Edge coding in human vision: a psychophysical and computational investigation

Keith Anthony May

Doctor of Philosophy

Aston University

December 2003

Aston University

Edge coding in human vision: a psychophysical and computational investigation

Keith Anthony May

Doctor of Philosophy

2003

This thesis presents a study of how edges are detected and encoded by the human visual system. The study begins with theoretical work on the development of a model of edge processing, and includes psychophysical experiments on humans, and computer simulations of these experiments, using the model. The first chapter reviews the literature on edge processing in biological and machine vision, and introduces the mathematical foundations of this area of research. The second chapter gives a formal presentation of a model of edge perception that detects edges and characterizes their blur, contrast and orientation, using Gaussian derivative templates. This model has previously been shown to accurately predict human performance in blur matching tasks with several different types of edge profile. The model provides veridical estimates of the blur and contrast of edges that have a Gaussian integral profile. Since blur and contrast are independent parameters of Gaussian edges, the model predicts that varying one parameter should not affect perception of the other. Psychophysical experiments showed that this prediction is incorrect: reducing the contrast makes an edge look sharper; increasing the blur reduces the perceived contrast. Both of these effects can be explained by introducing a smoothed threshold to one of the processing stages of the model. It is shown that, with this modification, the model can predict the perceived contrast and blur of a number of edge profiles that differ markedly from the ideal Gaussian edge profiles on which the templates are based. With only a few exceptions, the results from all the experiments on blur and contrast perception can be explained reasonably well using one set of parameters for each subject. In the few cases where the model fails, possible extensions to the model are discussed.

Key words: blur, contrast, derivative, scale-space, template

To Juliet

Acknowledgements

I would like to thank Mark Georgeson for his excellent supervision. I would also like to thank Andrew Schofield and Ed Powell for invaluable technical assistance during the early stages of my PhD.

Contents

Chapter 1 Edge processing in biological and machine vision: a tutorial review 27

1.1 What is an edge? 27

1.1.1 Edge detection by differentiation 28

1.1.2 Edge detection by template matching 29

1.2 First-generation edge detectors 29

1.2.1 Differential operators 29

1.2.2 Template matching operators 32

1.2.3 Overview of enhancement / thresholding edge operators 36

1.3 Noise reduction by surface-fitting 38

1.4 Noise reduction using large operators 39

1.4.1 Early large operators 39

1.4.2 Operators for detecting rectangular step edges 41

1.4.3 Regularization operators 44

1.4.4 Operators for detecting blurred edges 46

1.4.4.1 The Laplacian of Gaussian operator 46

1.4.4.2 Blurred template operators 47

1.5 The need for multiple spatial channels 50

1.5.1 Scale-space theory 52

1.6 The link with physiology and psychophysics 54

1.7 Edge location performance of Gaussian derivative operators 58

1.7.1 Edges flanked by ramps 58

1.7.2 Ramp edges 59

1.7.3 Bars and staircases 59

1.7.4 Corners 61

1.7.5 Finite-sized objects 61

1.7.6 Curved contours 62

1.7.7 Trihedral vertices 62

1.7.8 Contour geometry 62

1.7.9 False edges 63

1.7.10 Edge Coherence 63

1.8 Edge detection algorithms 63
1.8.1 Marr & Hildreth (1980) 64
1.8.2 Canny (1986) 69
1.8.3 Algorithms that correct spatial distortions 71
1.8.3.1 Bergholm (1987): edge focusing 71
1.8.3.2 Ulupinar & Medioni (1988, 1990) 72
1.8.3.3 Chen & Medioni (1989) 73
1.8.4 Algorithms that characterize edges in scale space 74
1.8.4.1 Korn (1988) 74
1.8.4.2 Williams & Shah (1993) 75
1.8.4.3 Zhang & Bergholm (1997) 75
1.8.5 Local scale selection algorithms 76
1.8.5.1 Rosenfeld & Thurston (1971) and Rosenfeld *et al.* (1972) 77
1.8.5.2 Marr (1976) 77
1.8.5.3 Perona & Malik (1990) 78
1.8.5.4 Jeong & Kim (1992) 79
1.8.5.5 Morrone, Navangione & Burr (1995) 80
1.8.5.6 Fleck (1992a) 81
1.8.5.7 Elder & Zucker (1996, 1998) 81
1.8.5.8 Marimont & Rubner (1998) 82
1.8.5.9 Lindeberg (1996, 1998) 83
1.8.6 Filter combination algorithms 84
1.8.6.1 Rosenfeld (1970) 84
1.8.6.2 Watt & Morgan: MIRAGE 84
1.8.6.3 Georgeson & Meese: filter combination 89
1.8.7 Phase congruency algorithms 90
1.8.7.1 Energy models 90
1.8.7.2 Kovesi (2000) 93
1.8.7.3 Peli (2002) 94
1.9 Edge characterization 95
1.10 Edge detection and localization by humans 99
1.11 The usefulness of edges 105
1.11.1 The information content of edge maps 105

1.11.2 The appropriateness of edge maps 106
Chapter 2 The Gaussian derivative template model 109
2.1 Stimulus definitions 114
2.2 The GDT model in 1D 115
2.2.1 Orientation/blur channels 115
2.2.2 Operators 116
2.2.3 The sequence of operations 117
2.2.4 The virtual template 119
2.2.5 Normalizing the correlation 120
2.2.5.1 Deriving an expression for the normalization factor 124
2.2.6 Deriving an expression for the edge strength in response to Gaussian edges 125
2.2.7 Proof that the template model correctly estimates the position and scale of a Gaussian edge 126
2.2.8 Recovering edge contrast from peak height 127
2.3 The GDT model in 2D 128
2.3.1 Orientation/blur channels 128
2.3.2 Operators 128
2.3.3 Sequence of operations 129
2.3.4 Correlation and peak-finding 130
2.4 Relations to other literature 131
2.4.1 Summary of previous template models 132
2.4.2 Conclusions 135
2.5 Appendix: a deeper look at the edge strength measure in the 2D model 135
Chapter 3 Implementing the GDT model 140
3.1 Introduction 140
3.2 Basic parameters 140
3.3 Operators 142
3.3.1 Combining operators 142
3.3.2 Spatial extent of operators 143
3.4 Finding peaks in the edge strength 145
3.5 Spurious peaks 146
Chapter 4 General psychophysical methods 152
4.1 Subjects 152

4.2 The matching paradigm 152
4.2.1 Responses 152
4.2.2 Staircases 153
4.2.3 Data analysis 154
4.3 Apparatus 157
Chapter 5 Effect of contrast on perceived blur 160
5.1 Introduction 160
5.2 Experiment 1: effect of contrast on perceived blur 160
5.2.1 Introduction 160
5.2.1.1 Experiment 1.1: main experiment 161
5.2.1.2 Experiment 1.2: effect of noise surround 161
5.2.1.3 Experiment 1.3: effect of inter-stimulus interval duration 161
5.2.1.4 Experiment 1.4: larger window diameter and longer inter-stimulus interval duration 162
5.2.2 Stimuli 162
5.2.2.1 Test edges 163
5.2.2.2 Comparison edges 165
5.2.2.3 Noise surround in experiments 1.2 and 1.3 165
5.2.2.4 Spatial envelope diameter and ISI duration in experiment 1.4 165
5.2.3 Methods 166
5.2.3.1 Experiments 1.1 and 1.2 166
5.2.3.2 Experiment 1.3 166
5.2.3.3 Experiment 1.4 167
5.2.4 Results 167
5.2.4.1 Experiment 1.1 167
5.2.4.2 Experiment 1.2 170
5.2.4.3 Experiment 1.3 172
5.2.4.4 Experiment 1.4 173
5.2.5 Discussion of experiment 1 174
5.3 Experiment 2: effect of contrast on perceived blur with fixed window size ... 177
5.3.1 Introduction 177
5.3.2 Stimuli 177
5.3.3 Methods 178

