Synergy of spatial frequency and orientation bandwidth in texture segregation

Cordula Hunt
Department of Psychology, Methods Section, Johannes Gutenberg-Universität, Mainz, Germany

Günter Meinhardt
Department of Psychology, Methods Section, Johannes Gutenberg-Universität, Mainz, Germany

Defining target textures by increased bandwidths in spatial frequency and orientation, we observed strong cue combination effects in a combined texture figure detection and discrimination task. Performance for double-cue targets was better than predicted by independent processing of either cue and even better than predicted from linear cue integration. Application of a texture-processing model revealed that the oversummative cue combination effect is captured by calculating a low-level summary statistic (ΔCE_m), which describes the differential contrast energy to target and reference textures, from multiple scales and orientations, and integrating this statistic across channels with a winner-take-all rule. Modeling detection performance using a signal detection theory framework showed that the observers' sensitivity to single-cue and double-cue texture targets, measured in d' units, could be reproduced with plausible settings for filter and noise parameters. These results challenge models assuming separate channeling of elementary features and their later integration, since oversummative cue combination effects appear as an inherent property of local energy mechanisms, at least for spatial frequency and orientation bandwidth-modulated textures.

Introduction

Humans categorize visual scenes rapidly (Fei-Fei et al., 2007; Schyns & Oliva, 1994), judging at a glance whether its content is natural or man made (Greene & Oliva, 2009a, 2009b; Loschky et al., 2007; Loschky & Larson, 2008; Oliva & Torralba, 2001). Also discriminating scene objects at the basic level of categorization seemingly works within the first 100 ms of processing without allocation of spatial attention (Grill-Spector & Kanwisher, 2005; Hershler & Hochstein, 2005; Rousselet et al., 2005; Thorpe et al., 1996). Rapid and preattentive scene and object categorization indicates that observers rely on cues provided by earlier stages of image analysis (Baddeley, 1997; Groen et al., 2013; VanRullen, 2006; Wichmann et al., 2006).

Likewise, rapid and preattentive discrimination of "textures," that is, artificial images with spatial regularity, also indicates involvement of early analysis mechanisms. First studies suggested that identity of the power spectrum decides whether two textures can be discriminated only with detailed scrutiny or immediately and preattentively. However, later studies found several counter examples, demonstrating that there are textures with equal power spectra, and even with equal third-order statistics, which segregate at a glance (Julesz et al., 1973, 1978). These results suggested that a Fourier-like image description that could be extracted from the responses of orientation and spatial frequency tuned cell populations found in striate cortex (Campbell et al., 1969; Daugman, 1980; Hubel & Wiesel, 1968) might only be an initial step in texture segregation (Sagi, 1995; Victor et al., 1995).

In one class of feedforward processing models, the output of orientation and spatial frequency selective filters is squared or full-wave-rectified to form a local energy measure (Landy, 2013; Lin & Wilson, 1996). This basic operation not only marks local contrast and luminance differences but also identifies texture regions that differ in arbitrary feature modulations by translating them into different local energy distributions (Bergen & Adelson, 1988; Landy & Bergen, 1991). Due to its simplicity, the model has widely been used to describe human sensitivity to feature modulation across space (Arsenault et al., 1999; Bergen & Adelson, 1988; Caelli, 1982; Prins & Kingdom, 2006; Rubenstein & Sagi, 1990; Sagi, 1995). It has also been used in recent attempts to explain the rapid extraction of a scene "gist" (Groen et al., 2013). In most implementations, there is a first filtering stage, followed by local energy computation, which again is followed by a secondary, larger-scale filtering stage, which enhances pooled regional differences in feature modulation. Models of this class are referred to as filter-rectify-filter (FRF) models (Landy & Bergen, 1991; Landy, 2013;

Citation: Hunt, C., & Meinhardt, G. (2021). Synergy of spatial frequency and orientation bandwidth in texture segregation. *Journal of Vision*, *21*(2):5, 1–25, https://doi.org/10.1167/jov.21.2.5.

https://doi.org/10.1167/jov.21.2.5 Received May 26, 2020; published February 9, 2021 ISSN 1534-7362

Landy & Oruc, 2002; Prins & Kingdom, 2006). FRF models can predict the close covariation of detection performance with degree of texture modulation (Landy & Bergen, 1991), the salience of texture boundaries (Malik & Perona, 1990; Landy & Bergen, 1991), and several asymmetries in human texture segregation (Rubenstein & Sagi, 1990). However, as pointed out by Prins and Kingdom (2006), direct evidence for the existence of local energy-based texture mechanisms is rare.

Malik and Perona (1990) claimed that a complete model of texture perception should satisfy three criteria: (1) It should be biologically plausible, (2) it should be general enough to be tested on any gray-scale image, and (3) it should yield a quantitative match with psychophysical data. Local energy-based models satisfy the first criterion of biological plausibility as they use first-order filters with weighting characteristics resembling the receptive field profiles of cortical simple cells in V1 (Daugman, 1980, 1985), which were found to be tuned to specific orientations and spatial scales (Campbell et al., 1969; Hubel & Wiesel, 1965). Further, V2 cells were shown to respond to locus and orientation of texture boundaries but were unresponsive to the luminance profiles of texture elements (von der Heydt et al., 1984), a behavior that closely matches the non-Fourier second-order squaring and pooling operation (Lin & Wilson, 1996). Thus, FRF mechanisms can be viewed as modeling processing chains of V1 and V2 cells. Local energy-based models also easily satisfy the second criterion as they can take an arbitrary gray-scale image as input.

The third criterion states that a test for local energy-based texture mechanisms requires a psychophysical benchmark. A strong challenge for feedforward local energy-based models would be prediction of the "feature synergy effect" arising in the discrimination of texture regions defined by feature modulation in two dimensions (Meinhardt & Persike, 2003; Meinhardt et al., 2004, 2006; Kida et al., 2011; Straube et al., 2010). Authors showed that sensitivity to target texture regions that varied from the surround in orientation and spatial frequency was much higher than predicted from the assumption of detecting texture modulation along either dimension independently. Such a multi cue advantage has been termed "synergy" (Kubovy & Cohen, 2001), and besides orientation and spatial frequency, it has been reported for color and form (Kubovy et al., 1999) and for color and orientation (Saarela & Landy, 2012). Strong cue summation was also found in contour integration, whereby additional feature contrast enhanced detection of contours defined by orientation alignment (Machilsen & Wagemans, 2011; Persike & Meinhardt, 2015).

Assuming feature independence implies the notion that the visual system analyzes visual scenes in terms of "features," using feature-specific modules acting in parallel (Kubovy & Cohen, 2001). Such models have been proposed to describe "pop out" in visual search, whereby each module signals spatiotemporal gradients for the feature it analyzes (Treisman & Gelade, 1980; Treisman, 1988). Local energy-based models are at odds with this notion, since their local nonlinearity marks any regional texture change regardless of its specific featural origin. Hence, in local energy-based models, there are no feature modules built from the outputs of the primary analyzers.

Predicting the feature synergy effect for textures jointly modulated in two features from local energy-based mechanisms would be strong evidence for this model class, since there are no free parameters for weighting how texture modulations from two features combine. Support for local energy-based mechanisms would be even stronger if the specific nature of the texture stimuli makes architectures with separate feature modules no likely candidates for explaining the synergy effect, even with additional assumptions for the way how module outputs are integrated.

In this sense, the present study provides evidence for local energy-based mechanisms in human texture segregation. We devised stimuli with no likely double-cue advantage, using textures with no feature contrast but different orientation and spatial frequency *bandwidths*. In these textures, orientation and spatial frequency differences of target and reference textures arise in regions of the parameter space where practically no overlap of the response characteristics of primary filters can be expected. However, we observed a strong double-cue benefit for these stimuli, which was at least as strong as the synergy effect for textures with orientation and spatial frequency contrast (Meinhardt et al., 2006; Persike & Meinhardt, 2008). Analyzing the textures with a plausible local energy model and combining the space-average responses with a simple **max**-rule replicated the synergy effect for detection, while no further assumptions or free parameters entered. The findings suggest that nonlinear saliency enhancement for jointly modulated textures is an inherent property of local energy mechanisms, which offers a parsimonious and straightforward explanation of the feature synergy effect in texture segregation.

Method

Experimental rationale

Varying the spread parameters of Gabor kernels for Landy-Bergen textures yields a synergy effect

Landy and Bergen (1991) constructed oriented noise textures by convolving a spatial noise image, having the gray value of each pixel sampled from

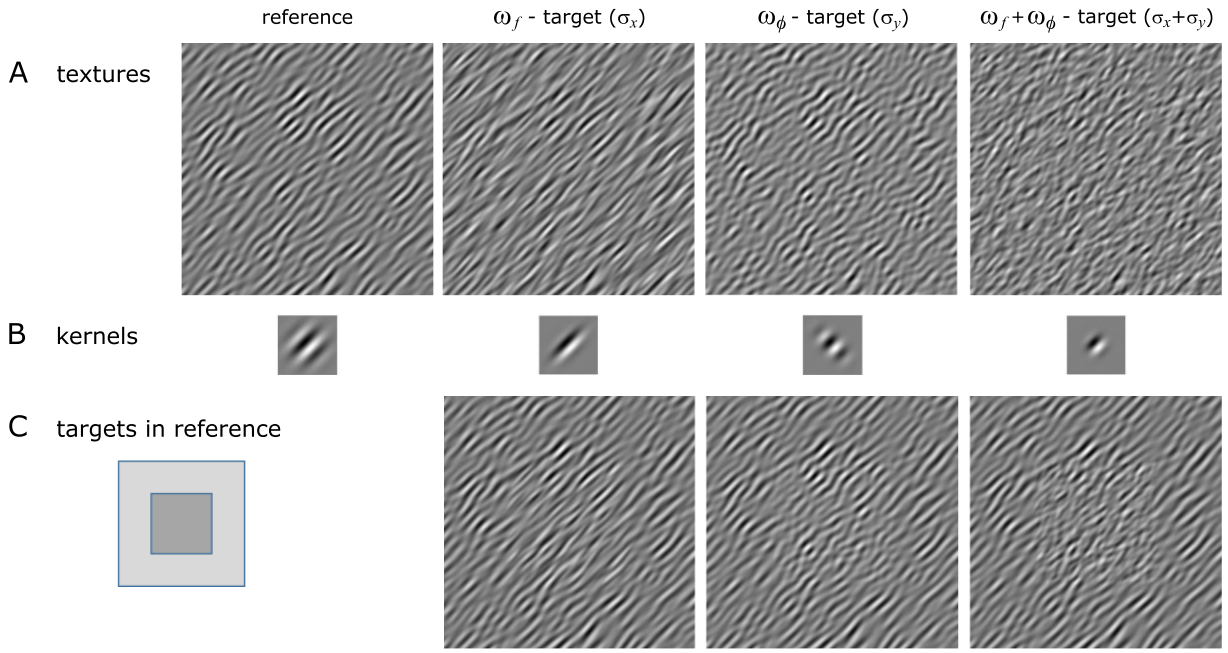

Figure 1. Reference and target textures (A), created by convolving Gabor-like kernels (B) with spatial noise images. The kernels were derived from the reference Gabor kernel by shrinking the spread parameter for the axis of luminance modulation (σ_x), perpendicular to it (σ_y), and doing both ($\sigma_x + \sigma_y$). In (C), the target textures are included in the middle square region, surrounded by the reference texture. Only for $\sigma_x + \sigma_y$, the middle region becomes salient as a square, while the texture figures for σ_x and σ_y manipulations are barely detectable. To avoid artificial texture edges, a cumulative Gaussian was used to smooth the transition from outer to inner square region.

independent and normally distributed random variables with the same mean, with a Gabor kernel with defined orientation and spatial frequency. Textures derived with this technique (*Landy-Bergen textures*) are shown in Figure 1A; the generating kernels are shown in Figure 1B. The first example shows a texture created with a Gabor kernel with equal-spread parameter of the Gaussian envelope in either direction, $\sigma_x = \sigma_y$, which defines the "reference" texture. If the spread parameter along the axis of sinusoidal luminance modulation, σ_x, is diminished, the textures' spatial frequency variability increases. Thus, line segments of similar average length but more inhomogeneous spatial periodicity than the reference appear ("ω_f-target"). If the spread parameter for the axis perpendicular to the axis of luminance modulation, σ_y, is reduced, textures with the same average spatial periodicity but more local orientation jitter shortening the average length result ("ω_ϕ-target"). Shrinking both σ_x and σ_y creates a texture that combines these two effects ("$\omega_f + \omega_\phi$-target"). The four individual textures look quite distinct in a side-by-side view. However, when each target texture is embedded into the reference texture as a smaller square region (Figure 1C), different perceptual effects result. While ω_f- and ω_ϕ-targets are just barely detectable, the $\omega_f + \omega_\phi$-target is strongly salient. These examples indicate that σ_x and σ_y kernel manipulations for Landy-Bergen textures are a further candidate for the synergy effect in texture segregation (Meinhardt et al., 2006).

Selectively changing bandwidths of principal frequency and orientation components

Analyzing the texture examples in terms of Fourier amplitude spectra shows the results of σ_x and σ_y manipulations, which can be expressed in the space spanned by spatial frequency (f) and orientation (ϕ) (see Figure 2).[1] The plots show that, compared to the reference, a wider range of spatial frequencies but not orientations is introduced in the ω_f-target, and a wider range of orientations but not spatial frequencies is introduced in the ω_ϕ-target. The points of maximum difference of targets and reference lie on the axes of principal spatial frequency and orientation, and are found at the same locations for the $\omega_f + \omega_\phi$-target (see red and yellow circles in Figure 2). Figure 3 shows the amplitude characteristics of Gabor filters tuned to the these positions in (f, ϕ)-space ($(f = 3.5, \phi = 26.4)$ and $(f = 5.4, \phi = 0)$) along their connection line (see dashed line in right panel of Figure 2). The plots show that each filter is nearly unresponsive to the (f, ϕ) parameters of the other filter.

