Higher order image structure enables boundary segmentation in the absence of luminance or contrast cues

Elizabeth Zavitz

McGill Vision Research, Department of Ophthalmology, McGill University, Montreal, Quebec, Canada
Department of Physiology, Monash University, Clayton, Victoria, Australia

Curtis L. Baker, Jr.

McGill Vision Research, Department of Ophthalmology, McGill University, Montreal, Quebec, Canada

Lower order image statistics, which can be described by an image's Fourier energy content, enable segmentation when they are different on either side of a boundary. We have previously demonstrated that the spatial distribution of the energy in an image (described by its higher order statistics or structure) could *influence* segmentation thresholds for contrast- and orientation-defined boundaries, even though it was the same on either side of the boundary and thus task irrelevant (Zavitz & Baker, 2013). Here we examined whether higher order statistics can also *enable* segmentation when boundaries are defined by differences in structure or density of texture elements. We used micropattern-based naturalistic synthetic textures to manipulate the sparseness, global phase alignment, and local phase alignment of carrier textures and measured segmentation thresholds based on forced-choice judgments of boundary orientation. We found that both global phase structure and sparseness, but not local phase alignment, enable segmentation and that local structure also has a small effect on segmentation thresholds in both cases. Simulations of a two-stage filter model with a compressive intermediate nonlinearity can reproduce the major features of the experimental data, segmenting boundaries defined by higher order statistics alone while capturing the influence of global image structure on segmentation thresholds.

Introduction

Boundary segmentation, one of the visual system's most fundamental tasks, is based upon detection of a discontinuity in one or more image properties such as luminance, orientation, or contrast. The lower order image statistics of textures, described by their Fourier amplitude spectrum, are well known to support segmentation when they differ across a boundary. We previously demonstrated that boundaries defined by differences in contrast or orientation of texture elements were easier to segment when the constituent textures were phase scrambled (Zavitz & Baker, 2013). Thus changes in the spatial distribution of the energy in a texture image (its higher order structure) can affect segmentation thresholds, even though these changes were the same on either side of the boundary and therefore irrelevant to the task. We also found differences in effects of different types of higher order statistics: Sparseness and global phase structure appeared to influence segmentation, but local phase structure did not. It remains to be seen to what extent higher order texture statistics, either global or local, might also *enable* segmentation when they vary on either side of the boundary. This issue is significant because naturally occurring textures are often broadband and rich in local structural features. Here we test the ability of textural structure to enable segmentation as a function of density of texture elements, as well as the ability of element density itself to enable segmentation for different kinds of local structure.

It is clear that scene and object perception rely on the visual system encoding the spatial distribution of energy, or *structure*, of an image, described by its Fourier phase spectrum (Hansen & Hess, 2007). Image statistics may be summarized in terms of the correlational relationships (moments) of image pixels. In this view, lower order image statistics are the mean and autocorrelation function (i.e., correlation between a

Citation: Zavitz, E., & Baker, C. L. Jr. (2014). Higher order image structure enables boundary segmentation in the absence of luminance or contrast cues. *Journal of Vision*, *14*(4):14, 1–14, http://www.journalofvision.org/content/14/4/14, doi:10.1167/14.4.14.

doi: 10.1167/14.4.14 Received October 13, 2013; published April 24, 2014 ISSN 1534-7362

given pixel intensity and one of its neighbors), or equivalently, Fourier amplitude spectrum: luminance and contrast energy at each orientation and spatial frequency component. Higher order statistics are described by higher order autocorrelation functions, i.e., how much a given pixel intensity is correlated with two or more neighboring pixels. The higher order statistics differ fundamentally from the lower order in that they are embodied in phase components of the Fourier transform of an image (Thomson & Foster, 1997). Thus higher order statistics capture image structure, or how the image energy is distributed across space. Structural information in a sufficiently broadband image may be removed by randomizing the phase spectrum (Oppenheim & Lim, 1981; Piotrowski & Campbell, 1982). For example, sparseness is a higher order image statistic because it describes how energy in an image is either clustered into locally high- and low-contrast regions (more sparse) or distributed so that contrast across the image is closer to the average (less sparse).

Here we will distinguish global from local structure. In this context, *global structure* refers to the higher order statistics that are derived across the entire image (such as sparseness), while *local structure* refers to characteristics of smaller regions of the image that are assessed independently of the average (such as local phase alignment). Textures composed of micropatterns are ideal for highlighting this distinction, because the micropatterns carry local structure, but their arrangement results in the global structure. This resembles the approach taken by Julesz (1981b) to show that not only the overall properties of an entire texture image but also those of the "textons" it contained could contribute to texture segmentation.

The set of texture statistics that can enable segmentation has previously been a subject of much research. Differences in dipole statistics (Julesz, 1962) were once considered the primary texture property that enabled segmentation, but later studies indicated that local differences, such as closure and corners that were not reflected in the dipole statistics, could also suffice (Julesz, 1981a, 1981b; Olson & Attneave, 1970). Bergen and Adelson (1988) demonstrated that differences in textures' overall Fourier energy enable segmentation and that the effects of many of the local texture properties posited in the aforementioned studies could be accounted for within this framework. However Malik and Perona (1990) and Motoyoshi and Kingdom (2007) showed that local contrast polarity can also enable segmentation even though the Fourier energy was uniform. This raises the question of whether other local properties not captured by the Fourier spectrum may also enable segmentation.

Arguably we should use natural texture photographs to begin this investigation to ensure inclusion of any potentially important higher order statistics (Arsenault, Yoonessi, & Baker, 2011). However in pilot tests using boundaries between pairs of natural textures, we observed examples of both improvements and impairments to segmentation thresholds after removing structure by randomizing the phase spectrum, depending on which individual textures were employed. Consequently such results were difficult to interpret because psychophysical performance depended not only on the individual textures but also on the interactions between them. Additionally, natural textures differ on many dimensions (e.g., orientation bandwidth, dominant orientation, structure), further complicating experimental interpretations. Instead, here we employ naturalistic synthetic textures that can capture important higher order statistics that affect segmentation performance (Zavitz & Baker, 2013).

Most existing models of human texture segmentation use a *filter-rectify-filter* (*FRF*) architecture: a bank of oriented filters at a range of high spatial frequencies followed by a pointwise nonlinearity and then summation following a second stage of linear low spatial frequency filtering (Figure 1). In this scheme, the first stage filters serve to characterize the fine-scale Fourier energy in the texture, the point-wise nonlinearity prevents cancellation between peaks and valleys of signals from the early filters, and the second filtering stage detects the coarse-scale boundary. These models have been shown to account for a wide range of texture segmentation phenomena (Landy & Graham, 2004). Such models are usually envisaged as segmenting regions of relatively homogeneous textures, as is the case in this study. However, because of the small scale of the first-stage filters relative to the scale of the boundary, a model with a sufficiently broadband or high spatial frequency second-stage filter can also segment boundaries defined only by differences in local contrast across a boundary even if the regions of texture themselves have the same average energy content (e.g., Nothdurft, 1991).

