The role of higher-order statistics in segmentation of natural textures

Elizabeth Arsenault

Doctor of Philosophy

Department of Psychology

McGill University

Montreal,Quebec

2012-12-06

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements of the degree of Doctor of Philosophy

ACKNOWLEDGEMENTS

I would like to extend my thanks, first and foremost, to my supervisor Dr. Curtis Baker for being generous with his time, helpful in his advice, and providing me with countless opportunities and resources. I would also like to thank my committee members, Dr. Frederick Kingdom and Dr. Aaron Johnson for their enthusiasm, interest, and invaluable feedback.

I wish to thank Ahmad Yoonessi for helping with many of the technical details required to get these studies off the ground when we began this project together, and everyone at McGill Vision Research for contributing to such an intellectually stimulating community. I am grateful for all I learned during lunch breaks and tea times. Thanks especially to Jason, Andrew, Jesse, Pi Chun, Simon, Alexandre, and Yeon Jin for talking me through bewildering results, teaching me new things, and sitting through hours of psychophysical testing.

I am grateful to my family, my parents Sara and Michael Arsenault, and grandparents Arthur and Barbara Arsenault, and Elizabeth, and Tom Ryan, for a lifetime of support, encouragement and pride in my achievements. I am also forever grateful to my fiancé David for his patience, countless trips to Montreal, and unwavering confidence in me.

ABSTRACT

Our perceptual experience of the visual world relies on successful segmentation of distinct regions in an image to delineate the boundaries between them. This thesis details a series of studies that begin to bridge the gap between standard energy models of texture segmentation and the properties of natural textures that affect human texture segmentation in ecological settings. Psychophysical and computational methods are applied to a combination of natural and naturalistic textures in an effort to identify the statistics found in natural textures that are relevant to human texture segmentation mechanisms, and to understand how they are processed by the visual system.

These studies draw a distinction between the energy present in an image of a texture (lower-order statistics, in the Fourier amplitude spectrum), and the spatial distribution of that energy (higher-order statistics, in the Fourier phase spectrum). The contribution of higher-order statistics to segmentation is assessed in a number of contexts by comparing psychophysical and model performance in the presence and absence of higher-order statistical information. As a whole, this work documents both the statistics that influence segmentation performance when they are the same on either side of the boundary, and those that enable performance when they define the boundary, while evaluating the extent to which present models of segmentation can take these statistics into account.

The first study suggests that the higher-order statistics present in natural texture photographs not only influence but impair contrast-boundary segmentation mechanisms, and that sparseness may be a relevant higher-order statistic in this task. The second study uses naturalistic synthetic

textures to assess individual image statistics experimentally. These experiments demonstrate that texture sparseness and global phase structure impair orientation and contrast boundary segmentation, but that local phase structure has little impact on segmentation. These findings can be accommodated by a standard filter-rectify-filter model in which the shape of the intermediate-stage nonlinearity is compressive. The third study suggests that global phase structure and texture sparseness can both enable and influence segmentation in the absence of any lower-order statistical cues for segmentation. The same model employed for contrast and orientation boundary segmentation data in the previous experiments can also account for the role of global structure and density, but not the role of local phase alignment, in these kinds of texture boundaries.

ABRÉGÉ

Notre expérience perceptuelle du monde visuel se base sur la bonne segmentation de régions distinctes dans une image pour en délimiter les bords. Cette thèse présente une série d'études qui tendent à combler le fossé entre les modèles d'énergie standards de segmentation de texture et les propriétés des textures naturelles qui affectent la segmentation de textures par l'homme en conditions écologiques. Des méthodes psychophysiques et computationnelles sont appliquées à une combinaison de textures naturelles et naturalistes afin d'identifier parmi les statistiques des textures naturelles, celles qui sont utiles aux mécanismes de segmentation de textures et comment elles sont traitées par le système visuel de l'homme.

Ces études distinguent l'énergie présente dans l'image d'une texture (statistiques d'ordre plus bas dans le spectre d'amplitude de Fourier), et la distribution spatiale de cette énergie (statistiques d'ordre plus haut dans le spectre de phase de Fourier). La contribution des statistiques d'ordre plus haut à la segmentation est étudiée dans différents contextes en comparant les performances psychophysiques à celles de modèles, en présence ou non d'information statistique d'ordre plus haut. Dans l'ensemble, ce travail présente à la fois les statistiques qui influencent la performance de segmentation lorsqu'elles sont les mêmes de chaque côté d'un contour et celles qui la permettent lorsqu'elles définissent le contour, en évaluant jusqu'à quel point les modèles de segmentation présentés peuvent prendre en compte ces statistiques.

La première étude suggère que les statistiques d'ordre plus haut présentes dans les photographies de textures naturelles non seulement influencent mais altèrent les mécanismes de segmentation de contours

de contraste, et que la dispersion peut être une statistique d'ordre plus haut pertinente pour cette tâche. La seconde étude utilise des textures synthétiques naturalistes pour évaluer expérimentalement les statistiques de chaque image. Ces études démontrent que la dispersion des textures et la structure de phase globale altèrent la segmentation de contours définis par le contraste ou l'orientation mais que la structure de phase locale n'a que peu d'impact sur la segmentation. Ces observations peuvent être accommodées par un modèle standard filtre-redresse-filtre dans lequel la non-linéarité du niveau intermédiaire est compressive. La troisième étude suggère que la structure de phase globale et la dispersion peuvent toutes deux permettre et influencer la segmentation en l'absence d'indice statistique d'ordre plus bas. Le même modèle employé pour la segmentation de contours définis par le contraste ou l'orientation peut aussi expliquer le rôle de la structure et de la densité globale mais pas le rôle de l'alignement local de phase dans ce type de contours de texture.

PREFACE

Original Contributions

This body of work makes novel use of natural and naturalistic textures as stimuli in human psychophysical tests of texture segmentation. To my knowledge, it is the first time in the field that naturalistic stimuli are used in concert with natural stimuli, and are used to fit a model of human psychophysical texture segmentation performance. These studies also make a novel distinction between image statistics that influence and those that enable segmentation.

Chapter 2 uses a novel method of imposing synthetic boundaries on natural texture stimuli to overcome some of the challenges of working with natural images. It demonstrates that segmentation mechanisms are sensitive to the higher-order statistics present in natural scenes, and makes two novel suggestions: first that removing information by phase scrambling can improve psychophysical performance, and second that sparseness and local edge structure are statistics that are potentially relevant to segmentation mechanisms.

The naturalistic texture stimuli developed in Chapter 3 are novel, and will be useful for a range of applications outside the scope of this research. Chapter 3 contains a novel demonstration of texture sparseness systematically impairing the segmentation of second-order boundaries. It is also the first time, to my knowledge, that a compressive intermediate nonlinearity has been found to be the best model of texture segmentation results.

The final manuscript, Chapter 4, provides a novel account of how global phase structure and texture density can enable segmentation. It also

provides evidence that local structure can influence structure and density segmentation though it cannot enable segmentation reliably. These findings themselves are novel, and the manuscript goes on to present a new two-stage filter model approach to successfully segment boundaries defined only by structure, as well as the first evidence of a single model for boundaries defined by either structure or energy.

Contributions of the Authors

All of the work presented here was completed under the supervision of Dr. Curtis Baker. The first study (Chapter 2) was conceptualized by Dr. Baker, and the experimental design for this study was a collaborative effort between Dr. Baker, fellow graduate student Ahmad Yoonessi, and myself. Subsequent experiments were conceived and designed by Dr. Baker and myself, with helpful input from my advisory committee Drs. Frederick Kingdom and Aaron Johnson.

Programs for running the experiments and creating texture modulated stimuli were developed by Dr. Baker and the code to run the Bits ++ device was written by Mr. Yoonessi. I wrote the scripts used to generate synthetic textures and to analyze the image statistics of both natural and synthetic textures. Dr. Johnson and Dr. Kingdom also contributed some of the image processing code we used, for rotationally averaging amplitude spectra and phase scrambling images respectively.

I was responsible for piloting and running all the experiments, except for those in the first study, which were piloted in a joint effort by Dr. Baker, Mr. Yoonessi, and myself. I performed all of the data analysis for the work presented here. I was also entirely responsible for the development, testing, evaluation, and optimization of the model presented in Chapters 3 and 4.

I took primary responsibility for writing and producing the manuscripts that make up Chapters 2, 3 and 4, with assistance from Dr. Baker. Mr. Yoonessi provided helpful input in the final stages of revising Chapter 2.

Dr. Baker provided facilities and equipment, and holds the NSERC grant that supported this research.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

ABSTRACT iii

ABRÉGÉ v

PREFACE vii

LIST OF TABLES xiii

LIST OF FIGURES xiv

1 Introduction 2

1.1 Visual Stimuli 4
1.1.1 Low-level stimuli 4
1.1.2 Natural Images 6
1.2 Linear and Nonlinear Behaviour: History, Models and Utility 10
1.2.1 The independent channels model and linear systems theory 11
1.2.2 Linear and Nonlinear Models of Visual Processing . 13
1.3 Texture 26
1.3.1 What is texture? 26
1.3.2 Texture as statistical vision 30
1.3.3 How is texture used by the visual system? 34
1.4 Texture Segmentation 36
1.4.1 Feature-based approaches 37
1.4.2 Energy-based approaches 38
1.4.3 Second-Order Vision 41
1.5 Organization of the Thesis 47

2 Higher order texture statistics impair contrast boundary segmentation 56

2.1 Abstract 56
2.2 Introduction 58
2.3 General Methods 65
2.3.1 Stimuli 65
2.3.2 Task 67
2.4 Experiment 1 69

2.4.1 Methods . . . 69
2.4.2 Results . . . 69
2.5 Experiment 2 . . . 70
2.5.1 Methods . . . 70
2.5.2 Results . . . 70
2.6 Experiment 3 . . . 73
2.6.1 Methods . . . 73
2.6.2 Results . . . 74
2.7 Discussion . . . 76
2.8 Conclusion . . . 84

3 Texture sparseness, but not local phase structure, impairs second-order segmentation . . . 96

3.1 Abstract . . . 96
3.2 Introduction . . . 97
3.3 General Methods . . . 100
3.3.1 Stimuli . . . 100
3.3.2 Boundary Creation . . . 103
3.3.3 Apparatus and Observers . . . 105
3.3.4 Task . . . 105
3.3.5 Data Analysis . . . 105
3.4 Experiment 1: Contrast boundary segmentation . . . 107
3.4.1 Methods . . . 107
3.4.2 Results . . . 107
3.5 Experiment 2: Orientation boundary segmentation . . . 110
3.5.1 Methods . . . 110
3.5.2 Results . . . 110
3.6 Model . . . 113
3.6.1 Filter-Rectify-Filter Model . . . 113
3.6.2 Simulation . . . 115
3.6.3 Optimization . . . 116
3.6.4 Results . . . 118
3.7 Discussion . . . 120

4 Texture segmentation is enabled by sparseness and global phase structure, but not local phase alignments . . . 136

4.1 Abstract . . . 136
4.2 Introduction . . . 137
4.3 General Methods . . . 140
4.3.1 Stimuli . . . 140
4.3.2 Apparatus and Observers . . . 142
4.3.3 Task . . . 142
4.3.4 Data Analysis . . . 143
4.4 Experiment 1: Structure Boundary Segmentation . . . 144

4.4.1 Methods . . . 144
4.4.2 Results . . . 145
4.5 Experiment 2: Density Boundary Segmentation . . . 147
4.5.1 Methods . . . 147
4.5.2 Results . . . 147
4.6 Model . . . 149
4.6.1 Filter-Rectify-Filter Model . . . 149
4.6.2 Simulation . . . 151
4.6.3 Simulation Results . . . 152
4.6.4 Optimization . . . 152
4.6.5 Optimization Results . . . 154
4.7 Discussion . . . 156
4.7.1 Psychophysics . . . 156
4.7.2 Model . . . 158
4.7.3 Sparseness . . . 159
4.7.4 Structure-defined boundary processing . . . 160
4.8 Conclusions . . . 161

5 Discussion . . . 172

5.1 Higher-order texture statistics . . . 173
5.2 Studying segmentation . . . 175
5.3 Role of nonlinearity . . . 177
5.4 Improving the model . . . 180

References . . . 185

LIST OF TABLES

Table page

3–1 Results of Bonferroni post-hoc tests for contrast boundary segmentation. t is the Bonferroni-corrected t statistic, p the significance level, d Cohen's d, and * indicates a statistically significant comparison. 109

3–2 Results of Bonferroni post-hoc tests for orientation boundary segmentation. t is the Bonferroni-corrected t statistic, p the significance level, d Cohen's d, and * indicates a statistically significant comparison. 112

4–1 Results of Bonferroni post-hoc tests for structure boundary segmentation. t is the Bonferroni-corrected t statistic, p the significance level, d Cohen's d, and * indicates a statistically significant comparison. 146

LIST OF FIGURES

Figure page

1–1 Sample of the types of boundaries that occur in natural scenes. Highlighted boundaries are defined by: (A) luminance and texture, (B) luminance, contrast, and texture scale, (C) mainly texture, (D) texture, luminance and contrast. 49

1–2 Evidence for spatial frequency- and orientation-selective processing in early vision. (A) Adaptation of a single spatial frequency channel. Pictured is a schematic of the results observed by Blakemore and Campbell (1969). The solid line shows sensitivity to gratings of a wide range of spatial frequencies. The dots show the sensitivity at a range of spatial frequencies following adaptation to a grating of the spatial frequency marked by the thin dashed line. Grey dashed curves show the spatial frequency-selective channels inferred from these results. The sensitivity of the channel closest to the adapted spatial frequency has been reduced. (B) Orientation selective neuron in visual cortex. This panel illustrates the type of results observed by Hubel and Wiesel (1959). The dashed lines in the shaded area show stimulus-specific responses to bars of light at different orientations (pictured at right). The neuron depicted prefers vertical stimuli. (C) Spatial frequency selective units in visual cortex. Each curve depicts the response amplitude of a different cortical simple cell. Each neuron is tuned to a different spatial frequency. Adapted from Maffei and Fiorentini (1973). 50

1–3 Relative contributions of amplitude and phase spectra to texture appearance. The 'Image' column shows two texture photographs (vines and grass), and their amplitude spectra are illustrated in the centre column. The grass is very broadband for orientation, and while the vines are broadband, most of the energy in the image is located in the vertical dimension (horizontal in the frequency domain). In the final column, the phase spectra are swapped between the two images. In the top image, the amplitude spectrum of the vines is paired with the phase spectrum of the grass. In the bottom image, the amplitude spectrum of the grass is paired with the phase spectrum of the vines. The vertical energy of the vines is still very apparent in the resulting image, though the continuous contours of the vines themselves are no longer present. The structure of the grass is visible, though faint, as relatively little energy is available to provide these features with contrast. In the bottom image, the leaves from the texture of vines appear to have been transplanted into a field of broadband noise. 51

1–4 Natural textures and their spatial frequency amplitude spectra. The amplitude for each spatial frequency component is plotted on the graph. These graphs are the same as those in the centre column of Figure 1–3, but averaged over the rotational axis (orientation). An inverse relationship between amplitude and spatial frequency with the most power in the lowest spatial frequencies is evident. There are some individual differences between textures, but the trend for a $1/f$ energy falloff is consistent. 52

1–5 Different basic rectifications. The linear input signal is shown as a dashed line with a slope of 1, and the output is shown in solid grey. The columns compare half- and full-wave rectification, and the rows show (from top to bottom): compressive (square root), linear (absolute value), and expansive (square law) nonlinearities. 53

1–6 Algorithm for second-order segmentation. (A) The luminance structure of a texture carrier with an oblique contrast modulation is detected by high-spatial frequency filters. (B) The responses of these filters cannot be used for contrast boundary segmentation by a linear filter of an appropriate spatial frequency and orientation because the mean response is the same between the high- and low-contrast regions. (C) Following rectification by a pointwise absolute value transformation, all filter responses are positive and so the mean response to the excitatory region is greater than that to the inhibitory region, and the boundary may be segmented. . . . 54

2–1 Examples of excluded (A) and included (B) natural textures. (A-top): Images that were excluded due to a subjective judgment that they were not sufficiently uniform, homogeneous, in focus, or contained prominent segmentable objects. (A-bottom): images that were excluded during computer screening due to inhomogeneity of luminance or contrast between two or more quadrants. (B): Images that were included in the texture corpus. 85

2–2 Examples of the stimuli used to determine modulation depth thresholds in experiments 1, 2, and 3 shown at three modulation depths (top-bottom: 75, 50 & 32). The envelope is a left- or right-oblique half-disc contrast modulation applied to an intact (left) or phase-scrambled (right) natural texture. 86

2–3 Modulation depth threshold results from Experiment 1 for four observers for intact and scrambled texture conditions. Thresholds were lower for the phase-scrambled textures (light bars) than for the intact textures (shaded bars). Error bars represent ± 1 standard error. 87

2–4 Modulation depth threshold results from Experiment 2 for three observers. Each symbol plots the phase-scrambled versus the intact threshold for a particular texture. In almost all 20 textures tested, for all three observers, the symbols lie below the 1:1 line (dashed), indicating that the intact threshold is higher than the phase-scrambled threshold. The amount of reduction, or the distance from the 1:1 line, is texture-dependent. Error bars show the standard error on each measurement. 88

2–5 Histogram of texture carriers used in Experiment 2, based on average magnitude of the effect of phase scrambling. The textures that show a larger change (> 2 dB) tend to have more prominent edges, and appear to be more sparse. 89

2–6 Stimuli used to determine carrier contrast thresholds in Experiment 3 shown at a range of carrier contrasts (top-bottom: 8,5 & 3% RMS contrast), all with 100% modulation depth. Thresholds were determined for intact (A) and scrambled (B) textures. 90

2–7 Carrier contrast threshold results from Experiment 3 for two observers. Each symbol plots the phase-scrambled vs the intact threshold for a particular texture. The dashed line indicates where a texture's intact and phase-scrambled thresholds correspond exactly. Carrier contrast thresholds are centered on the 1:1 line, suggesting no systematic change in detectability when a texture is phase scrambled. Variation in thresholds between observers, textures, and conditions is minimal. 91

2–8 Modulation depth threshold results from Experiment 3 for two observers. Each symbol plots the phase-scrambled versus the intact threshold for a particular texture, with each texture a fixed increment above its detection threshold. The dashed line indicates where a texture's intact and phase-scrambled thresholds correspond exactly. Modulation depth thresholds measured with these detectability-equated contrasts are still systematically lower for scrambled than for intact textures. 92

2–9 Relationship between image statistic indices and the change in segmentation threshold between intact and scrambled conditions in decibels. (A) Slope of fall-off of Fourier spectrum. (B) sparseness, as measured using intensity histogram kurtosis. (C) sparseness, as measured with Hansen and Hess (2007)'s LSSM a wavelet-based metric developed for natural scenes. (D) Edge density, modified from Bex (2010). Note the lack of relationship between threshold change and kurtosis, LSSM, or slope, but a clear correlation between edge density and threshold change. 93

3–1 Examples of properties that enable versus those that influence segmentation. In both of these textures, a contrast difference *enables* the percept of a right oblique boundary. The properties of the materials (leaves and bark) forming the carrier textures are different in structure which results in a difference in the strength of the boundary percept, even though the modulation depth of the contrast boundary is identical in both examples. In this example, the characteristics of the textures can be said to *influence* segmentation. . 124

3–2 Procedure for constructing naturalistic textures and orientation-modulated boundaries. (A) Micropattern types used to create synthetic textures. Top: Intact Gaussian-enveloped square wave, Bottom: Phase-scrambled square wave within the same Gaussian envelope. (B) Top: Variations of phase scrambled micropatterns, Bottom: Same instances, polarity reversed. (C) Procedure for quilting stimuli. Half-disc envelopes are multiplied with their corresponding carrier textures. These modulated halves are then combined. The modulation depth of the stimulus shown is 100%. 125

3–3 Examples of stimuli at varying levels of difficulty. Top: contrast modulations at (L-R) 30%, 50%, and 75% modulation depths. Bottom: orientation modulations at (L-R) 60%, 80%, and 100% modulation depths. 126

3–4 Examples of contrast modulated stimuli used for Experiment 1, shown at a modulation depth of 50%. The phase alignment conditions are arranged horizontally, while the density is increased vertically from top to bottom. The three phase alignment conditions are intact (INT), locally scrambled (LS), and globally scrambled (GS). Notice that the globally scrambled condition does not appear different at varying densities because density information is destroyed by phase scrambling. 127

3–5 Experiment 1 (contrast boundary segmentation) results for four observers (small graphs), and the average of these observers (large graph). The structure conditions are indicated by the data series: intact (INT), filled circles; locally scrambled (LS), open circles; and globally scrambled (GS), open triangles. Density increases along the abcissa. Note improved performance for phase scrambled carriers (GS) and lack of effect of density in this condition. In contrast, intact and locally scrambled conditions both result in higher thresholds at low densities than at high. Error bars indicate standard errors. 128

3–6 Examples of orientation modulated stimuli used for Experiment 2, shown at a modulation depth of 100%. The phase alignment conditions are arranged horizontally, while the density is varied vertically. The globally scrambled condition (GS) was only tested once, because density information is destroyed by phase scrambling and (as expected) no systematic effect of density was observed in the previous experiment. 129

3–7 Experiment 2 (orientation boundary segmentation) results for four observers (small graphs), and the average of these observers (large graph). The structure conditions are: intact (INT), filled circles; locally scrambled (LS), open circles; and globally scrambled (GS), open triangles. These results show improved performance for phase scrambled carriers (GS) as previously found for contrast segmentation. There is no systematic difference between the intact and locally scrambled conditions, as both result in higher thresholds at low densities than at high. Error bars indicate standard errors. 130

3–8 Architecture of our FRF-style model. Symbols at top correspond to equations in text. $\mathbf{G}_1$ and $\mathbf{G}_2$ stand for the first- and second-stage filters respectively, $\mathbf{F}_1$ and $\mathbf{F}_2$ stand for the filter outputs, $\mathbf{R}$ for the output of the rectification. The output of each filter following the injection of noise is indicated by o, and the result of the comparison between the files by d. 131

3–9 Model optimization process. (A) Two-parameter error space for the model. Intensity in each image is the disparity in decibels between model and human performance. Four performance metrics (INT slope, GS difficulty, INT-LS difference, and INT-GS difference) were computed for each combination of power-law exponent and decision noise level (increasing from left to right). The standard deviation of the decision noise distribution was different for each value of k, and its range was selected so that underestimates and overestimates GS difficulty were represented. (B) Error magnitude for each value of k in the three remaining metrics when the noise level is optimized on GS difficulty. 132

3–10 Results from model in Figure 9 with parameters k and a optimized for orientation boundary segmentation. (A) Orientation boundary segmentation results are similar to those for humans: lower thresholds for GS and higher thresholds for INT and LS that decrease as density increases. (B) Contrast boundary segmentation results are also similar to those for humans: lower thresholds for GS, higher thresholds for INT and LS that decrease as density increases. 133

3–11 Second-stage filter responses ($\mathbf{F}_2$) measured for each of the conditions tested in Experiment 1 using k=0.25. Responses are averaged over ten randomly-generated stimuli, and the error bars indicate standard error. (A) Second-stage filter responses to an unmodulated stimulus. The measured response reflects the second order structure present in the texture before modulation. Responses for both left (L, dark colours) and right (R, light colours) filters are the same, because whatever structure is present is non-oriented. (B) Second-stage filter responses to a stimulus modulated with a right-oblique boundary with a modulation depth 0.5 dB above threshold in each condition. In this case, the responses for the right oblique filters are higher signalling the presence of a boundary. 134

4–1 Procedure for quilting structure- and density-modulated boundaries. Half-disc envelopes are multiplied with their corresponding carrier textures. These modulated halves are then combined. The modulation depth of the stimulus shown is 100%. 162

4–2 An example of stimuli at varying levels of difficulty. This figure shows an INT-GS texture at 100%, 75%, and 50% modulation depth. 163

4–3 Structure modulated stimuli used for Experiment 1, shown at a modulation depth of 100%. The boundary types are organized into columns, while the density conditions are organized into rows. 164

4–4 Experiment 1 (structure boundary segmentation) results for four observers. The boundary conditions are: INT/GS, between an intact texture and a globally scrambled texture (filled circles); INT/LS, between an intact texture and a locally scrambled texture (open circles); and LS/GS, between a locally scrambled and globally scrambled texture (open triangles). The grey 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 do not appear to be dependent on density. Segmentation performance on the INT/GS and LS/GS conditions is dependent on density with very similar thresholds at low densities, but LS/GS becomes more difficult more rapidly at high densities. 165

4–5 Density boundary stimuli used in Experiment 2 at a modulation depth of 100%. Each stimulus is a modulation between a texture with 595 micropatterns and one with 2975 micropatterns. Because no visible (or statistical) boundary was formed in the GS condition, it was not tested. 166

4–6 Experiment 2 (density boundary segmentation) results for four observers. Density boundaries do not exist following phase scrambling, so the globally scrambled (GS) condition was not tested. Performance on the locally scrambled (LS) condition appears to be slightly but systematically better than on the intact (INT) condition. 167

4–7 Segmentation model architecture and modelling results alongside human data. (A) The architecture of the basic FRF model used in this study. (B) Results for model as fit using orientation segmentation data compared to human data. Human data is in purple, model in black. INT (intact)/GS (globally scrambled) (shaded circles) thresholds are estimated very well, though best at the lowest and highest densities. LS (locally scrambled)/GS (open circles) thresholds present the most difficulty, the model matches human performance at the highest density, but over-estimates thresholds at low and moderate densities. INT/LS (triangles) cannot be segmented reliably enough to compute a threshold for humans or the model, and are depicted here at 100% modulation depth. (C) Results for the model and humans in the density condition. The model performs very well when the textures are intact (INT), but overestimates thresholds when they are locally scrambled (LS). 168

4–8 Optimization procedure for structure and density boundaries and model results. (A) Matrices of SSE for each noise/exponent parameter value combination. Darker colours indicate lower error. Error is much lower for compressive nonlinearities. Circled points indicate the model evaluated for panels C and D. (B) Error with respect to noise level when the exponent of the nonlinearity is fixed at 0.25 for density and structure segmentation. The error functions have different minima, so a compromise (circled) was chosen for the model evaluated in panels C and D. (C) Structure segmentation results for the model with the noise fit to this particular data to minimize sum-of-squares error. INT/GS (shaded circles) is underestimated at all densities, but more so at low densities. LS/GS (open circles) thresholds are not far off any single estimate, but the rate at which thresholds increase with density does not accelerate as in the human data. INT-LS (triangles) can't be segmented by humans or the model well enough for thresholds to be computed. (D) Density shows a reasonably good estimate in the INT condition, but still over-estimates threshold in LS condition, if less severely. . . 169

4–9 Image energy following filtering and rectification, for different texture conditions. Each data point in this figure shows the spatially pooled, rectified, average (N=4) first-stage filter response to an unmodulated texture. Three types of textures are shown: an intact texture with a density of 595 micropatterns (purple diamonds), an intact texture with a density of 2975 (green circles), and a globally scrambled texture (blue triangles). The power-law exponent (k) is varied along the horizontal axis. All textures have approximately the same energy when the power law exponent is a square law because they are presented at a fixed RMS contrast. However, when the power law exponent is more expansive or compressive than 2, differences emerge between these textures that can enable segmentation. 170

1 Introduction

In this section I provide a thorough literature review of four sections: visual stimuli, linear and nonlinear models of visual processing, texture stimuli, and texture segmentation. In the section on visual stimuli, I introduce the historically important sine-wave grating along with the concept of spatial-frequency and orientation selective processing, followed by a brief summary of the properties of natural images and the efficient coding principles that make them informative stimuli. In the next section on models of visual processing, I review the linear systems approach to understanding visual perception as well as the independent channels model. I provide an outline of the early visual system, at each stage discussing the aspects of its behaviour that can be modelled using a linear approach and highlight those aspects that require nonlinear models. In the section on texture stimuli, I review historical and modern approaches to creating textures as well as those statistics used to quantify the properties of existing textures. Finally, I look into the problem and models of texture segmentation, focusing on the filter-rectify-filter model architecture and what has been inferred about its properties. I close the introduction with an overview of the organization of the thesis.

CHAPTER 1
Introduction

The world is made up of collections of objects arranged in space, and boundaries between these objects result in discontinuities in the visual image. Before many ecologically critical functions such as object recognition can be computed, discontinuities signalled by one or more changes in an image property, such as luminance, colour, contrast, or texture, must be detected and localized. Boundaries signalled by changes in texture are particularly interesting because textures contain local structure that itself is composed of fine-scale boundaries. The visual system is able to use some attributes of textures to *segment* these more complex boundaries, yet the mechanisms underlying this segmentation process are only partially understood.

Figure 1–1 illustrates some of the combinations of features that can define the boundaries occurring in natural scenes. In the figure (A) shows a boundary defined by luminance and texture, (B) a boundary defined by luminance, contrast, and texture scale, (C) a boundary defined mainly by texture, and (D) a boundary defined by texture, luminance and contrast. Ultimately we hope to refine our understanding of segmentation mechanisms so that given a natural scene, we could predict where boundaries of any type are perceived and how salient they are. Our current understanding of texture segmentation is focused on a limited number of texture properties (particularly contrast, orientation and spatial frequency). This thesis expands our understanding to include the role of texture *structure* in segmentation in two ways: first, by assessing how human psychophysical

performance in a segmentation task is affected by spatial structure, and then by revising a popular model of segmentation to account for the roles of different types of structure.

In this chapter, I will begin with an overview of the stimuli used by vision scientists to examine the visual system, and the relationship between these signals and visual encoding. Next, I will discuss the behaviour of the visual system as it constrains the linear and nonlinear models that are thought to describe visual processing from the retina through the lateral geniculate nucleus of the thalamus, to the primary visual cortex, and beyond. The stimuli used in this thesis, texture images, will next be discussed in detail, and then the particular nonlinear models thought to produce texture segmentation behaviour. Finally, I will provide an outline of the organization of the data chapters of the thesis.

1.1 Visual Stimuli

The problem of understanding the visual system is sometimes couched in terms of a classic system identification problem: if we know a given input (a stimulus) produces a given output (a spike train in the case of a single neuron, a response in the case of a human observer), what properties must the system have to generate this mapping of input to output? It is evident that what can be learned about this complex system depends a great deal on the stimuli that are employed. In this section, I will describe the stimuli that the visual system processes, and their relationship to our current understanding of how they are represented.

1.1.1 Low-level stimuli

Vision in the frequency domain

Any signal can be represented as a sum of sine waves, and the most classic example of this is the square wave: a sum of phase-aligned sine waves whose amplitudes decrease as spatial frequency increases in a $1/f$ relationship. In the same way, any image may be thought of as a two-dimensional signal, extending as a function of x and y in space. In the same manner as one-dimensional signals, any image can be represented as a sum of sine wave *gratings* simply by specifying their spatial frequency, orientation, amplitude, and phase. The two-dimensional Fourier transform is the tool used to represent images in this way, and will be discussed in more detail below. A detailed account of the use of frequency analysis for visual stimuli may be found in DeValois and DeValois (1990).

Campbell and Robson (1968) measured the ability of observers to discriminate between sine wave gratings and more complex wave forms. They found that the spatial frequency components of the stimulus critical

to discrimination had to reach their contrast detection thresholds independently of the other components present in the stimulus. In other words, they showed that–at low contrast–the energy in the stimulus appeared to be processed by independent detectors selective for spatial frequency. Blakemore and Campbell (1969) strengthened this idea by measuring the contrast sensitivity function (CSF), a plot of contrast sensitivity (sensitivity = 1/threshold) against grating spatial frequency (Figure 1–2 A). They had observers adapt to a particular spatial frequency by viewing it for a prolonged period of time, and then re-measured the CSF for each observer. They found that the sensitivity in the region of the CSF corresponding to the adapting frequency was decreased, while distant parts of the CSF were unaffected. This too suggested that early visual processing was based on a number of independent spatial frequency channels.

Earlier, Hubel and Wiesel (1959) recorded from single units in cat visual cortex while presenting oriented bars of light across the receptive field. They demonstrated that cortical neurons give the greatest number of responses to bars of a particular orientation, somewhat fewer responses to bars of nearby orientations, and few or no responses to distant orientations, showing orientation *selectivity* (Figure 1–2 B). Maffei and Fiorentini (1973) went on to adopt sine wave gratings as a neurophysiological stimulus and demonstrated that cortical neurons are selective for the spatial frequency of gratings (Figure 1–2 C) at their preferred orientation.

Sine wave gratings are a powerful stimulus for experiments in psychophysics and neurophysiology. This is largely due to the spatial frequency and orientation selective behaviour in early vision. Additionally, the properties of the optics of the eye are linear and thus sinusoidal stimuli may be

used to provide a good characterization (Hopkins, 1962; Campbell & Green, 1965).

The Fourier transform

The Fourier transform is a very useful tool for vision science because it enables the analysis of arbitrary stimuli in terms of their spatial frequency and orientation content. The Fourier transform of an image can be described in two parts: the amplitude spectrum and the phase spectrum (Figure 1–3). The amplitude spectrum on the captures how much contrast energy there is at each oriented spatial frequency component of the image. The phase spectrum captures the spatial relationships between the orientation and spatial frequency components, and thus describes the spatial layout of the energy in the amplitude spectrum. The phase spectrum is necessary to capture perceptually salient image features, such as broadband edges (Piotrowski & Campbell, 1982). Recognizable signal features and object identities both depend critically on the phase spectrum, and not on the amplitude spectrum (Oppenheim & Lim, 1981; Morgan, Ross, & Hayes, 1991).

1.1.2 Natural Images

As early as fifty years ago, both behavioural (Attneave, 1954) and physiological (Barlow, 1961) observations led to the idea that an important role of visual processing is to represent the incoming signal in an efficient manner. In this section, I will discuss the redundancies in natural image stimuli, followed by an overview of some efficient coding schemes and evidence for efficient coding at various levels of processing.

Redundancies in natural images

The set of natural images is much smaller than the set of all possible luminance combinations amongst all image pixels, because natural images

are projections of scenes constrained by the laws of the natural world (Simoncelli & Olshausen, 2001; Chandler & Field, 2007). These constraints permit generalizations about natural images, and in generalizations there is redundancy (Olshausen & Field, 2004). Two particular manifestations of these constraints that I will discuss in more detail are the power spectrum that is characteristic of natural images (lower-order statistics), and the redundancy of information in each image due to *structural sparseness* (higher-order statistics).

The most basic generalization about natural images is their power spectrum. The energy in natural images falls off as spatial frequency increases, in an approximately $1/f$ relationship Field (1987). A popular view is that this is because the individual broadband edges in an image have a $1/f$ fall-off (Figure 1–4) (Field, 1987). However, Ruderman (1997) proposed that the physics governing the distribution of discrete objects of a variety of sizes and distances as they are projected onto a given image plays an important role in determining the $1/f$ power spectrum as well. In this thesis, I take both these theories into account when creating synthetic natural stimuli in Chapter 3 and Chapter 4. The synthetic textures employed are composed of $1/f$ edgelet micropatterns of a variety of sizes. The proportion of micropatterns at each size is governed by a $1/f$ relationship.

The information in an image (such as an image falling on a camera sensor or the retina) is highly redundant (Barlow, 2001) because there is a strong correlation between image intensity values and their proximity in space (Simoncelli & Olshausen, 2001). This is partly because of the $1/f$ power spectrum, but is also due to the continuity of surfaces populating much of the visual world. For example, a pair of neighbouring image pixels

are far more likely to be part of the same surface than they are to straddle an object boundary. Because natural images tend to contain mainly surfaces with localized boundaries, they are said to be 'structurally sparse' (Hansen & Hess, 2007).

The impact that these regularities may have on coding will be examined below.

Coding

It is important to note that whether or not a code is 'efficient' depends entirely on the basis by which efficiency is judged. For example, the number of retinal ganglion cells used to represent an image is directly constrained by the size of the optic nerve, so a code that is efficient in that context will rely on a minimum of retinal ganglion cells. On the other hand, in the cortex, there is less need to minimize the number of neurons, but there are metabolic constraints such that an efficient code should minimize the number of action potentials. Neural impulses are expensive energetically (Attwell & Laughlin, 2001), so an economy of impulses would reduce energy consumption (Levy & Baxter, 1996). This implies that neural responses in the cortex should be relatively independent of one another to minimize redundancy: a *sparse* code. A sparse code is being employed if most neurons are inactive most of the time, because this means the minimum number of neurons are being used to represent a signal (Simoncelli & Olshausen, 2001).

In addition to the basis by which efficiency is judged, Field (1987) notes that whether a code is efficient also depends considerably on the structure of the image being encoded. A sine wave plaid is most efficiently represented in the form of its Fourier coefficients. On the other hand, an image of sparse dots is most efficiently represented as a list of position-intensity data pairs.

There is evidence for efficient coding in the vertebrate retina (Atick & Redlich, 1990), as well as evidence that the human visual system develops to represent the $1/f$ power spectrum in natural images in an efficient way (Ellemberg, Hansen, & Johnson, 2007, 2012). Further evidence of efficient coding has come from natural image analysis. Working backwards from large libraries of natural images, images of uncorrelated components of these scenes have been recovered using dimensionality reduction methods (Bell & Sejnowski, 1997; Olshausen & Field, 2004). The efficient components that they recover bear a strong resemblance to simple cell receptive fields at a variety of orientations, spatial frequencies, phases, and positions (Cadieu & Olshausen, 2012).

1.2 Linear and Nonlinear Behaviour: History, Models and Utility

Visual perception begins at the interface between air and the cornea, where incoming light is refracted and focused to form an image on the retina. Photoreceptors at the back of the retina transduce light into neural signals, which are propagated through retinal circuitry to retinal ganglion cells whose axons exit via the optic nerve and project to the lateral geniculate nucleus (LGN) of the thalamus. From there, visual signals are passed on to the primary visual cortex, followed by a multitude of increasingly specialized visual areas of extrastriate visual cortex.

In this section I will describe what we can infer about the properties of the visual system at various levels of processing from the retinal ganglion cells to extrastriate visual cortex. I will examine each level's behaviour, and note how these properties are represented in the form of models. Behaviour is presented in the form of single-unit responses, collected using neurophysiological methods, as well as perceptual measurements collected using psychophysical methods. It is important that the conclusions drawn from these separate approaches be considered individually, because the connection between the behaviour of individual neurons and the behaviour of a human subject is nontrivial to infer, even when similar phenomena are observed.

I will pay particular attention to the linear and nonlinear components of models of spatial processing, because of the nonlinearities that play a critical role in the topic of this thesis: resolving second-order discontinuities. Visual neurons also display nonlinear behavior in the temporal dimension of their spatiotemporal receptive fields (Dawis, Shapley, Kaplan, & Tranchina, 1984; Cai, Deangelis, & Freeman, 1997), and other temporal dynamics (e.g. adaptation (Bownds & Arshavsky, 1995; Purpura, Tranchina, Kaplan, &

Shapley, 1990) and response transients (Ikeda & Wright, 1975b, 1975a; Dean & Tolhurst, 1986)). As this thesis is primarily concerned with spatial texture processing, I will not cover matters of temporal dynamics in detail.

1.2.1 The independent channels model and linear systems theory

Two paradigms form the basis of many of the early ideas on the nature of visual computation: linear systems theory, and the 'independent channels' model. These concepts were typically used in concert, but are discussed separately here, because their strengths and weaknesses are not identical despite considerable overlap.

The independent channels model was developed based on psychophysical data (Campbell & Robson, 1968; Blakemore & Campbell, 1969; Graham, 1989; Graham & Nachmias, 1971) and suggests that the detection of visual stimuli is based on independent analyzers that fragment early visual processing into orientation and spatial frequency specific subsets of Fourier space. The adaptation experiment performed by Blakemore and Campbell (1969) to lead to this inference is described in Section 1.1.1, and the inferred independent spatial frequency channels are illustrated in Figure 1–2 A. Critically, it predicts that these detectors are 1) individually responsible for signalling the presence of the stimulus property for which they are selective; and 2) free from cross-channel interactions, at least at an early point in the stream of visual processing.

This theory also applies to neurophysiology, in that it suggests that a single neuron's responses should depend only on feedforward input. Each 'channel' can be imagined to consist of a population of spatial frequency and orientation selective neurons at an early stage of processing, which respond independently of one another. The simplest version of this model represents each neuron as a linear filter.

Linear systems theory was appealing to visual neurophysiologists because many receptive fields in early vision have excitatory and inhibitory zones that appear to summate linearly under many conditions. The theory is applied by estimating linear response properties for receptive fields at all stages of early visual processing: in the retina, the lateral geniculate nucleus of the thalamus (LGN), and in primary visual cortex. The response properties of these model neurons, because of their linearity, can be estimated based on the spatial layout of the excitatory and inhibitory regions of their receptive fields. The precise nature of these models will be examined in subsequent sections. This theory is also useful for interpreting psychophysical data by passing stimuli through a bank of linear filters that loosely match the selectivity of visual neurons and are thought to approximate the earliest stages of cortical processing.

The linear filter model of visual processing has appealed to vision scientists because approximately linear behaviour is evident in the optics of the eye, as well as at early levels of visual processing by the nervous system; and assuming a mostly linear system provides a starting point to provide a good first approximation that can be improved by adding nonlinear elements as needed.

Campbell and Green (1965) confirmed that the optics of the eye itself are linear, so visual stimuli are not transformed in a nonlinear way before being projected onto the retina. Linear filter models of early retinal, thalamic, and cortical processing have a long history. Spatially segregated excitatory and inhibitory regions were one of the earliest observations about receptive field structure in retinal ganglion cells (Kuffler, 1953), thalamus (Hubel & Wiesel, 1962), and cortex (Hubel & Wiesel, 1959). Later experiments quantitatively established that many of the neurons in these

areas sum approximately linearly; for example, in retinal ganglion X cells (Enroth-Cugell & Robson, 1966; Hochstein & Shapley, 1976), LGN (Cai et al., 1997) and cortical simple cells (Movshon, Thompson, & Tolhurst, 1978b).

Given the predictions of a linear system, it is easier to then identify the nonlinearities required fit the predictions to observed behaviour (DeValois & DeValois, 1990). A good example is seen in the results of Enroth-Cugell and Robson (1966) with the identification of Y-cell nonlinear subunits. Movshon, Thompson, and Tolhurst (1978a)'s model of complex cells as the nonlinear combination of simple cell outputs, and the Adelson and Bergen (1985) model of motion detection using quadrature pairs of linear filters are both examples of models built in a simple way from nonlinear summation of linear receptive field inputs. In a similar manner, models of gain control and second-order modulation are nonlinear, yet rely heavily on linear filter components. These classes of models will be discussed in depth below.

1.2.2 Linear and Nonlinear Models of Visual Processing

In this section, I will proceed from the retina, through the lateral geniculate nucleus, to the primary visual cortex and later cortical visual processing, describing the behaviour of single neurons in these areas and some of the models that capture that behaviour. Inferences about retinal and geniculate function are drawn entirely from neurophysiological evidence. In the cortex, some psychophysical phenomena are noted following the discussion of neurophysiology to demonstrate that many of the principles of encoding at the level of single neurons also have counterparts at the perceptual level.