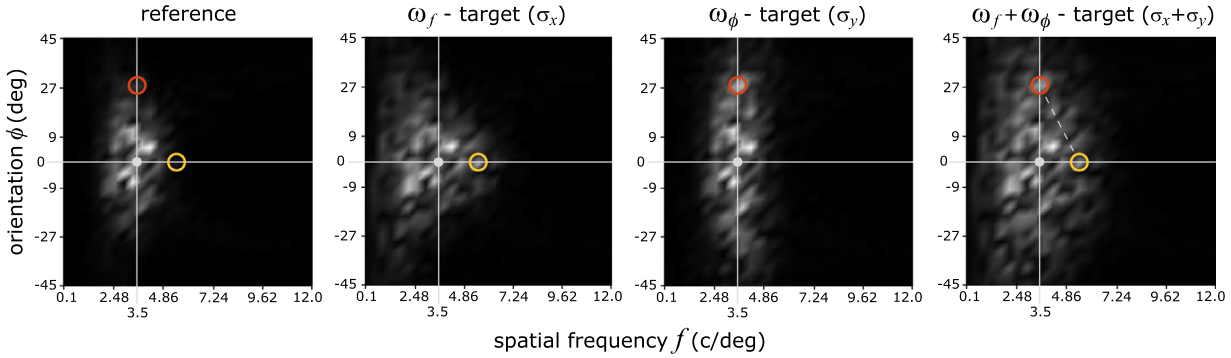

Figure 2. Amplitude spectra for the four texture stimuli shown in Figure 1, plotted in coordinates of spatial frequency (*x*-axis, f) and orientation (*y*-axis, ϕ). The major orientation of the textures (45°) was set to (0°), for convenience (horizontal axis). The major spatial frequency component (carrier frequency of the Gabor-kernel) was 3.5 cpd. The yellow circles mark the coordinates of maximum difference of ω_f-target textures and reference; red circles mark the corresponding coordinates for ω_ϕ-target textures and reference. The intersections of the major spatial frequency and orientation components (light gray dots) mark the spectral centroids, which coincide for all four textures.

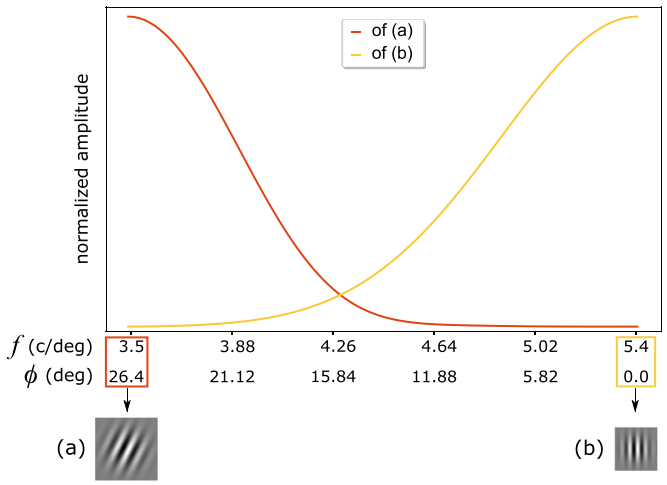

Figure 3. Normalized amplitude characteristics of Gabor filters located at the points of maximum difference of ω_f-target (a) and ω_ϕ-target (b) to the reference texture (see yellow and red circles in Figure 2), for (f, ϕ)-parameters along the connection line of both points in (f, ϕ)-space (dashed line in Figure 2). Spatial frequency filter bandwidths were about 0.5 octaves (see Model section). Filters (a) and (b) have moderate intersection and are mutually unresponsive at each others' (f, ϕ)-coordinates.

For a given choice of σ_x and σ_y, one can calculate spatial frequency and orientation bandwidths of the texture generated by a kernel with these σ parameters, defined as the half-amplitude range for f and ϕ when going along the principal carrier orientation and spatial frequency component (gray lines in Figure 2). Figure 4 shows either bandwidth measure. The oblique planes shown in Figures 4A and 4B illustrate that σ_x and σ_y manipulations affect spatial frequency bandwidth (ω_f) and orientation bandwidth (ω_ϕ) independently. Hence, by selectively manipulating either the σ_x or the σ_y parameter, it is possible to create textures with higher spatial frequency bandwidth but same orientation bandwidth than the reference or vice versa.

It is straightforward to study the psychophysical effect that results when both orthogonal bandwidth manipulations are combined in the $\omega_f + \omega_\phi$-target. In the present study, we report a synergy effect of spatial frequency and orientation bandwidth in texture figure detection and identification, following the same principal experimental setup as used in Meinhardt et al. (2006). Using bandwidth differences of target and reference in the order of magnitude of the detection threshold, we show that combining ω_f and ω_ϕ manipulations leads to strong improvement of form completion and figure-ground segregation in noisy environments.

Stimuli

Stimuli were Landy-Bergen textures (Landy & Bergen, 1991), consisting of a target texture embedded in a reference texture, while both differed in spatial frequency bandwidth, orientation bandwidth, or both. The whole texture display field comprised 16.35° × 16.35° visual angle (768 × 768 px). Within this area, a central square region of 10.9° × 10.9° visual angle (512 × 512 px) was used for displaying target textures in the spatial outline of a right-angle triangle with a leg length of 5.47° (256 px) (see Figure 5 for illustration). The positions of target textures in the central target region changed randomly across trials. Triangle diagonals could be leftward (−45°) or rightward (45°), occurring with equal frequency throughout the experiment.

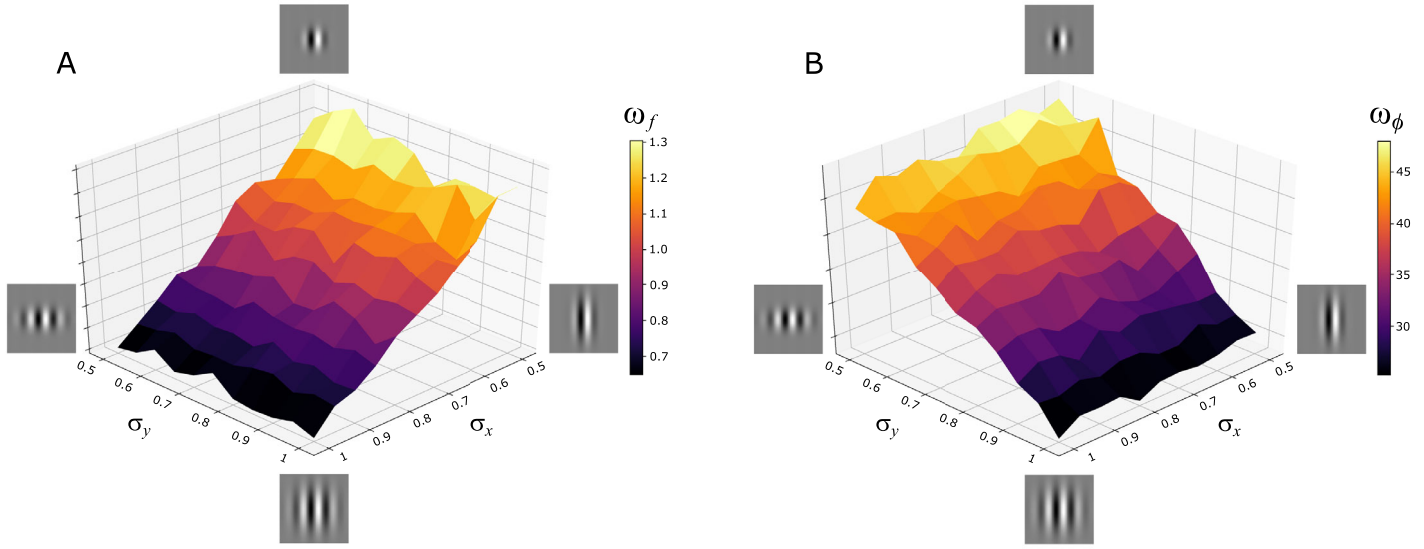

Figure 4. Bandwidth measures ω_f (A) and ω_ϕ (B), defined as half-amplitude parameter distance for f and ϕ for textures generated with given σ_x and σ_y parameters of the Gabor kernels. The *x*- and *y*-axes are normalized with respect to the σ parameters of the reference texture. Measure ω_f is given in octaves and ω_ϕ in degrees. The four edge points correspond to the four textures shown in Figure 1.

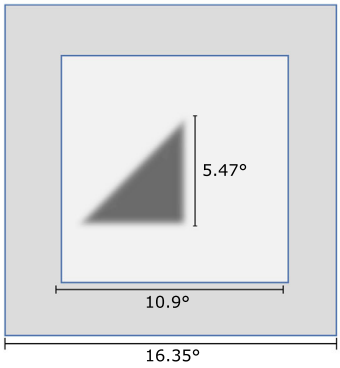

Figure 5. Illustration of spatial parameters of the display. The triangular target texture region could appear in the central portion of the stimulus and could be either variety of a leftward or a rightward diagonal triangle. Gaussian blur was used for the transition from target texture to reference texture.

For generating textures, normal spatial pixel noise was convolved with a Gabor kernel defined by

$$g\left(x, y, \sigma_x, \sigma_y\right) = \exp\left(-\frac{1}{2}\left(\left(\frac{x'}{\sigma_x}\right)^2 + \left(\frac{y'}{\sigma_y}\right)^2\right)\right) \times \sin\left(2\pi f x'\right) \quad (1)$$

with $x' = x\cos\left(\frac{\phi}{180}\pi\right) - y\sin\left(\frac{\phi}{180}\pi\right)$ and $y' = x\sin\left(\frac{\phi}{180}\pi\right) + y\cos\left(\frac{\phi}{180}\pi\right)$. Convolutions were computed via fast inverse Fourier transform, using a bounding box of $1.3^\circ \times 1.3^\circ$ (61 × 61 px) for the kernels. Carrier spatial frequency f was kept constant at 3.5 cycles per degree (cdp) of visual angle for all kernels, and the orientation ϕ was chosen randomly in 10° steps between 1° and 180° for each new trial but was always the same for reference and target kernels. The resulting reference and target textures did only differ in the size and form of the Gaussian window of their kernel controlled by the parameters σ_x and σ_y. Reference textures had a spatial frequency bandwidth of 0.60 octaves and an orientation bandwidth of 22.9 degrees. During the main experiment, the target spatial frequency bandwidth ranged between 0.73 and 2.0 octaves while the target orientation bandwidth ranged between 28.3 and 43.6 degrees. Target and reference textures were blended with smooth transition (standard deviation of Gauss smoothing $\sigma = 0.174^\circ$ [8 px]) as to avoid salient texture edges.

Psychophysical task and detection performance level

A two-alternative forced-choice (2AFC) task for target texture detection was followed by a target figure orientation discrimination task. Participants saw two subsequent stimulus frames, one of which contained a target figure. Subsequently, they indicated by button press whether the first or the second frame contained the target and whether the target was a triangle with leftward (-45°) or rightward (45°) diagonal.[2] Acoustical feedback was provided about correctness by two brief tone signals. Stimulus frame presentation was terminated by masking with static noise with a grain resolution of 1 px. The temporal order of events

was fixation (500 ms), blank (200 ms), first stimulus frame (280 ms), mask (400 ms), blank (200 ms), second stimulus frame (280 ms), mask (400 ms), fixation until response, and acoustical feedback.

The parameter difference in σ_x or σ_y parameters for reference and the target determines the degree of target detectability, as well as the degree to which the texture triangle orientation can be discriminated. For each cue, we used a parameter difference corresponding to a detection rate of 71.4% proportion of correct judgments ($d' = 0.8$), which was individually calibrated for each participant. For double-cue targets, the difference levels in both σ parameters were combined at the 71.4% level. Only one detection level was chosen since previous studies showed strong cue summation for two basic texture cues (spatial frequency and orientation contrast) at this level (Meinhardt et al., 2006; Persike & Meinhardt, 2008).

Participants

Nineteen undergraduate students and one of the authors, Cordula Hunt, served as observers. Seventeen were female and two male. All participants had normal or corrected-to-normal vision. The students were paid and not informed about specific hypotheses or expectations regarding the experimental tests until after participation. All participants signed a written consent form according to the World Medical Association's Helsinki Declaration. Prior to the experiment, they were informed about the procedures and the general intention of the study, and that they could withdraw from the experiment at any time without negative consequences. Further, participants were informed that their data would be collected and stored anonymously and that they could be shared with other researchers for scientific purposes only. After the measurements, a summary and a data explanation were provided for each participant.

Apparatus

Textures were generated using Python 3.6 and the psychopy-Toolbox (Peirce, 2008) and displayed on an EIZO ColorEdge CG2420 monitor. The pixel resolution of the monitor was 1,920 × 1,200 px with a refresh rate of 60 Hz, mean luminance was 100 cd/m^2, and gamma was 1.0 for all three colors. Gray values were taken from a gamma-corrected linear staircase consisting of 256 steps. The room was darkened so that the ambient illumination approximately matched the illumination on the screen. Patterns were viewed binocularly at a distance of 70 cm and participants gave their responses via the arrow keys of the computer keyboard.

Procedure

First, the participant's ability to detect and discriminate targets correctly was verified in a preceding training period in which the participants were made familiar with easily detectable targets. Then, a calibration phase followed, in which the parameters for the main experiment were estimated. These measurements served to determine perceptually equivalent parameter difference levels for both single cues in order to define double-cue targets with equally detectable components (Meinhardt et al., 2006; Persike & Meinhardt, 2008). Participants performed the two tasks at three arbitrarily chosen parameter difference levels. Proportion correct rates were calculated from 32 trials per difference level. The proportion correct data were fitted with Weibull functions, and the parameter differences that corresponded to a proportion of correct judgments of 71.4% ($d' = 0.8$) were estimated. These values were used in the main experiment, which was repeated three times per participant while previous to each run, the σ_x and σ_y parameters were slightly trimmed for better perceptual equivalence in single-cue performance. Every main experiment consisted of the same number of trials for each single cue and the double-cue targets, and all trials were randomly intermixed. Including the calibration the experiment encompassed a total of $192 + 3 \times 96 = 480$ trials. These were divided into two sessions, each lasting under 60 min. Measurements were done either on the same or on two consecutive days.

Measure of cue summation

In a yes-no task, the sensitivity measure from signal detection theory, d', is obtained from the hit rate (H) and the false alarm rate (FA) as

$$d' = \Phi^{-1}(\mathrm{H}) - \Phi^{-1}(\mathrm{FA}). \quad (2)$$

In a 2AFC task, it is uniquely related to the proportion of correct responses and can be obtained as

$$d' = \sqrt{2}\Phi^{-1}(p_c) \quad (3)$$

with $\Phi^{-1}(p)$ the p-th quantile of the standard normal distribution in both cases and p_c the proportion of correct responses (MacMillan & Creelman, 2009). Provided there is equal sensitivity to the individual orientation and spatial frequency bandwidth targets, we define the base sensitivity as the average single-cue sensitivity

$$d'_b = \frac{1}{2}(d'_{\omega_f} + d'_{\omega_\phi}). \quad (4)$$

There is some argument as to how to benchmark the integration of information from different stimulus commodities. The easiest and most intuitive way is to refer to the rules of vector summation as first derived by Tanner (1956). In the case of independent vectors, the two single features are mapped onto orthogonal random variables. The strength of the compound signal is given by the length of the two vectors added:

$$d'_{\perp} = \sqrt{(d'_{\omega_f})^2 + (d'_{\omega_\phi})^2} \quad (5)$$

(MacMillan & Creelman, 2009, see p. 158). This case is referred to as "dimensional orthogonality" (Ashby & Townsend, 1986). If the vectors are instead collinear, sensitivity to the double-cue signal is given by the linear sum of the two equal single-cue sensitivities, $2d'_b$. However, orthogonality is an overly optimal solution and not based in an actual signal detection theory framework. Another pair of measures, probability summation (PS) and additive summation (AS), remedy this. Recently, Kingdom et al. (2015) proposed flexible equations to compute PS and AS. PS assumes separate mechanisms to detect different stimuli in an independent way. Adding multiple features increases information and thus the chance that any one feature will be detected. AS, on the other hand, assumes a mechanism sensitive to the cue combination itself that pools the responses from different individual mechanisms.