5.3.4 Results and discussion 179
5.4 General discussion: explaining the effect of contrast of perceived blur 183
Chapter 6 Contrast threshold for edge detection 185
6.1 Introduction 185
6.2 Experiment 3: contrast threshold for edge detection 186
6.2.1 Introduction 186
6.2.2 Stimuli 189
6.2.3 Method 191
6.2.4 Results and discussion 192
6.2.4.1 Data analysis 192
6.2.4.2 Are edges of opposite polarity perfectly discriminated? ... 193
6.2.4.3 Variation of detection threshold with edge scale 198
Chapter 7 Modifying the rectifier 202
7.1 Threshold models in computer science, psychophysics and physiology 202
7.1.1 Computational threshold models 202
7.1.2 Psychophysical threshold models of contrast coding 203
7.1.3 Models of simple cells 204
7.1.3.1 Threshold-linear models 204
7.1.3.2 Power-law/normalization models 207
7.1.3.3 Reconciling threshold-linear and power-law models 209
7.2 Modifications to the half-wave rectifier 211
7.2.1 The power-function rectifier 212
7.2.1.1 Defining the function 212
7.2.1.2 Constructing the power-function rectifier from threshold-linear units 213
7.2.1.2.1 Case 1: $V < T_1$ 215
7.2.1.2.2 Case 2: $V > T_2$ 215
7.2.1.2.3 Case 3: $T_1 \leq V \leq T_2$ 216
7.2.1.3 Concluding remarks about the power-function rectifier ... 217
7.2.2 A simple threshold-linear model 218
7.2.2.1 Maximum threshold for detectability of a Gaussian edge. 219
7.2.3 A pure power-law transducer 220
7.2.4 A Legge-Foley transducer 221

7.2.5 Simulation 1: feature marking with a Legge-Foley transducer 222
7.2.5.1 Methods and stimuli 222
7.2.5.2 Results 223
7.2.6 Simulation 2: feature marking with the power-function rectifier 225
7.2.6.1 Methods and stimuli 225
7.2.6.2 Results 225
7.2.7 Summary of feature marking simulations 227
Chapter 8 General modelling methods 228
8.1 Introduction 228
8.2 Blur matching with a binary search method 228
8.3 Modified binary search termination criterion for contrast matching 229
8.4 Fitting the model to the data 231
8.5 1D and 2D simulations 232
Chapter 9 Simulating the effect of contrast on perceived blur 233
9.1 Simulation 3: simulating experiment 1.1 with a single threshold model 233
9.1.1 Methods and Stimuli 233
9.1.1.1 Simulation 3.1: effect of size of image matrix 234
9.1.1.2 Simulation 3.2: comparison of 1D and 2D models 236
9.1.2 Results and discussion of the main part of simulation 3 239
9.2 Simulation 4: varying the threshold with channel scale 240
9.2.1 Methods and stimuli 241
9.2.2 Results 242
9.3 Simulation 5: simulating experiment 1 with the power-function rectifier 243
9.3.1 Methods and stimuli 243
9.3.2 Results 244
9.4 Simulation 6: simulating experiment 1 with a Legge-Foley transducer 246
9.4.1 Methods and stimuli 246
9.4.2 Results 246
9.5 Simulation 7: simulating experiment 2 with a power-function rectifier 248
9.5.1 Methods and stimuli 248
9.5.1.1 Simulation 7.1: effect of size of image matrix 249
9.5.2 Results of the main part of simulation 7 250
9.6 Simulation 8: simulating experiment 2 with intrinsic blur 252
9.6.1 Methods and stimuli 252

9.6.1.1 Simulation 8.1: effect of size of image matrix 253
9.6.2 Results of the main part of simulation 8 253
9.7 Simulation 9: simulating experiment 2 with a Legge-Foley transducer 255
9.7.1 Methods and stimuli 255
9.7.2 Results 255
Chapter 10 Effect of edge scale on perceived contrast 257
10.1 Models of contrast coding 257
10.2 Experiment 4: effect of edge scale on perceived contrast 259
10.2.1 Methods and stimuli 259
10.2.2 Results 260
10.3 Simulation 10: simulating experiment 4 with a power-function rectifier 263
10.3.1 Methods and stimuli 263
10.3.2 Results 263
10.4 Simulation 11: simulating experiment 4 with a Legge-Foley transducer 265
10.4.1 Methods and stimuli 265
10.4.2 Results 265
10.5 General discussion of the perception of edge blur and contrast 267
Chapter 11 Effect of window width on perceived contrast 270
11.1 Introduction 270
11.2 Experiment 5: effect of edge scale and window width on perceived contrast 271
11.2.1 Methods and stimuli 271
11.2.2 Results 272
11.3 Simulation 12: simulating experiment 5 with a power-function rectifier 272
11.3.1 Methods and stimuli 272
11.3.1.1 Simulation 12.1: effect of image matrix size 274
11.3.1.2 Simulation 12.2: comparison of 1D and 2D models 274
11.3.2 Results of the main part of simulation 12 277
11.4 Simulation 13: simulating experiment 5 with a Legge-Foley transducer 278
11.4.1 Methods and stimuli 278
11.4.2 Results 278
11.5 General discussion of the effect of window width on perceived contrast 280
Chapter 12 Effect of linear ramp on edge perception 286
12.1 Introduction 286
12.2 Experiment 6: effect of linear ramp on perceived blur of Gaussian edges 290

12.2.1 Methods and stimuli 290
12.2.2 Results 291
12.3 Simulation 14: simulating experiment 6 with a power-function rectifier 294
12.3.1 Simulation 14.1: comparison of 1D and 2D models 294
12.3.2 Results of the main part of simulation 14 296
12.4 Simulation 15: simulating experiment 6 with a Legge-Foley transducer 298
12.4.1 Methods and stimuli 298
12.4.2 Results 298
12.5 Experiment 7: effect of linear ramp on perceived contrast of Gaussian edges 300
12.5.1 Methods and stimuli 300
12.5.2 Results 302
12.6 Simulation 16: simulating experiment 7 with a power-function rectifier 302
12.6.1 Methods and stimuli 302
12.6.1.1 Simulation 16.1: comparing 1D and 2D models. 303
12.6.2 Results of the main part of simulation 16 304
12.7 Simulation 17: simulating experiment 7 with a Legge-Foley transducer 305
12.7.1 Methods and stimuli 305
12.7.2 Results 305
12.8 General discussion: effect of linear ramps on edge perception 308
Chapter 13 Summary and conclusions 310
13.1 The GDT model 310
13.2 Experimental findings 312
13.3 Overall best-fitting parameters 313
13.3.1 Best-fitting parameters for the power-function rectifier 315
13.3.2 Best-fitting parameters for the Legge-Foley transducer 316
13.4 The model's performance 317
13.5 Future extensions to the GDT model 318
13.6 Conclusions 321

Table legends

Table 1.1. Compass gradient template kernels. 35

Table 3.1. The basic parameters used in the implementation of the GDT model. 141

Table 5.1. Scales, spatial window diameters and viewing distances of test stimuli in experiments 1.1 and 1.2. 164

Table 5.2. Variables manipulated within a session of experiments 1.1 and 1.2..... 166

Table 6.1. Compound propositions corresponding to the four possible response outcomes in experiment 3. 187

Table 6.2. Stimulus parameters for all conditions in experiment 3. 190

Table 6.3. Best-fitting parameters for predicting the contrast threshold in experiment 3, assuming that the contrast threshold corresponds to a gradient threshold of T after blurring with a Gaussian kernel with scale σ_b.. The two columns on the far right give the best-fitting T value when σ_b. is fixed at 1′ or 2′. The parameters in the "Mean" row are fitted to the geometric mean across the 3 subjects. 201