The Retina

The retina is the sheet of neurons at the back of the eye, onto which an image is formed. At the very back of the sheet are the photoreceptors: rods and cones. Rods and cones transduce photons into neural signals, and so are where visual information first enters the nervous system. The output layer of the retina, composed of ganglion cells, projects directly to the lateral geniculate nucleus of the thalamus via the optic nerve (a bundle of ganglion cell axons).

Retinal ganglion cells receive input from one or more bipolar cells. They have receptive fields with spatially segregated excitatory (ON) and inhibitory (OFF) regions in a centre-surround structure (Kuffler, 1953; Barlow, 1953). Due to the centre-surround structure, they exhibit some low-pass spatial frequency selectivity (De Monasterio & Gouras, 1975). Generally, they exhibit a relatively high level of spontaneous activity (De Monasterio & Gouras, 1975). There are examples of both linear and nonlinear behaviour among ganglion cells. For example, the X cells in cat behave linearly (Enroth-Cugell & Robson, 1966), while a second class, Y cells, display characteristic nonlinear behaviour (Hochstein & Shapley, 1976).

The receptive field shape of retinal ganglion cells has been modelled using difference-of-gaussian (DoG) functions (Rodieck, 1965). The shape of the DoG is representative of the concentric, antagonistic excitatory ON and inhibitory OFF zones that are typical of retinal and thalamic neurons. Linear convolution with a DoG model predicts X cell responses to a variety of stimuli. The nonlinear behaviour of the Y cells can be accounted for with a model including high-spatial frequency nonlinear subunits in addition to the linear DoG receptive field (Hochstein & Shapley, 1976).

Another, more pervasive, nonlinearity in retinal processing is contrast gain control (Shapley & Victor, 1978, 1981; Victor, 1987; Baccus & Meister, 2002), first evident in retinal ganglion cells, but present in later visual areas as well. If a neuron were to act as a linear filter, the firing rate of the neuron would increase linearly with contrast. However, this linear relationship between firing rate and contrast is not apparent in the responses of retinal ganglion cells (Shapley & Victor, 1978). Contrast response nonlinearities as described above are believed to be caused by contrast gain control. Contrast gain control is thought to depend on the standard deviation of local luminance intensities, consistent with the fact that it is independent of precise stimulus position (Benardete & Kaplan, 1999). This process of gain control is an example of *normalization*, a computation where neural responses are modulated (typically modelled as a divisive operation (Carandini & Heeger, 2012)) by the responses of a local neighbourhood. Contrast gain control is a distinct mechanism from light adaptation, which occurs primarily in the photoreceptors (Rushton, 1965; Ullman & Schechtman, 1982).

Lateral Geniculate Nucleus

The responses of neurons in the lateral geniculate nucleus (LGN) are very similar to those of retinal ganglion cells, leading to the suggestion that the LGN is a central 'relay' station, rather than a transformative processing stage (DeValois & DeValois, 1990). As in the retina, neurons in the LGN have a concentric, antagonistic excitatory and inhibitory receptive field structure. Many of the response properties of LGN cells can be explained by their RF structure (Cai et al., 1997), but there is evidence that nonlinear normalization processes, such as contrast gain control, are stronger in the LGN than in the retina (Bonin, Mante, & Carandini, 2005).

The conventional receptive field model for neurons in the lateral geniculate nucleus is a difference of gaussians (DoG) linear filter, plus half-wave rectification (Rodieck, 1965; Kuffler, 1953). Unlike retinal ganglion cells, neurons in the LGN have low levels of spontaneous activity, and so their responses are thresholded because they cannot decrease their firing rate below zero. In the simplest case, the response of a neuron is *half-wave rectified*: the positive components of the signal are maintained in the response, but the negative components are not (Figure 1–5 A). When a neuron responds equally to positive and negative input, there is a *full wave rectification*: negative inputs are transformed into proportionally positive outputs, so the response signals magnitude (i.e. absolute value), but not signal polarity (Figure 1–5 B).

In simple half- or full-wave rectification, the response magnitudes are linear with signal strength. However, rectification sometimes includes other nonlinear aspects such as expansive, compressive, or more complex saturating nonlinearities. An expansive power law (for example, a square law) is characterized by proportionally boosting larger responses (Figure 1–5 C, D). A compressive power law boosts smaller responses but saturates for larger responses (Figure 1–5 E, F). In some cases (as in the contrast response function of many single neurons), a sigmoidal nonlinearity is apparent, which boosts low contrast responses in an expansive way, but saturates at higher contrasts.

Spatial frequency selectivity is weak in the LGN because the surround is weaker than the centre, and low pass at low contrasts. Size selectivity is greater at high contrasts, an effect attributed to nonlinear normalization resulting in *surround suppression* (S. Solomon, White, & Martin, 2002; Bonin, Mante, & Carandini, 2003). In surround suppression, when the

region of the visual field that can evoke a response from a neuron (the *classical receptive field* or CRF) is stimulated at the same time as the region of the visual field that can influence the response of a neuron (the *extra receptive field* or ERF), the response is lower than when the CRF is stimulated alone.

In the above model the surround suppression is computed locally: the response to the linear CRF is divided by the standard deviation (local contrast) of the population of responses in the ERF (Bonin et al., 2005). This assumes the relevant contrast measure is the variance of the stimulus intensities, which is appropriate, because the response of the neurons is invariant to skew or kurtosis of the distribution of intensities (Bonin, Mante, & Carandini, 2006). Models of LGN responses have been used to predict actual responses to artificial and natural stimuli with considerable success (Mante, Bonin, & Carandini, 2008; Dan, Atick, & Reid, 1996).

Primary Visual Cortex

The primary visual cortex receives afferents from the lateral geniculate nucleus mainly onto neurons in layer 4 (Humphrey, Sur, Uhlrich, & Sherman, 1985). As in the retina and LGN, neurons that respond to input from similar areas in visual space are nearby in the cortical tissue (Talbot & Marshall, 1941). The receptive field structures of cortical neurons have been mapped using oriented bars of light (Hubel & Wiesel, 1959; Movshon et al., 1978b). They have also been recovered with the more agnostic methods of reverse correlation based on white noise stimulation (Jones & Palmer, 1987), and regression using natural images (Theunissen et al., 2001; Talebi & Baker, 2012).

Based on their receptive field properties, two broad classes of cortical neurons have been identified: simple and complex cells (Hubel & Wiesel,

1962). Simple-cell receptive fields contain adjacent, elongated excitatory and inhibitory regions. They are phase sensitive, i.e. their responses increase and decrease with the phase of a grating stimulus. Complex cell receptive fields also respond best to elongated stimuli, but they do not contain well-defined on and off regions (Hubel & Wiesel, 1962), and are therefore not sensitive to stimulus phase. Complex cells often exhibit frequency doubling: responses to both stimulus onset and offset (De Valois, Albrecht, & Thorell, 1982; Maffei & Fiorentini, 1973). Both simple and complex cells display orientation and spatial frequency selective behaviour.

Orientation selectivity first appears in the behaviour of visual cortical neurons, apparent from Hubel and Wiesel (1959)'s experiments onward in cat (Hammond & Andrews, 1978), and monkey (Hubel & Wiesel, 1968). Orientation tuning bandwidth is invariant to stimulus contrast (Sclar & Freeman, 1982). Anatomy following functional staining (Willmer & Rutter, 1977; Hubel, Wiesel, & Stryker, 1978) and optical imaging techniques (Grinvald, Lieke, Frostig, Gilbert, & Wiesel, 1986) have revealed in cats and monkeys that cells selective for the same orientations are arranged in columns, and that cells selective for similar orientations are in adjacent columns.

Following Hubel and Wiesel's (1962) observation of sensitivity to stimulus size, measurements of cortical neuron responses using sine wave gratings found that single units are tuned to spatial frequency (Campbell, Cooper, & Enroth-Cugell, 1969; De Valois et al., 1982). Maffei and Fiorentini (1973) measured spatial frequency tuning in the retina, LGN, and cortex, and showed that tuning sharpened with each stage of processing. They were struck by the improvement in tuning and so proposed that the visual cortex was a 'spatial frequency analyser'.

Simple cell receptive fields consist of adjacent, elongated, ON and OFF regions. They are believed to be created by combining centre-surround LGN cell responses, supported by evidence in ferret (Chapman, Zahs, & Stryker, 1991) and cat (Reid & Alonso, 1995; Alonso, Usrey, & Reid, 1996; Chung & Ferster, 1998; Alonso, Usrey, & Reid, 2001). Monosynaptic connections between LGN and cortical simple cells are likely if they share the following receptive field characteristics: position, size, timing, and sign (ON or OFF) (Alonso et al., 2001). LGN responses are thought to be arranged in the primary visual cortex with simple cell ON regions being both excited by ON responses and inhibited by OFF responses, and conversely, OFF regions are excited by OFF responses and inhibited by ON responses (Heeger, 1992b; Hirsch, 2003). This type of model has been termed 'push-pull' (Glezer, Tsherbach, Gauselman, & Bondarko, 1980, 1982).

The shape of the simple cell receptive field is typically modelled with a gabor-type function (Gabor, 1946): a sine wave grating windowed by a two-dimensional Gaussian function (Marčelja, 1980; Daugman, 1985). However, there is evidence that gabor functions do not accurately capture the spatial frequency selectivity observed in cortical neurons. In log units, spatial frequency tuning functions tend to be quite symmetrical, while gabor functions are symmetrical in linear units and have a heavy low spatial frequency bias when plotted in log units of spatial frequency (Hawken & Parker, 1987). Field (1987) proposed that 'log gabor' functions provide a useful alternative to gabor functions. They look very similar in the space domain, but in the frequency domain (on log axes), log gabor functions equally represent all spatial frequencies in natural images, while gabor functions overrepresent low spatial frequencies. Additionally, gabor functions are zero-balanced in odd phase, but in even phase a small mean level is

present. Log gabor functions are zero balanced in both even and odd phase. Regardless of its precise form, the gabor-like filter is typically followed by a half-wave rectification in most models of cortical simple cells (Movshon et al., 1978b; Carandini, 2006). As in the LGN, this rectification is used to account for the lack of spontaneous activity in cortical neurons.

The behaviour of complex and simple cells are alike in many ways, but because the complex cell receptive fields lack discrete inhibitory and excitatory regions (Hubel & Wiesel, 1962; Movshon et al., 1978a), they are not selective for the spatial phase of a stimulus. The popular model of a complex cell is the sum of the responses of several simple cells in misaligned phases (Hubel & Wiesel, 1962; Movshon et al., 1978a; Carandini, 2006), measuring stimulus energy in a particular band of spatial frequency and orientation regardless of phase. Both simple and complex cells have spatially localized classical receptive fields (CRF), within which a visual stimulus can elicit a response.

As in the LGN, cortical neurons are subject to nonlinear influences from within and outside the CRF. At least three major types of nonlinear behaviour are apparent in many cortical neurons (DeAngelis, Robson, Ohzawa, & Freeman, 1992; DeAngelis, Freeman, & Ohzawa, 1994; H. Tanaka & Ohzawa, 2009): (1) Contrast response saturation, (2) surround suppression, and (3) cross-channel suppression. All of these types of suppression can be captured (as in the LGN) using divisive suppression which is based on a pool that includes, or extends beyond, the classical receptive field (Albrecht & Geisler, 1991; Carandini, Heeger, & Movshon, 1997; Sceniak, Hawken, & Shapley, 2001; Cavanaugh, Bair, & Movshon, 2002a; Carandini & Heeger, 2012).

Contrast response saturation. As in earlier visual areas, the firing rate of cortical neurons does not increase linearly with contrast. When the firing rate of a neuron is plotted against stimulus contrast (i.e. a contrast response function), it is evident that response rates saturate quite heavily at high contrast (Albrecht & Hamilton, 1982). This saturation is thought to be the result of contrast gain control, a self-inhibition mechanism that adjusts the operating range of a neuron depending on local levels of contrast (Heeger, 1992b; Carandini, Heeger, & Anthony Movshon, 1999). Models of contrast gain control sum responses from within and immediately around the classical receptive field, and use this aggregate response as a divisor for the CRF response.

Surround suppression. A neuron will respond at its maximum when the entirety of its classical receptive field is covered by a grating of the preferred orientation and spatial frequency. If the neuron exhibits surround suppression, when the size of the stimulus is expanded into the extra receptive field (ERF), the firing rate is reduced. The shape of the ERF is not typically concentric with respect to the CRF as in the LGN, but instead asymmetric with clear regions of strong suppression (DeAngelis et al., 1994; Cavanaugh, Bair, & Movshon, 2002b; Walker, Ohzawa, & Freeman, 1999; H. Tanaka & Ohzawa, 2009). Also unlike the LGN, the suppressive ERF is orientation-selective (Blakemore & Tobin, 1972; DeAngelis et al., 1994), and thus established at least partially in cortex (Carandini et al., 1997; Carandini, 2004). Complex cells and simple cells appear to have equal amounts of suppression (DeAngelis, Ohzawa, & Freeman, 1995; Bonds, 1989), although reports to the contrary exist (Morrone, Burr, & Maffei, 1982). It is believed that surround suppression is the result of divisive normalization of the classical receptive field by the extra receptive field

(Carandini & Heeger, 2012). The ERF is modeled as the sum of neuronal responses in a localized area of the visual field outside the CRF, drawing from a relatively narrow range of orientation selectivities, but a wide range of phase selectivities.

Cross-channel suppression. A neuron will respond well to a grating of the preferred orientation and spatial frequency, and not at all to a grating well outside the preferred range of either of these parameters. However, when the non-preferred grating is presented in conjunction with a grating of the neuron's preferred orientation and spatial frequency, the response to the preferred grating is substantially reduced (De Valois & Tootell, 1983; Bonds, 1989; DeAngelis et al., 1992). This interaction is termed cross-orientation or cross-spatial frequency suppression, because a stimulus which elicits no response on its own can still modulate the response to a stimulus that does elicit a response. This suggests normalization interactions between spatial frequency and orientation selective channels. The functional role of cross-channel inhibition has been debated. It was proposed that it helps sharpen orientation tuning in the cortex (Morrone et al., 1982), supported by an apparent correlation between narrowband selectivity and the presence of suppressive effects (Foster, Gaska, Nagler, & Pollen, 1985). However, DeAngelis et al. (1992) determined that cross-channel suppression contributes negligibly to the sharpening of tuning. Priebe and Ferster (2006) suggest that this inhibition might not occur between orientation and spatial frequency-tuned channels, but instead may be inherited from the LGN. In that case, the ERF would be modeled over the same spatial extent as the CRF but with broadband orientation and spatial frequency tuning.

A clear comparison of the traditional linear model and the modern model of simple and complex cells is given in Rust and Movshon (2005).

Higher Visual Areas

Beyond the primary visual cortex, receptive fields increase both in size and the complexity of their selectivities. In humans and monkeys, following early visual cortical processing (V1/V2), information is divided into two largely independent pathways: the dorsal pathway which processes complex motion patterns, and contributes to motor planning, and object localization; and the ventral pathway which processes shape, colour, and object properties, and contributes to object recognition (Mishkin, Ungerleider, & Macko, 1983).

Early in the dorsal stream, the motion-selective neurons in area MT are relatively well-modeled and understood (Simoncelli & Heeger, 1998; Rust, Mante, Simoncelli, & Movshon, 2006; Nishimoto & Gallant, 2011). Models of higher level neurons are a challenge to construct, but Mineault, Khawaja, Butts, and Pack (2012) have recently made progress modelling the receptive fields of area MST neurons. MST neurons receive inputs from area MT, and are selective to complex motion patterns such as contraction, expansion, and rotation (Duffy & Wurtz, 1991; K. Tanaka et al., 1986). Mineault et al. (2012) model the MST receptive field with a compressive nonlinear sum of inhibitory and excitatory MT inputs.

Often, the output nonlinearities in models of neural responses have been described as being expansive (Heeger, 1992b). However, there is evidence that when outputs of lower visual areas are combined to create more complex selectivities (as from area MT to MST), the sum is less than what would be expected in a linear system (Mineault et al., 2012), rather than greater as an expansive nonlinearity would predict. This suggests that a compressive nonlinearity before summation (or 'sublinear summation') is employed by the nervous system beyond the primary visual cortex. It

has been shown that compressive nonlinear effects on neural responses (such as those caused by gain control or surround suppression) can result in selectivity for conjunctions of properties, such as the conjunction of orientation selective units to form curvature selective units (Peirce, 2007, 2011).

Psychophysics

There are psychophysical phenomena that exhibit many of the same properties as the single unit behaviour described above. Most fundamentally, there is psychophysical evidence for orientation- and spatial frequency-selective channels. These channels can be independently adapted to orientation (Georgeson & Harris, 1984; Bradley, Switkes, & De Valois, 1988) and spatial frequency (Pantle & Sekuler, 1968; Blakemore & Campbell, 1969; Georgeson & Harris, 1984; Bradley et al., 1988). Orientation-specific masking has been observed (G. Phillips & Wilson, 1984), as has a spatial frequency shift aftereffect (Blakemore & Sutton, 1969; Blakemore, Nachmias, & Sutton, 1970).

Cross-orientation masking, like cross-orientation suppression, occurs when two gratings of different orientations are superimposed and the contrast threshold of a target grating is elevated in the presence of an orthogonal, suprathreshold mask (Henning, Hertz, & Broadbent, 1975; Olzak, 1985; Burr & Morrone, 1987; Olzak & Thomas, 1991; Albrecht & Geisler, 1991; Ross & Speed, 1991; Ross, Speed, & Morgan, 1993; Foley, 1994; Meese & Holmes, 2007). Surround masking has been observed when the mask is presented in an annulus surrounding the target grating, or in more complex flanking configurations (Petrov, Carandini, & McKee, 2005).

Role of normalization processes and other nonlinearities

First and most critically, normalization serves to scale a neuron's response so that it will lie within its limited dynamic range of possible firing frequencies, while still signalling changes in a natural world with a dynamic range orders of magnitude larger. Second, tuning functions for spatial frequency (Albrecht & Hamilton, 1982) and orientation (Rose & Blakemore, 1974) have a consistent bandwidth, regardless of stimulus contrast. In a linear system with a threshold to prevent action potentials evoked by non-preferred orientations, the low threshold required to show selectivity at low contrasts would result in barely tuned responses at high contrasts. Contrast gain control forms a necessary (but not sufficient) component to a model that maintains selectivity bandwidth regardless of contrast (Carandini, 2007; Finn, Priebe, & Ferster, 2007). Third, there is evidence that the responses of populations of neurons are more effective at representing a stimulus than any single detector, and that the detectors that are not responding are just as important as those that are (Busse, Wade, & Carandini, 2009). A critical component of a population model that decodes stimulus orientation at the same level of performance as an ideal observer is an intermediate nonlinearity such as normalization (Deneve, Latham, & Pouget, 1999). Finally, normalization may also serve to reduce redundancy in visual coding. Wainwright, Schwartz, and Simoncelli (2002) showed that nonlinear transformations, such as divisive normalization, can make coding more efficient.

1.3 Texture

In this section, I will examine texture as a particular class of visual stimuli, touching on what sorts of images are typically considered 'textures', how texture might be represented by the visual system, and some types of visual tasks in which texture is an important property.

1.3.1 What is texture?

What precisely is considered visual texture is typically approached with either a great degree of rigour but limited applications, or more commonly a subjective assessment with each author having a slightly different dividing point in the dichotomy of texture versus non-texture patterns or applying different criteria. Portilla and Simoncelli (2000) took a quantitative approach and defined texture as any pattern described well enough by the statistics accounted for in their model to be synthesized with a high enough fidelity to be perceptually equivalent. Most authors have taken the route of a qualitative description, for example, Bergen and Adelson (1991) defined texture as the emergent statistical properties of an ensemble of texture elements. This sort of reasoning is common and descends from the Gestalt history of considering how perceptual systems group smaller units to form a larger perceptual whole (Beck, 1966; Olson & Attneave, 1970). Wilkinson and Wilson (1998) found that the perception of a pattern as a textured surface versus an array of discrete elements was related to element spacing, suggesting that it is likely that the dividing line between textured and non-textured surfaces is density-dependent. In this thesis, I remain agnostic to the limits of the set of patterns that can be defined as texture, and instead perform quantitative tests for spatial homogeneity as described in Chapter 2. Many types of textures have been

used as visual stimuli in recent decades. Here, I will review some of the most common.

Grid-based patterns with statistical constraints: This type of texture consists of a grid, in which each square is given a grey-level value depending on some rule defining the spatial correlational statistics of the texture (Julesz, 1962; Julesz, Gilbert, & Victor, 1978). In these textures, the distribution of grey values, sometimes binarized to black and white, depends on pixel histogram statistics (e.g. variance) (Julesz, 1962). Alternatively, the texture may consist of a noise background with some repetitively embedded pattern (Julesz, 1981a, 1981b). The higher-order statistics of these textures have been manipulated using a recursive formula that depends on n neighbouring pixels, where n is the order of statistic being manipulated (Julész, Gilbert, Shepp, & Frisch, 1973). These types of textures are historically important because they were employed in the first attempts to apply signal processing techniques to texture perception (Julesz, 1962; Bergen & Adelson, 1991), and in fact the first attempts to view texture as statistical vision. However, they have since largely fallen out of modern use due to a lack of biological relevance, and the emergence of alternative methods for manipulating texture statistics, evident in more modern types of textures described below.

Micropattern textures: This is a very commonly used class of textures with several sub-types, typically defined by micropattern type (dots, gabors, complex features), and micropattern arrangement (gridded, randomly positioned). In early studies, a micropattern was typically a complex feature element (Beck, 1966; Olson & Attneave, 1970; Julesz, 1986; Caelli & Julesz, 1978; Julesz, 1980). These elements were reasoned to be local features that were conspicuous because of their role as fundamental features extracted by

early visual processing, termed 'textons' by Julesz (Julesz & Bergen, 1983; Julesz, 1984). With the linear filtering models of early visual processing, it became more common to use texture elements matched to the shapes and selectivities of early visual neurons, such as gabors (Caelli & Moraglia, 1985), or the orientationally broadband and luminance-balanced difference-of-gaussians (DoGs). (Motoyoshi & Kingdom, 2007). Finally, some work on texture perception used element-arrangement patterns, where the layout of the micropatterns is not randomized, or gridded, and is itself an important parameter. The micropatterns might be placed so that their orientation or spatial frequency alternates in rows, columns, or a checkerboard pattern (reviewed in Beck (1983); Wolfson and Graham (2005)).

Micropattern textures are probably the most commonly used class of textures, and have been particularly useful in the field of texture segmentation. The synthetic textures used in this thesis are examples of micropattern textures.

Filtered noise: This is an alternative approach to creating textures with specific spectral characteristics, spurred on by the assumptions of the energy model of segmentation (as discussed in the next section). A random noise texture is typically filtered to a specific band of spatial frequency and/or orientation, and then may be modulated to create a boundary (e.g. Dakin & Mareschal, 2000; Landy & Oruç, 2002; El-Shamayleh, Movshon, & Kiorpes, 2010). Sometimes the amplitude spectrum of the noise is shaped to a specific slope (e.g. Schofield & Georgeson, 2003). It is mainly used by researchers studying segmentation, but it has also been used for texture discrimination (F. Phillips & Todd, 2010). These types of textures are the successors to the approach of studying vision using sine wave gratings. They are powerful stimuli for two main reasons: their phase spectra are random,

so they contain no spatial structure that might confound results, and it is easy to create an infinite number of stimuli with the same statistical properties.

Natural Textures: These textures are digital images of textured surfaces extracted from photographs of natural scenes, and are frequently employed in texture synthesis work (Portilla & Simoncelli, 2000; Zhu, Guo, Wang, & Xu, 2005), but rarely psychophysically (unlike photographs of entire natural scenes). However, natural textures are powerful stimuli, and some degree of parametric control is possible. For example, structural information may be gradually reduced by perturbing the phase spectrum (e.g. Emrith, Chantler, Green, Maloney, & Clarke, 2010). In the first data chapter of this thesis, I use photographs of natural textures as stimuli. These textures are appealing as stimuli because they are representative of the natural visual environment, but are limited in their use because they are difficult to parameterize, and developing a sufficiently large library is labour-intensive.

Synthesized Textures: A popular way of examining texture statistics, with many practical applications, is to employ an 'analysis-synthesis' algorithm. These algorithms take as input a representative sample of an exemplar texture, measure its properties, and using those measurements produce fresh samples of the 'same' texture. The extent to which the new samples appear to be the same depends on the sufficiency of the statistical measurements. There are several algorithms of this type, though none as widely used to generate texture stimuli as the Portilla and Simoncelli (2000) algorithm. The outputs of this process are sometimes used as stimuli with all or a subset of the available statistical constraints (e.g. Balas, 2006), or as intermediate cases between a natural texture and one synthesized without a complete set of constraints (e.g. Ziemba, Freeman, Movshon, & Simoncelli,

2012). Using synthesized natural textures it is much easier to build large libraries of varied stimuli.

1.3.2 Texture as statistical vision

A common thread throughout most work on texture is the idea that a texture image is not one-of-a-kind, but rather one instance of a class of possible images, constrained by statistical regularities rather than precise feature positions within a given instance of that class (Victor, 1994). The question of what statistics are required to fully describe a texture forms the foundation of the entire field of texture synthesis, but is also of interest to vision scientists in a slightly modified form: 'What image statistics capture the attributes of a texture that are relevant to perception?'.

Image statistics used to describe textures

When textures are described by their statistics, it is important to be clear about exactly what statistics are being used. This has been approached in a variety of ways, but following is a summary of the most notable. These are conceptually distinct, but in practical use are not mutually exclusive.

Julesz order statistics: First suggested by Julesz (1962), these texture statistics based on correlations between image pixels had a profound impact on the way texture was approached. 'First-order' statistics refer to the probability of any given pixel having a particular intensity value, which in practice determines the mean luminance of the image. Higher-order correlations increase the complexity of the spatial relationships between intensity values. For example, second-order statistics refer to the likelihood of a pixel having a given intensity value, contingent on the value of one other pixel (at any arbitrary spatial offset). Textures that are created from the same second-order probability distribution on average will have the

same autocorrelation function, and therefore the same Fourier amplitude spectrum. Third- and fourth-order correlations depend on two and three neighbouring values respectively.

Moments of a distribution: In many fields of study, the 'order' (n) of a statistic refers to the moment of the random distribution that is being measured (Kingdom, Hayes, & Field, 2001). For any distribution, the nth moment can be calculated as $\sum_{i=1}^{N}(x_i - \mu)^n/N$, where x_i is the ith pixel intensity, N is the number of sampled pixels, and μ is the mean pixel intensity (Bendat & Piersol, 1966). As in Julesz order statistics, the first moment of a distribution of image intensities is the mean luminance, computed as $\sum_{i=1}^{N} x_i/N$. Variance ($n = 2$) measures the spread of the intensity distribution around the mean. In terms of image appearance, this affects the contrast (Bex & Makous, 2002); a low contrast image will have its intensity values clustered around the mean, whereas a higher contrast image will have a much larger spread. The skewness ($n = 3$) of the distribution refers to whether the values in the pixel distribution are evenly divided on either side of the mean, or if there are a greater number of relatively light or dark intensities. Motoyoshi, Nishida, and Adelson (2005) have shown that skewness affects perception of surface properties such as gloss. Kurtosis ($n = 4$) describes the the extent to which the distribution is peaked around the mean, which in practice measures whether the variance is due to many moderate deviations from the mean, or relatively fewer extreme deviations. Perceptually, kurtosis has been shown to correspond with texture density (Kingdom et al., 2001), where distributions with a high peak and heavy tails are perceived as less dense. The basis (x) values for which the distribution is defined is very important. For example, a distribution of pixel intensity values may be less relevant to visual perception than the

distribution of gabor wavelet responses which mimic early visual cortical representation (Kingdom et al., 2001).

Feature correlations: Rather than capturing the contingencies between specific intensities, these statistics use more complex pre-processing (typically the outputs of linear filters), and define image 'features' in a variety of ways, but always dependent on the presence of multiple filter responses. Martin, Fowlkes, and Malik (2004) and Arbelaez, Maire, Fowlkes, and Malik (2011) use clustering to find sets of filter responses that frequently co-occur in spatial position, while Portilla and Simoncelli (2000) and Freeman and Simoncelli (2011) use local averages of products between pairs of filters in a bank of gabor filters, and scale the size of the neighbourhood depending on eccentricity. Feature correlations are similar to pixel correlations, but they are more physiologically relevant because they frame spatial information in a way that could be (and occasionally is) implemented in the context of generally accepted models of early vision.

Fourier spectrum: The Fourier transform and its usefulness in understanding visual processing was covered in the previous section. The amplitude spectrum corresponds to second-moment statistics, which can be measured given two points of reference (contrast, orientation, and spatial frequency). The phase spectrum corresponds to higher-order statistics, which require more than two points of reference for measurement (Oppenheim & Lim, 1981; Thomson & Foster, 1997). This definition of order statistics is appealing because it relates directly to the above definition (moments), and the broader study of texture boundaries. Like moment measurements and feature correlations, it generalizes easily to any type of

texture image because the Fourier transform can be applied to any two-dimensional signal. Additionally, it has a straightforward relationship to results obtained using traditional visual stimuli such as sinewave gratings.

Metamers, crowding, and the texture of a scene

The term 'metamer' is borrowed from colour vision where it is used to describe a pair of stimuli with different spectral components, but perceptually identical colours. Generally, metamers are stimuli that are not physically identical and yet appear the same, because they generate equivalent responses in some fundamental sensory representation. This property means that metamers, when they can be found, may provide useful stimuli for probing how information is preserved or discarded by the visual system.

While the term is not commonly used in the texture literature, the methodology of looking for perceptually indistinguishable stimuli forms a large part of the history of texture processing. Much of Julesz' work was focused on finding 'metamers' based on segmentation rather than appearance, and defined by orders of spatial correlational statistics (Julész et al., 1973). Richards and Polit (1974), looked for metamers in texture appearance by using sums of sine waves, and found that the presence of different spatial frequency components resulted in perceptually distinct textures, yet differences in the phase relationships between those components were metameric.

Our understanding of textures is constrained to those aspects of the visual world that appear to be texture-like in photographs: relatively homogeneous collections of like elements viewed at a sufficient distance to be considered 'stuff' (Adelson, 2001). However, there is evidence that scenes are themselves represented as 'summary statistics' (Balas, Nakano, & Rosenholtz, 2009), and in some circumstances (e.g., peripheral viewing) their

representation may be indistinguishable from what is traditionally considered texture (Oliva & Torralba, 2001; Lazebnik, Schmid, & Ponce, 2006; Walker Renninger & Malik, 2004). Oliva and Torralba (2007) propose that local scene details are abstracted out to form regions of relatively homogeneous texture, while only the global structure of the scene is represented in perception of scene 'gist'.

A longstanding observation about the visual system has been the underrepresentation of information in the visual periphery (Flom, Weymouth, & Kahneman, 1963; Andriessen & Bouma, 1976). When low-level explanations, such as differences in acuity, have been accounted for, this phenomenon is commonly termed 'crowding', because the presence of additional objects in the periphery distorts the appearance of objects that are identifiable in isolation (e.g. Levi, 2008). Crowding effects may be predicted by the same spatial correlations used in summary-statistical representations, suggesting that representation of the visual scene outside the fovea is statistical (Parkes et al., 2001; Freeman & Simoncelli, 2011). Balas et al. (2009) also show that a model using spatially localized summary statistics accounts well for effects of crowding. These results suggest that some aspects of visual processing are carried out in a statistical manner that we have typically associated with texture, and that the inferences we make about texture processing might actually have far broader applications.

1.3.3 How is texture used by the visual system?

Identification of 'stuff': Texture appearance can be helpful for identifying collections of things (DeValois & DeValois, 1990; Adelson, 2001), and recognizing surface material. This is particularly true when shape information is unhelpful: for example, when discerning long from short grass at a distance.

Surface affordance: Texture appearance, and the consequent inferences we can make about surfaces help with motor planning (Watt, 1995; Gibson, 1977, 1950). This is related to identification, but in this case texture is associated with its utility for possible subsequent actions by the observer. Is the surface wet? slippery? rugged? How we interact with different kinds of surfaces is often based on their visual texture (Adelson, 2001).

Depth, and depth gradients (slant): Texture gradients have long been considered a powerful monocular cue for depth (Knill, 1998), as well as surface slant (Marr, 1982).

Three-dimensional Shape: The projection of texture on 3-D objects is distorted in the eye's image plane, both because of slant and depth, and also because of occlusion (Aloimonos, 1988). These distortions contain information on object shape (Blake, Bülthoff, & Sheinberg, 1993), and it is believed that the gradients of orientation energy introduced by these distortions is particulary informative about 3D shape (Li & Zaidi, 2000).

Disambiguating surface changes from illumination changes: Contrast, one of the most basic texture properties, has been shown to have utility in disambiguating image gradients arising from illumination effects (such as shading) from those caused by changes in surface properties (Schofield, Rock, Sun, Jiang, & Georgeson, 2010).

Segmentation: Changes in texture provide a powerful cue for segmenting surface discontinuities. The ability of texture to support segmentation is the primary subject of this thesis, and will be discussed in detail in the next section.

1.4 Texture Segmentation

As the term is used in this thesis, *boundary segmentation* refers to a pre-attentive, involuntary process that detects and characterizes discontinuities defined by changes in some image property such as luminance, colour, stereo disparity, motion, or–of primary interest here–texture. Texture properties that enable segmentation include (but are not limited to) contrast (Burton, 1973), orientation (Nothdurft, 1991; Landy & Bergen, 1991; Kingdom, Keeble, & Moulden, 1995; Motoyoshi & Nishida, 2001), and spatial frequency (Caelli & Moraglia, 1985).

Psychophysical tasks to test segmentation performance include: discrimination of modulation spatial frequency (Smith & Scott-Samuel, 2001) or orientation (Morgan, Mason, & Baldassi, 2000), as well as detection (Landy & Bergen, 1991; Beck, 1983), identification (Nothdurft, 1985), or localization (Rubenstein & Sagi, 1996) of embedded shapes. For example, Caelli and Julesz (1979) created micropattern textures differing in dipole orientation variance, and presented pairs of textures as the left and right halves of a square stimulus. The magnitude of the difference in micropattern orientation variance was used to make the task easier or more difficult. To test the effect of micropattern density, they asked subjects whether two textures were present, or if the square was one homogeneous texture, and measured orientation variance thresholds for a range of densities. They found that thresholds were higher as density was increased. Caelli and Julesz (1979) did not make a distinction between texture segmentation and discrimination (see below) but the task and spatial relationship between the textures (abutting in each target stimulus) ensured that boundary segmentation mechanisms were being tested.

It is important to note that texture boundary segmentation should not be confused with texture *discrimination*. Discrimination is an appearance-based judgement that gauges how similar two textures appear, which is a task that may not necessarily rely on the same mechanism as segmentation (Northdurft, 1997; Landy & Graham, 2004), in spite of how the two terms have often been used interchangeably (e.g. Caelli & Julesz, 1978). This is evident when one considers how textures that appear very different, for example those composed of T- versus L-shaped micropatterns, do not produce a visible boundary when rendered side-by-side (Beck, 1966). Studies of texture discrimination are usually based on the statistics of the texture, with novel samples being generated for individual trials (e.g. Balas, 2006), though sometimes the same texture samples are used repeatedly, to measure how the features they contain are learned and used for discrimination (e.g. F. Phillips & Todd, 2010).

Objective criteria used to test discrimination include just-noticeable-difference measurements (Morgan, Chubb, & Solomon, 2008) and oddball detection (Balas, 2006) in addition to subjective criteria, such as visual inspection (Portilla & Simoncelli, 2000). Often, textures are presented over multiple intervals so that they may not be compared side-by-side.

1.4.1 Feature-based approaches

Early studies of texture perception noted that some pairs of textures, even those distinct in appearance, were segmented effortlessly while others required inspection (Julesz, 1962; Beck, 1966). Much effort was targeted towards characterizing texture pairs based on whether or not they supported segmentation, and then trying to divine the statistical rules that united the segmentable textures while excluding those that were not effortlessly segmented (Julesz, 1984 contains a summary). Early research on texture

segmentation suggested that certain texture statistics resulted in clusters of like-luminance pixels (Julesz, 1962). These clusters resembled image features matching hypothesized 'feature-detectors' in visual cortex, such as the report of neurons that appeared to behave like 'oriented line detectors' in primary visual cortex of the cat (Hubel & Wiesel, 1959).

How such feature clustering might occur was a problem that has only been approached recently. For example, Martin et al. (2004) propose a clustering/classification algorithm for texton extraction for segmentation in three stages: 1) detect the features, 2) link the features to form texture elements, 3) compare the texture elements present in many adjacent regions, 4) localize the boundaries. However, this and other clustering algorithms are not (and do not attempt to be) biologically plausible.

1.4.2 Energy-based approaches

Filter-rectify-filter model

Several researchers noticed a correspondence between the properties of some of the features that supported segmentation, the passbands of gabor filters, and the resemblance of gabor filters to early visual receptive fields (Bergen & Adelson, 1988; Jain & Farrokhnia, 1991; Fogel & Sagi, 1989; Malik & Perona, 1989, 1990).

A family of models constructed from differences of filter responses developed a consistent format similar to those models used to detect stimulus properties not evident in the amplitude spectrum (i.e. Malik & Perona, 1989). These models use a basic architecture consisting of two stages of filtering, separated by a rectifying nonlinearity (filter-rectify-filter, or FRF). The FRF model is able to detect differences of Fourier energy across a boundary (Bergen & Landy, 1991). This style of model is very widely used, and explains both neurophysiological (e.g. Zhou & Baker, 1994)

and psychophysical (e.g. Bergen & Adelson, 1988) results (Baker, 1999; Landy & Graham, 2004).

First-stage Filter: The first stage of filtering uses a bank of linear (often gabor) spatial filters (Heeger & Bergen, 1995) that span all orientations and a range of relatively high spatial frequencies, which extracts information about the texture's spatial frequency and orientation content (represented by which filters have the highest responses), and information about its contrast (conveyed by the strength of those responses). This stage is illustrated in Figure 1–6 A and can be thought of as being instantiated by the responses of early cortical neurons.

Rectify: There will be little difference in the mean response of the filter bank over an image containing a texture boundary, so linear filtering alone is not enough to detect these boundaries. However, a difference that might be detected is the difference in magnitudes of the responses. Regions with intensely positive responses also contain complementary intense negative responses due to the zero-balanced nature of the first-stage filters. Mutual cancellation of these opposite-sign responses is prevented by a nonlinearity, for example a rectification that either sets all negative values to zero (half wave) or converts them to positive responses of the same magnitude (full wave). The rectification traditionally takes the form of a square law, as a measure of image energy (Heeger, 1992a), which is relevant to the goal of analyzing the Fourier energy across the boundary of the texture. However, any nonlinearity that creates a difference in the mean responses across a texture boundary could enable segmentation by an FRF model. This stage is depicted in Figure 1–6 B. As discussed in section 1.2.2, rectifying nonlinearities are a pervasive attribute of visual processing in the LGN and visual cortex.

Second-stage Filter: The second stage of filtering responds to the difference in early-stage responses following the rectification. These filters must be at a coarser scale (to respond to the texture change rather than the details of the texture itself). Which filters in this bank are relevant depends on the task and stimulus. This stage is depicted in Figure 1–6 C.

The filter-rectify filter model predicts the following: 1) Boundaries between textures differing in contrast, orientation or size of elements should be segmentable; 2) If the rectification is a square-law nonlinearity, textures should not be segmentable if the energy in their Fourier amplitude spectra is the same (Bergen & Landy, 1991).

Many aspects of the model remain unspecified, such as how the early-stage channels are integrated (Mussap, 2001), what normalization processes affect the model at which stages (Ellemberg, Allen, & Hess, 2004; Wang, Landy, & Heeger, 2011), and the precise nature of the intermediate nonlinear rectification (J. Solomon & Sperling, 1994; Graham & Sutter, 1998). Beck, Sutter, and Ivry (1987) concluded that spatial frequency analysis is capable of performing texture segregation in many cases, but also noted that feature-level differences can also lead to segregation in the absence of mean differences in spatial frequency, orientation, or contrast. This seems problematic for the FRF model as it is currently conceived, because while it can and does accurately account for results in many circumstances, accounting for structural information has been shown to require more complex processing. For example, Graham, Sutter, and Venkatesan (1993) found that an additional layer of rectification and filtering was required to account for segmentation of textures that differed only in the arrangement of their elements.

1.4.3 Second-Order Vision

When sine wave gratings of slightly different spatial frequencies are added together, 'beats', periodic alternating regions of high and low contrast, appear to the viewer even though they are not represented in the Fourier spectrum (Henning et al., 1975). These contrast beats are visible, even though this seems surprising from the perspective of the independent channels model, because such a model predicts that only those frequencies present in the stimulus should be visible. It was apparent that some nonlinear process in perception was required to see these patterns (Burton, 1973). Human psychophysical studies showed that beat detection is not driven by early, 'trivial' nonlinearities as might occur in a computer monitor or in the photoreceptors, but nonlinearities at some stage of processing within the nervous system (Derrington & Badcock, 1986). It was also shown that the mechanism for beat detection was different from the mechanism used to detect modulations in luminance (Derrington & Badcock, 1985).

Beats appeared to be modulations in the amplitude of a luminance pattern, suggesting the creation of a new, more general class of psychophysical stimuli containing amplitude modulations. These stimuli were earlier called 'non-Fourier', later called 'second-order' (not to be confused with the 'order statistics' discussed earlier). Second-order stimuli contained a relatively high spatial frequency pattern, termed the 'carrier', which might be a simple sine wave grating or a complex texture. A second, 'envelope' pattern (which might also be a sine wave grating, or some other pattern) of a lower spatial frequency was applied multiplicatively as a contrast modulation. The envelope captured the appearance of the 'beats' evident in earlier stimuli, but could be used with both simpler grating carriers and more complex texture stimuli instead of being restricted to the addition of two luminance gratings.

Indeed, any texture boundary can be defined as a modulation in one or more texture properties.

Perception of both second-order texture boundaries and beat detection require a nonlinear mechanism. It is likely that the same or similar processes are responsible for the detection and segmentation of most types of discontinuities that rely on second-order comparisons. In this way, a number of historically disparate areas of research based on a variety of signals (beats, non-Fourier or second-order motion, texture segmentation, illusory contours) could be understood using a common mechanism: second-order processing. In this context, the filter-rectify-filter model is a natural candidate. The first stage of filtering is selective for the carrier pattern and the second stage of filtering is selective for the envelope pattern, so one can explicitly test the first-stage filter properties by varying the carrier, and examine the second-stage filter properties by varying the envelope.

Properties of Second-Order Mechanisms

In this section, I will review what has been inferred about the mechanisms that perform second-order segmentation, pulling from data obtained through neurophysiology and psychophysics, with an emphasis on how these findings relate to a model of second-order vision. I will examine: (1) the extent to which second-order segmentation mechanisms are independent from those that segment first-order boundaries; (2) invariance to different types of second-order cues; (3) the tuning properties of the first-stage filters; (4) the tuning properties of the second-stage filters; and finally (5) the relationship between the tuning properties of the first- and second-stage filters.