Here we will also investigate to what extent this model might account for human psychophysical performance on segmentation of structure boundaries—i.e., boundaries defined by differences in the spatial structure of the energy in adjacent textures rather than their amplitude spectra. As conventionally proposed (e.g., Adelson & Bergen, 1985; Malik & Perona, 1990), the FRF model has a square-law pointwise nonlinearity. Consequently when the responses of an early band-pass filter are averaged over space (as they are by the second-stage filter) the model measures differences in the energy present in the texture within that band. This measurement reflects differences in the Fourier amplitude spectrum (lower order statistics) but not those in the phase spectrum (higher

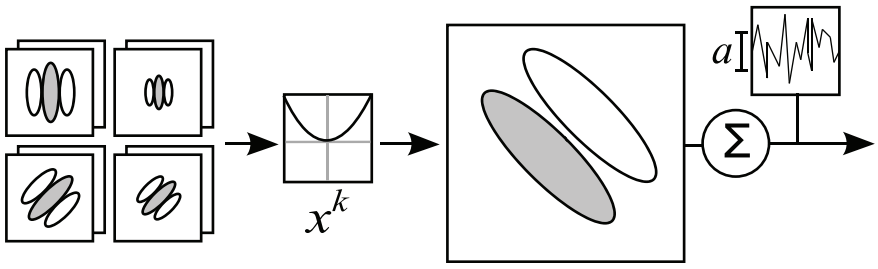

Figure 1. Schematic of a filter-rectify-filter style model. First-stage filters (at left) are applied in a range of orientations, spatial frequencies, and phases to characterize fine texture detail. The outputs of these filters are then subjected to a rectifying nonlinearity. A coarser scale second-stage filter reveals the modulation over the texture (e.g., in orientation or contrast). The shape of the nonlinearity can be parameterized using a power law *k*, and the amount of decision noise can be parameterized by the standard deviation of an additive noise distribution *a*.

order statistics). Recently we have demonstrated (Zavitz & Baker, 2013) that if the intermediate nonlinearity has a compressive shape, this type of model can also account for the influence of higher order statistics on contrast- and orientation-defined segmentation.

Here we will evaluate this model with boundaries defined by higher order texture statistics to assess its ability to capture the extent to which structure and sparseness can both influence and enable segmentation of boundaries defined by differences in texture structure. We will examine the effect of structure in general by completely randomizing textures' phase spectra, but we will also look at specific types of structure. We examine the effects of sparseness, a perceptually important global higher order statistic that reflects the distribution of energy in a texture, by manipulating the number of micropatterns placed in a fixed area, i.e., the texture density. Sparse textures are characterized by larger amounts of contrast energy concentrated in fewer locations, whereas dense textures have energy distributed more uniformly across image locations. We will also look at the effects of local phase-aligned edges by comparing performance for textures whose micropatterns consist of phase-aligned edges and those whose edges have the same bandwidth but whose phases are not aligned.

General methods

Stimuli

To create the stimuli for these experiments, we used pairs of texture patterns with identical Fourier amplitude spectra but with some higher order statistical difference between them, such as structure or density of texture elements, quilted together and windowed by a circular aperture to form a disc composed of two textured halves with a left- or right-oblique boundary between them. The details of this procedure are described in Zavitz and Baker (2013).

Textures

Synthetic micropattern textures were designed to capture attributes of natural textures such as a $1/f$ amplitude spectrum, sparseness, and local edge structure, while allowing us to parametrically control these and other image statistics. To create such a texture, a number of micropatterns in a range of sizes and orientations were randomly positioned on an oversize image without constraints on overlap. With this basic formulation we created three different types of texture that differ in structure.

Intact (INT) textures used broadband "edgelet" micropatterns consisting of phase-aligned sine waves producing a step edge, windowed by a Gaussian function (Figure 2A, top). Locally scrambled (LS) textures were very similar, but the sine-wave components were phase randomized before windowing to eliminate local edge structure (Figure 2A, bottom). To create the globally scrambled (GS) condition, the phase spectrum of an INT texture was replaced with the phase spectrum of white noise, removing all phase structure from the texture (Dakin, Hess, Ledgeway, & Achtman, 2002). In the INT and LS conditions, density was varied by changing the total number (595, 1,530, or 2,975) of micropatterns used to create the texture.

Textures were generated on a trial-by-trial basis and subjected to the same homogeneity constraints as described previously (Arsenault, Yoonessi, & Baker, 2011) to preclude spurious luminance or contrast boundaries caused by unevenly distributed micropattern placements. In the low-density condition, only about 12% of the generated textures passed this test; in

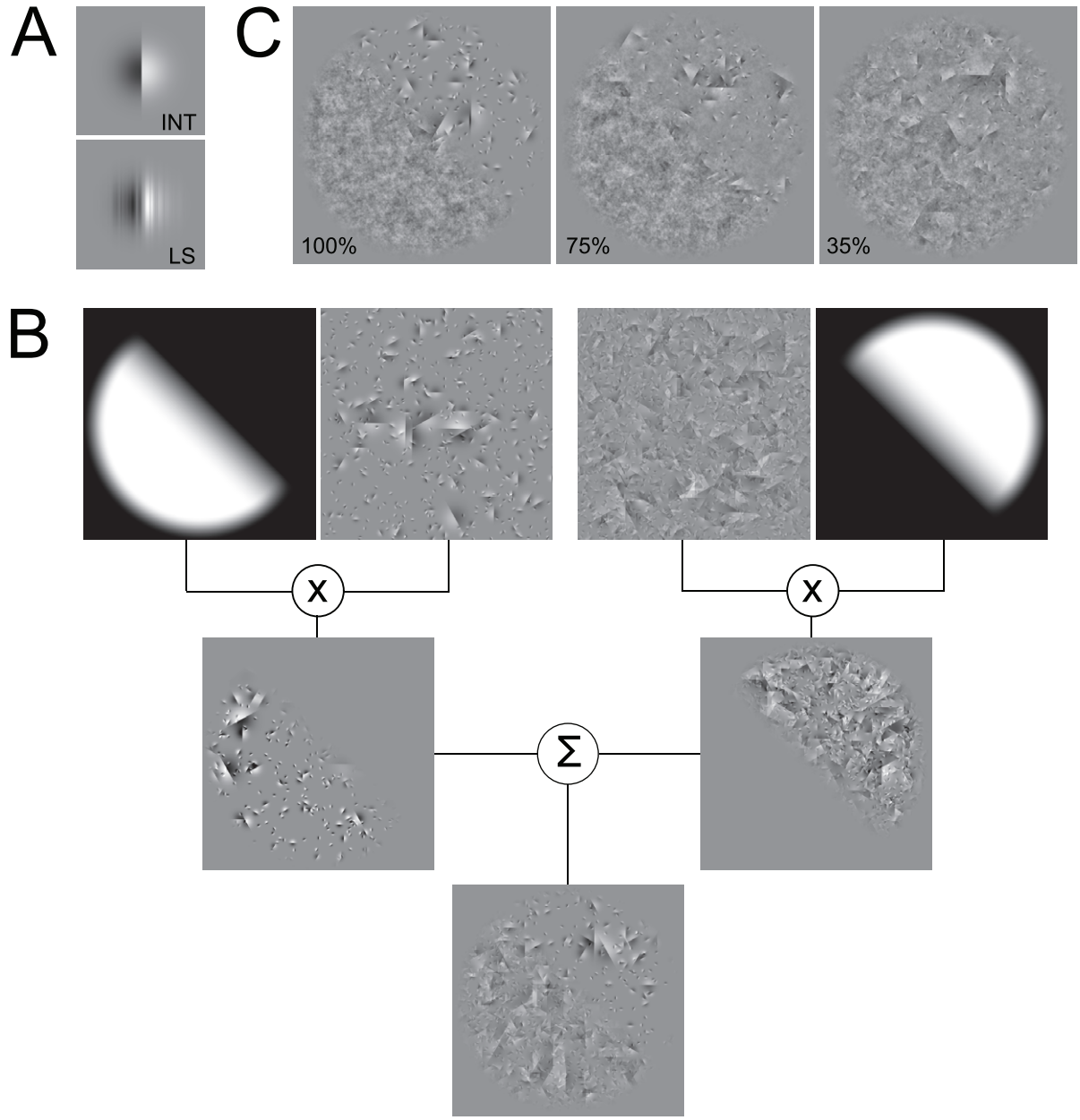

Figure 2. Construction of micropattern texture boundary stimuli. (A) Micropatterns used in texture creation. Top: Intact micropattern, bottom: one instance of a locally phase-scrambled micropattern. (B) Quilting method for creation of texture boundaries. Carrier textures are windowed and modulated separately, then these modulated halves are then combined, in this case, to produce a density-modulated boundary. The modulation depth of the stimulus shown is 100%. (C) Effect of modulation depth (*m*) on task difficulty. This figure depicts an INT-GS texture at 100%, 75%, and 35% modulation depths.

the medium density condition, about 47% of textures passed; and in the high-density condition, about 76% of textures passed. Each texture was scaled to have a mean value of zero, and its extreme luminance values were clipped at ±3 standard deviations and scaled to fit in the range of intensities between ±1.0. Textures were scaled to equalize root-mean-square (RMS) contrast prior to the boundary creation procedure.