Evaluating double-cue effects heavily depends on the benchmark prediction to compare behavioral performance. For a differential comparison of our results, we will consequently report all four measures where this is possible. While computation under the assumptions of dimensional orthogonality and collinearity is straightforward, PS as well as AS predictions need to be calculated under the signal detection framework, taking into account the specific experimental task and stimulus setting. In our study, we had a 2AFC task for detection, this means there were two intervals during each trial, two single cues had to be monitored, and two mechanisms were activated by our compound stimulus. However, a prerequisite for the computation with the formulae provided by Kingdom et al. (2015) is that the experimental setting follows a multiple AFC paradigm. In our study, this is true only for the detection task but not for the discrimination task, which is a yes-no task by structure. We therefore report all four measures for the detection task but only the traditional measures of prediction based on orthogonality and linear summation for the discrimination task. We compute predictions of PS and AS via the Palamedes Toolbox in MATLAB, which were contributed by Prins and Kingdom (2018) for these purposes. For a detailed discription of the utilized formulae, see Kingdom et al. (2015).

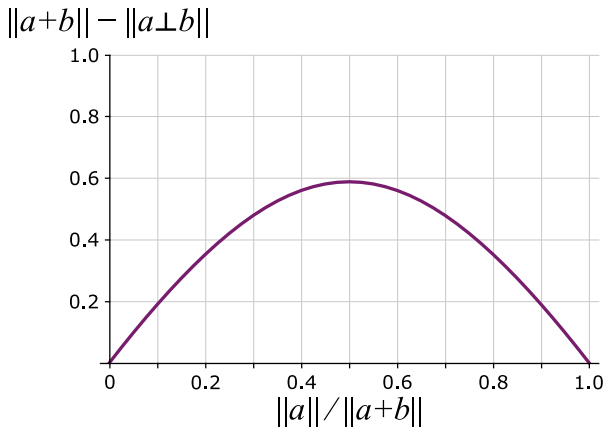

Figure 6. Difference of the sum vectors from collinear and orthogonal component vectors with complementary length $\|\mathbf{b}\| = 2 - \|\mathbf{a}\|$, as a function of the length ratio $|\mathbf{a}\|/\|\mathbf{a}+\mathbf{b}\|$.

Cue summation can be judged by considering the fraction formed by the sensitivity to the combined orientation and spatial frequency bandwidth components, $d'_c = d'_{\omega_f+\omega_\phi}$, and the base sensitivity d'_b

$$q = \frac{d'_c}{d'_b} \quad (6)$$

henceforth referred to as "summation ratio." As a rule of thumb for the strength of synergy effects, two summation ratios are noteworthy. Both express tangible changes in the summation process. Provided a stable baseline ($d'_{\omega_f} = d'_{\omega_\phi}$), $q = \sqrt{2}$ expresses the case of $d'_c = d'_{\perp}$, that is, cue orthogonality. Larger effects have previously been denoted as "synergy" (Kubovy & Cohen, 2001). $q = 2$ expresses the case of $d'_c = 2d'_b$, that is, linear summation. Larger effects violate the triangle inequality $\|\mathbf{a}+\mathbf{b}\| \leq \|\mathbf{a}\| + \|\mathbf{b}\|$ and can be denoted as "oversummative."

Perceptual equivalence of the single cues is an important constraint for measuring cue summation effects under all summation frameworks. If the single-cue sensitivities differ largely, true cue summation effects (synergy) can hardly be distinguished from cue independence. Figure 6 shows the difference of the sum vectors from collinear and orthogonal component vectors (directly pertaining to orthogonal summation, but the principle is the same for PS and AS). This difference is maximum, that is, $2 - \sqrt{2}$, for component vectors of equal length and rapidly declines the more the lengths of the component vectors differ.

Data clearing

Proportions correct for perfect performance were replaced by $1 - (2N)^{-1}$, where N is the number

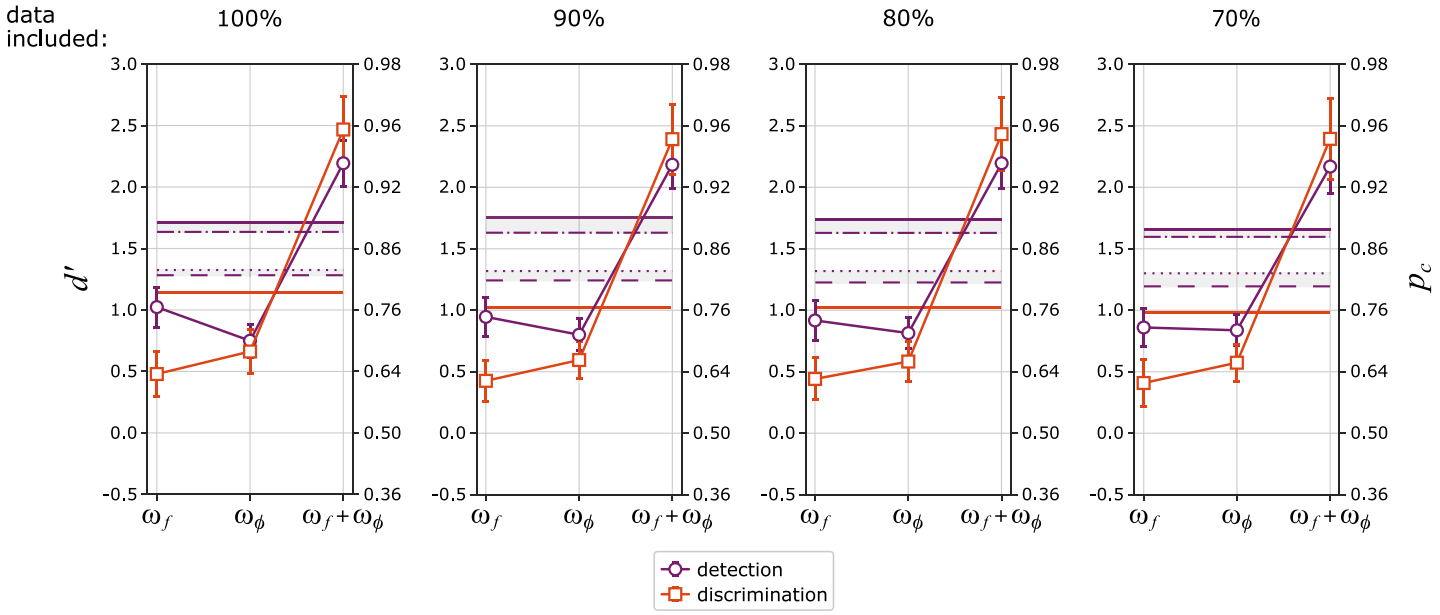

Figure 7. Mean d' sensitivities for target texture detection (purple circles) and target figure orientation discrimination (orange squares) for spatial frequency bandwidth (ω_f) targets, orientation bandwidth (ω_ϕ) targets, and double-cue ($\omega_f + \omega_\phi$) targets. The different panels refer to all data (left panel) and to 90%, 80%, and 70% of the data remaining in the sample after clearing for single-cue sensitivity deviations, $\Delta d'_b$. Bars denote 95% confidence intervals of the means. The right ordinate shows proportion correct (p_c) rates corresponding to d'. The lower dashed line indicates predicted double-cue sensitivity for independent cues, $\sqrt{2}d'_b$, while the lower dotted line indicates double-cue sensitivity as predicted by PS. Upper solid lines indicate predicted double-cue sensitivity for linear cue summation, $2d'_b$, while the upper dashed-and-dotted line indicates double-cue sensitivity as predicted by AS. For a better overview, only the most conservative criterion, $2d'_b$, is depicted for orientation discrimination.

of replications (MacMillan & Creelman, 2009, see p. 8). Analysis of single-cue sensitivities showed strong deviations from equality in places, albeit the σ- parameters were carefully adjusted for perceptual equivalence prior to the main experiment. To analyze the effects on the cue summation measure (6), we formed the distribution of $\Delta d'_b = d'_{\omega_f} - d'_{\omega_\phi}$ for each target detection and target figure orientation discrimination and sequentially removed all data beyond the 90th, 80th, and 70th percentiles in each distribution to achieve stepwise improvement of single-cue equivalence (see Results). Since this technique implies that missing data likely occur for almost every participant, we used linear mixed models for statistical analysis, which is more appropriate in these cases than repeated-measures analysis of variance (McCulloch & Searle, 2001).

Results

Figure 7 shows the results for the d' measure for target texture detection (purple circles) and discrimination (orange squares) for the complete sample and 4 degrees of clearing the sample for unequal single-cue sensitivity. With all data included, ω_f-targets were better detectable than ω_ϕ-targets, $t(df) = 2.67(56)$, $p = .010$. However, removing data with the most extreme 10% single-cue sensitivity deviations already resulted in just slightly unequal equal single-cue detection performance, $t(df) = 1.54(51)$, $p = .130$ (Figure 7, second panel from the left). There, the detection baseline d'_b was stable at the level that was targeted on the basis of the preparatory calibration measurements ($d' = 0.8$). The same was true for removing most extreme 20% and 30% single-cue deviations (20%: $t(df) = 1.25(45)$, $p = .217$; 30%: $t(df) = 0.3(39)$, $p = .764$). Target figure discrimination performance was somewhat lower than target detection performance. Including discrimination data only if the previous detection was successful stabilized the baseline throughout at slightly above $d' = 0.5$. Cue summation was reduced by this but remained markedly larger that under detection (for more information, see Figure 14 in Appendix A). Both tasks reflected a very strong double-cue advantage. Table 1 shows the q-ratios (6), which reflect the strong benefit of double-cue targets compared to the single-cue baseline. The summation ratio q results were quite stable across the different degrees of clearing for single-cue sensitivity deviations. For all four samples, the table

Detection

					Tanner (1956) measures				Kingdom et al. (2015) measures			
% incl.	*n*	$\bar{d}'_b$	$\bar{d}'_c$	*q*	$d'_\perp$	$d'_c - d'_\perp$	$2d'_b$	$d'_c - 2d'_b$	d'_{PS}	$d'_c - d'_{PS}$	d'_{AS}	$d'_c - d'_{AS}$
100	57	0.885	2.19	2.47	1.27	0.924	1.71	0.484	1.33	0.862	1.65	0.544
90	52	0.874	2.18	2.49	1.24	0.940	1.75	0.430	1.32	0.862	1.64	0.540
80	46	0.867	2.20	2.54	1.23	0.972	1.73	0.472	1.32	0.885	1.64	0.564
70	40	0.834	2.15	2.58	1.18	0.970	1.67	0.482	1.29	0.864	1.60	0.555

Discrimination

					Tanner (1956) measures			
% incl.	*n*	$\bar{d}'_b$	$\bar{d}'_c$	*q*	$d'_\perp$	$d'_c - d'_\perp$	$2d'_b$	$d'_c - 2d'_b$
100	57	0.571	2.47	4.32	0.817	1.65	1.14	1.33
90	52	0.510	2.39	4.69	0.731	1.66	1.02	1.37
80	46	0.510	2.43	4.76	0.728	1.70	1.02	1.41
70	40	0.490	2.39	4.88	0.703	1.69	0.980	1.41

Table 1. Sensitivity advantage of double-cue targets compared to the base sensitivity level for different percentages of data included. The table shows the number of data triplets included for each task *n*, mean base sensitivity level $\bar{d}'_b$, the mean sensitivity for double-cue targets $\bar{d}'_c$, and the ratio of double-cue to single-cue performance *q*. Further, the prediction of orthogonality $d'_\perp$, the prediction of linear summation $2d'_b$, the prediction of PS d'_{PS}, the prediction of AS d'_{AS}, and the corresponding differences of sensitivity for double-cue targets to each of the predictions are depicted.

shows summation ratios *q* of about 2.5, indicating stronger summation than expected from linear cue integration ($q = 2$) for target texture detection, while even summation ratios *q* of about 5 were reached in target figure orientation discrimination. Hence, results indicate "oversummative" cue combination effects for detection and discrimination of orientation and spatial frequency bandwidth targets.

Since the single cues were nearly equally detectable when data with at least the most extreme 10% single-cue sensitivity deviations were removed, we analyzed the samples with 90%, 80%, and 70% data included with a linear mixed model (LMM). Analyses were completed in R (R Core Team, 2012) with the lme4 package (Bates et al., 2015) and the MuMIn package (Barton, 2013) for estimating global model fit achieved with fixed and random effects (R^2). We predicted d' performance from a model including *Feature* (d'_b "baseline," d'_c "combined"), *Task* ("detection," "discrimination"), and their interaction as fixed effects, as well as random intercepts for participants and measurement times to account for the repeated measurements structure of our data. Table 2 summarizes the LMM analysis for the 90% sample, which explained 70.8% of d' variance. While the large slope for *Feature* (i.e., the average d' difference for single-cue and double-cue targets of 1.31 d' units) contributed the strongest significant effect, $t(df) = -13.26(187)$, $p < .001$, the d' difference between tasks of 0.44 d' units was also highly significant, $t(df) = -4.45(187)$, $p < .001$.

Factor	Estimate (SE)	*t* (*df*)	*p*	R^2
Intercept	2.18 (.117)	18.64 (14.7)	<.001	.708
Feature	−1.31 (.099)	−13.26 (187)	<.001	
Task	−.441 (.099)	−4.45 (187)	<.001	
Feature × task	−.074 (.140)	−.528 (187)	.598	

Table 2. Fixed-effects estimates and global model fit (R^2) of the LMM model.

The *Feature* × *Task* effect was not significant, $t(df) = -0.528(187)$, $p = .598$. The model residuals were analyzed with the Kolmogorov-Smirnov Lilliefors test (Lilliefors, 1967), indicating no violation of normality ($D = .049$, $p = .256$). Analyzing the 80% and 70% samples yielded practically the same results picture, with highly significant *Feature* and *Task* effects while the *Feature* × *Task* interaction was not significant. The latter finding indicated that there was an almost equal sensitivity decline for the baseline and the cue combination condition when comparing detection to discrimination. Constantly lower sensitivities raise the summation ratios *q* for the discrimination task. However, the absence of a *Feature* × *Task* interaction gives no support for concluding a larger cue summation effect in discrimination compared to detection. Also, for target figure detection, summation ratios *q* were found to be larger at lower base sensitivity levels (see Table 1 in Meinhardt et al., 2006).

