Vision Research 235 (2025) 108660

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Feature synergy enhances detection but not recognition of shape from texture cues

Cordula Hunt-Radej *, Anna-Lena Schubert, Günter Meinhardt

Johannes Gutenberg-University, Wallstrasse 3, 55122, Mainz, Germany

ARTICLE INFO

Keywords:
Cue combination
Texture segregation
Shape processing

ABSTRACT

Texture regions that differ from their surroundings in more than one local feature are more easily detected. Recent findings show that a low-level summary statistic, net contrast energy, predicts this double-cue advantage, suggesting early-stage integration during image analysis. We investigated whether this advantage also applies to more complex, texture-defined shape discrimination beyond figure-ground segregation. Using both a figure detection task and a more demanding shape identification task, we calibrated d' sensitivity to fixed baseline levels with single-cue targets defined by orientation or spatial frequency contrast. We then measured performance for double-cue targets at these baselines. Contrary to earlier results reported for simpler shape discriminations, we found a reduced double-cue advantage in the shape identification task. Specifically, double-cue sensitivity was notably lower than the algebraic sum of the single-cue sensitivities, a level achieved consistently in the detection task. Control tests with high feature contrast showed perfect detection performance for both single and combined cues. However, shape identification saturated at levels between 83%–90% accuracy, while gray-shaded figures yielded perfect performance, suggesting that unique shape representations could not be built from single or combined texture cues. These findings suggest that texture cue summation enhances texture segregation and segmentation but does not improve higher-level recognition of 2D texture shapes.

1. Introduction

A long-standing problem in human vision research is how the brain decomposes a visual scene into features and reintegrates them into unambiguous and stable object representations. When objects differ from the surround in several features, they can become more salient than expected from each individual feature, which is known as the "feature synergy" effect (Kubovy & Cohen, 2001). This effect has been shown for target texture regions defined by local changes in orientation and spatial frequency modulation (Kida et al., 2011; Meinhardt & Persike, 2003; Meinhardt et al., 2006, 2004; Straube et al., 2010), but also for color and shape (Kubovy et al., 1999), and color and orientation (Saarela & Landy, 2012). Strong cue summation effects were found not only for the detection of target textures in cluttered images, but also for the discrimination of their shape (Meinhardt et al., 2006; Persike & Meinhardt, 2008; Straube et al., 2010).

The extraction of local features and primary surface characteristics occurs at early retinotopic stages in V1 (Daugman, 1980; Hubel & Wiesel, 1974; Knierim & van Essen, 1992; Lamme et al., 1992; Maffei & Fiorentini, 1973; Zipser et al., 1996), and by V2 many cells already signal border ownership (Williford & von der Heydt, 2016; Zhou et al., 2000). Although V2 shows some sensitivity to surface textures, most V2 neurons remain preferentially tuned to luminance-defined contours, indicating that texture defined form is not fully analyzed before signals leave V2 (El-Shamayleh & Movshon, 2011). In macaque V4, probably corresponding functionally to human V4 and LOC (Pasupathy et al., 2019), subpopulations of neurons respond selectively to object shape or to object texture, while the majority encode both independently (Kim et al., 2019; Lieber et al., 2024). More advanced feature invariant shape and object coding then emerges in higher ventral areas such as LOC and IT (Grill-Spector et al., 1998; Kourtzi & Huberle, 2005; Kourtzi & Kanwisher, 2000). Because simultaneous coding of shape and texture begins in mid level areas like V4 and propagates downstream to build object representations that are tolerant to identity preserving transformations (Rust & DiCarlo, 2010) these mid to high level regions are likely candidates for feature integration.

Previous psychophysical studies suggest that texture feature synergy might augment at least rudimentary object representations, since results showed that synergy vanished when target elements did not form a closed 2D shape (Persike & Meinhardt, 2006), or did not form spatially coincident texture borders which uniquely separate two

* Corresponding author.
E-mail address: chunt@uni-mainz.de (C. Hunt-Radej).

https://doi.org/10.1016/j.visres.2025.108660
Received 20 February 2025; Received in revised form 3 June 2025; Accepted 27 June 2025
Available online 18 July 2025
0042-6989/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

objects (Kubovy & Cohen, 2001; Saarela & Landy, 2012). Rivest and Cabanagh (1996) showed that defining surface borders by feature contrasts in more than one attribute (luminance, color and texture) improved border localization precision significantly compared to a just luminance defined border, suggesting cue integration at a common site to improve object border representation. These results suggest that the integration of multiple spatially coincident texture cues may be bound into the functional chain of 2D shape processing.

Another body of research suggests that cues provided by early visual areas are used in scene and object categorization. Humans rapidly categorize visual scenes (Fei-Fei et al., 2007; Schyns & Oliva, 1994), judging at a glance whether their content is natural or artificial (Greene & Oliva, 2009a, 2009b; Loschky & Larson, 2008; Loschky et al., 2007; Oliva & Torralba, 2001), by using low-level image statistics computed at early stages of image analysis (Groen et al., 2013; Scholte et al., 2009). The extraction of such local energy-based measures has been modeled by filter-rectifier-filter (FRF) mechanisms that mimic cascades of V1 and V2 mechanisms (Caelli, 1995; Landy & Bergen, 1991; Lin & Wilson, 1996; Rubenstein & Sagi, 1990). A recent study by Hunt and Meinhardt (2021) showed that local energy computations in the first steps of visual coding are sufficient to explain the feature synergy effect for target detection in spatial frequency and orientation modulated textures, supporting the relevance of local energy computations for texture segmentation.

However, the modeling results of Hunt and Meinhardt (2021) do not cover the role of cue summation for processing the exact spatial outline of the target texture. Moreover, the double-cue benefits obtained in their target texture detection and shape discrimination task did not support stronger cue combination effects in either task, since the *Feature* × *Task* effect in their linear mixed model remained non-significant. The failure to find different task specific cue summation effects could be due to the fact that the shape discrimination task used by the authors was rather rudimentary, since the two alternative shapes were already discriminated by detecting a left-diagonal versus a right-diagonal texture border, which is close to detecting target presence by a texture segregation cue. This was indicated by only moderately lower sensitivities to single-cue targets in shape discrimination compared to target texture detection, suggesting that the task difficulty was not substantially different.

A first clue that there are different cue summation effects in texture figure detection and discrimination when the shape processing demands of the latter task are considerably larger than required for mere target detection comes from Persike and Meinhardt (2008). The authors used instances of two figure classes, defined in Gabor random fields. The shapes of one figure class were rather similar block shapes (“blocks”), which were barely distinguishable at various feature contrast levels sufficient for detection. However, this was possible with reasonable accuracy for the instances of the other figure class (“lozenges”). Results showed a strong double-cue advantage for both figure classes in detection. For shape discrimination, however, cue synergy was only observed for the easily discriminable shapes (lozenges), while there was only modest cue summation for the shape class with difficult to discriminate instances (blocks). This points towards strong cue summation in the detection of target texture presence and simple shape discriminations, but not for more complex shape discriminations. Unfortunately, the authors calibrated feature contrasts to achieve equal single-cue performance only for the detection task, resulting in significantly lower discrimination performance than detection across all experimental conditions. Thus, their finding of a potentially smaller cue summation benefit as shape discrimination becomes more difficult suffers from this confound.

The present study was designed to compare the double-cue benefit in target texture detection and complex shape discrimination (“identification”), avoiding confounds and ambiguities of former studies. First, to avoid confounds with task performance level, feature contrasts were adjusted to maintain the same single-cue accuracy in both tasks. Second, to warrant valid comparison of cue summation effects, these adjustments were made for several different accuracy levels within the possible range of performance which was not yet constrained by ceiling effects. Third, to effectively disentangle the two tasks, the shapes used in the identification task were designed such that they could not or just hardly be discriminated at the feature contrast levels used in the detection task, as it was the case for the block figures in the Persike and Meinhardt (2008) study. Since calibrating feature contrast for equal single-cue performance in detection and complex shape discrimination means that much larger feature contrasts are used in the shape discrimination task, measuring the double-cue advantage under these conditions is appropriate to test whether adding a second cue could improve shape processing routines that enter after the target area is already clearly distinguishable from the surround.

2. Methods

2.1. Study outline

To enable systematic comparison of cue summation effects in target texture detection and complex shape identification from texture figures, three experiments were executed, each requiring detection (D) and identification (I). In Experiment I_D observers saw a target texture embedded in a reference texture and indicated the quadrant where the target was presented (detection, tested via target localization). In Experiment I_I observers saw the same stimulus objects, but identified which one out of four possible alternatives was presented. Target textures deviated from the reference texture in spatial frequency (f), orientation (ϕ), or both ($f + \phi$). Before the main experiments I_D and I_I were executed, individual calibration measurements were run to determine the feature contrasts necessary to reach $40\%, 50\%$ and 60% accuracy with each single cue, f and ϕ, as three single-cue baseline levels for cue summation measurement in detection and identification. These feature contrasts were used to define double-cue targets with psychophysically equivalent feature components for each task. In the main experiments, single-cue and double-cue trials were randomly intermixed with feature contrasts taken from the three calibration levels.

Experiment II was a replication of Experiment I, but with feature contrasts interchanged across tasks: the detection task was run with the feature contrasts calibrated for identification, and the identification task with the feature contrasts calibrated for detection. Experiment II served as a control to test whether we succeeded to disentangle both tasks such that texture segregation is completed at the higher feature contrasts required to perform shape identification, while shape identification is precluded with the feature contrasts used for detection. In Experiment III we checked whether perfect performance was achieved in either task with maximal feature contrasts. In all experiments detection and identification tasks were executed as separate experimental units.

2.2. Psychophysical tasks

Detection and identification tasks had identical trial structure (see Fig. 1A), but the response screen differed. After fixation and blank intervals, the texture stimulus appeared for 400 ms, followed by static noise masking with a grain resolution of 1 px to erase afterimages, which might affect task performance depending on the feature contrast used in the target figure. Afterwards, the response screen was presented until subjects selected one of the four alternatives with the computer mouse. In the detection task, the stimulus frame contained a target at one of the four quadrant positions and participants responded by clicking on the supposed quadrant (see Fig. 1A1). For the shape identification task, the stimulus frame contained a target at randomized location around screen center and participants indicated its shape by selecting from the four alternatives shown on the response screen (see Fig. 1A2). Trial by trial feedback about correctness was provided by a brief tone signal (correct: “tack”; incorrect: “tack-tack”).