Boundary creation

Boundaries were created between textures using a quilting method (Landy & Oruç, 2002; Watson & Eckert, 1994) illustrated in Figure 2B. After two carrier textures ($\boldsymbol{C}_A$ and $\boldsymbol{C}_B$) were created, two complementary envelope patterns were generated to modulate the contrast of each of them in a half-disc pattern. A half-disc envelope function with a cosine taper at the boundary ($\boldsymbol{E}_{x,y}$) is

scaled to create modulators ($\boldsymbol{E}_A$ and $\boldsymbol{E}_B$):

$$\boldsymbol{E}_A = \boldsymbol{W}_{x,y}\sqrt{(1 + m\boldsymbol{E}_{x,y})/2} \quad (1)$$

$$\boldsymbol{E}_B = \boldsymbol{W}_{x,y}\sqrt{(1 - m\boldsymbol{E}_{x,y})/2}. \quad (2)$$

The window ($\boldsymbol{W}_{x,y}$) is a cosine-tapered circular aperture. The products of the window and each of the envelopes is a tapered disc bisected by an oblique boundary. The modulation depth parameter (m) determines the relative strength of a texture between the two halves (Figure 2C). With a modulation depth of 100% ($m = 1$), each texture is entirely confined to one half, but as the modulation depth is decreased, both textures are increasingly present in both halves, but with the two carrier textures weighted reciprocally in each half. When the modulation depth is 0%, the stimulus is a homogeneous blend of both textures (i.e., there is no boundary).

The carrier patterns are scaled to the specified carrier contrast with scaling factor c. The means of the carriers are adjusted so that the final stimulus will be luminance-balanced after the envelopes (Equations 1 & 2) have been applied.

$$\boldsymbol{C}'_A = c\boldsymbol{C}_A - \frac{\int\int c\boldsymbol{C}_A\boldsymbol{E}_A - 0.5}{\int\int \boldsymbol{E}_A} \quad (3)$$

$$\boldsymbol{C}'_B = c\boldsymbol{C}_B - \frac{\int\int c\boldsymbol{C}_B\boldsymbol{E}_B - 0.5}{\int\int \boldsymbol{E}_B}. \quad (4)$$

The final stimulus ($\boldsymbol{S}_{x,y}$) is the sum of the two carrier components, each spatially weighted by their respective envelopes, where $\boldsymbol{L}_o$ is the mean luminance:

$$\boldsymbol{S}_{x,y} = \boldsymbol{L}_o\{1 + \boldsymbol{C}'_A\boldsymbol{E}_A + \boldsymbol{C}'_B\boldsymbol{E}_B\}. \quad (5)$$

Apparatus and observers

The stimuli were presented on a CRT monitor (Sony Trinitron Multiscan G400, 81 cd/m^2, 75 Hz, 1024 × 768 pixels), gamma linearized with a digital video processor (Bits++, Cambridge Research Systems) for greater bit depth at low contrasts. Stimulus patterns appeared in a central 480 × 480 pixel patch on a mean gray background. Observers viewed the stimuli from a distance of 114 cm, resulting in a stimulus visual angle of approximately 6.5°. The experiment was run on a Macintosh (Desktop Pro, MacOSX) using Matlab and PsychToolbox (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). The experiment was performed under the approval of the McGill University Ethics Committee and conformed to the Helsinki Declaration for experiments with human subjects.

Task

Before beginning a block of trials, observers were informed as to what texture property would define the boundary. Within a block of trials, boundary and structure type were kept constant. At the beginning of a trial, observers fixated a mark at the center of the screen and initiated the trial with a button press. The stimulus was displayed for 100 ms, and the observer indicated with another button press whether the boundary between the textures was left or right oblique. The RMS contrast of the stimulus was 14%, which was well above threshold for all observers. Between stimulus presentations the screen was maintained at the gray level of the mean luminance. No feedback was provided, but all observers were given an opportunity to practice until they were comfortable with the stimuli and task.

We tested observers on five modulation depth levels, logarithmically spaced between 100% and 25%, with additional levels below 25% if necessary to reach chance performance. Observers were tested in blocks of 100 trials, with 20 trials per level. At least three of these blocks were run, staggered between different conditions, for a total of at least 300 trials per condition.

Data analysis

Percent-correct data were fit with a logistic psychometric function, and a threshold was interpolated at the 75% correct point. Prism (GraphPad Software, Inc.) was used for curve fitting and standard-error bootstrapping.

To test for significance we used two-way analyses of variance (ANOVAs) or paired-samples t tests with a criterion of $\alpha = 0.05$. Effect size (Klein, 2005) was measured using Cohen's d with the standardizer computed as:

$$s = \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{2}, \quad (6)$$

where σ_1 and σ_2 are the sample standard deviations of the compared conditions.

Experiment 1: Structure boundary segmentation

We have recently demonstrated that contrast and orientation boundary segmentation is influenced by the

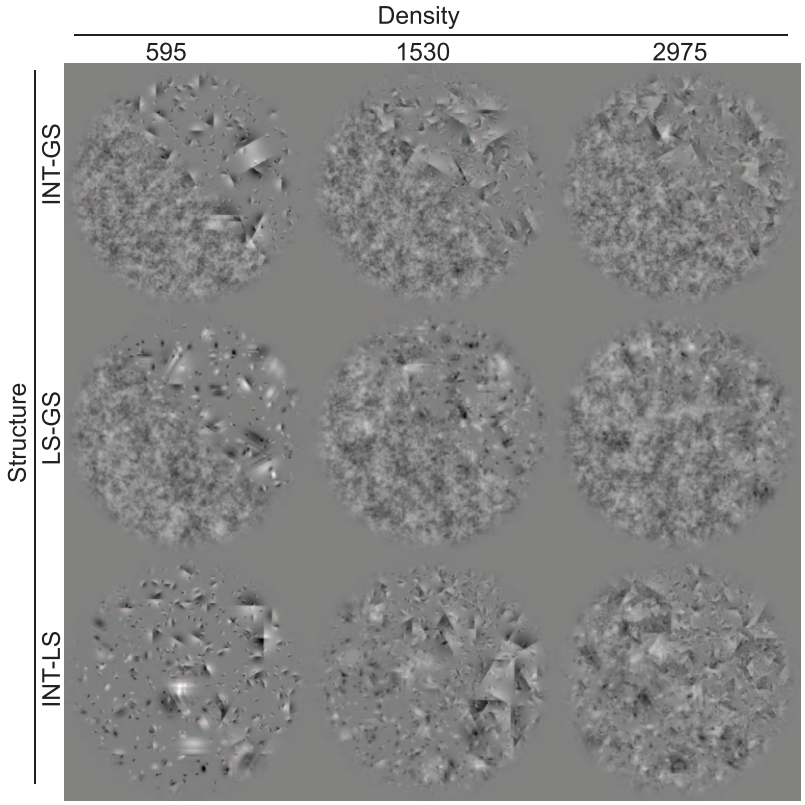

Figure 3. Examples of structure modulated stimuli used for Experiment 1, shown with left-oblique boundaries at a modulation depth of 100%. The boundary types are organized into rows, while the density conditions are organized into columns. INT-GS boundaries are a combination of intact and globally scrambled textures, LS-GS boundaries result from the combination of locally scrambled and globally scrambled textures, and INT-LS boundaries are created by pairing intact and locally scrambled textures.

presence of higher order image statistics (Zavitz & Baker, 2013). Here, we test whether differences in higher order statistics alone can enable segmentation and also examine the influence of sparseness and local phase structure on segmentation thresholds.