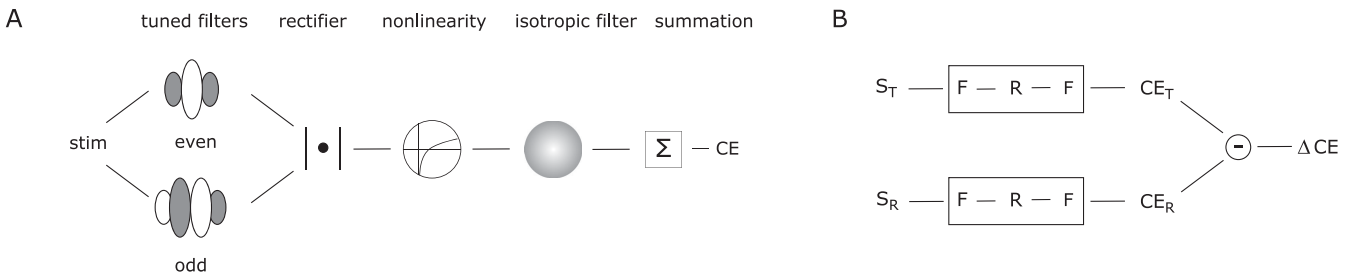

Figure 8. Scheme of the multiscale texture analysis model. (A) Filter-Rectify-Filter (FRF) chain. Texture input is convolved with a quadrature pair of even and odd Gabor filters selectively tuned to orientation and spatial frequency. Convolved images are rectified and passed through a compressive nonlinearity. The resulting gain-controlled local energy distribution is smoothed by a large-scale isotropic filter and then pooled into a space-average mean contrast energy, CE. (B) Subtracting the CE values for target and reference textures yields a net contrast energy measure, ΔCE.

Modeling synergy with an energy-based model

Outline

Experimental results reflected oversummative double-cue advantage in either task. In the modeling part, we show that oversummative double-cue advantage is captured by spatial frequency and orientation selective Filter-Rectify-Filter (FRF) mechanisms. We restrict modeling to the detection task, since modeling the discrimination task requires the assumption of further stages of shape processing (e.g., edge detection and form template matching [Landy & Bergen, 1991] or hierarchical feature coding [Riesenhuber & Poggio, 1999]), which is beyond the scope of this study.

In the first part, we demonstrate that the maximum energy difference to target and reference textures ("net contrast energy"), obtained from FR or FRF mechanisms at multiple scales and orientations, agrees fairly well with the summation ratio observed for double-cue texture targets ($q \approx 2.5$). In the second part, we include the effects of sensory noise on the energy distributions as well as on net contrast energy of each FRF chain and use a signal detection framework to predict detection performance.

Part I: Net contrast energy from multiple FRF mechanisms

We suggest a local energy-based model that combines the first stages of an FRF model (i.e., a bank of Gabor filter pairs in quadrature, local energy computation, and nonlinear compression) with the subsequent computation of contrast energy (CE), coding for contrast strength from local energy maps. Contrast energy has been shown to be associated with behavioral responses as well as single-trial event-related potentials during natural scene classification (Groen et al., 2013; Scholte et al., 2009). A model overview is depicted in Figure 8; details of the implementation are provided in Appendix B.

To identify the crucial stages for texture cue summation, we considered two varieties, the basic FR chain and the full FRF chain, which included the large-scale secondary filter. We first tried the FR and then the FRF variety.

Initially, the input was filtered with a bank of Gabor filters with variable spatial envelopes following a flexible scaling principle that smoothly increased spatial frequency resolution in the midrange of spatial frequency centered at the carrier spatial frequency of the texture stimuli. Filter bandwidths were narrowest ($\approx$ 0.5 octaves) for medium spatial frequencies between 2.5 and 5 cpd while they approached 1 octave for lower and higher spatial frequencies (see Table 3; see Appendix B for details of the bandwidth modulation). A dense frequency sampling was realized in the range of 0.7 cpd to 11.0 cpd, and orientation sampling was done in 3.5-degree steps. Local energy was computed from even and odd filter outputs and then passed through a compressive nonlinear transducer to account for gain control in texture discrimination (Legge & Foley, 1980; Rubenstein & Sagi, 1990; Motoyoshi & Nishida, 2001).

Thus, each even-odd filter pair localized in (f, ϕ)-space translated a texture into a gain-controlled local energy map, $E_{f,\phi}(x, y)$. Contrast energy, CE, was calculated from each map by taking its space-average mean value (Groen et al., 2013). Since also reference textures with no embedded target elicit contrast energy, we used the CE difference for a texture with target embedded (T) and reference texture (R), $\Delta CE = CE_T - CE_R$, to describe the differential response of each (f, ϕ)-tuned filter pair to target and reference textures ("net contrast energy") of a 2 AFC trial. The distribution of net contrast energy across filter spatial frequencies and orientations is shown

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
f_i	0.69	0.89	1.13	1.39	1.68	2.01	2.38	2.78	3.24	3.76	4.37	5.08	5.96	7.08	8.63	10.92
ω_i	0.89	0.77	0.69	0.62	0.57	0.53	0.5	0.48	0.47	0.47	0.48	0.5	0.55	0.64	0.79	0.99

Table 3. Center spatial frequencies and half-amplitude bandwidths (octaves) for the primary Gabor filters in the model.

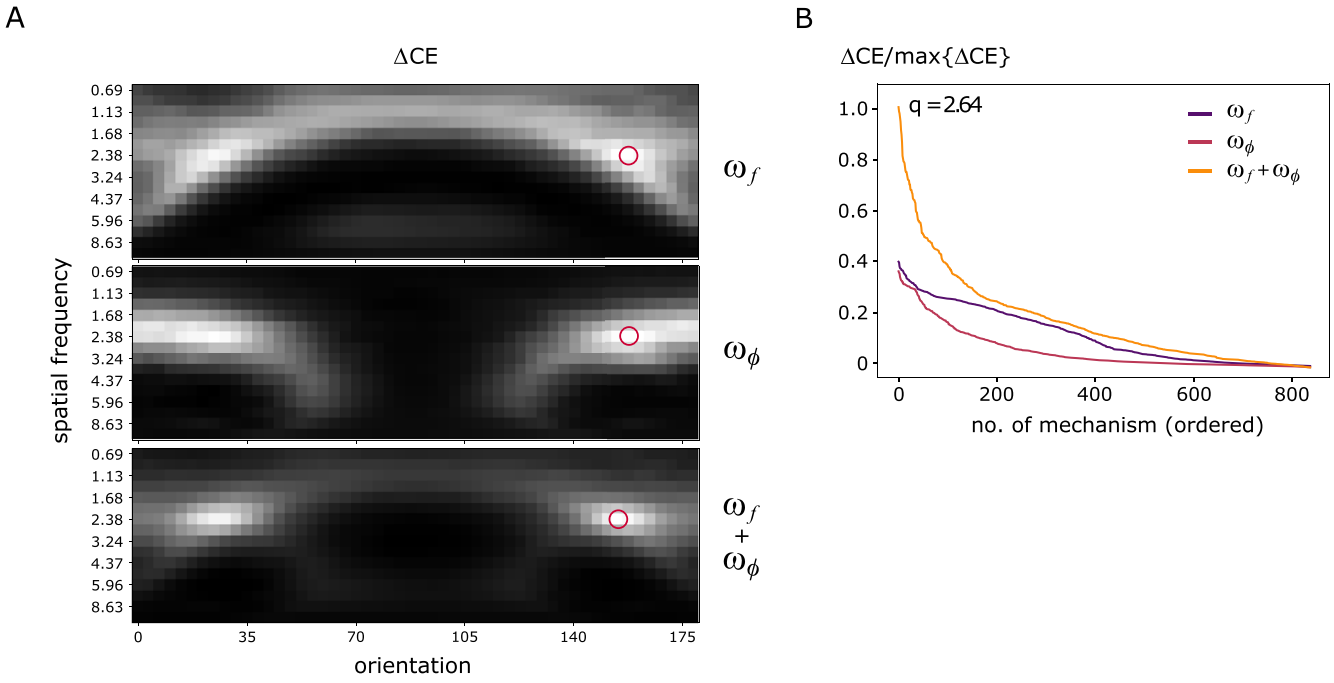

Figure 9. (A) Net contrast energy distributions obtained from the bank of filters jointly tuned to orientation (0°–180°, *x*-axis) and spatial frequency (0.69–10.92 cpd, *y*-axis). Net contrast energies are displayed as luminance values for ω_f-targets, ω_ϕ-targets, and the double-cue ($\omega_f + \omega_\phi$)-targets. The maxima of the distributions are indicated by red circles. (B) Size-ordered normalized net contrast energies from all FR mechanisms (scree plot). The summation ratio q calculated from the maxima of the three target conditions (mechanisms with element no. 0) was $q = 2.64$.

in Figure 9A. The FR mechanisms with maximum responses to ω_f-, ω_ϕ-, and $\omega_f + \omega_\phi$-targets are very close in (f, ϕ)-space. The distributions of net contrast energy are nearly symmetrical with respect to the 90° axis. Peaks are centered on 20° and 160° and at a spatial frequency of 2.38 cpd. For double-cue targets, the most responsive maps come from the small common peak region for both single-cue targets, forming a small subset in (f, ϕ)-space. Note that it is an inherent property of local energy models that the most responsive energy mechanisms are tuned to orientations and spatial frequencies that do not match those most prominently present in the stimulus (Prins & Kingdom, 2006). Here, the strongest net contrast energy comes from primary filters with about half an octave lower carrier spatial frequency than the texture stimuli. This is in line with results from adaptation experiments showing that units centered at lower spatial frequencies than the target stimuli mediate discrimination performance (Regan & Beverley, 1983), a finding that is explained by multichannel models with response pooling and compressive nonlinearity (Wilson & Regan, 1984).

The scree plot of net contrast energy for $\omega_f + \omega_\phi$ targets declines sharply, while the curves for both single-cue targets fall at slower rates (see Figure 9B). This indicates an oversummative double-cue advantage only for the maximally responding FR mechanism, or just few FR mechanisms around the maximum. A way to substantiate a **max**-decision rule ("winner-take-all" rule) is to apply a p-norm to a data set and fitting the exponent, p. A large value of p indicates that only the largest values of a set determine the p-norm (Micko & Fischer, 1970), which means that there is practically no summation among set items.[3] Using this approach, we calculated the final model output, ΔCE_m, as a p-norm of all net contrast energies

$$\Delta CE_m = \left(\sum_{f,\phi} |\Delta CE|^p \right)^{1/p} \quad (7)$$

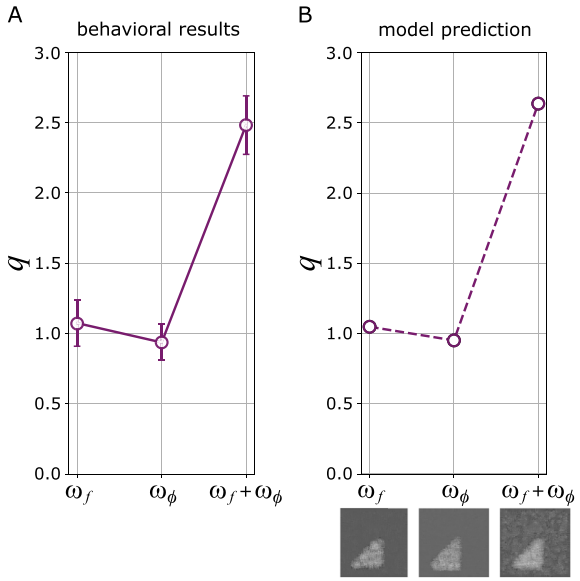

Figure 10. Behavioral detection data (A) compared to the model prediction (B). The empirical cue summation ratio was $q = 2.49$ (90% data included; see Figure 7), while the model predicted a gain of $q = 2.64$. Below the abscissa, the most responsive energy maps for each target texture type are depicted. Data were normalized at baseline sensitivity, d'_b, for convenience. The bars indicate 95% confidence intervals of the means for behavioral data. For the model predictions, five independent runs with newly computed random textures for target and reference were carried out. The data show mean normalized CE_m values of these five runs. Confidence intervals are quite small and fall within the circle symbols.

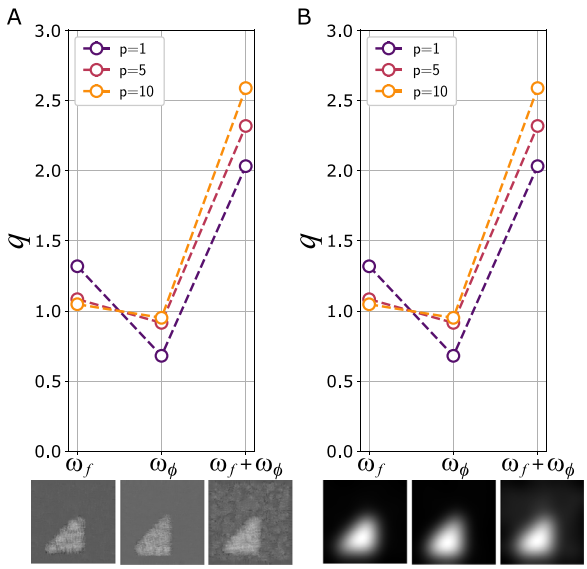

Figure 11. Effect of different p-norms for integrating net contrast energies, ranging from linear summation ($p = 1$) to large p values ($p = 10$) approaching a winner-take-all rule, for models without (A) and including (B) a second stage of isotropic Gaussian filtering. Best responsive energy maps are displayed for each target kind.

and fitted the summation exponent, p. To do so, we normalized the behavioral d' data at the mean base sensitivity level, $\bar{d}'_b$, and also normalized ΔCE_m values at the mean ΔCE_m of the two single-cue conditions. This means that performance is expressed in units of the summation ratio, q (6), both for the psychophysical data and FR model predictions. Varying p from 1 to 15 showed settled agreement of normalized data and normalized model ΔCE_m output for values of $p > 10$, confirming a **max**-rule for the integration of net contrast energies (see Figure 10). While participants showed a ratio of $q = 2.49$ (see Table 1 for 90% data included) for double-cue targets, the model predicted a close ratio of $q = 2.64$. The model also reproduced the slight advantage in baseline performance for ω_f- compared to ω_ϕ-textures.[4] These results substantiate that oversummative double-cue advantage comes from only few maximally responding FR mechanisms, while there is practically no summation across mechanisms.

Since the standard FRF model includes a second-order filtering stage (Landy, 2013), we also added second-order filtering, using a large-scale isotropic Gaussian with a standard deviation of $\sigma = 0.713°$. Figure 11 shows results for normalized net contrast energy computation without (A) and with (B) second-order filtering. While the second-order filter clearly enhanced the target region from the background, practically the same summation ratios q resulted from the FR and the FRF versions of the multiscale local energy model. Also the effects of different p-norms for integrating net contrast energies in (f, ϕ)- space were nearly identical.