Table 7.1. Parameters of the Legge-Foley transducer for the six conditions of simulation 1. 223

Table 7.2. Parameters of the power-function rectifier for the four conditions in simulation 2. 225

Table 9.1. Width of the spatial envelope used in simulation 3, in arc min and pixels. 234

Table 9.2. Estimated blur and contrast of the 0.2 contrast test edges in simulation 3.1. All values rounded to 3 significant figures. The border of zero values made no

difference to the perceived blur, and made a difference of less than 1×10^{-8} to the perceived contrast of each stimulus..... 235

Table 9.3. The model and stimulus parameters used in simulation 3.2. See Table 3.1 for an explanation of the rightmost two columns..... 236

Table 9.4. Comparison between 1D and 2D models with zero threshold. The shaded areas give the models' estimated blur for each of the test edges. 237

Table 9.5. Comparison between 1D and 2D models with threshold of 0.05. The shaded areas give the models' estimated blur for each of the test edges. 237

Table 9.6. Widths of image matrices used in simulation 7, in arc min and pixels. All stimuli the experiment 2 had window widths of 320', but smaller stimuli were used in the simulation to reduce processing time. 249

Table 9.7. Results of simulation 7.1..... 249

Table 9.8. Width of stimuli used in simulation 8, in arc min and pixels..... 252

Table 9.9. Results of simulation 8.1, using image matrices with size 1280 pixels, or the sizes given in Table 9.8 (the results were the same to 3 significant figures). 253

Table 11.1. Comparison between 1D and 2D models with two different parameter settings. The shaded areas give the models' estimated contrast for each of the test edges. 275

Table 11.2. Comparison between 1D and 2D models with two different parameter settings. The shaded areas give the models' estimated blur for each of the test edges. 275

Table 12.1. Variables manipulated within a block of experiment 6. 291

Table 12.2. The model and stimulus parameters used in simulation 14.1..... 295

Table 12.3. Comparison between 1D and 2D models with unbiased half-wave rectifier. The shaded areas give the models' estimated blur for each of the test edges in simulation 14. 295

Table 12.4. Comparison between 1D and 2D models with a power-function rectifier with parameters $T_1 = 0.04$, $T_2 = 0.3$, $n = 2.38$. The shaded areas give the models' estimated blur for each of the test edges in simulation 14. 296

Table 12.5. Variables manipulated within a session of experiment 7. 302

Table 12.6. Comparison between 1D and 2D models with unbiased half-wave rectifier. The shaded areas give the models' estimated contrast for each of the test edges in simulation 16. 303

Table 12.7. Comparison between 1D and 2D models with a power-function rectifier with parameters $T_1 = 0.04$, $T_2 = 0.3$, $n = 2.38$. The shaded areas give the models' estimated contrast for each of the test edges in simulation 16. 303

Table 13.1. Power-function rectifier parameters for subject KAM. 315

Table 13.2. Power-function rectifier parameters for subject MAG. 315

Table 13.3. Power-function rectifier parameters for subject PAA. 315

Table 13.4. Legge-Foley transducer parameters for subject KAM. 316

Table 13.5. Legge-Foley transducer parameters for subject MAG. 316

Table 13.6. Legge-Foley transducer parameters for subject PAA. 316

Figure legends

Figure 2.1. Processing stages in an early version of the template model. The top row shows the processing of an edge with a positive gradient; the bottom row shows the processing of an edge with a negative gradient. The four columns on the left show the processing in the blur channel that matches the scale of the edge; in these graphs, the vertical axis represents response level. The right hand column shows scale-space representations that give the template response for all blur channels; here, the vertical axis represents scale, and the response level is represented by the grey level at each point. With the positive edge, a single peak occurs in the scale-space representation, indicating the position and scale of the edge. With the negative edge, two peaks occur either side of the main edge. These can be removed by placing a half-wave rectifier after the 1st derivative operation. The negative edge is then removed from the channel. The rectifier causes each channel to process only edges that have a gradient that is positive relative to the channel's orientation. 110

Figure 2.2. The processing stages in Kovasznay & Joseph's (1955) edge detection algorithm. (a) Image luminance; (b) 1st derivative; (c) half-wave rectification with bias; (d) 2nd derivative of parts of signal that pass through the rectifier; (e) trigger driven by the 2nd derivative; (f) pulse formed from the trigger. 111

Figure 2.3. The output from Kovasznay & Joseph's (1955) edge detection algorithm. The top row shows the processing of a geometric figure. The bottom row shows that, even back in 1955, the test stimulus of choice was still a woman in an alluring pose. Depending on how the apparatus was set up, the edge map could be black on white (centre) or white on black (right). 112

Figure 3.1. The bars show the central 9 pixels of a Gaussian 1st-derivative operator with scale equal to 1 pixel, the smallest operator scale used in the simulations. The height of the operator is scaled to give the gradient in units/deg when the image

resolution is 240 pixels/deg. The curve shows the corresponding continuous function..... 141

Figure 3.2. Gaussian 1st-derivative operators were constructed with scale (σ) ranging from 0.6 to 1.4 pixels in steps of 0.01 pixels. Each operator was 31 pixels wide. These operators were used to estimate the gradient of a straight line with gradient 1. The graph shows the estimated gradient minus 1..... 142

Figure 3.3. Gaussian 1st-derivative operators were constructed with scale (σ) ranging from 2 to 3 pixels in steps of 0.01 pixels. Taking the central pixel position to be 0, each operator extended spatially from -round($n\sigma$) pixels to round($n\sigma$) pixels, where $n = 4$, 5, or 6, and round(x) is x, rounded to the nearest integer. These operators were used to estimate the gradient of a straight line with gradient 1. The graphs show the error, i.e. the estimated gradient minus 1. As σ increases, the gradient is underestimated by a progressively larger amount, until the operator size steps up by 1 pixel on each side. At this point, there is a jump in the estimated gradient. This jump occurs whenever $n\sigma = m + 0.5$, for $m \in \mathbb{Z}$, because these values of $n\sigma$ mark the boundary between two different values of round($n\sigma$).... 144

Figure 3.4. The ten configurations of pixels that were counted as peaks. In each configuration, all the grey pixels have the same value, which is greater than that of any of the white pixels; the dark grey pixel indicates the pixel that was chosen as the location of the peak..... 146

Figure 3.5. (a) The stimulus, a vertical Gaussian edge with scale 16′, and contrast 1. The image spanned 1200 pixels left-to-right, which corresponded to 5 deg of visual angle. (b) The edge strength representation in the positive channel (the channel that detects edges that are dark on the left); the white disks mark all the edges found in this channel. (c) The same edge strength representation without the edges marked. The model was run with the parameter values given in Table 3.1. 147

Figure 3.6. (a) An enlargement of the bottom-left corner of Figure 3.5(c), with brightness re-scaled so that black represents -4.64×10^{-14} and white represents 4.64×10^{-14}. (b) The corresponding region of the rectified 1st derivative, with white representing 1.01×10^{-14}. In both images, blur channels 2, 4, 9, 13, 18, and 19 have a value of exactly zero at each spatial position. 148

Figure 5.1. (a) An example of the stimuli used in experiment 1.1. (b) An example of the stimuli used in experiment 1.2, in which the stimuli were surrounded by binary noise. These edges have an orientation of 0°. 163

