

Investigating shape perception by classification images

Ilmari Kurki

Jussi Saarinen

Aapo Hyvärinen

Institute of Behavioral Sciences,
University of Helsinki, Helsinki, Finland
Department of Computer Science and HIIT,
University of Helsinki, Helsinki, Finland



Institute of Behavioral Sciences,
University of Helsinki, Helsinki, Finland

Institute of Behavioral Sciences,
University of Helsinki, Helsinki, Finland
Department of Computer Science and HIIT,
University of Helsinki, Helsinki, Finland

Radial frequency (RF) patterns are circular contours where the radius is modulated sinusoidally. These stimuli can represent a wide range of common shapes and have been popular for investigating human shape perception. Theories postulate a multistage model where a global contour integration mechanism integrates the outputs of local curvature-sensitive mechanisms. However, studies on how the local contour features are processed have been mostly based on indirect experimental manipulations. Here, we use a novel way to explore the contour integration, using the classification image (a psychophysical reverse-correlation) method. RF contours were composed of local elements, and random “radial position noise” offsets were added to element radial positions. We analyzed the relationship between trial-to-trial variations in radial noise and corresponding behavioral responses, resulting in a “shape template”: an estimate of the contour parts and features that the visual system uses in the shape discrimination task. Integration of contour features in a global template-like model explains our data well, and we show that observer performance for different shapes can be predicted from the classification images. Classification images show that observers used most of the contour parts. Further analysis suggests linear rather than probability summation of contour parts. Convex forms were detected better than concave forms and the corresponding templates had better sampling efficiency. With sufficient presentation time, we found no systematic preferences for a certain class of contour features (such as corners or sides). However, when the presentation time was very short, the visual system might prefer corner features over side features.

Introduction

Mechanisms of shape perception

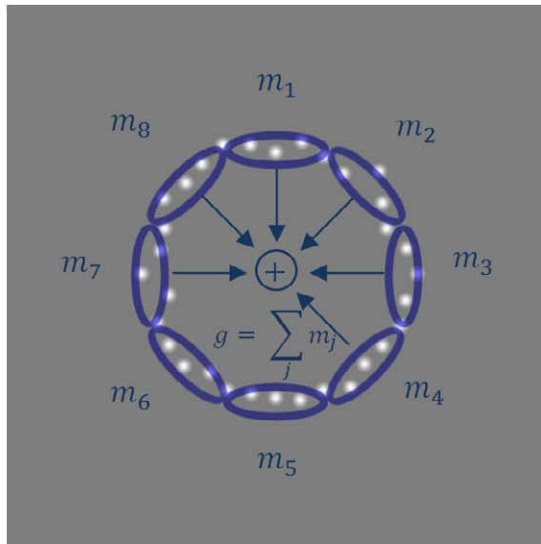
The overall shape of an object is an important cue for object recognition, as many objects are perfectly recognizable from the shape cues alone, e.g., in the case of silhouettes and line drawings. In general, shape integration in humans seems to be fast and largely invariant to such low-level physical properties of stimuli as contrast or gross size.

Neural computations in shape integration have been under intensive research. Shape perception has various stages; lower levels analyze the local contour parts and features feeding the signals to higher levels, which are sensitive to global shapes. The first stage in the shape analysis takes place already in the first cortical visual area V1. V1 neurons are capable of extracting local contour information, being sensitive to stimulus orientation, spatial scale, and polarity (Campbell & Robson, 1968; Hubel & Wiesel, 1959). However, V1 receptive field sizes are very small, and typically V1 neurons have unimodal orientation tuning, implying that all but the simplest shape discriminations require higher level integration of responses over a population of neurons.

Many computational models suggest the existence of an intermediate stage of analysis, where more complex local *contour features*, such as local angles and curves, are extracted (see Figure 1). Output of these mechanisms could serve as shape primitives for

Citation: Kurki, I., Saarinen, J., & Hyvärinen, A. (2014). Investigating shape perception by classification images. *Journal of Vision*, 14(12):24, 1–19, <http://www.journalofvision.org/content/14/12/24>, doi:10.1167/14.12.24.

A



B

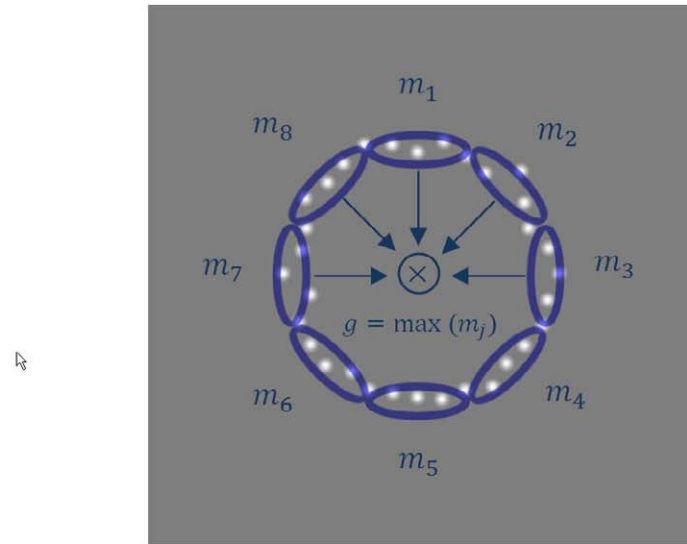


Figure 1. Multistage model for contour integration. The first stage (represented by the blue oval) analyzes the local contour curvature. Here, we assume that this stage can be represented by matching the contour element locations and contour templates. The global shape analysis stage (A) then integrates (sums) these responses globally. Another possible scheme (B) is probability summation, where shape integration is based not on the global analysis stage but on maximum local responses.

later global stages in shape analysis. Psychophysical evidence for mechanisms selective to contour curvature has been obtained by adaptation paradigms (*shape-frequency aftereffect* and *shape-amplitude aftereffect*; Gheorghiu & Kingdom, 2007, 2009) as well as by polarity-specific integration of the contour information (Bell, Gheorghiu, Hess, & Kingdom, 2011). These models are supported by the neurophysiological evidence for neurons specialized for local contour feature encoding. Neurons in macaque area V2 seem to code local orientation combinations (Anzai, Peng, & Van Essen, 2007; Willmore, Prenger, & Gallant, 2010). V4 neurons often have bimodal orientation tuning (David, Hayden, & Gallant, 2006) and respond well to curvilinear shapes (Gallant, Braun, & Van Essen, 1993; Gallant, Connor, Rakshit, Lewis, & Van Essen, 1996; Nandy, Sharpee, Reynolds, & Mitchell, 2013). Many neurons in V4 are tuned to contour features or specific combinations of contour features, such as convex features in the upper left with a concavity in the bottom, but are not necessarily selective for specific global shapes (Connor, Brincat, & Pasupathy, 2007; Pasupathy & Connor, 1999, 2001). The last global stages of shape perception are assumed to take place in a network that consists of higher ventral visual areas such as human V4 (Wilkinson et al., 2000) and LOC (see, e.g., Grill-Spector, Kourtzi, & Kanwisher, 2001).

The major empirical support for the third stage, global shape mechanisms, comes from psychophysical contour integration studies. Radial frequency (RF) patterns (Wilkinson, Wilson, & Habak, 1998)—i.e., closed circular contours where the radius is modulated by a sinusoidal function of the polar angle (see Figure 2 and Methods)—have been a popular stimulus in shape processing studies. By varying the radial frequency and phase of the pattern, it is possible to generate many geometrical shapes, such as a triangle (RF3), a square/diamond (RF4), a pentagon (RF5), and so on. Even more complex shapes can be represented by a sum of elementary radial frequencies. Observers have been shown to have extreme sensitivity to detecting the radial frequency modulation, as the discrimination thresholds versus perfect circles were near the hyperacuity range (Wilkinson et al., 1998).

Spatial integration of radial frequency modulation has been studied by varying the number and extent of contour features (Bell & Badcock, 2008; Hess, Wang, & Dakin, 1999; Ivanov & Mullen, 2012; Loffler, Wilson, & Wilkinson, 2003; Schmidtmann, Kennedy, Orbach, & Loffler, 2012). Many studies suggest global or “optimal” integration, i.e., integration of feature information in a mechanism that sums information from local contour part detectors in a manner that is nearly linear, rather than probability summation from independent detectors (as discussed later in more detail).