Methods

In this experiment we measured modulation depth thresholds for observers segmenting boundaries defined by differences in phase structure. We tested all pairwise combinations of three phase structure conditions (Figure 3): intact (INT) carriers using edgelet micropatterns consisting of a phase-aligned broadband edge in a Gaussian window; globally scrambled (GS) texture carriers generated by phase-scrambling intact textures; and locally scrambled (LS) textures using the same broadband micropatterns but with a phase-scrambled edge (Figure 2A). Density was varied parametrically for the INT and LS textures by changing the number of micropatterns being randomly placed on the texture canvas (595, 1,530, or 2,975). Pairs of such textures were quilted as described in the General methods and illustrated in Figure 2B to create a unique stimulus for each trial. We created boundaries testing all combinations of the phase-alignment conditions for three structure boundary conditions: INT-LS, INT-GS, and LS-GS (rows of Figure 3), and tested each at three density levels (columns of Figure 3). All observers had normal or corrected-to-normal vision, and JH, JB, and AR were naïve to the purpose of the experiment.

Results

Structure boundary segmentation results are shown in Figure 4. While this was a challenging task, observers experienced no difficulty segmenting the boundaries at the greatest modulation depths, with the exception of the INT-LS condition for two of the observers (upper right and lower left panels of Figure 4, open circles).

Thresholds for boundaries between textures with and without global structure (INT-GS and LS-GS, open triangles and filled circles in Figure 4) were considerably better. In both conditions the thresholds increased with density. Thresholds in these conditions were

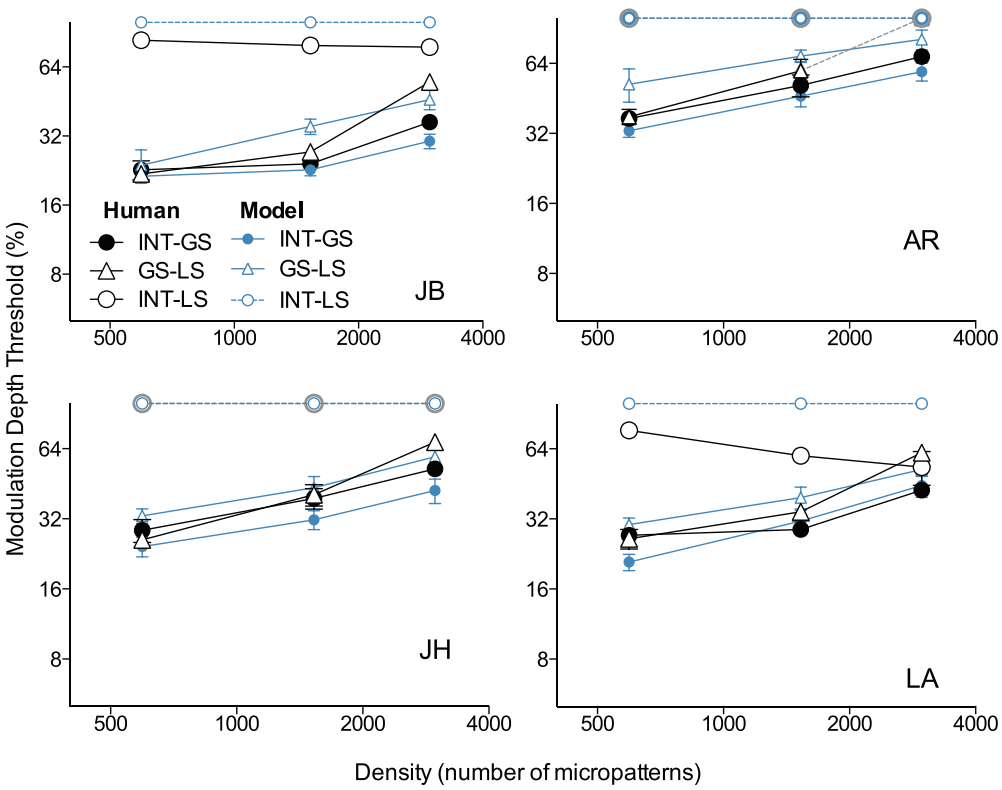

Figure 4. Structure boundary segmentation results for individual observers. Error bars indicate standard error. The boundary conditions are: INT-GS, between an intact texture and a phase-scrambled texture (filled circles); LS-GS, between a locally scrambled and globally scrambled texture (open triangles); and INT-LS, between an intact texture and a locally scrambled texture (open circles). The gray symbols for observers AR and JH indicate above-chance performance only at a modulation depth of 100%, so no threshold could be estimated. These results show elevated thresholds for the INT-LS condition that display little or no dependence on density. In this condition, human performance was slightly above chance and good enough to fit thresholds in some cases (LA and JB). Segmentation performance on the INT-GS and LS-GS conditions is relatively impaired with increasing density, with very similar thresholds at low densities, but LS-GS exhibiting more difficulty at high densities. Blue symbols and lines indicate predictions of a two-stage model.

almost identical at the lowest density, but the thresholds for LS-GS boundaries increased more quickly with density than those in the INT-GS condition. A two-way ANOVA was run on only the LS-GS and INT-GS conditions, because thresholds could not be reliably estimated for all observers and conditions in the INT-LS condition. This confirmed a main effect of structure, $F(1, 9) = 41.01$, $p < 0.05$, and a main effect of density, $F(2, 9) = 6.67$, $p < 0.05$, as well as a significant interaction between structure and density, $F(2, 9) = 26.87$, $p < 0.05$. Post-hoc Bonferroni tests (Table 1) demonstrated that the significant difference between the INT-GS and LS-GS conditions was at the highest density only ($t = 9.52$, $p < 0.05$, $d = 1.24$).

Density	*t*	*p*	*d*
INT-GS to LS-GS			
595	0.429	> 0.05	0.30
1,530	2.005	> 0.05	0.69
2,975*	9.516	< 0.05	2.46

Table 1. Results from Bonferroni post-hoc tests (*t*, *p*) and effect sizes (*d*) for the data presented in Experiment 1. Statistically significant differences ($p < 0.05$) are indicated with an asterisk.