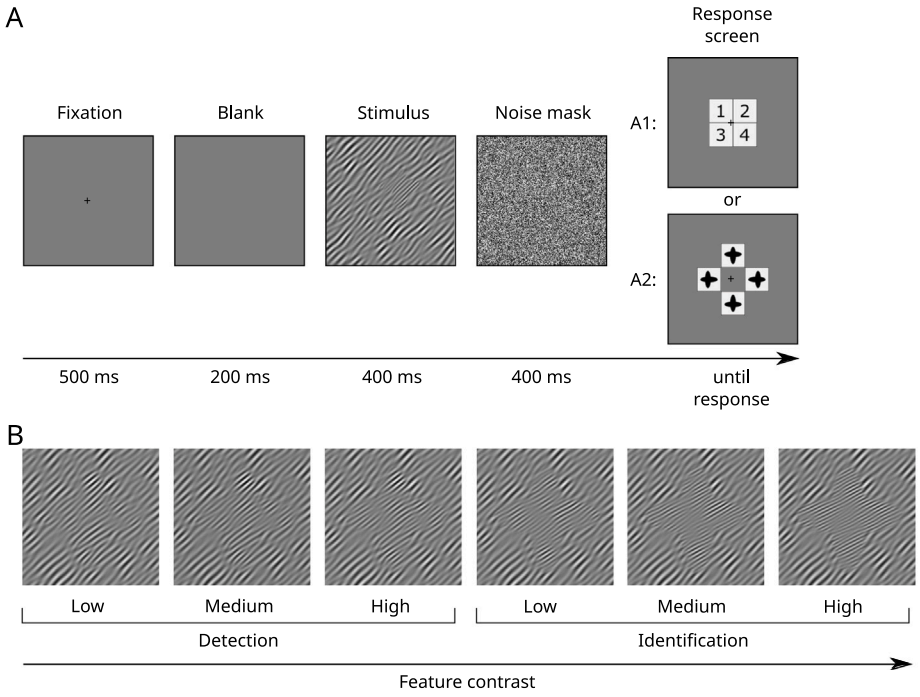

Fig. 1. Trial sequence used in all three experiments (A) and illustration of changing target figure saliency of a double-cue target with rising difference to the surround in spatial frequency and orientation over the levels used in figure detection (left) and shape identification (right) (B). Feature differences were adjusted for illustrative purposes.

2.3. Stimuli

2.3.1. Target texture shapes

The guiding objective for constructing target textures was that a precise representation of 2D figure shape was necessary for their correct identification, avoiding that target shapes were distinguishable just by local form cues. To achieve this, we crossed a horizontal and a vertical ellipse with a slight eccentric displacement upward, and generated three further shapes by clockwise rotation in 90° steps. The four target forms are easily distinguished as line-drawings or shaded figures (see upper row of Fig. 2). However, when their shape has to be extracted from regional texture differences (see lower three rows of Fig. 2), they are hardly discriminable, while target texture presence within the embedding surround is obvious. In order to study texture cue summation we generated textures with different principal spatial frequency or orientation ("single-cue targets"), or both ("double-cue targets") in the target region compared to the surround. The stimulus examples in Fig. 2 show that double-cue targets appear more salient than single-cue targets, but the advantage is not impressively large, as it is obvious on a first glance for detecting the presence of target textures in differently textured surroundings (see Meinhardt et al., 2006, Fig. 1; Persike & Meinhardt, 2008, Fig. 12; Hunt & Meinhardt, 2021, Fig. 1).

2.3.2. Texture generation

Textures were generated by convolving uniformly distributed spatial pixel noise with a Gabor kernel defined by

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

with $x' = x\cos\left(\frac{\phi}{180}\pi\right) - y\sin\left(\frac{\phi}{180}\pi\right)$ and $y' = x\sin\left(\frac{\phi}{180}\pi\right) + y\cos\left(\frac{\phi}{180}\pi\right)$. Spatial parameters σ_x and σ_y were kept constant so that the bounding box of 1.35° × 1.35° (61 × 61 px) contained a same-sized, symmetric Gaussian window for all kernels. Convolutions were computed via fast inverse Fourier transform. Reference textures had a spatial frequency $f = f_R$ of 3.5 cpd and an orientation $\phi = \phi_R$ of either +45° or −45°. Depending on the feature condition, target textures were defined by either increasing the spatial frequency parameter f to f_T, increasing the orientation parameter (i.e., clockwise rotation) ϕ to ϕ_T, or doing both (double-cue condition). The resulting target textures differ from the reference texture in their space-average means, i.e. there is *feature contrast* in spatial frequency, $f = f_T - f_R$, or orientation, $\phi = \phi_T - \phi_R$, or in both parameters, respectively.

To avoid transient texture edges, the target texture was blended into the reference texture using a target form template and a cumulative Gaussian transition function for combining the gray values from target and reference regions. The Gaussian had a standard deviation of σ = 0.177°, corresponding to about 8 px, which means that the transition from reference to target region was practically complete after 32 px.

2.3.3. Spatial parameters

Target shapes had a length of 5.66° (256 px), a width of a third of that, 1.89° (85 px), and were defined by a 0.332° (15 px) shift. Targets were localized within a central square region of 8.5° × 8.5° visual angle (384 × 384 px). For the detection task, targets were placed at the four most eccentric corner positions which still fell within the central square of the stimulus. This entailed a 2.83° × 2.83° visual angle (128 × 128 px) overlap at the point of fixation in the middle of the screen ensuring that participants saw the target with foveal vision and still leaving enough differential local information as to allow a clear statement in which of the four quadrants the target was. For the target shape identification task, target positions were randomized across trials within the central square region. Fig. 3 illustrates the spatial outline of the display and target positioning.

2.4. Participants

The same 26 participants (22 female) served as observers in Experiments I and II. One participant completed only the calibration to the identification task, and suspended participation afterwards. All participants had normal or corrected-to-normal vision. Participants were paid and not informed about specific hypotheses or expectations regarding the experimental tests until after participation. All signed

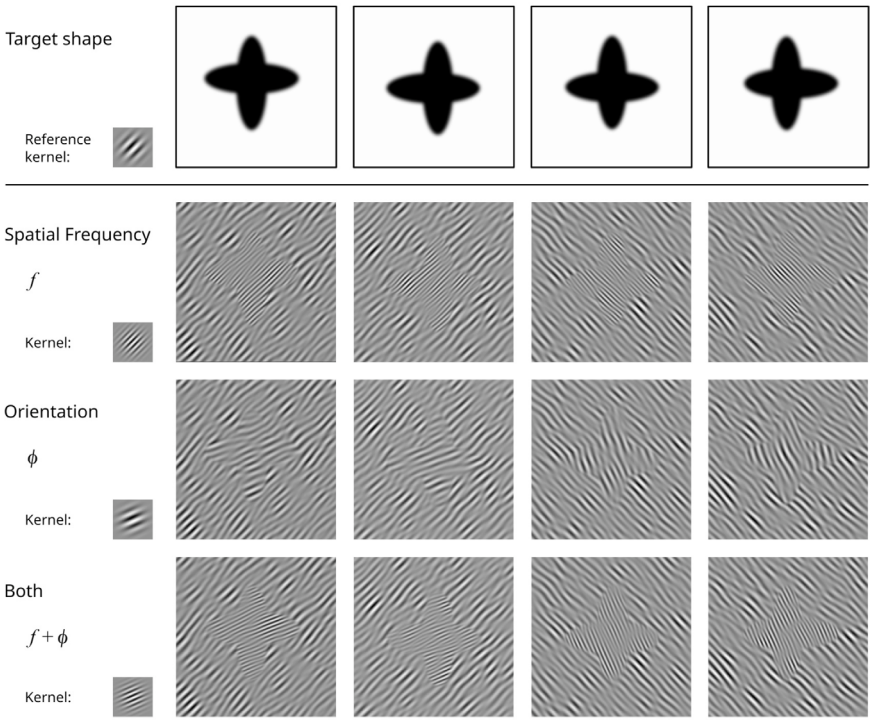

Fig. 2. Texture figures used in detection and shape identification tasks. The first row shows the four target shapes as shaded figures, rows 2 to 4 show the target shapes as texture regions embedded in a surrounding (reference) texture. Textures were generated by convolving spatial pixel noise with a Gabor kernel using specific spatial frequency and orientation parameters for target and reference region. Target textures were defined by changing the kernel's carrier spatial frequency (f, second row), orientation (ϕ, third row), or both ($f+\phi$, last row) compared to the reference kernel (see left column). The spatial transitions of reference and target region were smoothed by a cumulative Gaussian in order to avoid spatial transients at the texture borders. Reference and single feature parameters are adapted in magnitude for illustration purposes.

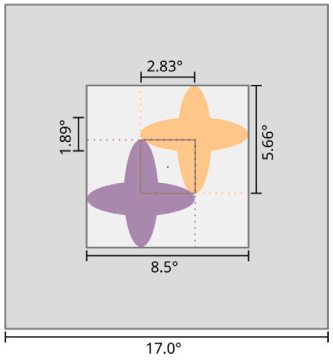

Fig. 3. Spatial parameters of the display and target presentation area. Targets appeared randomly at one of the four quadrant positions of the central 8.5° × 8.5° stimulus region in the detection task, and at any randomly chosen position within this region in the identification task. The figure outlines examples of target positioning in the upper right quadrant (orange) and in the lower left quadrant (purple). A dot marking screen center is included to aid judgment of geometry. Note, that during shape identification, the four outer quadrants were possible but not likely target positions.

a written consent form according to the World Medical Association Helsinki Declaration. Prior to the experiment they were informed about the procedures and the general intention of the studies, 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 it could be shared with other researchers for scientific purposes only. After the measurements, a summary and data explanation was provided for each participant. Experiment III was executed with a subset of 19 participants, who were willing to participate in a further extra lab session at a consecutive day. The procedures in this study were approved by the Ethics Committee of the Department of Psychology, Johannes Gutenberg University, Mainz, Germany (reference number: 2024-JGU-psychEK-014).