These results show that net contrast energy, obtained from the maximally responding spatial frequency and orientation selective local energy mechanisms, captures the oversummative double advantage in the order of magnitude observed in the detection task. Comparing the results for FR and FRF chains shows that oversummative cue integration occurs in the first steps, joint spatial frequency and orientation selective coding, followed by a nonlinearity.

Part II: A signal detection theory framework

The modeling results of Part I show that a multiscale local energy model can account for oversummative double-cue effects in terms of net contrast energy. However, a more comprehensive approach would model the observers' detection performance in terms of

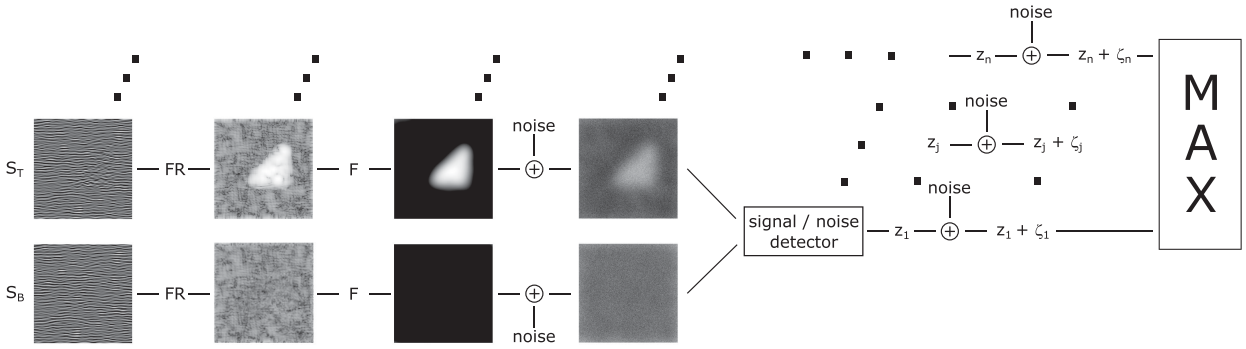

Figure 12. Outline of the detection model based on multiple FRF mechanisms. Texture is passed through the FR stage, resulting in spatial local energy distributions for target and reference textures. The second-order filter averages over larger regions, thus filtering out much of the spurious local energy fluctuations while enhancing the target region. Since an FRF chain is assumed behind each retinal location, the output for a definite spatial coordinate (x, y) is contaminated by noise that originates from spontaneous fluctuations within each FRF chain. A detector unit linked to each FRF chain averages across the spatial energy distributions and evaluates their mean difference relative to the measured spatial variability, yielding a standard score for signal-to-noise distance as the final output of the j-th FRF mechanism. After adding noise to each score, all scores are integrated according to a **max**-rule at the final decision stage.

measured sensitivity, d'. To do this requires to consider the effects of sensory noise.

The detection model assumes two classes of basic neural units operating at distinct levels, each with its own noise source. First, we assume spatial frequency and orientation selective FRF mechanisms as the neural units behind each retinal location. If random activation generated in these units adds to their local energy response, the local energy distributions are overlaid by *spatial noise*. Second, we assume that the response of each j-th FRF mechanism is integrated by a later stage unit, which calculates the CE difference of target and reference local energy distributions (see Part I of modeling section). In the presence of spatial noise, this response is measured in units of the standard deviation of local energy, thus yielding a signal-to-noise ratio score, z_j. Noise that originates at the level of these later units may add as well (*channel noise*). The observer is assumed to rely on the j-th FRF channel with the largest response. The model is outlined in Figure 12.

Contrast energy detection in the presence of noise

Neural systems show fluctuating responses to repeated presentations of the same visual stimulus (Tolhurst & Dean, 1983), while the variance is level dependent (Roufs, 1974). To account for level dependency with a simplified assumption, we used Crozier's law (Crozier, 1936), which claims constancy of the ratio of standard deviation to the mean, c_v, known as variation coefficient. Constrained by linearity assumptions in the vicinity of detection thresholds, Roufs (1974) showed that constancy of the variation coefficient implies detector operation at a constant signal-to-noise ratio for Gaussian noise. Assuming

$$\frac{\sigma}{\mu} = c_v = const, \quad (8)$$

we set $\sigma_j = c_v \mu \left[\hat{E}_{j,R}\right]$, with $\mu \left[\hat{E}_{j,R}\right]$ the space-average energy of the j-th FRF chain for a reference texture (see Appendix B for the definition of $\hat{E}$). Here, j is a running index for (f, ϕ) combinations, $j \in [n]$, $[n] = \{1, \ldots, n\}$, n the number of mechanisms. The value of c_v was chosen such that the distribution function of the noise had a slope parameter of $\beta = 3$ if the Weibull function was used as an approximation to a Normal distribution. Setting $\beta = 3$ results in a variation coefficient of $c_v = 0.384$ (see Equation 33 in Appendix C). A value of $\beta = 3$ is a good overall estimate for the slope parameter of psychometric curves in many psychophysical tasks (Robson & Graham, 1981; Watson, 1982; Graham, 1989; Wallis et al., 2013; Kingdom et al., 2015). To mimic spatial noise, we added a sample $\xi_j(x, y)$ from a Normal distribution $N(0, c_v \mu_j)$ to each point of the local energy distributions for target and reference stimuli of the j-th FRF chain. To control noise strength while maintaining a constant variation coefficient, each sample was multiplied by a spatial noise factor, n_x. The resulting energy distributions,

$$\tilde{E}_j(x, y) = \hat{E}_j(x, y) + n_x \xi_j(x, y), \quad (9)$$

have two independent sources of random variation, one stemming from the local energy response to a spatial random signal (see subsection "Stimuli" in Method section) and the other from external Gaussian noise.[5] Since in target intervals, local energy is increased in a given spatial region while the background is generated with the same rule as for nontarget intervals (see example pictures in Figure 12), a classical signal-to-noise ratio detector (Green & Swets, 1966/1988) would measure the separation of energy means in units of spatial energy variation in the reference (nontarget) distribution:

$$z_j = \frac{\mu\left[\tilde{E}_{j,T}\right] - \mu\left[\tilde{E}_{j,R}\right]}{\sqrt{\mathrm{VAR}\left(\hat{E}_{j,R}(x,y)\right) + n_x^2 \mathrm{VAR}\left(\xi_j(x,y)\right)}}. \quad (10)$$

The z_j score can be considered a d' measure, signal-to-noise ratio, calculated by a contrast energy detector operating on the spatial energy distributions of the j-th FRF chain. Now, since $\mu[\xi_j] = 0$ for all $j \in [n]$, the numerator of (10) reduces to the mean difference of energies, that is, $\mu\left[\tilde{E}_{j,T}\right] - \mu\left[\tilde{E}_{j,R}\right] = \mu\left[\hat{E}_{j,T}\right] - \mu\left[\hat{E}_{j,R}\right]$. This means that

$$z_j = \frac{CE_{j,T} - CE_{j,R}}{\sqrt{\mathrm{VAR}\left(\hat{E}_{j,R}(x,y)\right) + n_x^2 \mathrm{VAR}\left(\xi_j(x,y)\right)}} \quad (11)$$

measures the signal-to-noise ratio of net contrast energy, ΔCE, from the j-th FRF chain.

Also, in these later-stage neural units, spontaneous random activity may be generated (*channel noise*). This can be modeled by adding independent Gaussian noise to each z_j score in a given experimental trial. Since the detector outputs are standard scores sharing the same scale, we used the mean of all z_j baseline scores to estimate the standard deviation of the noise, again assuming a variation coefficient of $c_v = 0.384$. Noise samples ζ_j from $N(0, c_v\mu_z)$ were scaled by a factor n_c to control the magnitude of the noise. As in Part I of modeling, a **max**-rule is suggested for the integration of all detector outputs,

$$d = \max_{j\in[n]}\{z_j + n_c\zeta_j\}, \quad (12)$$

which means that the maximum of all randomly confounded signal-to-noise ratios determines the observers' response.

Modeling results

To calculate model predictions, we used 128 independent runs with newly computed random textures. For creating target texture sets, we used the across-subjects median values for ω_f and ω_ϕ parameters. With this basic setup, several settings for the noise amplitude parameters n_x and n_c were tried. Model calculations were done both for the FR model and the FRF model. Figure 13 gives an overview of the results for the full FRF model.

Figure 13A, panel b shows results for $n_x = 0.125$ and $n_c = 1.0$. The scree plots and the results plot in the format of Figure 7 indicate that the FRF model outlined in Figure 12 fairly well explains the behavioral detection data with these settings. Measured d' and model d predictions nearly coincide for ω_f and ω_ϕ baselines and the $\omega_f + \omega_\phi$ double-cue condition. The summation ratio $q \approx 2.5$ could be exactly replicated by the model.

The major effect of increasing spatial noise amplitudes (compare a to d in Figure 13A) is a continuous drop in the signal-to-noise ratio, d. This is illustrated by the target energy distributions in the upper row of the figure, which show a fading triangle with increasing spatial noise. Apparently, this affects baseline and double-cue conditions to equal degrees, since the summation ratio q is not affected, while performance in all three conditions gradually falls. Hence, the oversummative cue summation effect proves to be robust against spatial noise. The model suggested here would predict oversummative cue integration over the whole usable range of baseline performance in a cue summation experiment, from lowest (as low as $d = 0.25$) to high (as high as $d = 1.4$) baseline performance likewise.

The major effect of increasing channel noise amplitudes (compare a to d in Figure 13B) is a perturbation of the distribution of detector responses in (f, ϕ)-space. The density plots illustrate that the distribution of $z + n_c\zeta$ becomes more and more random due to the increase of the random part $n_c\zeta$ relative to z. The density plots illustrate a randomly sampled picture for one trial, which means that each (f, ϕ)-tuned cell was overlaid by one random sample $n_c\zeta$. Apparently, the maxima d come from the same (f, ϕ)-tuned FRF mechanisms in baselines and the double-cue condition for $n_c \leq 1$ (see red circles), while this no longer holds for $n_c \geq 2$. Generally, the distribution of detector responses in (f, ϕ)-space appears to be more resilient against channel noise for the double-cue condition compared to the single-cue condition. To further explore the effects of increasing channel noise, we varied n_c while n_x was constant at $n_x = 0.125$. From 128 trial replications, we determined the maxima d for double-cue and baseline conditions, as well as their fraction, q. Table 4 shows means and standard deviations. The summation ratio q falls from 2.70, obtained without channel noise, to 1.97 for $n_c = 4$. With increasing n_c, the maxima of $z + n_c\zeta$ are more and more determined by the channel noise component $n_c\zeta$ while the z component remains at its local energy-driven level. Increasing the

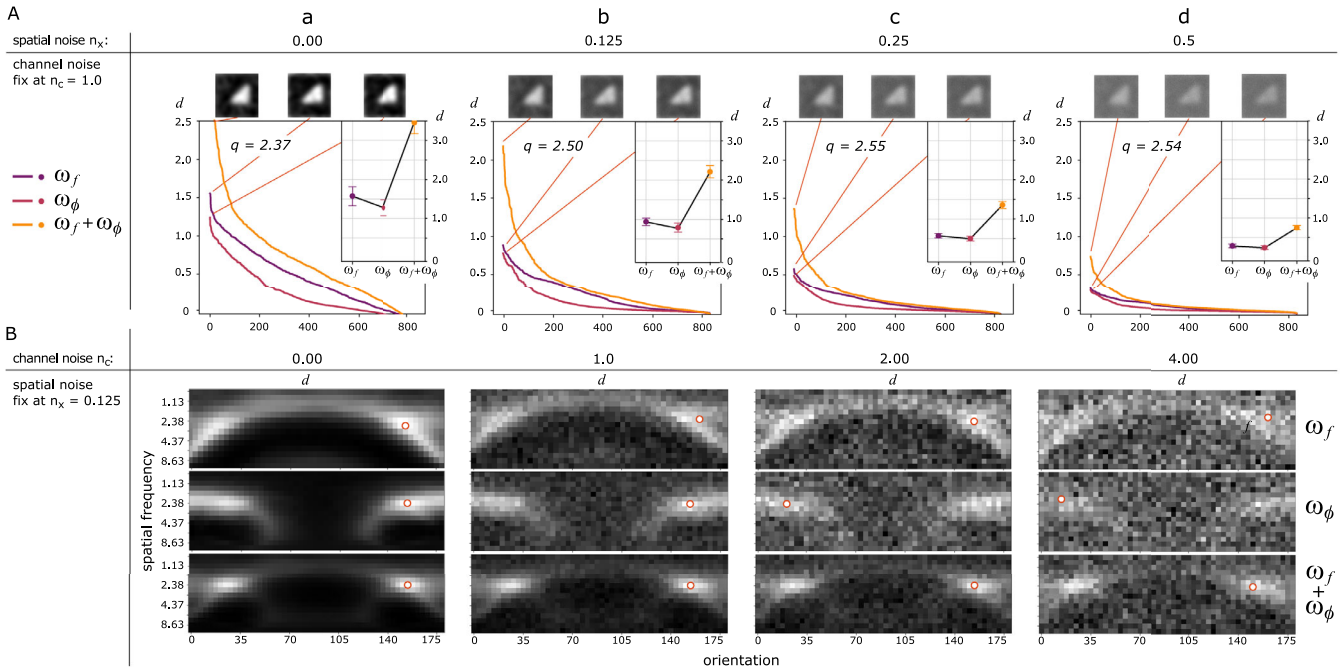

Figure 13. Overview of the model simulation results. (A) a to d show the effects of increasing the spatial noise factor n_x for the maxima d in the resulting scree plots, the summation ratio q, and the results plots in the format of Figure 7. The channel noise factor was held constant at $n_c = 1.0$ in all calculations. Horizontal bars indicate 95% confidence intervals of the means from 128 replications. (B) The effects of increasing the channel noise factor n_c on the distribution of detector outputs $z_j + n_c\zeta_j$ in (f, ϕ)-space for a single trial. Red circles mark the maxima for each condition. The spatial noise factor was held constant at $n_x = 0.125$ in all calculations.

n_c	0	0.5	1	2	4
$\bar{d}_b$	0.808	0.829	0.871	0.975	1.22
σ_{d_b}	0.000	0.023	0.035	0.062	0.103
$\bar{d}_c$	2.18	2.18	2.19	2.25	2.40
σ_{d_c}	0.000	0.028	0.051	0.086	0.141
$\bar{q}$	2.70	2.63	2.51	2.31	1.97
σ_q	0.000	0.056	0.082	0.133	0.153

Table 4. Average d measures for baseline and double-cue conditions, summation ratio, q, and standard deviations of all measures, for increasing channel noise factor, n_c. Means were calculated from 128 trial replications.

random component thus progressively masks the local energy-rooted differences of double-cue and baseline conditions. The data in Table 4 therefore indicate that stronger channel noise is not compatible with the observed double-cue advantage in the detection task.