These results demonstrate that differences in higher order statistics, particularly in global phase structure, can enable segmentation. Globally scrambled (GS) textures have particularly low sparseness compared to the locally scrambled (LS) and intact (INT) textures with which they have been paired. From inspection it seems likely that an important factor enabling segmentation in these stimuli is a difference in sparseness on either side of the boundary. These results also suggest that for highly sparse textures (i.e., INT and LS at low density), differences in local phase structure do not influence segmentation, as we did not find a difference between thresholds in the INT-GS and LS-GS conditions. At high densities the presence of local structure can facilitate segmentation, as evidenced by lower thresholds for the INT-GS condition relative to

the LS-GS condition. However, differences in local phase structure alone (i.e., INT-LS) appear to provide poor support for segmentation. It is possible that locally scrambled micropatterns may not have the same sparseness as the intact micropatterns, because their energy is distributed evenly over their area instead of being concentrated at a single edge. These relatively subtle variations in sparseness may have impacted our results, and we further consider this idea in the Discussion.

Experiment 2: Density boundary segmentation

In Experiment 1 the task was more difficult at higher densities for boundaries between the intact or locally scrambled textures and the globally scrambled textures, suggesting that the difference in sparseness across the boundary might have been an important segmentation cue.

While density has been shown to be a salient cue for segmentation (e.g., Wilkinson & Lessard, 1995), previous demonstrations of this effect have been confounded by concomitant luminance gradients. Here we test directly whether observers can segment boundaries defined solely by a difference in density of micropatterns, i.e., with luminance and RMS contrast equated across the boundary. The experiment also examines the influence of local structure on density segmentation.

Methods

We created density boundaries for two types of textures, intact (INT) and locally scrambled (LS). For each type, density boundaries were formed by quilting a low-density texture (595 micropatterns) with a high density texture (2,975 micropatterns) of the same type, as shown in the left and middle stimuli of Figure 5. Globally scrambled textures have a maximal sparseness regardless of the original micropattern density, so that no boundary is formed (Figure 5, right-most stimulus), and hence this condition was not considered further. The same four observers were tested as in Experiment 1.

Results

Density boundary segmentation results are shown in Figure 6. All four observers were able to segment these boundaries, with well-defined thresholds for all observers and conditions. We found that density boundaries are approximately as difficult to segment as the

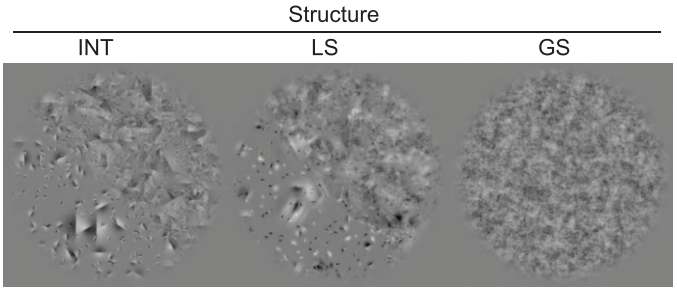

Figure 5. Examples of density boundary stimuli used in Experiment 2, shown at a modulation depth of 100%. Each stimulus is a modulation between a texture with 595 micropatterns and one with 2,975 micropatterns, either intact structured edgelets (INT) or locally phase-scrambled edgelets (LS). In the globally scrambled (GS) condition each texture was phase scrambled before being combined. Because no visible boundary was formed in the GS condition, it was not tested.

INT-GS and LS-GS structure boundaries in the previous experiment. The thresholds for the locally scrambled (LS) condition were lower than those for the intact (INT) condition by a small but systematic difference in each of the observers. A two-tailed paired-samples t test finds this difference statistically significant $t(3) = 8.01$, $p < 0.05$, but the effect size ($d = 0.96$) is modest. It appears that the density boundary is slightly easier to segment in locally scrambled textures.

These results demonstrate that differences in micropattern texture density can support segmentation without any other differences in global phase structure and support the idea that density modulation was a very important contributor to performance for the structure boundaries in Experiment 1. We found a small but consistent improvement in density-based segmentation for locally scrambled textures compared to intact textures. This was unexpected, because in the previous experiment the performance difference due to local structure showed impairment, rather than improvement, for locally scrambled textures.

Model

Results from the above psychophysical experiments were compared to those obtained by simulating identical experiments with a filter-rectify-filter style model acting as the observer (Figure 1), as described in detail in Zavitz and Baker (2013).

Methods

In the first filtering stage, a given stimulus image was convolved with a bank of log-Gabor filters (Kovesi,

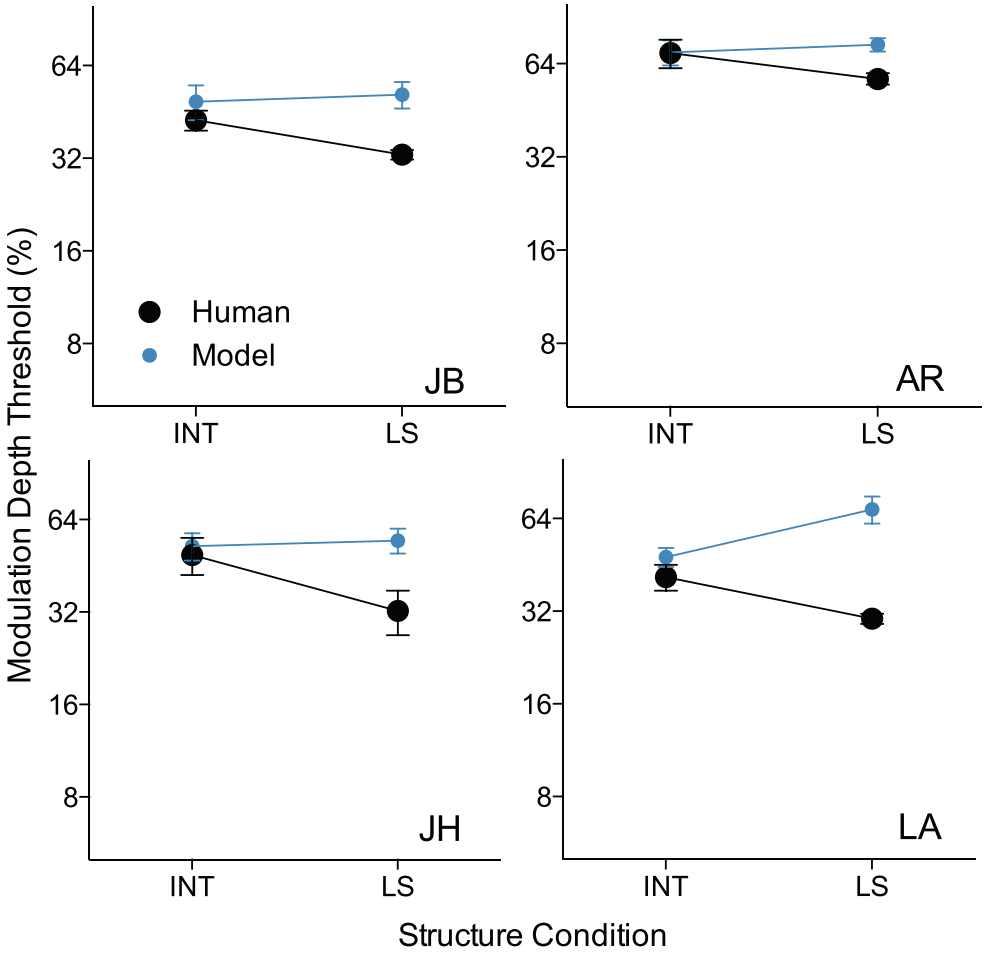

Figure 6. Results for individual observers in density segmentation task. Overall levels of performance were similar to those for the structure boundary stimuli in the first experiment. All observers show a slight, but consistent, decrease in thresholds for the locally scrambled (LS, light bars) condition, relative to the intact condition (INT, shaded bars). Model predictions shown in blue. Error bars indicate standard error.

