha$ deg$^2$ of visual angle and which is centered at $\lambda$ degrees of eccentricity, can be approximated as follows:

$$A(\alpha, \lambda) \approx \alpha [M(\lambda)]^2 \approx \frac{64\alpha}{(0.4\lambda + 1)^2} \quad (6.2)$$

Assuming that the uncertainty about the value of a stimulus is inversely proportional to both the size of its cortical representation and its contrast, we obtain the following relation between stimulus uncertainty $\sigma^2$ on the one hand, and stimulus eccentricity $\lambda$, contrast $c$, and size $\alpha$ on the other hand:

$$\sigma^2 \propto \frac{1}{A(\alpha, \lambda)c} \approx \frac{(0.4\lambda + 1)^2}{64\alpha c} \quad (6.3)$$

$$\sigma \propto \frac{1}{\sqrt{A(\alpha, \lambda)c}} \approx \frac{0.4\lambda + 1}{8\sqrt{\alpha c}} = \frac{0.4(\lambda + 2.5)}{8\sqrt{\alpha c}} \approx \frac{\lambda + 2.5}{\sqrt{\alpha c}} \quad (6.4)$$

In our model, we choose a constant of proportionality of 0.4:

$$\sigma = 0.4 \left( \frac{\lambda + 2.5}{\sqrt{\alpha c}} \right) \quad (6.5)$$
which gives, for example, \( \sigma = 1 \text{deg} \) for a foveally presented stimulus with a size of 1 deg\(^2\) and a contrast of 1, and \( \sigma = 4 \) for the same stimulus at 7.5 degrees of eccentricity.

In order to account for the circularity of the orientation domain, we define the input distributions to be circular normal (von Mises) distributions. More specifically, the distribution over orientation \( s \) for a stimulus \( S = (\theta, \alpha, \vec{\lambda}, c) \) is defined as:

\[
p(s|\theta^*, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp \left[ \kappa \cos(s - \theta^*) \right], -\pi < s < \pi
\]

(6.6)

where \( I_0(\kappa) \) is the modified Bessel function of order 0, \( \kappa = \sigma^{-2} \) is an inverse measure of statistical dispersion, and \( \theta^* \) is a value drawn from the normal distribution \( N(\theta, \sigma^2) \) over \( s \). In the simulation experiments we map the stimulus domain \([-90, 90) \text{deg}\) to \([-\pi, \pi)\). The tuning curves \( f_i(s) \) of the cells are defined as circular normal functions over \( s \):

\[
f_i(s) = g(c, \alpha) \exp \left[ \frac{\cos(s - s_i) - 1}{2\sigma_t^2} \right], -\pi < s < \pi
\]

(6.7)

where \( s_i \) is the preferred orientation of cell \( i \), \( \sigma_t \) the width of the tuning curves, and \( g(c, \alpha) \) an S-shaped function that defines how cell gain relates to the contrast \( c \) and size \( \alpha \) of a stimulus, defined as follows:

\[
g(c, \alpha) = g_{\text{max}} \left( 1 - \frac{q}{1 + c} \right)
\]

(6.8)

where \( g_{\text{max}} \) is the maximum firing rate of a cell, \( q = 1.16 \) is a constant, and \( t = a(\alpha c - b) \), with constants \( a = 1.5 \) and \( b = 1.2 \) (see Figure S1 for a plot of this function). Note that stimulus contrast and size have the same effect on response gain.

Following the DPC scheme, we compute the average response of cell \( i \) to a stimulus \( S = (\theta, \alpha, \vec{\lambda}, c) \) as follows

\[
\langle r_i \rangle = r_{\text{base}} + \int p(s|\theta^*, \kappa) f_i(s) ds
\]

(6.9)

where \( r_{\text{base}} \) is the level of spontaneous activity and \( \theta^* \) drawn from a normal distribution with mean \( \theta \) and a standard deviation \( \sigma \). In order to evaluate this integral numerically, we approximate the input distributions \( p(s|\theta^*, \kappa) \) by histograms \( \vec{\Theta} = \{\Theta_1, \ldots, \Theta_J\} \) and the tuning functions \( f_i(s) \) by histograms \( \vec{F} = \{F_{i1}, \ldots, F_{ij}\} \), both with bin centres linearly spaced in the range \([-\pi, \pi)\). Hence, we can rewrite equation (6.9) to:

\[
\langle r_i \rangle = r_{\text{base}} + \sum_j \Theta_j F_{ij}
\]