Since the modeling results of Part I showed that both the FR and the FRF variety could explain the summation ratio q, we also tested the FR model in the presence of noise. The calculations showed very low baselines of $d \approx 0.25$ for the FR model even with no spatial noise. The variability in the local energy response to a spatial random signal, obvious as "salt and pepper" noise in the local energy distributions (see density plots of FR output in Figure 13), was too strong to allow larger signal-to-noise ratios (Equation 11). Hence, the large-scale second-order filter was necessary to raise the signal-to-noise ratio for single texture cues into the empirically observed range. The largely increased signal-to-noise separation thus motivates one to assume a larger-scale second-order filtering stage after the primary local energy extraction.

Discussion

We investigated interactive nonlinear cooperation (synergy) in texture figure detection and discrimination using target textures defined by bandwidth enlargement in spatial frequency and orientation. Performance for double-cue targets was better than predicted by independent processing of either cue and even better than predicted from linear cue integration. This indicates that texture cues are not handled as independent "features." Particularly, the finding of oversummative cue combination effects suggests highly effective cue integration in a specific mechanism rather than separate sensory channeling and later integration.

Application of an energy-based texture-processing model revealed that the oversummative double-cue

advantage is captured by assuming feedforward Filter-Rectify-Filter processing chains at multiple spatial scales and orientations, differential response calculation of pooled contrast energy for stimulus alternatives, and a **max**-decision rule for the integration of channel outputs. Modeling results showed that the observers' sensitivity to single-cue and double-cue texture targets, measured in d' units, could be reproduced with plausible settings for filter and noise parameters. Comparing FR and FRF model varieties indicated that the oversummative cue combination advantage roots in the early energy extraction (FR) stage, while replication of detection performance in terms of signal-to-noise ratio requires including a secondary, large-scale isotropic Gaussian filter. These results suggest that the synergy effect in texture segregation roots in the computation of a low-level summary statistic, net contrast energy (ΔCE_m), and does not require higher-order integration of separately encoded first- or second-order feature information.

Synergy for bandwidth defined textures

Studies on the synergy effect in visual segmentation have so far used stimuli with "feature contrast" in different dimensions (Nothdurft, 2000). Studies reporting synergy of color and form (Kubovy et al., 1999), color and orientation (Saarela & Landy, 2012), and the synergy effect for spatial frequency and orientation (see Introduction) all used target and reference textures that had different mean values in two feature dimensions. In this study, target and reference regions had the same centroids in (f, ϕ)-space, which means that there were no mean differences in spatial frequency or orientation. Instead, ω_f- and ω_ϕ-targets had larger bandwidths, being not so well localized than the reference texture in this space. However, we observed a strong oversummative synergy effect for $\omega_f + \omega_\phi$-targets. This shows that defining target textures by feature contrast in elementary feature dimensions is not necessary for the synergy effect. In an earlier study, Wolfson and Landy (1998) postulated different texture analysis mechanisms for discriminating textures with and without orientation contrast of the abutting areas. Our findings suggest that assuming two distinct texture analysis mechanisms may not be necessary, since the ΔCE_m measure can explain detection of regional texture variation without feature contrast.

Local energy-based models predict the synergy effect

Local energy-based models have widely been used for predicting human texture segregation (Rubenstein & Sagi, 1990; Landy, 2013; Prins & Kingdom, 2006), and their predictive value has also been demonstrated for more complex stimuli like natural scenes (Groen et al., 2013). However, local energy-based models have not yet been applied to predict the synergy effect. Our results show that the ΔCE_m measure is a suitable and accurate predictor of the amount of cue summation. Beyond its capability of accounting quantitatively for the synergy effect in texture segmentation, the local energy-based model offers a biologically plausible and parsimonious explanation for cue combination effects by feedforward processing chains at early visual processing stages. Our modeling results showed that maximum net contrast energy occurred in the same or closely neighbored mechanisms for the single-cue conditions, ω_f and ω_ϕ, and the double-cue condition, $\omega_f + \omega_\phi$ (see Figure 9A, net contrast energy distributions in (f, ϕ)-space, and Figure 9B, sharply declining scree plots). Hence, the model predicts cue combination effects of bandwidth-modulated textures as a result of local energy computation and a subsequent nonlinearity in the *same units*. Thus, as a surprising result, oversummative cue combination effects appear as an inherent property of local energy mechanisms.

The calculations in Part I of the modeling section showed that adding a large-scale isotropic secondary filter (FRF model) yielded practically the same cue summation predictions than the mere FR model (see Figure 11). This indicates that the oversummative cue combination effect roots in the much stronger local energy values for double-cue targets, which are extracted by local filtering and the subsequent nonlinearity. The secondary, large-scale filtering stage is functional in other respects. It enhances figure-ground separation of target and reference texture by suppressing salt-and-pepper noise in the energy maps, makes the processing chain completely nonresponsive to the original properties of the texture elements, and is also functional for subsequent edge detection (Landy & Bergen, 1991). Modeling in a signal detection theory framework (Part II of modeling section) clearly showed the enhancement of figure-ground separation by the secondary large-scale Gauss filter, which lifted the baseline performance compared to the mere FR model remarkable on the signal-to-noise ratio (d) scale. The data also showed that this effect concerned single-cue and double-cue conditions to equal degrees, since the summation ratio q was the same in FR and FRF variety, and also increasing spatial noise did not affect the value of q. Cells that responded to locus and orientation of texture boundary but were insensitive to local texture features were found in V2 (von der Heydt et al., 1984; von der Heydt & Peterhans, 1989). A full FRF model is thus biologically more plausible than the FR model, and the fact that observers' detection performance could only be reproduced by a full FRF model is a strong indication for including the secondary

filtering stage. It cannot be omitted in a reasonable early mechanism model of texture segregation.

The model introduced here is not a comprehensive one—our intention was to demonstrate that a simple feedforward local energy model can explain the synergy effect in texture segregation with spatial frequency and orientation bandwidth-modulated textures, both in terms of net contrast energy and observers' sensitivity. Therefore, its scope is narrowly focused on the prediction of the oversummative double-cue advantage. In a more comprehensive approach, Olzak and Thomas (1999) introduced a five-stage model that includes higher-level mechanisms at later stages, which allow flexible extraction of features for task-driven comparisons of information provided by the earlier levels. In series of experiments, they tested whether discrimination decisions root in direct access to early layers or need transformations by higher-level units. Configural effects in orientation discrimination with compound $f + 3f$ gratings, concluded from performance decline in cue summation conditions, showed that the observers' decisions were not based on a direct access to two independent grating components, but their information was used to evaluate the whole stimulus configuration as conflicting or congruent when the components were tilted in opposite or same directions. These effects could only be modeled by higher-level units and are evidence against a direct-access account. This points to limitations of the direct-access FRF model proposed here. Constructing target stimuli that deviate from the reference texture in two orientations or two well-separated ranges of spatial frequency, both equal in contrast energy but forming different texture shapes, would pose problems for a simple feedforward FRF model with a winner-take-all decision rule.

Local energy from joint coding challenges the "feature module" account of feature integration

Bergen and Adelson (1988) gave the first evidence that human segregation of textures could be captured by size-tuned units, without reference to any feature-like properties of the textures. This "feature blindness" in evaluating regional texture changes is one of the conceptual strengths of the energy computing algorithm. To give explanations for oversummative cue combination effects for the given bandwidth-modulated textures in terms of feature-processing models requires much more theoretical assumptions. A "feature module" approach assumes that observers compare feature content when they discriminate target and reference, for example, by referring to local differences in saliency maps for each visual feature separately (Kubovy & Cohen, 2001; Itti & Koch, 2001; Treisman & Gelade, 1980; Treisman, 1988). To explain cue combination effects, a rule has to be suggested how the differential results from different feature maps, or modules, are combined into a general map that topographically encodes saliency from different features (Itti & Koch, 2001; Koene & Zhaoping, 2007). Several integration rules have been proposed, ranging from probability summation and/or dimensional orthogonality for the case of feature independence to linear summation for the case of response integration in the same mechanism (Kubovy et al., 1999; Machilsen & Wagemans, 2011; Saarela & Landy, 2012). However, cue summation effects that are larger than the sum of sensitivities to the two single cues, as found here, raise problems for the feature module account, since this requires one to assume special conjunctive mechanisms tuned to both features, but not passive summation rules (Koene & Zhaoping, 2007; Zhaoping & Zhe, 2012). As argued by these authors, early V1 units specifically tuned to combinations of features can explain strong saliency summation effects of double-cue targets, while they make an early feature channeling – later integration architecture superfluous (Koene and Zhaoping, 2007, p. 10). If there are units optimally tuned to feature combinations, decisions about target presence can simply be mediated by a **max**-rule (winner-take-all rule) among all early coding mechanisms on the same layer, and no further response integration rules are necessary. Likewise, the local energy mechanisms modeled here are better tuned to $\omega_f + \omega_\phi$-targets than to the single-cue targets, so applying the **max**-rule in (f, ϕ)-space is all that is necessary to explain the oversummative cue combination effect.

Hypothetical site

The strong local energy increase caused by combined cues points to improved segregation through enhancement of local contrast between figure and background elements as the major effect of cue combination. Since local energy computation on different spatial scales with non-Fourier second-order squaring and pooling could be implemented by feedforward chains from V1 to V2 (Daugman, 1980; Lin & Wilson, 1996; von der Heydt et al., 1984; see Introduction), it is straightforward to presume these early sites behind the synergy effect in texture segmentation for orientation and spatial frequency bandwidth-modulated targets. Indeed, area V2 seems to be anatomically and functionally ideal for local segregation processes (Shipp & Zeki, 2002a, 2002b). An early site is also suggested by electrophysiological results of Bach et al. (2000), who found that the "tsVEP," an early (100–300 ms) electrophysiological correlate of preattententive and effortless texture segregation, showed stronger activation by combined spatial frequency and orientation cues, albeit the cue

combination effect was not impressive in the tsVEP amplitude.

However, nature of stimuli and psychophysical tasks imply that local segregation of texture regions is only an initial processing step to more global processes of shape and object coding, whereby cue combination might be functionally involved. In Gabor random fields, strong cue combination effects were observed only when the stimulus elements together formed simple shapes (rectangles, squares, or lozenges) but not for randomly positioned feature contrast targets that could not be spatially grouped (Meinhardt et al., 2004; Persike & Meinhardt, 2006, 2006). Further, the cue combination effect is generally strong for barely detectable targets and declines rapidly as target detectability increases (Meinhardt & Persike, 2003; Meinhardt et al., 2004). These results indicate that cue combination could be a mechanism that augments figure-ground segregation, serving to enable and to stabilize object identification in cluttered images and noisy surrounds (Persike & Meinhardt, 2006; Straube et al., 2010). Such an enhancement of figure-ground segregation could potentially be mediated by early sites (Lee et al., 1998). However, results from psychophysics and neuroimaging indicate that, despite a strong coupling, local target detection and form identification are distinct processes that differentially activate cortical areas (Straube & Fahle, 2011). Using Gabor texture figures with adjusted feature contrast to equate sensory performance across tasks, the authors found that figure detection and identification equally activated V1 and V2, but the identification task led to much stronger activation in the lateral occipital complex and posterior fusiform gyrus, which are known to be object-selective areas with relatively feature-invariant shape coding (Grill-Spector et al., 1998, 2000).

Because enhancement of local contrast between figure and background elements improves both target detection as well as identification of texture figure shape, the cue combination effect of spatial frequency and orientation-defined textures should have correlates on early retinotopic sites, as well as on higher object-related areas. Disappointingly, current electrophysiological and neuroimaging studies could yet not clearly localize the cue combination effect in texture figure perception. Single-unit recordings from V1 showed no further enhancement of firing rates when texture disks were redundantly defined by feature contrast in several dimensions (Zipser et al., 1996). However, authors used high feature contrasts for the single-feature disks, which does not take into account the nonlinear gain characteristic of cue summation effects (Meinhardt & Persike, 2003; Meinhardt et al., 2004). The same applies to a study on feature summation effects in V1 and V2 neurons (Kastner & Pigarev, 1999). Also, Bach et al. (2000) used quite high feature contrast levels for spatial frequency and orientation-defined texture checkerboards but observed an at least modest cue summation effect at mid-occipital electrodes, which points to generators in V1 and V2. The only electrophysiological studies that used small feature contrast levels for the single-cue targets were contributed by Straube et al. (2010) and Kida et al. (2011). Recording EEG from 26 standard electrode positions, Straube et al. (2010) found that single- and double-cue conditions elicited a negative amplitude shift, influencing mainly the peak amplitude of the posterior P2 component at about 200 ms, which all signaled target presence, but unspecific for the cue conditions. Psychophysical data showed very strong cue summation effects at two different feature contrast levels. Kida et al. (2011) recorded from 62 locations and found a long-lasting negative deflecting starting at 130 ms, which was specific for double-cue targets, at electrodes around the inferior temporal region. At central occipital electrodes, no double-cue specific potentials were found. These results indicated a cue summation effect in object-related areas of the ventral visual stream but not in striate cortex and V2. Control experiments at different feature contrast levels showed that the long-lasting enhancement at inferior temporal electrodes was not directly related to discriminability but appeared as a specific double-cue marker. Hence, current attempts to localize cue combination effects just revealed correlates of enhanced shape and figure perception in higher object-related areas, while there is currently no evidence for correlates in retinotopic areas, which have been shown to identify texture borders (Zipser et al., 1996; Kastner & Pigarev, 1999; Zhaoping, 2003), discriminate figure and ground (Lee et al., 1998), and encode object ownership (Zhou et al., 2000).

Keywords: cue combination, spatial frequency bandwidth, orientation bandwidth, texture segregation, local energy

Acknowledgments

Commercial relationships: none.
Corresponding author: Cordula Hunt.
Email: chunt@uni-mainz.de.
Address: Department of Psychology, Johannes Gutenberg University, Mainz, Germany.

Footnotes

[1]We calculated amplitude spectra values for a dense sampling of spatial frequency and orientation components instead of the usual $|F(u, v)|$ amplitudes to achieve a representation that matches the dimensions of stimulus parameter manipulation.
[2]While the detection task has the standard 2AFC format, the discrimination task does not. Discrimination performance does not rest on stimulus mappings from two stimulus frames, but just from one, so it is a yes-no task by structure. Hence, we treated the detection task as 2AFC in the signal detection framework (MacMillan & Creelman, 2009, Chapter 7), but handled d' calculation for the discrimination task with the

standard formulae derived for the single-frame yes-no task (MacMillan & Creelman, 2009, Chapter 1).

[3]Applying the p-norm in multidimensional scaling, Micko and Fischer (1970) showed that increasing the exponent p more and more narrows the focus on a specific direction in perceptual space (see Micko & Fischer, 1970, Figure 1). Mathematically, the maximum of a set is obtained by letting $p \to \infty$. Define $\mathrm{l}_p(x) = \left(\sum_i |x_i|^p\right)^{1/p}$. Let $X \subset \mathbb{R}^+$ and $x_j = \max\{X\}$. Since $\lim_{p\to\infty} \left(\frac{|x_i|}{|x_j|}\right)^p = 0$ for $i \neq j$, it follows $\lim_{p\to\infty} \mathrm{l}_p\left(\frac{x}{x_j}\right) = 1$, that is, $\lim_{p\to\infty} \mathrm{l}_p(x) = x_j$.