Figure 5.2. Results of experiment 1.1. Each row of panels shows data for a particular test edge scale. Each of the three leftmost columns of panels shows data for one subject; filled symbols show data from trials in which the test edge preceded the comparison edge, and open symbols represent trials with the opposite order. The right hand column shows data for all three subjects, collapsed across stimulus order. Within each panel, the horizontal axis represents the test edge contrast, and the vertical axis represents the ratio of the comparison and test edge scales when they perceptually matched in blur. The dashed line represents the null hypothesis that the edges would match in blur when their scales were equal... 168

Figure 5.3. Data from the right hand column of Figure 5.2 (i.e. collapsed across stimulus order), re-plotted to show how test edge scale interacts with contrast in determining the σ-ratio. 169

Figure 5.4. Comparison between KAM's results from experiments 1.1 and 1.2. Circles represent data from experiment 1.1, with no noise; squares represent data from experiment 1.2, in which stimuli were surrounded by a border of binary noise. The symbol colours correspond to the colours used in Figure 5.2..... 171

Figure 5.5. Results of 1.3. Subject was KAM. Each panel shows data for a different test edge scale. Filled symbols show data from trials in which the test edge preceded the comparison edge, and open symbols represent trials with the opposite order. Test edge contrast was 0.05 in all conditions..... 172

Figure 5.6. Comparison of results of experiment 1.4 with the corresponding conditions of experiment 1.1. Large symbols represent the data from experiment 1.4, which used a 300′ window diameter, and small symbols represent data from experiment 1.1, which used a 160′ window diameter with the test edge scales used in experiment 1.4. The symbol colours correspond to the colours used in Figure 5.2 and Figure 5.4. The top two rows of panels show data from subject KAM; the bottom two rows show data from MAG. 173

Figure 5.7. Data from the right hand column of Figure 5.6 (i.e. collapsed across stimulus order), re-plotted to show how test edge scale interacts with contrast in determining the σ-ratio. 174

Figure 5.8. Georgeson's (1994) data, re-plotted in the form used in this thesis. Each row of panels shows data for a different test edge scale, and the columns show data from subjects MAG, TCAF, and the geometric mean across the two subjects. 176

Figure 5.9. Results of the main part of experiment 2. One data point could not be plotted on these axes, because it was too wayward. This was PAA's data for the test edge with scale 32′ and contrast 0.05, when the test edge came 2nd. The σ-ratio for this point was 2.96. 180

Figure 5.10. Results of the three versions of experiment 2 that KAM performed. The left column shows the results of the main experiment (taken from the left column of Figure 5.9). The middle column shows the results with a longer ISI duration (1200ms) and the right column shows the results with orthogonal test and comparison edges (ISI duration 500ms, as in the main experiment). 181

Figure 5.11. Results KAM's 3 versions of experiment 2, plotted so that the corresponding conditions in the three versions can be compared directly. All the data from Figure 5.10 appear in the two leftmost columns of this figure, represented by the same symbols. The right column shows the results of fitting the psychometric functions after collapsing across stimulus order. Also shown are KAM's data from the 8′, 16′ and 32′ conditions of experiment 1.1, which were identical to the corresponding conditions of experiment 2, apart from the window shape and/or

width. The data from experiment 1.1 are represented by circles, which are the same symbols used to represent these data in Figure 5.2. 182

Figure 5.12. An illustration of why half-wave rectification with a threshold makes low-contrast edges appear sharper. Each row shows the processing of an edge of different contrast. Each column shows a different processing stage. The threshold level is indicated by the dashed line in the 1st derivative stage. The threshold causes the 2nd derivative to be truncated. The truncation is more severe with the low-contrast edge, so the 2nd derivative is narrower, and fits a narrower template, and therefore looks sharper than the high-contrast edge. 184

Figure 6.1. An example of the control stimuli used in experiment 3. The central edge is completely occluded by an opaque strip of binary noise. The semicircular regions have uniform luminance $L_0(1 \pm C)$, , where C is the Michelson contrast. 190

Figure 6.2. Results of experiment 3. Black symbols represent detection thresholds, and red symbols represent identification thresholds. Circles represent Gaussian edges, and squares represent the control stimuli. The dashed lines pass through the square symbols, to facilitate comparisons between the control thresholds and thresholds for each of the Gaussian edges. 195

Figure 6.3. Data from trials in experiment 3 in which the target preceded the non-target. Symbol colour and shape as in Figure 6.2. 196

Figure 6.4. Data from trials in experiment 3 in which the non-target preceded the target. Symbol colour and shape as in Figure 6.2. 197

Figure 6.5. Circles represent detection thresholds for Gaussian edges from Figure 6.2, and the horizontal dashed lines represent the detection thresholds for the control stimulus. The solid lines represent the thresholds predicted from the best-fitting parameters in Table 6.3. Sloping dotted lines and dashed lines represent the best-fitting predictions when σ_b. is fixed at 1′ and 2′, respectively. 200

Figure 7.1. The power-function rectifier. The graph shows two sets of three functions. Within each set, the T_1 and T_2 parameters are the same, and the exponent, n, varies, as indicated by the line style. The set on the left has $T_1 = 0$ and $T_2 = 1$; the set on the right has $T_1 = 2$ and $T_2 = 3$. With exponent, $n = 1$, the function is a half-wave rectifier with threshold T_1. With $n > 1$, the function is a power function of order n between T_1 and T_2, and a straight line for values $V > T_2$. As n increases, the function approaches a half-wave rectifier with threshold T_2, but it always remains smooth..... 211

Figure 7.2. Results of simulation 1: feature marking with different Legge-Foley transducer parameters. The vertical axis represents the scale of the marked edges, and the horizontal axis represents their positions. The background of each axis shows the sine wave grating stimulus, scaled to fit the horizontal axis. White symbols represent positive (dark-light) edges, and dark symbols represent negative (light-dark) edges..... 224

Figure 7.3. Results of simulation 2: feature marking with different power-function rectifier parameters. The top row shows results using a compressive square-root function. The middle row shows a squaring function. The bottom row shows the power-function rectifier which is a straight line for inputs greater than 0.3. The left column shows models with the T_1 parameter set to 0. The right column shows models with the T_1 parameter set to 0.05. For an explanation of the figure, see the legend of Figure 7.2. 226

Figure 9.1. Results of the main part of simulation 3, which simulated experiment 1 with a fixed-threshold rectifier. Each row of panels shows data for a particular test edge scale. Each column of panels shows data for a particular subject. The circles show the subject's data, collapsed across stimulus order. The data for MAG and PAA are the same as those shown in the right hand column in Figure 5.2. The data for KAM are the geometric mean of KAM's data from experiments 1.1 and 1.2 (both shown in the right hand column of Figure 5.4). Each subject's data are plotted against the same set of coloured lines, which show the model's predictions with different threshold values, T..... 238

Figure 9.2. The thick grey line shows the gradient profiles of a 6′ edge on the left and 32′ edge on the right. Both edges have a contrast of 0.05. A threshold of 0.05 (solid black line) substantially reduces the width of the 32′ edge's gradient profile, but only makes a small difference to the width of the 6′ profile. A threshold of 0.25 (dashed black line) substantially reduces the width of the 6′ profile, but the gradient of the 32′ edge falls well below this threshold. 240

Figure 9.3. Results of simulation 4. The graph on the left shows the effect of contrast on perceived blur of a 16′ edge when the threshold, T, is inversely proportional to the virtual template scale, σ. The graph on the right shows the effect when the threshold is inversely proportional to σ^2..... 242

Figure 9.4. Output of the rectifier stage of the GDT model with a scaled threshold. The stimulus was the 16′ Gaussian edge from simulation 4, with a threshold equal to $0.08/\sigma$, where σ is the virtual template scale. 243