Independence: Mechanisms of second-order segmentation are functionally distinct and independent of first-order edge detection mechanisms. They are considered independent psychophysically because noise in one channel

does not mask the other (Schofield & Georgeson, 2003; Allard & Faubert, 2007), and because psychophysical adaptation does not transfer between first and second order stimuli (Langley, Fleet, & Hibbard, 1996). In fMRI studies, adaptation does not transfer from luminance to orientation modulations (Larsson, Landy, & Heeger, 2006). The sensitivity functions for first- and second-order vision are quite different, with the classic contrast sensitivity function peaking around 4 cycles per degree (Campbell & Robson, 1968) and the modulation sensitivity function being more low-pass (Sutter, Sperling, & Chubb, 1995). Neurophysiological evidence, on the other hand, demonstrates that second-order responsive neurons are also selective to first-order modulations, and that the preferred orientation (Mareschal & Baker, 1999) and, to some extent, spatial frequency (Zhou & Baker Jr, 1996; Mareschal & Baker, 1999), for these two kinds of stimuli are tightly correlated. Sensitivity to second order stimuli is lower than to first order in both psychophysical (Schofield & Georgeson, 1999) and neurophysiological (Ledgeway, Zhan, Johnson, Song, & Baker, 2005) contexts. In a model of single-unit responses, this is accounted for with two pathways: a linear pathway that conforms to the typical model of simple-cell responses, and a nonlinear pathway with some type of filter-rectify-filter arrangement (Zhou & Baker Jr, 1996).

Cue Invariance: There is some question whether the same cue-invariant mechanism serves all types of second-order boundaries, or if different boundary types are segmented by different mechanisms. Based on psychophysical evidence, Motoyoshi and Nishida (2004) suggest that orientation and contrast-defined boundaries are segmented by different, but ultimately connected pathways. Song and Baker (2007) showed that individual neurons

can segment different types of second-order boundaries (contrast modulations and illusory contours). Given an appropriate bank of first-stage filters, the FRF model is capable of segmenting boundaries defined by a variety of cues (i.e. spatial frequency, orientation, illusory contours, and contrast).

First stage filter tuning: Psychophysical measurements of carrier frequency dependence have generally found that there is no specific tuning to spatial frequency (Schofield & Georgeson, 2003) or orientation (A. Arsenault, Wilkinson, & Kingdom, 1999), though there have been observations of broad orientation tuning (Langley et al., 1996). This suggests either that 1) the first stage filters feeding into each second-stage filter have diverse preferred spatial frequencies and orientations, or 2) each instance of a second-stage filter is tuned to a specific carrier, but that tuning varies across the population of responses. This is connected to studies of cue invariance in an important way: an FRF model with a limited range of orientations in the bank of first stage filters will have gaps in its ability to segment orientation-defined boundaries, but no gaps in its ability to segment contrast-defined boundaries. For this reason, and because there is little psychophysical evidence for orientation tuning, FRF models aimed at capturing psychophysical performance typically cover all orientations in the first bank of filters, but at relatively high spatial frequencies.

Single unit recordings in visual cortex found that second-order neurons were always tuned to the carrier spatial frequency. In addition, they were sometimes tuned to the orientation of the carrier texture, but the bandwidth of that tuning was highly variable from one neuron to another (Mareschal & Baker, 1999; Song & Baker, 2006). This suggests that an FRF model of a single unit response should use a smaller range of first-stage filter frequencies, selective for a variable range of orientations, depending on the

neuron being modelled (Mareschal & Baker, 1999). Testing these neurons on orientation-defined boundaries would help to determine whether neurons with broad carrier selectivity have a pool of many tightly-tuned orientation-selective filters, or fewer broadly tuned first-stage filters.

Second stage filter tuning A. Arsenault et al. (1999) measured the psychophysical tuning of second-order segmentation mechanisms using a masking paradigm like those used to measure tuning in first order vision. They found that second-stage filters were tuned both to orientation and spatial frequency, though in both cases this tuning was quite broad.

Neurophysiological results show consistent selectivity for the orientation and spatial frequency of the second-order boundary (Baker, 1999), typically (as mentioned previously) at the same preferred orientation and similar spatial frequency as the neuron's first-order responses (Zhou & Baker Jr, 1996; Mareschal & Baker, 1999). This tuning is characterized in FRF models by using an orientation and spatial-frequency selective filter in the second stage.

By definition, the spatial frequency of the carrier (and thus the filters selective for it) is higher than that of the envelope. There is no neurophysiological evidence of a systematic coupling (e.g., fixed ratio) between the preferred carrier and envelope spatial frequencies in single units (Mareschal & Baker, 1999; Song & Baker, 2006), though the preferred spatial frequency of the carrier has consistently been found to be much higher than that of the envelope (Mareschal & Baker, 1999).

In single units, a consequence of a suppressive ERF is sensitivity to contrast modulations (H. Tanaka & Ohzawa, 2009). This is not to be confused with the segmentation performed by the filter-rectify-filter mechanisms inferred by the research presented here. The ERF-mediated

segmentation differs from FRF-mediated segmentation in two important ways: 1) FRF-mediated segmentation evokes responses from contrast modulations within the classical receptive field, and 2) the ratio of carrier to envelope spatial frequency is between 6:1 and 40:1 for FRF-mediated segmentation (Mareschal & Baker, 1999), but only around 2:1 in ERF-mediated segmentation (H. Tanaka & Ohzawa, 2009).

1.5 Organization of the Thesis

This thesis first uses photographs of natural textures to psychophysically demonstrate that texture segmentation mechanisms are dependent on higher-order statistics, then uses naturalistic synthetic stimuli to examine how specific statistics influence and enable segmentation. An FRF-style model is introduced and modified to account for the role of global texture statistics in texture orientation, contrast, and structure.

In the first study (Chapter 2), photographs of natural textures, and the same photographs phase-randomized, are employed as textures over which contrast modulations are applied. Human segmentation thresholds are found to be much lower for the phase-randomized textures, implying that the presence of higher-order structure impairs the performance of segmentation mechanisms. Based upon a rank-ordering of the textures according to the magnitude of the threshold reduction, sparseness and local edge structure appear to be candidates for higher-order statistics that might impair the segmentation of contrast boundaries.

In the second study (Chapter 3), naturalistic synthetic textures are created to assess individual image statistics experimentally. Segmentation thresholds for contrast and orientation are compared between textures whose local edge structure and sparseness are systematically manipulated. These experiments demonstrate that texture sparseness and global phase structure impair orientation and contrast boundary segmentation, but that local phase structure has little impact on segmentation. The experiment is simulated using a standard filter-rectify-filter model, which is optimized with a compressive nonlinearity to accommodate the psychophysical data.

In the third study (Chapter 4), the same synthetic textures are paired to determine the extent to which differences in texture structure can enable

segmentation. The results suggest that global phase structure and texture sparseness can both enable and influence segmentation in the absence of any lower-order statistical cues for segmentation. The same model employed for contrast and orientation boundary segmentation data in the previous experiments is found to also account for the role of global structure and density, but not the role of local phase alignment in these kinds of texture boundaries.

Finally, the impact of these results, some of the logistical issues we encountered while modelling these results, and future areas for potential study are considered.

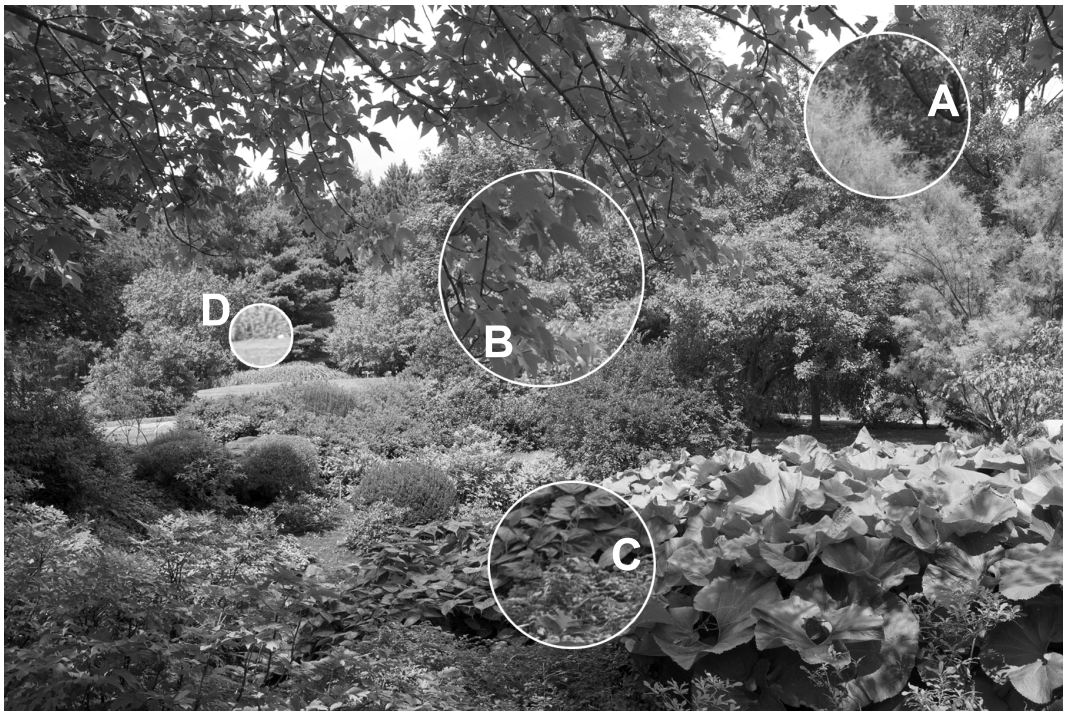

Figure 1–1: Sample of the types of boundaries that occur in natural scenes. Highlighted boundaries are defined by: (A) luminance and texture, (B) luminance, contrast, and texture scale, (C) mainly texture, (D) texture, luminance and contrast.

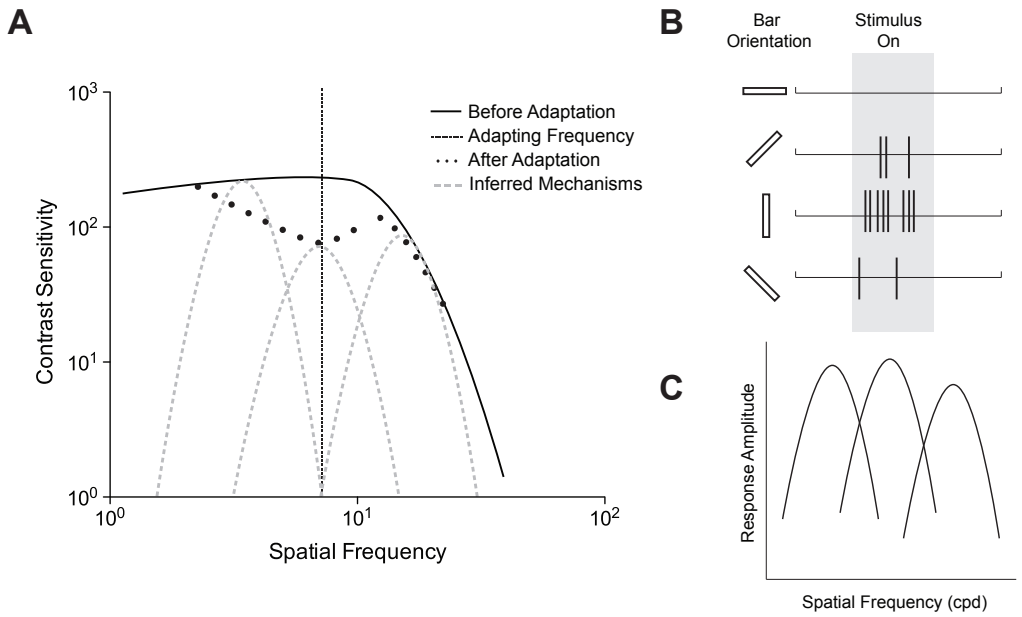

Figure 1–2: Evidence for spatial frequency- and orientation-selective processing in early vision. (A) Adaptation of a single spatial frequency channel. Pictured is a schematic of the results observed by Blakemore and Campbell (1969). The solid line shows sensitivity to gratings of a wide range of spatial frequencies. The dots show the sensitivity at a range of spatial frequencies following adaptation to a grating of the spatial frequency marked by the thin dashed line. Grey dashed curves show the spatial frequency-selective channels inferred from these results. The sensitivity of the channel closest to the adapted spatial frequency has been reduced. (B) Orientation selective neuron in visual cortex. This panel illustrates the type of results observed by Hubel and Wiesel (1959). The dashed lines in the shaded area show stimulus-specific responses to bars of light at different orientations (pictured at right). The neuron depicted prefers vertical stimuli. (C) Spatial frequency selective units in visual cortex. Each curve depicts the response amplitude of a different cortical simple cell. Each neuron is tuned to a different spatial frequency. Adapted from Maffei and Fiorentini (1973).

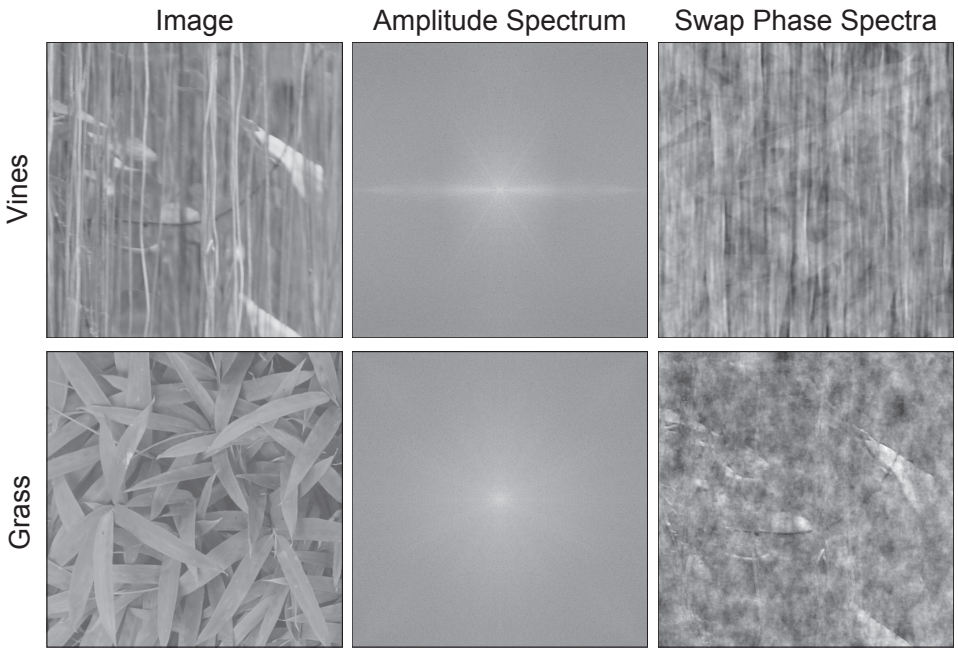

Figure 1–3: Relative contributions of amplitude and phase spectra to texture appearance. The 'Image' column shows two texture photographs (vines and grass), and their amplitude spectra are illustrated in the centre column. The grass is very broadband for orientation, and while the vines are broadband, most of the energy in the image is located in the vertical dimension (horizontal in the frequency domain). In the final column, the phase spectra are swapped between the two images. In the top image, the amplitude spectrum of the vines is paired with the phase spectrum of the grass. In the bottom image, the amplitude spectrum of the grass is paired with the phase spectrum of the vines. The vertical energy of the vines is still very apparent in the resulting image, though the continuous contours of the vines themselves are no longer present. The structure of the grass is visible, though faint, as relatively little energy is available to provide these features with contrast. In the bottom image, the leaves from the texture of vines appear to have been transplanted into a field of broadband noise.

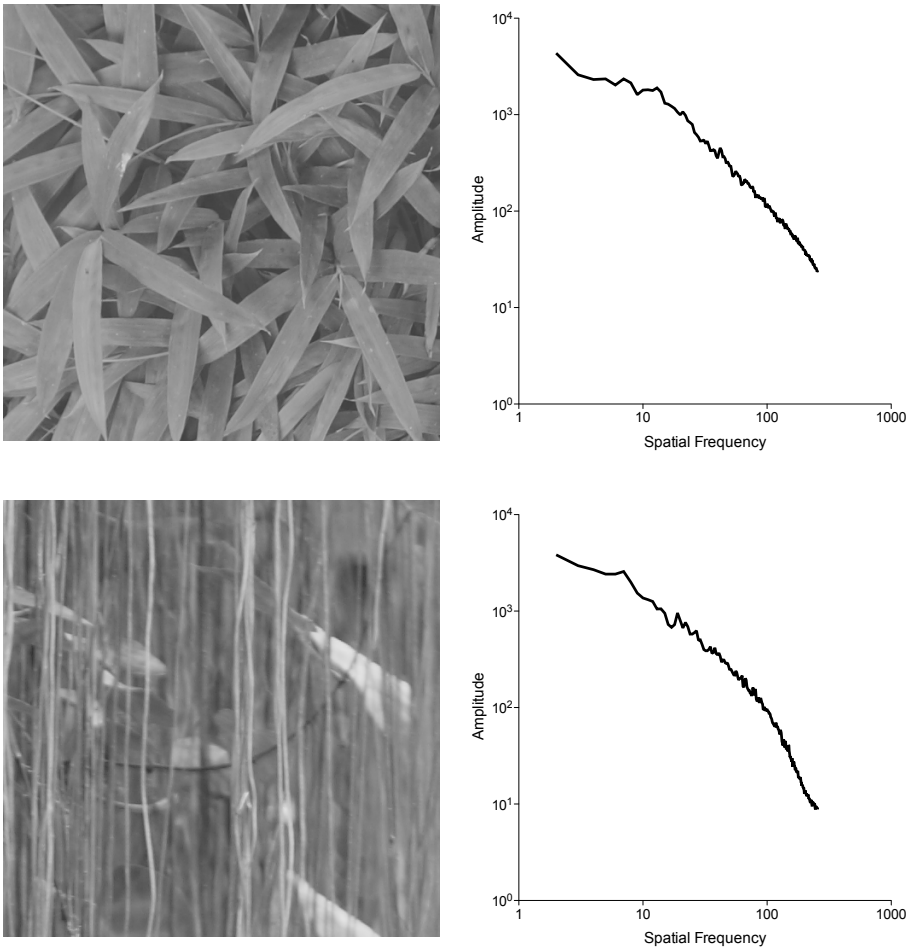

Figure 1–4: Natural textures and their spatial frequency amplitude spectra. The amplitude for each spatial frequency component is plotted on the graph. These graphs are the same as those in the centre column of Figure 1–3, but averaged over the rotational axis (orientation). An inverse relationship between amplitude and spatial frequency with the most power in the lowest spatial frequencies is evident. There are some individual differences between textures, but the trend for a $1/f$ energy falloff is consistent.

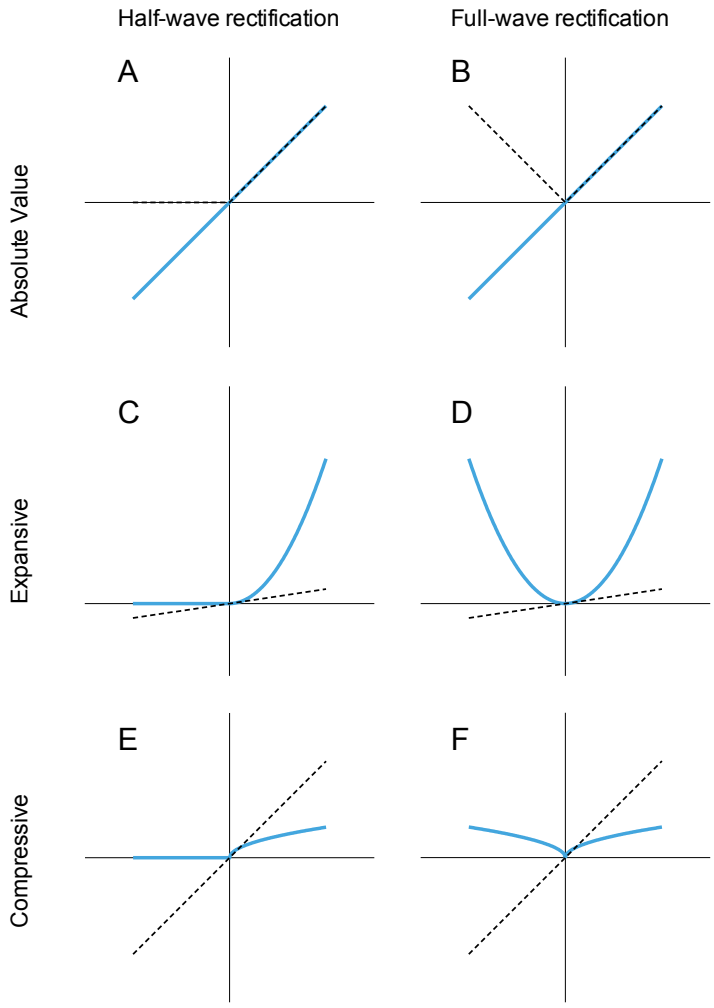

Figure 1–5: Different basic rectifications. The linear input signal is shown as a dashed line with a slope of 1, and the output is shown in solid grey. The columns compare half- and full-wave rectification, and the rows show (from top to bottom): compressive (square root), linear (absolute value), and expansive (square law) nonlinearities.

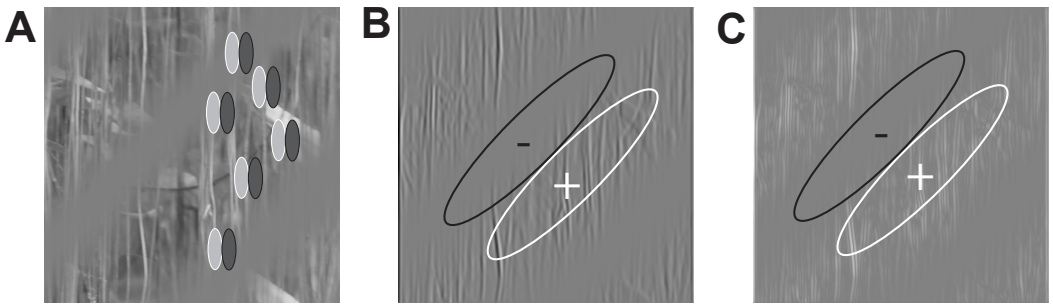

Figure 1–6: Algorithm for second-order segmentation. (A) The luminance structure of a texture carrier with an oblique contrast modulation is detected by high-spatial frequency filters. (B) The responses of these filters cannot be used for contrast boundary segmentation by a linear filter of an appropriate spatial frequency and orientation because the mean response is the same between the high- and low-contrast regions. (C) Following rectification by a pointwise absolute value transformation, all filter responses are positive and so the mean response to the excitatory region is greater than that to the inhibitory region, and the boundary may be segmented.

2 Higher order texture statistics impair contrast boundary segmentation

In the following manuscript I employ photographs of natural textures, and the same photographs phase-randomized as textures over which contrast modulations were applied. I found that human segmentation thresholds were much lower for the phase-randomized textures, which implies that the presence of higher-order structure impairs the performance of segmentation mechanisms. Following rank-ordering the textures based on the magnitude of the threshold reduction, sparseness and local edge structure appear to be candidates for higher-order statistics that impair the segmentation of contrast boundaries.

CHAPTER 2
Higher order texture statistics impair contrast boundary segmentation

This chapter has been published as: Arsenault, E., Yoonessi, A. & Baker Jr., C. (2011). Higher order texture statistics impair contrast boundary segmentation. Journal of Vision, 11(10), 14.

2.1 Abstract

Texture boundary segmentation is conventionally thought to be mediated by global differences in Fourier energy, i.e., low-order texture statistics. Here, we have examined the importance of higher-order statistical structure of textures in a simple second-order segmentation task. We measured modulation depth thresholds for contrast boundaries imposed on texture samples extracted from natural scene photographs, using forced-choice judgments of boundary orientation (left vs. right oblique). We compared segmentation thresholds for contrast boundaries whose constituent textures were either intact or phase scrambled. In the intact condition, all the texture statistics were preserved, while in the phase-scrambled condition the higher-order statistics of the same texture were randomized, but the lower order statistics were unchanged. We found that (1) contrast boundary segmentation is impaired by the presence of higher-order statistics; (2) every texture shows impairment but some substantially more than others; and (3) our findings are not related to scrambling-induced changes in detectability. The magnitude of phase-scrambling effect for individual textures was uncorrelated with variations in their amplitude spectra, but instead we

suggest that it might be related to differences in local edge structure or sparseness.

2.2 Introduction

Our rich perceptual experience of the shapes, objects, and surfaces that make up the visual world relies on successful segmentation of distinct regions in an image to delineate the boundaries between them. The visual system can detect boundaries defined by changes in a number of properties, commonly divided into two categories: those that can be distinguished based on a point-to-point comparison of simple intensive properties such as luminance or colour (first-order); and those that require two-stage processing to distinguish, such as orientation, spatial frequency, or contrast of textures (second-order). Processing of these first- and second-order boundaries is widely thought to be mediated by distinct mechanisms (e.g. Schofield & Georgeson, 1999; Allard & Faubert, 2007). First-order processing is relatively well modeled in terms of linear Gabor-like spatial filters that ostensibly represent V1 receptive fields. Second-order boundaries are inherently more complex, and how they are segmented has been a continuing subject of investigation.

We use the term 'segmentation' not to refer to a specific task, but to refer to the process by which the visual system detects second-order boundaries. In this paper, we examine second-order vision through its simplest manifestation: contrast boundary segmentation. It is well known that contrast boundary segmentation performance depends on some of the properties of the texture over which the contrast gradient is defined, i.e., the carrier. In particular, carrier orientation orthogonal to the contrast boundary facilitates contrast boundary detection at low spatial frequencies (Dakin & Mareschal, 2000), and higher spatial frequency carriers have been found to show an advantage as well (Sutter et al., 1995; Dakin & Mareschal, 2000). However with relatively broadband noise, spatial frequency content

was found to have little impact on detection (Schofield & Georgeson, 2003). These studies were restricted to simple filtered noise carriers, and the full extent to which a texture's appearance is relevant to the operation of second-order mechanisms remains to be seen. In this paper we address this issue by imposing contrast modulations on textures sampled from natural images to begin our examination of the importance of a wide range of statistical structure on second-order vision.

It has long been clear that only a limited subset of a texture's properties are used by the visual system to segment it from another texture. A classic demonstration of this is our inability to segment pairs of textures whose elements are readily discriminated for example, upright and inverted chevrons (Olson & Attneave, 1970) or Ts and Ls (Beck, 1966; Bergen & Julesz, 1983). Such observations led naturally to the idea that textures should be thought of in fundamentally statistical terms, and that their segmentation is based on a representation in which only some image statistics are preserved. In the last two decades, most of the work on mechanisms of texture segmentation has been couched in terms of two-stage filtering models (Bergen & Adelson, 1988; Landy & Graham, 2004) that can be thought of as comparing the global Fourier energy across a boundary. These models have only been evaluated using simple synthetic textures, and it is unclear how adequately such models can account for human segmentation of textures, and boundaries defined over textures, that contain a wider variety of local features. Texture segmentation is an important example of the emerging general idea that many of our perceptual abilities seem to be based not on a perfect translation of the retinal image, but on "summary statistics", a compressed statistical representation in which only some attributes of the retinal image are retained (Chong & Treisman, 2003; Rosenholtz,

2011). There is evidence that such a representation is automatically and pre-attentively computed (Oliva & Torralba, 2007; Oliva & Torralba, 2001), and it appears as though we make some judgments based only on a subset of the statistics available in the image (Ariely, 2001; Chong & Treisman, 2003; Alvarez & Oliva, 2008). In at least some contexts, such as peripheral vision (Balas et al., 2009), a statistical summary of the information in the stimulus may be just as relevant to perception as the stimulus itself. Finding the most appropriate summary statistics for a given task is both informative about the mechanisms involved and important to consider when evaluating the results of past studies, or designing stimuli for future experiments.

In this work we employ a particularly useful and popular way of describing image statistics using a Fourier decomposition of the image. We can distinguish between lower- and higher-order statistics based on the Fourier power spectrum and phase spectrum. The lower-order statistics represented in the power spectrum describe the global energy present in the image: luminance, contrast, spatial frequency, and orientation. The phase spectrum embodies higher-order statistics that describe the spatial distribution of that energy (Oppenheim & Lim, 1981; Thomson & Foster, 1997). For example, step edges in luminance occur when Fourier components of the same orientation over a range of spatial frequencies are phase-aligned in their zero-crossings; such broadband edges are considered to be of particular interest in statistics of natural images (Olshausen & Field, 1996; Thomson, 1999).

While randomizing an image's phase structure will severely handicap identification of image content (Hansen & Hess, 2007), some of the textural aspects of the image's appearance are preserved - particularly for textures with a high degree of structural regularity, or periodicity (Emrith et al.,

2010). While some aspects of overall texture and shading may be captured by the power spectrum (Tadmor & Tolhurst, 1993), several studies have shown that phase spectral information contributes to human perception of isolated textures. Kingdom et al. (2001) manipulated parameters of synthetic micropattern textures to modify their contrast, skew and kurtosis - they found that human observers could most efficiently discriminate textures differing only in their fourth-order statistics (kurtosis). Demonstrations of texture synthesis (Portilla & Simoncelli, 2000) showed that a variety of higher-order statistics are required to capture a texture's appearance when they are attentively examined, though evidently only a subset of these higher-order statistics are necessary for pre-attentive discrimination of textures (Balas, 2006). Motoyoshi and Kingdom (2010) demonstrated that discrimination of random paired-Gabor textures was enhanced by a co-circular relationship between nearby orientational structures.

Even though information in the phase spectrum is critical to higher-level tasks such as texture appearance judgments, and can aid the discrimination of one texture from another, its relevance to a pre-attentive, low-level task such as texture segmentation remains unclear. The popular conception of an energy model of segmentation emphasizes global comparisons of lower-order statistics, but other models have been based upon different sets of statistics, some of which are higher-order. Julesz (1962) conjectured that texture segmentation mechanisms might operate on only a subset of available statistics, i.e. the relationship between the luminance values of any two pixels at a given distance from one another. This theory was later expanded to include relationships between triplets (Julesz et al., 1978) and quadruplets (Julesz, 1981a) of pixels. Graham et al. (1993) demonstrated element arrangement patterns created with oriented gabor patches that

can be readily segmented along boundaries defined only by differences in the relative positions of the texture elements, implying a mechanism that is sensitive to phase information. To achieve human-like segmentation in natural scenes by a computer vision algorithm, Martin et al. (2004) and later Arbelaez et al. (2011) made use of higher-order texture statistics along with other boundary cues. They classified each pixel of an image as belonging to one of a small collection of 'textons' based on the responses of a range of co-localized oriented filters, followed by a second stage operator that compares the texton histograms on opposing sides of a putative boundary. Phase scrambling would remove the spatial co-localization of filter responses that define these textons, and so texton-based segmentation would be impossible. Thus there is evidence suggesting that higher-order statistics influence segmentation, but a systematic study is difficult because what constitutes a 'higher-order statistic' is unbounded, and defined only by exclusion to consist of anything that is not a lower-order statistic. In this work, we use natural image photographs to sample higher-order texture statistics that are likely to be critical to ecological vision, and explore the relationship between these statistical regularities and human performance on a texture segmentation task.

The texture statistics most ecologically relevant to segmentation are those occurring on either side of boundaries that occur in natural images. However photographs of natural texture boundaries would make poor experimental stimuli for a number of reasons: (1) the texture boundaries in natural images most often arise from occlusions of one object by another, which typically are accompanied by coincident luminance changes, and therefore are not purely second-order (Johnson & Baker Jr., 2004); (2) experimentally manipulating the textures on either side of a boundary is

problematic without affecting the boundary itself; and (3) boundaries in images from natural scenes (excluding man-made structures) are rarely straight, further complicating the preceding difficulty. Instead we approach the problem with photographs of natural textures, which we can individually manipulate and use as carrier patterns to construct synthetic envelope boundaries. This semi-natural approach gives us the same access to the higher-order texture statistics that are present in photographs of the real world, while affording the benefits of using synthetic boundaries: experimentally controllable texture statistics and a consistent boundary shape without luminance artifacts.

We explore segmentation of boundaries defined by contrast differences imposed across individual textures, rather than segmentation of a boundary between two distinct textures, for two reasons. Firstly, contrast gradients are the simplest form of texture boundary, and thus more amenable to analysis. Secondly, this approach allows us to deal with individual textures one at a time, affording a better opportunity to investigate the effects of individual differences in texture statistics.

Note that most previous studies of higher-order texture statistics and segmentation have explored whether it was possible to segment boundaries defined by differences in these statistics, such that they were necessary to do the task (e.g., Julesz et al., 1978). On the other hand, in these experiments the higher-order texture statistics are, in principle, irrelevant to the task; instead we ask whether their presence facilitates or impairs segmentation performance.

To evaluate the role of higher-order carrier statistics in contrast boundary segmentation, we look at psychophysical performance under two conditions: natural textures with all the statistics preserved ('intact'

condition), or phase-scrambled versions of the same natural textures in which the higher-order statistics have been randomized but the lower-order statistics remain the same ('scrambled' condition). If the power spectrum provides the basis of segmentation, we would expect to find no differences in psychophysical performance between the intact and scrambled conditions. If any higher-order information is utilized by the visual system in this task, we would expect to see performance impaired in the scrambled condition. On the other hand the boundary might be obscured by higher-order information in the texture, in which case we would expect improved performance in the scrambled condition.

2.3 General Methods

2.3.1 Stimuli

The natural textures used in this experiment were acquired from high-resolution photographs (3888 × 2592 pixels) taken in a variety of locations such as parks, beaches, and botanical gardens. A digital SLR camera (Canon Digital Rebel XTi) was used to take the photos in RAW format with a linear gamma profile, which were then converted to 16-bit TIFF and imported into Matlab. From each of these photographs, we manually extracted candidate texture regions of 480 × 480 pixel squares.

The candidate images were then screened subjectively by the authors to evaluate the extent to which they exemplified key characteristics of 'texture': uniformity of lightness, contrast, and granularity (Wilkinson, 1990; Bergen & Adelson, 1991; Wilkinson & Wilson, 1998; Portilla & Simoncelli, 2000; Kingdom et al., 2001). We used these characteristics to define our acceptance criteria for textures as images that appeared to be relatively uniform and composed predominantly of a single type of material (such as grass, bark, or ripples in sand) or a homogeneous mixture of materials (e.g. branches and leaves). We also required the detail of the texture to be in focus, and free of prominent segmentable objects. Textures of man-made materials such as bricks, concrete, or tiles were excluded. Examples of textures excluded in this stage are shown in Figure 2–1A (top).

The textures that passed the subjective screenings were converted to greyscale (using the Matlab function rgb2gray) and further screened objectively for internal homogeneity by comparing the luminance and RMS contrast (Kingdom et al., 2001; Bex & Makous, 2002), of four quadrants of the texture. If there were any pairwise differences greater than 3dB the texture was excluded (Figure 2–1A (bottom)). Approximately 64% of

the hand-selected textures passed this test, providing a database of 239 natural texture images. Four examples of these textures are displayed in Figure 2–1B.

The stimuli for all of our experiments used the textures from this database as carrier patterns. Texture stimuli for the baseline ('intact') condition were created using the texture as described above, to measure segmentation with all the higher- and lower-order statistics present in greyscale natural photographs. In the second ('scrambled') condition we phase scrambled the intact texture to remove structure of the higher-order statistics. We created scrambled textures by applying a Fourier transform to both the intact texture and a white-noise image of the same size. The phase values in the natural texture were replaced with those of the white noise and inverse-transformed, thus leaving the power spectrum unchanged while completely randomizing the phases (Dakin, Hess, Ledgeway, & Achtman, 2002).

To create the carriers, each texture was scaled to have a mean value of 0, and its extreme values were clipped at ± 3 standard deviations and scaled to fit in the range of intensities between ± 1.0. This texture carrier was then modulated by an envelope pattern to create a synthetic contrast boundary. For our envelope, we used a half-disc pattern with an oblique orientation boundary, graduated over 20% of the image width with a cosine taper (Figure 2–2). The final stimulus, $\mathbf{S}_{x,y}$, is the product of the stimulus window $\mathbf{W}_{x,y}$, the carrier, $\mathbf{C}_{x,y}$, and the envelope, $\mathbf{E}_{x,y}$, scaled by the modulation depth, m:

$$\mathbf{S}_{x,y} = L_o\{1 + c\mathbf{C}_{x,y}\mathbf{W}_{x,y}((1 + m\mathbf{E}_{x,y})/2)\} \tag{2.1}$$

where $|\mathbf{C}_{x,y}| \leq 1.0$, $|\mathbf{E}_{x,y}| \leq 1.0$, and $0 \leq \mathbf{W}_{x,y} \leq 1$. L_o is the mean luminance, m is the modulation depth, and c is a contrast scaling factor which is adjusted to produce the desired RMS contrast.

We used these stimuli to measure threshold values of modulation depth (m) or carrier contrast, for intact (Figure 2–2, left) and phase scrambled (Figure 2–2, right) natural textures using an envelope orientation judgment ($\pm$45 deg) in a two-alternative forced-choice task. We presented the stimuli at a suprathreshold contrast in all experiments unless otherwise specified.

To prevent observers from performing the task by monitoring the contrast of only one quadrant of the texture, the phase of the envelope was randomly shifted 180 degrees from trial to trial. To further diversify the stimulus appearance, and impair observers' ability to learn and use specific texture features, carrier textures were randomly flipped vertically and/or horizontally on each trial, prior to applying the contrast envelope.

The stimuli were presented on a CRT monitor (Sony Trinitron Multiscan G400, 81 cd/m2, 75 Hz, 1024 $\times$ 768 pixels), gamma-linearized with a digital video processor (Bits++, Cambridge Research Systems) that allowed us to present low contrast stimuli without binarizing artifacts by increasing the bit-depth from 8 to 14 bits. Stimulus patterns appeared in a central 480 $\times$ 480 pixel patch on a mean grey background. Observers viewed the stimuli from a distance of 114 cm, resulting in a stimulus visual angle of approximately 6.5 degrees. The experiments were run on a Macintosh (Desktop Pro, MacOSX) using Matlab and PsychToolbox (Brainard, 1997; Pelli, 1997).

2.3.2 Task

At the beginning of each trial observers were presented with a central fixation point, and used a button press to initiate each 100-millisecond

stimulus presentation. The envelope boundary was oriented 45 degrees either left or right oblique, and observers indicated with a button press the perceived orientation of the boundary. Feedback was not provided as a precaution against aiding spurious cue learning. The screen was maintained at the mean grey background between stimulus presentations.

We measured thresholds using a method of constant stimuli with five logarithmically spaced level values, chosen to span an appropriate range as determined from pilot experiments for each observer and texture that spanned 5% to 60% modulation depth. A minimum of three blocks of 100 trials, with 20 trials per level, were run for each condition to yield a total of at least 60 trials per level. Percent-correct data from a total of 600 trials were fit with a logistic function, and a threshold was interpolated for 75% correct. Curve fitting was performed by the statistics package Prism (GraphPad Software, Inc.), and standard error measurements were estimated with its bootstrapping algorithm.

We tested the significance of the difference between thresholds in the intact and scrambled conditions using a two-tailed paired-samples t-test with a criterion $\alpha = 0.05$, and measured the effect size (D. Klein, 2005) using Cohen's d with the standardizer s computed as:

$$s = \sqrt{\sigma_1^2 + \sigma_2^2}/2 \tag{2.2}$$

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

2.4 Experiment 1

This experiment examined whether an observer's ability to segment contrast boundaries is affected by higher-order statistics of carrier textures drawn from a large sample of texture appearances. This was accomplished by comparing modulation depth thresholds for contrast boundaries with natural texture ('intact') carriers, and those with phase-scrambled ('scrambled') carriers.

2.4.1 Methods

To obtain a general picture of the contribution of higher-order statistics to segmentation, we measured modulation depth thresholds for the texture library as a whole. On each trial, a carrier texture was selected from the database randomly without replacement within each block of 100 trials. At a suprathreshold carrier RMS contrast of 14.5%, we measured modulation depth thresholds for each observer in the intact and scrambled conditions. We collected data from four experienced psychophysical observers, three of whom (JB, AM, JH) were naive to the hypotheses of the experiment.

2.4.2 Results

The results for Experiment 1 are shown in Figure 2–3. Modulation depth thresholds for phase-scrambled textures (light bars) are substantially lower than those for intact textures (shaded bars) for each observer. We found a large, statistically significant effect of phase scrambling ($t(3) = 14.71, p < .05, d = 2.86$) with the average observer's intact threshold 2.36 dB above their scrambled threshold. These results not only suggest that the presence of higher-order statistics in natural textures is a relevant factor in performance on this task, but also that segmentation improves when higher-order statistics are removed.

2.5 Experiment 2

In the previous experiment, we observed a difference in thresholds for ensembles of intact and ensembles of phase-scrambled textures, providing evidence for a role for higher-order texture statistics in boundary segmentation. But since our textures vary widely in appearance, it is unclear to what extent our result is uniformly representative across textures, or if some textures demonstrate a greater effect of phase scrambling than others. In this experiment, by comparing modulation depth thresholds for individual intact and phase-scrambled textures, we aimed to determine what effect individual differences in texture appearance have on modulation depth thresholds of contrast boundaries.

2.5.1 Methods

This experiment was conducted in the same manner as Experiment 1 in almost every respect. However, rather than randomly selecting textures on each trial, modulation depth thresholds were measured in separate blocks for each of twenty individual textures chosen to span a wide range of appearances and represent a variety of scales, materials and environments. For each threshold measurement, a single texture was used on every stimulus presentation, so that modulation depth thresholds, and therefore any difference between the intact and phase-scrambled conditions, could be assessed separately for each texture.

A modulation depth threshold was determined for each texture in the intact and scrambled conditions. Data were collected for three observers, two of whom (JH & AM), were naive to the hypotheses of the experiment.

2.5.2 Results

The results from this experiment are shown in Figure 2–4, where each symbol indicates the thresholds for the scrambled versus the intact

conditions for a particular texture. The dashed line indicates the 1:1 ratio between the two thresholds, which is where we would expect the data to fall if there were no effect of phase scrambling. The thresholds for all textures tested fall below the 1:1 line, indicating that the intact thresholds are higher than the scrambled thresholds, in agreement with the results from Experiment 1. On average, intact thresholds are 2.25 dB (SD = 0.84 dB) higher than scrambled for observer LA, 2.48 dB (SD = 1.08 dB) higher for JH, and 2.44 dB (SD = 1.29 dB) higher for AM. Overall, thresholds for all subjects show a substantial, statistically significant reduction after the carrier is phase scrambled : LA ($t(19) = 8.46, p < .05, d = 2.00$), AM ($t(19) = 4.89, p < .05, d = 2.00$) and JH ($t(19) = 7.3, p < .05, d = 1.99$).

From the scatter plots in Figure 2–4 it is apparent that while thresholds for all textures are affected by phase scrambling to some extent, some thresholds are reduced substantially more than others. One contributing factor appears to be the magnitude of the intact threshold; textures with higher intact thresholds seem to show more reduction than those with lower intact thresholds. A one-tailed Spearman correlation shows a significant, positive correlation between the intact threshold and the threshold reduction in decibels: LA ($r(20) = 0.72, p < .05$), AM ($r(20) = 0.60, p < .05$), and JH ($r(20) = 0.84, p < .05$). Thus the textures that are more difficult on the segmentation task are the ones that benefit most from phase scrambling.