2.5. Apparatus

Textures were generated using python 3.6 and the psychopy-Toolbox (Peirce, 2008). They were displayed on an EIZO ColorEdge CG2420 monitor with a pixel resolution of 1920 × 1200 pixel, a refresh rate of 60 Hz, and gamma set to 1.0 for all three colors. Gray values were taken from this gamma-corrected 8 bit linear staircase (lookup-table with 256 entries). The room had no windows and was illuminated indirectly through wall and ceiling reflections, realized by directing two brightness-adjustable stand spotlights towards the ceiling. This ensured that no direct light fell onto the screen. The ambient illumination was adjusted to approximately match the lightness of the computer display, which had a mean luminance (luminance of the lookup-table value 128) of 100 cd/m^2, controlled by a build-in photometer. Patterns were viewed binocularly at a distance of 70 cm. Participants gave their responses via the left mouse button by clicking on one of the possible response alternatives presented at the end of each trial.

2.6. Procedure

2.6.1. Experiment I

First, participants were familiarized with the stimulus material, the target shapes, and the tasks during a training period employing well discriminable texture targets with high feature contrasts. Afterwards, a calibration phase followed during which feature contrast parameters were estimated such that equal single-cue accuracy could be expected in the main experiment at three different performance levels (*single-cue baselines*). After these parameters were found for a given task, the main experiment was executed, which comprised 3 experimental units. In a counter-balanced manner, half of the participants started the calibration — main experiment sequence with the detection task (Experiment

Table 1
Feature parameter differences of target and reference textures (feature contrasts) for the three single-cue baselines and the calibration to either the target detection task or the shape identification task. The table lists the median values, Q_{50}, of the difference to the reference texture parameters, together with the inter-quartile ranges (IQRs), taken from all participants and samples of Experiment I. Values are given in octaves for spatial frequency and in degrees of clockwise rotation away from the background for orientation.

Task	Feature	Low: p_c = 40%/d' = .52		Medium: p_c = 50%/d' = .84		High: p_c = 60%/d' = 1.16	
		Q_{50}	IQR	Q_{50}	IQR	Q_{50}	IQR
Detection	f	0.193	0.107	0.263	0.138	0.330	0.120
	ϕ	7.00	3.00	9.00	3.00	11.00	4.00
Identification	f	0.536	0.216	0.697	0.217	0.848	0.327
	ϕ	16.00	6.00	19.00	8.81	22.33	9.00

I_D), and the other half with the identification task (Experiment I_I) in random allocation. An experimental session with calibration and following 3 experimental units lasted about 90 min. To complete the two tasks, two experimental sessions were held on two consecutive days.

Calibration measurements. Separately for each task, three feature parameter values, each for spatial frequency and orientation, were chosen arbitrarily. The initial choice of the parameter values was guided by results from self-tests and accumulating experience during the course of testing. After measurement with 32 trials per feature contrast level, the resulting accuracy data were transformed into the d' sensitivity measure, and the feature contrast versus d' data were fitted with linear functions. These functions were used to estimate the feature contrasts yielding d' sensitivities d' = 0.52, d' = 0.84 and d' = 1.16, which correspond to accuracies p_c = 0.4, p_c = 0.5 and p_c = 0.6, respectively (see Section "Relating proportion correct to d'" below). The feature contrasts in f and ϕ corresponding to each of these single-cue baseline levels were saved for defining single-cue and double-cue targets in the main experiments. The calibration measurements comprised 8 replications × 4 shapes × 3 feature contrast levels = 96 trials for each of the 2 feature conditions, which were administered as separate experimental blocks. Executing both blocks took about 15 min in total.

Main experiments. In the main experiments, the feature contrast parameters estimated to reach equal single-cue performance at each baseline level were used to define single-cue and double-cue texture targets, resulting in 288 trials for one experimental unit with the same number of replications as used in the calibration measurement, since $f+\phi$ was added. After one unit was completed the experimenter briefly screened the results. If a participant's single-cue sensitivities deviated substantially from the targeted baseline levels, feature contrasts were slightly adjusted to ensure that single-cue performance met the targeted calibration levels in the next measurement unit. This compensated individual sensitivity changes, arising mostly as a result of learning effects in the early course of experimentation, and could affect orientation and spatial frequency features differently. Indeed, some adaption of parameters was necessary in nearly all cases before a new unit could be started. An experimental unit lasted about 25 min.

Table 1 shows the median values of the feature contrasts used in Experiment I. These were more than doubled for shape identification compared to detection at the same baseline levels. This reflects the large difference in required figure-background separation of physical parameters necessary to reach equal psychophysical performance in these two tasks. Fig. 1B shows an illustration of example stimuli for the low, medium, and high levels. At medium and high levels in identification, a salient texture segmentation into figure and surround is reached.[1]

[1] While notably higher feature contrasts are necessary to reach the same performance level in shape identification as in target detection, it is important to note that the feature contrasts used in the identification task are still insufficient to warrant separated processing of target and background by different primary coding mechanisms. Orientation feature contrasts in the range of 16° to 22° (see Table 1) obviously fall within the orientation bandwidth estimates for primary orientation selective mechanisms, since orientation half-bandwidth estimates have been found to decrease with increasing spatial frequency, ranging from about 30 degrees at 0.5 cpd to 15 degrees at 11.3 cpd (Wilson et al., 1983). Using an adaptation paradigm, Snowden (1992) obtained quite similar estimates in terms of the standard deviation of best fitting Gaussian tuning curves. Similarly, spatial frequency separation of target and background in the range of 0.5 to 0.85 octaves is in the range of bandwidth estimates of primary coding mechanisms. Using a suprathreshold oblique masking technique, Wilson et al. (1983) estimated half-amplitude bandwidths to range between 2.5 octaves at very low spatial frequencies below 1 cpd to about 1.25 octaves at very high spatial frequencies of 22 cpd. Using a subthreshold summation contrast detection technique and taking into account effects of probability summation, Watson (1982) arrived at an average bandwidth estimate of about half an octave. Generally, a spatial frequency separation of 1.6 octaves (f-3f experiments, see Graham, 1989) is regarded as necessary to warrant independent processing of grating stimulus components.

2.6.2. Experiment II

Subsequent to Experiment I, we tested each participant's detection performance with their feature contrasts used in the final experimental unit of the identification task, and, vice versa, we also tested identification performance with the individual feature contrast settings used in the last experimental unit of the detection task. This was done to verify whether the substantially different levels of feature contrast required for equated performance in both tasks would render one task trivial (in the case of detection) and the other infeasible (in the case of identification). We reduced the number of replications to 16 to relieve participants. An experimental unit of Experiment II comprised 16 replications (4 of each shape) × 3 feature contrast levels × 3 feature conditions = 144 trials, which lasted about 12 min. Participants were free to complete the two units still in the second experimental session or to come for a third session the next day.

2.6.3. Experiment III

To achieve maximal feature contrasts, target spatial frequency was increased about 2.1 octaves relative to background, resulting in a target texture spatial frequency of 15 cpd, which is 4.3 times the spatial frequency of the reference texture. Target orientation was orthogonal (rotated by 90°) to the reference texture (compare Fig. 8B for examples of such highly salient target textures). Experiment III consisted of brief tests comprising 3 feature conditions × 32 replications = 96 trials for each task. These were run three times to highly stable proportion correct rates, resulting in 288 trials per task. Additionally, both tasks were executed with high contrast shaded figures (see Fig. 8B, left column). These tests were run only once, and with 96 trials for each task. Experiment III was conducted in an own experimental session, which lasted about 45 min.

2.7. Relating proportion correct to d'

Performance was registered as the proportion of correct judgments, p_c. Elliot (1964) first related proportion correct to the sensitivity measure d' from signal detection theory (SDT) for the general case of M stimulus alternatives, assuming that these activate separate, orthogonal

channels and that the observer monitors them without bias. Within the framework of SDT, the case of M stimulus alternatives is analytically equivalent to the case of M spatial or temporal intervals with possible stimulus presence (M-AFC task, see MacMillan & Creelman, 2005, p. 246–249). Proportion correct is then related to d' by

$$P(d') = \int_{-\infty}^{\infty} \varphi(x - d')\Phi(x)^{M-1} dx, \tag{2}$$

which is the probability that the random variable in the signal distribution exceeds the values of the remainder $M - 1$ identically and normally distributed noise variables (Green et al., 1966, p. 69; Hacker & Ratcliff, 1979). In Eq. (2), φ is the normal probability density, and Φ the corresponding distribution function. Hacker and Ratcliff (1979) used Eq. (2) to provide detailed tables for different numbers of alternatives. To receive a continuous measure, these tables can be approximated by normal quantile functions, $d'_{\text{M-AFC}} = \Phi^{-1}(p_c; \mu_M, \sigma_M)$. Here, μ_M and σ_M are the mean and standard deviation of the Gaussian and unique for each number of alternatives M. In a four-alternative forced-choice task (4-AFC), d' can be obtained from p_c with $\mu_4 = 0.840$ and $\sigma_4 = 1.25$ as

$$d'_{\text{4-AFC}} = \Phi^{-1}(p_c; 0.840, 1.25). \tag{3}$$

Comparing results of Eq. (3) with the 4-AFC tabulation of Hacker and Ratcliff (1979) reveals that just small deviations in the second decimal place of d' occur, and only for very low and very high values of p_c, which correspond to less than one correct or incorrect judgment, respectively. We therefore used Eq. (3) to relate proportion correct to d' throughout this study.