2000) at two phases (even and odd), six orientations (evenly spaced with bandwidths chosen for approximately uniform coverage), and four spatial frequencies (160, 80, 40, and 20 cycles per image, spanning a range equivalent to approximately 24.6 to 3.1 cycles/° for human observers, each with a bandwidth of approximately 1.5 octaves to provide coverage in frequency space). The outputs of these filters were then full-wave rectified and raised to a power k. We used $k = 0.5$, in accordance with our previous study (Zavitz & Baker, 2013) demonstrating that compressive intermediate nonlinearities best capture the influence of texture structure on segmentation performance. These signals were summed to create orientation and spatial frequency band-pass responses, which were convolved with each of a pair of second stage filters: low spatial frequency, odd-phase Gabors that were matched to the possible boundary orientations (left and right oblique). These responses were integrated over their area, orientation, and spatial frequency to produce a scalar response, raised to a power $1/k$, and subjected to additive noise drawn from a distribution with standard deviation a to produce a pair of labeled outputs. The boundary direction decision was determined according to the larger of the outputs: left or right oblique.

Here we used the same model power law exponent (k) that best-predicted human performance for orientation and contrast boundary segmentation data in the previous paper (Zavitz & Baker, 2013) and fit the noise parameter (a) to minimize the sum of squares error for each observer. We will address the effect of different values of k in the Discussion.

Results

Figure 4 illustrates the model predictions (blue) of the data for individual observers (black) in the structure segmentation task. In the INT-LS condition (open circles), where the only difference between

textures is in their local structure, the model performs at chance, which is in line with the poor human performance observed in the same condition. In the INT-GS condition (filled circles), for both human and model observers, performance is best when the INT texture is at low densities and thresholds increase with density. Results in the LS-GS condition (open triangles) are similar but with a slightly greater impairment in performance at high densities. The model correctly predicts that LS-GS texture boundaries would be harder to segment than INT-GS boundaries at high densities.

The model predictions and average human data in the density segmentation experiment are shown in Figure 6. The model predicts the overall level of performance reasonably well but predicts no difference between the intact (INT) and locally scrambled (LS) textures, while the human observers have consistently lower thresholds in the LS condition.

Discussion

In the first experiment, we observed that differences in global phase structure (INT-GS and LS-GS), but not local phase alignments (INT-LS), were sufficient to enable segmentation robustly. We also found that segmentation thresholds for boundaries between locally scrambled and globally scrambled (LS-GS) textures increased more quickly than in the intact and globally scrambled (INT-GS) condition as density was increased. In the second experiment, we demonstrated that a density difference alone could enable segmentation and, in addition, that randomizing local phase alignments (LS) led to slightly better performance. A filter-rectify-filter style model with a compressive intermediate nonlinearity was capable of segmenting boundaries defined by structure or sparseness and of predicting the influence of structural sparseness, though not that of local structure. Further, we demonstrated that the same parameter values that fit orientation and contrast boundary segmentation in our previous study also predicted most aspects of performance on boundaries defined by higher order structure, such as the effect of global structure.

Psychophysics

For the structure segmentation condition in which the boundary consisted only of a difference in local structure (INT-LS), all of our observers were able to segment the boundaries at above chance performance when the modulation depth was 100%, but performance was not sufficient to provide a well-defined psychometric function and in some cases did not reach the threshold level (75% correct) at all. The boundaries in the INT-LS condition are evident by inspection (Figure 3, bottom row) and therefore probably could be segmented more reliably with a longer viewing time. However, a segmentation-by-inspection approach might be due to some other (potentially top-down) mechanism. The varying ability of observers to segment the INT-LS boundaries might be due to practice effects. The most practiced observers were able to segment this condition at the highest modulation depths (100%) at the low densities and at slightly lower modulation depths at higher densities. Less practiced observers struggled with this condition at all densities, and performance was at chance when the modulation depth was decreased from 100%.

A major consequence of phase scrambling is a reduction in sparseness. Consequently in the INT-GS and LS-GS conditions, the difference in sparseness between the two halves was large (i.e., a sparse INT or LS texture paired with a GS texture), and segmentation thresholds were lower than when the density difference between these halves was smaller. Given this relationship, it would make sense that performance degrades as the density of the INT or LS texture is increased, since the difference in sparseness would thereby be decreased. In Experiment 2 we went on to demonstrate that density differences alone could indeed enable segmentation.

The influence of local structure, however, seems more difficult to understand. Orientation and contrast boundary segmentation performance is the same for the intact and locally scrambled conditions at a given density (Zavitz & Baker, 2013), which suggests that local phase alignment is not encoded by the mechanisms that segment these boundaries. Yet in Experiment 1, performance is the same in the INT-GS and LS-GS conditions at low densities but different at the highest density tested. Furthermore, in the density segmentation task, we see a small but consistent performance *impairment* in the INT condition. Because phase scrambling reduces sparseness even on a small scale, LS micropatterns may be effectively slightly less sparse than INT micropatterns. It may even be the case that because this difference in sparseness arises from the micropatterns themselves, the more micropatterns are added (as in the high density conditions), the greater the difference in sparseness between INT and LS textures having the same number of micropatterns. If this were the case, in Experiment 1 we would expect to see segmentation occurring in the INT-LS conditions. It does occur for some observers, in whom we see what might be a slight trend toward lower thresholds in higher density conditions as would then be expected. We would also expect that the LS-GS condition would be more difficult than the INT-GS condition at higher

densities, which appears to be the case. Finally, in Experiment 2, the LS condition would be easier if the effective sparseness difference between its constituent halves is greater than implied simply by the difference between their numbers of micropatterns. Thus local differences in sparseness might explain these effects of local structure in human observers, though it is unclear why the model is not sensitive to these differences. One possibility might be that we have used inappropriate spatial frequencies or bandwidths for the model's first-stage filters. It is unlikely that the model is using insufficiently high spatial frequencies to match those of human vision, because the highest spatial frequency filter used in the model (160 cycles per image, about 25 cycles/°) is well into the high frequency roll-off of the human contrast sensitivity function (Campbell & Robson, 1968). There are no gaps in the model's first-stage spatial frequency representation, so within the range of 3–25 cycles/°, no information is lost. However, the optimal tuning characteristics of the first stage filters remain an interesting question—see for example, Westrick and Landy (2013).

Model

We modeled our psychophysical results using the same two-stage filter model as used previously to predict thresholds for orientation and contrast boundary segmentation (Zavitz & Baker, 2013). This model, using the same shape of nonlinearity ($k = 0.5$) as in our previous work, successfully captured the ability of global phase structure to enable segmentation as well as the influence of global structure on those thresholds, though it did not account for the differing effects of local phase structure on either structure or density segmentation thresholds or the (very limited) ability to segment boundaries in the INT-LS condition.

The model's ability to segment boundaries defined by differences in textural structure is due to the non-square-law nonlinearity applied before early stage responses are spatially pooled. The stimuli are RMS balanced, so with a square law the model's segmentation performance does not exceed chance. A power-law exponent greater than two produces a larger response for the high local contrast of the sparse textures, while a value less than two boosts the relatively low local contrasts of dense textures. This differential response depending on the distribution of contrast energy (whether it is spread out and therefore dense, or clumped and sparse) allows the model to segment textures that differ in the kinds of structure tested here. A consequence of translating density differences into energy-like differences is that there should be a contrast decrement one could apply to one of the textures to null the appearance of the density boundary (as it is nulled for the model when $k = 2$). This might not necessarily be something we would expect to observe psychophysically, since the overall percept might be mediated by a population of FRF-like neurons having a distribution of nonlinearities and thus different null points. For example, Mineault, Khawaja, Butts, and Pack (2012) inferred just such a distribution of intermediate nonlinearities between MT and MST.