To get an idea of what specific texture attributes might contribute to the differing thresholds, we sorted the textures into a histogram (Figure 2–5) based on the threshold change for each texture averaged across the three observers. The textures that have a small effect of phase scrambling tend to be made up of densely packed, smaller features or markings, while the

textures that showed a large effect of phase scrambling tend to be composed of larger elements with longer continuous contours.

2.6 Experiment 3

In the previous experiments, the textures were all equated for RMS contrast, and this metric (like other low-order image statistics) is preserved after phase scrambling. Nevertheless, it is conceivable that our results could be explained by systematic differences in detectability between the intact and phase-scrambled texture conditions. If the scrambled textures were easier to detect than their intact counterparts, they might be at an advantage in the contrast boundary segmentation task. Here, the same task was undertaken as before on a representative subset of the textures from Experiment 2, but using stimuli constructed from textures at fixed contrast increments above their individually measured detection thresholds.

2.6.1 Methods

In this experiment, two thresholds were determined in separate blocks for each condition: first the carrier contrast threshold, and then the modulation depth threshold. As before, 'modulation depth' (m in Equation 2.1) refers to the extent to which the envelope, in this case a contrast change, is applied. 'Carrier contrast' refers to the RMS contrast level of the unmodulated carrier.

We measured carrier contrast thresholds using a method of constant stimuli for each condition, texture and observer. Five logarithmically spaced carrier contrast levels were tested at a modulation depth of 100% (Figure 2–6) using the same left- or right-oblique segmentation task. Then, to compare modulation depth thresholds as directly as possible, we presented the stimuli at 6dB above each observer's carrier contrast threshold for that particular texture. We tested eight natural textures from the previous subset of twenty for this experiment. Carrier contrast, and then modulation depth

thresholds, were measured for two observers, one of whom (JB) was nave to the purposes of the experiment.

2.6.2 Results

The carrier contrast threshold results are shown in Figure 2–7, where each symbol indicates the scrambled and intact thresholds for a particular texture. The points fall very close to the equality line, suggesting that there is no systematic effect of phase scrambling on detectability. We found no statistically significant differences between the carrier contrast thresholds of intact and scrambled textures for either observer LA ($t(7) = 0.412, p > .05$) or JB ($t(7) = 2.038, p > .05$). We also found relatively little variability between textures; the axes illustrated in Figure 2–7 span a range of only one octave, compared with a four-octave range illustrated in Figure 2–4. This finding of very similar detection thresholds for different RMS contrast-equated textures, whether intact or scrambled, is consistent with the report of Bex and Makous (2002) that RMS contrast provides a good contrast metric for natural images.

Modulation depth thresholds for the detectability-equated intact and phase-scrambled textures are shown in Figure 2–8. All points lie below the 1:1 line as in Experiment 2, indicating that thresholds were again lowered following phase scrambling. Comparing the average effect of phase scrambling, we find that intact thresholds are still 2.17dB (SD = 0.83) higher than scrambled for observer LA, and 2.53dB (SD = 1.28) higher for observer JB. The difference between the thresholds in the intact and phase scrambled conditions remains statistically significant for both observers: LA ($t(7) = 5.734, p < .05, d = 2.59$) and JB ($t(7) = 4.326, p < .05, d = 2.11$). Furthermore, the large effect sizes (d) reported here are similar to those found in Experiment 2, as are the average changes in threshold,

indicating that the effect observed in Experiments 1 and 2 is not the result of differences in effective RMS contrast for intact and phase-scrambled textures.

2.7 Discussion

In this study we found that the presence of higher-order statistics impaired performance on a basic texture segmentation task. In Experiment 1 we used an ensemble of more than 200 natural texture photographs to show that it is more difficult to segment contrast boundaries imposed on intact textures than those imposed on phase-scrambled textures. We extended this result in Experiment 2, showing that this effect occurs in varying degrees for different individual textures. Finally, in Experiment 3 we showed that intact and scrambled textures are about equally detectable, and that scaling the carrier contrast to the detection thresholds of individual textures and observers does not eliminate or even reduce the observed effect. Based on these results, we cannot rule out the possibility that some kinds of higher-order statistics could contribute positively to segmentation; we simply conclude that whatever help some statistics might contribute, they do not overcome the impairment imposed by other statistics.

Our finding that higher-order information impairs performance runs contrary to what has been found in many non-segmentation tasks such as texture discrimination (F. Phillips & Todd, 2010), spectral slope discrimination (Thomson & Foster, 1997) and scene recognition (Hansen & Hess, 2007), where higher-order statistics improve performance. However, higher-order statistics have previously been found to impair the detection of distortions in natural scenes (Bex, 2010). As in the work described here, other studies have found that perception depends on more than simply the presence or absence of higher-order statistics; it depends on some statistics more than others, and the degree of their importance varies from image to image for reasons that are not entirely clear (Hansen & Hess, 2007; Bex, Solomon, & Dakin, 2009; F. Phillips & Todd, 2010).

In the past, different investigators have considered various kinds of 'higher-order' statistics - excellent reviews can be found in Kingdom et al. (2001) and Landy and Graham (2004). Julesz et al. (1978) emphasized the importance of considering higher-order statistics in segmentation models, but their use of the term is not congruent with the more conventional Fourier-based statistics that we employed in this study. By controlling n-th order correlations, one can create images with identical autocorrelation functions, and therefore identical Fourier amplitude spectra in an ensemble average (Julesz, 1962; Julesz et al., 1978; Victor, Chubb, & Conte, 2005). The Julesz constraint that second-order correlations be identical does not preclude individual samples of these populations from differing in their second-order statistics (Yellott Jr, 1993; Chubb & Yellott, 2000). Though Victor (1994) argued that texture statistics, by nature, characterize a population rather than individual samples, segmentation mechanisms have access to only a pair of samples at any given moment and so sample statistics cannot be ignored. Furthermore, these statistics are difficult to examine in the context of the linear filtering models that are prevalent in modern vision theory, because the Julesz statistics are not maintained following linear filtering (S. Klein & Tyler, 1986).

Which higher-order statistics impair segmentation?

What image statistics might be at the root of our results remains unclear. In the histogram of our stimuli (Figure 2–5), there is a strong visual impression of a difference in structural appearance between the textures whose thresholds are least affected by phase scrambling (left side) and those most affected by phase scrambling (right side). Some specific apparent differences are relative amounts of high and low spatial frequency

information, structural sparseness, and local edge structure, which we will now consider.

Spectral slope. While we find an effect of higher-order statistics on segmentation, the magnitude of the threshold change could be associated with individual differences in the amplitude spectra of the textures. It appears as though the textures whose thresholds are most affected by phase scrambling might have relatively less energy in the high spatial frequencies and proportionately more energy at lower frequencies. This difference in the proportions of low and high spatial frequency information could have an impact on segmentation mechanisms. To assess the relative amounts of high and low spatial frequencies in our stimuli, we measured their spectral slopes by fitting a linear regression to the log-log plot of Fourier amplitude vs. spatial frequency (Thomson & Foster, 1997; Bex & Makous, 2002) for each texture - steeper negative slopes would indicate relatively more energy in the high spatial frequencies. The results are plotted in Figure 2–9A as a function of the change in threshold between intact and scrambled conditions - note that most of the spectral slopes were close to -1, as expected for natural images (Field, 1987; Ruderman, 1997). There does not appear to be any relationship between threshold change and spectral slope, and a Pearson correlation on these variables failed to find any significant correlation ($r(20) = -0.39, p > .05$). This lack of relationship suggests that an explanation of our results based on relative differences in high vs. low spatial frequencies is unlikely.

Sparseness. Colloquially, structural sparseness can be defined in terms of the amount of 'stuff' that appears to be in a texture (Adelson, 2001). A collection of 50 leaves seen close up can be considered a 'sparse' texture, while a field of millions of blades of grass seen from a distance appears less

sparse. Textures that have been phase scrambled do not appear sparse because there are no local concentrations of energy (e.g., small edges or other texture markings) and complementary regions of blank space forming discrete objects. Upon phase scrambling, the pixel and wavelet distributions become normal (Bex & Makous, 2002), rather than the kurtotic distribution that is a signature of sparseness (Kingdom et al., 2001). Sparseness is well known as a key attribute of natural scenes (Ruderman, 1997; Field, 1987), but it is also a primary property of textures. Victor and Conte (1996) proposed "granularity" as an important higher-order distinction between textures, which they investigated using textures formed with a range of element sizes. Computer science and image statistical methods have described 'coarseness' as a major dimension along which textures vary (Rubner & Tomasi, 1998). Durgin (1995, 2008) showed that density is a primitive texture feature for which adaptation effects can be measured, and Kingdom et al. (2001) demonstrated that textures can be discriminated based on sparseness. However, none of these previous efforts examined the impact of sparseness on boundary segmentation.

To measure sparseness, Kingdom et al. (2001) suggested intensity histogram or wavelet-based kurtosis, and Hansen and Hess (2007) developed a wavelet-based measure of kurtosis, the LSSM, to assess the sparseness of natural images. We computed these metrics for each texture, and plotted them against the textures' change in threshold (Figure 2–9B, C) - in both cases, there is no systematic relationship, as confirmed by the lack of significant correlation between threshold change and either pixel kurtosis ($r(20) = -0.15, p > .05$) or LSSM ($r(20) = -0.07, p > .05$). These results suggest either that sparseness is not a relevant higher-order statistic, or that the sparseness metrics we employed are not sufficiently sensitive to

sparseness. Note that textures appear relatively dense compared to images of scenes, and it may be that these sparseness metrics perform less well in this context.

Edge structure. It might not be the global arrangement of the energy (sparseness) that determines the difference in the thresholds, but the varying density of broadband features within the image (local edge structures). The textures that were more affected by phase scrambling (Figure 2–5, right side) appear to have more prominent local edges. We assessed the amount of local edge structure using a modified version of the method for computing edge density outlined in Bex (2010). We integrated a Canny edge map (constructed using Matlab's canny edge detector), and normalized by the number of pixels in the texture to obtain an index of edge density. In a plot of this edge density index against the textures' change in threshold (Figure 2–9D), we can see a systematic relationship: the textures with greatest effect of phase scrambling had higher edge density indices, while those with least effect had lower edge densities. This relationship was confirmed by a significant correlation between this rough measure of edge density and threshold change ($r(20) = -0.80, p < .05$), suggesting that some aspect of both broadband edges and the density in which they occur may impact the effect of higher-order image statistics on segmentation performance. However this result is also consistent with a role for sparseness, since sparser textures would produce smaller indices of edge density. Untangling these factors may be problematic with natural texture photographs, but using synthetic textures, where both sparseness and local edge structure can be controlled, is a clear way forward.

Higher-order statistics impair segmentation performance

Why the presence of higher-order statistics might impair the segmentation of contrast boundaries is an open question.

Camouflage or masking. It could be the case that luminance-defined contours in intact textures camouflage the target boundary - however luminance noise has little impact on contrast boundary segmentation (Allard & Faubert, 2007), so we do not think that this causes our results. Aside from luminance variation, sparse images contain low- and high-contrast regions, and this spatial modulation of contrast could mask the modulation that observers are tasked with identifying (Allard & Faubert, 2007). We consider this unlikely for three reasons: (1) the observer knows that the edge will be in one of two positions, so there is very little positional uncertainty; (2) the textures are randomly flipped from trial-to-trial, so any texture features that happen to appear along the envelope boundary will only affect some (25%) of the trials; and (3) second-order masking is spatial-frequency dependent (Hutchinson & Ledgeway, 2004), which suggests that such high spatial frequency contrast noise should not affect performance on our low spatial-frequency boundary. Low spatial-frequency contrast noise should not present a problem because we specifically excluded textures that were too coarse or had large-scale contrast gradients. However the precise bandwidths of the noise, boundary, and second-order mechanisms are ill-defined, so while it seems unlikely we cannot rule out the possibility that second-order masking plays a role in determining our results.

Subjective texture selection. One might argue that including those textures we excluded based on their subjective characteristics could somehow reduce or nullify the observed effect of phase-scrambling. Because we find a strong correlation between threshold and Canny edge density, we performed the analysis of Figure 2–9D on the 49 textures excluded in the subjective

stage of texture screening. We found that they have systematically lower edge density (M=0.060, SD= 0.02) than the 20 we tested in Experiment 2 (M=0.095, SD=0.03), and thus their inclusion would in fact have been more likely to increase the size of the effect we observe.

Edge vs. region processing. Finally, it may be that phase-scrambled and dense images provide more support along the contrast boundary itself, and so are easiest to segment. This explanation supposes that the mechanism responsible for segmentation preferentially uses information near the texture boundary (an edge-based process) rather than integrating information throughout the entire stimulus (a region-based process). This is possible, but is a departure from the common conception of texture segmentation mechanisms as two-stage filter models with large second-stage filters that can operate across the entire region. In the context of this model, a strong reliance on edge support would be surprising, particularly for contrast modulations.

The regional summation in the second stage of a standard energy model is not explicitly selective for the local features encoded in higher-order statistics, but there are various ways to adapt this model to enable such selectivity. The simplest modification of a standard FRF model would entail changing the nonlinear function separating the first- and second-stage filters. Graham and Sutter (1998) found that an expansive power-law nonlinearity with an exponent between 3 and 4 better accounted for their findings with element arrangement patterns. By using a nonlinearity that is more expansive than a square-law, textures with localized areas of high energy (such as sparse natural images) would give a greater response. A more serious modification to the FRF scheme would be to use first-stage filters that act like nonlinear 'feature-detectors' - for example, Martin et

al. (2004) used histograms of different types of local features, defined by the co-localizations of wavelet responses, to segment texture boundaries. Finally, one might add an additional nonlinear process beyond the second stage. Graham et al. (1993) argued that an additional stage of a filter-rectify cascade (i.e., Filter-Rectify-Filter-Rectify-Filter rather than FRF) was necessary to segment some element arrangement patterns that differed only in their higher-order statistics.

2.8 Conclusion

We have shown that texture segmentation mechanisms are sensitive to the information in the phase spectrum of an image. From these findings we cannot be certain which specific statistics contribute, but it appears as though sparseness and local edge structure in particular might be relevant statistics in this task. To address these issues, we intend to use synthetic stimuli to isolate these specific higher-order statistics, and to gauge their impact on segmentation independently. It is not yet clear whether current models of texture segmentation account for how humans process higher-order statistics, but testing and modifying those models may prove to be informative. A detailed examination of the effects of higher-order statistics as an ensemble and as individual exemplars (e.g., sparseness) on segmentation will be useful for refining models to the point where we can begin to apply them to biologically relevant stimuli. We can conclude that while textures are segmented using a limited subset of the information they contain, this subset must be expanded to include higher-order statistics in some capacity.

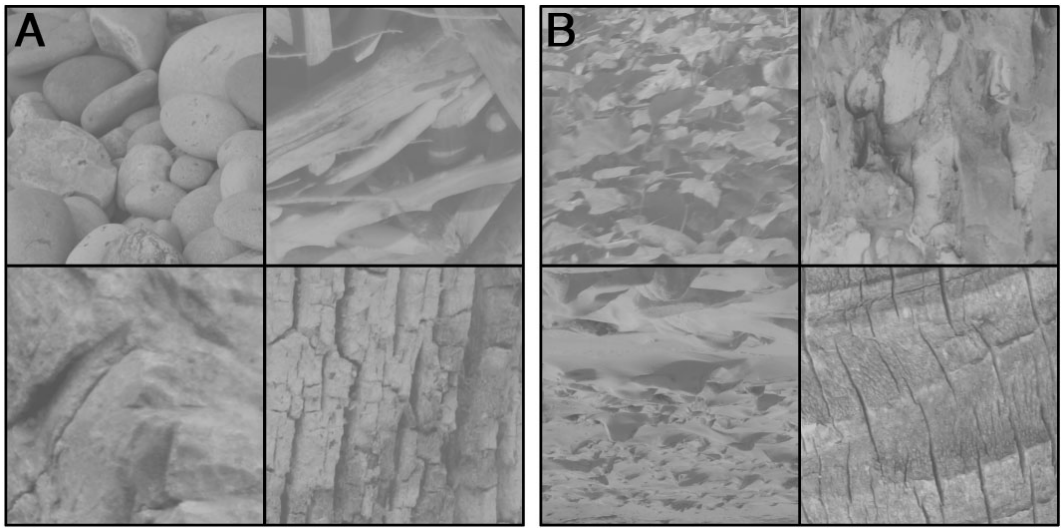

Figure 2–1: Examples of excluded (A) and included (B) natural textures. (A-top): Images that were excluded due to a subjective judgment that they were not sufficiently uniform, homogeneous, in focus, or contained prominent segmentable objects. (A-bottom): images that were excluded during computer screening due to inhomogeneity of luminance or contrast between two or more quadrants. (B): Images that were included in the texture corpus.

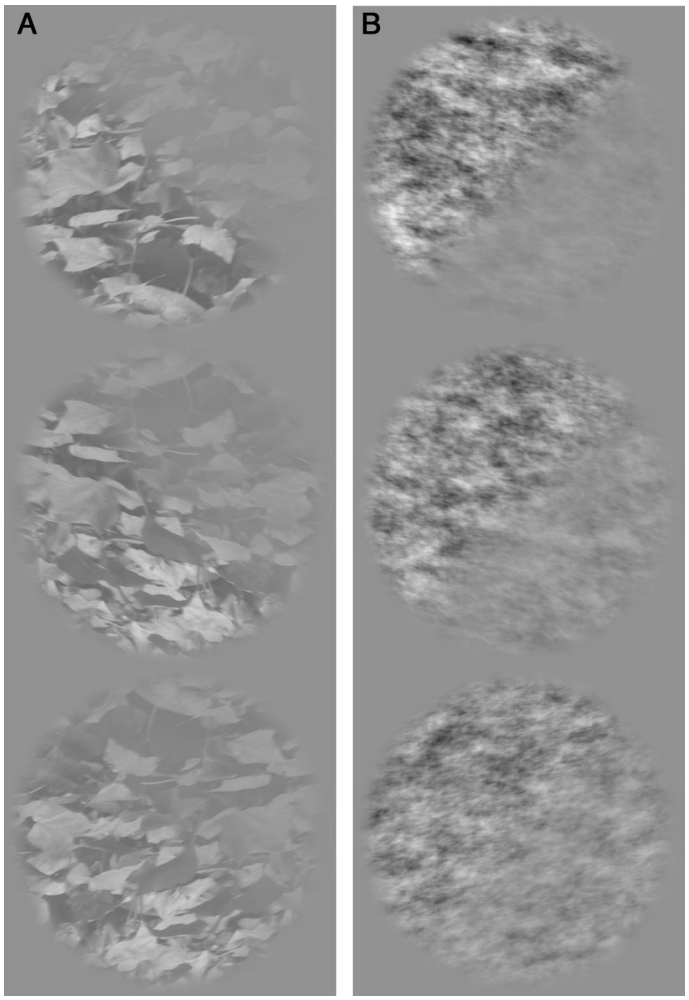

Figure 2–2: Examples of the stimuli used to determine modulation depth thresholds in experiments 1, 2, and 3 shown at three modulation depths (top-bottom: 75, 50 & 32). The envelope is a left- or right-oblique half-disc contrast modulation applied to an intact (left) or phase-scrambled (right) natural texture.

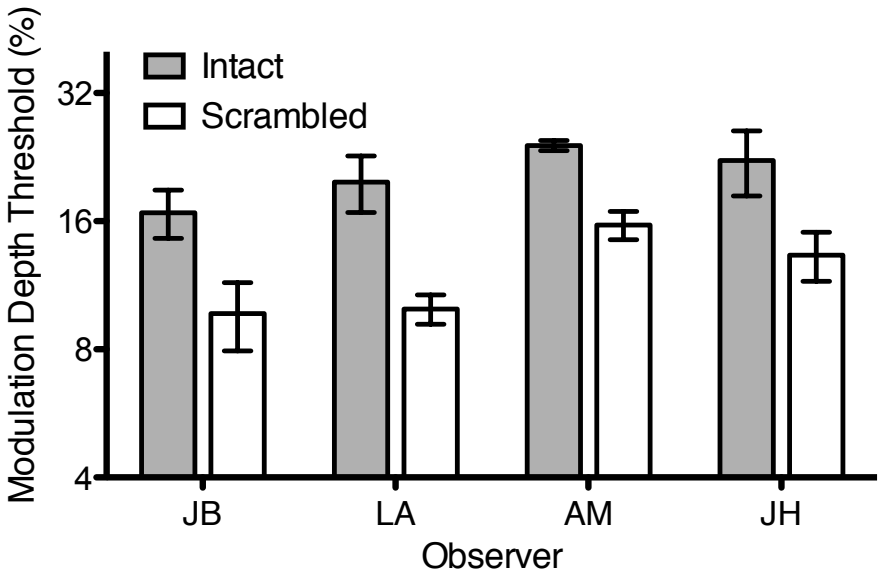

Figure 2–3: Modulation depth threshold results from Experiment 1 for four observers for intact and scrambled texture conditions. Thresholds were lower for the phase-scrambled textures (light bars) than for the intact textures (shaded bars). Error bars represent ± 1 standard error.

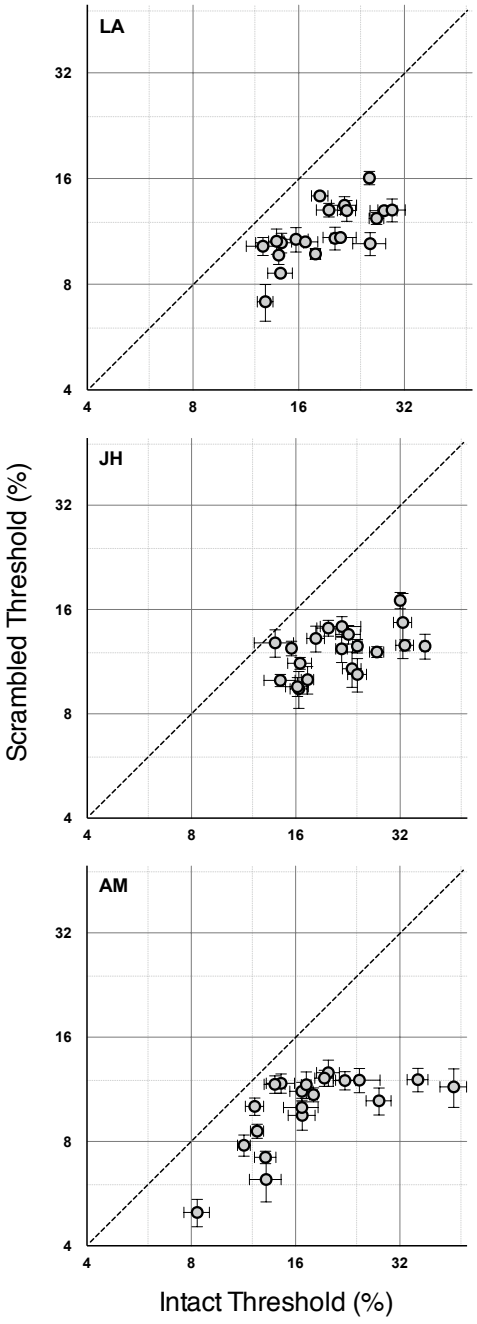

Figure 2–4: Modulation depth threshold results from Experiment 2 for three observers. Each symbol plots the phase-scrambled versus the intact threshold for a particular texture. In almost all 20 textures tested, for all three observers, the symbols lie below the 1:1 line (dashed), indicating that the intact threshold is higher than the phase-scrambled threshold. The amount of reduction, or the distance from the 1:1 line, is texture-dependent. Error bars show the standard error on each measurement.

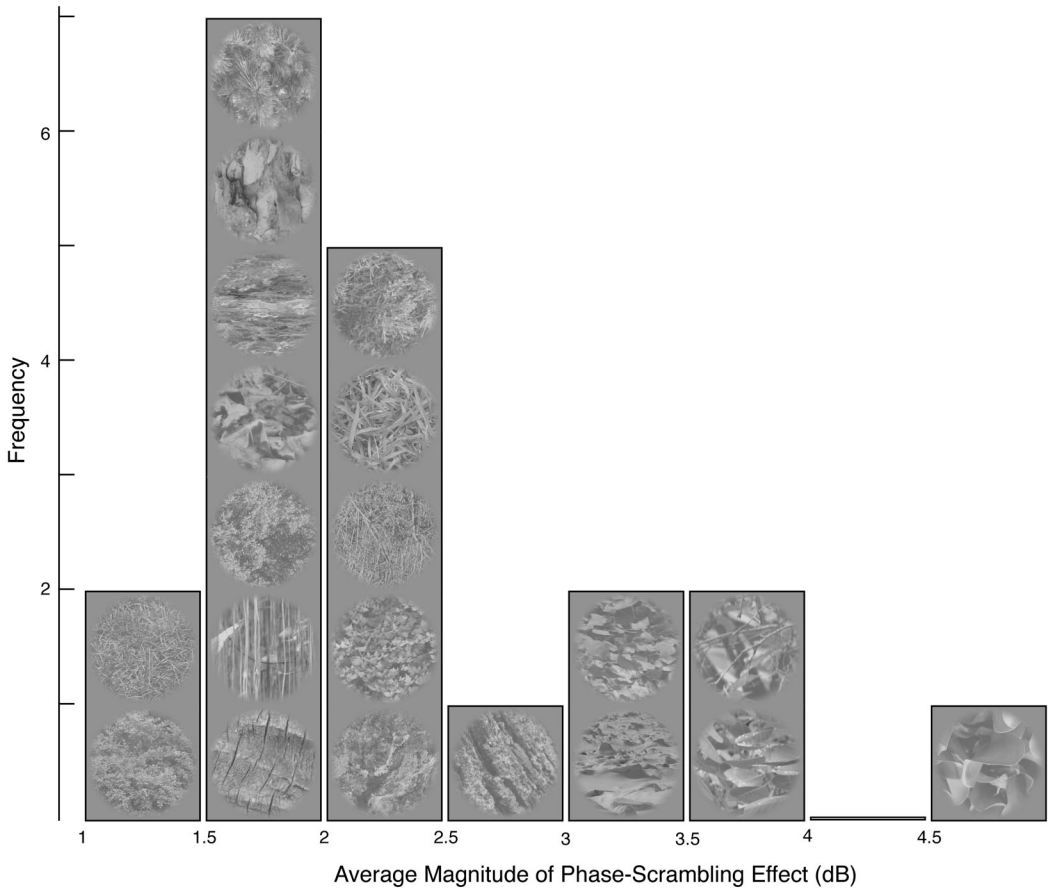

Figure 2–5: Histogram of texture carriers used in Experiment 2, based on average magnitude of the effect of phase scrambling. The textures that show a larger change (> 2 dB) tend to have more prominent edges, and appear to be more sparse.

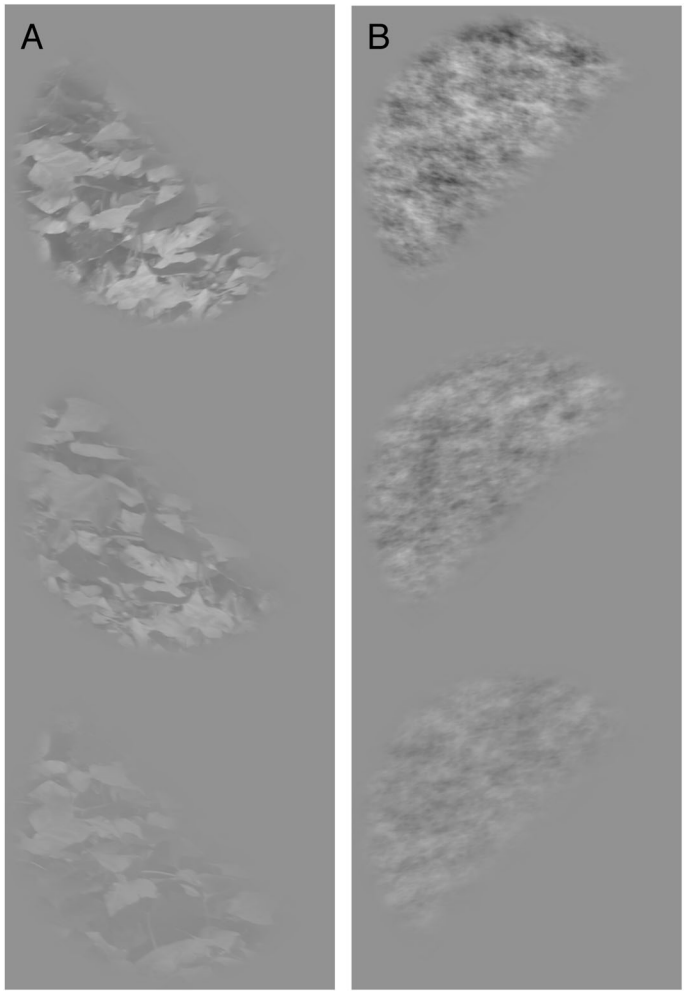

Figure 2–6: Stimuli used to determine carrier contrast thresholds in Experiment 3 shown at a range of carrier contrasts (top-bottom: 8,5 & 3% RMS contrast), all with 100% modulation depth. Thresholds were determined for intact (A) and scrambled (B) textures.

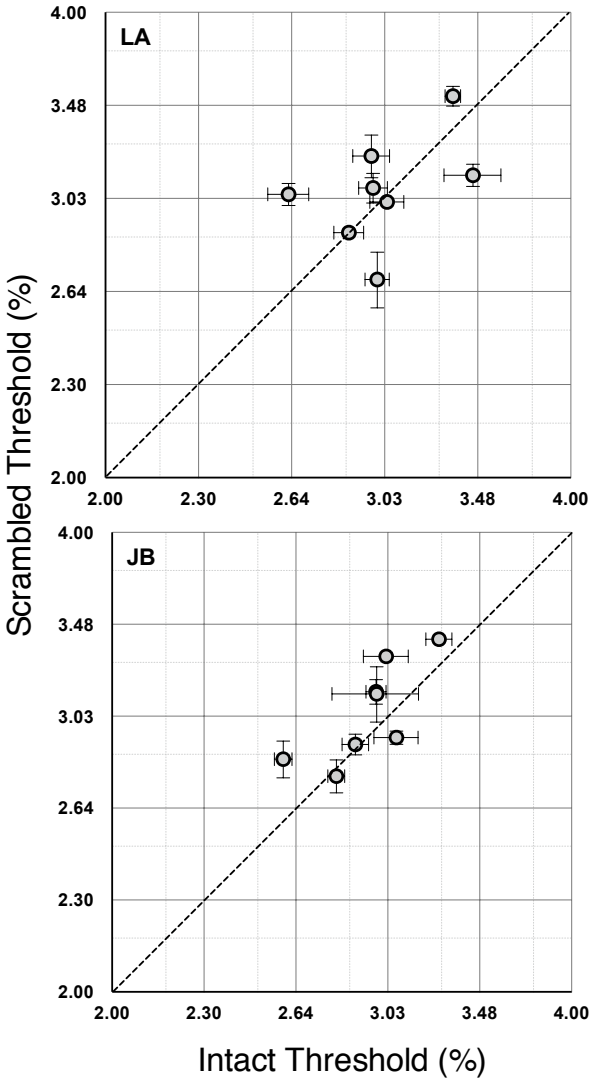

Figure 2–7: Carrier contrast threshold results from Experiment 3 for two observers. Each symbol plots the phase-scrambled vs the intact threshold for a particular texture. The dashed line indicates where a texture's intact and phase-scrambled thresholds correspond exactly. Carrier contrast thresholds are centered on the 1:1 line, suggesting no systematic change in detectability when a texture is phase scrambled. Variation in thresholds between observers, textures, and conditions is minimal.

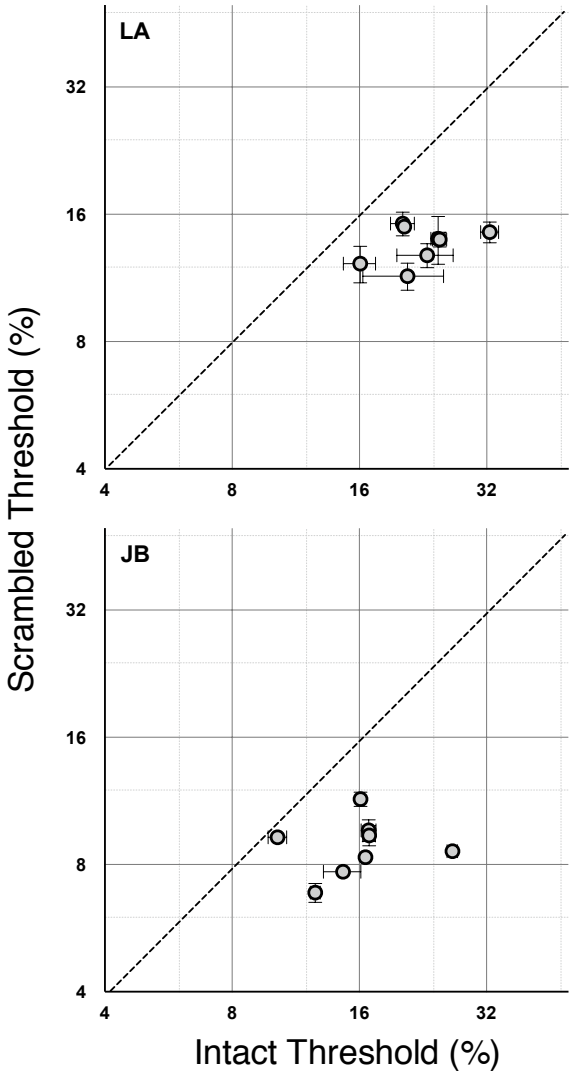

Figure 2–8: Modulation depth threshold results from Experiment 3 for two observers. Each symbol plots the phase-scrambled versus the intact threshold for a particular texture, with each texture a fixed increment above its detection threshold. The dashed line indicates where a texture's intact and phase-scrambled thresholds correspond exactly. Modulation depth thresholds measured with these detectability-equated contrasts are still systematically lower for scrambled than for intact textures.

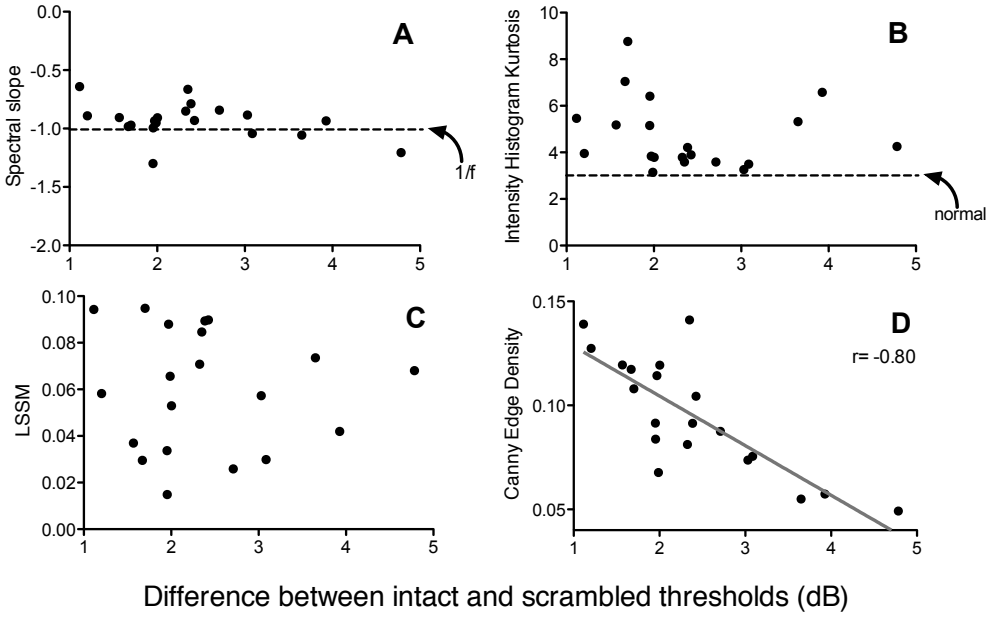

Figure 2–9: Relationship between image statistic indices and the change in segmentation threshold between intact and scrambled conditions in decibels. (A) Slope of fall-off of Fourier spectrum. (B) sparseness, as measured using intensity histogram kurtosis. (C) sparseness, as measured with Hansen and Hess (2007)'s LSSM a wavelet-based metric developed for natural scenes. (D) Edge density, modified from Bex (2010). Note the lack of relationship between threshold change and kurtosis, LSSM, or slope, but a clear correlation between edge density and threshold change.

3 Texture sparseness, but not local phase structure, impairs second-order segmentation

In the previous study I showed that higher-order statistics impaired texture segmentation, and used a rank-ordering method to infer that sparseness and local phase alignment were important statistics. In this study, I create naturalistic synthetic textures to manipulate these individual image statistics experimentally. I compare contrast and orientation boundary segmentation thresholds between textures varying in their structure and sparseness. These experiments demonstrate that texture sparseness and global phase structure impair orientation and contrast boundary segmentation, but that local phase structure has little impact on segmentation. I simulate the experiment using a standard filter-rectify-filter model, which I optimize to accommodate the psychophysical data by employing a compressive nonlinearity.

CHAPTER 3
Texture sparseness, but not local phase structure, impairs second-order segmentation

3.1 Abstract

Texture boundary segmentation is typically modeled by comparing differences in Fourier energy (i.e. low-order texture statistics) on either side of a boundary, but in a previous study (Arsenault, Yoonessi, and Baker, 2011) we showed that the distribution of that energy within the texture (higher-order statistics) also influences contrast boundary segmentation. Here we examine which higher-order texture statistics influence the segmentation of contrast- and orientation-defined boundaries. We use naturalistic synthetic textures to manipulate the sparseness, global phase structure, and local phase alignments of carrier textures, and measure segmentation thresholds based on forced-choice judgments of boundary orientation. We find a similar pattern of results for both contrast and orientation boundaries: (1) removing all structure by globally phase scrambling the texture reduces thresholds substantially, (2) decreasing sparseness also reduces thresholds, and (3) removing local phase alignments has no effect on segmentation thresholds. A two-stage filter model with an interposed compressive nonlinearity can account for the pattern of contrast and orientation boundary segmentation data. From these results, it is apparent that sparseness, created by global phase structure, impairs second-order segmentation.

3.2 Introduction

Boundaries between objects result in discontinuities in a variety of image properties, among which changes in texture are a particularly interesting example because the means by which they are segmented is not yet well understood. Texture can be represented in terms of spatial statistics, but it is unclear what subset of these statistics is actually employed by segmentation mechanisms. Research on texture segmentation usually focuses on those statistics that are different on either side of a boundary and that are sufficient to *enable* segmentation, such as orientation. However textures contain, and their neuronal representation may include, many other statistics that are constant on either side of the boundary. For example, in Figure 3–1 the ivy forms a different percept than the bark, even though both contain contrast-defined boundaries. Such texture statistics may not vary across the boundary so they cannot *enable* segmentation but, if the form they take affects the efficiency of segmentation mechanisms, their presence could *influence* segmentation.

Much previous research has aimed to determine the precise statistical differences that enable segmentation, but few have considered which texture statistics influence performance when they are common to the textures on either side of the boundary. Caelli (1980) examined the influence of a box-shaped feature common throughout the stimulus on segmentation of a boundary defined by a difference in the orientation of the line segments within the boxes. He found that segmentation was more difficult when the boxes were present than when the line segments were presented alone. E. Arsenault, Yoonessi, and Baker Jr (2011) used contrast modulations applied to natural textures to show that higher-order statistics, though they are not relevant to the segmentation task, impair contrast boundary

segmentation. Following a rank-ordering of textures based on the difference in threshold between the intact and phase-scrambled conditions, we noticed that textures with a greater difference appeared to be more sparse as well. We applied a number of image statistical measures that have been used in the literature to measure density of textures or natural scenes, and found that a measure of edge density correlated strongly with the difference between thresholds. From this, we suggested that local edge structure and sparseness were two candidates for texture properties that might cause such a performance difference. It could be the case that the overall contrast boundary is masked by local contrast modulations caused by the regions of high-contrast features that form the structure of sparse textures (Allard & Faubert, 2007), if they are of a low enough spatial frequency to interfere with the task boundary (Hutchinson & Ledgeway, 2004). Additionally, changes in higher-order texture properties could produce different responses following a non-square-law rectification (i.e. a sparse texture will have higher peaks that would be accentuated by an expansive nonlinearity). Given these possibilities, sparseness and local broadband edges are particularly logical statistics of interest, because both result in localized concentrations of image energy.

Previous efforts to understand segmentation of naturalistic textures have faced some serious limitations. It is difficult to target specific statistics using natural textures because individual properties cannot be varied independently or manipulated parametrically. While we have reason to suspect that sparseness or local edge structure might be important influencing statistics, our results are only correlational. In the following experiments, we addressed these challenges by creating synthetic textures consistent with observations of the statistical properties of natural textures

using broadband edge-segment micropatterns. These textures allowed us to not only manipulate global structure through phase scrambling as before, but also the presence of local structure (by phase-scrambling individual micropattens) and the sparseness (by changing the number of micropatterns).

Here, we first aim to verify that these synthetic textures contain the relevant properties of natural textures by demonstrating again the effect of phase-scrambling on contrast boundary segmentation thresholds as in E. Arsenault et al. (2011). By varying texture density and phase structure, we are also able to differentiate the influence of local phase alignments, global phase alignments, and sparseness in segmentation of both contrast and orientation boundaries. We chose to study contrast boundaries because they are the simplest kind of texture boundary, and orientation boundaries because we have observed that natural textures are frequently narrowband for orientation and this type of boundary has been widely studied (Landy & Oruç, 2002; Meso & Hess, 2011). Finally, we present a filter-rectify-filter model with a compressive intermediate nonlinearity that accounts for the pattern of both our contrast and orientation boundary segmentation results.

3.3 General Methods

3.3.1 Stimuli

Each stimulus consisted of a single texture pattern that was contrast-modulated with a half-disc envelope, or two texture patterns 'quilted' together to form a disc with distinct halves (Figure 3–2C). The textures we used were designed to mimic the image statistics of natural textures, while allowing for control of specific texture properties. These synthetic textures were constructed by summing together a large number of randomly scattered *edgelet* micropatterns.

Micropatterns

To emulate natural textures, we used edgelet micropatterns that contained a spatially localized edge composed of phase-aligned Fourier components. The edge was created by adding together the Fourier components of a half-cycle of a square wave ($f, 3f, 5f, , nf$ where n = size/4), with decreasing amplitudes (scaled by $1/f$) and aligned in sine-phase ($\phi = 0$). One cycle of the lowest spatial frequency pattern was combined with like-oriented in-phase harmonics of gratings ($\mathbf{G}$) to form a square wave 'edge' ($\mathbf{D}$) (Equation 3.1):

$$\mathbf{D}_{x,y}(\theta, s) = \sum_{j=0}^{s/4} \frac{1}{f} \mathbf{G}_{x,y}(\theta, \phi, f, s), f = 2j + 1 \tag{3.1}$$

Once the edge had been created, we applied a Gaussian window, whose sigma was 1/8 of the size of the micropattern, for the final edgelet ($\mathbf{D}'$) (Figure 3–2A - top) (Equation 3.2):

$$\mathbf{D}'_{x,y} = \mathbf{D}_{x,y} + e^{-\left(\frac{(x-s/2)^2}{2\sigma_x^2} + \frac{(y-s/2)^2}{2\sigma_y^2}\right)} \tag{3.2}$$

To generate novel textures rapidly, we created a library of 48 such micropatterns at four widths (16, 32, 64, 128 pixels, or 0.22, 0.44, 0.87, and 1.74 degrees of visual angle), each at twelve orientations evenly spaced in 30 degree increments.