(6.10)
A population code $\vec{r}_h = \{r_{h1}, r_{h2}, \ldots, r_{hJ}\}$ representing a stimulus $S_h = (\theta_h, \alpha_h, \vec{\lambda}_h, c_h)$ is constructed by drawing responses $r_{hi}$ from Poisson distributions $P[r_{hi} | \vec{\Theta}_h] = e^{-\langle r_{hi} \rangle} \langle r_{hi} \rangle^{r_{hi}} / r_{hi}! \quad (6.11)$

### 6.4.2 Signal integration (second model layer)

The second layer of the model spatially integrates the stimulus representations in the first layer. The layer-2 population code $\vec{R}_h = \{R_{h1}, \ldots, R_{hJ}\}$ that is associated with position $\vec{\lambda}_h$ is computed as a weighted sum over the population code representations of all $N$ input stimuli:

$$R_{hi} = \sum_{k=1}^{N} w(\vec{\lambda}_h, \vec{\lambda}_k) r_{ki} \quad (6.12)$$

where $w(\vec{\lambda}_h, \vec{\lambda}_k)$ is a 2D Gaussian weight function that represents the cortical integration fields.

These weights are a function of the distance between the (center of an) integration field and the stimulus positions. Since these distances are computed in cortical space, we first compute the cortical locations (in primary visual cortex) of both the integration field center and the stimuli. For this, we use the complex log mapping from [Schira et al., 2007] to map visual field locations $\vec{\lambda}$ to cortical locations:

$$u(E, \varphi) = k \log(E e^{i\varphi} + a) \quad (6.13)$$
where $E$ and $\varphi$ are the eccentricity and angle of the polar representation of $\lambda_h$, $k = 19.2$, $a = 0.77$ are constants, and $f_a(E, \varphi)$ is a shear function that is approximated by

$$f_a(E, \varphi) = \text{sech}(\varphi) \text{sech}(\log(E/a)S_1)S_2$$

where $S_1 = 0.76$ and $S_2 = 0.18$ are constants.

The weight function that determines how strongly the response of a cell that encodes a stimulus at position $\lambda_h$ contributes to the response of an 'integration cell' associated with position $\lambda_k$, is defined by a 2D Gaussian function of cortical distance between both cells:

$$w(\lambda_h, \lambda_k) = \exp\left(-\frac{[d_{\text{rad}}(\lambda_h, \lambda_k)]^2}{2\sigma^2_{\text{rad}}} - \frac{[d_{\text{tan}}(\lambda_h, \lambda_k)]^2}{2\sigma^2_{\text{tan}}}ight)$$

where $d_{\text{rad}}(\lambda_h, \lambda_k)$ and $d_{\text{tan}}(\lambda_h, \lambda_k)$ give the radial and tangential distance between the cortical locations of the cells, and $\sigma_{\text{rad}}$ and $\sigma_{\text{tan}}$ determine the size of the integration field in the radial and tangential direction, respectively (see Figure S2 for computation of radial and tangential distances).

### 6.4.3 Signal decoding

Several of our simulation experiments require that a task response is generated. In those experiments, a Bayesian decoder is used to estimate the stimulus probability distribution $\Psi$ that is encoded in the post-integration population code $R$ associated with the target position. Subsequently, the orientation with the highest probability is interpreted as representing the most likely orientation of the target, and chosen for response. We assume that the distribution encoded in a post-integration population code is a mixture model of $K$ von Mises distributions, which we approximate by a histogram $\Psi = \sum_{k=1}^{K} \Psi_k$, with each component $\Psi_k = \{\Psi_{k1}, \ldots, \Psi_{kJ}\}$ specified as follows:

$$\Psi_{kj} = \frac{p_k}{2\pi I_0(\kappa_k)} \exp[\kappa_k \cos(s_j - \mu_k)]$$

where $I_0(\kappa)$ is the modified Bessel function of order 0, $\kappa = \sigma^{-2}$ is an inverse measure of statistical dispersion, $s_j$ is the orientation corresponding to the $j$-th bin of the histograms, and $p_k$ are the mixing proportions, $\mu_k$ the means, and $\sigma^2_k$ the variances of the mixture components.