2.8. Measures of cue summation

In SDT, it is assumed that a stimulus s maps onto a random variable ζ, composed of the sum of a deterministic system response u and a normally distributed ($N(\mu_0; \sigma)$) random variable ξ, representing the spontaneous activity of the system. Hence, $\zeta = u + \xi$ is a linear combination with mean $\mu(\zeta) = u + \mu_0$. The mean difference of random variables ζ for a stimulus present ($u > 0$, signal + noise) and absent ($u = 0$, noise), expressed in units of σ, is defined as the measure of sensitivity. It reflects the deterministic system response scaled by σ, i.e. $d' = (\mu(\zeta) - \mu_0)/\sigma) = u/\sigma$. Since the d' measures for two stimulus components s_1 and s_2 are $d'_1 = u_1/\sigma$ and $d'_2 = u_2/\sigma$, the sensitivity measure for a compound stimulus $s_{1+2} = s_1 + s_2$ is $d'_{1+2} = (\mu(u_1 + u_2 + \xi) - \mu_0)/\sigma = (u_1 + u_2)/\sigma$, which means that $d'_{1+2} = d'_1 + d'_2$. As outlined in Tanner's theory of recognition (Tanner, 1956), mapping the compound stimulus onto a single random variable yields the sum of the component sensitivities, $d'_\Sigma = d'_{1+2}$, as the theoretical upper limit imposed by the triangle inequality for linear combinations (see also MacMillan & Creelman, 2005, pp. 191; Persike & Meinhardt, 2006). If, however, the two components of the compound stimulus are mapped onto two orthogonal random variables ζ_1, ζ_2, the resulting mean difference of the two-dimensional distributions for signal + noise and noise is $d'_\perp = \sqrt{(d'_1)^2 + (d'_2)^2}$, a case Tanner called "dimensional orthogonality". Additive summation of sensitivities and dimensional orthogonality arise as the two extremes in the linear combinations of the mean vectors from two, generally correlated, random variables ζ_1, ζ_2. Linear summation implies perfect correlation and dimensional orthogonality independent random activity elicited by the two stimulus components.

Since comparison with the theoretical maximum for linear combinations of random vectors is a useful benchmark for evaluating the strength of cue summation, we define

$$q = \frac{d'_{f+\phi}}{d'_f + d'_\phi} \tag{4}$$

as the *summation ratio*. If q reaches 1, the observed double-cue sensitivity is the linear sum of both single-cue sensitivities. If q equals $1/\sqrt{2} \approx 0.7071$, it coincides with the prediction of dimensional orthogonality for two equal single-cue d' sensitivities.

2.9. Data handling and outlier clearing

As required for d' analysis, proportion correct rates for perfect performance were replaced by $1 - 1/(2N)$. With $N = 32$ the number of replications (MacMillan & Creelman, 2009, see p. 8). The corrected accuracy maximum was $p_c = 0.984$, limiting the maximum d' sensitivity to 3.526. For Experiments II and III proportion correct rates were used for statistical analysis, so no replacements for perfect or null performance were made. Inspection of sample q-q-plots showed just two outliers, one in the d'_f sample at the low single-cue baseline and one in the d'_ϕ sample at the mid-level single cue baseline of Experiment I_D. We decided to keep them in the sample, since testing with and without these values showed practically same results. For statistical interference analysis rmANOVA designs were used. Analyses were completed in R (Team, 2012) and access to the data and statistical analysis code is provided via the Open Science Framework (https://osf.io/h6g5u/).

3. Results

3.1. Experiment I: Cue summation in target detection and shape identification with equal single-cue baselines

Fig. 4 shows d' means for all feature conditions at the three calibrated single-cue performance levels (baseline $d' \approx 0.52$, 0.84 and 1.16, corresponding to panels A, B, and C, respectively) of Experiment I_D (purple circles) and Experiment I_I (orange triangles). These levels were achieved by adjusting the single-cue stimulus contrasts individually for each participant. Sensitivity equivalence of single cues across features and tasks was met with good accuracy at all three levels. Accordingly, none of these single-cue baselines showed pairwise significant differences (Low: $p_{b_D} = 0.385$, $p_{b_I} = 0.929$, Medium: $p_{b_D} = 0.793$, $p_{b_I} = 0.728$, High: $p_{b_D} = 0.641$, $p_{b_I} = 0.674$) which were excellent prerequisites for comparing the cue combination effect across tasks. Table 2 shows the results of testing mean d' double-cue sensitivity against the predictions from additive summation and dimensional orthogonality. The data reflect a unique pattern of cue summation effects relative to the baseline d' levels established for single-cue conditions. As Fig. 5 shows, an almost constant summation ratio of $q \approx 1$ in detection indicates that the double-cue sensitivity remained consistently at the level expected from additive summation of feature cues across all three baseline levels. Double-cue sensitivity was always significantly greater than the level expected from dimensional orthogonality. One-sample tests confirmed these results (see Table 2, $p_\perp$ and p_Σ entries for detection). In contrast, a summation ratio decreasing from a value of 1.15 at the low single-cue baseline to a value of 0.74 at the high baseline indicated decreasing cue summation in identification, starting slightly above the level expected from additive summation but ending at the level expected from dimensional orthogonality at the high single-cue baseline. Again, one-sample tests confirmed these findings (see Table 2, $p_\perp$ and p_Σ entries for identification).

To test the different trends of cue summation in detection and identification, we compared the effects of *Feature* (d'_f, d'_ϕ, $d'_{f+\phi}$), *Task* ('detection', 'identification'), *Measurement* (3 runs of each main experiment), and their interactions in repeated measurement ANOVAs for each single-cue baseline. Table 3 shows the results. As expected, *Feature* always gave the strongest single effect with large to very large effect sizes. Most importantly, the *Feature* × *Task* interaction changed with increasing baseline level, being non-significant for the low baseline, but strongly significant at the medium and high baselines. Effect sizes of the interaction term increased from a small effect ($\eta_p^2 = 0.054$) to a large effect for the medium ($\eta_p^2 = 0.292$) and an even larger effect for the high ($\eta_p^2 = 0.477$) baseline. These results for the *Feature* × *Task* effect thus confirm that detection and identification results differed in the double-cue effect when the single-cue baseline increased, since single-cue performance agreed fairly well across tasks at each baseline level (see Fig. 4 and p_b results above). The main effect of *Task* for the high

Table 2
Results overview for testing the double-cue advantage against cue summation predictions from dimensional orthogonality and additive summation among spatial frequency and orientation cues. Results are shown for low, medium, and high single-cue baselines, both for the detection task (Experiment I_D), and the identification task (Experiment I_I). The table shows the single-cue sensitivities, $\bar{d}'_f$ and $\bar{d}'_\phi$ and the mean sensitivity for double-cue targets, $\bar{d}'_{f+\phi}$. The following columns show sensitivity predictions derived from dimensional orthogonality, $d'_\perp = \sqrt{\bar{d}'^2_f + \bar{d}'^2_\phi}$, and additive summation, $d'_\Sigma = \bar{d}'_f + \bar{d}'_\phi$, with the corresponding probabilities that these predictions fall within the 95% confidence interval of $\bar{d}'_{f+\phi}$, calculated from two-tailed, one sample t- tests (columns $p_\perp$ and p_Σ).

Low baseline ($d' = .52/p_c = 40\%$)							
Task	$\bar{d}'_f$	$\bar{d}'_\phi$	$\bar{d}'_{f+\phi}$	Dimensional orthogonality ($\perp$)		Additive summation (Σ)	
				$d'_\perp$	$p_\perp$	d'_Σ	p_Σ
Detection	0.516	0.580	1.15	0.776	0.009	1.10	0.518
Identification	0.565	0.571	1.31	0.803	<0.001	1.14	0.019
Medium baseline ($d' = .84/p_c = 50\%$)							
Task	$\bar{d}'_f$	$\bar{d}'_\phi$	$\bar{d}'_{f+\phi}$	Dimensional orthogonality ($\perp$)		Additive summation (Σ)	
				$d'_\perp$	$p_\perp$	d'_Σ	p_Σ
Detection	0.878	0.898	1.89	1.26	<0.001	1.78	0.140
Identification	0.940	0.923	1.56	1.32	0.085	1.86	<0.001
High baseline ($d' = 1.16/p_c = 60\%$)							
Task	$\bar{d}'_f$	$\bar{d}'_\phi$	$\bar{d}'_{f+\phi}$	Dimensional orthogonality ($\perp$)		Additive summation (Σ)	
				$d'_\perp$	$p_\perp$	d'_Σ	p_Σ
Detection	1.25	1.28	2.45	1.79	<0.001	2.53	0.292
Identification	1.23	1.21	1.80	1.73	0.727	2.44	<0.001

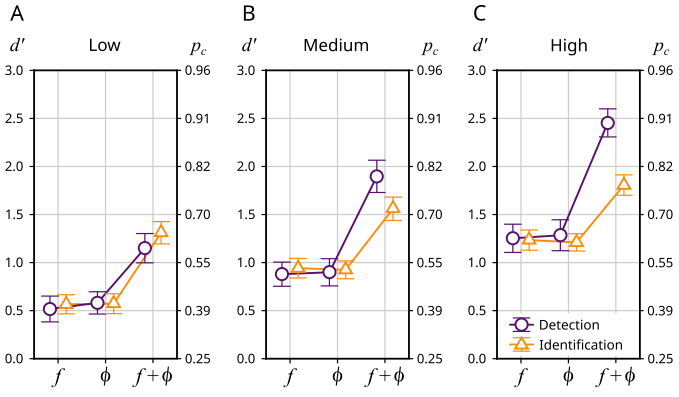

Fig. 4. Mean d' sensitivities for target texture detection (purple circles) and target shape identification (orange triangles) for spatial frequency (f), orientation (ϕ), and double-cue ($f + \phi$) targets. The three panels show the data for low (A), medium (B), and high (C) single-cue sensitivity levels. Bars indicate 95% confidence intervals of the means.

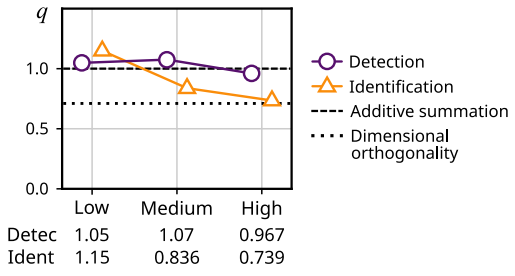

Fig. 5. Summation-ratios (q) of the mean d' sensitivities shown in Fig. 4 for target texture detection (purple circles) and target shape identification (orange triangles). The dotted line marks the q-ratio predicted under dimensional orthogonality ($q = 1/\sqrt{2} \approx 0.707$), the dashed line indicates the prediction under additive cue summation ($q = 1$).

single-cue baseline also reached significance (large effect, $\eta_p^2 = 0.321$). Since the single-cue sensitivities were equal across tasks, this effect, too, is owed to the increasing separation of tasks just on double-cue sensitivity.