In addition to square-wave edgelets, we created phase-scrambled edgelets in a similar manner but with the components' phases (ϕ) randomized rather than aligned (Figure 3–2A - bottom). A total of 50 versions of the phase-scrambling (e.g., Figure 3–2B) were added to the library at each micropattern size and orientation.

Textures

Edgelets drawn from the library were randomly positioned on a 544×544 pixel canvas and summed where they overlapped. While the square-wave edgelets are luminance-balanced with equally sized light and dark regions, the random phases of the scrambled edgelets may result in net mean luminance differences from the gray background. To luminance-balance the texture stimuli, when a phase-scrambled edgelet was randomly selected for inclusion in a texture, it had a 50% chance of having its polarity reversed before being drawn (Figure 3–2B). To obtain an approximately $1/f$ amplitude spectrum, four sizes of micropatterns (16, 32, 64, and 128 pixels) were added in proportions necessary to achieve equal coverage for each spatial frequency (Ruderman, 1997; Kingdom et al., 2001). Thus, for each 128-pixel micropattern, 4, 16, and 64 of the progressively smaller micropatterns were be added - so each texture contained an integer multiple of 85 micropatterns. The possible positions of the micropatterns were constrained to lie entirely within a 544×544 canvas, which was subsequently

cropped to the central 480×480 region after all the micropatterns had been drawn.

Three density conditions were created by varying the number of micropatterns within each texture stimulus. The low density condition had 595 micropatterns, medium density 1530, and high density 2975. These conditions were chosen to result in textures that were qualitatively different in appearance, while satisfying the above constraints to produce an approximately $1/f$ amplitude spectrum.

For each density, three *structure* conditions, where higher-order structure was randomized, were created. *Intact* (INT) textures were composed from square-wave edgelets as described above - these textures were rich in both global (arrangement of micropatterns) and local (broadband edges) structure. By one-dimensionally phase-scrambling the individual edgelets we produced a *locally scrambled* (LS) texture that had an equivalent amount of global structure but lacked local phase alignments (broadband edges). We created *globally scrambled* (GS) textures by applying a Fourier transform to both the intact texture and a white-noise image of the same size. The phase values in the original texture were replaced with those of the white noise and inverse-transformed, thus leaving the power spectrum unchanged while completely randomizing the phases (Dakin et al., 2002).

Due to the random arrangement of micropatterns, some of the INT and LS textures exhibited substantial inhomogeneity, and thus were unsuitable to use as carrier patterns. To circumvent this problem we rejected textures having differences in luminance or RMS contrast greater than 3dB between quadrants of the texture (E. Arsenault et al., 2011). In the low density condition, only about 12% of the generated textures passed this test; in 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 0, and its extreme luminance values were clipped at ± 3 standard deviations and scaled to fit in the range of intensities between ± 1.0.

3.3.2 Boundary Creation

Contrast boundaries: To create contrast-defined boundaries, textures were contrast-modulated by an envelope pattern, consisting of an obliquely oriented half-disc, graduated with a cosine taper. The final stimulus, $\mathbf{S}_{x,y}$, is the product of the texture carrier, $\mathbf{C}_{x,y}$, and the envelope, $\mathbf{E}_{x,y}$, scaled by the modulation depth, m:

$$\mathbf{S}_{x,y} = L_o\{1 + c\mathbf{C}_{x,y}\mathbf{W}_{x,y}((1 + m\mathbf{E}_{x,y})/2)\} \tag{3.3}$$

where $|\mathbf{C}_{x,y}| \leq 1.0$, $|\mathbf{E}_{x,y}| \leq 1.0$, and $0 \leq \mathbf{W}_{x,y} \leq 1$, L_o is the mean luminance, m is the modulation depth, and c is a contrast scaling factor which is adjusted to produce the desired RMS contrast. The top row of Figure 3–3 shows three examples of stimuli with contrast boundaries at different modulation depths.

Orientation boundaries: Orientation-defined boundaries were created between different textures using a method of 'quilting' described by Watson and Eckert (1994) and Landy and Oruç (2002), and illustrated in Figure 3–2C. To modulate two texture carriers ($\mathbf{C}_A$ and $\mathbf{C}_B$) with respect to one another we used a half-disc envelope function ($\mathbf{E}_{x,y}$), scaled to create the carrier A's modulator ($\mathbf{E}_A$), and scaled and inverted to create carrier B's modulator ($\mathbf{E}_B$):

$$\mathbf{E}_A = \sqrt{(1 + m\mathbf{E}_{x,y})/2} \tag{3.4}$$

$$\mathbf{E}_B = \sqrt{(1 - m\mathbf{E}_{x,y})/2} \tag{3.5}$$

The modulation depth parameter (m) scales the difference between the envelope halves.

The luminance-balanced carrier textures ($\mathbf{C}_A$ and $\mathbf{C}_B$) are scaled to yield the desired contrast with scaling factor c, and their means are adjusted so that the final stimulus will be luminance balanced after the envelope has been applied.

$$\mathbf{C}'_A = c\,\mathbf{C}_A - \frac{\iint c\,\mathbf{C}_A\mathbf{E}_A - 0.5}{\iint \mathbf{E}_A} \tag{3.6}$$

$$\mathbf{C}'_B = c\,\mathbf{C}_B - \frac{\iint c\,\mathbf{C}_B\mathbf{E}_B - 0.5}{\iint \mathbf{E}_B} \tag{3.7}$$

The final stimulus ($\mathbf{S}_{x,y}$) is the sum of the two carriers, each spatially weighted by their respective envelopes:

$$\mathbf{S}_{x,y} = L_0\{1 + \mathbf{C}'_A\mathbf{E}_A + \mathbf{C}'_B\mathbf{E}_B\} \tag{3.8}$$

The weight is specified by the modulation depth, m in Equations 3.4 and 3.5. At a modulation depth of zero, the resulting stimulus is a homogeneous blend of the two textures. At a modulation depth of 100%, one half of the disc is entirely $\mathbf{C}_A$, and the other half entirely $\mathbf{C}_B$, with a smooth taper between them at the boundary. The bottom row of Figure 3–3 shows three examples of stimuli with orientation boundaries at different modulation depths.

3.3.3 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 grey background. Observers viewed the stimuli from a distance of 114 cm, resulting in a stimulus visual angle of approximately 6.5 degrees. The experiment was run on a Macintosh (Desktop Pro, MacOSX) using Matlab and PsychToolbox (Brainard, 1997; Pelli, 1997).

3.3.4 Task

Observers were presented with a central fixation point and initiated each 100-millisecond stimulus presentation with a button press. The stimulus contained a boundary that was oriented 45 degrees, either left or right oblique, and observers indicated the perceived orientation with a button press. Feedback was not provided. The screen was maintained at the mean grey background between stimulus presentations.

We determined an appropriate range of testing values from pilot experiments for each observer, and used a method of constant stimuli over five logarithmically spaced level values of modulation depth to measure each threshold. All stimuli were presented at a suprathreshold carrier RMS contrast of 14.5%. A minimum of three blocks of 100 trials, with 20 trials per level, were run for each condition to yield a total of at least 60 trials per level.

3.3.5 Data Analysis

Percent-correct data from a total of 600 trials were fit with a logistic function, and a threshold was interpolated at the 75% correct point. Curve-fitting was performed using the statistics package Prism (GraphPad

Software, Inc.), and standard errors were estimated with its bootstrapping algorithm.

We used two-way ANOVAs to test for significance with a criterion $\alpha = 0.05$. We measured the effect size (D. Klein, 2005) using Cohen's d with the standardizer s computed as:

$$s = \sqrt{\sigma_1^2 + \sigma_2^2}/2. \tag{3.9}$$

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

3.4 Experiment 1: Contrast boundary segmentation

In this experiment we set out to (1) test whether our finding that higher-order statistics impair contrast boundary segmentation in natural textures (E. Arsenault et al., 2011) could be replicated using synthetic edgelet textures and, if so, (2) investigate the influence of sparseness and local edge structure on segmentation thresholds. Given our previous findings, we expect better performance in the globally scrambled (GS) condition than in the intact (INT) condition, at least for some values of density. If local edge structure influences segmentation, we expect a difference between the LS and INT conditions; if sparseness influences segmentation, we expect a decrease in threshold as density is increased in the INT and LS conditions.

3.4.1 Methods

This experiment measured the modulation depth threshold for segmentation of contrast-defined boundaries over a number of synthetic textures, created as described in the general methods. We tested intact, locally, and globally scrambled textures at each of three density levels. These are depicted in Figure 3–4 with the structure changes (INT, LS, GS) varying across columns and the density increasing down each column. Thresholds were measured for four experienced psychophysical observers with normal or corrected-to-normal vision, three of whom (JH, AR, PCH) were naive to our hypotheses.

3.4.2 Results

Contrast boundary segmentation results are shown in Figure 3–5, for individual observers in the upper four graphs, and as a group average in the lower graph. For all observers the global scramble (GS) condition (open triangles) yielded the lowest thresholds, particularly at low and

moderate densities. This result indicates that higher-order statistics impair segmentation, in agreement with our earlier results for natural textures (E. Arsenault et al., 2011). Because phase scrambling eliminates any effect of density, it was not surprising that we found no difference in thresholds across different density levels in the GS condition (Figure 3–5, open triangles). On average we found no difference between segmentation thresholds for the intact (INT, filled circles) and locally scrambled (LS, open circles) texture carriers. The impact of density on performance for these textures was modest, but there was a trend towards higher thresholds at lower densities.

A two-way ANOVA confirmed a main effect of structure (whether the texture is in the GS, LS, or INT condition) $F(2, 18) = 28.41, p < .05$, but not density $F(2, 9) = 1.510, p > .05$, and an interaction between structure and density $F(4, 18) = 3.236, p < .05$. The interaction between structure and density suggests that there is an effect of density in the INT and LS conditions, but not in the GS condition (as expected). Post-hoc Bonferroni tests (Table 3–1) indicated no difference between the INT and LS conditions, but both are significantly different from the GS condition at the low density of 595, and the moderate density of 1530. The trend for decreasing effect sizes (d) with increasing density supports our observation that thresholds in the INT and LS conditions are lower at higher densities.

From these results we conclude that our synthetic textures capture at least some of the image statistics that are related to the phase scrambling effect we observed earlier in natural textures (E. Arsenault et al., 2011). In both experiments, we observed substantially lower segmentation thresholds in textures that were completely phase-scrambled compared to textures that were completely intact or only locally scrambled. Furthermore, because

Density	t	p	d
	INT vs. GS		
595*	4.546	$< .05$	1.58
1530*	2.783	$< .05$	1.34
2975	2.217	$> .05$	2.02
	LS vs. GS		
595*	6.999	$< .05$	4.41
1530*	3.402	$< .05$	2.22
2975	2.087	$> .05$	2.38
	INT vs. LS		
595	2.453	$> .05$	0.76
1530	0.618	$> .05$	0.32
2975	0.130	$> .05$	0.10

Table 3–1: Results of Bonferroni post-hoc tests for contrast boundary segmentation. t is the Bonferroni-corrected t statistic, p the significance level, d Cohen's d, and * indicates a statistically significant comparison.

segmentation thresholds decrease as density is increased it appears that sparseness and global structure impair contrast boundary segmentation, but because the intact and locally scrambled conditions are the same, it appears that local phase structure does not.

3.5 Experiment 2: Orientation boundary segmentation

In the previous experiment we found that global phase structure, specifically sparseness, influenced contrast boundary segmentation. This experiment aimed to extend those findings to texture boundaries defined by orientation. The structure of the models used to segment orientation-defined boundaries are very similar or the same as those used to segment contrast boundaries - two stages of linear filtering separated by a pointwise nonlinearity - so we expect the two conditions will produce a similar pattern of results.

3.5.1 Methods

In this experiment, we measured modulation depth thresholds for observers segmenting boundaries defined by orientation. To maximize orientation contrast we used pairs of textures, each texture narrowband for orientation with micropatterns oriented at 0 and 90 degrees respectively to form a 'herringbone' along the $\pm$ 45 degree boundary in the quilted stimulus. As in the contrast boundary segmentation task, thresholds were measured for intact and locally scrambled textures at each of three density levels (Figure 3–6). The globally scrambled stimulus is only depicted once because density information is destroyed following phase scrambling so it was not tested at various density levels. Thresholds were measured for four experienced psychophysical observers with normal or corrected-to-normal vision, three of whom (JH, AR, JB) were naive to our hypotheses.

3.5.2 Results

The segmentation thresholds from this experiment are plotted in Figure 3–7, with individual results in the four top graphs, and group-average results at bottom. The global scramble (GS) condition was tested once and cannot vary with density. These results are qualitatively consistent with

those from contrast boundaries (Experiment 1). Removing global phase (GS) decreased thresholds substantially relative to the intact texture thresholds, but removing only local phase alignments (LS) had no systematic effect on segmentation. In both the intact and LS conditions, increasing micropattern density reduced modulation depth thresholds. Quantitatively, thresholds were much higher in every condition than those for contrast modulations, suggesting that orientation boundary segmentation is more difficult than contrast boundary segmentation. We found that contrast modulation depth thresholds (Figure 3–5) ranged from 7-40% across all conditions for all observers tested while orientation modulation depth thresholds ranged from 15-80% across all conditions and observers. In any given condition, for any given observer who participated in both experiments, the orientation modulation depth threshold was higher. This is consistent with Motoyoshi and Nishida (2004), who also found this quantitative difference when comparing orientation- and contrast-defined boundary segmentation directly.

A two-way ANOVA confirmed main effects of structure (GS, LS, vs. INT), $F(2, 18) = 151.53, p < .05$, and density, $F(2, 9) = 21.32, p > .05$, and an interaction between structure and density, $F(4, 18) = 23.09, p < .05$. The interaction between structure and density results from the decline in the INT and LS conditions with increasing density. Post-hoc Bonferroni tests (Table 3–2) confirmed the lack of difference between the INT and LS conditions, but showed that both were significantly different from the GS condition at every density. The trend for decreasing effect sizes (d) with increasing density again suggests that thresholds in the INT and LS conditions decline at higher densities, as we found in the Experiment 1.

Taken together with the results from Experiment 1, we conclude that performance measurements for segmenting orientation and contrast

Density	t	p	d
	INT vs. GS		
595*	15.12	$< .05$	8.65
1530*	7.088	$< .05$	2.99
2975*	3.701	$< .05$	1.55
	LS vs. GS		
595*	15.26	$< .05$	7.88
1530*	7.328	$< .05$	4.08
2975*	3.722	$< .05$	2.01
	INT vs. LS		
595	0.133	$> .05$	0.08
1530	0.239	$> .05$	0.11
2975	0.021	$> .05$	0.01

Table 3–2: Results of Bonferroni post-hoc tests for orientation boundary segmentation. t is the Bonferroni-corrected t statistic, p the significance level, d Cohen's d, and * indicates a statistically significant comparison.

boundaries are influenced by the presence of higher-order statistics in the same way: global phase structure, particularly the structure that results in sparseness, impairs segmentation, but local edge structure appears to have no effect.

3.6 Model

Second-order boundary segmentation has often been understood in terms of a 'filter-rectify-filter' model using early high spatial-frequency filters to capture the texture, followed by a nonlinearity and a late, low spatial-frequency filter that recovers the boundary (e.g. Chubb & Sperling, 1988; Malik & Perona, 1990; Landy & Graham, 2004). Here we implement a model with a filter-rectify-filter architecture in order to see whether the observed effects of structure and density can be accounted for using such a model, and if so, for what configuration and parameter values.

3.6.1 Filter-Rectify-Filter Model

We implemented a basic filter-rectify-filter model (Figure 3–8), as described in the previous study. First the stimulus ($\mathbf{S}_{x,y}$) was convolved ($*$) with a bank of linear filters ($\mathbf{G}_1$) that varied in orientation (θ), spatial frequency (ω), and phase (ϕ) (Eq. 3.10):

$$\mathbf{F}_1(\theta, \omega, \phi, x, y) = \mathbf{G}_1(\theta, \omega, \phi) * \mathbf{S}(x, y). \tag{3.10}$$

The filters were log-gabors, generated using code provided by Kovesi (2000), at two phases (even and odd), six orientations (evenly spaced with their bandwidths chosen for approximately uniform coverage), and four spatial frequencies (160, 80, 40, and 20 cpi, each with a bandwidth of approximately 1.5 octaves). Log gabors are spatial frequency and orientation bandpass functions that were selected instead of typical gabor functions because they are zero-balanced in both even and odd phase, and because they are symmetric in the frequency domain (Field, 1987). The output of each of these filters ($\mathbf{F}_1$) was weighted (w_f), full-wave rectified and raised

(pointwise) according to a power-law of order k, then pooled over phase (Eq. 3.11):

$$\mathbf{R}(\theta, \omega, x, y) = \sum_{\phi} |\mathbf{F}_1(\theta, \omega, \phi, x, y) \cdot w(f)|^k. \tag{3.11}$$

The values of w_f were chosen to equalize responses across spatial scales for a stimulus with a $1/f$ spectral falloff, as in the synthetic textures used here, and is typical of natural images on average (Field, 1987). To this end, the responses to the higher spatial frequency channels were magnified relative to the responses to lower spatial frequencies, using a weighting function $w(f) = 2f$, where f is an index of spatial frequency with $f = 1$ designating the lowest spatial frequency. Dot products were computed between these responses ($\mathbf{R}_{x,y}$) and two second-stage filters ($\mathbf{G}_2$) in the form of low spatial frequency sine-phase gabor functions that match the two possible orientations of the boundary in the stimulus (45 and -45 degrees), as well as its central position:

$$\mathbf{F}_2(45, \theta, \omega) = \mathbf{G}_2(45) \bullet \mathbf{R}(\theta, \omega, x, y) \tag{3.12}$$

$$\mathbf{F}_2(-45, \theta, \omega) = \mathbf{G}_2(-45) \bullet \mathbf{R}(\theta, \omega, x, y). \tag{3.13}$$

The outputs (o) were computed by pooling the magnitudes of the late-stage filter responses across the orientations and spatial frequencies of the early-stage filters. The pooled response magnitudes were raised to a power (the reciprocal of k), and then combined with additive decision noise for the final output value. (Equations 3.14 & 3.15). The noise values n_1

and n_2 were drawn from a normal distribution with a mean of 0 and whose standard deviation, or amplitude, a is a free parameter of the model.

$$o_{45} = \left(\sum_{\theta,\omega} |\mathbf{F}_{2_{45}}(\theta,\omega)| \right)^{1/k} + n_1 \tag{3.14}$$

$$o_{-45} = \left(\sum_{\theta,\omega} |\mathbf{F}_{2_{-45}}(\theta,\omega)| \right)^{1/k} + n_2 \tag{3.15}$$

These outputs were compared, and the late-stage filter (left- or right-oblique) with the strongest response determined the decision (d) of the model (Equation 3.16):

$$d(o) = \begin{cases} 45 & \text{if } o_{45} \geq o_{-45} \\ -45 & \text{if } o_{-45} > o_{45} \end{cases} \tag{3.16}$$

3.6.2 Simulation

We tested the model to determine its segmentation thresholds in much the same manner as we tested our human participants. The model made left- or right-oblique decisions in 60 trials for each of the stimulus conditions illustrated in Figure 3–4 and Figure 3–6 on 12 logarimithically-spaced modulation depth levels that spanned chance to perfect performance. We measured the percent-correct for each level and stimulus condition, and then fit a cumulative Gaussian function using Matlab to determine the model's threshold. Because the stimuli are randomly generated on each trial, model results varied from one simulation to another. For this reason, we simulated the experiment four times and averaged the thresholds. Standard errors were determined based on variability between the four runs.

3.6.3 Optimization

The model, as specified, has two free parameters: k, the order of the power-law nonlinearity, and a the amplitude of the decision noise distribution from which n_1 and n_2 are selected. We chose to optimize the model based on the human orientation boundary segmentation data, because of its smaller variability between subjects.

We simulated the experiment as described above for each of five power-law exponents and six noise levels. We evaluated the model's responses based on four metrics, chosen to capture key aspects of the pattern of our data for human performance on this experiment (Figure 3–9A): (1) The slope of the intact performance with respect to density (INT slope) was computed between the thresholds at the lowest and highest density values. This metric captures the rate at which performance improves with density. (2) The difficulty of the globally scrambled condition (GS difficulty) compares the GS threshold between the model and humans. (3) The difference between the intact and locally scrambled conditions (INT-LS difference) was consistent over density, and so was measured here at the lowest density as the difference in decibels between these conditions. This should reflect the contribution (or lack thereof) of local structure to the threshold. (4) The difference between the intact and globally scrambled conditions (INT-GS difference) was measured in the same manner as INT-LS difference, at the lowest density. This reflects the magnitude of the reduction in threshold that results from phase scrambling. These metrics were computed for both human and model performance, and error was measured as the decibel difference between the two. Positive error values reflect the model over-estimating these metrics, and negative values reflect under-estimates.

We created two-dimensional error maps (Figure 3–9A) to help visualize how each of the metrics are affected by the two parameters. Each vertical axis refers to the values of k, while the horizontal axis refers to the range of six noise values used for each value of k. The value of k has a substantial effect on the magnitude of the second-stage filter responses and therefore the magnitude of the noise values that are required to affect performance. A different set of six noise levels was chosen for each value of k. The colour of each cell shows the decibel difference between model and human for a given value of k, noise level, and metric. Green cells are closest to zero, blue signal under-estimates and red over-estimates, so green cells signal lowest error and the optimal model parameters.

The slope of the intact thresholds (INT slope) with respect to density, and the difference between the intact and locally scrambled conditions (INT-LS difference) are primarily affected by the nonlinear power-law exponent and hardly at all by the decision noise, while the difficulty of the globally scrambled condition is strongly affected by the decision noise. The difference between the intact and globally scrambled conditions is determined by an interaction between these parameters, in which increasing values of the nonlinear power-law exponent yield minimal error for decreasing levels of decision noise.

We chose to optimize the difficulty of the globally scrambled condition first, because that graph shows a clear minimum for performance at each value of nonlinear power-law exponent. A noise level within the range illustrated in Figure 3–9A, but chosen with more precision so that the difficulty of the globally scrambled condition was correct within ± 0.5 dB, was determined for every value of k. When the model was evaluated at each nonlinear power law exponent with the level of noise fixed, a simple

relationship between error the other three error metrics and the value of k emerged (Figure 3–9B): all the metrics improve as k decreases (i.e.: the nonlinearity becomes more compressive) and error appears to reach asymptote at a k below 0.5.

3.6.4 Results

We evaluated the model at k=0.25 with the noise level optimized as described above to match the difficulty of the globally scrambled condition on the orientation boundary segmentation task, and used these parameters to predict performance on the contrast boundary segmentation task. The average thresholds from four model runs are show in Figure 3–10.

The orientation boundary segmentation results for the model (black) and humans (grey) are shown in Figure 3–10A. The model matches the quantitative human data in the GS condition very well, as it should because the noise parameter was optimized for these data. As in the human data, the model shows little difference between the INT and LS conditions at any density, and the rate at which thresholds in these conditions decrease with density is very similar to the human results. The model appears to slightly, but consistently, underestimate thresholds for the INT and LS conditions at all densities.

The contrast boundary segmentation results are plotted in Figure 3–10B, determined using the model parameters that were optimized for the orientation boundary segmentation data. The model, like human observers, shows a decrease in segmentation thresholds for globally scrambled textures relative to the similar intact and locally scrambled texture conditions, and matches the human thresholds reasonably well. Thresholds for intact and locally scrambled textures are again very similar and the model consistently predicts declining threshold with density, but the model considerably

overestimates the rate at which thresholds decrease with density in the INT and LS conditions. The threshold predictions at low densities are too high, and those at high densities are too low.

Both the model and human observers had higher thresholds for orientation than for contrast boundary segmentation, which is readily explainable in terms of the model. The information for orientation segmentation is only in two orientation channels of the early filters (vertical and horizontal), whereas contrast information is available in all of the first-stage spatial frequency and orientation channels. Overall, the model performs very well given that only two free parameters were optimized. With a compressive nonlinearity and an appropriate amount of decision noise, the model performance depends upon global higher order statistics in the same way that human performance does, yet also (like humans) is insensitive to local phase structure.

3.7 Discussion

Our previous work (E. Arsenault et al., 2011) indicated that higher-order statistics impair segmentation but only suggested that sparseness and local edge structure might be important statistics. Here we have identified sparseness as a critical texture property and ruled out a role for local phase alignments, and extended our findings on contrast boundary segmentation to orientation boundary segmentation. The results of this work demonstrate that higher-order statistics influence the segmentation of second-order boundaries defined by contrast or by orientation. In both cases the presence of higher-order statistics was found to impair segmentation; in particular sparser textures impaired segmentation more. For both orientation and contrast-defined boundaries, the presence of local phase alignments did not affect segmentation. We found that these results can be accounted for using an FRF-style model with a compressive pointwise nonlinearity and decision noise.

Any number of high-level explanations for our data can be ruled out using the model. Because whatever causes density and global phase scrambling to influence performance is present in both human observers and the model, we are able to quickly narrow down the possible causes of the density and phase scrambling effects to one common factor: second-order noise. When sparseness is increased, while overall RMS contrast is held constant, energy is clumped into higher local contrast regions with lower local contrast regions between them. These changes in local contrast are themselves contrast modulations that could act as noise for the mechanism segmenting the main boundary (Allard & Faubert, 2007). We were able to test this possibility with the model by measuring the average second-stage filter response to unmodulated textures for ten randomly generated textures

in each condition (Figure 3–11A). We found that, though the second-stage filter was narrow-band, its response to the texture was affected by the texture's second-order variations. Second-stage filter responses were lowest (little second-order interference) in the GS condition, and highest (greatest second-order interference) in the INT & LS-595 condition. The average response decreased parametrically as density was increased to 1530 and then to 2975. When a modulation 0.5 dB above threshold was introduced, the pattern persisted (Figure 3–11B); responses were higher for intact and locally scrambled textures than for globally scrambled textures. The right oblique filter, responding to the modulation, has a greater output than the left in every condition, though the difference between the left and right filters (before the addition of noise) is more for the GS textures than INT or LS.

That the contrast noise affects contrast boundary segmentation does not obviously imply that it would have the same effect on orientation boundary segmentation. In this case, we can see that because the model processes orientation and contrast boundaries within the same pathway, the sparseness (which could result from local contrast, orientation, or spatial frequency changes) provides image features that interfere with the second-stage filter in the same way.

We determined that a compressive intermediate nonlinearity provided the best fit for our results, which has not been previously suggested in segmentation literature, but is in agreement with diverse models of various visual processes. In texture segmentation tasks, a square law has been the traditional nonlinearity shape, because it conceptually corresponds to the Fourier energy present in the stimulus (Malik & Perona, 1990). Graham and Sutter (1998) used a local contrast summation paradigm to

estimate that this nonlinearity was even more expansive than k=2, but probably less expansive than k=4. Otherwise, models of the visual system in the literature tend to favour compressive nonlinearities. Mineault et al. (2012) estimated that the nonlinearity which transforms MT outputs before they are combined in MST is usually compressive, with values of k falling between 0.2 and 0.4. Nishimoto and Gallant (2011) found that the model that best-predicted responses of MT neurons utilized a compressive nonlinearity (k=0.5) between V1 outputs and MT. Thus our model is consistent with what may be an emerging principle of compressive nonlinear summation of sub-units. This compressive summation might be the result of gain control mechanisms implemented by divisive normalization, or by localized surround suppression.

The current model fits the data reasonably well with two free parameters, but there are a number of biologically relevant changes that could improve the model, e.g., interactions across space in the form of surround suppression (H. Tanaka & Ohzawa, 2009), or between channels in the form of cross-orientation inhibition (Motoyoshi & Nishida, 2004). Any spatial interactions are critical to account for because spatial arrangement of information is inherently higher-order. Likewise it is important to capture cross-channel interactions when examining broadband stimuli (David, Vinje, & Gallant, 2004; Bex, Mareschal, & Dakin, 2007).

In the contrast boundary segmentation task, unlike the orientation boundary segmentation task, there was a substantial amount of variability between observers. This was unexpected because pilot data was far less variable. We believe that this variability is due mainly to the range of values tested, because we had difficulty selecting a range of modulation depths that spanned floor and ceiling performance. Had we increased the spacing

between the level values, or tested more levels, we expect that these data would have been much cleaner.

Based on our results using contrast modulations of natural textures (E. Arsenault et al., 2011), we expected to find an effect of local edge structure– a difference between the intact and locally scrambled textures– because our results in that study correlated well with the Canny edge density metric (Bex, 2010). We applied the same metric to the synthetic textures used in these experiments, and found that it estimates a higher edge density for LS textures than for INT because it picks up on the narrowband striations in each scrambled edgelet. For this reason, the Canny edge density metric would predict a slight reduction in segmentation thresholds for locally scrambled textures.

There are texture properties that we know to be relevant to segmentation that edgelet-based textures did not allow us to manipulate. For instance, we did not consider local contrast polarity, though it can enable segmentation (Malik & Perona, 1990; Motoyoshi & Kingdom, 2007), or the higher-order spatial properties that are known to be relevant to texture appearance (Portilla & Simoncelli, 2000). However, these textures provide a useful tool for studying the impact of specific texture properties on segmentation in a parametric way. They allow control over the shape of the amplitude spectrum, which we fixed at $1/f$, and allowed us to vary sparseness while separating global from local phase alignments. This enabled us to characterize the distinct effects (or lack thereof) of local phase alignments, global phase alignments, and density on second-order segmentation mechanisms.

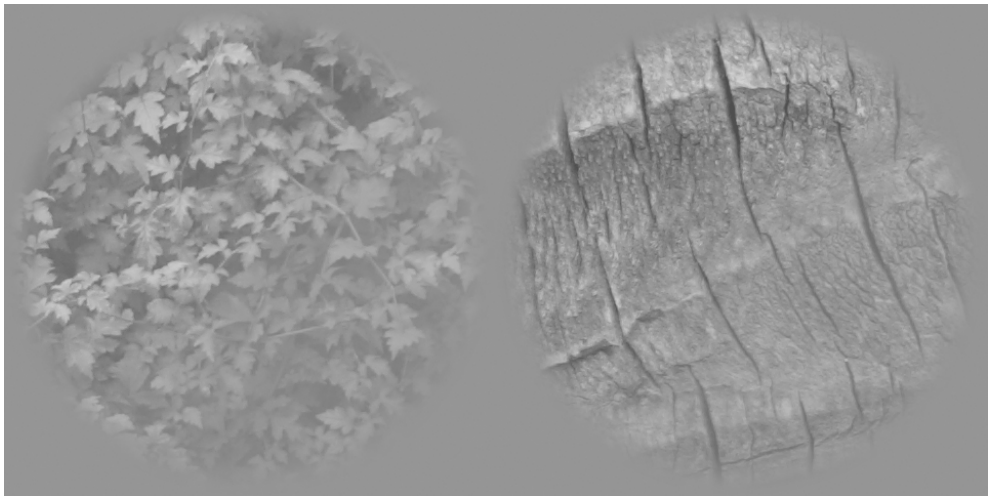

Figure 3–1: Examples of properties that enable versus those that influence segmentation. In both of these textures, a contrast difference *enables* the percept of a right oblique boundary. The properties of the materials (leaves and bark) forming the carrier textures are different in structure which results in a difference in the strength of the boundary percept, even though the modulation depth of the contrast boundary is identical in both examples. In this example, the characteristics of the textures can be said to *influence* segmentation.

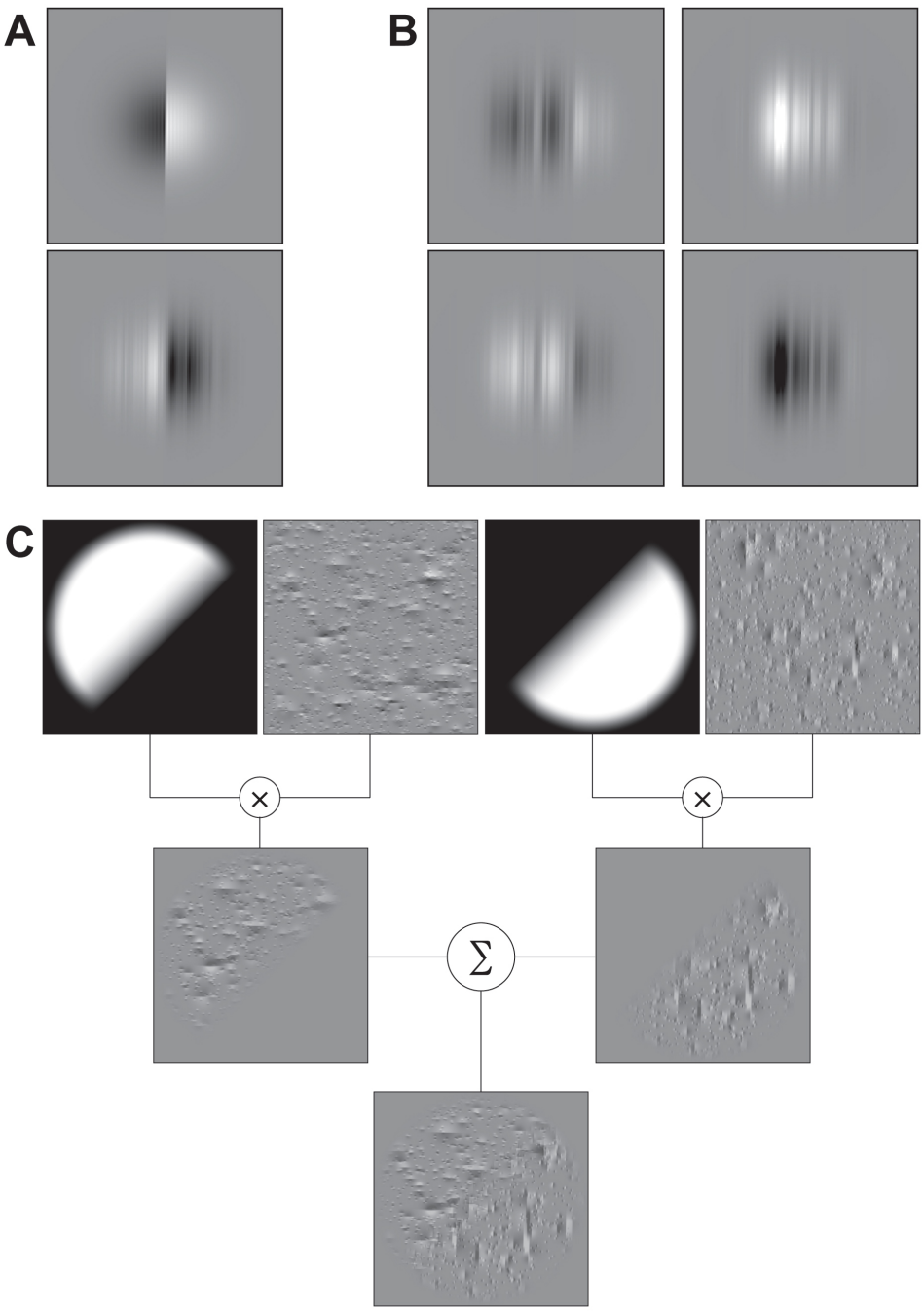

Figure 3–2: Procedure for constructing naturalistic textures and orientation-modulated boundaries. (A) Micropattern types used to create synthetic textures. Top: Intact Gaussian-enveloped square wave, Bottom: Phase-scrambled square wave within the same Gaussian envelope. (B) Top: Variations of phase scrambled micropatterns, Bottom: Same instances, polarity reversed. (C) Procedure for quilting stimuli. Half-disc envelopes are multiplied with their corresponding carrier textures. These modulated halves are then combined. The modulation depth of the stimulus shown is 100%.

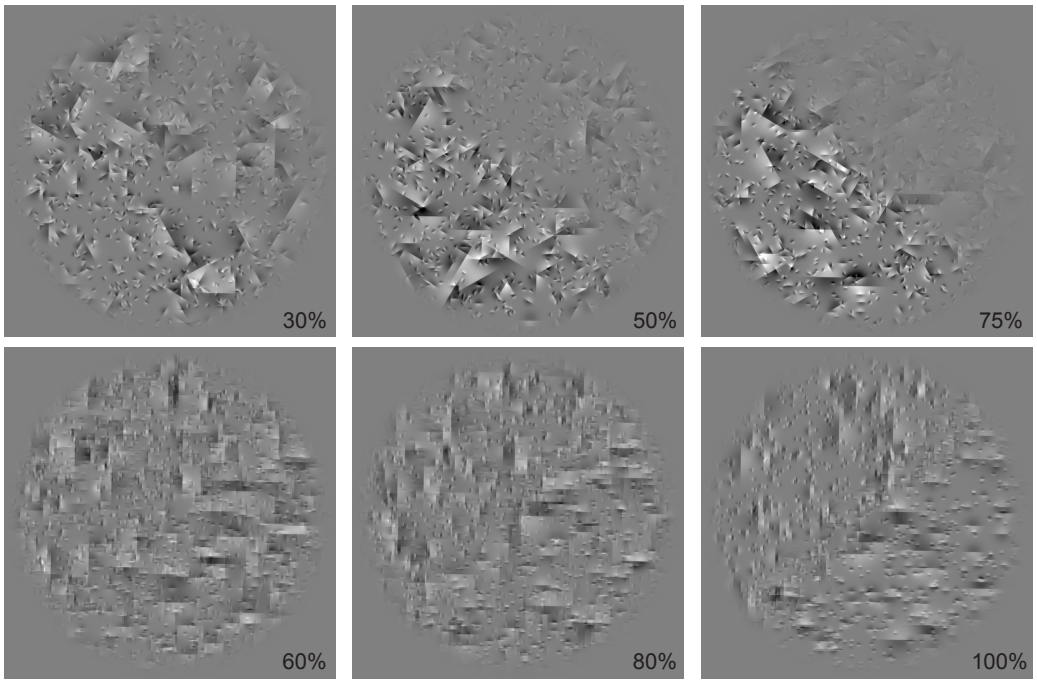

Figure 3–3: Examples of stimuli at varying levels of difficulty. Top: contrast modulations at (L-R) 30%, 50%, and 75% modulation depths. Bottom: orientation modulations at (L-R) 60%, 80%, and 100% modulation depths.

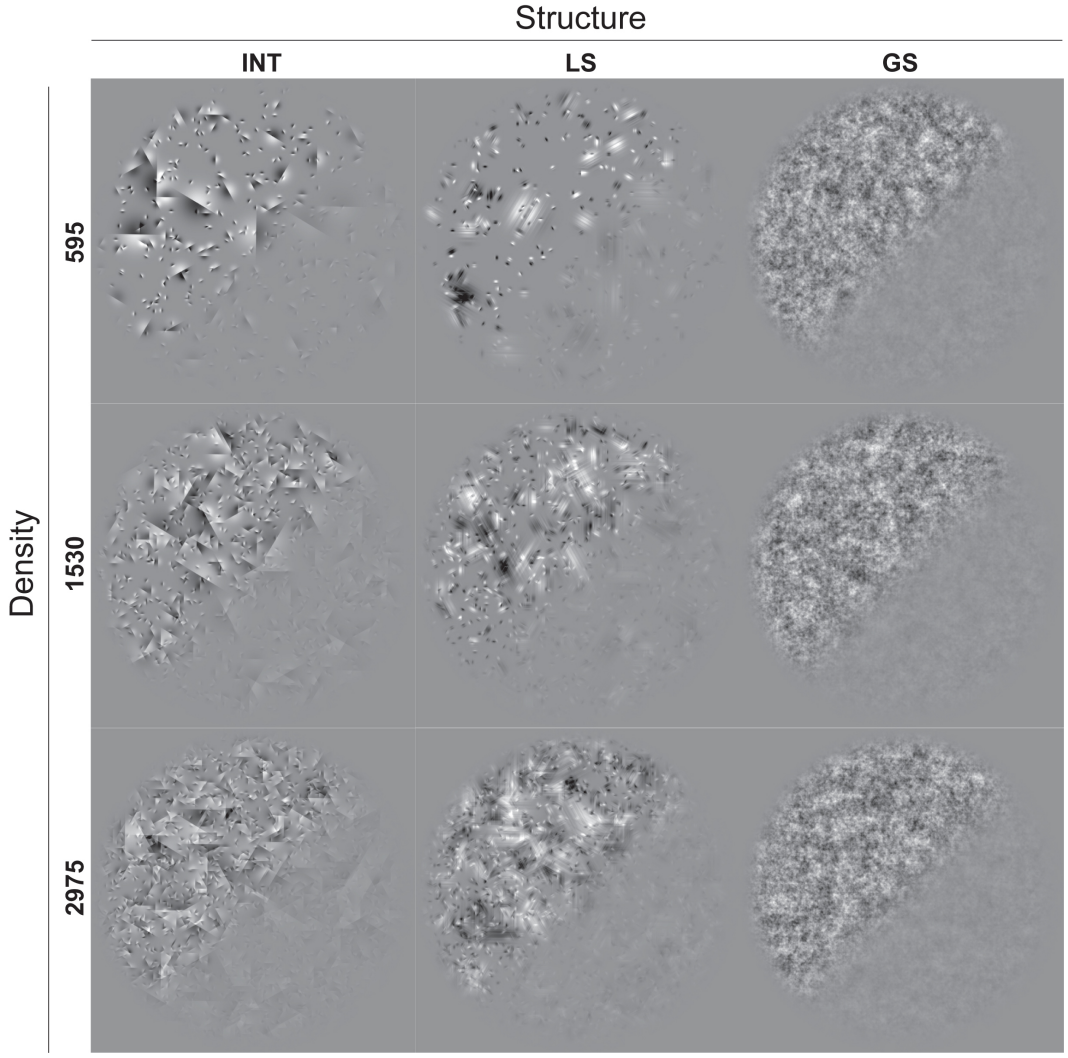

Figure 3–4: Examples of contrast modulated stimuli used for Experiment 1, shown at a modulation depth of 50%. The phase alignment conditions are arranged horizontally, while the density is increased vertically from top to bottom. The three phase alignment conditions are intact (INT), locally scrambled (LS), and globally scrambled (GS). Notice that the globally scrambled condition does not appear different at varying densities because density information is destroyed by phase scrambling.