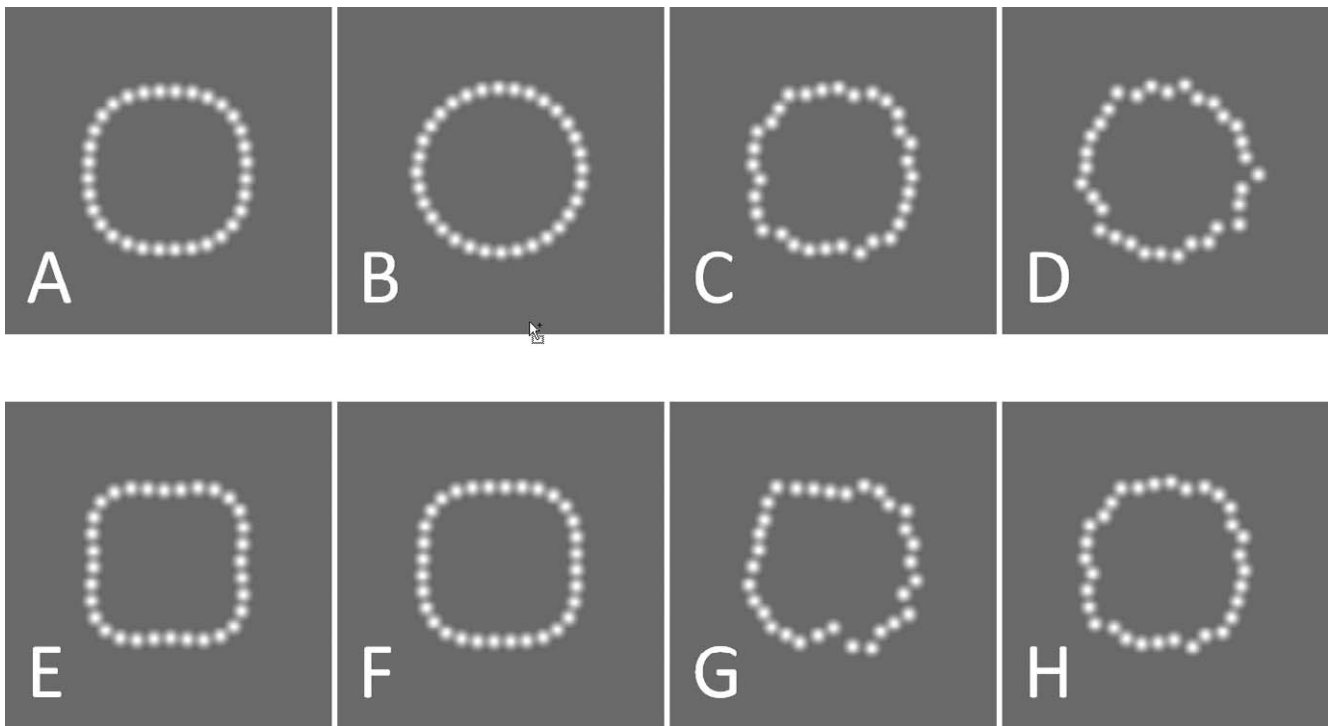


Figure 2. Stimuli were composed of DoG elements. The RF4 pattern used here can be thought to have four (convex) corner features, and depending on amplitude four side features that can be either convex (A), straight (F), or concave (E). Convex and concave side processing was tested in separate experimental conditions. (A) The target RF4 pattern with convex sides (convex condition) without position noise. (B) A perfect circle with no RF modulation; the baseline shape in the convex condition. (C) The target pattern with convex sides with noise. (D) The baseline shape in the convex condition with noise. (E) The target RF4 pattern with concave sides (concave condition) without noise. (F) The baseline shape in the concave condition, a “square” shape with straight edges. (G) The target concave RF pattern with noise. (H) The baseline shape in the concave condition with noise. The task in the convex conditions was to discriminate between instances of (C) and (D); in concave conditions, the task was to discriminate between instances of (G) and (H).

An important paradigm has used a circular stimulus that contains both radially modulated and nonmodulated parts (Hess et al., 1999; Loffler et al., 2003; Schmidtman et al., 2012). Strong improvement of discriminability as a function of modulated contour length has been found, at least when the underlying RF patterns have low radial frequency ($RF < 10$). This has been interpreted as evidence for a global mechanism that sums information across the whole contour, as probability summation alone cannot explain this phenomenon. However, a study where contour parts were presented in isolated spatial windows (Mullen, Beaudot, & Ivanov, 2011) reported that the discrimination thresholds for a single feature were close to the discrimination threshold for the whole pattern (but see Dickinson, McGinty, Webster, & Badcock, 2012). Moreover, global integration has been reported to be dependent on low-level factors such as stimulus contrast (Ivanov & Mullen, 2012).

Similar global shape mechanisms have been proposed to support the perception of global forms in Glass patterns, where strong summation of local stimulus information has been reported (Wilson &

Wilkinson, 1998; Wilson, Wilkinson, & Asaad, 1997; but see Kurki, Laurinen, Peromaa, & Saarinen, 2003). However, the mechanisms are not necessarily the same (Bell & Badcock, 2008).

Processing of contour features

In addition to the global processing hypothesis, another central question has been whether the efficiency of detecting certain contour features (such as corners, convex sides, and concave sides) is higher than that for other contour features. Hess et al. (1999) reported that masking the square-like RF4 pattern with orientation-filtered luminance noise increased the threshold more when the orientation band in the noise matched the orientation of the side contours in the RF4 pattern, compared with the corners. This was interpreted as evidence that the visual system relies mostly on the side contour features. However, results from lateral RF masking studies have been interpreted as showing that curvature maxima (corners) features are

more important than sides for shape perception (Habak, Wilkinson, Zakher, & Wilson, 2004; Poirier & Wilson, 2007).

Another line of research has investigated possible differences in processing of convex and concave features. Concavity and convexity are defined by curvature relative to the object center (here, the middle of the screen), and the visual system may use them as a cue in object segregation. It has been reported that figure/ground segregation prefers an interpretation where figures have convex sides (Kanizsa, 1979). Also, visual sensitivity has been reported to be higher in convex parts of the shape (see, e.g., Bertamini & Mosca, 2004; Driver & Baylis, 1995; Loffler et al., 2003; but see Barenholtz, Cohen, Feldman, & Singh, 2003). Moreover, functional magnetic resonance imaging studies have suggested that convex contours may elicit stronger responses in higher ventral areas such as lateral occipital complex (LOC) (Haushofer, Baker, Livingstone, & Kanwisher, 2008).

Classification images and contour integration

Here, we used the classification-image or psychophysical reverse-correlation method (Beard & Ahumada, 1998; Eckstein & Ahumada, 2002; Levi & Klein, 2003; Li, Levi, & Klein, 2004; Murray, 2011; Solomon, 2002) to investigate shape integration. The method enables direct estimation of the contour parts and features processed by the visual system in a shape discrimination task. It is based on analyzing the trial-to-trial relationship between noisy stimulus values that were shown in the experiment and corresponding behavioral responses. We use a novel design, utilizing the “position noise” paradigm (Li et al., 2004) that has been previously used to study perceptual learning and amblyopia in a Vernier offset task. Observers’ task was to discriminate the target radial frequency shape from a “baseline” (circle or square) shape. Shape stimuli were composed of local difference-of-Gaussian (DoG) elements. The radial positions of the elements were sinusoidally modulated (see Figure 2). To enable classification-image estimation, we further added a slight random offset (position noise) to the radii of the elements. These offsets were independent and randomly generated anew in every trial. Addition of radial noise makes the presence of features that define the shape (corners; convex or concave sides) stochastic, i.e., they are present only on average and not on every stimulus instance. Thus, the visual system needs to compute a measure of similarity between noisy stimuli and the target shapes. Applying the radial position noise paradigm allowed us to use very low-dimensional and efficient stimuli and reduce the number of trials per classification image to around 2,000.

The classification-image method is based on the linear template-matching model in signal-detection theory (Peterson, Birdsall, & Fox, 1954), also known as the noisy cross-correlator model. In this straightforward model, the outcome of the visual processing is modeled by linear cross correlation between the stimulus information and an internal “shape template” that determines which parts and features of the stimuli are sampled and processed in the visual system (see Methods). In any trial, noise can either increase or decrease the radial modulation of each contour feature (noise radial offsets match the radial modulation or are opposite to the radial modulation).

We assume that the visual system makes the perceptual decision by first analyzing the local contour information (radial offsets) and then integrating (cross correlating) these data in an internal shape template¹ for the behavioral decision. The relationship between the known stimuli, the unknown internal weights, and the observed responses can be represented as a regression problem, which can be solved using the generalized linear model (GLM) probit regression.

Template-matching models have previously been found to explain many key features in low-level visual experiments, especially the detectability of a stimulus with the presence of external noise (Li et al., 2004; Neri & Levi, 2006; Pelli & Farell, 1999; Solomon & Pelli, 1994). The shape radial position noise paradigm allows comparison of human performance with an ideal observer that is able to use the shape feature information in an optimal way. Cross correlating the stimulus information (i.e., radial offsets) with the target stimulus profile is equal to obtaining the likelihood ratio of the target presence and is thus an ideal detection and discrimination strategy (Green & Swets, 1974).

The classification-image method enables a direct estimation of critical features in shape discrimination, complementing findings obtained from indirect masking techniques such as lateral contour masking. It is able to provide more detailed information on relative weight and type of contour features used, as opposed to relatively coarse addition/subtraction of features. Further, we show that it is possible to test quantitative models on how the contour features are integrated by comparing the observed and predicted performance of a noisy linear model and building a model observer and comparing its responses to the same stimulus shown to human observers.

Research questions

We derive classification images for radial frequency patterns, aiming to shed light on contour shape integration mechanisms. We use a square-like RF4

pattern, as it a relatively simple and recognizable pattern and is both horizontally and vertically symmetrical. By changing the RF4 amplitude, it is possible to create circular (zero RF amplitude) and square patterns with convex sides (medium amplitude) and finally square patterns with concave sides (high amplitude). First, we test how feasible a template-matching model for shape integration is by comparing how well a linear template-matching model that uses the estimated template (classification image) can predict the observed discrimination efficiency (i.e., squared detectability index divided by an ideal observer's detectability index). Following Murray, Bennett, and Sekuler (2005), we show that it is possible to directly predict the discrimination efficiency from a classification image that has been estimated with GLM. A good match between predicted and observed efficiency suggests that a template-matching model can quantitatively explain (at least to an approximation) the contour integration performance on the task.

Second, we estimate the relative efficiency of contour detectors that process the corner features and different kinds of side features. The RF4 pattern can be thought to have eight contour features: four corners (convex maxima) and 4 side features. Depending on the RF amplitude of the pattern, the sides or RF troughs are either convex—i.e., curved outwards (low-amplitude RF4 pattern)—or concave (high-amplitude RF4 pattern; see Figure 1), i.e., curved inwards. In the RF4 pattern, RF peaks form corners or convex maxima which are at the diagonals of the pattern.

In the first and the last experiments, the target pattern sides have convex curvature or “convex RF” with respect to the center of the display. The baseline pattern was a circle with constant curvature. In the second experiment, we investigate the discrimination of shape with concave sides, “concave RF,” which is compared with a baseline shape with straight sides. We then estimate the efficiency of both classes of contour feature discrimination by estimating the average local template efficiency by means of an ideal observer. The stimuli were designed so that the relative shape difference between the baseline and target shape was identical in both concave and convex conditions. Thus the ideal observer was exactly the same.

Third, we investigate the efficiency of processing contour features and how they are summed at the global stage (see Figure 1). In addition to standard classification images, i.e., internal weighting of stimulus elements, we analyze weighting of each of the eight contour features, corresponding to half cycles of RF modulation (i.e., a peak or a trough). Classification images by themselves could show lack of global processing: If classification images contain just one or a few contour parts, this can be used as a strong argument against global contour integration. On the

other hand, if contour integration is a global process, the classification image should contain all contour parts. However, even if a classification image has all the contour parts, that does not necessarily imply “optimal” linear integration; another possibility is that the visual system uses a nonlinear, maximum-of-output kind of integration from a single best-fitting contour feature detector. This “probability summation” (see, e.g., Loffler et al., 2003; Mullen et al., 2011) would still use all contour features at the early stage, but in a probabilistic manner; the best-matching detector varies from trial to trial, and outputs from other detectors are not used. In this scheme, using the standard linear method to analyze the data would still result in a classification image containing signatures from all of the contour detectors over all trials, even when their responses are not summed together linearly (see also Tjan & Nandy, 2006). Therefore, we use the classification-image data to test different integration models. Following the multistage contour integration model (Figure 1), we assume that the global integration stage integrates responses of local contour feature detectors (instead of single contour elements). Responses to contour features can be approximated from linear classification-image data. We test the global/local processing by feeding the stimuli to a model that sums the outputs of local contour detectors with Minkowski summation. The Minkowski summation model (see, e.g., Graham, 1989) is a generalization of the maximum-of-outputs rule, where a summation parameter controls the nature of summation from linear to maximum-of-outputs. We test the model with different summation parameters from linear to maximum-of-outputs to estimate which parameter predicts the observer responses the best.