In the two ANOVAs for the medium and high baseline levels, the *Measurement* effect reached significance with moderate effect sizes.

Table 3
Results of separate ANOVAs for each single-cue baseline level. The table lists degrees of freedom, sums of squares for the effect (SS) and error (SS_E) terms, F- values, Greenhouse–Geisser corrected significance levels p, and partial η_p^2 as effect sizes.

Low baseline ($d' = .52/p_c = 40\%$)						
Factor	df	SS	SS_E	F	p	η_p^2
Feature	(2,48)	44.8	7.09	151	<0.001***	.863
Task	(1,24)	.397	11.3	.843	.368	.034
Measurement	(2,48)	.839	28.5	.705	.486	.029
Feature × Task	(2,48)	.552	9.61	1.38	.261	.054
Feature × Measurement	(4,96)	.445	15.4	.693	.586	.028
Task × Measurement	(2,48)	.282	21.7	.312	.729	.013
Feature × Task × Measurement	(4,96)	1.57	13.8	2.72	.042*	.102
Medium baseline ($d' = .84/p_c = 50\%$)						
Factor	df	SS	SS_E	F	p	η_p^2
Feature	(2,48)	67.3	8.01	202	<0.001***	.894
Task	(1,24)	1.09	13.2	1.98	.173	.076
Measurement	(2,48)	4.74	32.5	3.49	0.038*	.127
Feature × Task	(2,48)	3.49	8.49	9.88	<0.001***	.292
Feature × Measurement	(4,96)	.31	14.0	.540	.707	.022
Task × Measurement	(2,48)	1.24	21.7	1.37	.264	.054
Feature × Task × Measurement	(4,96)	.71	13.9	1.22	.307	.048
High baseline ($d' = 1.16// p_c = 60\%$)						
Factor	df	SS	SS_E	F	p	η_p^2
Feature	(2,48)	77.9	6.96	269	<0.001***	.918
Task	(1,24)	6.97	14.7	11.3	0.003**	.321
Measurement	(2,48)	5.62	25.4	5.32	0.008**	.181
Feature × Task	(2,48)	9.13	10.0	21.9	<0.001***	.477
Feature × Measurement	(4,96)	1.26	17.9	1.69	.158	.066
Task × Measurement	(2,48)	.52	23.3	.536	.588	.022
Feature × Task × Measurement	(4,96)	1.59	15.9	2.41	.055	.091

Due to imperfect individual calibration and learning effects, single-cue performance in the first experimental unit often exceeded the levels intended by the calibration procedure. To address this, we adjusted parameters across the three runs of Experiment I to better align individual single-cue performance with the targeted baselines. As a result, and partly due to a tendency to overcompensate for perceptual learning, single-cue performance declined slightly across runs, starting above the intended baseline in the first run and ending slightly below it in the last. However, these technical measurement-related effects were not overly strong.

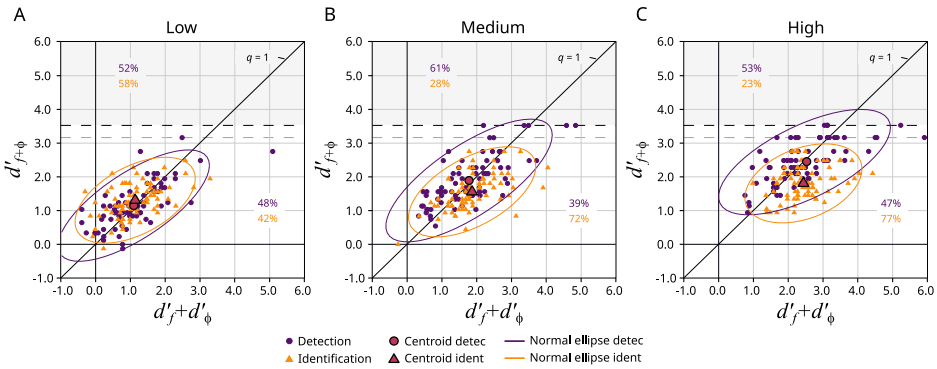

Fig. 6. Double-cue sensitivity ($d'_{f+\phi}$, y-axis) plotted against the sum of both single-cue sensitivities ($d'_f + d'_\phi$, x-axis) for Experiment I_D (detection, purple dots) and I_I (identification, orange triangles), for low, medium, and high single-cue baselines. The diagonal indicates a summation ratio of $q = 1$. Ellipses indicate the 95% quantiles of the bivariate normal distributions for each task. Task centroids are marked by larger red symbols and numbers denote the percentage of data points above and below the diagonal for the respective task. Upper horizontal dashed lines (dark gray) correspond to the largest $d'_{f+\phi}$ value possible after correcting for perfect performance. Lower horizontal dashed lines (light gray) mark the d' sensitivity that corresponds to 31 correct judgments out of 32 trials.

Scatterplots of the raw data points jointly on the two measures of the linear sum of single-cue sensitivities, $d'_f + d'_\phi$, and double-cue sensitivity, $d'_{f+\phi}$, illustrate the unique trend (see Fig. 6). While the sum of single-cue sensitivities agreed well across tasks (same levels of x-components, see also Table 2), the task difference in the double-cue performance (y-component) increased with the single-cue baseline, moving the identification distribution more and more below the detection distribution. Note, that data points beyond the diagonal correspond to an over-summative double-cue benefit in this data triplet, while data points below the diagonal indicate a double-cue sensitivity less than the linear sum of the single-cue values. The detection data remain about equally distributed above and below the diagonal as the single-cue baseline increases (low: 52:48, medium: 61:39, high: 53:47; see Fig. 6). In contrast, data distributions for the identification task start equivalently (low: 58:42), but the ratio continuously decreases with increasing single-cue baseline (medium: 28:72, high: 23:77). This illustrates that a double-cue benefit consistent with additive summation for detection and a strongly declining double-cue benefit for identification are also found at the level of individual data.

Occasionally, there was ceiling performance in the detection of double-cue stimuli (see 5 data points on the upper dashed line in Fig. 6B and 3 data points in 6C), while cases that observers missed detection of just one target were frequent at the high baseline level. For shape identification, however, none of the data triplets displayed either perfect performance or just a single error. Participants consistently made at least two errors for double-cue targets at medium and high baseline levels. This stresses how, for equally difficult single-cue judgments, the double-cue target consistently conveys a strong benefit for detection while barely improving shape judgments any further when single-cue performance as such gets better and better.

3.2. Experiment II: Testing detection and identification performance across tasks

Fig. 7 shows the mean accuracies achieved in Experiment II, given as percent values of the proportion of correct judgments. At medium and high baseline levels, accuracy in detection tested with the feature contrasts used for identification fell within the expected range of saturated performance where maximally one error was made. At the low baseline level, single-cue performance had not yet saturated but was close to that level (see arrows in 7A), while double-cue accuracy had already saturated. Accuracy rates in identification tested with the feature contrasts calibrated for detection fell within the expected range of chance performance, except for two double-cue targets that yielded modestly higher rates at medium and high baseline levels ($p_c = 0.349$ and $p_c = 0.398$, see arrows in 7B and 7C). Results confirm that the

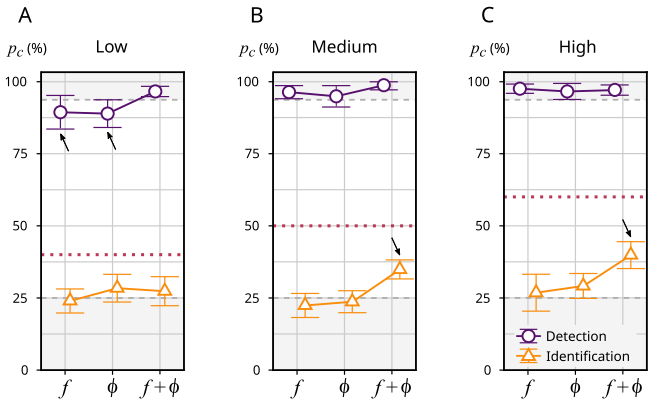

Fig. 7. Accuracy in Experiment II, measured as percentage of correct responses, for the three single-cue baselines. Error bars indicate 95% confidence intervals of the mean proportions. The dotted red lines indicate the calibration baseline levels in each panel. The upper gray zone marks the area corresponding to near perfect accuracy, with less than one wrong judgment. The lower gray zone marks the area below chance performance.

feature contrasts calibrated for identification raise salient texture segregation. Vice versa, the feature contrasts calibrated for detection almost preclude better than chance shape identifications. The above chance identification rates for double-cue targets suggest that segmentation of texture into target and non-target regions is sufficient for elementary shape recognition performance at medium and high single-cue baselines. Overall, detection and identification appear as well separated perceptual tasks over the whole range of investigated feature contrasts.