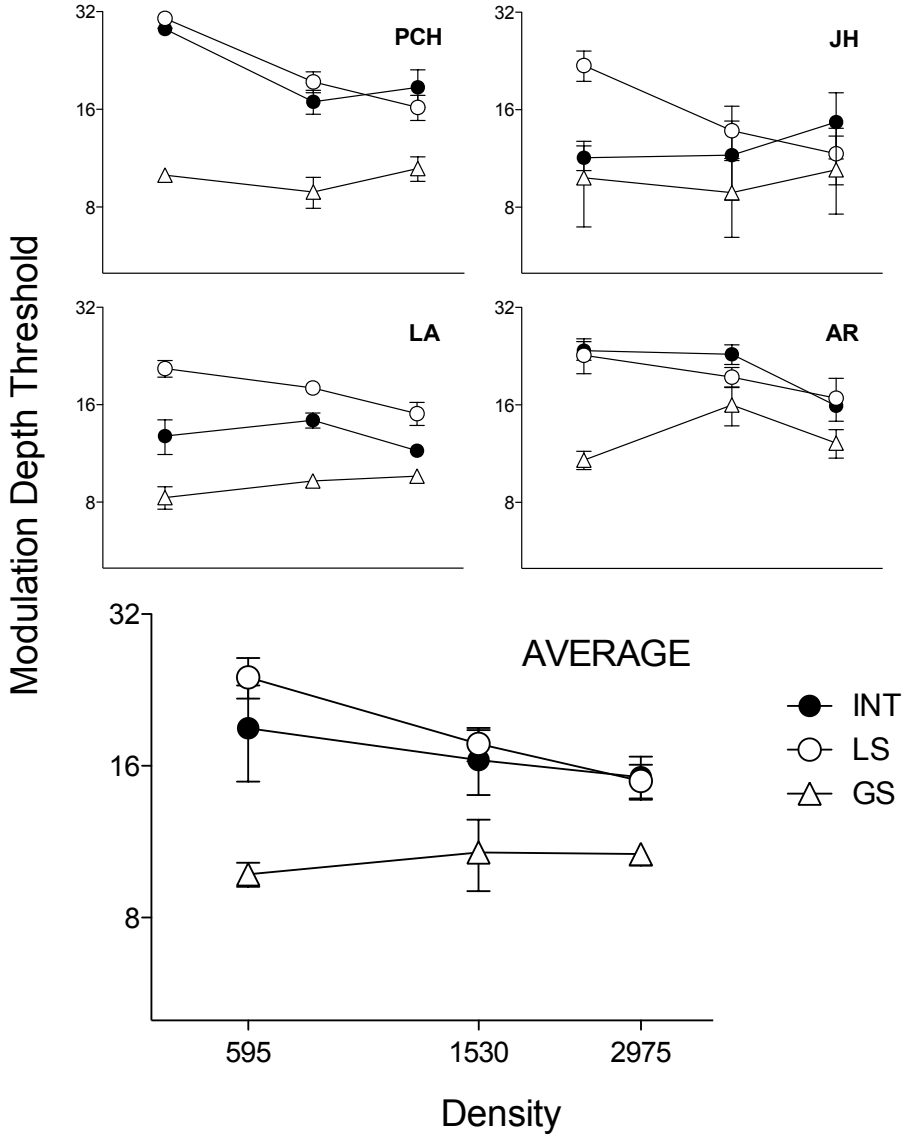

Figure 3–5: Experiment 1 (contrast boundary segmentation) results for four observers (small graphs), and the average of these observers (large graph). The structure conditions are indicated by the data series: intact (INT), filled circles; locally scrambled (LS), open circles; and globally scrambled (GS), open triangles. Density increases along the abcissa. Note improved performance for phase scrambled carriers (GS) and lack of effect of density in this condition. In contrast, intact and locally scrambled conditions both result in higher thresholds at low densities than at high. Error bars indicate standard errors.

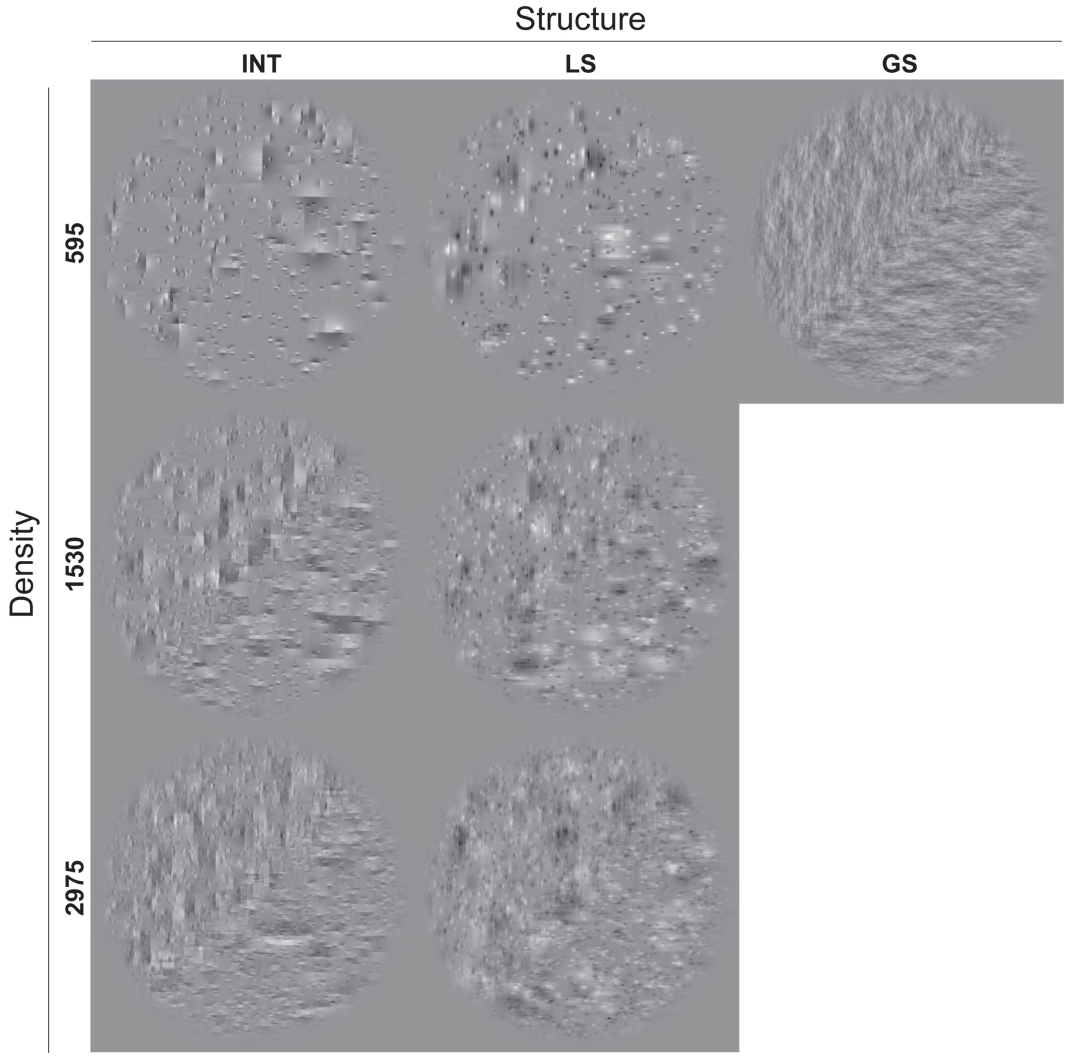

Figure 3–6: Examples of orientation modulated stimuli used for Experiment 2, shown at a modulation depth of 100%. The phase alignment conditions are arranged horizontally, while the density is varied vertically. The globally scrambled condition (GS) was only tested once, because density information is destroyed by phase scrambling and (as expected) no systematic effect of density was observed in the previous experiment.

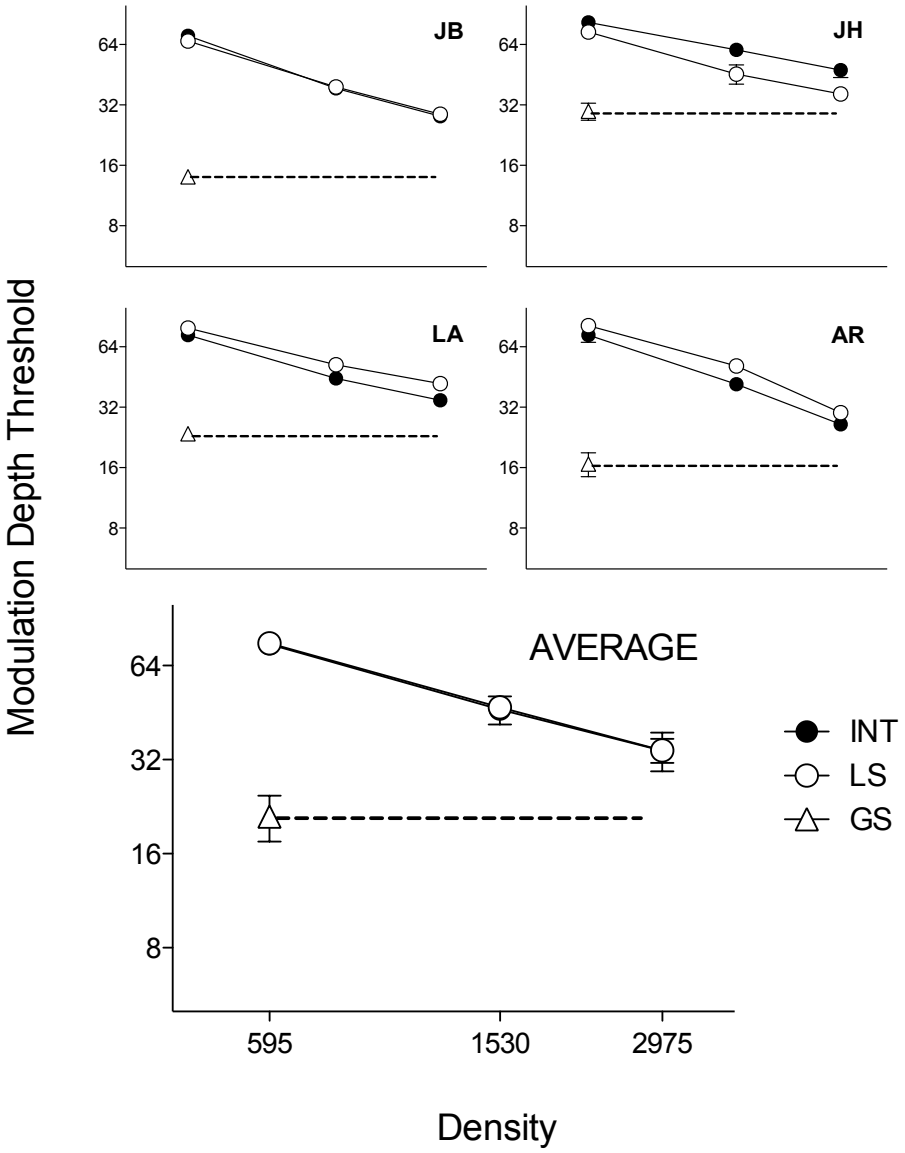

Figure 3–7: Experiment 2 (orientation boundary segmentation) results for four observers (small graphs), and the average of these observers (large graph). The structure conditions are: intact (INT), filled circles; locally scrambled (LS), open circles; and globally scrambled (GS), open triangles. These results show improved performance for phase scrambled carriers (GS) as previously found for contrast segmentation. There is no systematic difference between the intact and locally scrambled conditions, as both result in higher thresholds at low densities than at high. Error bars indicate standard errors.

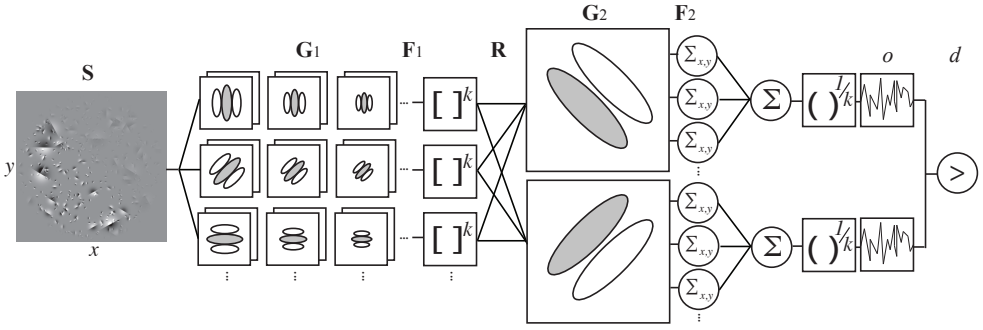

Figure 3–8: Architecture of our FRF-style model. Symbols at top correspond to equations in text. $\mathbf{G}_1$ and $\mathbf{G}_2$ stand for the first- and second-stage filters respectively, $\mathbf{F}_1$ and $\mathbf{F}_2$ stand for the filter outputs, $\mathbf{R}$ for the output of the rectification. The output of each filter following the injection of noise is indicated by o, and the result of the comparison between the files by d.

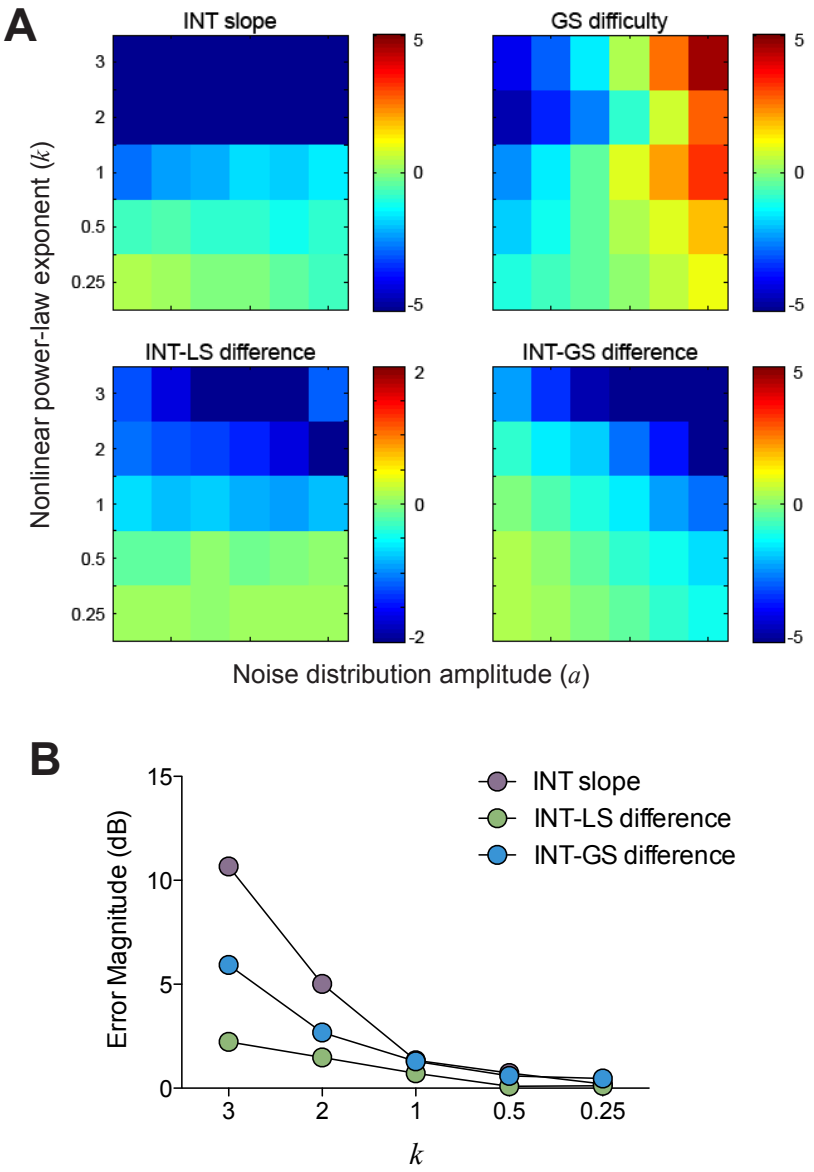

Figure 3–9: Model optimization process. (A) Two-parameter error space for the model. Intensity in each image is the disparity in decibels between model and human performance. Four performance metrics (INT slope, GS difficulty, INT-LS difference, and INT-GS difference) were computed for each combination of power-law exponent and decision noise level (increasing from left to right). The standard deviation of the decision noise distribution was different for each value of k, and its range was selected so that underestimates and over-estimates GS difficulty were represented. (B) Error magnitude for each value of k in the three remaining metrics when the noise level is optimized on GS difficulty.

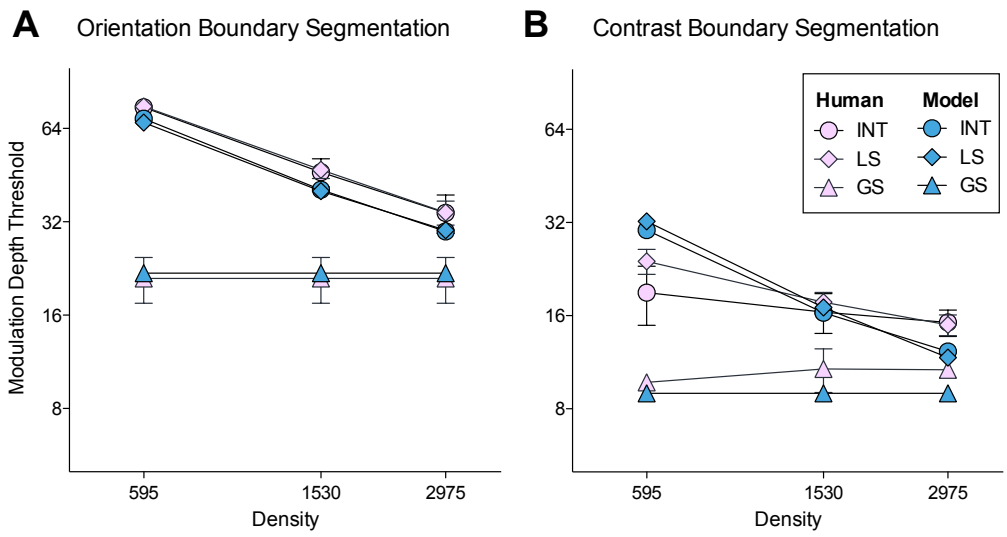

Figure 3–10: Results from model in Figure 9 with parameters k and a optimized for orientation boundary segmentation. (A) Orientation boundary segmentation results are similar to those for humans: lower thresholds for GS and higher thresholds for INT and LS that decrease as density increases. (B) Contrast boundary segmentation results are also similar to those for humans: lower thresholds for GS, higher thresholds for INT and LS that decrease as density increases.

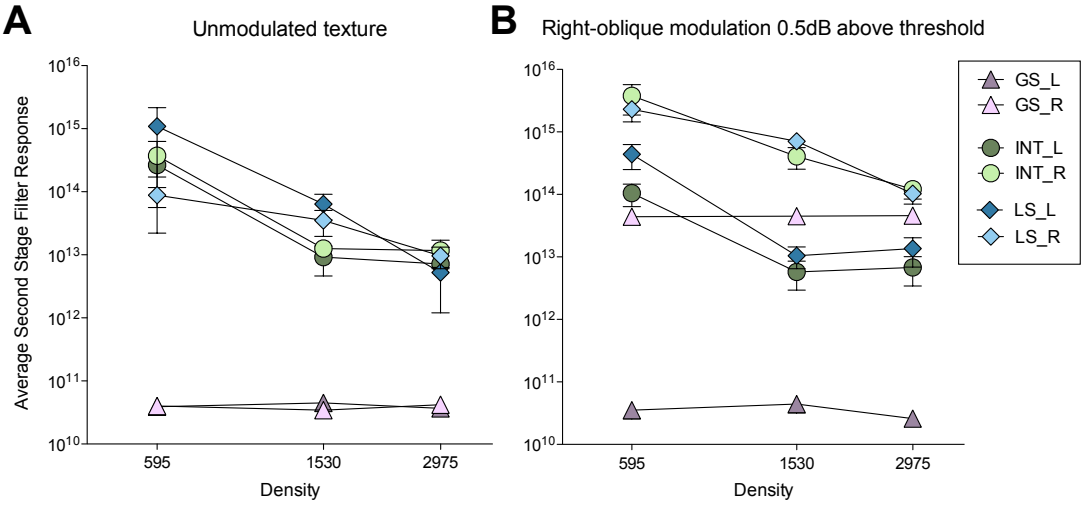

Figure 3–11: Second-stage filter responses ($\mathbf{F}_2$) measured for each of the conditions tested in Experiment 1 using k=0.25. Responses are averaged over ten randomly-generated stimuli, and the error bars indicate standard error. (A) Second-stage filter responses to an unmodulated stimulus. The measured response reflects the second order structure present in the texture before modulation. Responses for both left (L, dark colours) and right (R, light colours) filters are the same, because whatever structure is present is non-oriented. (B) Second-stage filter responses to a stimulus modulated with a right-oblique boundary with a modulation depth 0.5 dB above threshold in each condition. In this case, the responses for the right oblique filters are higher signalling the presence of a boundary.

4 Texture segmentation is enabled by sparseness and global phase structure, but not local phase alignments

In the previous chapter, I demonstrated that sparseness influences segmentation, and that this influence could be modeled using a filter-rectify-filter model with a compressive rectifying nonlinearity. In the following chapter, I use pairs of the same synthetic textures to determine the extent to which differences in texture structure enable segmentation. The results suggest that global phase structure and texture sparseness can both enable and influence segmentation in the absence of any lower-order statistical cues for segmentation. I found that the same model employed for contrast and orientation boundary segmentation data in the previous experiments can also account for the role of global structure and density, but not the role of local phase alignment, in segmentation performance on these kinds of texture boundaries.

CHAPTER 4
Texture segmentation is enabled by sparseness and global phase structure, but not local phase alignments

4.1 Abstract

Second-order properties, which describe an image's energy content, enable segmentation when they are different on either side of a boundary. We have demonstrated that the spatial distribution of the energy in an image (higher-order statistics), particularly sparseness and global phase structure, influences segmentation of contrast and orientation boundaries (Chapter 3), but that local phase structure does not. Here, we examine whether higher-order statistics enable segmentation with boundaries defined by differences in structure and density. We used naturalistic synthetic textures to manipulate the sparseness, global phase structure, and local phase alignments of carrier textures, and measured segmentation thresholds based on forced-choice judgments of boundary orientation. We found that both global structure and sparseness enable segmentation, and that local structure affects segmentation thresholds in both these cases. Using a two-stage filter model, we demonstrated that the same compressive intermediate nonlinearity best accounts for the segmentation of boundaries defined by contrast, orientation, structure and density. However the model architecture we employed, regardless of parameter values, does not properly account for the influence of local structure on structure or boundary segmentation.

4.2 Introduction

Boundary segmentation, one of the visual system's most fundamental tasks, occurs when segmentation mechanisms detect a discontinuity in at least one of several first- or second-order properties such as luminance, orientation or contrast. Second-order properties, which describe an image's energy content, *enable* segmentation when they are different on either side of a boundary. We have demonstrated that the spatial distribution of the energy in an image (higher-order statistics), particularly sparseness and global phase structure, *influences* segmentation of contrast and orientation boundaries (Chapter 3), but that local phase structure does not. Occlusion boundaries in the natural world often have broadband texture on both sides of the boundaries, which is not typically accounted for in the studies of contrast and orientation boundary segmentation. It remains to be seen whether higher-order statistics, either global or local, enable segmentation. Here, we test the ability of global and local structure to enable segmentation as a function of density, as well as the ability of sparseness itself to enable segmentation for different kinds of local structure.

The set of texture statistics that can enable segmentation has been a subject of intense research in the past. 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 be sufficient (Olson & Attneave, 1970; Julesz, 1981a). 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 could be accounted for with this analysis. However, local properties not captured by the Fourier amplitude spectrum have been shown to

affect segmentation. Malik and Perona (1990) and Motoyoshi and Kingdom (2007) showed that local contrast polarity can enable segmentation, and Graham et al. (1993) used element arrangement patterns of oriented grating patches to show that the spatial configuration of local features alone could enable segmentation.

Ideally, we would use natural texture photographs to begin this investigation without eliminating any potentially important statistics, but testing boundaries between pairs of natural textures presents several difficulties. In our pilot work, attempting to use natural textures, we observed both improvements and impairments to segmentation thresholds after removing higher-order statistics, which suggests that individual differences in higher- and lower-order statistics are a critical factor. These results tended to be of an anecdotal nature because psychophysical performance may have depended on the pairing of properties as well as the individual texture statistics. Additionally, because natural textures differ on many dimensions (e.g. orientation bandwidth, dominant orientation, structure), it is not clear how to create boundaries between them with only higher-order differences. To overcome these complications, we employed the same naturalistic synthetic textures as in the previous study, because we know that they capture higher-order statistics that affect segmentation performance.

Most existing models of texture segmentation use a 'filter-rectify-filter' architecture: a bank of oriented filters at a range of spatial frequencies, followed by a pointwise nonlinearity, and finally, a second stage of linear filtering. In this scheme, the role of the first stage filters is to characterize the energy present in the texture, the pointwise nonlinearity prevents peaks and valleys of a luminance-balanced texture from cancelling each other out, and the second stage of filtering serves to detect the boundary. These

models have been shown to account for a wide range of texture segmentation phenomena (Landy & Graham, 2004), and we have shown (Chapter 3) that if the intermediate nonlinearity has a compressive shape, this type of model can also account for the influence of higher-order statistics on second-order texture segmentation.

Here, we will also investigate the ability of the model to account for human psychophysical performance on segmentation of structure boundaries; boundaries defined by the structure of the energy in the texture rather than the type of energy that is present. As initially proposed (Bergen & Adelson, 1988; Malik & Perona, 1990), the FRF model has a square-law pointwise nonlinearity that measures the energy present in the texture, essentially its lower-order statistics. Previously, we found that a compressive nonlinearity in the second-order segmentation mechanism accounted for the influence of higher-order statistics. Here we will evaluate this model with boundaries defined by higher-order statistics, to assess its ability to capture the extent to which structure and sparseness influence and enable structure boundary segmentation.

4.3 General Methods

4.3.1 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, 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.

Textures

Synthetic micropattern textures were created in the manner described in Chapter 3. They 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 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. Locally scrambled (LS) textures were very similar, but the sine wave components were phase-randomized before windowing to eliminate local edge structure. 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 et al., 2002). Within the INT and LS conditions, density was varied by changing the total number (595, 1530, or 2975) of micropatterns used to create the texture.

Textures were generated on a trial-by-trial basis and subjected to the same homogeneity constraints as in the previous chapters to preclude luminance or contrast boundaries caused by unfortunate micropattern placements. In the low density condition, only about 12% of the generated textures passed this test; in 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 0, and its extreme luminance values were clipped at ± 3 standard deviations and scaled to fit in the range of intensities between ± 1.0.

Boundary Creation

Boundaries were created between textures using a quilting method (Watson & Eckert, 1994; Landy & Oruç, 2002) illustrated in Figure 4–1. After two carrier textures were created ($\mathbf{C}_A$ and $\mathbf{C}_B$), 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 ($\mathbf{E}_{x,y}$) is scaled to create modulators ($\mathbf{E}_A$ and $\mathbf{E}_B$):

$$\mathbf{E}_A = \sqrt{(1 + m\mathbf{E}_{x,y})/2} \tag{4.1}$$

$$\mathbf{E}_B = \sqrt{(1 - m\mathbf{E}_{x,y})/2}. \tag{4.2}$$

The modulation depth parameter (m) determines the relative strength of a texture between the two halves (Figure 4–2). 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 present in both halves, but with carrier texture weighted more heavily in one 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 4.1 & 4.2) have been applied.

$$\mathbf{C}'_A = c\,\mathbf{C}_A - \frac{\iint c\,\mathbf{C}_A\mathbf{E}_A - 0.5}{\iint \mathbf{E}_A} \tag{4.3}$$

$$\mathbf{C}'_B = c\,\mathbf{C}_B - \frac{\iint c\,\mathbf{C}_B\mathbf{E}_B - 0.5}{\iint \mathbf{E}_B} \tag{4.4}$$

The final stimulus ($\mathbf{S}_{x,y}$) is the sum of the two carriers, each spatially weighted by their respective envelopes, where L_0 is the mean luminance:

$$\mathbf{S}_{x,y} = L_0\{1 + \mathbf{C}'_A\mathbf{E}_A + \mathbf{C}'_B\mathbf{E}_B\}. \tag{4.5}$$

4.3.2 Apparatus and Observers

The stimuli were presented on a CRT monitor (Sony Trinitron Multiscan G400, 81 cd/m2, 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 grey background. Observers viewed the stimuli from a distance of 114 cm, resulting in a stimulus visual angle of approximately 6.5 degrees. The experiment was run on a Macintosh (Desktop Pro, Mac OSX) using Matlab and PsychToolbox (Brainard, 1997; Pelli, 1997).

4.3.3 Task

Observers were instructed to fixate a mark at the centre of the screen and initiate the trial with a button press. The stimulus was displayed for 100ms and the observers indicated whether the boundary between the

textures was left or right oblique, with a button press. 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.

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. Subjects 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 600 trials per condition.

4.3.4 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 ANOVAs or paired-samples t-tests with a criterion of $\alpha = 0.05$. Effect size (D. Klein, 2005) was measured using Cohen's d with the standardizer computed as:

$$s = \sqrt{\sigma_1^2 + \sigma_2^2}/2. \tag{4.6}$$

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

4.4 Experiment 1: Structure Boundary Segmentation

That changes in orientation and contrast enable segmentation is well-documented, and we have demonstrated that contrast and orientation boundary segmentation mechanisms are affected by higher-order image statistics (Chapter 3). Whether this effect extends to the ability to enable segmentation based on differences in higher-order texture statistics has yet to be examined. Here, we test directly whether differences in higher-order statistics alone can enable segmentation, and examine the influence of sparseness and local phase structure on segmentation thresholds.

4.4.1 Methods

In this experiment we measured modulation depth thresholds for observers segmenting boundaries defined by phase structure. We tested all pairwise combinations of three phase structure conditions. The 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. Density was varied parametrically for the INT and LS textures by randomly placing 595, 1530, or 2975 micropatterns on a texture canvas. Two textures were generated for each trial, and quilted as described in the General Methods and illustrated in Figure 4–1 to create a unique stimulus. We created boundaries testing all combinations of the phase alignment conditions for three structure boundary conditions: INT/LS, INT/GS, and LS/GS, and tested them at three density levels each (see Figure 4–3 for examples). All observers had normal or corrected-to-normal vision, and JH, JB, and AR were naive to the purpose of the experiment.

4.4.2 Results

Structure boundary segmentation results are shown in Figure 4–4. 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 (Figure 4–4, open circles), regardless of density. Because only half of the observers could perform the task at any modulation depth less than 100%, the INT/LS condition was excluded from further analysis. Performance for boundaries between textures with and without global structure (LS/GS and INT/GS, open triangles and filled circles in Figure 4–4) were considerably better. In both conditions the thresholds increased with density. Thresholds in the INT/GS and LS/GS conditions were 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 run on only the LS/GS and INT/GS conditions confirmed a main effect of structure $F(1,9) = 41.01, p < .05$ and a main effect of density $F(2,9) = 6.67, p < .05$, as well as a significant interaction between structure and density $F(2,9) = 26.87, p < .05$. Post-hoc Bonferroni tests (Table4–1) confirmed that the significant difference between the INT/GS and LS/GS conditions was at the highest density only ($t = 9.52, p < .05, d = 1.24$).

Globally scrambled (GS) textures have particularly low sparseness, and so it is likely that at least one of the properties enabling segmentation is a difference in sparseness on either side of the boundary. These results indicate that segmentation by structure is easiest when there is a large difference in sparseness between the textures being segmented. They suggest that for highly sparse textures (i.e., INT and LS at low density), differences in local phase structure have little impact on segmentation, but at high

Density	t	p	d
INT/GS vs. LS/GS			
595	0.429	$> .05$	0.30
1530	2.005	$> .05$	0.69
2975*	9.516	$< .05$	2.46

Table 4–1: Results of Bonferroni post-hoc tests for structure boundary segmentation. t is the Bonferroni-corrected t statistic, p the significance level, d Cohen's d, and * indicates a statistically significant comparison.

densities a lack of local phase structure impairs segmentation. In this Experiment, we demonstrated that higher-order statistics, particularly global phase alignments, can enable segmentation. We showed that at low densities, local edge structure is not relevant to segmentation, but its presence may improve performance at high densities where global structural differences are diminished. Part of the reason the higher densities are more difficult is likely because decreased sparseness is a major consequence phase scrambling when density is increased the boundary becomes harder to segment. We also demonstrated that differences in local phase structure alone (i.e., INT/LS) provide poor support for segmentation.

4.5 Experiment 2: Density Boundary Segmentation

In Experiment 1, we found that boundaries defined by a difference in global structure enable segmentation by creating boundaries between locally scrambled or intact and globally scrambled textures. Segmentation became more difficult as the density in the intact or locally scrambled textures was increased by including more micropatterns in the texture. This suggested that the difference in sparseness across the boundary was an important segmentation cue. Here, we isolate density to determine the extent to which it enables segmentation, and examine the influence of local structure on density segmentation.

4.5.1 Methods

In this experiment, we used two types of stimuli: intact textures with density boundaries and locally scrambled textures with density boundaries. Globally scrambled textures were not considered because their sparseness does not vary. To create density boundaries, we quilted two carrier textures: one with 595 micropatterns, and the other with 2975 micropatterns. We measured segmentation thresholds for density boundaries in intact and locally scrambled textures, as shown in Figure 4–5. All observers had normal or corrected-to-normal vision, and JH, JB, and AR were naive to the purpose of the experiment.

4.5.2 Results

Density boundary segmentation results for four observers are shown in Figure 4–6. All of the observers were able to segment the boundaries, and we had no difficulty obtaining thresholds for any observer or condition. We found that density boundaries enable segmentation at approximately the same level of difficulty as structure segmentation. 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 $t(3) = 8.01, p < .05, d = 0.96$, but the effect size is modest. It appears that the density boundary is slightly easier to segment for locally scrambled textures.

In this experiment, we demonstrated that differences in micropattern texture density support segmentation without any other differences in global phase structure. Unexpectedly, we found a small but consistent improvement in density-based segmentation for locally scrambled textures compared to intact textures.

4.6 Model

In the past, three-stage models with an additional intermediate nonlinearity (FRFRF) have been proposed to account for the role of structure (Graham et al., 1993; Kingdom et al., 2001; Landy & Graham, 2004). In Chapter 3, we determined that a modified filter-rectify-filter (FRF) model with a compressive intermediate nonlinearity, a final expansive nonlinearity, and decision noise could account for the influence of structure in our orientation and contrast boundary segmentation data. Here, we examine whether that model can also account for segmentation of boundaries defined by structure or density, or if another model better accounts for performance.

4.6.1 Filter-Rectify-Filter Model

We implemented a basic filter-rectify-filter model (Figure 4–7A), as described in the previous study. First the stimulus ($\mathbf{S}_{x,y}$) was convolved ($*$) with a bank of linear filters ($\mathbf{G}_1$) that varied in orientation (θ), spatial frequency (ω), and phase (ϕ) (Equation 4.7):

$$\mathbf{F}_1(\theta, \omega, \phi, x, y) = \mathbf{G}_1(\theta, \omega, \phi) * \mathbf{S}(x, y). \tag{4.7}$$

The filters were log-gabors (Kovesi, 2000), at two phases (even and odd), six orientations (evenly spaced with their bandwidths chosen for approximately uniform coverage), and four spatial frequencies (160, 80, 40, and 20 cpi, each with a bandwidth of approximately 1.5 octaves). The output of each of these filters ($\mathbf{F}_1$) was weighted (w_f), full-wave rectified and raised (pointwise) according to a power-law of order k, then pooled over phase (Equation 4.8):

$$\mathbf{R}(\theta, \omega, x, y) = \sum_{\phi} |\mathbf{F}_1(\theta, \omega, \phi, x, y) \cdot w(f)|^k. \tag{4.8}$$

The values of w_f were chosen to equalize responses across spatial scales for a stimulus with a $1/f$ spectral falloff (Field, 1987). To this end, the responses to the higher spatial frequency channels were magnified relative to the responses to lower spatial frequencies, using a weighting function $w(f) = 2f$, where f is an index of spatial frequency with $f = 1$ designating the lowest spatial frequency. Dot products were computed between these responses ($\mathbf{R}_{x,y}$) and two second-stage filters ($\mathbf{G}_2$) in the form of low spatial frequency sine-phase gabor functions that match the two possible orientations of the boundary in the stimulus (45 and -45 degrees), as well as its central position:

$$\mathbf{F}_2(45, \theta, \omega) = \mathbf{G}_2(45) \bullet \mathbf{R}(\theta, \omega, x, y) \tag{4.9}$$

$$\mathbf{F}_2(-45, \theta, \omega) = \mathbf{G}_2(-45) \bullet \mathbf{R}(\theta, \omega, x, y). \tag{4.10}$$

The outputs (o) were computed by pooling the magnitudes of the late-stage filter responses across the orientations and spatial frequencies of the early-stage filters. The pooled response magnitudes were raised to a power (the reciprocal of k), and then combined with additive decision noise for the final output value. (Equations 4.11 & 4.12). The noise values n_1 and n_2 were drawn from a normal distribution with a mean of 0 and whose standard deviation, or amplitude, a is a free parameter of the model.

$$o_{45} = \left(\sum_{\theta,\omega} |\mathbf{F}_{2_{45}}(\theta, \omega)| \right)^{1/k} + n_1 \tag{4.11}$$

$$o_{-45} = \left(\sum_{\theta,\omega} |\mathbf{F}_{2_{-45}}(\theta,\omega)| \right)^{1/k} + n_2 \tag{4.12}$$

These outputs were compared, and the late-stage filter (left- or right-oblique) with the strongest response determined the decision (d) of the model (Eq. 4.13):

$$d(o) = \begin{cases} 45 & \text{if } o_{45} \geq o_{-45} \\ -45 & \text{if } o_{-45} > o_{45} \end{cases} \tag{4.13}$$

4.6.2 Simulation

We used the same method to determine segmentation thresholds for the model as we used for our human participants. The model made left- or right-oblique decisions for each of 60 trials for nine structure boundary and two density boundary conditions on 10 logarimithically-spaced modulation depth levels producing performance ranging from chance to near-perfect. We measured a percent-correct value for each level and stimulus condition, and then fit a cumulative Gaussian function to determine the model's threshold. Because the stimuli were randomly generated on each trial, the model's results varied from one simulation to another. For this reason, we simulated the experiment four times and averaged the resulting thresholds. Standard error was determined across the four model runs. Each run required 36 hours of computational time. The summed square error was calculated as:

$$SSE = \sum_i (\log h_i - \log m_i)^2, \tag{4.14}$$

where h is the series of human thresholds averaged over observers and m is the series of model thresholds averaged over the four simulations. This

model described above has two free parameters: the order of the power-law nonlinearity (k), and the noise amplitude (a) (i.e., standard deviation of the distribution from which the decision noise is drawn). Previously (Chapter 3), we determined optimal values of these parameters based on segmentation performance for orientation boundaries. We used those optimal values to predict human performance in segmenting boundaries defined by structure and density.

4.6.3 Simulation Results

The results of the structure boundary simulation are shown in Figure 4–7B. Thresholds for boundaries between intact and globally scrambled textures (INT/GS) are estimated very well, especially at the lowest and highest densities. Thresholds for boundaries between locally scrambled and globally scrambled textures (LS/GS) present a bigger problem for the model. The model matches human performance well at the highest density, but over-estimates thresholds at low and moderate densities. Like human observers, the model could not segment INT/LS boundaries reliably enough to estimate a threshold. The total sum-of-squares error for this model was 0.315. Figure 4–7C shows the results of the density boundary segmentation simulation. Thresholds for density boundaries between intact textures are well estimated, but thresholds for boundaries between locally scrambled textures were substantially over-estimated. The total sum-of-squares error for this model was 0.136.

4.6.4 Optimization

Using the model parameters estimated from the orientation boundary segmentation data of the previous study (Chapter 3), the model produces good threshold predictions for structure and density segmentation. However with these parameters the model somewhat over-estimates thresholds

for sparse, locally scrambled textures in both the structure and density experiments. Here, we tested whether values of the two free parameters could be found that reduce the sum-of-squares error of all the data, as well as provide a better qualitative fit to the human data.

Model performance was simulated for five values of the nonlinear power-law exponent ($k = [3\ 2\ 1\ 0.5\ 0.25]$), and a series of noise levels, chosen separately for each value of k to span the range of under-estimating and over-estimating segmentation thresholds. This was necessary because smaller values of k resulted in much larger responses, which meant that larger noise values needed to be injected to have an effect.

The sum-of-squares error (Equation 4.14) was computed for the model's thresholds for each of these combinations of the two free parameters, for both the structure and density segmentation conditions. The resulting two-dimensional error space is illustrated in Figure 4–8A as a colour map in which darker matrix cells indicate lower error. From this visualization it can be seen that compressive nonlinearities (i.e., $k < 1$) at any level of noise outperform expansive nonlinearities, and that the optimal noise values for density segmentation are lower than those for structure segmentation. The most compressive nonlinearity ($k = 0.25$), as before, produced the lowest error in both the structure and density conditions. Figure 4–8B shows how the error at $k = 0.25$ in these two conditions changed as a function of the noise level. The density condition has the lowest error for the smallest value of noise, but the error is minimized in the structure condition at a higher noise level. The noise level circled in red was selected as a good overall compromise 'optimal' value for these conditions.

4.6.5 Optimization Results

The results of the structure boundary simulation with the revised parameters are shown in Figure 4–8C, with the sum-of-squares error for this condition improved to 0.21. Thresholds for boundaries between intact and globally scrambled textures (INT/GS) are somewhat under predicted at all densities, perhaps slightly more so at lower densities. Thresholds for boundaries between locally scrambled and globally scrambled (LS/GS) textures are estimated better– no single threshold is very far off. However the rate at which thresholds increase with density does not appear to accelerate as in the human data. Boundaries between intact and locally scrambled textures (INT/LS) could not be segmented by humans or by the model. Thresholds for density boundaries are also better fit using the revised parameters, with the sum-of-squares error for this condition improved to 0.136. The thresholds are well-predicted in the intact (INT) condition, but they are still substantially over-predicted in the locally scrambled (LS) condition. The sum-of-squares error for this condition was improved to 0.136. The difference between the model's predictions when optimized for orientation boundaries and applied to all boundary types, and its predictions when optimized for the present conditions is minimal, and there is little change in the parameters as well. The optimal value of k is the same, so the sum-of-squares error is reduced simply by reducing the amplitude of the decision noise (a).

These results demonstrate that a simple two-stage model is capable of segmenting boundaries defined by structure or sparseness, but not the influence of local structure. More importantly, these results show that with small adjustments to the noise amplitude, the same nonlinear power-law exponent and model architecture are capable of predicting segmentation

thresholds to boundaries defined by contrast, orientation, structure, and sparseness.

4.7 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. We also found that segmentation thresholds for boundaries between locally scrambled and globally scrambled (LS/GS) textures became higher more quickly than in the INT/GS condition as density was increased. In the second experiment we demonstrated that density alone enabled segmentation, and that a randomizing local phase alignments (LS) led to slightly better performance. We modelled these results using the same two-stage filter model as used to predict thresholds for orientation and contrast boundary segmentation and found that the same, highly compressive, nonlinearity best predicts these results as well. However to minimize the sum-of-squares error in the threshold predictions for these experiments the amplitude of the decision noise was decreased slightly from the value fitted to orientation boundary segmentation. This model captured the ability of global phase structure to enable segmentation as well as the influence of global structure on those thresholds, yet did not account for the influence of local phase structure on either structure or density segmentation thresholds.