In the last experiment, the temporal dynamics in shape integration are investigated using a convex side stimulus similar to one in the first experiment, but with very brief stimulus presentation time of 20 ms: “short-duration RF.” We wanted to study contour integration in a time-limited condition, as it is often thought that short stimulus durations allow only a coarse analysis of more complex stimulus features.

Methods

Participants

Five observers participated in the study. IK is one of the authors, and VS is a nonnaïve observer. All of the observers had normal or corrected-to-normal visual acuity and extensive experience in psychophysical experiments. The experimental procedure was in accordance with the Declaration of Helsinki and

approved by the Ethics Committee of the Institute of Behavioral Sciences, University of Helsinki. All subjects were volunteers and gave their written consent.

Apparatus and stimuli

Experiments were conducted in a dimly lit room. Stimuli were created using Matlab 2008b (Mathworks Inc., Natick, MA) using custom software and the PsychToolbox 3 extensions (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). Stimuli were displayed on a Mitsubishi Diamond Pro 2070 SB monitor using a Cambridge Research Systems (Cambridge, UK) ViSaGe stimulus generation system that enables 15-bit luminance resolution.

The stimuli were generated by using a two-step procedure. The radius r_θ of the radial frequency pattern (modulation around the mean radius) is defined in polar coordinates by

$$r_\theta = a \sin(f\theta + \phi), \quad (1)$$

where a is the amplitude of the radial modulation, θ is the polar angle, f is the radial frequency (here: 4), and ϕ is the radial phase. Radial frequency patterns were presented always at the same phase. A virtual radial frequency contour where elements were later placed was computed first. Then the locations of the elements were chosen by sampling the virtual contour at 32 locations, keeping the distance between the elements along the contour constant. Generation of the stimulus is explained in more detail in Appendix A. This method ensures that the radial modulation of the contour does not change the local element density, as would be the case if the element positions had been sampled by using a constant polar angular separation. Let r_T be the vector that contains the target shape radii and t a binary vector that encodes the target presence on each trial of the experiment. Radial modulation of the stimulus elements around the baseline shape trial k was coded by vector r_k (Equation 2), where n_k denotes a vector of random radial position noise that is added to the radii of elements. This position noise was independent Gaussian random noise ($SD = 6.2^\circ$), generated using the “randn” function of Matlab.² Thus, we get

$$r_k = t(k)r_T + n_k. \quad (2)$$

Finally, the baseline radius r_b was added. For experiments with convex shapes, this was the mean radius (i.e., perfect circle). For concave-shape experiments, the baseline shape was a square RF pattern (see Figure 2). Thus, radii for the final shape r_s were given by

$$r_s = r_k + r_b. \quad (3)$$

The global position of the pattern on the screen was then randomized by displacing the center of the pattern

by an offset drawn from a uniform random distribution.

The mean radius of the pattern was 1.5° . Elements were difference-of-Gaussian (DoG) envelopes. The standard deviation of the center was 5.6 arcmin and of the surround, 16.9 arcmin. Thus the surround was rather large and had a low peak contrast (see Figure 2). The backward mask stimulus was a $6 \times 6^\circ$ random (spatial) luminance noise patch, which was spatially convoluted with the element DoG envelope. Root-mean-square contrast of the noise was 25%.

Procedure

A one-interval shape discrimination task with a four-point confidence rating procedure (see also Li, Klein, & Levi, 2006; Li et al., 2004) was used. The rationale for using a four-point rating procedure is that it provides more information about outcome of perceptual processing compared to yes/no answers; effectively measuring with a 4-point scale leads to a smaller estimation error in classification images (Murray et al., 2002). The trial started with a blank get-ready period (250 ms), followed by the fixation screen (250 ms). The stimulus was then displayed (200 ms, except for the short-duration experiment, where it was 20 ms), followed by a backward luminance noise mask (500 ms). The observers' task was to indicate whether the noisy RF stimulus resembled more the target (RF) shape or the baseline (circle/square) shape. The observer gave the confidence rating response of seeing the target versus the baseline (sure target, unsure target, unsure baseline, sure baseline). An audio feedback was provided immediately for incorrect answers. The feedback was not dependent on the confidence level. After that, the next trial started immediately. The shape amplitude of the RF target stimulus was adjusted so that the proportion of correct responses was about 75%, using the QUEST procedure (Watson & Pelli, 1983). Experiments were conducted in blocks of 100 trials, with different conditions in balanced order. Before starting the experiments, subjects practiced at least two blocks for every condition.

For every trial, the radial position noise sample and the response of the observer were saved for later analysis. Observers ran between 2,000 and 3,000 trials in each condition. Final thresholds were computed from the mean of the last 20 experiment blocks (last 10 experiment blocks in the short-duration condition, where some thresholds were initially very high).

Linear observer model and data analysis

We analyzed both standard classification images and contour feature weights.

We assume that the response variable (visual system’s internal response) in trial k is a function of the linear cross correlation g_k between the radial modulation r_k of shape elements and the internal shape template w , which determines how information at each location of the contour is used for the perceptual decision (here, we assume that w has a unit length, i.e., $w^T w = 1$). As the baseline shapes’ (circle or a square RF pattern) radii r_b were added to all stimuli within an experimental condition, and since the cross-correlation operation is linear, the cross correlation with the template adds just a constant in all trials. We can therefore subtract the baseline shape r_b from the stimulus r_s , i.e., assume that only modulations around the baseline shape are used. Lastly, to represent the internal noise in the visual system, a normally distributed scalar e_k is added to the cross correlation between the shape template and the stimulus radial information. Thus, we have

$$g_k = r_k^T w + e_k. \quad (4)$$

The observer responds with four confidence ratings by comparing the response with a set of internal criteria. An observer’s confidence rating response a_k on trial k is l when (if and only if) response g_k falls between criteria $c = [-\infty, c_2, c_3, c_4, \infty]$ c_a and c_{a+1} :

$$a_k = l, \text{ iff } c_l \leq g_k < c_{l+1}. \quad (5)$$

Since the target shape r_T was not varied in an experimental condition and its amplitude was almost constant, we approximate that the match between the template and the target is a constant. The expectation for a positive “target seen” response, i.e., a response exceeding criteria c , can be written as Equation (6). The term $n_k(i)$ represents the position noise of the i th element on the k th trial. Noise is weighted by elements 1–32 of vector of covariates w . Since the match between the template and the target is a constant, we use the 33rd regressor β to represent the constant target presence response (coded as a dummy variable) in those trials where the target was present. This gives

$$E(y > c) = \phi \left(w_1 n_k(1) + w_2 n_k(2) + w_3 n_k(3) + \dots + w_{32} n_k(32) + \beta t(k) - c \right), \quad (6)$$

where ϕ is the standard cumulative normal distribution function. This formulation allows estimation of a classification image using only presented noise masks, in both target and baseline trials without the target pattern and four-point observer responses. This ensures that any patterns in classification images that resemble the target cannot be caused by the target stimulus. We use the generalized linear model (GLM) to estimate the weights w (the classification image). More specifically, we use the ordered probit regression. Recent studies

have shown that using maximum-likelihood GLM can lead to a smaller estimation error than the standard weighted sums method (Knoblauch & Maloney, 2008; Murray, 2011). The Matlab Statistics Toolbox function “`mnrfit`” was used for GLM. This procedure returns an estimate of the template up to a scale factor, which is proportional to internal noise (Knoblauch & Maloney, 2008; Murray, 2011).

Finally, we tested the statistical significance for shape templates as well as the template difference (for the difference between concave and convex pattern) using a nested hypothesis likelihood ratio test (Knoblauch & Maloney, 2008, 2012). In the template significance test (T1), we compared an unconstrained GLM model with classification-image weights as well as a regressor for the stimulus presence to a constrained model with just one regressor for the stimulus presence. For the template difference test (T2), we compared an unconstrained model in which a single classification image is estimated on both conditions to a model that had two separate classification images for the convex and concave conditions. A likelihood ratio test can then be used to reject the hypotheses that the extra regressors for the classification-image weights do not improve the likelihood of the model in a statistically significant way (T1) and that extra regressors for two separate classification images (for each condition), instead of one, do not improve the likelihood in statistically significant way (T2)—i.e., to test if the classification images in the convex and concave conditions were significantly different.

Estimating template absolute efficiency

Absolute efficiency F_o^2 is a measure of observed discrimination performance relative to the ideal observer, the maximum theoretical limit. It can be obtained by comparing the ratio of the observer’s discrimination performance d'_o and the ideal observer’s discrimination performance d'_* with the same stimuli, here the stimulus at 75% threshold shape amplitude:

$$F_o^2 = (d'_o/d'_*)^2. \quad (7)$$

We asked how well the classification image is able to predict the observed absolute efficiency (see also Murray et al., 2005). The performance of a linear observer is dependent on two factors: the sampling efficiency or the match between the template and the stimulus, and the amount of internal noise.