3.3. Experiment III: Testing detection and identification at maximal feature contrasts

The fact that participants did not, and not even occasionally, reach perfect or near perfect performance in identification while this happened frequently in detection (see Fig. 6) let us surmise that large feature differences among texture regions are sufficient for detecting target presence, but might be insufficient for a unique recognition of shape. To test this conjecture, we executed Experiment III, using maximal feature contrasts for the texture targets and also using shaded figures as controls. Similar to the results of Experiment II, perfect performance was reached in detection, with just occasional lapses a few subjects made on single trials (see Fig. 8). This held for high feature contrast target textures and shaded figures alike. In identification, however, the picture was different. With shaded figures there was

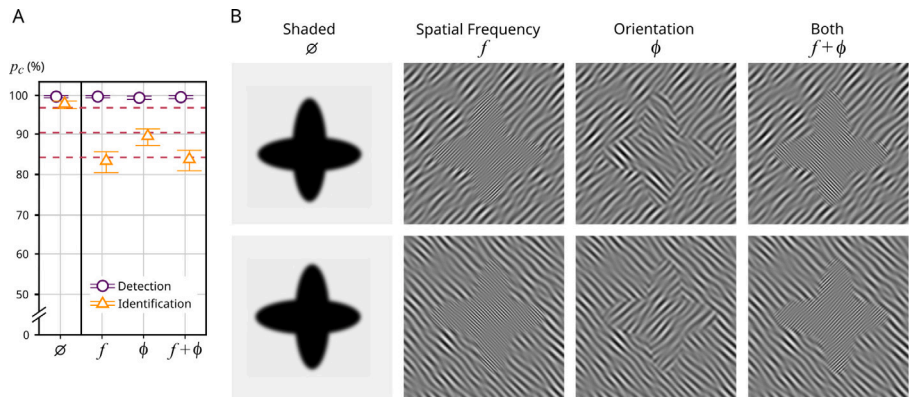

Fig. 8. Results (A) and stimulus examples (B) of Experiment III. (A) Accuracy in detection and identification (percentage correct) for shaded figures (∅ symbol) and the three texture feature conditions at maximal feature contrasts. Error bars indicate 95% confidence intervals of the proportions. Dashed lines indicate the accuracy rates corresponding to 1, 3, and 5 errors. (B) Sample pictures of two target shapes as shaded figures and for high feature contrasts in spatial frequency, orientation, and both features.

near perfect performance, but with some more lapses compared to detection (average percentage correct = 97.6%, one error from 32 trials corresponds to 3.125%). For high feature contrast targets, however, accuracy of target shape identification from texture was clearly reduced compared to shaded figure identification, with accuracies of 83.2% (f), 89.5% (ϕ), and 83.6% ($f + \phi$; see Fig. 8A). An ANOVA substantiated these differential task effects. There was a highly significant *Task* effect, showing that detection accuracy was larger than identification accuracy ($F(1, 56) = 188.6, p < 0.001$). Also an effect of *Feature* ($F(2, 112) = 19.35, p < 0.001$) was found, but the latter stemmed from the *Feature* × *Task* interaction ($F(2, 112) = 22.35, p < 0.001$), which showed a larger accuracy in orientation compared to spatial frequency or double-cue targets only in the identification task. Here, post-hoc comparisons showed that orientation reached higher accuracy than spatial frequency ($\Delta\mu = 6.3\%, t(56) = 5.43, p < 0.001$) and their combination ($\Delta\mu = 5.9\%, t(56) = 5.87, p < 0.001$), while performance with the latter was not better compared to spatial frequency ($\Delta\mu = 0.4\%, t(56) = 0.43, p = 0.672$; Bonferroni corrected 1% $\alpha = 0.0034$).

Results thus showed that there is near perfect shape identification with shaded figures, while feature contrast alone is insufficient to reach this level in shape identification from texture. This clearly suggests that recognizing shape from texture cues either engages partially distinct mechanisms or that texture interferes with the mechanisms typically involved in recognizing luminance-defined shapes. Further, accuracy slightly above 80% seems to be a limit for the combination of spatial frequency and orientation, which shows no advantage to the spatial frequency cue alone. Instead, accuracy is higher with orthogonal textures of the same spatial frequency band, indicating a configural effect of cue summation for 2D figure identification with highly salient target textures: Even with non-matched single-cue performances we would expect double-cue performance to be as least as good as the best of both single cue performances. However, double-cue performance agrees with the worst of both.

4. Discussion

Measuring the double-cue benefit for detection and identification of 2D texture shapes, having single-cue sensitivity equated across tasks and features, showed quite different results for either task in the range that allowed for double-cue sensitivity measurement not seriously constrained by ceiling effects. Over this range, double-cue sensitivity was accurately predicted by the sum of the single-cue sensitivities in the detection task (summation ratio $q \approx 1$), suggesting integration by a single mechanism. In the identification task, however, double-cue sensitivity started with the same strength as observed for detection at the lowest sensitivity baseline, but declined to levels not better than expected from independent cue processing at the largest baseline level.

Together with the results from control Experiments II and III, these differential findings for the double-cue benefit in detection and identification indicate that cue summation enhances segregation and segmentation of texture regions, while it does not seem to improve more specific processes involved in the extraction and recognition of shape from texture cues.

4.1. Cue summation in texture segregation: a saliency based mechanism

Early research on texture segregation has shown that the human capability to segregate and to segment differently textured regions is achieved by automatic and parallel processing at all spatial coordinates of the visual scene without serial scanning of image parts (Bergen & Julesz, 1983). Neural correlates of such massive parallel processing have been found in area V1 (Kastner et al., 1997; Knierim & van Essen, 1992; Lamme et al., 1992), indicating that the fundamental steps of figure-ground segregation occur early on the visual pathway (Lamme, 1995; Lamme et al., 1992).

The psychophysical correlate of texture segregation is saliency, which scales with the magnitude of local feature differences between two areas (Nothdurft, 1991). Scaling of saliency by feature contrast has been measured by pedestal versus increment functions (Naka-Rushton: Motoyoshi & Nishida, 2001; Weber-Fechner: Meinhardt & Persike, 2003), cross-modal matching (Nothdurft, 2000), and also indirectly via threshold presentation time (Bach & Meigen, 1999). As in this study, Straube and colleagues measured texture figure detection via a localization task and also shape identification (Straube et al., 2010), while they additionally recorded ERPs. In each task, authors observed a negative shift of the posterior P2 at about 200 ms latency unspecific of the feature condition (spatial frequency contrast, orientation contrast, or both) in P2 latency and in the overall ERP time course. The negativity of the P2 amplitude correlated strongly with the d' measure from the psychophysical task, indicating stronger negativity for larger target-background feature differences and also for double-cue targets. The latter showed advantage in the d' measure, which agreed fairly well with the magnitude of the double-cue effects reported here. Similar findings were reported by Bach et al. (2000), who showed stronger negativity in the texture-segregation visual evoked potential (tsVEP, Bach and Meigen (1992)) for double-cue (spatial frequency + orientation) textures, with otherwise same temporal response characteristics as for the single-cue tsVEPs. These results indicate that summation of spatial frequency and orientation gradients is a mechanism which enhances the saliency of segregation at early levels of stimulus processing. Since detection or relatively crude localization of a target texture does not require more than registering a segregation signal, which is absent in the reference stimulus, a mechanism augmenting local differences of texture regions is sufficient to explain the double-cue advantage in

detection and localization tasks. As shown recently (Hunt & Meinhardt, 2021), the cue summation effect of spatial frequency and orientation is captured by a multi-channel local energy-based model (Bergen & Adelson, 1988; Landy & Bergen, 1991; Malik & Perona, 1990), implementing the non-Fourier second-order squaring and pooling operation found in chains of V1 and V2 cells (Von der Heydt et al., 1984; Lin & Wilson, 1996). It relates the double-cue advantage to an enhancement of the segregation-related maximum net contrast energy, which measures the difference of the summed local energy responses to target and reference textures. Closer analysis showed that the cue summation effect already arises in the first processing steps of local filtering and nonlinear rectification, but not at the later second order filtering stage (Hunt & Meinhardt, 2021, p. 16).

Local energy filter-rectify-filter (FRF) chains respond to spatial texture modulation, while the nature of the modulated first order feature is lost in final output (Caelli, 1995; Lin & Wilson, 1996). To qualitatively account for human texture discrimination, FRF models must be supplemented with a *labeling* mechanism for the first-order input, as observers are typically aware of the type of feature modulation when detecting and discriminating textures (Prins & Kingdom, 2003). To test integration of sub- and suprathreshold texture modulation by a common mechanism, Kingdom et al. (2003) measured discrimination thresholds for various pedestal feature contrasts for textures modulated in spatial frequency, orientation, luminance contrast and their pairwise cross-modal combinations. When test and reference patterns were modulated in the same feature, increment thresholds as functions of the varied pedestal contrasts had the typical *dipper* shape, indicating interaction within a single mechanism. However, when the test pattern added feature contrast from a different feature (cross-modal), the increment thresholds coincided with the detection threshold of this specific feature, independent of whether the texture modulation of the pedestal was salient or barely visible, indicating that reference and test did not interact. Authors viewed this as evidence for feature independence of texture mechanisms. However, since the task was to identify the test pattern, observers could have responded just to the new 1st order feature in the display, but not to a potentially stronger saliency of the conjoint texture modulation.[2]

To avoid confounding texture modulation and saliency with first-order feature discriminability, saliency could be matched against a third standard, or discriminated among conjoint feature contrast targets with different component strengths. Using a luminance matching standard, Nothdurft (2000) found that supra-threshold texture singletons with feature contrast in two dimensions (pairwise combinations of color, motion, orientation and luminance) were judged as more salient than the single cues, albeit saliency effects across features did not add linearly. Meinhardt and Persike (2003) used classical Weber–Fechner scaling for briefly presented (125 ms) spatial frequency and orientation singletons in Gabor arrays. The feature contrasts reached at the fourth JND (*just noticable difference*) for each single feature were combined to create double-cue stimuli. These stimuli were then used as reference stimuli against which double-cue targets with stronger feature contrast components were to be discriminated.[3] For these quite salient targets, results showed that the threshold for double-cue targets was 79.7% of the single-cue threshold, which is a degree of summation expected from an inclusive-or rule for independent feature-specific mechanisms (Graham, 1989). At detection (JND 0), the double-cue threshold was 63% of the single-cue threshold, indicating strong component summation. Albeit there are no measurements for intermediate supra-threshold saliency levels of double-cue texture stimuli, these results suggest that saliency summation across features continuously declines after the detection threshold (JND 0), falling to levels expected from independent cue interaction when texture figure saliency rises.

4.2. Cue summation in the identification of texture figures reflects just the saliency based advantage

Straube and Fahle (2011) have pointed out that it is difficult to distinguish detection and identification conceptually, since the demands for either task may be similar or highly different, depending on the experimental settings. As stated in the Introduction, we pursued to maximally separate the task demands by defining a shape identification task which could be handled only if detection of the target texture had already occurred (Barlasov-Ioffe & Hochstein, 2008; Sagi & Julesz, 1984). Our data show that we succeeded well in establishing this setting. Post-testing detection with the feature contrasts used in identification showed high to near perfect detection rates (see Fig. 7), indicating that texture figure segregation had already settled at these feature contrasts. On the other hand, the feature contrasts calibrated for detection practically precluded better than chance performance in shape identification, thus ruling out that mechanisms specifically tuned to shape processing underlie the double-cue benefit in detection.