The filter-rectify-filter model is typically imagined as having a square-law intermediate nonlinearity, corresponding to segmentation based simply on Fourier energy (Heeger, 1992). The only study that has quantitatively evaluated the shape of the nonlinearity, to our knowledge, is that of Graham and Sutter (1998), who found a value greater than two. In our previous investigation using orientation- and contrast-defined boundaries we found that, while both expansive ($k > 2$) and compressive ($k < 1$) nonlinearities could explain the results broadly, a compressive nonlinearity provided a better quantitative fit to the data (Zavitz & Baker, 2013).

We tested the present model with other values of k (Figure 7) and measured its ability to predict both individual and average observer thresholds by measuring sum-of-squares error. We again found that lower values of k tend to produce better model performance. There are two main reasons we might have found an optimal $k < 2$: (a) stimulus contrast and (b) stimulus conditions and modeling approach.

The contrast response functions of early visual neurons typically have a sigmoidal shape (Albrecht & Hamilton, 1982), wherein they are expansive at low contrast and compressive at higher contrasts. We conducted our experiments well above contrast detection thresholds, which could result in responses at the compressive part of contrast response functions. It is possible that the stimuli used by Graham and Sutter (1998) were at sufficiently low contrasts to evoke responses at the expansive part of the relevant neurons' classical receptive field. However, because the contrast measures, stimuli, and tasks employed by Graham and Sutter (1998) were quite different from those used in this study, it is difficult to make a direct comparison between their work and ours.

Another difference from previous studies is the consideration of local effects such as the comparison of intact and locally scrambled texture conditions, which provided a major constraint for the model. All of the power law exponents we tested (except for a perfect square law) were able to account for our primary finding that sparseness impairs segmentation. However, an expansive nonlinearity predicts segmentation thresholds in the INT-LS condition approximately the same as those in the INT-GS or LS-GS conditions, which does not agree with our finding of much higher thresholds for INT-LS.

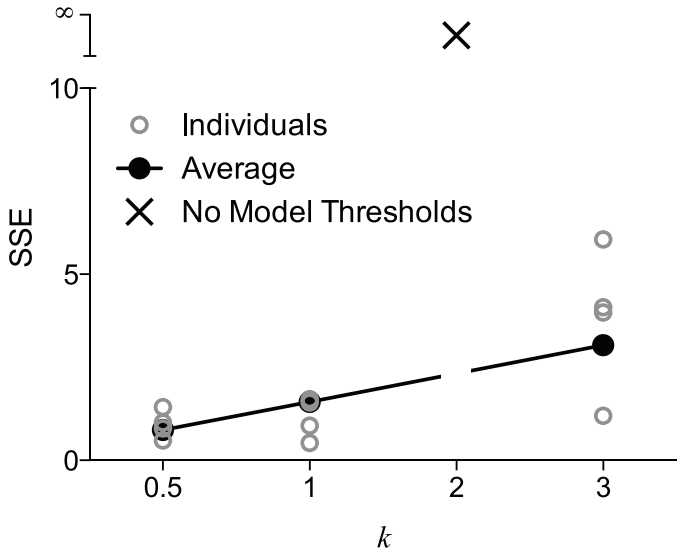

Figure 7. Minimum sum of squares error (SSE) for the model with three values of the power law exponent in the intermediate nonlinearity, compared to the average human data. The SSE was computed using thresholds from both the structure and density tasks (except for the INT-LS condition where thresholds could not be consistently estimated). SSE computed separately on the results from the structure and density tasks follows the same trend. Results are not shown for $k = 2$ (X), because model thresholds could not be obtained in this condition as performance did not exceed chance. Overall, SSE is lowest with a compressive nonlinearity (in this case $k = 0.5$) and gets worse as the nonlinearity becomes more expansive. Minimum SSE for the model compared to individual observer data is indicated with gray open circles at each value of k.

To our knowledge, this is the first time a two-stage filter model has successfully been applied to segmentation of textures that differ only in their phase structure. However many of the attributes of our model are very common in the literature. Filter-rectify-filter models can be thought of as a type of summation model, in this case summing across early stage filters and over the spatial support of the second-stage filter. We chose to use Minkowski summation (Shepard, 1964), which raises the summation arguments to a power with the resulting sum raised to the reciprocal of the same power. Minkowski summation has often been employed as a model of psychophysical channel summation (e.g., Meese & Summers, 2007; To, Baddeley, Troscianko, & Tolhurst, 2010), though in most cases with an exponent of three to four. The compressive nonlinearity that best fits our results is however in agreement with a number of recent studies showing evidence of subadditive summation in visual processing demonstrated in a variety of ways: cascade models with compressive nonlinearities (Mineault, Khawaja, Butts, & Pack, 2012), divisive normalization (Rust, Mante, Simoncelli, & Movshon, 2006), or surround suppression (Tsui, Hunter, Born, & Pack, 2010).

We have found that the same model, with the same parameters, can predict boundary segmentation thresholds for orientation and contrast boundaries defined over synthetic or natural textures very well (Zavitz & Baker, 2013). This model also predicts boundary segmentation thresholds, and the influence of global structure on those thresholds, for boundaries defined by higher order structure.

Sparseness

Sparseness can be described as the extent to which the energy in the image is spatially aggregated into local high-contrast regions, separated by regions relatively empty of local content (Hansen & Hess, 2007). It has been considered an independent texture property because it is a specifically adaptable feature (Durgin, 1995), dissociable from luminance spatial frequency adaptation (Durgin & Huk, 1997). Dakin, Tibber, Greenwood, Kingdom, and Morgan (2011) suggest that perception of sparseness can be measured by the ratio of the modulation depth of higher spatial frequency contrast to the modulation depth of lower spatial frequency contrast, which is essentially a third-order comparison (i.e., requiring an additional stage of processing beyond an FRF-type model). The results of this study indicate that, with an appropriate intermediate summation nonlinearity, density can act as an intensive property that, like contrast, affects the response strength without the necessity of a third processing stage.

With respect to segmentation, Julesz (1981b) suggested that "texton density" was the primary statistical comparison that enabled segmentation. However his formulation has not been reconciled with an energy-based approach to segmentation. His definition of texton density—the spacing between features that match internal texton templates—is unlike the concept of density we used to construct the textures used in this study. In particular, our micropatterns are placed so that they overlap. To Julesz, this would destroy the identity of the texton.

Structure-defined boundary segmentation

Boundaries defined by structure in the absence of differences in amplitude spectrum have seldom been examined. Julesz (1981a) and Victor, Chubb, and Conte (2005) used textures that differed only in their third- or fourth-order spatial correlations, but this meant that the amplitude spectra of these "even" and "odd" textures were the same only when the categories

were taken as ensembles. Any given pair of such textures contained a difference in the orientation bandwidth of the amplitude spectrum, and thus (unlike the stimuli used here) they could be segmented on the basis of lower order statistics (Turner, 1986).

Graham, Sutter, and Venkatesan (1993) used element arrangement patterns that differed in only their higher order statistics, as we have defined the term. However the element sizes were quite large and relatively few were arrayed to form each region, making it questionable to what extent they would be processed by the visual system as "texture" (Wilkinson, 1990). Graham et al. (1993) used a three-stage model (FRFRF) to account for the extra comparison that seemed to be necessary to segment the element patterns. Here we demonstrated that a third stage of the model may not be necessary, and a single presummation compressive nonlinearity can be sufficient for differences in the arrangement of the energy in the texture to enable segmentation.