To quantify the sampling efficiency ρ_w^2 of a shape template, we compare how well the relative weighting of the estimated template matches the ideal weighting. This was done by comparing the normalized classification image \hat{w}_n , normalized so that $\sqrt{\hat{w}_n^T \hat{w}_n} = 1$, and the ideal template w_* , which is the normalized target

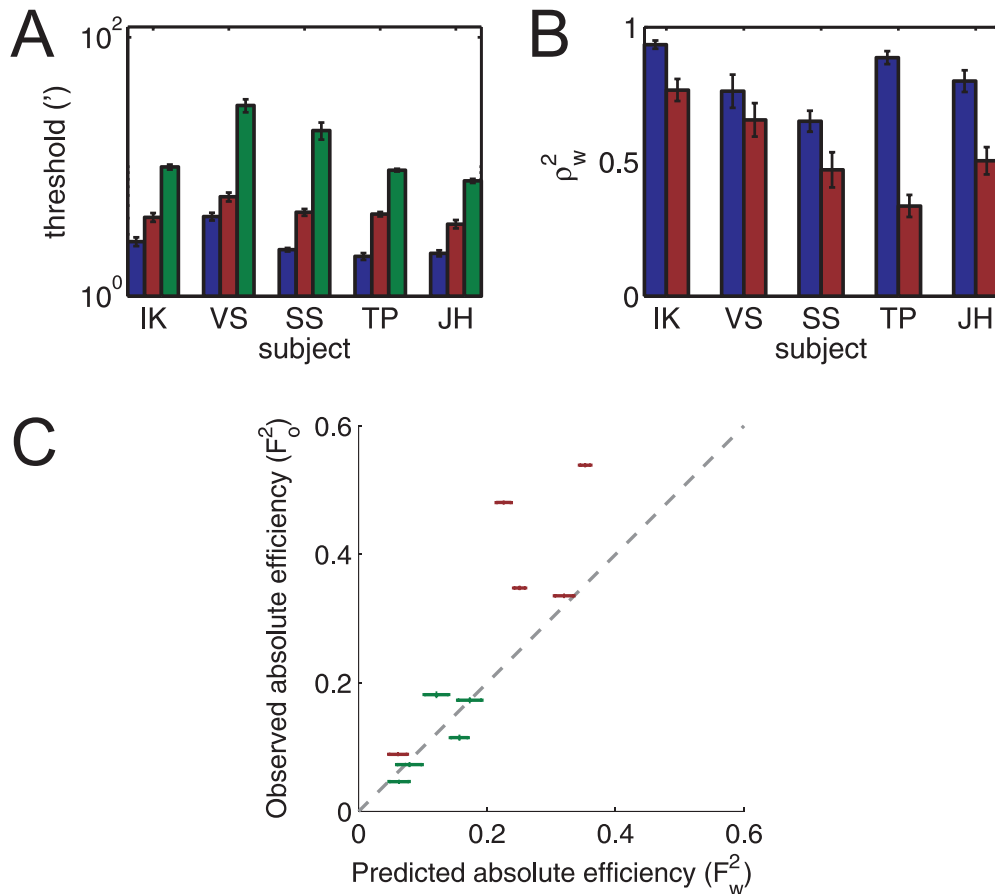


Figure 3. (A) Discrimination thresholds for all subjects and conditions in arc minutes ($'$). Thresholds are lowest in the convex and short-duration conditions. (B) Template sampling efficiency for different stimuli and conditions (blue: convex; red: concave). Error bars represent one standard error. Sampling efficiency is higher in the convex and lower in the concave condition. (C) Comparison of predicted absolute efficiency for a linear observer using classification image and the observed absolute efficiency. Correlation is high ($r = 0.89$). The average predicted efficiency is about 10% lower than the observed.

profile (see dashed lines in Figure 4):

$$\rho_w^2 = (\hat{w}_n^T w_*)^2. \quad (8)$$

The expected absolute efficiency F_w^2 for a linear observer with sampling efficiency ρ_w^2 , internal noise standard deviation $\hat{\sigma}_i$, and external position noise standard deviation $\hat{\sigma}_e$ (Burgess, Wagner, Jennings, & Barlow, 1981; see also Eckstein, Pham, & Shimozaki, 2004; Murray et al., 2005) is

$$F_w^2 = \frac{\rho_w^2}{\left(1 + \frac{\hat{\sigma}_i^2}{\hat{\sigma}_e^2}\right)}. \quad (9)$$

From the definition of the probit model and (internal) shape template, it follows that for a linear observer, the expected GLM classification image is proportional to the observer's shape template divided by the amount of internal noise (see also Knoblauch & Maloney, 2008; Murray, 2011). Thus we estimate $\hat{\sigma}_i$ from the length of the nonnormalized classification image \hat{w} :

$$\hat{\sigma}_i = \frac{1}{\sqrt{\hat{w}^T \hat{w}}}. \quad (10)$$

Computer simulations (results not shown) were used to test the method, and we found that it is reasonably accurate when the internal-to-external noise ratio is within the range of 0.5–2, as was found to be the case here. A good match between the predicted efficiency F_w^2 and observed efficiency F_o^2 suggests that the linear model is able to explain the shape discrimination performance.

Feature weight analysis

We investigated the relative weight that the observer gives for contour features: corners and sides. More specifically, we divided the target RF4 pattern of 32 elements into eight parts, each consisting of one contour feature.³ The contour features here each contained three elements, as every fourth element was

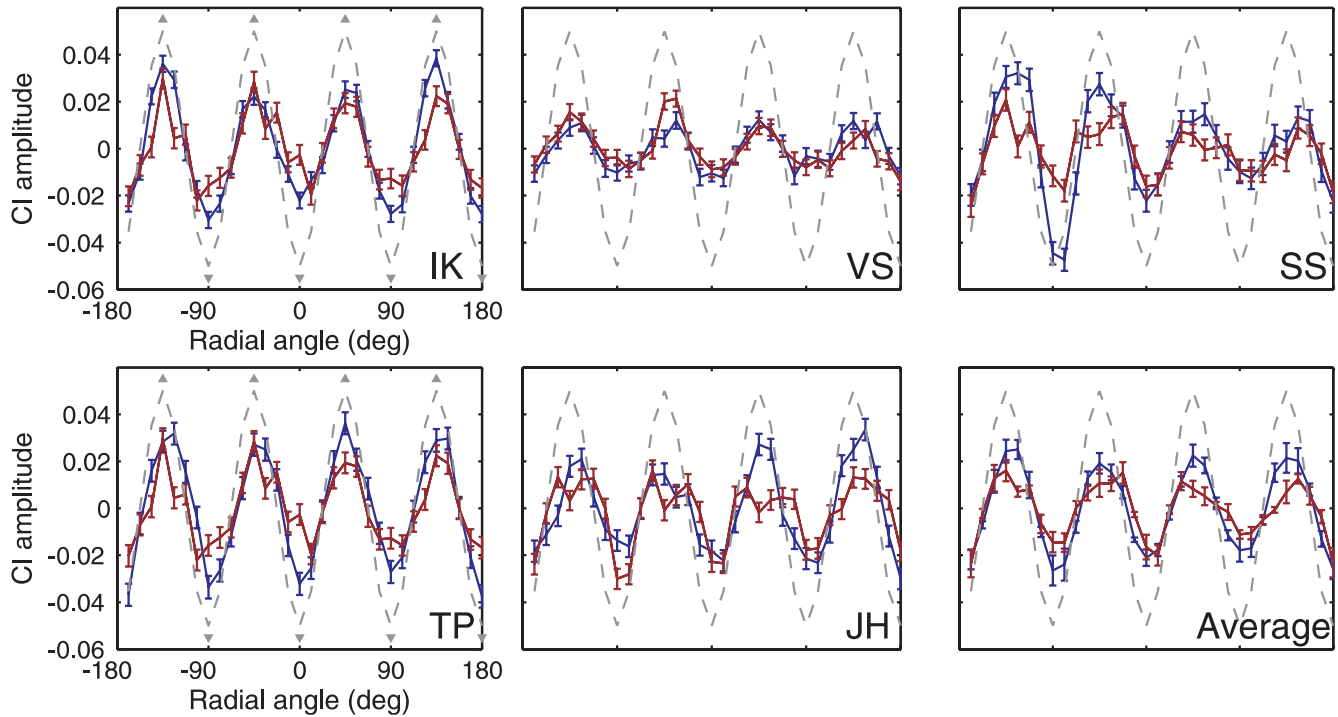


Figure 4. Classification images for a convex (blue) and concave (red) RF4 pattern plotted against radial angle. The estimated internal template (element information weighting) is plotted against the radial angle of the elements. RF peaks correspond to corners of the pattern; RF troughs correspond to sides (convex/concave) of the pattern. The target pattern, which is also an ideal template (not normalized in these plots), is shown by a dashed line. Classification images are plotted in arbitrary scale; error bars represent 1 standard error of the mean. Panels represent different subjects; the average classification image is on the bottom right.

a zero crossing and thus did not contain radial modulation. Let $\mathbf{p} = [1, 4, 8, \dots, 32]$ be the vector containing element position indices. The contour feature weight $\mathbf{h}(i)$ of the i th contour feature was computed by cross correlating the elements of the normalized classification image $\hat{\mathbf{w}}_n$ at the locations of the j th feature (starting from location $\mathbf{p}(j)$) and the corresponding locations on the ideal shape template \mathbf{w}_* :

$$\mathbf{h}(i) = \sum_{j=\mathbf{p}(i)}^{\mathbf{p}(i)+2} \left(\hat{\mathbf{w}}_n(j) \mathbf{w}_*(j) \right). \quad (11)$$

An ideal detector would weight all features with 1/8 weights. However, it should be noted that as the normalization operation takes a sum of all contour features, this analysis does not necessarily reveal the true efficiency of a single feature detector independently of others. Thus, the absolute weighting is not meaningful. The goal here is to compare classes of feature detectors.

Nonlinear integration analysis

As an alternative to the linear integration model (Equation 4), we consider the Minkowski summation

model (Figure 1) with nonlinear integration of contour feature detector outputs. The first feature detector stage has eight feature detectors that code the local curvature information (corners and sides) by cross correlating the local stimulus information in an area spanning three elements. The response of the i th local feature detector $\mathbf{f}(i)$ to pattern \mathbf{r}_k , which was presented in the k th trial, is then

$$\mathbf{f}(i) = \sum_{j=\mathbf{p}(i)}^{\mathbf{p}(i)+2} \left(\hat{\mathbf{w}}_n(j) \mathbf{r}_k(j) \right). \quad (12)$$

The global response g_k in the k th trial is then computed by taking the Minkowski nonlinear sum of local matches,

$$g_k = \left(\sum_j \mathbf{f}(j)^\gamma \right)^{1/\gamma} + e_k, \quad (13)$$

where γ is the parameter controlling the nonlinearity of summation. A value of $\gamma = 1$ equals linear integration of contour features, being equivalent to the linear model (Equation 4). With increasing γ , the best-matching feature becomes more dominant. A value of $\gamma = \infty$ means a winner-takes-all scheme, where the response is determined by just the best-matching feature. The

internal response is translated to the observer's response using Equation 5.