The only instance where the double-cue effect in identification met the double-cue effect in detection occurred at the low single-cue baseline (see Fig. 4A). Post-testing detection with these feature contrasts showed that single-cue detection was still not perfect (see arrows in Fig. 7A), while accuracy saturated for double-cue targets. Apparently, cue summation still led to improved texture segregation and segmentation, notably boosting shape recognition performance from the low levels achieved with single cues.

Increasing the target-background feature differences for identification further showed near perfect detection performance already with single cues (see Figs. 7B and 7C), while the double-cue advantage for identification continuously declined (see Figs. 4B and 4C and 5). This indicates that, once segmentation into figure and surround is complete, a saliency-based mechanism becomes less important for shape identification.[4] This conclusion is also supported by comparing the results of Experiments I and III. Note that the maximal accuracy in identification was p_c = 77.9%, reached with double-cue targets (see Fig. 4C). Here, feature contrasts were 0.848 octaves in spatial frequency and 22.3° in orientation (see Table 1). In Experiment III, we increased spatial frequency contrast to 2.1 octaves and orientation contrast to 90°, but this improved double-cue performance to just 83.6%. Yet, identification accuracy for shaded figures was near perfect (see Fig. 8A), highlighting that strong texture segmentation alone does not guarantee successful shape identification when shape must be inferred from texture-defined boundaries, as opposed to luminance-based contours.

[2] See Fig. 2 in Prins and Kingdom (2003). The conjoint OM+FM texture grating (e) is more salient than the OM and FM single feature texture gratings (a, OM) and (b, FM), but observers could also attend the single feature which is present in (e) but not in (a) or (b) if the task is to distinguish single-cue from double-cue modulation.

[3] Although there was considerable deviation among subjects, the feature contrasts of the 4th JND for texture singleton discrimination (see ibid, Table 1) are in a range which is comparable to the feature contrasts used at the high single-cue baseline in the shape identification task of this study (see Table 1).

[4] Zipser et al. (1996) recorded from 140 V1 cells of awake macaque monkeys viewing texture figures (disks) modulated in disparity, color, luminance, orientation and their combination. They found differential response modulation according to the presence or absence of a texture figure after a latency of about 80 to 120 ms, indicating that feedback from extrastriate areas beyond cortical area V2 may be responsible for the contextual modulation of V1 responses, which acts as a mechanism for coding figure-ground segregation (Lamme, 1995; Lee et al., 1998). Using maximal feature contrasts for each single cue, authors did not find stronger effects in response amplitude or contextual modulation for figures defined by multiple cues, indicating that figure-ground segregation was already fully established with the single cues. As our study also used closed, convex texture figures as targets, it is suggested that the decline in cue summation with increasing feature contrast level after detection relates to the feed-forward component of the functional chain of figure-ground segregation, specifically the segregation and segmentation of texture regions.

4.3. Cue summation beyond saliency?

The differential cue combination results for detection and shape identification shed a new light on earlier studies about the functional role of cue combination for figure-ground segregation and shape recognition. First, they confirm earlier results of Persike and Meinhardt (2008), who reported reduced cue combination effects for more difficult shape judgments which require higher target saliency in Gabor random fields to be feasible. Straube et al. (2010) measured shape identification performance, but since both the psychophysical and the ERP results completely aligned with the results for mere target localization (Straube & Fahle, 2010), there is no indication that the double-cue advantage, obvious both as a sensitivity advantage and a stronger posterior P2 negativity, reflects something beyond a strengthened saliency of texture figure segregation in either task. As long as target detection still improves with increasing feature contrast, a double-cue advantage in identification performance is no conclusive evidence that mechanisms specifically tuned to shape processing, particularly form completion and contour grouping, benefit from cue combination.

The results of Experiment III give some important clues to the role of texture cue combination for specific shape processing routines besides texture segmentation and figure-ground segregation, since there the latter processes were completed as indicated by perfect target detection (see Fig. 8A). Accuracy for identification was notably lower than for detection, reaching accuracy levels of just above 80% to scarcely 90%. Most strikingly, results showed a reversal of the double-cue benefit at high feature contrasts, since identification performance was best with just the orientation cue. For highly salient targets additionally changing the spatial frequency band of the target area is apparently not apt to increase the accuracy of shape judgments. Instead, local orthogonal elements in the same spatial frequency band are better suited. This indicates that processes underlying contour extraction and surface formation for building 2D shape representations from texture do not benefit from combining these cues.

Note that shaded figures have real contours, whereas figures defined by regional texture differences rely on illusory contour formation (see Fig. 8B). The advantage for orthogonal orientations within the same spatial frequency band could therefore stem from the fact that this configuration supports stronger illusory contours. This interpretation is supported by recordings from end-stopped cells in V1, which show increased responses when contrast polarity in bar stimuli is inverted at the ends (Yazdanbakhsh & Livingstone, 2006). Together with corner-responsive units, end-stopped cells are thought to contribute to illusory contour perception (Von der Heydt et al., 1984; Peterhans & Von der Heydt, 1991). The results of Experiment III further suggest that these mechanisms likely do not benefit from increased differences in both orientation and spatial frequency, as varying the spatial frequency band of the target seems to cancel the local cues necessary for effective contour formation.

Beyond V1, shape perception mechanisms for luminance-defined and texture-defined targets diverge along the ventral visual stream. Recent research suggests that these two types of shape signals engage overlapping but temporally and functionally distinct pathways. Luminance-defined shapes are processed rapidly by early visual areas, with V2 neurons signaling border ownership (Williford & von der Heydt, 2016) and showing strong selectivity for luminance-based contours (El-Shamayleh & Movshon, 2011). Further downstream, in areas such as V4 and possibly LOC, neurons encode more complex boundary features, such as specific curvature at particular positions relative to the stimulus center (Pasupathy & Connor, 2001). In contrast, texture-defined regions require additional processing: V2 shows only weak selectivity for texture-defined orientation (El-Shamayleh & Movshon, 2011), and texture tuning in V4 emerges more slowly (approximately 30 ms later) than shape tuning (Kim et al., 2019), suggesting a greater reliance on lateral or feedback mechanisms to integrate surface cues into shape representations (Pasupathy et al., 2019). Despite these temporal differences, neurons in V4 and possibly LOC often independently encode both shape and texture (Kim et al., 2019), with a greater dynamic range of responses to images containing clear object boundaries compared to texture-like stimuli lacking defined outlines. Lieber et al. (2024) showed that V4 responses were highly sensitive to even minimal image scrambling that disrupted object boundaries while preserving basic texture statistics. Neural activity in V4 also appears to be modulated by local contour features that define figure shape (Shishikura et al., 2025), further emphasizing the importance of boundary information. Additionally, textured surfaces are less preferred in higher-order visual regions such as LOC and FFA when compared to smooth surfaces (Echavarria et al., 2016). This indicates that a robust contour representation emerges when the contour coincides with a local surface difference (e.g., in texture or color), while the surface information itself may emerge later and rely more on feedback or lateral processing compared to the contour alone (Pasupathy et al., 2019). Behaviorally, observers rely on both contour and surface cues to detect shapes, with a clear double-cue benefit when both cues are aligned (Machilsen & Wagemans, 2011). Overall, luminance-based shape processing appears to be more immediate and robust, whereas texture-defined shape perception depends on slower, integrative processing and thus reflects the added complexity of encoding surface structure alongside object boundaries. This view is consistent with our findings from Experiment III, where luminance-defined shapes were identified with perfect accuracy, whereas shapes defined by texture differences were notably more difficult to discriminate.

4.4. Conclusion

Studying the effects of cue combination in texture figure detection and more difficult shape identification revealed different cue summation effects in either task. A double-cue sensitivity as large as the sum of both single-cue sensitivities over a wide range of feature contrast levels in detection suggests a low-level mechanism which enhances segregation and segmentation of texture regions into target and surround. Since identification performance also benefits from cue combination at feature contrasts which still do not allow perfect target detection, both tasks ground in this saliency-based mechanism. The reversal of the double-cue advantage at high feature contrasts, reflected by superior performance with the orientation cue alone, suggests that 2D shape extraction from texture involves contour and surface formation processes that do not benefit from joint feature contrast in spatial frequency and orientation, but require other task-specific cues for optimal functioning. Taken together, these results suggest that the functional role of cue summation among basic texture cues is limited to enhancing texture segregation and segmentation as the basic steps in establishing stable figure-ground segregation. Since the mechanisms underlying these operations have been localized in early retinotopic cortices (Bergen & Julesz, 1983; Knierim & van Essen, 1992; Lamme, 1995; Lamme et al., 1992; Zipser et al., 1996), an involvement of higher ventral areas processing shape in a feature-invariant manner (Kourtzi & Huberle, 2005; Kourtzi & Kanwisher, 2000) is not suggested.

CRediT authorship contribution statement

Cordula Hunt-Radej: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Anna-Lena Schubert:** Project administration, Funding acquisition, Conceptualization. **Günter Meinhardt:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Funding: This study was supported by the Deutsche Forschungsgemeinschaft fund, Germany 471755627 given to Anna-Lena Schubert and Günter Meinhardt.

Data availability

We have included the link to the data and code in the Methods section of the manuscript.

References

Bach, M., & Meigen, T. (1992). Electrophysiological correlates of texture segregation in the human visual evoked potential. *Vision Research*, *32*(3), 417–424. http://dx.doi.org/10.1016/0042-6989(92)90233-9.

Bach, M., & Meigen, T. (1999). Electrophysiological correlates of human texture segregation, an overview. *Documenta Ophthalmologica*, *95*, 335–347. http://dx.doi.org/10.1023/a:1001864625557.

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. http://dx.doi.org/10.1016/S0042-6989(00)00195-4.