[4]To achieve this, the spatial frequency bandwidth of the Gabor filter array was modulated accordingly. See Appendix B.

[5]Energy distributions could be indexed to distinguish the FR and the full FRF model (see Appendix B). We resign from doing so here and consider only the full FRF model, since it turned out in the simulations that an FR model could not adequately describe the detection data in the context of noise.

References

Arsenault, A. S., Wilkinson, F., & Kingdom, F. A. (1999). Modulation frequency and orientation tuning of second-order texture mechanisms. *Journal of the Optical Society of America A, 16*(3), 427–435, doi:10.1364/JOSAA.16.000427.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93*(2), 154–179, doi:10.1037/0033-295X.93.2.154.

Bach, M., Schmitt, C., Quenzer, T., Meigen, T., & Fahle, M. (2000). Summation of texture segregation across orientation and spatial frequency: electrophysiological and psychophysical findings. *Vision Research, 40*(26), 3559–3566, doi:10.1016/S0042-6989(00)00195-4.

Baddeley, R. (1997). The correlational structure of natural images and the calibration of spatial representations. *Cognitive Science, 21*(3), 351–372, doi:10.1207/s15516709cog2103_4.

Barton, K. (2013). MuMIn: Multi-model inference [R package]. Vienna, Austria: Comprehensive R Archive Network (CRAN).

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48, doi:10.18637/jss.v067.i01.

Bergen, J. R., & Adelson, E. H. (1988). Early vision and texture perception. *Nature, 333*(26), 363–364.

Caelli, T. (1982). On discriminating visual textures and images. *Perception & Psychophysics, 31*(2), 149–159, doi:10.3758/BF03206215.

Campbell, F. W., Cooper, G. F., & Enroth-Cugell, C. (1969). The spatial selectivity of the visual cells of the cat. *Journal of Physiology, 203*(1), 223–235, doi:10.1113/jphysiol.1969.sp008861.

Crozier, W. J. (1936). On the sensory discrimination of intensities. *Proceedings of the National Academy of Sciences of the United States of America, 22*, 412–416.

Daugman, J. (1980). Two dimensional spectral analysis of cortical receptive field profiles. *Vision Research, 20*, 847–856.

Daugman, J. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A, 2*(7), 1160–1169, doi:10.1364/JOSAA.2.001160.

Du Buf, J. M. H. (1993). Responses of simple cells: Events, interferences and ambiguities. *Biological Cybernetics, 68*, 321–333, doi:10.1007/BF00201857.

Ellemberg, D., Allen, H. A., & Hess, R. F. (2006). Second-order spatial frequency and orientation channels in human vision. *Vision Research, 46*(17), 2798–2803, doi:10.1016/j.visres.2006.01.028.

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision, 7*(10), 1–29, doi:10.1167/7.1.10.

Graham, N. (1989). *Visual pattern analysers*. Oxford, UK: Oxford University Press.

Green, D. M., & Swets, J. A. (1988). *Signal detection theory and psychophysics*. New York, NY: John Wiley. (Original work published 1966).

Greene, M. R., & Oliva, A. (2009a). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology, 58*(2), 137–176, doi:10.1016/j.cogpsych.2008.06.001.

Greene, M. R., & Oliva, A. (2009b). The briefest of glances: The time course of natural scene understanding. *Psychological Science, 20*(4), 464–472, doi:10.1111/j.1467-9280.2009.02316.x.

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science, 16*(2), 152–160, doi:10.1111/j.0956-7976.2005.00796.x.

Grill-Spector, K., Kushnir, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron, 21*(1), 191–202, doi:10.1016/S0896-6273(00)80526-7.

Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience, 3*(8), 837–843, doi:10.1038/77754.

Groen, I. I. A., Ghebreab, S., Prins, H., Lamme, V. A. F., & Scholte, H. S. (2013). From image statistics to scene gist: Evoked neural activity reveals transition from low-level natural image structure to scene category. *The*

Journal of Neuroscience, 33(48), 18814–18824, doi:10.1523/JNEUROSCI.3128-13.2013.

Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Research, 45*(13), 1707–1724, doi:10.1016/j.visres.2004.12.021.

Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology, 28*, 229–289, doi:10.1152/jn.1965.28.2.229.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology, 195*, 215–243.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2*(3), 194–203, doi:10.1038/35058500.

Julesz, B., Gilbert, E., & Shepp, L. (1973). Inability of humans to discriminate between visual textures that agree in second-order statistics—revisited. *Perception, 2*(4), 391–405.

Julesz, B., Gilbert, E., & Victor, J. D. (1978). Visual discrimination of textures with identical third-order statistics. *Biological Cybernetics, 31*, 137–140.

Kastner, N. H. C., & Pigarev, S. I. (1999). Neuronal responses to orientation and motion contrast in cat striate cortex. *Visual Neuroscience, 16*(3), 587–600, doi:10.1017/s095252389916317x.

Kida, T., Tanaka, E., Takeshima, Y., & Kakigi, R. (2011). Neural representation of feature synergy. *NeuroImage, 55*(2), 669–680, doi:10.1016/j.neuroimage.2010.11.054.

Kingdom, F. A. A., Baldwin, A. S., & Schmidtmann, G. (2015). Modeling probability and additive summation for detection across multiple mechanisms under the assumptions of signal detection theory. *Journal of Vision, 15*(5), 1, doi:10.1167/15.5.1.

Koene, A. R., & Zhaoping, L. (2007). Feature-specific interactions in salience from combined feature contrasts: Evidence for a bottom-up salience map in V1. *Journal of Vision, 7*, 1–14, doi:10.1167/7.7.6.

Kubovy, M., & Cohen, D. J. (2001). What boundaries tell us about binding. *Trends in Cognitive Sciences, 5*(3), 93–95, doi:10.1016/S1364-6613(00)01604-1.

Kubovy, M., Cohen, D. J., & Hollier, J. (1999). Feature integration that routinely occurs without focal attention. *Psychonomic Bulletin & Review, 6*(2), 183–203, doi:10.3758/BF03212326.

Landy, M. S. (2013). Texture analysis and perception. In: J. S. Werner, & L. M. Chalupa (Eds.), *The new visual neurosciences* (pp. 639–652). Cambridge, MA: MIT Press.

Landy, M. S., & Bergen, J. R. (1991). Texture segregation and orientation gradient. *Vision Research, 31*(4), 679–691, doi:10.1016/0042-6989(91)90009-T.

Landy, M. S., & Oruc, I. (2002). Properties of second-order spatial frequency channels. *Vision Research, 42*, 2311–2329.

Lee, T. S., Mumford, D., Romero, R., & Lamme, V. (1998). The role of the primary visual cortex in higher level vision. *Vision Research, 38*(15–16), 2429–2454, doi:10.1016/s0042-6989(97)00464-1.

Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America, 70*(12), 1458–1471, doi:10.1364/JOSA.70.001458.

Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association, 62*(318), 399, doi:10.2307/2283970.

Lin, L.-M., & Wilson, H. R. (1996). Fourier and non-Fourier pattern discrimination compared. *Vision Research, 36*(13), 1907–1918, doi:10.1016/0042-6989(95)00260-X.

Loschky, L. C., & Larson, A. M. (2008). Localized information is necessary for scene categorization, including the natural/man-made distinction. *Journal of Vision, 8*(1), 4, doi:10.1167/8.1.4.

Loschky, L. C., Sethi, A., Simons, D. J., Pydimarri, T. N., Ochs, D., & Corbeille, J. L. (2007). The importance of information localization in scene gist recognition. *Journal of Experimental Psychology: Human Perception and Performance, 33*(6), 1431–1450, doi:10.1037/0096-1523.33.6.1431.

Machilsen, B., & Wagemans, J. (2011). Integration of contour and surface information in shape detection. *Vision Research, 51*(1), 179–186, doi:10.1016/j.visres.2010.11.005.

MacMillan, N. A., & Creelman, C. D. (2009). *Detection theory: A user's guide* (2nd ed.). New York, NY: Psychology Press.

Malik, J., & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A, 7*(5), 923, doi:10.1364/JOSAA.7.000923.

McCulloch, C. E., & Searle, S. R. (2001). *Generalized, linear, and mixed models*. New York, NY: Wiley.

Meinhardt, G., & Persike, M. (2003). Strength of feature contrast mediates interaction among feature domains. *Spatial Vision, 16*(5), 459–478.

Meinhardt, G., Persike, M., Mesenholl, B., & Hagemann, C. (2006). Cue combination in a combined feature contrast detection and figure

identification task. *Vision Research, 46*(23), 3977–3993, doi:10.1016/j.visres.2006.07.009.

Meinhardt, G., Schmidt, M., Persike, M., & Röers, B. (2004). Feature synergy depends on feature contrast and objecthood. *Vision Research, 44*(16), 1843–1850, doi:10.1016/j.visres.2004.04.002.

Micko, H. C., & Fischer, W. (1970). The metric of multidimensional psychological spaces as a function of the differential attention to subjective attributes. *Journal of Mathematical Psychology, 7*(1), 118–143, doi:10.1016/0022-2496(70)90061-1.

Mortensen, U., & Suhl, U. (1991). An evaluation of sensory noise in the human visual system. *Biological Cybernetics, 66*, 37–47.

Motoyoshi, I., & Nishida, S. (2001). Visual response saturation to orientation contrast in the perception of texture boundary. *Journal of the Optical Society of America A, 18*(9), 2209–2219, doi:10.1364/JOSAA.18.002209.

Nothdurft, H.-C. (2000). Salience from feature contrast: additivity across dimensions. *Vision Research, 40*(10–12), 1183–1201, doi:10.1016/S0042-6989(00)00031-6.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42*(3), 145–175, doi:10.1023/A:1011139631724.

Olzak, L. A., & Thomas, J. P. (1999). Neural recoding in human pattern vision: Model and mechanisms. *Vision Research, 39*(2), 231–256, https://doi.org/10.1016/S0042-6989(98)00122-9.

Peirce, J. W. (2008). Generating stimuli for neuroscience using psychopy. *Frontiers in Neuroinformatics, 2*, 10, doi:10.3389/neuro.11.010.2008.

Persike, M., & Meinhardt, G. (2006). Synergy of features enables detection of texture defined figures. *Spatial Vision, 19*(1), 77–102, doi:10.1163/156856806775009214.

Persike, M., & Meinhardt, G. (2008). Cue summation enables perceptual grouping. *Journal of Experimental Psychology: Human Perception and Performance, 34*(1), 1–26, doi:10.1037/0096-1523.34.1.1.

Persike, M., & Meinhardt, G. (2015). Cue combination anisotropies in contour integration: The role of lower spatial frequencies. *Journal of Vision, 15*(5), 17, doi:10.1167/15.5.17.

Prins, N., & Kingdom, F. A. A. (2006). Direct evidence for the existence of energy-based texture mechanisms. *Perception, 35*(8), 1035–1046, doi:10.1068/p5546.

Prins, N., & Kingdom, F. A. A. (2018). Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes toolbox. *Frontiers in Psychology, 9*, 1250, doi:10.3389/fpsyg.2018.01250.

R Core Team. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Regan, D., & Beverley, K. I. (1983). Spatial-frequency discrimination and detection: Comparison of postadaptation thresholds. *Journal of the Optical Society of America A, 73*(12), 1684–1690, doi:10.1364/JOSA.73.001684.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*(11), 1019–1025, doi:10.1038/14819.

Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Reserach, 21*, 409–418.

Roufs, J. A. J. (1974). Dynamic properties of vision: VI. Stochastic threshold fluctuations and their effect on flash-to-flicker ratio. *Vision Research, 14*, 871–888, doi:10.1016/0042-6989(74)90150-3.

Rousselet, G., Joubert, O., & Fabre-Thorpe, M. (2005). How long to get to the gist of real-world natural scenes? *Visual Cognition, 12*(6), 852–877, doi:10.1080/13506280444000553.

Rubenstein, B. S., & Sagi, D. (1990). Spatial variability as a limiting factor in texture-discrimination tasks: Implications for performance asymmetries. *Journal of the Optical Society of America A, 7*(9), 1632, doi:10.1364/JOSAA.7.001632.

Saarela, T. P., & Landy, M. S. (2012). Combination of texture and color cues in visual segmentation. *Vision Research, 58*, 59–67, doi:10.1016/j.visres.2012.01.019.

Sachs, M. B., Nachmias, J., & Robson, J. G. (1971). Spatial-frequency channels in human vision. *Journal of the Optical Society of America, 61*(9), 1176–1186, doi:10.1364/JOSA.61.001176.

Sagi, D. (1995). The psychophysics of texture segmentation. In: T. V. Papathomas, C. Chubb, A. Gorea, & E. Kowler (Eds.), *Early vision and beyond* (pp. 69–78). Cambridge, MA: MIT Press.

Scholte, H. S., Ghebreab, S., Waldorp, L., Smeulders, A. W. M., & Lamme, V. A. F. (2009). Brain responses strongly correlate with Weibull image statistics when processing natural images. *Journal of Vision, 9*(4), 29.1–29.15, doi:10.1167/9.4.29.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological*

Science, 5(4), 195–200, 10.1111/j.1467-9280.1994.tb00500.x.

Shipp, S., & Zeki, S. (2002a). The functional organization of area V2, II: The impact of stripes on visual topography. *Visual Neuroscience, 19*(2), 211–231, doi:10.1017/s0952523802191176.

Shipp, S., & Zeki, S. (2002b). The functional organization of area V2, I: Specialization across stripes and layers. *Visual Neuroscience, 19*(2), 187–210, doi:10.1017/s0952523802191164.

Straube, S., & Fahle, M. (2011). Visual detection and identification are not the same: Evidence from psychophysics and fMRI. *Brain and Cognition, 75*(1), 29–38, doi:10.1016/j.bandc.2010.10.004.

Straube, S., Grimsen, C., & Fahle, M. (2010). Electrophysiological correlates of figure ground segregation directly reflect perceptual saliency. *Vision Research, 50*(5), 509–521, doi:10.1016/j.visres.2009.12.013.

Tanner, W. P. (1956). Theory of recognition. *Journal of the Acoustical Society of America, 28*, 882–888, https://doi.org/10.1121/1.1908504.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520–522, doi:10.1038/381520a0.

Tolhurst, M. J. A., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research, 23*, 775–785, doi:10.1016/0042-6989(83)90200-6.

Treisman, A. (1988). Features and objects: The Fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology Section A, 40*(2), 201–237, doi:10.1080/02724988843000104.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97–136, doi:10.1016/0010-0285(80)90005-5.