Nonlinear summation was investigated by using the same stimuli shown to subjects as an input to the model (Equations 12 and 13) and comparing how well the model explained the observer responses. Weight functions of the local mechanisms were taken from the linear classification image. As observer confidence rating scale responses were categorical in nature, we thresholded the model responses so that the number of response categories and their relative frequencies were matched, allowing the use of rank (categorical) correlation as a metric for the goodness of fit (model–observer correlation). More specifically, we (1) computed the cumulative distribution for observer responses from the relative confidence rating frequencies, using a standard signal-detection approach; (2) thresholded the model responses to category responses, seeking the thresholds by matching the quantiles in the model's response distribution with psychophysical data; and (3) quantified the match between the categorical model and observer responses by using Spearman's rho rank correlation. This procedure does not require model responses to be normal, and moreover, it is not dependent on an observer's internal criteria (like simple percentage of matches). We computed the matches separately for target-present and target-absent data. For estimating the best fits, we took the mean of target-present and target-absent fits.

We tested the accuracy of the γ estimation method using simulations (see Appendix B). We found it to be reasonably good near the range of best-fitting γ , although some downward bias was present at higher values.

Results

Performance

We found that mean discrimination thresholds (Figure 3A) for different contour shapes varied systematically, being highest for the short-duration contour (14.4'), next highest for the concave contour (4.3'), and lowest for the convex contour (2.5'). The difference in thresholds across all conditions was statistically significant across the subjects, repeated-measures ANOVA $F(2, 8) = 9.55$, $p = 0.008$. The difference between concave and (long-duration) convex conditions was also significant across the subjects, $t(4) = -10.39$, $p = 0.005$.

We then computed the absolute efficiencies, comparing human performance with the ideal observer's performance. This is shown in Figure 3B. The average absolute efficiency was highest for the convex contour

(0.36) and somewhat lower for the concave contour (0.12), $t(4) = -3.72$, $p = 0.02$.

Classification images

Classification images represent the estimated weight that the visual system places on each contour element for perceptual decisions. These are represented as graphs where the y-axis displays the estimated internal shape template weight of each element, plotted against the radial angle (position). Target RF pattern modulation (which is also an ideal template) is plotted as a reference. In this representation, RF peaks correspond to the corner contour feature locations, and troughs to side contour feature locations. Classification-image weights in Figure 4 are nonnormalized regression weights. Individual statistics are presented in Table 1.

Statistical testing with short-duration classification images showed that not all templates were statistically significant as a whole, so we restricted the analysis of these data to differences between corner and side features and excluded it from other template comparisons.

We then used concave- and convex-condition classification images to make a prediction for the discrimination efficiency of a linear observer (Figure 3C). A good match between these two suggests that a template-matching model can explain quantitatively (at least to an approximation) the contour integration performance on the task. We found a high correlation between the predicted efficiency and observed efficiency, $r = 0.89$, $p = 0.0005$. However, the predicted efficiency is some 10% lower than what was observed.

Convex and concave RF4 contours

In the convex condition, all subjects show nonzero weights for nearly all contour features (convex sides and corners, corresponding to RF troughs and RF peaks; Figure 5, blue plots). This suggests that nearly all parts of the contours are processed. We quantified the possible differences in contour feature preference by taking the mean of estimated feature weights (see Equation 9). Average contour feature weights (Figure 6) show that no systematic preference in using corner (RF peak) or side (RF trough) features can be seen in any observer. Estimated templates resemble the target closely; this was reflected in high sampling efficiency (on average 80% in the convex condition; Figure 3B).

For concave contours, most observers still tend to process at least most of the contour parts (Figure 5, red plots). However, contour feature preference is no longer even. Especially naïve observers (TP, JH, SS) rely predominantly on a set of local features or a single

Observer	IK	VS	SS	TP	JH
T1 (cv)					
Dev M11/M10	6,063/7,081	7,082/7,258	3,918/4,414	3,835/4,472	3,106/3,532
$\chi^2(32)$	1,023	175	496	637	429
p	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$
T1 (cc)					
Dev M11/M10	4,020/4,350	7,092/7,283	4,163/4,341	4,219/4,695	3,729/4,035
$\chi^2(32)$	330	191	178	476	306
p	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$
T2 (cv, cc)					
Dev M21/M20	10,092/10,208	14,249/14,296	8,178/8,282	8,224/8,437	6,957/7,068
$\chi^2(32)$	116	47	104	212	112
$p(>\chi^2)$	$<10^{-6}$	0.045	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$
Th (cv, cc)					
$t(19)$	−4.80	−5.00	−11.19	−20.09	−8.53
p	0.00012	0.00008	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$

Table 1. Statistical tests for individual data in the convex and concave conditions. We used nested hypothesis likelihood ratio tests. “T1” is the test for significance of classification-image weights. “Dev” is the deviance (twice the negative log likelihood) for a model. The full model (M11) is the unconstrained GLM model with classification-image weights. It was compared with a model that had only one regressor for target presence (M10). “T2” is the test for template difference. There, an unconstrained model (M21) had separate classification-image weights for convex and concave conditions. It was compared with a model (M20) where just one set of weights was estimated for both conditions. Rows marked p give the p -value for χ^2 test statistics. “Th” is the t -test value for the discrimination threshold difference.

contour feature and use the others less. Interestingly, the average contour feature weights reveal that observers also show a preference for contours of a certain feature type (concave sides or RF troughs, corners or RF peaks), but this is not systematic across the observers. IK and VS weight more corner features, whereas TP, JH, and SS prefer side features (Figure 6). Average template sampling efficiency was lower than in the convex condition, $t(4) = 3.31$, $p = 0.03$, but still rather high (average 55%).

No systematic geometric location preference (such as left upper corner) seems to exist across subjects in either of the conditions (see average data in Figures 4 and 5).

Short-duration processing

As short-duration classification images were noisy, we restrict our analysis to average weights that the observers assign to corner and (convex) side features. All observers show more efficient processing of corner features relative to side features, $t(4) = 7.20$, $p = 0.006$ (Figure 6).

Nonlinear integration

Goodness of fit between model and observer responses (model–observer response correlation) at various levels of nonlinear integration (γ) is shown in Figure 7. Fit curves are clearly unimodal; curves peak

around $\gamma = 1$, which is equivalent to linear integration. There is little evidence for any nonlinear integration, as higher γ values—i.e., maximum-of-outputs integration—provide a low correlation with human responses.

Discussion

Mechanisms and integration of contour information

The main purpose of this study was to shed light on psychophysical mechanisms underlying contour integration, by using the position noise classification-image paradigm. We found that shape discrimination performance could be predicted quite accurately by a template-matching model using the estimated classification image. The correlation between estimated and predicted templates was high; predicted efficiencies were just about 10% lower than observed. It should be noted that absolute efficiency is proportional to the square of the discrimination index (d'), a more commonly used measure for performance. Expressed in d' , the underprediction would be only about 5%. In previous studies, it has also been found that classification images slightly underpredict the true efficiency (see also Murray et al., 2005). This may be caused by factors such as phase or (here perhaps) feature location uncertainty. We conclude that the high correlation between the model prediction and observed efficiency,

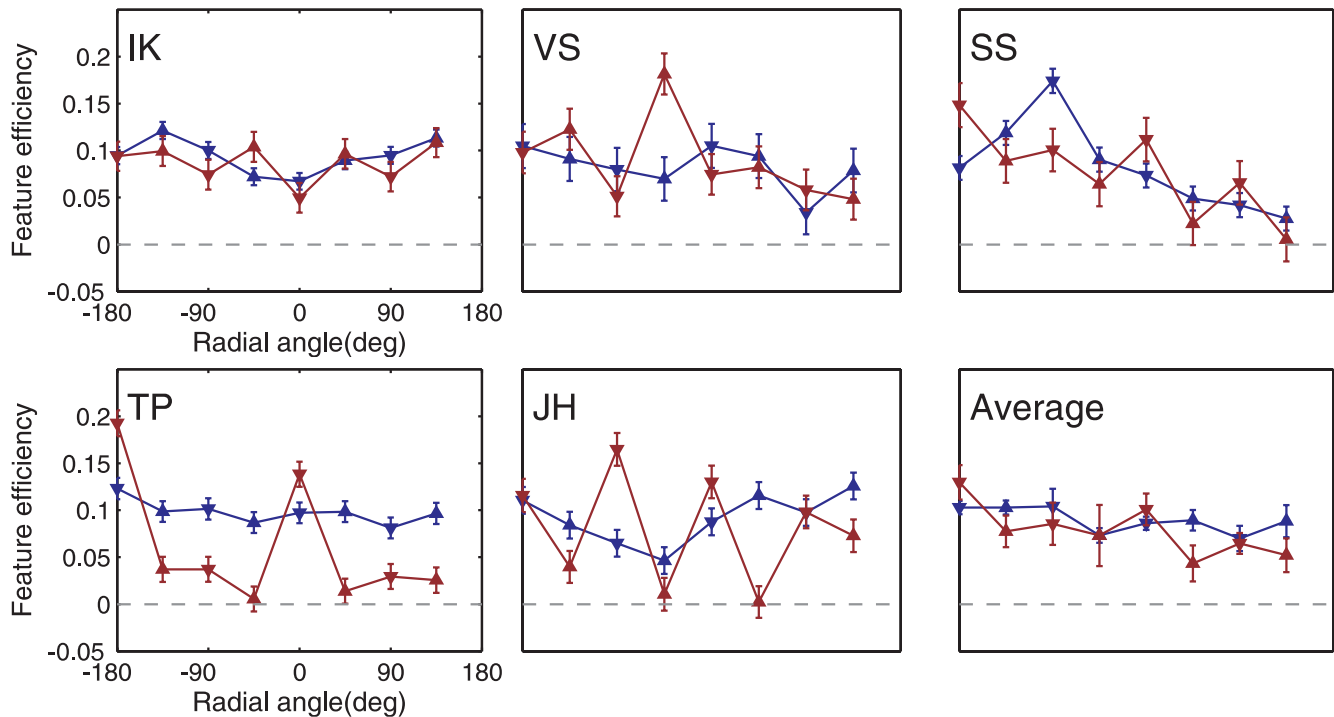


Figure 5. Comparison of average contour feature weights in convex (blue) and concave (red) RF patterns. Bars with upward triangles show the average corner feature weights, and bars with downward triangles show the average weighting of the side features. Amplitudes are relative to $1/8$, which indicates a perfect match between the feature in the classification image and the feature in the ideal template. “Average” shows the average across the observers. Error bars represent 1 standard error of the mean.

as well as relatively accurate overall performance, suggests that a simple model matching local contour parts and an internal shape template can capture some key aspects of contour integration.