4.7.1 Psychophysics

In the structure segmentation condition where the boundary was signalled only by the presence or absence of local phase alignments (INT/LS), performance varied substantially between participants. All participants were able to segment the boundaries above chance performance when the modulation depth was 100%, but performance was not sufficiently high to provide a psychometric function, and in some cases did not reach the threshold level (75% correct). The boundaries in the INT/LS condition are

evident by inspection (Figure 4–3), and may be segmented more reliably with a longer viewing time, but this segmentation-by-inspection might mean that some other (potentially more cognitive) mechanism is at work. This borderline case that cannot be reliably segmented under a 50ms exposure casts doubt on the early segmentation work where the extent to which a texture pair could be segmented depended on unlimited viewing time and subjective assessments of effortlessness (e.g. Caelli & Julesz, 1978; Julesz, 1981b). There is a substantial amount of variability between subjects in this condition, which may be due to practice effects. The most practiced subjects 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, but less practiced subjects 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. This means that it is difficult to assess the effect of sparseness in the INT/GS and LS/GS conditions because it is part or all of what enables segmentation. Given this relationship, it makes sense that performance degrades as the density of the INT or LS texture is increased. The effect of local structure, however, is difficult to understand. Orientation and contrast boundary segmentation performance is the same for the intact and locally scrambled conditions at a given density, which suggests that local phase alignment is discarded by the mechanisms that segment these boundaries. On the other hand, in structure segmentation, performance is the same in the INT/GS and LS/GS conditions at low densities, but different at the highest density tested. This suggests that although local structure is ignored

by segmentation mechanisms when global differences (sparseness in particular) are sufficient for segmentation, it can be used to boost performance when the global differences are weak. Furthermore, in the density segmentation task, we see a small but consistent performance improvement in the LS condition. Overall, we have consistent evidence of structure segmentation mechanisms being selective for local structure differences while no such evidence exists for orientation or contrast segmentation mechanisms.

4.7.2 Model

The model's ability to segment structure is due to the compressive nonlinearity employed. The stimuli are RMS balanced, so there is little difference in the energy on either side of the boundary (which is evident in the model's extremely poor performance when the intermediate nonlinearity is a square-law). A power-law exponent other than two applied before responses are pooled over space either produces a bigger response for the high local contrast of the sparse textures (expansive), or boosts the relatively low local contrasts of dense textures (compressive). Figure 4–9 shows the effect of the power-law exponent on the image energy that is passed into the second-stage filter. The amount of energy following a power-law exponent other than 2 is dependent on density, and thus enables segmentation based on structural sparseness. That the LS/GS condition is more difficult for the model to segment than the INT/GS condition might stem from the fact that intact (INT) edgelets are slightly more sparse than locally scrambled (LS) edgelets.

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 are an example of summation

models, because ultimately the information that is processed is summed - in this case, both across channels and over space. We chose to use Minkowski summation, which raises the arguments to the summation to a particular power, and then the resulting sum is raised to the reciprocal of the same power. Minkowski summation is a common technique used when combining the outputs of several channels (Meese & Summers, 2007). The compressive nonlinearity we employ is in agreement with a number of recent studies showing evidence of sub-additive summation in visual processing, conceptualized in a variety of ways: compressive nonlinearities (Mineault et al., 2012), divisive normalization (Rust et al., 2006), and surround suppression (Tsui, Hunter, Born, & Pack, 2010).

4.7.3 Sparseness

Density has long been considered an independent texture property due to its status as an adaptable feature (Durgin, 1995), dissociable from luminance spatial frequency adaptation (Durgin & Huk, 1997). It is evident from the results covered here and in Chapter 3 that texture sparseness is extremely important to consider with respect to second-order segmentation mechanisms. Sparseness can be described as the extent to which the energy in the image is clumped into local, high contrast regions. The spatial modulation that results from sparse contrast is, of course, a contrast modulation and as such contributes to the response of second-order neurons. Dakin, Tibber, Greenwood, Kingdom, and Morgan (2011) suggest that perception of sparseness can be well-measured by the ratio of higher to lower contrast spatial frequency information. The results of this study indicate that, with a compressive or expansive intermediate nonlinearity, density can, in some circumstances, be considered an intensive property that, like contrast, affects strength of the response but is not a tuned property.

4.7.4 Structure-defined boundary processing

Boundaries defined by structure in the absence of differences in the amplitude spectrum have not been the subject of much study. Julesz (1981b) and Victor et al. (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 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 et al. (1993) used element arrangement patterns that did differ in only their higher-order statistics, as we have defined the term. They used a three-stage model (FRFRF) to account for the extra comparison that would need to be made to segment the textures. Here, we demonstrated that a third stage of the model is not necessary, and a single pre-summation compressive nonlinearity is sufficient for differences in the arrangement of the energy in the texture to enable segmentation. Although they used textures composed of elements, their element size was quite large and relatively few were arrayed to form each region. It is questionable whether images composed of these large elements would be processed as 'texture'.

4.8 Conclusions

In these experiments we demonstrated that differences in the structure, rather than simply the composition, of energy in a texture can enable and influence segmentation performance. We presented a two-stage model that accounts for the ability of global phase structure and density to enable segmentation, but not the influence of local phase structure on these thresholds. This model is consistent with the model we used to predict orientation and contrast boundary segmentation thresholds in Chapter 3. Using only two free parameters we have demonstrated that a two-stage filter model can segment boundaries defined only by differences in higher-order statistics, and account for the role of global structure in segmentation.

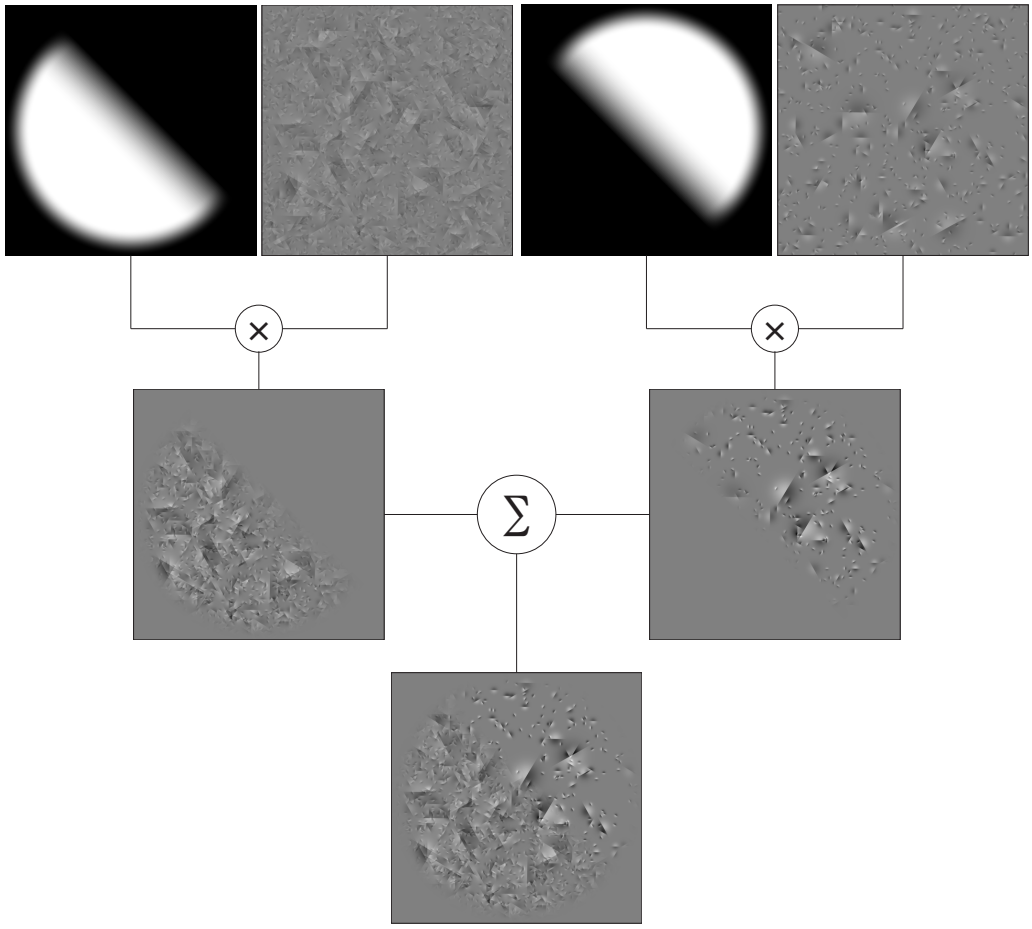

Figure 4–1: Procedure for quilting structure- and density-modulated boundaries. Half-disc envelopes are multiplied with their corresponding carrier textures. These modulated halves are then combined. The modulation depth of the stimulus shown is 100%.

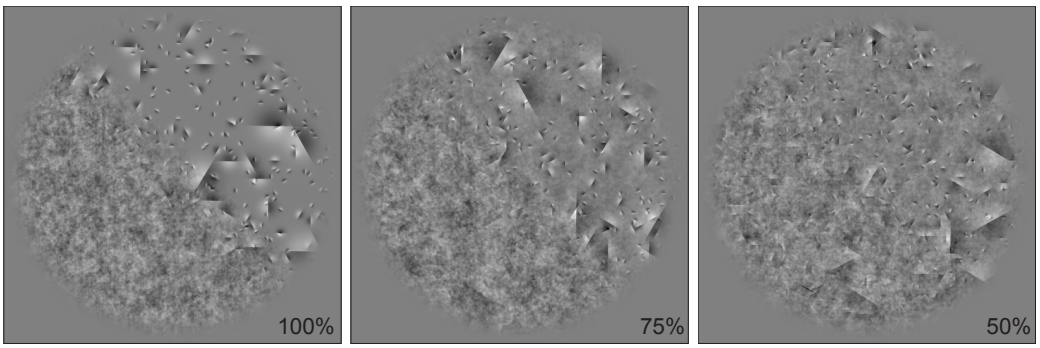

Figure 4–2: An example of stimuli at varying levels of difficulty. This figure shows an INT-GS texture at 100%, 75%, and 50% modulation depth.

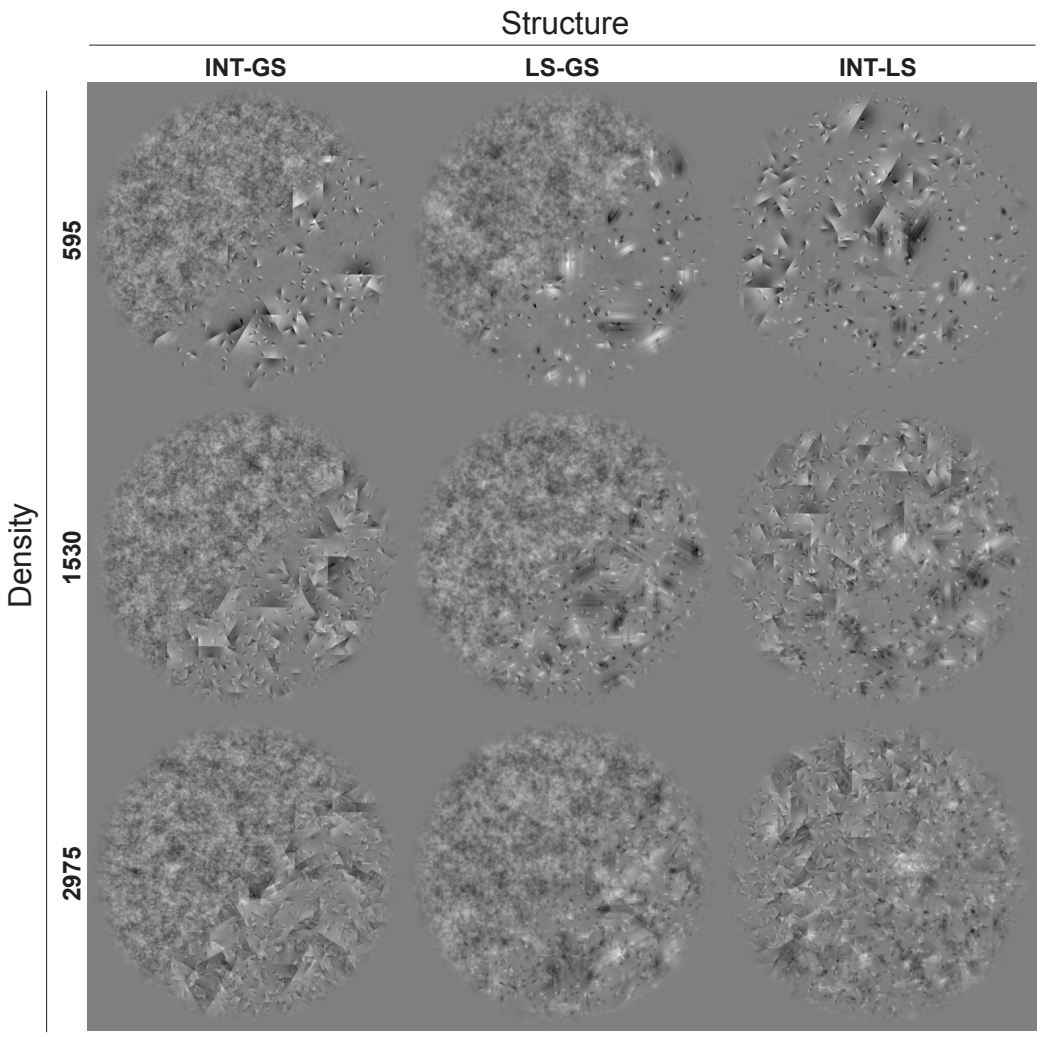

Figure 4–3: Structure modulated stimuli used for Experiment 1, shown at a modulation depth of 100%. The boundary types are organized into columns, while the density conditions are organized into rows.

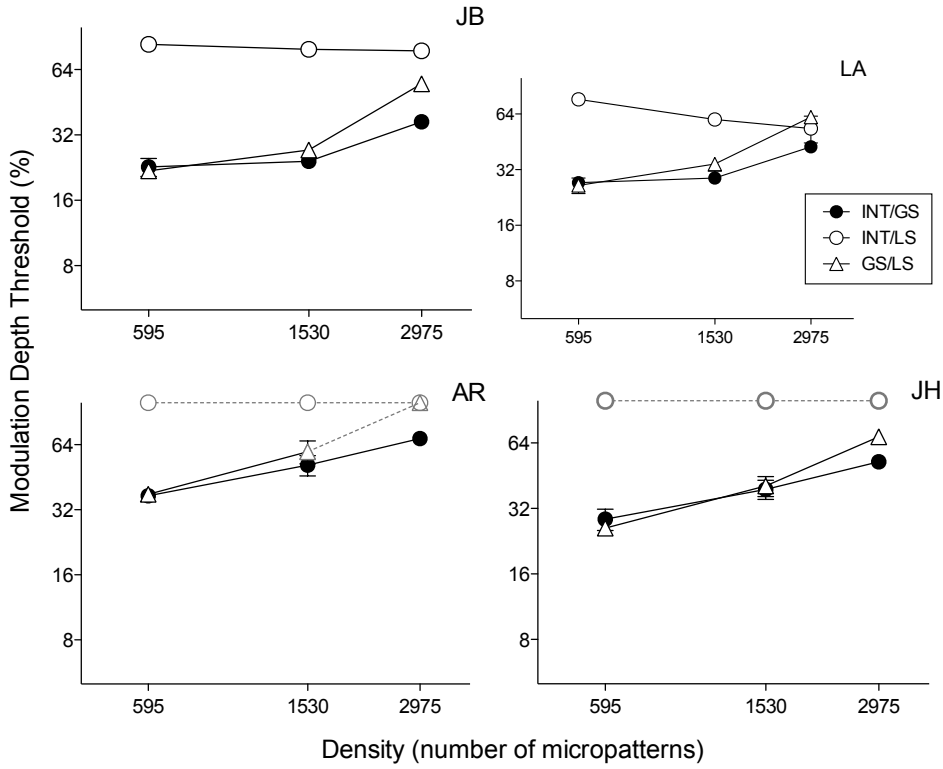

Figure 4–4: Experiment 1 (structure boundary segmentation) results for four observers. The boundary conditions are: INT/GS, between an intact texture and a globally scrambled texture (filled circles); INT/LS, between an intact texture and a locally scrambled texture (open circles); and LS/GS, between a locally scrambled and globally scrambled texture (open triangles). The grey 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 do not appear to be dependent on density. Segmentation performance on the INT/GS and LS/GS conditions is dependent on density with very similar thresholds at low densities, but LS/GS becomes more difficult more rapidly at high densities.

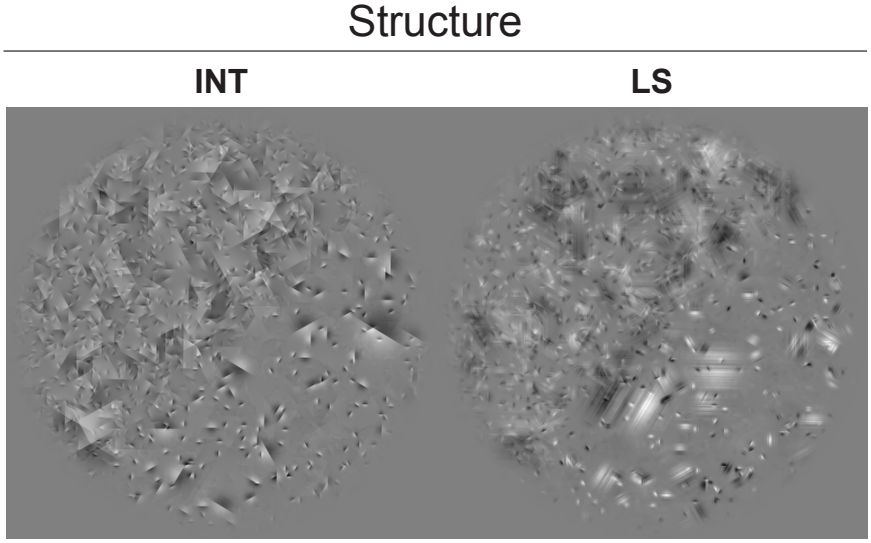

Figure 4–5: Density boundary stimuli used in Experiment 2 at a modulation depth of 100%. Each stimulus is a modulation between a texture with 595 micropatterns and one with 2975 micropatterns. Because no visible (or statistical) boundary was formed in the GS condition, it was not tested.

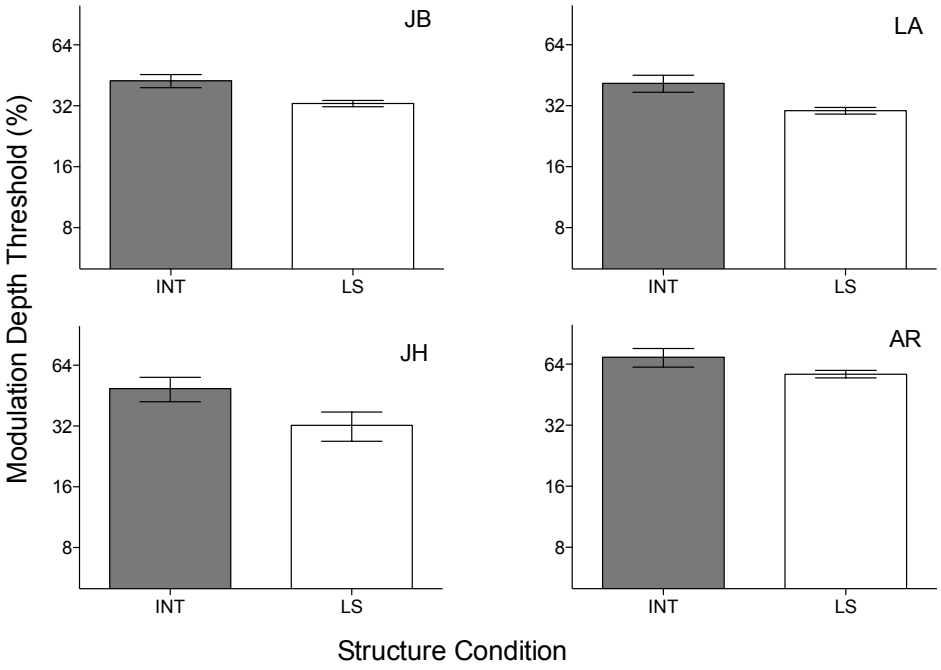

Figure 4–6: Experiment 2 (density boundary segmentation) results for four observers. Density boundaries do not exist following phase scrambling, so the globally scrambled (GS) condition was not tested. Performance on the locally scrambled (LS) condition appears to be slightly but systematically better than on the intact (INT) condition.

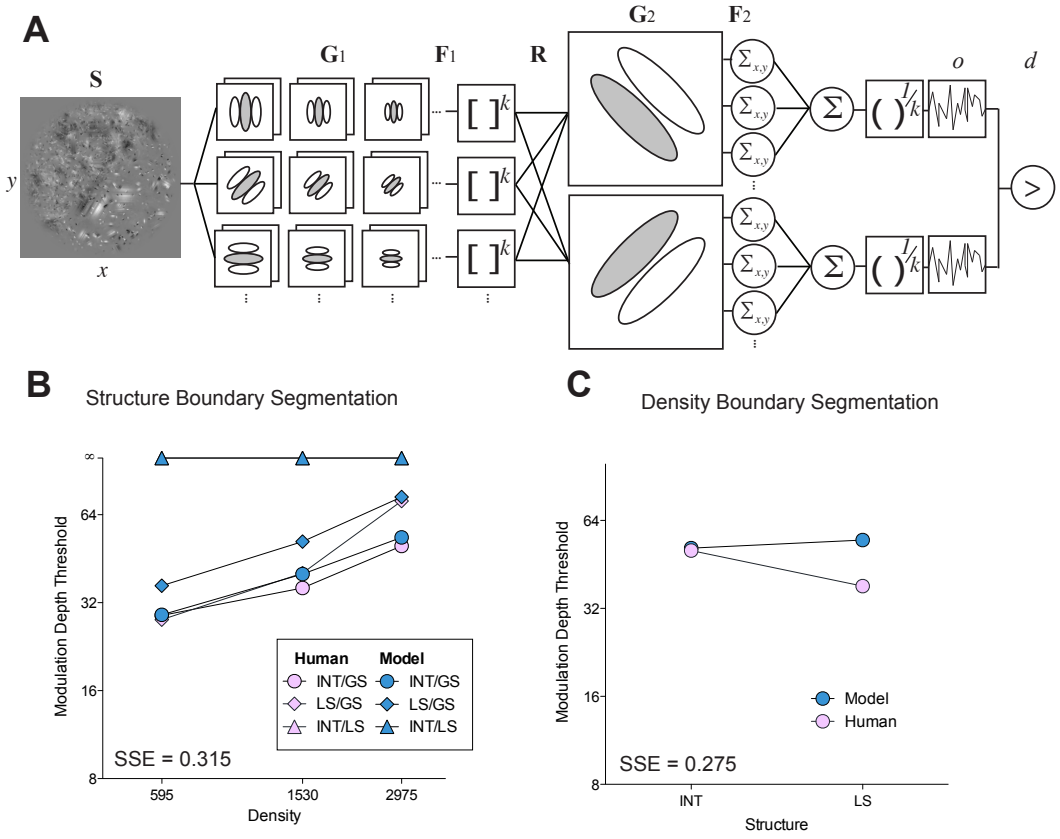

Figure 4–7: Segmentation model architecture and modelling results alongside human data. (A) The architecture of the basic FRF model used in this study. (B) Results for model as fit using orientation segmentation data compared to human data. Human data is in purple, model in black. INT (intact)/GS (globally scrambled) (shaded circles) thresholds are estimated very well, though best at the lowest and highest densities. LS (locally scrambled)/GS (open circles) thresholds present the most difficulty, the model matches human performance at the highest density, but over-estimates thresholds at low and moderate densities. INT/LS (triangles) cannot be segmented reliably enough to compute a threshold for humans or the model, and are depicted here at 100% modulation depth. (C) Results for the model and humans in the density condition. The model performs very well when the textures are intact (INT), but overestimates thresholds when they are locally scrambled (LS).

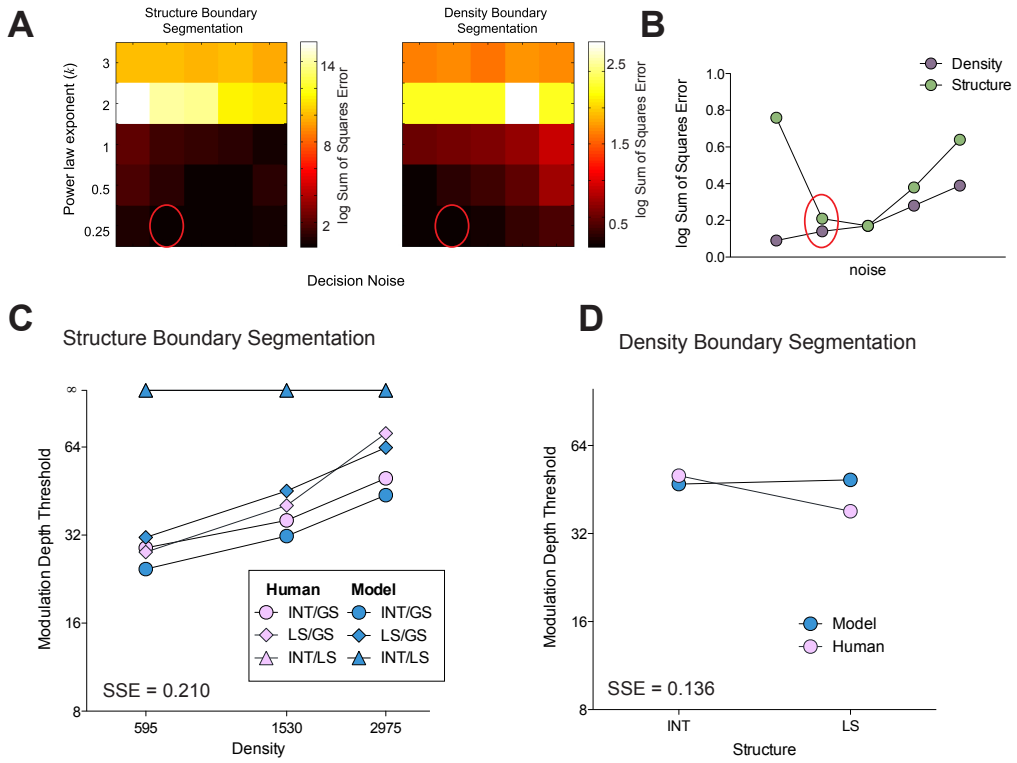

Figure 4–8: Optimization procedure for structure and density boundaries and model results. (A) Matrices of SSE for each noise/exponent parameter value combination. Darker colours indicate lower error. Error is much lower for compressive nonlinearities. Circled points indicate the model evaluated for panels C and D. (B) Error with respect to noise level when the exponent of the nonlinearity is fixed at 0.25 for density and structure segmentation. The error functions have different minima, so a compromise (circled) was chosen for the model evaluated in panels C and D. (C) Structure segmentation results for the model with the noise fit to this particular data to minimize sum-of-squares error. INT/GS (shaded circles) is underestimated at all densities, but more so at low densities. LS/GS (open circles) thresholds are not far off any single estimate, but the rate at which thresholds increase with density does not accelerate as in the human data. INT-LS (triangles) can't be segmented by humans or the model well enough for thresholds to be computed. (D) Density shows a reasonably good estimate in the INT condition, but still over-estimates threshold in LS condition, if less severely.

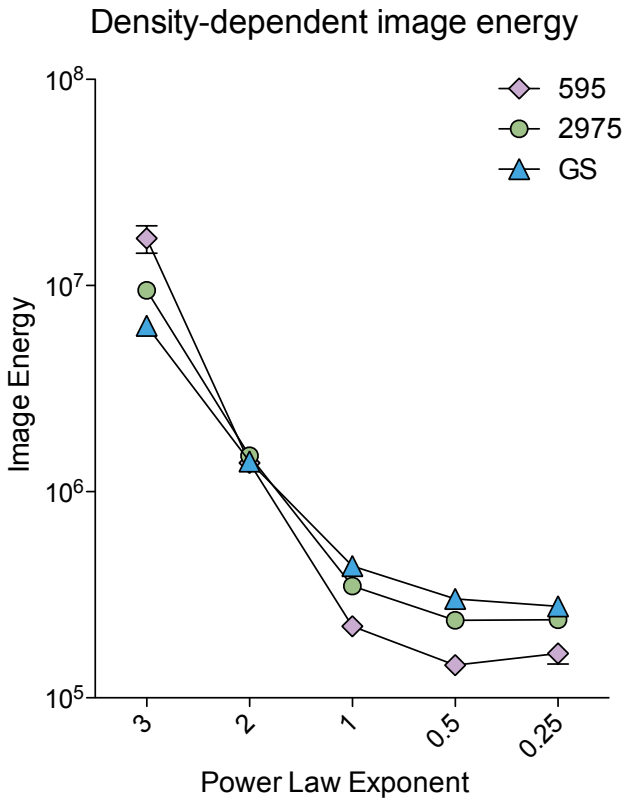

Figure 4–9: Image energy following filtering and rectification, for different texture conditions. Each data point in this figure shows the spatially pooled, rectified, average (N=4) first-stage filter response to an unmodulated texture. Three types of textures are shown: an intact texture with a density of 595 micropatterns (purple diamonds), an intact texture with a density of 2975 (green circles), and a globally scrambled texture (blue triangles). The power-law exponent (k) is varied along the horizontal axis. All textures have approximately the same energy when the power law exponent is a square law because they are presented at a fixed RMS contrast. However, when the power law exponent is more expansive or compressive than 2, differences emerge between these textures that can enable segmentation.

5 Discussion

Here I present a summary of the findings of the work collected in this thesis, highlighting the strengths and weaknesses of the approach used, along with potential avenues of future study. The results of the thesis are discussed in the context of higher-order statistics, edge and region based processing, the effect of working above contrast threshold on measuring nonlinearities in vision, and work on nonlinear processing in general.

CHAPTER 5
Discussion

These studies form a comprehensive examination of the role of higher-order texture statistics in human segmentation of second-order boundaries, and make strong inferences about the mechanisms that underlie performance. In Chapter 2, natural and phase scrambled natural textures were used to demonstrate that higher-order statistics impair segmentation of contrast boundaries. By correlating image statistical measures against the magnitude of the impairment it was determined that sparseness and local phase alignments were important factors to examine. In Chapter 3 naturalistic synthetic textures were used to assess individual image statistics experimentally. It was shown that texture sparseness and global phase structure impair orientation and contrast boundary segmentation, but that local phase structure has little impact on second-order boundary segmentation. An FRF model with two free parameters (exponent of power-law nonlinearity and decision noise) was fit to these data, and it was determined that a compressive rectifying nonlinearity is required to fit the segmentation results. In Chapter 4 it was demonstrated that global phase structure and density can enable segmentation in the absence of any lower-order cues. The same model with the same nonlinear power-law exponent employed for contrast and orientation boundary segmentation data in Chapter 3 could also account for the roles of global structure and density, though not the effects of local phase alignments in these data.

As stated at the outset of this thesis, the goal of a great deal of the research in vision is to understand how different stages of visual processing

respond to natural scenes. Much of the past work using carefully parameterized and controlled stimuli has done a very good job of characterizing the aspects of the visual system that we know to look for, so Touryan and Dan (2001) suggest that using natural stimuli will enable us to uncover further properties of processing that are free of some of these assumptions (Rumsfeld (2011)'s famous 'unknown unknowns'). Natural images, however, are challenging stimuli to work with because they are difficult to manipulate systematically, and have spatial properties we have no established way of quantifying. Because of these challenges, Rust and Movshon (2005) argue that the complexity of natural images leaves them useful only as a benchmark for model performance. In this thesis, I have used a combination of natural and synthetic stimuli to circumvent their respective limitations. To uncover potential phenomenology, I used natural textures and a correlational analysis to pinpoint specific texture statistics for further analysis. These phenomena formed the basis of new hypotheses that I then tested explicitly using synthetic stimuli designed to address the texture statistics of interest. I developed and fit a model that accounts for much of the variance in human performance on synthetic texture segmentation. The final step in this cycle, and a future direction for this work, is to test this model against results with natural stimuli.

5.1 Higher-order texture statistics

The most important higher-order statistic found for segmentation, identified using the method outlined above, is texture sparseness. There is a widespread misconception that sparseness refers to the spatial frequency content of an image (e.g., DeValois & DeValois, 1990), in particular that sparser images contain more low spatial frequency information. This is not necessarily the case, and the work in this thesis follows Boulton and Baker

(1993) and Kingdom et al. (2001) in completely dissociating the structure of an image (i.e. its density) from its amplitude spectrum. This is not the first work to identify the importance of density as a basic image property to which humans are sensitive (Durgin, 1995; Nothdurft, 1985; Boulton & Baker, 1993; Nothdurft, 2000; Wilkinson & Wilson, 1998), or the first to note that it is a dimension on which natural images differ from random noise patterns (Hansen & Hess, 2007), but it serves to suggest that there is a clear relationship between texture density and segmentation, and that this relationship affects segmentation of natural textures. It shows that future studies of segmentation should account for or control texture density, and ensure that it is not inadvertently varied along with a manipulated parameter, or present as a confounding factor when comparing results between experiments. This could be a particular problem when using natural texture photographs, but Dakin et al. (2011) suggest a promising method for computing the density of any arbitrary image.

Sparseness is only one of many higher-order statistics that could be important to segmentation. The statistics examined here are by no means an exhaustive exploration of higher-order image statistics, and identifying others is an important goal for future studies. Portilla and Simoncelli (2000) found a set of image statistics that were sufficient to explain natural texture appearance, but the relationship between appearance judgements and segmentation mechanisms is unclear. Instead, it could be fruitful to pursue other structural image properties such as collinearity (Field, Hayes, & Hess, 1993), regularity (Morgan, Mareschal, Chubb, & Solomon, 2012), contrast polarity (Malik & Perona, 1990), or co-circularity (Motoyoshi & Kingdom, 2010) to determine their influence on as well as their ability to enable segmentation.

For example, contrast polarity might be of particular interest. The locally scrambled micropatterns used in this thesis are created by randomizing the phase of each of their Fourier components. When the lowest spatial frequency is randomized, some micropatterns end up mostly light or mostly dark, so that the textures composed of locally scrambled micropatterns have a greater polarity variance than those with the micropatterns intact. In retrospect, it would have been more unambiguous to keep the lowest spatial frequency component centered in the gaussian window, and scramble only the higher spatial frequency components to dissociate polarity variance and local phase alignment.

Second-order mechanisms are often probed with regular stimuli, such as modulated sine wave gratings or checkerboards, that are far simpler than what is typical of statistically-defined textures. These stimuli are entirely regular, but in fact regularity exists on a continuum (Morgan et al., 2012) and might interact with density in interesting ways in both segmentation and in other texture-related tasks.

As Figure 1-1 shows, in natural images segmentation cues do not typically occur in isolation. Most boundaries in the natural world are defined by changes in more than one image property. Restricting the statistics enabling segmentation was one of the simplifications that made this research tractable. Future experiments should examine how boundary cues are combined, particularly cues of different statistical order such as luminance and density, or orientation and local structure.

5.2 Studying segmentation

In early work on texture segmentation (Julesz, 1962), the criterion for segmentation was a subjective judgement of whether or not the region defined by the boundary seemed to be segmented preattentively, or (in

later papers) 'popped out'. This term was borrowed from the visual search literature to describe the percept of salient objects that were apparent in an array of distractor objects, without any effortful searching required to localize them. However, using a subjective measure such as salience may be misleading, especially in cases when the boundary salience is qualitatively higher than in the 'no segmentation' cases, yet lower than in the pop-out cases. This grey area was dealt with by measuring segmentation thresholds, to obtain a quantitative estimate of how easy a boundary was to segment. To prevent observers from searching for a boundary, while avoiding subjective measurements, we used a very brief stimulus presentation time (100ms) along with a fixation point to limit eye movements. Bergen and Julesz (1983) attempted to show that salience in visual search is related to segmentation because single texture elements that popped out of a distractor array would typically also form regions that segmented from the same distractors when presented as a textured patch. This is a tempting relationship to draw, but is difficult to support or refute.

Models of boundary segmentation usually proceed in one of two ways: edge-based processing or region-based processing. Region-based processing, which was implicitly assumed throughout this thesis, weights the entirety of each texture region more-or-less equally in segmentation (in fact, the precise weighting is accounted for in the shape of the second-stage filter). Edge-based processing, on the other hand, uses local discontinuities at the boundary. Wolfson and Landy (1998) attempted to examine how information along edges or regions affected segmentation thresholds by using half-disc stimuli where the halves were either abutting or separated. This is somewhat misguided, because in the abutting condition they are measuring texture segmentation, but in the separated condition, they measure texture

discrimination. As in the famous example with Ts and Ls, and emphasized in the introduction to this thesis, these tasks are fundamentally different and easily dissociated (Beck, 1966). We instead addressed the question of edge- versus region-based processing with a pilot experiment using 'bubbles' windows (Gosselin & Schyns, 2001), and found that observers relied on the entire stimulus image equally. Of course, there are circumstances under which region-based segmentation fails–in particular, illusory contours from abutting gratings–though it has been shown that the same filter-rectify-filter mechanism (Wilson & Richards, 1992) and the same single neurons (Song & Baker, 2007) can segment both these types of boundaries.

5.3 Role of nonlinearity

Decades of early work in the psychophysics of visual perception was performed with stimuli presented near their contrast threshold. This was for a number of reasons, one of which is pertinent here: as contrast is increased the effect of nonlinearities often increases, so using low-contrast stimuli helps maintain linearity of responses and enable quantitative modelling. In this thesis, the primary effect of structure is attributed to the shape of the intermediate nonlinearity in the two-stage model of texture segmentation. Recall, however, that in Chapter 2, Experiment 3 (Figure 2–7), when carrier contrast (rather than modulation depth) thresholds were measured, no effect of structure was observed. This highlights the importance of understanding the nonlinear processing that underlies the representation of structure, and suggests that when reviewing the evidence for the effect of structure on visual perception it is important to consider the contrast of the stimuli employed.

This work demonstrates that the FRF-style of model can segment boundaries based on properties other than those specified by the selectivity

of the first-stage filters (i.e., orientation and spatial frequency) without additional nonlinear stages. This is an alternative to the idea that additional stages of rectification and filtering would be required (Graham et al., 1993; Kingdom et al., 2001). This finding also serves to ameliorate some of the tension between the popular energy-based view of segmentation, and demonstrations that segmentation can occur without differences in the amplitude spectrum (e.g., Malik & Perona, 1990; Graham et al., 1993). Still, it is not a complete model, because some influences of local texture properties in our results are not accounted for.

An intermediate nonlinearity is required for second order segmentation to occur at all. In any form tested, an intermediate nonlinearity both enabled segmentation and accounted for the overall qualitative finding that sparseness impairs contrast and orientation boundary segmentation. This is because the model's representation of the signal (the second-order boundary) is embedded in a noisier background when the texture is sparse. The background is noisy because sparse images are characterized by local contrast modulations. This is evident in Figure 3–11A, which illustrates the greater response of the second stage filter to sparse, unmodulated textures. The shape of the nonlinearity was important to account for the lack of effect of local structure, and the degree to which sparseness affects segmentation.

When the boundary is defined by higher-order statistics, such as density, the interpretation of the model's responses is somewhat more complex. As illustrated in Figure 4–9, because the textures that define the boundary have been matched for RMS contrast, a square law nonlinearity will not enable segmentation between boundaries defined by density (Figure 4–9, power-law exponent = 2). However, any other power law nonlinearity will enable segmentation between these conditions. Relative to a square law, a

more expansive nonlinearity boosts the areas of high local contrast, and so sparser textures are stronger stimuli. On the other hand, a more compressive nonlinearity boosts the areas of low contrast and suppresses the high contrast areas, so a denser texture presents a stronger stimulus. In the case of structure modulations, the observed effect of increasing density leading to higher thresholds is almost certainly due to how this decreases the energy difference between the halves of the stimulus. The modest effect of local structure in these cases is not replicated by the model, and is thus difficult to interpret.

It is possible, and even logical, to instantiate both compressive and expansive nonlinearities in the same model. The best model for our results employed an intermediate (compressive) nonlinearity between the first- and second-stage filters, as well as an (expansive) output nonlinearity following the second stage of filtering and pooling. The sigmoidal function that is commonly used to represent the contrast response function of a neuron is expansive at low contrasts and compressive at high contrasts. It is plausible that the first-stage responses are strong enough to be affected by gain control mechanisms, and thus compressive, while the second-stage responses are affected by the lower, expansive, portion of the contrast response function.

An important aspect of sensitivity to higher-order image statistics is some sort of 'conjunction-detection', i.e. an enhanced (supralinear) response to combinations of signals. Methods of conjunction detection have been proposed on the feature level, explicitly using products of early filter responses (Martin et al., 2004; Freeman & Simoncelli, 2011). While in Freeman and Simoncelli (2011) this clustering occurs locally, Martin et al. (2004) use a k-means procedure, which performs clustering on the over

the entire image. Peirce (2007) suggests that an intermediate compressive nonlinearity can serve to create conjunction selectivity. Mathematically, we would expect a compressive nonlinearity and pooling followed by an expansive nonlinearity to perform something approximating a product, as a compressive nonlinearity can approximate a logarithm and an expansive nonlinearity an exponential function (following the laws of logarithms: $\log xy = \log x + \log y$).

5.4 Improving the model

A large part of this work included modelling psychophysical thresholds by simulating our experiment. We fit a model with only two parameters to our data quite well– though a model architecture with more parameters might be desirable, fitting additional parameters presented a challenge. Running a simulation was very computationally intensive for a variety of reasons, including generating random stimuli for each trial, and stochastic stimuli require a large number of trials.

Every micropattern texture was unique, with a new one generated randomly for each trial as in the psychophysical experiments. The reason for this was to avoid biasing our results with a relatively small subset of textures. However it meant that every trial the model ran required generating at least one texture in the case of contrast modulations, and at least two textures when quilting was required to impose the modulation. In addition to texture generation, each trial required the carrier or carriers to go through the enveloping procedure and– particularly time consuming– be filtered by our bank of log-gabors.

A consequence of using stochastic stimuli was that the model gave different results on repeated runs of a given simulation, depending on the stimuli that had been generated. This required running as many trials as

were run for human observers in order to average out the case-by-case variations that could and did impact the best-fitting model parameters. Note that even if the textures were pre-generated to alleviate the computational strain of generating stimuli on the fly, a large number of trials would still be needed to ensure sufficiently general model results. In future studies of this kind, to cut down on the time it takes to run each simulation and enable us to fit more model parameters or examine different model architectures, it may be desirable to pre-compute a stimulus library and associated first stage filter responses.