Figure 9.5. Results of simulation 5, which simulated experiment 1 with a power-function rectifier (PFR). The solid line in each graph shows the results of fitting the PFR to the data from experiment 1: for each subject, one set of parameters was fitted to the whole data set. The dashed line in each graph shows the results of fitting the PFR to each subject's data from all the experiments..... 245

Figure 9.6. Results of simulation 6, which simulated experiment 1 with a Legge-Foley transducer (LFT). The predictions of the best-fitting Legge-Foley transducer are shown in green. These are compared with the predictions of the power-function rectifier (PFR), which are shown in black, as in Figure 9.5. Solid lines show the best-fitting parameters for experiment 1, and dashed lines show the best-fitting parameters for all the experiments..... 247

Figure 9.7. Results of simulation 7, which simulated experiment 2 with a power-function rectifier (PFR). The solid line shows the predictions from the parameters that fit best to the data from experiment 2, and the dashed line shows the predictions from the parameters fitted to all the experiments. The circles indicate the

experimental data. In the case of KAM, these data are the geometric mean of all three versions of this experiment that KAM performed..... 251

Figure 9.8. Results of simulation 8, which simulated experiment 2 using a power-function rectifier and intrinsic blur of 1′. The results of simulation 8 are shown in red, and are compared with the results of simulation 7, which are plotted in black, as in Figure 9.7. Circles show the same data as Figure 9.7..... 254

Figure 9.9. Results of simulation 9, which simulated experiment 2 with a Legge-Foley transducer (LFT). The predictions of the best-fitting Legge-Foley transducer are shown in green; predictions of the power-function rectifier (PFR) are shown in black, as in Figure 9.7. Solid lines show the best-fitting parameters to experiment 2, and dashed lines show the best-fitting parameters to all the experiments. Circles show the same data as Figure 9.7. 256

Figure 10.1. Results of experiment 4. Each row of panels shows data for a particular test edge contrast. The three leftmost columns of panels show data for the individual subjects; filled symbols show data from trials in which the test edge preceded the comparison edge, and open symbols represent trials with the opposite order. The right hand column shows data for all three subjects, collapsed across stimulus order. Within each panel, the horizontal axis represents the test edge scale, and the vertical axis represents the ratio of the comparison and test edge contrast when they perceptually matched in contrast. The dashed line represents the null hypothesis that the edges would match in contrast when their contrasts were equal. 261

Figure 10.2. Results of experiment 4. The data are identical to those in Figure 10.1, but they are plotted differently, to be consistent with the graphs in Chapter 5 and Chapter 9. Each row of panels shows data for a particular test edge scale. Within each panel, the horizontal axis represents the test edge contrast, and the vertical axis represents the ratio of the comparison and test edge contrasts when they perceptually matched in contrast. 262

Figure 10.3. Results of simulation 10, which simulated experiment 4 with a power-function rectifier. 264

Figure 10.4. Results of simulation 11, which simulated experiment 4 with a Legge-Foley transducer. The black lines show the results predicted by the power-function rectifier in simulation 10 (Figure 10.3). 266

Figure 10.5. Isoscale contours. Each line gives the estimated contrast and scale for an edge of the given physical scale, as its contrast is varied from 0.05 to 0.4. The model was run with KAM's overall best-fitting parameters, $T_1 = 0.04$, $T_2 = 0.3$, $n = 2.38$. The data for the 2′ and 4′ test edges came from simulation 7, and the remaining data came from simulation 5. 269

Figure 10.6. Isocontrast contours. This figure is plotted using the same data as Figure 10.5, but each line shows the estimated contrast and scale for an edge of the given physical contrast as its scale is varied from 2′ to 32′. 269

Figure 11.1. Results of experiment 5. Each row of panels shows data for a particular test stimulus width. Otherwise, the data presentation follows the same pattern as Figure 10.1. The trace along the bottom of each panel shows the profile of the test stimulus in each condition. 273

Figure 11.2. Results of simulation 12, which simulated experiment 5 with a power-function rectifier. The black lines show the best-fitting models. The blue line shows the predictions of the pure form of the GDT model, described in Chapter 2. The red line shows the predicted results if the edges match in perceived contrast when they match in Michelson contrast. 276

Figure 11.3. Results of simulation 13, which simulated experiment 5 with a Legge-Foley transducer. The black lines show the results from simulation 12 (Figure 11.2). 279

Figure 12.1. Effect of adding a linear ramp to an edge: luminance and gradient profiles. 287

Figure 12.2. Predicted effects of adding a positive or negative ramp to a positive edge, using the GDT model with an unbiased half-wave rectifier. The positive ramp makes no difference to the 2nd derivative signature, so the edge should have the same perceived blur and contrast as the standard edge. The negative ramp causes the gradient profile to be truncated by the rectifier, and this leads to a narrower 2nd derivative signature, which fits a narrower template, making the edge look sharper. 288

Figure 12.3. Predicted effects of adding a positive or negative ramp to a positive edge, using the GDT model with a biased half-wave rectifier. Unlike in Figure 12.2, the 2nd derivative signature of the standard edge is truncated slightly because the rectifier has a threshold, making the edge look sharper. The positive ramp lifts the gradient signal above the threshold, so it is no longer subject to the sharpening effect, and so it looks more blurred. 289

Figure 12.4. Results of experiment 6, using stimuli with a binary noise surround. 292

Figure 12.5. Comparison of KAM's results with and without surrounding binary noise in experiment 6. 293

Figure 12.6. Results of simulation 14, which simulated experiment 6 with a power-function rectifier. 297

Figure 12.7. Results of simulation 15, which simulated experiment 6 with a legge-foley transducer. The black lines show the predicted results using the power-function rectifier in simulation 14 (Figure 12.6) 299

Figure 12.8. Results of experiment 7 301

Figure 12.9. Results of simulation 16, which simulated experiment 7 with a power-function rectifier. The blue lines show the predictions using an unbiased half-wave rectifier. The contrast ratios predicted by the Michelson contrast model were too high to plot on these axes. The ratios for the 0.25 and 0.50 ramp gradient

conditions were 0.399 and 0.599 for the 6′, 12′, and 24′ stimuli, and 0.466 and 0.732 for the 8′, 16′ and 32′ stimuli..... 306

Figure 12.10. Results of simulation 17, which simulated experiment 7 with a Legge-Foley transducer. 307

Chapter 1
Edge processing in biological and machine vision: a tutorial review

1.1 What is an edge?

There is little consensus on what edges are, or even how useful it is to detect them. This chapter describes the different ways that researchers have defined edges, and the different kinds of algorithms that have been proposed for detecting them, both in biological and machine vision systems. A discussion of the usefulness of detecting edges is postponed until section 1.11.

Many researchers have defined an edge as a change in light intensity. Detection of intensity changes is considered to be useful because these changes correspond to physical changes that may be important. Marr & Hildreth (Marr & Hildreth, 1980; Marr, 1982; Hildreth, 1983) drew a distinction between properties of the image and properties of the external world: "a change in intensity is the phenomenon that we will detect and describe in the image; edges are the physical changes that give rise to these intensity changes" (Hildreth, 1983, p. 2). Intensity changes may be caused by changes in surface illumination, geometry, or reflectance (Marr & Hildreth, 1980; Marr, 1982). These three types of physical change give rise to intensity changes that are, in general, locally indistinguishable (Elder & Zucker, 1998), so the same intensity change could arise from a number of different physical sources. Thus, Marr and Hildreth's distinction between "edges" and "intensity changes" is more than merely a trivial issue of terminology. Although it is useful to keep in mind the distinction between image properties and the

physical changes that give rise to them, in this chapter, the term "edge" will be used more loosely, to describe both of these things. The reason for this is that, with no agreement over which image properties should be detected, there is no term, such as "intensity change", that is in sufficiently general use.