Template-matching models have not been previously tested in shape perception, presumably because most of the studies have used static (nonstochastic) stimuli with tasks that measure fine visual discrimination processing. Our results complement these findings, showing that for stochastic shape stimuli, the visual system calculates the linear match between noisy input and shape template in a nearly global manner. We do not aim to present a biologically plausible model here, but we point out that certain features of previous biologically inspired models (Poirier & Wilson, 2007, 2011), such as global pooling at the integration stage, are in line with our results, as well as postulated high-level shape templates.

Integration of contour features in shape perception

Classification images in both conditions show that observers used all or most of the contour features (sides, corners), suggesting that all contour parts contribute to shape processing. We argue that classi-

fication images will provide a major advantage over many previous designs that have sought to investigate global/local integration by adding/subtracting the number of contour features (Bell & Badcock, 2008; Hess et al., 1999; Schmidtman et al., 2012), since observers could switch to a more local processing strategy when the stimulus has only certain parts. It is not trivial to generalize these results to the global integration task. Moreover, it has been speculated that the number of stimulus fragments may have an influence on how attention is distributed to contour information (Mullen et al., 2011). With classification images, we can use the whole intact stimulus while still being able to infer how its different parts are processed.

We found that nonlinear, maximum-of-outputs integration of contour features did not provide a better fit to the data, but instead that linear summation consistently gave the best response correlation between the model and observers. Also, we found that weighting of all contour features was rather even. This supports the idea of global and linear integration of shape feature information, even when shape phase was not randomized (but the location of the pattern was). Our results are in line with studies where nearly optimal integration of RF contour fragments has been reported (Dickinson et al., 2012; Hess et al., 1999; Loffler et al., 2003). This further supports the idea that shape

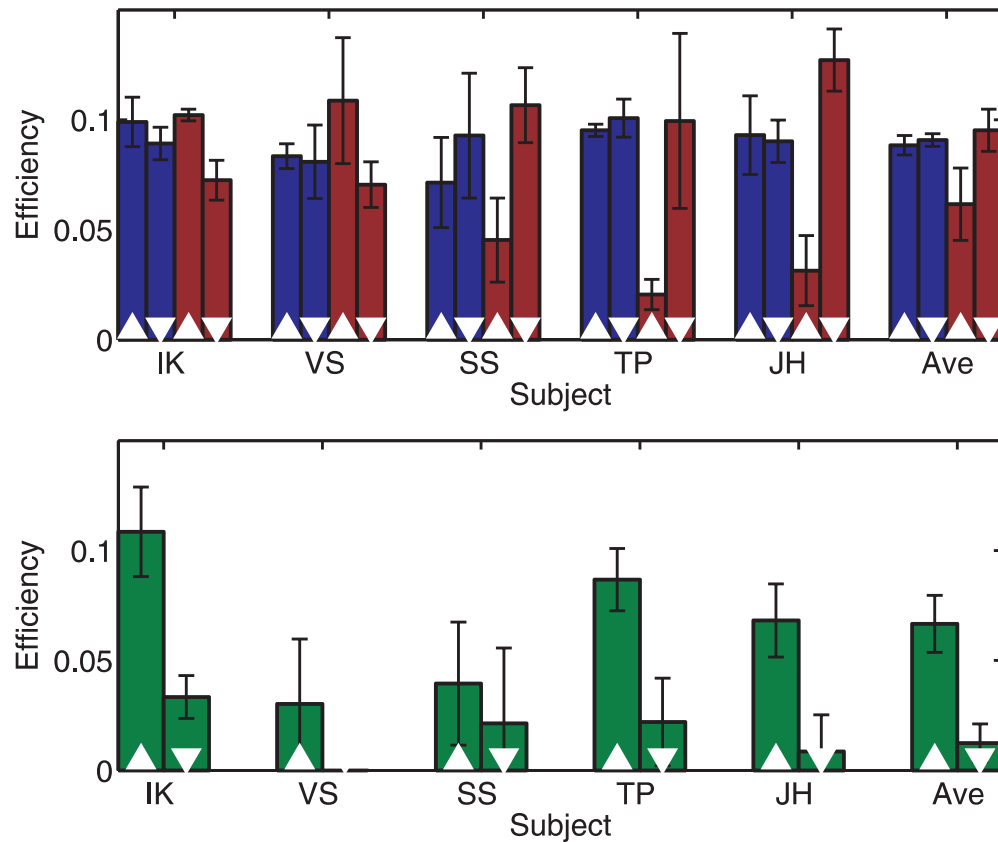


Figure 6. Average feature weights across locations; upward triangles are corner features and downward triangles are side features. “Ave” shows the average across the observers. Top panel: Blue bars show average feature weights for convex and red bars for concave patterns. Differences between weightings of the contour features are not systematic across the observers. Bottom panel: Average feature weights for the short-duration convex pattern. Error bars represent 1 standard error of the mean. A preference for corner features (RF peaks) can be seen.

processing is mediated through neural mechanisms that are sensitive to more global and object-centered properties of the stimulus instead of local cues (such as single corner or side), in line with neurophysiological evidence for neurons at higher stages of the ventral processing stream that are specialized for extraction of contour and shape information (Gallant et al., 1996; Nandy et al., 2013; Pasupathy & Connor, 1999, 2002; Yau, Pasupathy, Brincat, & Connor, 2013).

Side and corner features

We derived the classification images for convex and concave RF4 shape targets. We found that sampling efficiency of processing was better for the target with convex sides compared to the target with concave sides, even when their ideal observer was identical. This is likely to be explained by a poorer sampling and processing of contour information (instead of a difference in internal noise level), as the shape-template efficiency was also lower. This is in line with the idea

that the visual system has increased sensitivity for convex contour information (Bertamini & Mosca, 2004; Driver & Baylis, 1995; Loffler et al., 2003). However, in the concave condition the weighting was not systematically inferior for concave side features compared to corner features. Another possibility is that contour templates in the convex-shape condition reflect more typical and better learned shapes, and thus the templates are more efficient for that reason.

With the convex target, we found no systematic differences in processing of corners (RF peaks) and convex side features. Further, all contour parts were used with a fairly even weighting. However, when the target was a concave shape, we found inferior processing of concave side features (RF troughs) compared with corners (RF peaks) in three of the five observers. The other two observers showed a preference for concave sides over corner features.

As noted previously, noise-masking studies (Hess et al., 1999) and pattern-masking studies (Habak et al., 2004; Poirier & Wilson, 2007) have provided conflicting results on the relative importance of corner and side

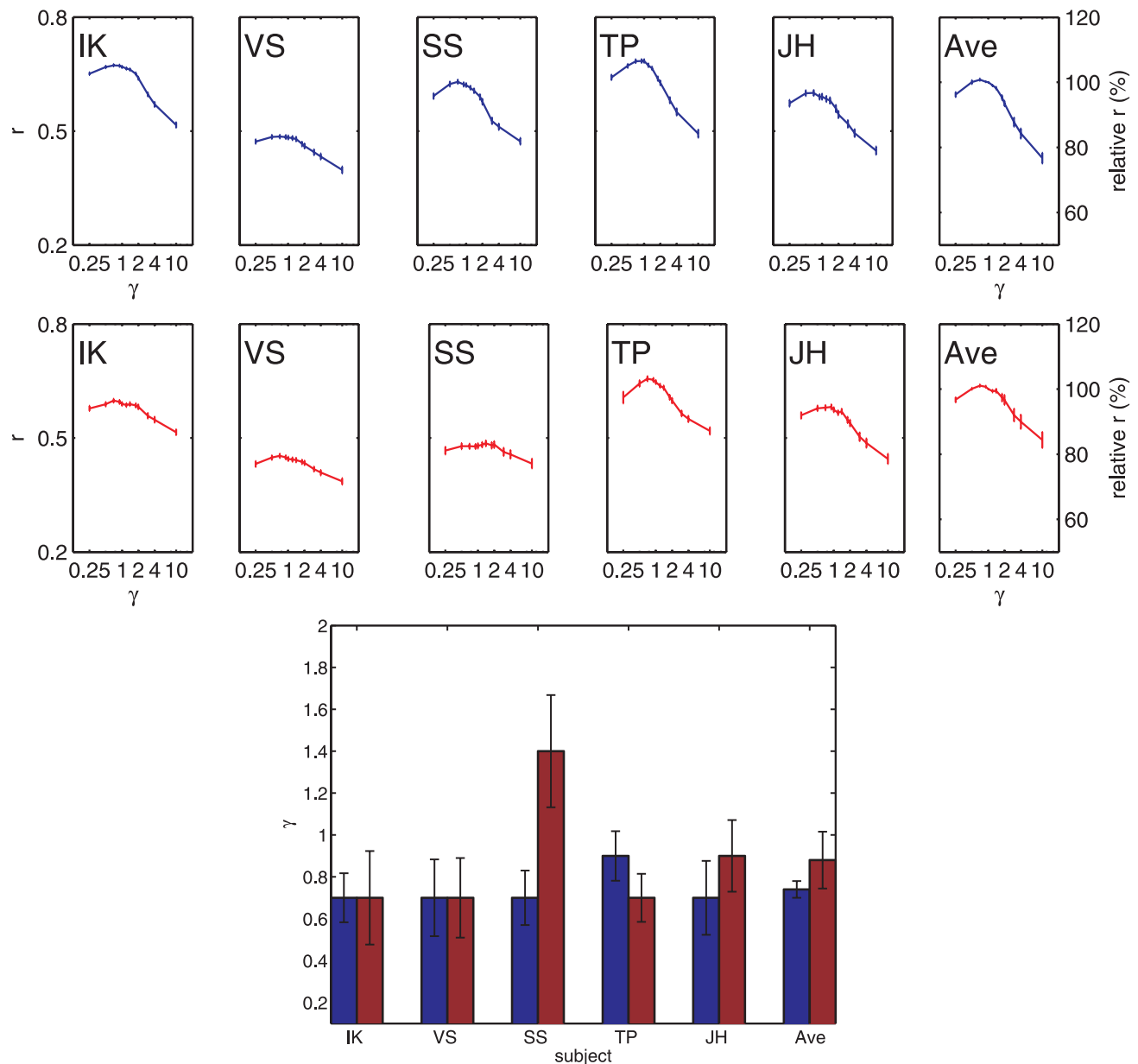


Figure 7. Summation analysis. Contour feature integration was investigated by comparing how well a nonlinear Minkowski contour feature integration model could predict the observer responses at various levels of nonlinearity. The two top panels show the model performance r (model–human response correlation) against γ , the parameter controlling the nonlinearity of summation. In the average panel, the correlation is expressed as a percentage with respect to $\gamma = 1$ for each subject. Blue curves: convex condition; red curves: concave condition. Results peak near $\gamma = 1$, implying linear integration. Bottom panel: best-fitting summation values. Error bars represent 1 standard error of the mean, obtained by bootstrap resampling.