The work in this thesis uses what is commonly termed a *discrimination paradigm* (Graham, 1989) (not to be confused with texture discrimination), because the observer must discriminate between two non-blank stimulus conditions, such as vertical versus horizontal boundary orientation. This is a common psychophysical paradigm in studies of second-order processing, but is not frequently addressed by models. In contrast, many psychophysical studies of early visual processing employ a *detection* paradigm, where the observer must identify whether or not a target stimulus was present.

The effects of testing observers' ability to determine if the boundary was left- or right-oblique, rather than simply present are evident in the modelling: for the latter, the model thresholds can be anticipated based on the strength of the response of the second stage filters. For the former, how to predict the thresholds in any way other than a complete simulation is unclear. When the model must determine which of two possible stimuli was present, its threshold is not proportional to the strength of the second-stage filter response, nor to the difference between the strengths of the left- and right- oblique responses. This is well illustrated in Figure 3–11B. All of the filter responses depicted result in the same slightly-above-threshold

performance, but this could not be predicted from either the absolute filter response, or the difference between the left- and right-oblique filter responses. For this reason, it is necessary to perform a simulation of the experiment and evaluate the model based on segmentation thresholds rather than use some proxy such as filter responses. This is the first time, to our knowledge, that segmentation has been modeled using complete simulations of individual trials, and these results indicate that care should be taken when using metrics (e.g. difference in filter responses) to model segmentation performance, depending on the task.

In the case of orientation and contrast boundary segmentation, the model accounted for the roles of density and global phase structure, as well as the lack of a role for local phase alignments. When applied to structure and density boundaries, the model continued to account well for the roles of density and global structure, however it did not predict the role of local phase structure at higher texture densities that we see in human psychophysical data. This suggests that the model is missing some component that is selective for local structure. Adapting the model so that the differences in local phase structure are properly predicted could possibly be accomplished by implementing local interactions across spatial frequency channels in the bank of first-stage filters. It is also possible that the differences could be predicted by accounting for the difference in polarity variance.

The model is somewhat unrealistic, particularly in the uniform weightings between carrier-selective channels. However, the model performs extremely well in this simplified form, so a more complex set of experiments would be required to fit any additional parameters.

While developing our model, we investigated a range of architectural possibilities. Because the conditions under which a model does not work can be informative, we present them here:

The model required some decision noise. Without decision noise, model performance spanned a wide enough range to obtain thresholds in all conditions however in the absence of decision noise the model gives thresholds lower than the human observers, particularly in the globally scrambled condition. The model also gave poor fits when all spatial frequency channels were weighted equally: when the weights were equal, the model produced a difference in thresholds between the intact and locally scrambled conditions in the contrast and orientation segmentation experiments. However, after applying a $1/f$ weighting to the channels, the model matched human performance by generating equal thresholds for the intact and locally scrambled conditions. This weighting is what we expect for the efficient coding of natural images (Field, 1987).

The model also required zero-balanced first stage filters. We found that models using conventional gabor filters (i.e., not zero-balanced) produced a difference between the intact and locally scrambled segmentation thresholds in the contrast and orientation boundary segmentation tasks. This difference was restricted to the even-phase gabor channel, indicating that zero-balancing was indeed the problem. Zero-balancing the first stage filters by using log-gabor filters is arguably the most appropriate way to model early cortical responses to broadband images (Field, 1987).

Surround suppression between first-stage filters had undesirable side effects. Surround suppression is a pervasive property of visual processing (Fitzpatrick, 2000), and one that is appealing for density encoding, because dense textures will induce much more surround suppression than sparse

textures. We implemented a surround suppression algorithm based on Grigorescu, Petkov, and Westenberg (2004), and the extra-receptive field properties documented in H. Tanaka and Ohzawa (2009), but found the resulting models' performance very unlike those of our human participants. Most problematically, surround suppression enabled intact/locally scrambled (INT/LS) segmentation with thresholds as low as those for other structure boundaries, while human observers could not perform this task well enough to obtain thresholds.

To simplify the model, I chose to only represent two second-stage units with filters (left- and right-oblique, at a low spatial frequency) in a winner-takes-all arrangement. Realistically, there should be a population of filters at this stage, at a range of spatial frequencies and orientations. This change could be implemented and evaluated based on the human data presented here because the cosine-tapered boundary used in the envelope construction is relatively broadband. Instead of a simple decision rule, a more biologically plausible network approach could be employed, as in Deneve et al. (1999).

References

Adelson, E. (2001). On seeing stuff: The perception of materials by humans and machines. In *Proceedings of the spie* (Vol. 4299, pp. 1–12).

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

Albrecht, D., & Geisler, W. (1991). Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Visual neuroscience*, *7*(6), 531–546.

Albrecht, D., & Hamilton, D. (1982). Striate cortex of monkey and cat: contrast response function. *Journal of Neurophysiology*, *48*(1), 217–237.

Allard, R., & Faubert, J. (2007). Double dissociation between first-and second-order processing. *Vision Research*, *47*(9), 1129–1141.

Aloimonos, J. (1988). Shape from texture. *Biological Cybernetics*, *58*(5), 345–360.

Alonso, J., Usrey, W., & Reid, R. (1996). Precisely correlated firing in cells of the lateral geniculate nucleus. *Nature*, *383*(6603), 815–819.

Alonso, J., Usrey, W., & Reid, R. (2001). Rules of connectivity between geniculate cells and simple cells in cat primary visual cortex. *The Journal of Neuroscience*, *21*(11), 4002–4015.

Alvarez, G., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*(4), 392–398.

Andriessen, J., & Bouma, H. (1976). Eccentric vision: Adverse interactions between line segments. *Vision Research*, *16*(1), 71–78.

Arbelaez, P., Maire, M., Fowlkes, C., & Malik, J. (2011). Contour detection and hierarchical image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*(99), 1–1.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162.

Arsenault, A., Wilkinson, F., & Kingdom, F. (1999). Modulation frequency and orientation tuning of second-order texture mechanisms. *JOSA A*, *16*(3), 427–435.

Arsenault, E., Yoonessi, A., & Baker Jr, C. (2011). Higher order texture statistics impair contrast boundary segmentation. *Journal of Vision*, *11*(10).

Atick, J., & Redlich, A. (1990). Towards a theory of early visual processing. *Neural Computation*, *2*(3), 308–320.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*(3), 183.

Attwell, D., & Laughlin, S. (2001). An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow & Metabolism*, *21*(10), 1133–1145.

Baccus, S., & Meister, M. (2002). Fast and slow contrast adaptation in retinal circuitry. *Neuron*, *36*(5), 909–919.

Baker, C. (1999). Central neural mechanisms for detecting second-order motion. *Current Opinion in Neurobiology*, *9*(4), 461–466.

Balas, B. (2006). Texture synthesis and perception: Using computational models to study texture representations in the human visual system. *Vision Research*, *46*(3), 299–309.

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, *9*(12).

Barlow, H. (1953). Summation and inhibition in the frog's retina. *The Journal of Physiology*, *119*(1), 69–88.

Barlow, H. (1961). Possible principles underlying the transformations of sensory messages.

Barlow, H. (2001). Redundancy reduction revisited. *Network: Computation in Neural Systems*, *12*(3), 241–253.

Beck, J. (1966). Effect of orientation and of shape similarity on perceptual grouping. *Attention, Perception, & Psychophysics*, *1*(5), 300–302.

Beck, J. (1983). Textural segmentation, second-order statistics, and textural elements. *Biological Cybernetics*, *48*(2), 125–130.

Beck, J., Sutter, A., & Ivry, R. (1987). Spatial frequency channels and perceptual grouping in texture segregation. *Computer Vision, Graphics, and Image Processing*, *37*(2), 299–325.

Bell, A., & Sejnowski, T. (1997). The 'independent components' of natural scenes are edge filters. *Vision Research*, *37*(23), 3327.

Benardete, E., & Kaplan, E. (1999). The dynamics of primate m retinal ganglion cells. *Visual Neuroscience*, *16*(2), 355–368.

Bendat, J., & Piersol, A. (1966). *Measurement and analysis of random data*. Wiley New York.

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

Bergen, J., & Adelson, E. (1991). Theories of visual texture perception. *Spatial Vision*, *10*, 114–134.

Bergen, J., & Julesz, B. (1983). Parallel versus serial processing in rapid

pattern discrimination. *Nature*, *303*(5919), 696–698.

Bergen, J., & Landy, M. (1991). Computational modeling of visual texture segregation. *Computational Models of Visual Processing*, 253–271.

Bex, P. (2010). (in) sensitivity to spatial distortion in natural scenes. *Journal of Vision*, *10*(2).

Bex, P., & Makous, W. (2002). Spatial frequency, phase, and the contrast of natural images. *Journal of the Optical Society of America A*, *19*(6), 1096–1106.

Bex, P., Mareschal, I., & Dakin, S. (2007). Contrast gain control in natural scenes. *Journal of Vision*, *7*(11).

Bex, P., Solomon, S., & Dakin, S. (2009). Contrast sensitivity in natural scenes depends on edge as well as spatial frequency structure. *Journal of Vision*, *9*(10).

Blake, A., Bülthoff, H., & Sheinberg, D. (1993). Shape from texture: Ideal observers and human psychophysics. *Vision Research*, *33*(12), 1723–1737.

Blakemore, C., & Campbell, F. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *The Journal of Physiology*, *203*(1), 237–260.

Blakemore, C., Nachmias, J., & Sutton, P. (1970). The perceived spatial frequency shift: evidence for frequency-selective neurones in the human brain. *The Journal of Physiology*, *210*(3), 727–750.

Blakemore, C., & Sutton, P. (1969). Size adaptation: A new aftereffect. *Science*, *166*(3902), 245–247.

Blakemore, C., & Tobin, E. (1972). Lateral inhibition between orientation detectors in the cat's visual cortex. *Experimental Brain Research*, *15*(4), 439–440.

Bonds, A. (1989). Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. *Visual Neuroscience*, *2*(01), 41–55.

Bonin, V., Mante, V., & Carandini, M. (2003). Origins of size tuning in LGN neurons. *Journal of Vision*, *3*(9), 19–19.

Bonin, V., Mante, V., & Carandini, M. (2005). The suppressive field of neurons in lateral geniculate nucleus. *The Journal of Neuroscience*, *25*(47), 10844–10856.

Bonin, V., Mante, V., & Carandini, M. (2006). The statistical computation underlying contrast gain control. *The Journal of Neuroscience*, *26*(23), 6346–6353.

Boulton, J., & Baker, C. (1993). Different parameters control motion perception above and below a critical density. *Vision Research*, *33*(13), 1803–1811.

Bownds, M., & Arshavsky, V. (1995). What are the mechanisms of photoreceptor adaptation? *Behavioral and Brain Sciences*, *18*(3), 415–424.

Bradley, A., Switkes, E., & De Valois, K. (1988). Orientation and spatial frequency selectivity of adaptation to color and luminance gratings. *Vision Research*, *28*(7), 841–856.

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

Burr, D., & Morrone, M. (1987). Inhibitory interactions in the human vision system revealed in pattern-evoked potentials. *The Journal of Physiology*, *389*(1), 1–21.

Burton, G. (1973). Evidence for non-linear response processes in the human visual system from measurements on the thresholds of spatial beat

frequencies. *Vision Research*, *13*(7), 1211–1225.

Busse, L., Wade, A., & Carandini, M. (2009). Representation of concurrent stimuli by population activity in visual cortex. *Neuron*, *64*(6), 931.

Cadieu, C., & Olshausen, B. (2012). Learning intermediate-level representations of form and motion from natural movies. *Neural computation*, *24*(4), 827–866.

Caelli, T. (1980). Facilitative and inhibitory factors in visual texture discrimination. *Biological Cybernetics*, *39*(1), 21–26.

Caelli, T., & Julesz, B. (1978). On perceptual analyzers underlying visual texture discrimination: Part i. *Biological Cybernetics*, *28*(3), 167–175.

Caelli, T., & Julesz, B. (1979). Psychophysical evidence for global feature processing in visual texture discrimination. *Journal of the Optical Society of America A*, *69*(5), 675–678.

Caelli, T., & Moraglia, G. (1985). On the detection of gabor signals and discrimination of gabor textures. *Vision Research*, *25*(5), 671–684.

Cai, D., Deangelis, G., & Freeman, R. (1997). Spatiotemporal receptive field organization in the lateral geniculate nucleus of cats and kittens. *Journal of Neurophysiology*, *78*(2), 1045–1061.

Campbell, F., Cooper, G., & Enroth-Cugell, C. (1969). The spatial selectivity of the visual cells of the cat. *The Journal of Physiology*, *203*(1), 223.

Campbell, F., & Green, D. (1965). Optical and retinal factors affecting visual resolution. *The Journal of Physiology*, *181*(3), 576.

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

Carandini, M. (2004). Receptive fields and suppressive fields in the early visual system. *The Cognitive Neurosciences III, Ed*, *3*, 313–326.

Carandini, M. (2006). What simple and complex cells compute. *The Journal of Physiology*, *577*(2), 463–466.

Carandini, M. (2007). Melting the iceberg: contrast invariance in visual cortex. *Neuron*, *54*(1), 11–13.

Carandini, M., & Heeger, D. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*(1).

Carandini, M., Heeger, D., & Anthony Movshon, J. (1999). Linearity and gain control in V1 simple cells. *Cerebral Cortex*, 401–443.

Carandini, M., Heeger, D., & Movshon, J. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *The Journal of Neuroscience*, *17*(21), 8621–8644.

Cavanaugh, J., Bair, W., & Movshon, J. (2002a). Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *Journal of Neurophysiology*, *88*(5), 2530–2546.

Cavanaugh, J., Bair, W., & Movshon, J. (2002b). Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons. *Journal of Neurophysiology*, *88*(5), 2547–2556.

Chandler, D., & Field, D. (2007). Estimates of the information content and dimensionality of natural scenes from proximity distributions. *Journal of the Optical Society of America A*, *24*(4), 922–941.

Chapman, B., Zahs, K., & Stryker, M. (1991). Relation of cortical cell orientation selectivity to alignment of receptive fields of the geniculocortical afferents that arborize within a single orientation column in ferret visual cortex. *The Journal of Neuroscience*, *11*(5), 1347–1358.

Chong, S., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404.

Chubb, C., & Sperling, G. (1988). Drift-balanced random stimuli- a general basis for studying non-fourier motion perception. *Journal of the Optical Society of America A*, *5*, 1986–2007.

Chubb, C., & Yellott, J. (2000). Every discrete, finite image is uniquely determined by its dipole histogram. *Vision Research*, *40*(5), 485–492.

Chung, S., & Ferster, D. (1998). Strength and orientation tuning of the thalamic input to simple cells revealed by electrically evoked cortical suppression. *Neuron*, *20*(6), 1177–1190.

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

Dakin, S., & Mareschal, I. (2000). Sensitivity to contrast modulation depends on carrier spatial frequency and orientation. *Vision Research*, *40*(3), 311–329.

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

Dan, Y., Atick, J., & Reid, R. (1996). Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *The Journal of Neuroscience*, *16*(10), 3351–3362.

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

David, S., Vinje, W., & Gallant, J. (2004). Natural stimulus statistics alter the receptive field structure of V1 neurons. *The Journal of Neuroscience*, *24*(31), 6991–7006.

Dawis, S., Shapley, R., Kaplan, E., & Tranchina, D. (1984). The receptive

field organization of X-cells in the cat: spatiotemporal coupling and asymmetry. *Vision Research*, *24*(6), 549–564.

Dean, A., & Tolhurst, D. (1986). Factors influencing the temporal phase of response to bar and grating stimuli for simple cells in the cat striate cortex. *Experimental Brain Research*, *62*(1), 143–151.

DeAngelis, G., Freeman, R., & Ohzawa, I. (1994). Length and width tuning of neurons in the cat's primary visual cortex. *Journal of Neurophysiology*, *71*(1), 347–374.

DeAngelis, G., Ohzawa, I., & Freeman, R. (1995). Receptive-field dynamics in the central visual pathways. *Trends in Neurosciences*, *18*(10), 451–458.

DeAngelis, G., Robson, J., Ohzawa, I., & Freeman, R. (1992). Organization of suppression in receptive fields of neurons in cat visual cortex. *Journal of Neurophysiology*, *68*(1), 144–163.

De Monasterio, F., & Gouras, P. (1975). Functional properties of ganglion cells of the rhesus monkey retina. *The Journal of Physiology*, *251*(1), 167–195.

Deneve, S., Latham, P., & Pouget, A. (1999). Reading population codes: a neural implementation of ideal observers. *Nature Neuroscience*, *2*(8), 740.

Derrington, A., & Badcock, D. (1985). Separate detectors for simple and complex grating patterns? *Vision Research*, *25*(12), 1869–1878.

Derrington, A., & Badcock, D. (1986). Detection of spatial beats: non-linearity or contrast increment detection? *Vision Research*, *26*(2), 343–348.

De Valois, K., & Tootell, R. (1983). Spatial-frequency-specific inhibition in cat striate cortex cells. *The Journal of Physiology*, *336*(1), 359–376.

De Valois, R., Albrecht, D., & Thorell, L. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, *22*(5), 545–559.

DeValois, R., & DeValois, K. (1990). *Spatial vision* (No. 14). Oxford University Press, USA.

Duffy, C., & Wurtz, R. (1991). Sensitivity of MST neurons to optic flow stimuli. i. a continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology*, *65*(6), 1329–1345.

Durgin, F. (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. (2008). Texture density adaptation and visual number revisited. *Current Biology*, *18*(18), R855–R856.

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

Ellemberg, D., Allen, H., & Hess, R. (2004). Investigating local network interactions underlying first-and second-order processing. *Vision Research*, *44*(15), 1787–1798.

Ellemberg, D., Hansen, B., & Johnson, A. (2007). Discrimination of amplitude spectrum slope of natural scenes during childhood. *Journal of Vision*, *7*(9), 962–962.

Ellemberg, D., Hansen, B., & Johnson, A. (2012). The developing visual system is not optimally sensitive to the spatial statistics of natural images. *Vision Research*.

El-Shamayleh, Y., Movshon, J., & Kiorpes, L. (2010). Development of sensitivity to visual texture modulation in macaque monkeys. *Journal of Vision*, *10*(11).

Emrith, K., Chantler, M., Green, P., Maloney, L., & Clarke, A. (2010). Measuring perceived differences in surface texture due to changes in higher order statistics. *Journal of the Optical Society of America A*, *27*(5), 1232–1244.

Enroth-Cugell, C., & Robson, J. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *The Journal of Physiology*, *187*(3), 517–552.

Field, D. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, *4*(12), 2379-2394.

Field, D., Hayes, A., & Hess, R. (1993). Contour integration by the human visual system: Evidence for a local 'association field'. *Vision Research*, *33*(2), 173–193.

Finn, I., Priebe, N., & Ferster, D. (2007). The emergence of contrast-invariant orientation tuning in simple cells of cat visual cortex. *Neuron*, *54*(1), 137.

Fitzpatrick, D. (2000). Seeing beyond the receptive field in primary visual cortex. *Current Opinion in Neurobiology*, *10*(4), 438–443.

Flom, M., Weymouth, F., & Kahneman, D. (1963). Visual resolution and contour interaction. *Journal of the Optical Society of America A*, *53*(9), 1026–1032.

Fogel, I., & Sagi, D. (1989). Gabor filters as texture discriminator. *Biological Cybernetics*, *61*(2), 103–113.

Foley, J. (1994). Human luminance pattern-vision mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America A*, *11*(6), 1710–1719.

Foster, K., Gaska, J., Nagler, M., & Pollen, D. (1985). Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of

the macaque monkey. *The Journal of Physiology*, *365*(1), 331–363.

Freeman, J., & Simoncelli, E. (2011). Metamers of the ventral stream. *Nature Neuroscience*, *14*(9), 1195–1201.

Gabor, D. (1946). Theory of communication. part 1: The analysis of information. *Electrical Engineers-Part III: Radio and Communication Engineering, Journal of the Institution of*, *93*(26), 429–441.

Georgeson, M., & Harris, M. (1984). Spatial selectivity of contrast adaptation: models and data. *Vision Research*, *24*(7), 729–741.

Gibson, J. (1950). The perception of the visual world.

Gibson, J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing* (pp. 67–82). Lawrence Erlbaum Associates.

Glezer, V., Tsherbach, T., Gauselman, V., & Bondarko, V. (1980). Linear and non-linear properties of simple and complex receptive fields in area 17 of the cat visual cortex. *Biological Cybernetics*, *37*(4), 195–208.

Glezer, V., Tsherbach, T., Gauselman, V., & Bondarko, V. (1982). Spatio-temporal organization of receptive fields of the cat striate cortex. *Biological Cybernetics*, *43*(1), 35–49.

Gosselin, F., & Schyns, P. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271.

Graham, N. (1989). *Visual pattern analyzers* (No. 16). Oxford University Press, USA.

Graham, N., & Nachmias, J. (1971). Detection of grating patterns containing two spatial frequencies: A comparison of single-channel and multiple-channels models. *Vision Research*, *11*(3), 251–IN4.

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.

Grigorescu, C., Petkov, N., & Westenberg, M. (2004). Contour and boundary detection improved by surround suppression of texture edges. *Image and Vision Computing*, *22*(8), 609–622.

Grinvald, A., Lieke, E., Frostig, R., Gilbert, C., & Wiesel, T. (1986). Functional architecture of cortex revealed by optical imaging of intrinsic signals. *Nature*, *324*.

Hammond, P., & Andrews, D. (1978). Orientation tuning of cells in areas 17 and 18 of the cat's visual cortex. *Experimental Brain Research*, *31*(3), 341–351.

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

Hawken, M., & Parker, A. (1987). Spatial properties of neurons in the monkey striate cortex. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *231*(1263), 251–288.

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

Heeger, D. (1992b). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, *9*(02), 181–197.

Heeger, D., & Bergen, J. (1995). Pyramid-based texture analysis/synthesis. In *Proceedings of the 22nd annual conference on computer graphics and interactive techniques* (pp. 229–238).

Henning, B., Hertz, G., & Broadbent, D. (1975). Some experiments bearing

on the hypothesis that the visual system analyses spatial patterns in independent bands of spatial frequency. *Vision Research*, *15*(8), 887–897.

Hirsch, J. (2003). Synaptic physiology and receptive field structure in the early visual pathway of the cat. *Cerebral Cortex*, *13*(1), 63–69.

Hochstein, S., & Shapley, R. (1976). Linear and nonlinear spatial subunits in y cat retinal ganglion cells. *The Journal of Physiology*, *262*(2), 265–284.

Hopkins, H. (1962). The application of frequency response techniques in optics. *Proceedings of the Physical Society*, *79*(5), 889.

Hubel, D., & Wiesel, T. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, *148*(3), 574–591.

Hubel, D., & Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*(1), 106.

Hubel, D., & Wiesel, T. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*(1), 215–243.

Hubel, D., Wiesel, T., & Stryker, M. (1978). Anatomical demonstration of orientation columns in macaque monkey. *The Journal of comparative neurology*, *177*(3), 361–379.

Humphrey, A., Sur, M., Uhlrich, D., & Sherman, S. (1985). Termination patterns of individual x-and y-cell axons in the visual cortex of the cat: Projections to area 18, to the 17/18 border region, and to both areas 17 and 18. *The Journal of Comparative Neurology*, *233*(2), 190–212.

Hutchinson, C., & Ledgeway, T. (2004). Spatial frequency selective masking of first-order and second-order motion in the absence of off-frequency

'looking'. *Vision research*, *44*(13), 1499–1510.

Ikeda, H., & Wright, M. (1975a). Retinotopic distribution, visual latency and orientation tuning of 'sustained' and 'transient' cortical neurones in area 17 of the cat. *Experimental Brain Research*, *22*(4), 385–398.

Ikeda, H., & Wright, M. (1975b). Spatial and temporal properties of 'sustained' and 'transient' neurones in area 17 of the cat's visual cortex. *Experimental Brain Research*, *22*(4), 363–383.

Jain, A., & Farrokhnia, F. (1991). Unsupervised texture segmentation using gabor filters. *Pattern Recognition*, *24*(12), 1167–1186.

Johnson, A., & Baker Jr., C. (2004). First-and second-order information in natural images: a filter-based approach to image statistics. *Journal of the Optical Society of America A*, *21*(6), 913–925.

Jones, J., & Palmer, L. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, *58*(6), 1187–1211.

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

Julesz, B. (1980). Spatial nonlinearities in the instantaneous perception of textures with identical power spectra. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *290*(1038), 83–94.

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

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

Julesz, B. (1984). A brief outline of the texton theory of human vision. *Trends in Neurosciences*, *7*(2), 41–45.

Julesz, B. (1986). Texton gradients: The texton theory revisited. *Biological Cybernetics*, *54*(4), 245–251.

Julesz, B., & Bergen, J. (1983). Textons, the fundamental elements in preattentive vision and perception of textures. *Bell System Technical Journal*.

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

Julesz, B., Gilbert, E., & Victor, J. (1978). Visual discrimination of textures with identical third-order statistics. *Biological Cybernetics*, *31*(3), 137–140.

Kingdom, F., Hayes, A., & Field, D. (2001). Sensitivity to contrast histogram differences in synthetic wavelet-textures. *Vision Research*, *41*(5), 585–598.

Kingdom, F., Keeble, D., & Moulden, B. (1995). Sensitivity to orientation modulation in micropattern-based textures. *Vision Research*, *35*(1), 79–91.

Klein, D. (2005). *Beyond significance testing: Reforming data analysis methods in behavioral research* (Vol. 162) (No. 3). Am Psychiatric Assoc.

Klein, S., & Tyler, C. (1986). Phase discrimination of compound gratings: generalized autocorrelation analysis. *Journal of the Optical Society of America A*, *3*(6), 868–879.

Knill, D. (1998). Discrimination of planar surface slant from texture: human and ideal observers compared. *Vision Research*, *38*, 1683–1712.

Kovesi, P. (2000). Matlab and octave functions for computer vision and image processing. Retrieved November 2011, from

www.csse.uwa.edu.au/∼pk/Research/MatlabFns/

Kuffler, S. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, *16*(1), 37–68.

Landy, M., & Bergen, J. (1991). Texture segregation and orientation gradient. *Vision Research*, *31*(4), 679–691.

Landy, M., & Graham, N. (2004). Visual perception of texture. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (Vol. 2, pp. 1106–1118). MIT.

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

Langley, K., Fleet, D., & Hibbard, P. (1996). Linear filtering precedes nonlinear processing in early vision. *Current Biology*, *6*(7), 891–896.

Larsson, J., Landy, M., & Heeger, D. (2006). Orientation-selective adaptation to first-and second-order patterns in human visual cortex. *Journal of Neurophysiology*, *95*(2), 862–881.

Lazebnik, S., Schmid, C., & Ponce, J. (2006). Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Computer vision and pattern recognition, 2006 ieee computer society conference on* (Vol. 2, pp. 2169–2178).

Ledgeway, T., Zhan, C., Johnson, A., Song, Y., & Baker, C. (2005). The direction-selective contrast response of area 18 neurons is different for first-and second-order motion. *Visual Neuroscience*, *22*(01), 87–99.

Levi, D. (2008). Crowding-an essential bottleneck for object recognition: A mini-review. *Vision Research*, *48*(5), 635.

Levy, W., & Baxter, R. (1996). Energy efficient neural codes. *Neural Computation*, *8*(3), 531–543.

Li, A., & Zaidi, Q. (2000). Perception of three-dimensional shape from

texture is based on patterns of oriented energy. *Vision Research*, *40*(2), 217–242.

Maffei, L., & Fiorentini, A. (1973). The visual cortex as a spatial frequency analyser. *Vision Research*, *13*(7), 1255–1267.

Malik, J., & Perona, P. (1989). A computational model of texture segmentation. In *Computer vision and pattern recognition, 1989. proceedings cvpr'89., ieee computer society conference on* (pp. 326–332).

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

Mante, V., Bonin, V., & Carandini, M. (2008). Functional mechanisms shaping lateral geniculate responses to artificial and natural stimuli. *Neuron*, *58*(4), 625–638.

Marčelja, S. (1980). Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America A*, *70*(11), 1297–1300.

Mareschal, I., & Baker, C. (1999). Cortical processing of second-order motion. *Visual Neuroscience*, *16*(03), 527–540.

Marr, D. (1982). Vision. *W. H. Freeman and Company*.

Martin, D., Fowlkes, C., & Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color, and texture cues. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *26*(5), 530–549.

Meese, T., & Holmes, D. (2007). Spatial and temporal dependencies of cross-orientation suppression in human vision. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1606), 127–136.

Meese, T., & Summers, R. (2007). Area summation in human vision at

and above detection threshold. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1627), 2891–2900.

Meso, A., & Hess, R. (2011). Orientation gradient detection exhibits variable coupling between first-and second-stage filtering mechanisms. *Journal of the Optical Society of America A*, *28*(8), 1721–1731.

Mineault, P., Khawaja, F., Butts, D., & Pack, C. (2012). Hierarchical processing of complex motion along the primate dorsal visual pathway. *Proceedings of the National Academy of Sciences*.

Mishkin, M., Ungerleider, L., & Macko, K. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, *6*, 414–417.

Morgan, M., Chubb, C., & Solomon, J. (2008). A 'dipper' function for texture discrimination based on orientation variance. *Journal of Vision*, *8*(11).

Morgan, M., Mareschal, I., Chubb, C., & Solomon, J. (2012). Perceived pattern regularity computed as a summary statistic: implications for camouflage. *Proceedings of the Royal Society B: Biological Sciences*.

Morgan, M., Mason, A., & Baldassi, S. (2000). Are there separate first-order and second-order mechanisms for orientation discrimination? *Vision Research*, *40*(13), 1751–1763.

Morgan, M., Ross, J., & Hayes, A. (1991). The relative importance of local phase and local amplitude in patchwise image reconstruction. *Biological Cybernetics*, *65*(2), 113–119.

Morrone, M., Burr, D., & Maffei, L. (1982). Functional implications of cross-orientation inhibition of cortical visual cells. i. neurophysiological evidence. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *216*(1204), 335–354.

Motoyoshi, I., & Kingdom, F. (2007). Differential roles of contrast polarity

reveal two streams of second-order visual processing. *Vision Research*, *47*(15), 2047–2054.

Motoyoshi, I., & Kingdom, F. (2010). The role of co-circularity of local elements in texture perception. *Journal of Vision*, *10*(1).

Motoyoshi, I., & Nishida, S. (2001). Temporal resolution of orientation-based texture segregation. *Vision Research*, *41*(16), 2089–2105.

Motoyoshi, I., & Nishida, S. (2004). Cross-orientation summation in texture segregation. *Vision Research*, *44*(22), 2567–2576.

Motoyoshi, I., Nishida, S., & Adelson, E. (2005). Image statistics as a determinant of reflectance perception. *Journal of Vision*, *5*(8), 569–569.

Movshon, J., Thompson, I., & Tolhurst, D. (1978a). Receptive field organization of complex cells in the cat's striate cortex. *The Journal of Pphysiology*, *283*(1), 79–99.

Movshon, J., Thompson, I., & Tolhurst, D. (1978b). Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *The Journal of Physiology*, *283*(1), 53–77.

Mussap, A. (2001). Orientation integration in detection and discrimination of contrast-modulated patterns. *Vision Research*, *41*(3), 295–311.

Nishimoto, S., & Gallant, J. (2011). A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. *The Journal of Neuroscience*, *31*(41), 14551–14564.

Northdurft, H. (1997). Different approaches to the coding of visual segmentation. In M. Jenkin & L. Harris (Eds.), *Computational and psychophysical mechanisms of visual coding.* Cambridge University Press.

Nothdurft, H. (1985). Sensitivity for structure gradient in texture discrimination tasks. *Vision Research*, *25*(12), 1957–1968.

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

Nothdurft, H. (2000). Salience from feature contrast: variations with texture density. *Vision Research*, *40*(23), 3181-3200.

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.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*(12), 520–527.

Olshausen, B., & Field, D. (1996). Natural image statistics and efficient coding. *Network: Computation in Neural Systems*, *7*(2), 333–339.

Olshausen, B., & Field, D. (2004). Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, *14*(4), 481–487.

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

Olzak, L. (1985). Interactions between spatially tuned mechanisms: converging evidence. *Journal of the Optical Society of America A*, *2*(9), 1551–1559.

Olzak, L., & Thomas, J. (1991). When orthogonal orientations are not processed independently. *Vision Research*, *31*(1), 51–57.

Oppenheim, A., & Lim, J. (1981). The importance of phase in signals. *Proceedings of the IEEE*, *69*(5), 529–541.

Pantle, A., & Sekuler, R. (1968). Size-detecting mechanisms in human vision. *Science*, *162*(3858), 1146–1148.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., Morgan, M., et al.

(2001). Compulsory averaging of crowded orientation signals in human vision. *Nature neuroscience*, *4*(7), 739–744.

Peirce, J. (2007). The potential importance of saturating and supersaturating contrast response functions in visual cortex. *Journal of Vision*, *7*(6).

Peirce, J. (2011). Nonlinear summation really can be used to perform AND operations: Reply to May and Zhaoping. *Journal of Vision*, *11*(9).

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

Petrov, Y., Carandini, M., & McKee, S. (2005). Two distinct mechanisms of suppression in human vision. *The Journal of Neuroscience*, *25*(38), 8704–8707.

Phillips, F., & Todd, J. (2010). Texture discrimination based on global feature alignments. *Journal of Vision*, *10*(6).

Phillips, G., & Wilson, H. (1984). Orientation bandwidths of spatial mechanisms measured by masking. *Journal of the Optical Society of America A*, *1*(2), 226–232.

Piotrowski, L., & Campbell, F. (1982). A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase. *Perception*, *11*(13), 337–346.

Portilla, J., & Simoncelli, E. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, *40*(1), 49–70.

Priebe, N., & Ferster, D. (2006). Mechanisms underlying cross-orientation suppression in cat visual cortex. *Nature Neuroscience*, *9*(4), 552–561.

Purpura, K., Tranchina, D., Kaplan, E., & Shapley, R. (1990). Light adaptation in the primate retina: analysis of changes in gain and

dynamics of monkey retinal ganglion cells. *Visual Neurosciences*, *4*(1), 75–93.

Reid, R., & Alonso, J. (1995). Specificity of monosynaptic connections from thalamus to visual cortex. *Nature*, *378*(6554), 281–283.

Richards, W., & Polit, A. (1974). Texture matching. *Biological Cybernetics*, *16*(3), 155–162.

Rodieck, R. (1965). Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, *5*(12), 583–601.

Rose, D., & Blakemore, C. (1974). Effects of bicuculline on functions of inhibition in visual cortex. *Nature*, *249*, 375–377.

Rosenholtz, R. (2011). What your visual system sees where you are not looking. In *Proceedings of spie: Human vision and electronic imaging* (pp. 7865–7910).

Ross, J., & Speed, H. (1991). Contrast adaptation and contrast masking in human vision. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *246*(1315), 61–70.

Ross, J., Speed, H., & Morgan, M. (1993). The effects of adaptation and masking on incremental thresholds for contrast. *Vision Research*, *33*(15), 2051–2056.

Rubenstein, B., & Sagi, D. (1996). Preattentive texture segmentation: the role of line terminations, size, and filter wavelength. *Attention, Perception, & Psychophysics*, *58*(4), 489–509.

Rubner, Y., & Tomasi, C. (1998). Texture metrics. In *Systems, man, and cybernetics, 1998. 1998 ieee international conference on* (Vol. 5, pp. 4601–4607).

Ruderman, D. (1997). Origins of scaling in natural images. *Vision Research*, *37*(23), 3385–3398.

Rumsfeld, D. (2011). *Known and unknown: a memoir*. Penguin.

Rushton, W. (1965). Bleached rhodopson and visual adaptation. *The Journal of Physiology*, *181*(3), 645.

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

Rust, N., & Movshon, J. (2005). In praise of artifice. *Nature Neuroscience*, *8*(12), 1647–1650.

Sceniak, M., Hawken, M., & Shapley, R. (2001). Visual spatial characterization of macaque V1 neurons. *Journal of Neurophysiology*, *85*(5), 1873–1887.

Schofield, A., & Georgeson, M. (1999). Sensitivity to modulations of luminance and contrast in visual white noise: Separate mechanisms with similar behaviour. *Vision Research*, *39*(16), 2697–2716.

Schofield, A., & Georgeson, M. (2003). Sensitivity to contrast modulation: The spatial frequency dependence of second-order vision. *Vision Research*, *43*(3), 243–259.

Schofield, A., Rock, P., Sun, P., Jiang, X., & Georgeson, M. (2010). What is second-order vision for? discriminating illumination versus material changes. *Journal of Vision*, *10*(9).

Sclar, G., & Freeman, R. (1982). Orientation selectivity in the cat's striate cortex is invariant with stimulus contrast. *Experimental Brain Research*, *46*(3), 457–461.

Shapley, R., & Victor, J. (1978). The effect of contrast on the transfer properties of cat retinal ganglion cells. *The Journal of Physiology*, *285*(1), 275–298.

Shapley, R., & Victor, J. (1981). How the contrast gain control modifies

the frequency responses of cat retinal ganglion cells. *The Journal of Physiology*, *318*(1), 161–179.

Simoncelli, E., & Heeger, D. (1998). A model of neuronal responses in visual area MT. *Vision research*, *38*(5), 743–761.

Simoncelli, E., & Olshausen, B. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, *24*(1), 1193–1216.

Smith, A., & Scott-Samuel, N. (2001). First-order and second-order signals combine to improve perceptual accuracy. *Journal of the Optical Society of America A*, *18*(9), 2267–2272.

Solomon, J., & Sperling, G. (1994). Full-wave and half-wave rectification in second-order motion perception. *Vision Research*, *34*(17), 2239–2257.

Solomon, S., White, A., & Martin, P. (2002). Extraclassical receptive field properties of parvocellular, magnocellular, and koniocellular cells in the primate lateral geniculate nucleus. *The Journal of Neuroscience*, *22*(1), 338–349.

Song, Y., & Baker, C. (2006). Neural mechanisms mediating responses to abutting gratings: Luminance edges vs. illusory contours. *Visual Neuroscience*, *23*(02), 181–199.

Song, Y., & Baker, C. (2007). Neuronal response to texture and contrast-defined boundaries in early visual cortex. *Visual Neuroscience*, *24*(01), 65–77.

Sutter, A., Sperling, G., & Chubb, C. (1995). Measuring the spatial frequency selectivity of second-order texture mechanisms. *Vision Research*, *35*(7), 915–924.

Tadmor, Y., & Tolhurst, D. (1993). Both the phase and the amplitude spectrum may determine the appearance of natural images. *Vision Research*, *33*(1), 141–145.

Talbot, S., & Marshall, W. (1941). Physiological studies on neural mechanisms of visual localization and discrimination. *American Journal of Ophthalmology*, *24*(11), 1255–1264.

Talebi, V., & Baker, C. (2012). Natural versus synthetic stimuli for estimating receptive field models: A comparison of predictive robustness. *The Journal of Neuroscience*, *32*(5), 1560–1576.

Tanaka, H., & Ohzawa, I. (2009). Surround suppression of v1 neurons mediates orientation-based representation of high-order visual features. *Journal of Neurophysiology*, *101*(3), 1444–1462.

Tanaka, K., Hikosaka, K., Saito, H., Yukie, M., Fukada, Y., & Iwai, E. (1986). Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey. *The Journal of Neuroscience*, *6*(1), 134–144.

Theunissen, F., David, S., Singh, N., Hsu, A., Vinje, W., & Gallant, J. (2001). Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network: Computation in Neural Systems*, *12*(3), 289–316.

Thomson, M. (1999). Visual coding and the phase structure of natural scenes. *Network: Computation in Neural Systems*, *10*(2), 123–132.

Thomson, M., & Foster, D. (1997). Role of second-and third-order statistics in the discriminability of natural images. *Journal of the Optical Society of America A*, *14*(9), 2081–2090.

Touryan, J., & Dan, Y. (2001). Analysis of sensory coding with complex stimuli. *Current Opinion in Neurobiology*, *11*(4), 443–448.

Tsui, J., Hunter, J., Born, R., & Pack, C. (2010). The role of v1 surround suppression in mt motion integration. *Journal of Neurophysiology*, *103*(6), 3123–3138.

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

Ullman, S., & Schechtman, G. (1982). Adaptation and gain normalization. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *216*(1204), 299–313.

Victor, J. (1987). The dynamics of the cat retinal x cell centre. *The Journal of Physiology*, *386*(1), 219–246.

Victor, J. (1994). Images, statistics, and textures: implications of triple correlation uniqueness for texture statistics and the julesz conjecture: comment. *Journal of the Optical Society of America A*, *11*(5), 1680–1684.

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

Victor, J., & Conte, M. (1996). The role of high-order phase correlations in texture processing. *Vision Research*, *36*(11), 1615–1631.

Wainwright, M., Schwartz, O., & Simoncelli, E. (2002). *Natural image statistics and divisive normalization* (R. Rao, B. Olshausen, & M. Lewicki, Eds.). MIT Press.

Walker, G., Ohzawa, I., & Freeman, R. (1999). Asymmetric suppression outside the classical receptive field of the visual cortex. *The Journal of Neuroscience*, *19*(23), 10536–10553.

Walker Renninger, L., & Malik, J. (2004). When is scene identification just texture recognition? *Vision Research*, *44*(19), 2301–2311.

Wang, H., Landy, M., & Heeger, D. (2011). Psychophysical evidence for normalization in second-order mechanisms. *Journal of Vision*, *11*(11), 1173–1173.

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

Watt, R. (1995). Some speculations on the role of texture processing in visual perception. In T. Papathomas, C. Chubb, A. Gorea, & E. Kowler (Eds.), *Early vision and beyond* (p. 59). Massasschusetts Institute of Technology.

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., & Wilson, H. (1998). Measurement of the texture-coherence limit for bandpass arrays. *Perception*, *27*, 711–728.

Willmer, G., & Rutter, J. (1977). Orientation columns in macaque monkey visual cortex demonstrated by the 2-deoxyglucose autoradiographic technique. *Nature*, *269*, 329.

Wilson, H., & Richards, W. (1992). Curvature and separation discrimination at texture boundaries. *Journal of the Optical Society of America A*, *9*(10), 1653–1662.

Wolfson, S., & Graham, N. (2005). Element-arrangement textures in multiple objective tasks. *Spatial Vision*, *18*(2), 209.

Wolfson, S., & Landy, M. (1998). Examining edge-and region-based texture analysis mechanisms. *Vision Research*, *38*(3), 439–446.

Yellott Jr, J. (1993). Implications of triple correlation uniqueness for texture statistics and the julesz conjecture. *Journal of the Optical Society of America A*, *10*(5), 777–793.

Zhou, Y., & Baker, C. (1994). Envelope-responsive neurons in areas 17 and 18 of cat. *Journal of Neurophysiology*, *72*(5), 2134–2150.

Zhou, Y., & Baker Jr, C. (1996). Spatial properties of envelope-responsive cells in area 17 and 18 neurons of the cat. *Journal of Neurophysiology*, *75*(3), 1038–1050.

Zhu, S., Guo, C., Wang, Y., & Xu, Z. (2005). What are textons? *International Journal of Computer Vision*, *62*(1), 121–143.

Ziemba, C., Freeman, J., Movshon, J., & Simoncelli, E. (2012, February). Selectivity and invariance are greater in macaque v2 than v1. In *Computational systems neuroscience (CoSyNe).*