Conclusions

In these experiments we demonstrated that differences in the structure, rather than simply the spectrum of Fourier energy, in a texture could both enable and influence segmentation performance. We have shown that a two-stage model with a compressive intermediate nonlinearity can predict thresholds for a wide range of stimuli: contrast boundaries defined over natural textures and contrast and orientation boundaries defined using naturalistic textures (Zavitz & Baker, 2013), as well as the structure boundaries in naturalistic textures modeled here. This work also raises interesting questions about the effect of local texture structure, which undeniably influences performance but does so in ways that are not predicted under current models.

Keywords: texture, segmentation, second order, filter rectify filter

Acknowledgments

The authors would like to thank Fred Kingdom and Aaron Johnson for helpful comments and discussions, Ahmad Yoonessi for contributions to software development, and our observers for the generous donation of their time. This research was supported by NSERC grant OPG0001978 to CB.

Commercial relationships: none.
Corresponding author: Elizabeth Zavitz.
Email: elizabeth.zavitz@monash.edu; elizabeth.arsenault@mail.mcgill.ca.
Address: Department of Physiology, Monash University, Clayton, Victoria, Australia.

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A, 2*(2), 284–299.

Albrecht, D. G., & Hamilton, D.B. (1982). Striate cortex of monkey and cat: Contrast response function. *Journal of Neurophysiology, 48,* 217–237.

Arsenault, E., Yoonessi, A., & Baker, C., Jr. (2011). Higher order texture statistics impair contrast boundary segmentation. *Journal of Vision*, *11*(10): 14, 1–15, http://www.journalofvision.org/content/11/10/14, doi:10.1167/11.10.14. [PubMed] [Article]

Bergen, J. R., & Adelson, E. H. (1988). Early vision and texture perception. *Nature*, *333*(6171), 363–364.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *The Journal of Physiology*, *197*(3), 551.

Dakin, S. C., Hess, R. F., Ledgeway, T., & Achtman, R. L. (2002). What causes non-monotonic tuning of fMRI response to noisy images? *Current Biology*, *12*(14), 476.

Dakin, S. C., Tibber, M. S., Greenwood, J. A., Kingdom, F. A. A., & Morgan, M. J. (2011). A common visual metric for approximate number and density. *Proceedings of the National Academy of Sciences, USA, 108*(49), 19552–19557.

Durgin, F. H. (1995). Texture density adaptation and the perceived numerosity and density of texture. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 149–169.

Durgin, F. H., & Huk, A. C. (1997). Texture density aftereffects in the perception of artificial and natural textures. *Vision Research, 37*(23), 3273–3282.

Graham, N., & Sutter, A. (1998). Spatial summation in simple (Fourier) and complex (non-Fourier) texture channels. *Vision Research*, *38*(2), 231–257.

Graham, N., Sutter, A., & Venkatesan, C. (1993). Spatial-frequency-and orientation-selectivity of simple and complex channels in region segregation. *Vision Research*, *33*(14), 1893–1911.

Hansen, B. C., & Hess, R. F. (2007). Structural sparseness and spatial phase alignment in natural scenes. *Journal of the Optical Society of America A*, *24*(7), 1873–1885.

Heeger, D. J. (1992). Half-squaring in responses of cat striate cells. *Visual Neuroscience*, *9*(05), 427–443.

Julesz, B. (1962). Visual pattern discrimination. *IEEE Transactions on Information Theory, 8*(2), 84–92.

Julesz, B. (1981a). Textons, the elements of texture perception, and their interactions. *Nature, 290*(5802), 91–97.

Julesz, B. (1981b). A theory of preattentive texture discrimination based on first-order statistics of textons. *Biological Cybernetics, 41*(2), 131–138.

Klein, D. F. (2005). Beyond significance testing: Reforming data analysis methods in behavioral research. *162*.

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3. *Perception*, *36*(14).

Kline, R. B. (2004). *Beyond significance testing: reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association.

Kovesi, P. D. (2012). *MATLAB and octave functions for computer vision and image processing*. Retrieved from http://www.csse.uwa.edu.au/~pk/research/matlabfns/ January 2012.

Landy, M. S., & Graham, N. (2004). Visual perception of texture. In L. M. Chalupa, J. S. Werner, and C. J. Barnstable, (Eds.), *The visual neuroscience*, (pp. 1106–1118). Cambridge, MA: MIT Press.

Landy, M. S., & Oruç, I. (2002). Properties of second-order spatial frequency channels. *Vision Research*, *42*(19), 2311–2329.

Malik, J., & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A*, *7*(5), 923–932.

Meese, T. S., & Summers, R. J. (2007). Area summation in human vision at and above detection threshold. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1627), 2891–2900.

Mineault, P. J., Khawaja, F. A., Butts, D. A., & Pack, C. C. (2012). Hierarchical processing of complex motion along the primate dorsal visual pathway. *Proceedings of the National Academy of Sciences, USA 109*(16), E972–E980.

Motoyoshi, I., & Kingdom, F. A. A. (2007). Differential roles of contrast polarity reveal two streams of second-order visual processing. *Vision Research*, *47*(15), 2047–2054.

Nothdurft, H. C. (1991). Texture segmentation and pop-out from orientation contrast. *Vision Research*, *31*(6), 1073–1078.

Olson, R. K., & Attneave, F. (1970). What variables produce similarity grouping? *The American Journal of Psychology, 83*(1), 1–21.

Oppenheim, A. V., & Lim, J. S. (1981). The importance of phase in signals. *Proceedings of the IEEE, 69,* 529–541.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Piotrowski, L. N., & Campbell, F. W. (1982). A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase. *Perception*, *11*, 337–346.

Rust, N. C., Mante, V., Simoncelli, E. P., & Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. *Nature Neuroscience*, *9*(11), 1421–1431.

Shepard, R. N. (1964). Attention and the metric structure of stimulus space. *Journal of Mathematical Psychology*, *1,* 54–87.

Thomson, M. G. A., & Foster, G. H. (1997). Role of second- and third-order statistics in the discriminability of natural images. *Journal of the Optical Society of America A, 14,* 2081–2090.

To, M. P. S., Baddeley, R. J., Troscianko, T., & Tolhurst, D. J. (2010). A general rule for sensory cue summation: Evidence from photographic, musical, phonetic and cross-modal stimuli. *Proceedings of the Royal Society B*, *278*(1710), 1365–1372.

Tsui, J. M. G., Hunter, J. N., Born, R. T., & Pack, C. C. (2010). The role of V1 surround suppression in MT motion integration. *The Journal of Neuroscience*, *103*(6), 3123–3138.

Turner, M. R. (1986). Texture discrimination by Gabor functions. *Biological Cybernetics*, *55*(2), 71–82.

Victor, J. D., Chubb, C., & Conte, M. M. (2005). Interaction of luminance and higher-order statistics in texture discrimination. *Vision Research*, *45*(3), 311–328.

Watson, A. B., & Eckert, M. P. (1994). Motion-contrast sensitivity: Visibility of motion gradients of various spatial frequencies. *Journal of the Optical Society of America A, 11*(2), 496–505.

Westrick, Z. M., & Landy, M. S. (2013). Pooling of first-order inputs in second-order vision. *Vision Research, 91,* 108–117.

Wilkinson, F. (1990). Texture segmentation. In W. C. S. M. A. Berkley (Ed.), *Comparative Perception* (Vol. 2, pp. 125–156). New York: John Wiley.

Wilkinson, F., & Lessard, J. (1995). Orientation, density and size as cues to texture segmentation in kittens. *Vision Research*, *35*(17), 2463–2478.

Zavitz, E., & Baker, C. L. (2013). Texture sparseness, but not local phase structure, impairs second-order segmentation. *Vision Research, 91,* 45–55.