features. Here, we used a more direct approach and found no systematic preference for any of the observers. Similar results have recently been reported using an adaptation paradigm (Bell, Hancock, Kingdom, & Peirce, 2010). Moreover Mullen et al. (2011) observed no systematic difference in discrimination thresholds for isolated corner and side features. Nevertheless, data from the concave condition show

that observers may prefer either the side or the corner features. This may indicate that the visual system is able to prefer a set of contour features in a contour through an attentional strategy or a similar mechanism, which may explain why different indirect experimental manipulations, like RF or luminance masking, have different effects on the relative efficiency of side and corner feature processing.

In the time-limited short-duration condition, we found that estimated templates were very noisy, but observers still did use corner features quite systematically. A reason for investigating this condition was that results of the convex and concave RF conditions did not show consistent differences between different feature types (corners/sides). Yet, we found that templates in these long-duration experiments were overall rather efficient, i.e., the observers used contour features in an optimal manner. A motivation in the short-duration experiment was to test whether higher sensitivity for certain contour features would manifest only in conditions where the shape integration was challenging. The results show that in a short-duration condition, corner processing can be more efficient than side feature processing. However, it is unclear whether this reflects processing at the global integration stage, as the classification images were highly noisy and do not clearly support global processing. We stress that previous studies have provided conflicting results on possible preference for side or corner features, and we conclude that preference for certain contour features may be dependent on factors such as stimulus duration and RF amplitude (which was also highest in the short-duration experiment).

General conclusions

In this study, we used an efficient variant of the classification-image method to directly estimate the shape templates for different radial frequency shapes and showed that the midlevel visual task of contour integration can be explained—at least to a first approximation—by a template-matching model that computes a linear sum between the noisy stimulus information and the internal shape template that determines how the shape features are weighted. Discrimination efficiency for different shapes could be predicted from the classification-image template. Shape templates contained most or even all of the contour parts (especially in the convex condition), suggesting global integration of contour information, in line with previous evidence. Further, we analyzed the nature of feature integration in radial patterns by comparing responses of nonlinear models and human data to the same stimulus, and found that linear summation of features was more likely than nonlinear probability summation.

Unlike previous masking studies and studies with contour parts, we used a more direct approach to estimate the relative contribution of corner and side features and found no systematic preference for either one, except in the short-duration condition, which may suggest a preference for corners in time-limited processing.

Lastly, this study shows how the classification-image method can be further developed for research questions like contour integration that lie outside the basic luminance noise-masking paradigm. The method allows formulation of exact computational models on how complex midlevel tasks like contour integration work, as well as measurement of the information the observer uses in such a task without relying on indirect stimulus manipulations.

Keywords: radial frequency pattern, classification image, shape perception, global processing

Acknowledgments

A. H. was supported by the Academy of Finland, Centres-of-Excellence in Inverse Problems Research and Algorithmic Data Analysis (Algodan). Preliminary results from this study were presented at the Vision Sciences Society Meeting 2010 at Naples, FL. We thank Dr. Viljami Salmela and three anonymous reviewers for helpful comments.

Commercial relationships: none.

Corresponding author: Ilmari Kurki.

Email: ilmari.kurki@helsinki.fi.

Address: Institute of Behavioral Sciences, University of Helsinki, Helsinki, Finland.

Footnotes

¹It should be noted that even when we manipulate and analyze the stimulus in the position domain, we do not need to assume that shape coding is based on analyzing explicitly the position of the elements. Positional perturbation also changes the local orientations and spatial phase in the stimulus, and it is likely that positional signal is in fact extracted in the visual system by a local orientation and phase analysis.

²The prime symbol refers to a visual angle of 1 arc minute.

³Note that each feature in the ideal template had the same total energy (sum of squared deviations).

References

- Anzai, A., Peng, X., & Van Essen, D. C. (2007). Neurons in monkey visual area V2 encode combinations of orientations. *Nature Neuroscience*, *10*(10), 1313–1321, doi:10.1038/nn1975.
- Barenholtz, E., Cohen, E. H., Feldman, J., & Singh, M.

- (2003). Detection of change in shape: An advantage for concavities. *Cognition*, 89(1), 1–9, doi:10.1016/S0010-0277(03)00068-4.
- Beard, B., & Ahumada, A. J. (1998). A technique to extract the relevant features for visual tasks. In B. E. Rogowitz & T. N. Pappas (Eds.), *SPIE proceedings 3299* (pp. 79–85). Bellingham, WA: SPIE.
- Bell, J., & Badcock, D. R. (2008). Luminance and contrast cues are integrated in global shape detection with contours. *Vision Research*, 48(21), 2336–2344, doi:10.1016/j.visres.2008.07.015.
- Bell, J., Gheorghiu, E., Hess, R. F., & Kingdom, F. A. (2011). Global shape processing involves a hierarchy of integration stages. *Vision Research*, 51(15), 1760–1766, doi:10.1016/j.visres.2011.06.003.
- Bell, J., Hancock, S., Kingdom, F., & Peirce, J. W. (2010). Global shape processing: Which parts form the whole? *Journal of Vision*, 10(6):16, 1–13, <http://www.journalofvision.org/content/10/6/16>, doi:10.1167/10.6.16. [PubMed] [Article]
- Bertamini, M., & Mosca, F. (2004). Early computation of contour curvature and part structure: Evidence from holes. *Perception*, 33(1), 35–48.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.
- Burgess, A., Wagner, R., Jennings, R., & Barlow, H. (1981, October 2). Efficiency of human visual signal discrimination. *Science*, 214(4516), 93–94, doi:10.1126/science.7280685.
- Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology (London)*, 197, 551–566.
- Connor, C. E., Brincat, S. L., & Pasupathy, A. (2007). Transformation of shape information in the ventral pathway. *Current Opinion in Neurobiology*, 17(2), 140–147.
- David, S. V., Hayden, B. Y., & Gallant, J. L. (2006). Spectral receptive field properties explain shape selectivity in area V4. *Journal of Neurophysiology*, 96(6), 3492–3505.
- Dickinson, J. E., McGinty, J., Webster, K. E., & Badcock, D. R. (2012). Further evidence that local cues to shape in RF patterns are integrated globally. *Journal of Vision*, 12(12):16, 1–17, <http://www.journalofvision.org/content/12/12/16>, doi:10.1167/12.12.16. [PubMed] [Article]
- Driver, J., & Baylis, G. C. (1995). One-sided edge assignment in vision: 2. Part decomposition, shape description, and attention to objects. *Current Directions in Psychological Science*, 4(6), 201–206, doi:10.1111/1467-8721.ep10772645.
- Eckstein, M. P., & Ahumada, A. J. (2002). Classification images: A tool to analyze visual strategies. *Journal of Vision*, 2(1):i, <http://www.journalofvision.org/content/2/1/i>, doi:10.1167/2.1.i. [PubMed] [Article]
- Eckstein, M. P., Pham, B. T., & Shimozaki, S. S. (2004). The footprints of visual attention during search with 100% valid and 100% invalid cues. *Vision Research*, 44(12), 1193–1207, doi:10.1016/j.visres.2003.10.026.
- Gallant, J. L., Braun, J., & Van Essen, D. C. (1993, January 1). Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science*, 259(5091), 100–103.
- Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W., & Van Essen, D. C. (1996). Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *Journal of Neurophysiology*, 76(4), 2718–2739.
- Gheorghiu, E., & Kingdom, F. A. (2009). Multiplication in curvature processing. *Journal of Vision*, 9(2):23, 1–17, <http://www.journalofvision.org/content/9/2/23>, doi:10.1167/9.2.23. [PubMed] [Article]
- Gheorghiu, E., & Kingdom, F. A. (2007). The spatial feature underlying the shape-frequency and shape-amplitude after-effects. *Vision Research*, 47(6), 834–844, doi:10.1016/j.visres.2006.11.023.
- Graham, N. (1989). *Visual pattern analyzers*. New York: Oxford University Press.
- Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics* (Reprint ed.). New York: John Wiley and Sons.
- Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, 41(10–11), 1409–1422, doi:10.1016/S0042-6989(01)00073-6.
- Habak, C., Wilkinson, F., Zakher, B., & Wilson, H. R. (2004). Curvature population coding for complex shapes in human vision. *Vision Research*, 44(24), 2815–2823, doi:10.1016/j.visres.2004.06.019.
- Haushofer, J., Baker, C. I., Livingstone, M. S., & Kanwisher, N. (2008). Privileged coding of convex shapes in human object-selective cortex. *Journal of Neurophysiology*, 100(2), 753–762, doi:10.1152/jn.90310.2008.
- Hess, R. F., Wang, Y. Z., & Dakin, S. C. (1999). Are judgements of circularity local or global? *Vision Research*, 39(26), 4354–4360.
- Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, 148(3), 574–591.
- Ivanov, I. V., & Mullen, K. T. (2012). The role of local