Assuming statistical independence of cell responses, applying Bayes’ rule and assuming a flat prior over $\Psi$, we obtain the following likelihood function:
Figure 6.10: A graphical illustration of how the radial and tangential distance between an integration field and stimulus are computed. (a) Visualization of the right visual hemifield. The red marker indicates the center location of an integration field. The blue marker indicates the location of a stimulus. (b) Cortical representation of the visual hemifield. (c) The cortical distance between the integration field center and the stimulus along the eccentricity axis is defined as the radial distance. The distance along the orthogonal axis is defined as the tangential distance.
\[ P[\vec{\Psi} | \vec{R}] \propto P[\vec{R} | \vec{\Psi}] = \prod_{i=1}^{N} e^{-\langle r_{ij} \rangle \frac{R_i}{r_{ij}!}} \]

\[
\propto \prod_{i=1}^{N} \exp \left[ -r_{\text{base}} - \sum_{jk} \Psi_{kj} F_{ij} \right] \left[ r_{\text{base}} + \sum_{jk} \Psi_{kj} F_{ij} \right]^{R_i} \]

\[
= \exp \left[ -Nr_{\text{base}} - \sum_{ij} \Psi_{ij} F_{ij} \right] \prod_{i} \left[ r_{\text{base}} + \sum_{jk} \Psi_{kj} F_{ij} \right]^{R_i} \]

Taking logs, we find:

\[
\log(P[\vec{\Psi} | \vec{R}]) = \log \left( \exp \left[ -Nr_{\text{base}} - \sum_{ij} \Psi_{ij} F_{ij} \right] \prod_{i} \left[ r_{\text{base}} + \sum_{jk} \Psi_{kj} F_{ij} \right]^{R_i} \right) + C \]

\[
= \sum_{i} R_i \log \left[ r_{\text{base}} + \sum_{jk} \Psi_{ij} F_{ij} \right] - Nr_{\text{base}} - \sum_{ij} \Psi_{ij} F_{ij} + C \]

where \( C \) is a constant. The maximum likelihood parameters were estimated by using a gradient descent method. The partial derivatives of the log likelihood function are as follows:

\[
\frac{\partial \log(P[\vec{R} | \vec{\Psi}])}{\partial p_{iq}} = \sum_{i} R_i \left( \frac{1}{r_{\text{base}} + \sum_{jk} \Psi_{ij} F_{ij}} - \Psi_{ij} F_{ij} \right) \]

\[
\frac{\partial \log(P[\vec{R} | \vec{\Psi}])}{\partial \mu_{iq}} = \sum_{i} R_i \left( \frac{\kappa_{ij} \sin(s_j - \mu_q)}{r_{\text{base}} + \sum_{jk} \Psi_{ij} F_{ij}} - \kappa_{ij} \Psi_{ij} F_{ij} \right) \]

\[
\frac{\partial \log(P[\vec{R} | \vec{\Psi}])}{\partial \kappa_{iq}} = \sum_{i} R_i \left( \frac{\cos(s_j - \mu_q) - \frac{I_1(\kappa_{ij})}{I_0(\kappa_{ij})}}{r_{\text{base}} + \sum_{jk} \Psi_{ij} F_{ij}} \right) \Psi_{ij} F_{ij} \]
where $I_1(q)$ is the first derivative of $I_0(q)$ with respect to $q$.

Determining the most likely number of stimuli encoded in a population code is an open problem in the theoretical neurosciences. Therefore, here we simply assume that in the brain a mechanism exists to estimate this number, without concerning ourselves with the neural implementation of this mechanism. We approximate the most likely number of stimuli represented in a population code by decoding to mixture distributions with 1, 2, and 3 components, and computing the Bayesian Information Criterion (BIC) for each of these mixtures:

$$\text{BIC} = -2 \log L + k \log J \quad (6.25)$$

where $L$ is the likelihood value for an estimated model, $k$ the number of parameters of the associated model, and $J$ the number of data points (i.e., the length of the population code). The mixture model with the lowest BIC was chosen as the most likely stimulus distribution encoded by the respective population code.

### 6.4.4 Model parameters

The parameter settings of the model were as follows. In all simulations, the width of the tuning curves $\sigma_t$ was set to 15°, the number of neurons $J$ comprising one population code was set to 90, the spontaneous firing rate $r_{\text{base}}$ was set to 5 spikes/s, and the maximum firing rate was set to 90 spikes/s. The only parameters that varied between simulations were $\sigma_{\text{rad}}$ and $\sigma_{\text{tan}}$, which determine the integration field width in the 'radial' and 'tangential' direction, respectively. These were set to 2.5 and 1.0 mm, respectively, in all simulations, except the one in which we estimated critical regions (Figure 6.3d), where the values were set to 1.6 and 1.1mm, respectively. This difference is motivated by the observation that the human data in Figure 6.3d are from a subject with an unusually small critical spacing (approximately 0.3 times the target eccentricity).