Barlasov-Ioffe, A., & Hochstein, S. (2008). Perceiving illusory contours: Figure detection and shape discrimination. *Journal of Vision*, *11*(14), 1–15. http://dx.doi.org/10.1167/8.11.14.

Bergen, J. R., & Adelson, E. H. (1988). Early vision and texture perception. *Nature*, *333*(6171), 363–364. http://dx.doi.org/10.1038/333363a0.

Bergen, J. R., & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature*, *303*(5919), 696–698. http://dx.doi.org/10.1038/303696a0.

Caelli, T. (1995). A brief overview of texture processing in machine vision. *Early Vision and Beyond*, 79–87.

Daugman, J. G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, *20*(10), 847–856. http://dx.doi.org/10.1016/0042-6989(80)90065-6.

Echavarria, C., Nasr, S., & Tootell, R. (2016). Smooth versus textured surfaces: Feature-based category selectivity in human visual cortex. *ENeuro*, *3*(5), http://dx.doi.org/10.1523/ENEURO.0051-16.2016.

El-Shamayleh, Y., & Movshon, J. A. (2011). Neuronal responses to texture-defined form in macaque visual area V2. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(23), 8543–8555. http://dx.doi.org/10.1523/JNEUROSCI.5974-10.2011.

Elliot, P. B. (1964). Tables of d'. In J. A. Swets (Ed.), *Signal detection and recognition by human observers*. New York: Wiley.

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*(1), 10. http://dx.doi.org/10.1167/7.1.10.

Graham, N. S. (1989). *Visual pattern analyzers*. Oxford University Press, http://dx.doi.org/10.1093/acprof:oso/9780195051544.001.0001.

Green, D. M., Swets, J. A., et al. (1966). *Signal detection theory and psychophysics*: *Vol. 1*, Wiley New York.

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. http://dx.doi.org/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. http://dx.doi.org/10.1111/j.1467-9280.2009.02316.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. http://dx.doi.org/10.1016/S0896-6273(00)80526-7.

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. http://dx.doi.org/10.1523/JNEUROSCI.3128-13.2013.

Hacker, M. J., & Ratcliff, R. (1979). A revised table of d' for M-alternative forced choice. *Perception & Psychophysics*, *26*(2), 168–170. http://dx.doi.org/10.3758/BF03208311.

Von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science (New York, N.Y.)*, *224*(4654), 1260–1262. http://dx.doi.org/10.1126/science.6539501.

Hubel, D. H., & Wiesel, T. N. (1974). Sequence regularity and geometry of orientation columns in the monkey striate cortex. *The Journal of Comparative Neurology*, *158*(3), 267–293. http://dx.doi.org/10.1002/cne.901580304.

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

Kastner, S., Nothdurft, H. C., & Pigarev, I. N. (1997). Neuronal correlates of pop-out in cat striate cortex. *Vision Research*, *37*(4), 371–376. http://dx.doi.org/10.1016/S0042-6989(96)00184-8.

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

Kim, T., Bair, W., & Pasupathy, A. (2019). Neural coding for shape and texture in macaque area V4. *The Journal of Neuroscience*, *39*(24), 4760–4774. http://dx.doi.org/10.1523/JNEUROSCI.3073-18.2019.

Kingdom, F. A. A., Prins, N., & Hayes, A. (2003). mechanism independence for texture-modulation detection is consistent with a filter-rectify-filter mechanism. *Visual Neuroscience*, *20*(1), 65–76. http://dx.doi.org/10.1017/s0952523803201073.

Knierim, J. J., & van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology*, *67*(4), 961–980. http://dx.doi.org/10.1152/jn.1992.67.4.961.

Kourtzi, Z., & Huberle, E. (2005). Spatiotemporal characteristics of form analysis in the human visual cortex revealed by rapid event-related fMRI adaptation. *NeuroImage*, *28*(2), 440–452. http://dx.doi.org/10.1016/j.neuroimage.2005.06.017.

Kourtzi, Z., & Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *The Journal of Neuroscience*, *20*(9), 3310–3318. http://dx.doi.org/10.1523/JNEUROSCI.20-09-03310.2000.

Kubovy, M., & Cohen, D. J. (2001). What boundaries tell us about binding. *Trends in Cognitive Sciences*, *5*(3), 93–95. http://dx.doi.org/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. http://dx.doi.org/10.3758/BF03212326.

Lamme, V. A. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *The Journal of Neuroscience*, *15*(2), 1605–1615. http://dx.doi.org/10.1523/JNEUROSCI.15-02-01605.1995.

Lamme, V. A., Dijk, B. W. V., & Spekreijse, H. (1992). Texture segregation is processed by primary visual cortex in man and monkey. Evidence from VEP experiments. *Vision Research*, *32*(5), 797–807. http://dx.doi.org/10.1016/0042-6989(92)90022-B.

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

Lee, T. S., Mumford, D., Romero, R., & Lamme, V. A. (1998). The role of the primary visual cortex in higher level vision. *Vision Research*, *38*(15), 2429–2454. http://dx.doi.org/10.1016/S0042-6989(97)00464-1.

Lieber, J. D., Oleskiw, T. D., Palmieri, L., Simoncelli, E. P., & Movshon, J. A. (2024). Responses of neurons in macaque V4 to object and texture images. http://dx.doi.org/10.1101/2024.02.20.581273, BioRxiv 2024–2.

Lin, L. M., & Wilson, H. R. (1996). Fourier and non-Fourier pattern discrimination compared. *Vision Research*, *36*(13), 1907–1918. http://dx.doi.org/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), 4.1–9. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/10.1016/j.visres.2010.11.005.

MacMillan, N. A., & Creelman, C. D. (2005). *Detection Theory* (2nd ed.). Lawrence Erlbaum Inc.

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

Maffei, L., & Fiorentini, A. (1973). The visual cortex as a spatial frequency analyser. *Vision Research*, *13*(7), 1255–1267. http://dx.doi.org/10.1016/0042-6989(73)90201-0.

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

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. http://dx.doi.org/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. http://dx.doi.org/10.1016/j.visres.2004.04.002.

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. http://dx.doi.org/10.1364/JOSAA.18.002209.

Nothdurft, H. (1991). Texture segmentation and pop-out from orientation contrast. *Vision Research*, *31*(6), 1073–1078. http://dx.doi.org/10.1016/0042-6989(91)90211-M.

Nothdurft, H. (2000). Salience from feature contrast: Additivity across dimensions. *Vision Research*, *40*(10–12), 1183–1201. http://dx.doi.org/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. http://dx.doi.org/10.1023/A:1011139631724.

Pasupathy, A., & Connor, C. E. (2001). Shape representation in area V4: Position-specific tuning for boundary conformation. *Journal of Neurophysiology*, *86*(5), 2505–2519. http://dx.doi.org/10.1152/jn.2001.86.5.2505.

Pasupathy, A., Kim, T., & Popovkina, D. V. (2019). Object shape and surface properties are jointly encoded in mid-level ventral visual cortex. *Current Opinion in Neurobiology*, *58*, 199–208. http://dx.doi.org/10.1016/j.conb.2019.09.009.

Peirce, J. W. (2008). Generating stimuli for neuroscience using psychopy. *Frontiers in Neuroinformatics*, *2*, 10. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/10.1037/0096-1523.34.1.1.

Peterhans, E., & Von der Heydt, R. (1991). Subjective contours–bridging the gap between psychophysics and physiology. *Trends in Neuroscience*, *14*(3), 112–117. http://dx.doi.org/10.1016/0166-2236(91)90072-3.

Prins, N., & Kingdom, F. A. A. (2003). Detection and discrimination of texture modulations defined by orientation, spatial frequency, and contrast. *Journal of the Optical Society of America A*, *20*(3), 401–410. http://dx.doi.org/10.1364/JOSAA.20.000401.

Rivest, J., & Cabanagh, P. (1996). Localizing contours defined by more than one attribute. *Vision Research*, *36*(1), 53–66. http://dx.doi.org/10.1016/0042-6989(95)00056-6.

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. http://dx.doi.org/10.1364/JOSAA.7.001632.

Rust, N. C., & DiCarlo, J. J. (2010). Selectivity and tolerance (“invariance”) both increase as visual information propagates from cortical area V4 to IT. *Journal of Neuroscience*, *30*(39), 12978–12995.

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

Sagi, D., & Julesz, B. (1984). Detection versus discrimination of visual orientation. *Perception*, *13*(5), 619–628. http://dx.doi.org/10.1068/p130619.

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), 29.1–15. http://dx.doi.org/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. http://dx.doi.org/10.1111/j.1467-9280.1994.tb00500.x.

Shishikura, M., Machida, I., Tamura, H., & Sakai, K. (2025). Local contour features contribute to figure-ground segregation in monkey V4 neural populations and human perception. *Neural Networks*, *181*, Article 106821.

Snowden, R. J. (1992). Orientation bandwidth: The effect of spatial and temporal frequency. *Vision Research*, *32*(10), 1965–1974. http://dx.doi.org/10.1016/0042-6989(92)90056-O.

Straube, S., & Fahle, M. (2010). The electrophysiological correlate of saliency: Evidence from a figure-detection task. *Brain Research*, *1307*, 89–102. http://dx.doi.org/10.1016/j.brainres.2009.10.043.

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. http://dx.doi.org/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. http://dx.doi.org/10.1016/j.visres.2009.12.013.

Tanner, W. P. (1956). Theory of recognition. *The Journal of the Acoustical Society of America*, *28*(5), 882–888. http://dx.doi.org/10.1121/1.1908504.

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

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

Williford, J. R., & von der Heydt, R. (2016). Figure-ground organization in visual cortex for natural scenes. *ENeuro*, *3*(6), http://dx.doi.org/10.1523/ENEURO.0127-16.2016.

Wilson, H. R., Mcfarlane, D. K., & Phillips, G. C. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research*, *23*(9), 873–882. http://dx.doi.org/10.1016/0042-6989(83)90055-X.

Yazdanbakhsh, A., & Livingstone, M. S. (2006). End stopping in V1 is sensitive to contrast. *Nature Neuroscience*, *9*(5), 697–702. http://dx.doi.org/10.1038/nn1693.

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

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