1.1.1 Edge detection by differentiation

Marr and Hildreth's view that physical edges can be detected by detecting intensity changes in the image has been held by many researchers since the field of image processing first appeared. In the earliest approaches to edge detection by machine (e.g. Kovasznay & Joseph, 1953, 1955; Kirsch *et al.*, 1957), edges were considered to be sudden changes, i.e. points of high intensity gradient. This suggested that edges could be found by differentiating the image signal with respect to space, to give the intensity gradient at each point; edge locations would then correspond to the gradient peaks, or simply locations with high gradient. Very many edge detectors find edges by detecting peaks or high values in the 1st derivative.

Peaks in the 1st derivative give rise to zero-crossings (ZCs) in the 2nd derivative. Marr & Hildreth's (1980) algorithm detected these ZCs. A possible problem with this approach is that ZCs in the 2nd derivative also arise from *minima* in the magnitude of the 1st derivative, i.e. "points that are as non-edge-like as possible" (Clark, 1989, p. 44). These *phantom* or *false* edges can be eliminated by taking the second derivative in the direction of the gradient (i.e. the direction in which the 1st derivative *increases* the most sharply); then ZCs with negative slopes correspond precisely to the peaks in the gradient magnitude (Haralick, 1984). Equivalently, this approach has been expressed as marking ZCs at points where the product of 1st and 3rd derivatives is negative (Berzins, 1984; Ulupinar & Medioni, 1988, 1990; Clark, 1989). The latter formulation is most natural for 1D signals, where the derivatives are traditionally taken towards the right; in 2D images, the derivatives are usually taken in the gradient direction, in which case the 1st derivative is always positive, so it is only necessary to consider the sign of the 3rd derivative, as in Haralick's approach. It may be that, in modelling human vision, it is not necessary to remove phantom edges: Georgeson & Freeman (1997) reported that subjects perceived edges at gradient minima, as well as maxima.

1.1.2 Edge detection by template matching

Another approach is to have a set of edge templates, and mark edges where the image intensity profile provides a good fit to one of the templates. The edges can then be characterized by the parameters of the best-fitting template. Section 1.2.2 shows that, despite the superficial difference between differential and template algorithms, they can be formally equivalent.

1.2 First-generation edge detectors

The first generation of edge detectors are often referred to as *enhancement/thresholding* edge detectors. The general approach was to process the image in such a way that edges gave rise to high values in the processed image. Any values above a threshold were then taken to be edge points. Enhancement/thresholding edge detectors fall into two categories: differential operators and template matching operators.

1.2.1 Differential operators

Kovasznay & Joseph's (1955) edge detector (described in Chapter 2) used analogue electronics to perform the differentiation, but most subsequent systems have used digital computers, in which images are digitized into pixels. In digital implementations, the first derivative can be approximated by the *finite difference*, which is found by subtracting the values of adjacent pixels. This can be understood by considering the definition of the 1st derivative. The first derivative, dI/dx of a 1D image signal, I, at spatial position, x, is defined as follows:

$$\frac{dI}{dx} = \lim_{\Delta x \to 0} \frac{I(x+\Delta x) - I(x)}{\Delta x}. \qquad (1.1)$$

If we take Δx to be 1 pixel then, from (1.1), we have

$$\frac{dI}{dx} \simeq I(x+1) - I(x). \tag{1.2}$$

An alternative approximation is found by subtracting the values either size of x:

$$\frac{dI}{dx} \simeq \frac{I(x+1) - I(x-1)}{2}. \tag{1.3}$$

The 2 in the denominator of (1.3) can be dropped, since the correspondence between the luminance range and the range of grey-level pixel values is somewhat arbitrary, and so the absolute values of the gradient are unimportant.

These derivative approximations can be expressed as linear filtering operations[1]. The result, R, of filtering the image, I, with a *kernel*, or *mask*, f, is defined as follows:

$$R(x) = \sum_{i \in A} f(i-x) I(i), \tag{1.4}$$

where A is the domain of I. The response at position, x, is found by centring the filter kernel on that position in the image, multiplying the values of superimposed pixels in the image and kernel, and then adding these products together. The derivative approximations given by (1.2) and (1.3) can be achieved by filtering the image with the kernels given by $f = (-1 \quad 1)$ and $f = (-1 \quad 0 \quad 1)$, respectively[2].

So far, the analysis has been in one dimension. Since images are 2D, we need to consider partial derivatives along a particular direction. If we think of an edge as being a high luminance gradient along a particular direction, the best direction to choose at each point will be the direction along which the gradient is the largest. Suppose we define two

---

[1] A function, g, is a linear operator if and only if $g(ax_1 + bx_2) = ag(x_1) + bg(x_2)$.

[2] In order to index these kernels in a way compatible with equation (1.4), the central pixel (or left-of-centre pixel in the case of an even number of pixels) should have an index of 0, and the kernel is assumed to have a value of 0 for any index not explicitly defined.

orthogonal axes, x and y, arranged so that positive values on the x-axis are mapped onto positive values on the y-axis by an anticlockwise rotation of 90° about the origin. A standard result from vector calculus is that the partial derivative in a direction θ (measured anticlockwise from the x-axis) is given by

$$\frac{\partial I}{\partial x}\cos\theta + \frac{\partial I}{\partial y}\sin\theta,$$

where $\partial I/\partial x$ and $\partial I/\partial y$ are the partial derivatives in the x and y directions, respectively. From this result, it is straightforward to prove that the maximum gradient occurs along a direction, θ_{max}, given by

$$\theta_{max} = \arctan\left(\frac{\partial I}{\partial y}\Big/\frac{\partial I}{\partial x}\right). \tag{1.5}$$

The magnitude of this gradient, M_{max}, is given by

$$M_{max} = \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2}. \tag{1.6}$$

Equations (1.5) and (1.6) show that, in order to find the direction and magnitude of the steepest gradient at any point, it is only necessary to take partial derivatives in two orthogonal directions. Kovasznay & Joseph (1955) were probably the first to suggest using equation (1.6) in the context of edge detection. Since then, this approach has been used with many digital enhancement/thresholding gradient operators. The first was probably that described by Roberts (1965). He estimated the partial derivatives using the following two filter kernels:

$$f_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad f_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

The image can be filtered with these 2-dimensional kernels using a version of equation (1.4), modified to use two indices (or coordinates) for each point:

$$R(x,y) = \sum_{(i,j)\in A\times B} f\big((i,j)-(x,y)\big)I(i,j), \qquad (1.7)$$

where A and B are the sets of horizontal and vertical indices, and the Cartesian product, $A\times B$, is the set of all ordered pairs of image coordinates, i.e. the domain of the image, I. Filtering the image separately with kernels f_1 and f_2 finds the finite differences between adjacent pixels in the two diagonal directions, which approximate the partial derivatives in these directions.

More commonly, the partial derivatives have been estimated in the horizontal and vertical directions using 3×3 filter kernels:

$$f_1 = \begin{pmatrix} -1 & 0 & 1 \\ -w & 0 & w \\ -1 & 0 & 1 \end{pmatrix}, \qquad f_2 = \begin{pmatrix} 1 & w & 1 \\ 0 & 0 & 0 \\ -1 & -w & -1 \end{pmatrix}.$$

In these kernels, w usually takes a value of 1 (Prewitt, 1970), 2 (Sobel - see Duda & Hart, 1973, p. 271), or $\sqrt{2}$ (Frei & Chen, 1977). With all of these operators, the magnitude and direction of the luminance gradient at each point can be estimated using equations (1.5) and (1.6).