### 6.4.5 Simulations

#### Estimation of target identification thresholds and critical spacing

Several simulation experiments involved estimation of target contrast thresholds for a tilt identification task. In those experiments, the procedure on a single trial was as follows. The target and flanker stimuli were encoded and their representations integrated, as described above. Subsequently, the post-integration population code associated with the target position was decoded to a mixture of normal distributions. The sign of the orientation associated with the peak location in the returned probability distribution was compared with the sign of the input target. Performance was considered correct if the signs were the same, and incorrect otherwise. Performance estimates were made for several target contrasts, by simulating 50 trials for each contrast. Finally, a sigmoid function with a mean $a$ and a width $b$: 
A POPULATION CODING MODEL OF VISUAL CROWDING

\[ g(x; a, b) = \frac{50}{1 + e^{-(x-a)/b}} + 50 \]  

(6.26)

was fit to these data, in order to obtain an estimate of the target contrast that yields 75%-correct performance (see Figure 6.3b for an example).

In the simulation experiments that estimated critical spacing, the above procedure was repeated to obtain 75%-correct thresholds for several target-flanker spacings. A clipped line was fit to these thresholds in order to estimate critical spacing (see Figure 6.3c for an example).

Estimation of critical regions (Figure 6.3)

Input stimuli consisted of a +/-10° tilted target and two 30° tilted flankers, positioned at opposite sides of the target. Flanker contrast and the size of both the target and flankers were set to 1. Using the procedure described above, critical spacing was estimated for the same target and flanker positions as in the psychophysical experiment by [Pelli et al., 2007].

Effect of stimulus properties on critical spacing (Figure 6.4)

The input stimuli consisted of a +/-10° tilted target, one −30° tilted flanker, and one +30° tilted flanker. Flanker contrast and the size of both the target and flankers were set to 1. Critical spacing was determined for flankers positioned along the radial axis, on opposite sides of the target.

Compulsory averaging of crowded orientation signals (Figure 6.5)

In the first simulation (Figure 6.4a), input stimuli consisted of a 0° tilted target and two flankers with 10° tilt in the first condition and 50° tilt in the second condition. The target was positioned at 2.5 deg of eccentricity. The flankers were positioned on opposite sides of the target, with a spacing of 0.5 deg of eccentricity. The contrast and size of all stimuli were set to 1. Stimuli used in the second simulation (Figure 6.4b) were similar to those used in the psychophysical experiment by [Parkes et al., 2001]: \(N\) tilted targets and 9 − \(N\) vertical flankers (first condition) or no vertical flankers (second condition), with a central position of 2.5 deg of eccentricity and a spacing of 0.5 deg between the central stimulus and surrounding stimuli. The contrast and size of the stimuli were set to 0.5. On a single trial, the post-integration population code associated with the central stimulus position was decoded to a unimodal stimulus distribution. The sign of the orientation with the highest probability was compared with the sign of the target. If they were the same, performance on that trial was considered correct. We measured performance over 100 trials for varying target tilts. Based on these data, 75%-correct performance thresholds were determined. This procedure was repeated for different values of \(N\).
**Foveal-peripheral flanker anisotropy (Figure 6.6)**

Input stimuli consisted of a +/-10° tilted target without a flanker (condition 1), with a 30° tilted foveal flanker (condition 2), or a 30° tilted foveal flanker (condition 3). Flanker contrast and the size of both the target and flankers were set to 1. For all three conditions, 75%-correct target contrasts were estimated for a range of target-flanker spacings. Threshold elevations $TE_{\text{foveal}}$ and $TE_{\text{peripheral}}$ were defined as described in the main text.

**Model response to visual contours (Figure 6.7)**

The input stimuli consisted of a set of oriented bars, comprising three contours within a field of randomly oriented bars. The circle contour consisted of 35 equally spaced segments, was centered at (0,10) degrees of eccentricity and had a radius of 4 degrees of visual angle. The other four contours consisted of 23 line segments each, with a spacing of 0.7 degrees of visual angle between every two neighboring segments. The randomly oriented line segments were placed on a grid with a radius of 18 degrees of eccentricity and a grid spacing of 2.0 degrees. The contrast and size of all line segments was set to 0.8.