- features in shape discrimination of contour- and surface-defined radial frequency patterns at low contrast. *Vision Research*, 52(1), 1–10, doi:10.1016/j.visres.2011.10.002.
- Kanizsa, G. (1979). *Organization in vision*. New York: Praeger.
- Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3? *Perception*, 36(ECVP Abstract Supplement).
- Knoblauch, K., & Maloney, L. T. (2008). Estimating classification images with generalized linear and additive models. *Journal of Vision*, 8(16):10, 1–19, <http://www.journalofvision.org/content/8/16/10>, doi:10.1167/8.16.10. [PubMed] [Article]
- Knoblauch, K., & Maloney, L. T. (2012). *Modelling psychophysical data in R*. New York: Springer.
- Kurki, I., Laurinen, P., Peromaa, T., & Saarinen, J. (2003). Spatial integration in glass patterns. *Perception*, 32(10), 1211–1220.
- Levi, D. M., & Klein, S. A. (2003). Noise provides new signals about the spatial vision of amblyopes. *Journal of Neuroscience*, 23(7), 2522–2526.
- Li, R. W., Klein, S. A., & Levi, D. M. (2006). The receptive field and internal noise for position acuity change with feature separation. *Journal of Vision*, 6(4):2, 311–321, <http://www.journalofvision.org/content/6/4/2>, doi:10.1167/6.4.2. [PubMed] [Article]
- Li, R. W., Levi, D. M., & Klein, S. A. (2004). Perceptual learning improves efficiency by re-tuning the decision “template” for position discrimination. *Nature Neuroscience*, 7(2), 178–183.
- Loffler, G., Wilson, H. R., & Wilkinson, F. (2003). Local and global contributions to shape discrimination. *Vision Research*, 43(5), 519–530.
- Mullen, K. T., Beaudot, W. H., & Ivanov, I. V. (2011). Evidence that global processing does not limit thresholds for RF shape discrimination. *Journal of Vision*, 11(3):6, 1–21, <http://www.journalofvision.org/content/11/3/6>, doi:10.1167/11.3.6. [PubMed] [Article]
- Murray, R. F. (2011). Classification images: A review. *Journal of Vision*, 11(5):2, 1–25, <http://www.journalofvision.org/content/11/5/2>, doi:10.1167/11.5.2. [PubMed] [Article]
- Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2005). Classification images predict absolute efficiency. *Journal of Vision*, 5(2):5, 139–149, <http://www.journalofvision.org/content/5/2/5>, doi:10.1167/5.2.5. [PubMed] [Article]
- Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2002). Optimal methods for calculating classification images: Weighted sums. *Journal of Vision*, 2(1):6, 79–104, <http://www.journalofvision.org/content/2/1/6>, doi:10.1167/2.1.6. [PubMed] [Article]
- Nandy, A. S., Sharpee, T. O., Reynolds, J. H., & Mitchell, J. F. (2013). The fine structure of shape tuning in area V4. *Neuron*, 78(6), 1102–1115, doi:10.1016/j.neuron.2013.04.016.
- Neri, P., & Levi, D. M. (2006). Receptive versus perceptive fields from the reverse-correlation viewpoint. *Vision Research*, 46(16), 2465–2474.
- Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience*, 5(12), 1332–1338.
- Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of Neurophysiology*, 82(5), 2490–2502.
- Pasupathy, A., & Connor, C. E. (2001). Shape representation in area V4: Position-specific tuning for boundary conformation. *Journal of Neurophysiology*, 86(5), 2505–2519.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of Optical Society of America*, 16(3), 647–653.
- Peterson, W., Birdsall, T., & Fox, W. (1954). The theory of signal detectability. *IRE Professional Group on Information Theory*, 4(4), 171–212, doi:10.1109/TIT.1954.1057460.
- Poirier, F. J., & Wilson, H. R. (2011). A biologically plausible model of human shape symmetry perception. *Journal of Vision*, 10(1):9, 1–16, <http://www.journalofvision.org/content/10/1/9>, doi:10.1167/10.1.9. [PubMed] [Article]
- Poirier, F. J., & Wilson, H. R. (2007). Object perception and masking: Contributions of sides and convexities. *Vision Research*, 47(23), 3001–3011.
- Schmidtman, G., Kennedy, G. J., Orbach, H. S., & Loffler, G. (2012). Non-linear global pooling in the discrimination of circular and non-circular shapes. *Vision Research*, 62, 44–56, doi:10.1016/j.visres.2012.03.001.
- Solomon, J. A. (2002). Noise reveals visual mechanisms of detection and discrimination. *Journal of Vision*, 2(1):7, 105–120, <http://www.journalofvision.org/content/2/1/7>, doi:10.1167/2.1.7. [PubMed] [Article]
- Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature*, 369(6479), 395–397.
- Tjan, B. S., & Nandy, A. S. (2006). Classification images with uncertainty. *Journal of Vision*, 6(4):8, 387–413, <http://www.journalofvision.org/content/6/4/8>, doi:10.1167/6.4.8. [PubMed] [Article]

- Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception and Psychophysics*, 33, 113–120.
- Wilkinson, F., James, T. W., Wilson, H. R., Gati, J. S., Menon, R. S., & Goodale, M. A. (2000). An fMRI study of the selective activation of human extrastriate form vision areas by radial and concentric gratings. *Current Biology*, 10(22), 1455–1458.
- Wilkinson, F., Wilson, H. R., & Habak, C. (1998). Detection and recognition of radial frequency patterns. *Vision Research*, 38(22), 3555–3568.
- Willmore, B. D. B., Prenger, R. J., & Gallant, J. L. (2010). Neural representation of natural images in visual area V2. *Journal of Neuroscience*, 30(6), 2102–2114, doi:10.1523/JNEUROSCI.4099-09.2010.
- Wilson, H. R., & Wilkinson, F. (1998). Detection of global structure in Glass patterns: Implications for form vision. *Vision Research*, 38(19), 2933–2947.
- Wilson, H. R., Wilkinson, F., & Asaad, W. (1997). Concentric orientation summation in human form vision. *Vision Research*, 37(17), 2325–2330.
- Yau, J. M., Pasupathy, A., Brincat, S. L., & Connor, C. E. (2013). Curvature processing dynamics in macaque area V4. *Cerebral Cortex*, 23(1), 198–209, doi:10.1093/cercor/bhs004.

Appendix A: Stimulus generation

Radial frequency patterns are conventionally defined in polar coordinates. In a perfect circle ($a = 0$), the spatial distances between the elements in the contour are exactly proportional to the angular distances between the elements. However, in general this is not true. Spacing elements by equal polar angle could cause a local density cue to the corners and sides of the pattern. Therefore, we designed the stimulus so that the Cartesian distance between elements at the radial frequency contour remains constant regardless of the pattern.

In polar coordinates, a radial frequency pattern is defined as a sinusoidal modulation of the radius as a function of polar angle. Let r_m be the mean radius of the pattern, f the radial frequency, ϕ the phase, and a the amplitude. Then

$$r(\theta) = a \sin(f\theta + \phi) + r_m. \quad (\text{A1})$$

We first computed a virtual radial frequency contour by generating the set of possible element locations in Cartesian coordinates. Contour locations $\mathbf{p}(k)$ were added by the following iterative algorithm:

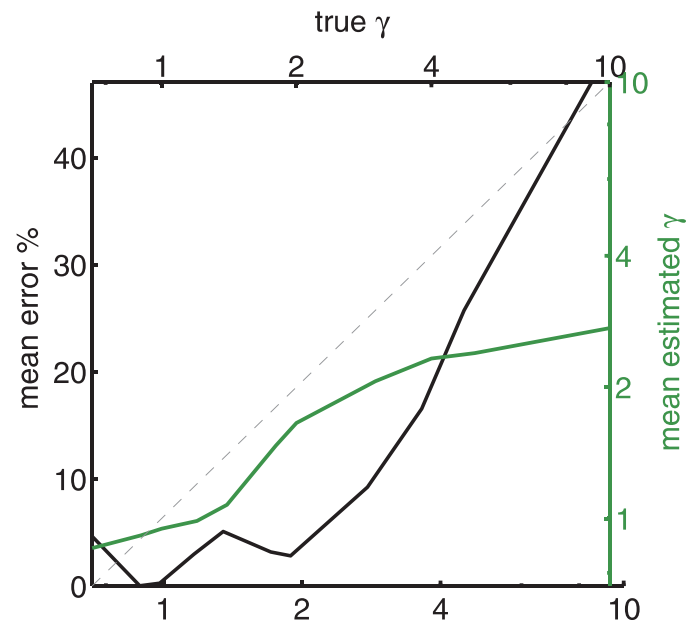


Figure A1. Nonlinear summation analysis accuracy simulation. The estimated Minkowski summation parameter γ is plotted against true γ . The black line represents the mean error (mean squared estimation error as a percentage of squared true value); the green line represents the mean estimated value. At low γ values, the method is reasonably accurate, but at higher γ values it has a downward bias.

$$\mathbf{p}(k+1) = \mathbf{p}(k) + \frac{1}{\sqrt{x'(\mathbf{p}(k))^2 + y'(\mathbf{p}(k))^2}}, \quad (\text{A2})$$

where $x'(\mathbf{p}(k))$ and $y'(\mathbf{p}(k))$ are the derivatives of the RF pattern at point $\mathbf{p}(k)$:

$$x'(\mathbf{p}(k)) = af \cos(f\theta + \phi) \cos(\theta) - (a \sin(f\theta + \phi) + r_m) \sin(\theta), \quad (\text{A3})$$

$$y'(\mathbf{p}(k)) = af \cos(f\theta + \phi) \cos(\theta) + (a \sin(f\theta + \phi) + r_m) \sin(\theta). \quad (\text{A4})$$

New locations were added to the virtual contour until the polar angle of the last location was 2π radians from the first location.

Next, 32 elements were placed on the virtual contour locations so that the distance between the elements was constant. The radial angle of the first element was randomized. Finally, the global position of the pattern was jittered by adding a random spatial offset.

Appendix B: Precision of the nonlinear summation analysis

We tested the accuracy and bias of the nonlinear summation analysis that was used in the analysis (Equation 13). The observer model summed the responses to eight local features (see Methods) non-linearly, using Minkowski summation using a noisy version of an optimal template. The parameters of the observer model (template efficiency, performance, internal noise ratio, etc.) were selected so that they would match the parameters in the experiment as closely as possible. We then estimated the Minkowski

summation parameter γ by using the same procedure as with the empirical data.

The number of trials was set to 2,000. We tested summation parameter values from 0.5 to 10 with 1,000 repetitions each. It should be noted that the Minkowski model's behavior does not change much when values of γ exceed about 4.

Figure A1 shows the accuracy of the method (mean prediction error as a percentage of true value) plotted against true value of γ . The green curve represents the average estimated γ . Accuracy is reasonably good when γ is around 1, but at large values, there is a downward bias, i.e., the model tends to underestimate the real summation.