1.2.2 Template matching operators

The template matching approach uses a set of edge templates that represent the luminance profiles of ideal edges at a number of different orientations (usually eight). At each point, the best-fitting template is found, and the estimated orientation is set to the orientation of that template. The correlation between the image and the template at each point can be found by filtering the image with a kernel that has the shape of the template. This can be understood by considering the filter kernel, f, to be a vector in an N-dimensional vector space, where N is the number of pixels in the filter kernel. Let h be the region of the image covered by the filter kernel centred on a particular pixel. Then h

is also an N-dimensional vector. The result, R, of the filtering process for that pixel is given by

$$R = \sum_{i=1}^{N} f_i h_i \,. \tag{1.8}$$

R is called the *dot product* of f and h, written $f \cdot h$ (alternative terms are *scalar product* and *inner product*). A standard result is

$$f \cdot h = |f||h| \cos\theta \,, \tag{1.9}$$

where θ is the angle between these vectors. $|f|$ and $|h|$ are the magnitudes of these vectors, which, according to Pythagorus's theorem, are given by

$$|f| = \sqrt{\sum_{i=1}^{N} f_i^2} \,, \tag{1.10}$$

$$|h| = \sqrt{\sum_{i=1}^{N} h_i^2} \,. \tag{1.11}$$

Thus, from (1.8) to (1.11) we can deduce

$$\cos\theta = \frac{\sum_{i=1}^{N} f_i h_i}{\sqrt{\left(\sum_{i=1}^{N} f_i^2\right)\left(\sum_{i=1}^{N} h_i^2\right)}} \,. \tag{1.12}$$

The term $\cos\theta$ in (1.12) is known as the *normalized correlation coefficient*. It can be used as a measure of the similarity between the vectors f and h. When the vectors are identical, except for a scaling factor (i.e. perfectly correlated), their orientations will be identical, so $\theta = 0°$ and $\cos\theta = 1$; when the vectors are orthogonal, i.e. completely uncorrelated, $\theta = 90°$, giving $\cos\theta = 0$; when the vectors are exact opposites of each other, $\theta = 180°$, giving $\cos\theta = -1$. The normalized correlation coefficient is very closely related to the Pearson product-moment correlation coefficient, r, given by

$$r = \frac{N\sum_{i=1}^{N} f_i h_i - \sum_{i=1}^{N} f_i \sum_{i=1}^{N} h_i}{\sqrt{\left[N\sum_{i=1}^{N} f_i^2 - \left(\sum_{i=1}^{N} f_i\right)^2\right]\left[N\sum_{i=1}^{N} h_i^2 - \left(\sum_{i=1}^{N} h_i\right)^2\right]}} \qquad (1.13)$$

From (1.12) and (1.13), it is easily seen that, when f and h have zero means (and hence zero sums), we have $r = \cos\theta$ so, in this case, the normalized correlation coefficient is exactly equivalent to the Pearson correlation coefficient. When either vector does not have a zero mean, the two correlation coefficients are not equivalent: adding a constant value to all the elements of one of the vectors will, in general, change the angle between the vectors, yet the Pearson correlation coefficient will be unaffected.

Since filtering the image with a kernel, f, finds the correlation between f and the image at each point, the filtering process is a form of template matching, with the kernel, f, acting as the template. Consider filtering the image with a set of templates of different orientation: the correlation between each template and the image region beneath it is given by (1.12). For a particular image region, $|h|$ will be the same for each template, as long as the templates all have the same size and shape. If the set of templates is constructed so that $|f|$ is the same for each template, then the denominator in (1.12) can be ignored, as long as we are just interested in finding the best-fitting template across orientations at a particular position. Thus, the simple filtering process described by equation (1.7) indicates the best-fitting template at each position. The value at each point in the image obtained by filtering without dividing by the denominator in (1.12) is called the *unnormalized correlation coefficient*.

Filtering with an operator that has the shape of the feature to be detected is often described as *matched filtering*. If white noise is added to the signal, then the matched filter is optimal in the sense that it maximizes the signal-to-noise ratio (SNR) at the feature position. Given almost any noise distribution, it is possible to derive the linear filter that maximizes the signal-to-noise ratio (see Carlson *et al.*, 2002, pp. 388-391).

Table 1.1 shows a selection of templates described in the literature, commonly known as *compass gradient operators*. Two templates are shown out of each set of eight, one for

edges that vary towards the right, one for edges that vary towards top-right. In each case, the set of eight templates is formed by taking one template and rotating the sequence of eight outer pixels to each possible position. Thus, trivially, $|f|$ is the same for each template.

Each template is a model of a particular edge profile. But, because of the shape of an edge profile, these templates can also be considered gradient operators. In fact, Robinson's (1977) vertical and horizontal 3-level and 5-level templates are identical to the Prewitt and Sobel gradient operators described in the previous section. After filtering with the set of eight templates, the best-fitting (i.e. most strongly responding) template is selected at each point (giving the orientation), and the luminance gradient is given by the response level of that template. Thus, the template matching approach has a great deal in common with the differential gradient approach.

Name	Kernel for rightwards direction	Kernel for top-rightwards direction
Prewitt (1970) compass	$\begin{pmatrix} -1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 1 \\ -1 & -2 & 1 \\ -1 & -1 & 1 \end{pmatrix}$
Kirsch (1971)	$\begin{pmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{pmatrix}$	$\begin{pmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{pmatrix}$
Robinson (1977) 3-level	$\begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{pmatrix}$
Robinson (1977) 5-level	$\begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{pmatrix}$

Table 1.1. Compass gradient template kernels.

1.2.3 Overview of enhancement / thresholding edge operators

For detailed empirical and theoretical studies of enhancement/thresholding operators, see Abdou & Pratt (1979), Kitchen & Rosenfeld (1981), Kittler (1983), Kittler *et al.* (1983) Davies (1984) and Lyvers & Mitchell (1988). The standard Sobel operator is surprisingly accurate, with a maximum gradient magnitude error of 9.32%, and a maximum orientation error of only 2.90° (Lyvers & Mitchell, 1988). Davies (1984) and Lyvers & Mitchell (1988) argue that the profile of the Sobel operator is close to being optimal for this type of operator. The Prewitt and Sobel gradient operators give better magnitude and orientation estimates than the 2×2 Roberts operator (Abdou & Pratt, 1979). The template operators clearly give a much cruder estimate of orientation.

The analyses of the accuracy of these operators in estimating magnitude and direction of gradients would be irrelevant if the detectors could not distinguish edges from non-edges. The usual approach is to set a threshold, so that the points with a gradient magnitude above the threshold are considered to be edge elements. Without further processing, though, the results are disappointing. Roberts (1965) summed up the situation as follows:

> The hopelessness of this procedure is easily seen when one looks at typical pictures and considers them as a three-dimensional surface where the z [gradient magnitude] values are used as the height. Even a very clean input picture when viewed in this way looks like a bumpy, hilly landscape, with a broken-down stone wall representing the lines, and where some hills are higher than the top of other stone walls. If we imagine the clip [threshold] level as a flood over this landscape, there is no water level which covers all the hills and yet does not submerge some stone walls. In fact, even by adjusting the water level to be optimal for a particular area, a line will look like stepping stones in a rock-strewn brook rather than a smooth dam. (p. 170)

After thresholding, the edge maps can be cleaned up in two ways: non-maximum suppression (to find the gradient peaks), and edge linking (to fill the gaps). The non-maximum suppression algorithm usually takes each edge pixel and rejects it if there is a higher gradient within a certain distance in a direction perpendicular to the edge orientation (e.g. Rosenfeld & Thurston, 1971; Rosenfeld *et al.*, 1972). In this way, peaks

are found in a direction across an edge, but edge elements along the edge contour do not suppress each other. Edge linking algorithms are usually based on the assumption that edges should be smooth and continuous. This often involves consideration of the orientation of edge elements (e.g. Robinson, 1977; Nevatia & Babu, 1980), but not always (e.g. Roberts, 1965). Any edge pixels that cannot be linked into a smooth continuous boundary are discarded.