VanRullen, R. (2006). On second glance: Still no high-level pop-out effect for faces. *Vision Research, 46*(18), 3017–3027, doi:10.1016/j.visres.2005.07.009.

Victor, J. D., Conte, M. M., Purpuran, K., & Katz, E. (1995). Isodipole textures: a window on cortical mechanisms of form processing. In: T. V. Papathomas, C. Chubb, A. Gorea, & E. Kowler (Eds.), *Early vision and beyond* (pp. 99–107). Cambridge, MA: MIT Press.

von der Heydt, R., & Peterhans, E. (1989). Mechanisms of contour perception in monkey visual cortex: Lines of pattern discontinuity. *Journal of Neuroscience, 9*(5), 1731–1748, doi:10.1523/JNEUROSCI.09-05-01731.1989.

von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science, 224*, 1260–1262, doi:10.1126/science.6539501.

Wallis, S. A., Baker, D. H., Meese, T. S., & Georgeson, M. A. (2013). The slope of the psychometric function and non-stationarity of thresholds in spatiotemporal contrast vision. *Vision Research, 76*, 1–10, https://doi.org/10.1016/j.visres.2012.09.019.

Watson, A. B. (1982). Summation of grating patches indicates many types of detector at one retinal location. *Vision Research, 22*(1), 17–25, https://doi.org/10.1016/0042-6989(82)90162-6.

Wichmann, F. A., Braun, D. I., & Gegenfurtner, K. R. (2006). Phase noise and the classification of natural images. *Vision Research, 46*(8), 1520–1529, doi:10.1016/j.visres.2005.11.008.

Wilson, H. R., & Regan, D. J. (1984). Spatial frequency adaptation and grating discrimination: Predictions of a line-element model. *Journal of the Optical Society of America, 1*, 1091–1096, doi:10.1364/josaa.1.001091.

Wolfson, S. S., & Landy, M. S. (1998). Examining edge- and region-based texture analysis mechanisms. *Vision Research, 38*(3), 439–446, doi:10.1016/S0042-6989(97)00153-3.

Zhaoping, L. (2003). V1 mechanisms and some figure-ground and border effects. *Journal of Physiology, Paris, 97*(4–6), 503–515, doi:10.1016/j.jphysparis.2004.01.008.

Zhaoping, L., & Zhe, L. (2012). Properties of V1 neurons tuned to conjunctions of visual features: Application of the V1 salience hypothesis to visual search behavior. *PLosOne, 7*(6), e36223, doi:10.1371/journal.pone.0036223.

Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *Journal of Neuroscience, 20*(17), 6594–6611, doi:10.1523/JNEUROSCI.20-17-06594.2000.

Zipser, K., Lamme, V. A. F., & Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *Journal of Neuroscience, 16*(22), 7376–7389, doi:10.1523/JNEUROSCI.16-22-07376.1996.

Appendix A

Conditioning discrimination on detection

Behavioral results showed lower baseline sensitivity for orientation discrimination than target detection. Conditioning discrimination on previous correct detection has only little effect on the different feature sensitivities. It slightly raises sensitivity to spatial frequency bandwidth, thus stabilizing the baseline

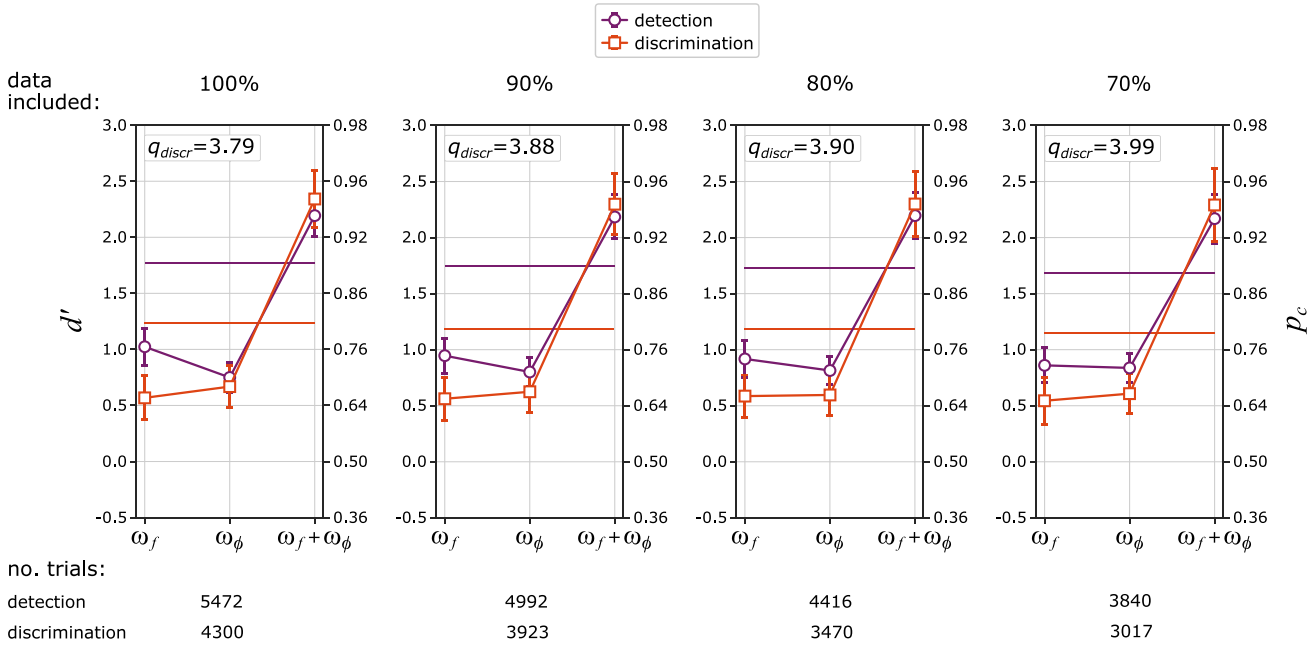

Figure 14. Display analogous to Figure 7 but discrimination data were only included if the previous detection was successful. This reduces data on the level of individual trials; hence, included trials are also indicated (maximum number of trials is participants × number of repetitions × trials in main experiment, that is, 19 × 3 × 96 = 5, 472). As a measure of cue summation, q_{discr} is given for detection q is identical to the values shown in Table 1. For better overview, only solid lines indicating the most conservative prediction of linear summation are included.

at slightly above $d' = 0.5$, and marginally reduces sensitivity for the feature conjunction. This leads to a reduction of the q-ratio, but its value nevertheless remains markedly larger for discrimination than for detection (compare Figure 14).

Appendix B

Specific modeling of the Filter-Rectify-Filter process

We defined the primary filtering stage by a bank of even and odd Gabor filters, each described by

$$h_{s;f,\theta}(x_0, y_0) = \exp\left(-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2(f)}\right) \times \begin{cases} \cos\left(2\pi f((x-x_0)\cos\theta - (y-y_0)\sin\theta)\right), & s = \text{even} \\ \sin\left(2\pi f((x-x_0)\cos\theta - (y-y_0)\sin\theta)\right), & s = \text{odd} \end{cases} \quad (13)$$

In Equation 13, f is the filter spatial frequency; θ the orientation in radiants, that is, $\theta = \frac{\phi}{180}\pi$, with ϕ the rotation angle in degrees, and x_0, y_0 the spatial filter coordinates. We implemented a flexible scaling scheme

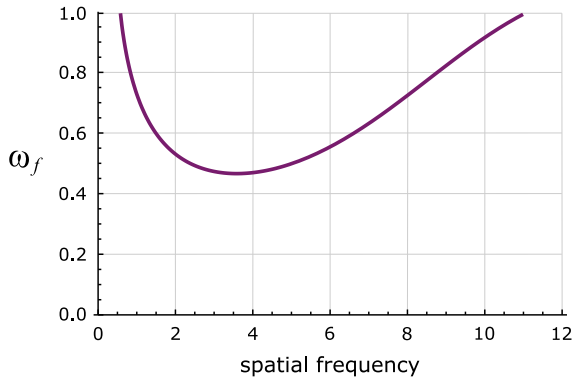

Figure 15. Half-amplitude spatial frequency bandwidth, ω_f, measured in octaves, of the quadrature filter pairs as a function of carrier spatial frequency f across the spectrum of frequencies used in the model.

to keep filter bandwidth narrow ($\leqslant 0.5$ octaves) and thereby frequency information precisely localized for medium frequencies between 2.5 and 5 cpd. For extreme frequencies at both ends of our tested frequency range, filter bandwidth approached one octave; thus, frequency information was less certain for lower and higher frequencies (compare Table 3 and Figure 15). Channels are typically thought to be broader for lower spatial frequencies (Sachs et al., 1971), but it is less clear

if their bandwidths increase again for higher spatial frequencies. A study by Ellemberg et al. (2006) might indicate as much for second-order channels, but a dipped relationship between spatial frequency channel and its bandwidth or a monotonically decreasing one is biologically plausible. We decided on the former scheme because it worked slightly better within our architecture, and it also predicted the same slight instability of the baseline as in the behavioral results. The scheme was implemented by applying a normal scaling principle (Du Buf, 1993)

$$\sigma(f) = \frac{a}{f}, \quad (14)$$

while controlling the parameter a via a Weibull function

$$y(u) = 0.2 \cdot 1.75 \cdot (0.2 \cdot (u - 0.5))^{(1.75-1)} \cdot e^{-(0.2\cdot(u-0.5))^{1.75}} \quad (15)$$

$$a(u) = 0.67 \cdot \frac{y(u)}{\max(y(u))} + 0.5. \quad (16)$$

The spacing along the spatial frequency continuum was such that each new filter was centered at the half-amplitude frequency of the previous filter. Since spatial filter bandwidths were narrow in the range from about one octave below and one octave above the texture carrier frequency of 3.5 cpd, this maintained dense spacing in this range, allowing us to localize the maximally responding local energy mechanisms with good precision. Table 3 lists the center spatial frequencies and corresponding wavelengths. Orientations were sampled from the [0,180[interval in 3.5° steps.

In the modeling sequence, a stimulus was first convolved with each of the even and odd filters tuned to a specific pair of spatial frequency f and orientation ϕ. The stimulus was coded as a luminance distribution $L(x, y)$, and thus the first stage filter output was

$$u_{s;f,\phi}(x, y) = L(x, y) * h_{s;f,\phi}(x, y). \quad (17)$$

In the next stage, the outputs of the even and odd filter were rectified to one response by a local energy computation

$$L_{f,\phi}(x, y) = u^2_{\text{even};f,\phi}(x, y) + u^2_{\text{odd};f,\phi}(x, y) \quad (18)$$

followed by a logarithmic, compressive nonlinear transducer

$$E_{f,\phi}(x, y) = \left(\log\left[\sqrt{L_{f,\phi}(x, y)} + 1\right]\right)^{0.5}. \quad (19)$$

The typical nonlinear transducer in local energy models is a simple logarithmic. Here, we implemented a slightly stronger transducer because it proved useful to put further constraint on spurious responses in nonoptimal channels of our model architecture. The energy distributions $E_{f,\phi}(x, y)$ are the result of the FR stages of the model. In the full FRF model, these energy distributions are convolved with a second-stage, large-scale isotropic Gaussian filter. We used an isotropic Gaussian $G(x, y)$ with a standard deviation of $\sigma = 0.713°$. Hence, the final local energy distributions were obtained by

$$\hat{E}_{f,\phi}(x, y) = E_{f,\phi}(x, y) * G(x, y). \quad (20)$$

To model texture segmentation, contrast energy was computed by taking the mean μ of each local energy distribution (Groen et al., 2013; Scholte et al., 2009), which was compared for target and reference and integrated across the parameter space using a winner-take-all rule (see modeling section and Figure 8).

Appendix C

The variation coefficient for a Weibull approximation to Normal distributions

If x is a $N(\mu, \sigma)$ distributed random variable with distribution function

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} \exp\left(-\frac{1}{2}\left(\frac{v - \mu}{\sigma}\right)^2\right) dv, \quad (21)$$

the slope of F at μ is proportional to $(\sigma)^{-1}$

$$\delta = \left.\frac{dF}{dx}\right|_{\mu} = \left(\sqrt{2\pi}\sigma\right)^{-1}, \quad (22)$$

that is, σ is

$$\sigma = \left(\delta\sqrt{2\pi}\right)^{-1}. \quad (23)$$

The ratio of σ to μ

$$c_v = \frac{\sigma}{\mu} = \frac{\delta^{-1}}{\mu}\left(\sqrt{2\pi}\right)^{-1} \quad (24)$$

is called the *variation coefficient*. Since σ is inversely proportional to the slope at μ, the variation coefficient reflects the spread of the density $f(x)$ relative to the location on the scale.

It is possible to define a *generalized variation coefficient* (Mortensen & Suhl, 1991) by evaluating the

ratio of the reciprocal of slope in a particular quantile x_0 relative to this quantile of a distribution function. The ratio

$$q_0 = \frac{\delta^{-1}}{x_0} = (\delta x_0)^{-1} \quad (25)$$

with $x_0 = F^{-1}(p_0)$ and $\delta = dF/dx|_{x_0}$ is proportional to the variation coefficient for a normal distribution with $p_0 = 0.5$ (see Equation 24) and can be used with other distribution functions, as the Weibull. For a Weibull distribution

$$F(x) = 1 - \exp\left(-\alpha x^{\beta}\right) \quad (26)$$

with density

$$f(x) = \frac{dF}{dx} = \alpha\beta \exp\left(-\alpha x^{\beta}\right) x^{\beta-1}, \quad (27)$$

the generalized variation coefficient takes the form

$$q_0 = (\delta x_0)^{-1} = \frac{e}{\beta} \quad (28)$$

for $p_0 = 1 - 1/e$ (≈ 0.632). Equations 28 and 32 show that the variation coefficient taken from Weibull functions is independent of the scale parameter; just the slope parameter β enters.

For the quantile corresponding to $p_0 = 0.5$, one obtains from (26)

$$x_0 = \ln(2)^{1/\beta}\alpha^{-1/\beta} \quad (29)$$

and the value of $f(x)$ in x_0 is

$$\begin{aligned}\delta &= \alpha\beta \exp\left(-\alpha\left(\ln(2)^{1/\beta}\alpha^{-1/\beta}\right)^{\beta}\right) x_0^{\beta} x_0^{-1} \\ &= \alpha\beta e^{-\ln(2)} \ln(2)\alpha^{-1} x_0^{-1} \quad (30) \\ &= \beta e^{-\ln(2)} \ln(2) x_0^{-1} \quad (31)\end{aligned}$$

and $q_0 = (\delta x_0)^{-1}$ becomes

$$q_0 = \frac{e^{\ln(2)}}{\beta \ln(2)}. \quad (32)$$

Finally, in view of (24),

$$c_v = \frac{q_0}{\sqrt{2\pi}} \quad (33)$$

one obtains the variation coefficient c_v from an approximation with the Weibull distribution if the slope parameter β is known.