Even with a good edge linking algorithm, the quality of the initial map of edge elements is important. Abdou & Pratt (1979) used a heuristic measure based on the positional accuracy of edge elements and found that the 3×3 operators are clearly superior to the Roberts operator. Kitchen & Rosenfeld (1981) used an *edge coherence* measure, which measured how well-formed the edges were, and found similar results to Abdou & Pratt, with the 3×3 operators producing much more coherent edges than the Roberts operator. The Robinson 3-level template operator consistently produced the most coherent edges.

The edge detectors described so far perform reasonably well on noise-free images containing sharp edges, but they can perform very badly on images containing noise or blurred edges, including luminance ramps. Noise generally introduces high-frequency perturbations in the luminance, which may give rise to isolated pixels with high gradients. Blurred edges and ramps can give rise to large areas with above-threshold gradients. High-frequency noise is a particular problem for differential operators because differentiation amplifies high frequencies. Let y be a sine wave, with amplitude, a, and frequency, f:

$$y = a\sin(2\pi fx).$$

Then

$$\frac{d^n y}{dx^n} = a(2\pi f)^n \sin(2\pi fx + n\pi/2). \tag{1.14}$$

Taking *a* and *f* to be the noise amplitude and frequency, if *f* is very high, any derivative of the noise could be high in amplitude, however low *a* is. The next two sections describe different approaches that have been used to overcome these problems.

1.3 Noise reduction by surface-fitting

We can assume that the image results from the addition of noise to an underlying function. Ideally, the edge detection mechanism should find derivatives of the underlying function, not the noisy one. One approach, first suggested by Prewitt (1970), is to estimate the underlying function by fitting a surface to the image data before taking the gradient. This idea has been developed extensively by Haralick and co-workers, who have described a number of *facet models*, in which a smooth surface, or facet, is fitted to each $K \times K$ block of pixels. Within each facet, the grey level is described by a polynomial function. The degree of the polynomial determines the kind of surface used: a degree 0 polynomial gives a uniform grey level, a degree 1 polynomial gives a sloped plane, and degree 2 and 3 polynomials give quadratic and cubic surfaces, respectively.

The *sloped facet model* (polynomial of degree 1) assumes that, for each facet, *n*, the image region, *I*, can be described by the following equation:

$$I(x,y) = \alpha_n x + \beta_n y + \gamma_n , \tag{1.15}$$

where α_n, β_n, and γ_n are the parameters of the sloped plane in facet *n*. The parameters of the best-fitting surface are found for each block of $K \times K$ pixels. Equation (1.15) makes clear that α_n is the partial derivative (i.e. gradient) along the horizontal direction, and β_n is the partial derivative along the vertical direction. Thus, these parameters can be estimated using the derivative operators described previously. Haralick & Watson (1981) derived equations to find the best-fitting parameters for each facet (in the least-squares sense) and showed that, for 3×3 facets, the solution is equivalent to filtering with the Prewitt gradient operators (which are also Robinson's 3-level horizontal and vertical

templates). In addition, Haralick (1980) showed that fitting a planar surface to a 2×2 window is equivalent to filtering with the Roberts gradient operators. Thus, for small operators, there is a formal equivalence between three superficially different approaches to image processing: differentiation, edge template matching, and the fitting of a planar surface[3].

Haralick (1984) used cubic polynomial facets, which, like planar surfaces, can be fitted to the image by filtering with a set of kernels. Having fitted the surface to the image, derivatives could be found analytically. Rather than finding peaks the first derivative, Haralick located zero-crossings in the 2nd derivative. Haralick derived equations that allowed the locations of zero-crossings to be found from the coefficients of the fitted polynomials. These locations were marked as edges. The statistical reliability of these edge assertions could be assessed using a statistical test based on the error of the fit of the polynomial surface to the actual image data. Unreliable edges could then be rejected.

1.4 Noise reduction using large operators

1.4.1 Early large operators

One of the reasons that 3×3 operators give more coherent edges than the Roberts operator (Kitchen & Rosenfeld, 1981) is simply that the 3×3 operators are larger, and the effects of noise are averaged over several pixels. Around 1970, several researchers started to use much larger operators in order to reduce the effects of noise. Rosenfeld and colleagues (Rosenfeld, 1970, 1971; Rosenfeld & Thurston, 1971; Rosenfeld *et al.*, 1972) used difference-of-boxes (DOB) operators, which are rectangular operators consisting of two adjacent rectangles with opposite sign: the mean intensity on one side is subtracted from the mean on the other side. They found that these operators could locate edges in very noisy images. Macleod (1970, 1972) used operators that were the difference of two

[3] Schachter & Rosenfeld (1978) also showed that the magnitude of the gradient obtained from the Prewitt Gradient operator was equal to the Euclidean distance between the centroid and centre of the 3×3 neighbourhood of pixels covered by the operator.

offset Gaussian curves, termed DOOG by Young (1985, 1986) . Like the DOB operators, these have adjacent positive and negative regions, and subtract a weighted average on one side of the operator from a weighted average on the other side; unlike the DOB operators, the DOOG operators are smooth, with a distinct maximum and minimum in their profile.

The DOB and DOOG operators can both be considered differential operators because they find the difference between intensity levels in adjacent regions. In particular, the DOOG operator approximates the first derivative of the image after blurring with a Gaussian kernel. This is due to the fact that the DOOG operator has a profile very similar to the first derivative of a Gaussian (Young, 1985, Figure 4; Young, 1986, Figure 9) – in fact, the first derivative of the Gaussian is defined as the limit of the DOOG function as the size of the offset tends towards zero (Young *et al.*, 2001). Differentiation commutes with convolution[4], i.e. convolving with a kernel and then differentiating is equivalent to convolving with the derivative of the original kernel. Thus, convolving with a Gaussian derivative kernel (approximated by a DOOG operator) is equivalent to convolving with a Gaussian blurring kernel and then differentiating, or differentiating first, and then convolving with a Gaussian blurring kernel.

Around the same time that Rosenfeld and Macleod were developing their large differential operators, Hueckel (1971, 1973, 1974) was developing a large template operator. The templates were circular regions through which passed a straight, rectangular step edge[5]. The template was described by five parameters: three parameters described the equation of the straight line along the edge, and the other two parameters gave the intensity level on each side of the edge. It would have been unfeasibly slow to filter the image with each possible template, as was done with the template algorithms described in section 1.2.2. Instead, each circular region of the image was subjected to a kind of Fourier analysis in which it was represented as a linear combination of eight basis functions. These basis functions were specially constructed images, which allowed a low-pass filtered version of the circular image region to be reconstructed by adding together multiples of each basis

---

4 Convolution is filtering with the kernel rotated by 180° or, equivalently, reflected in both axes. Convolution has many convenient mathematical properties not shared by filtering, such as its commutativity with differentiation.

5 The term *rectangular step edge* will be used to refer to edges that consist of a discontinuity, with areas of locally uniform intensity either side. Often, the term *step edge* is used for this purpose, but some authors use this term to refer to any monotonically increasing edge function (e.g., Zhang & Bergholm, 1997).

function. Thus, the circular image region could be represented by a set of eight numbers, which were the multipliers of the basis functions. Hueckel showed that the parameters of the best-fitting template could be found analytically from these eight numbers, with only a small amount of computation. At each position, an edge was marked if the fit was sufficiently good.

Fram & Deutsch (1975) attempted a quantitative evaluation of the Rosenfeld, Macleod and Hueckel operators, using noisy edge stimuli. They derived two performance measures: P_1 was an estimate of the SNR in the operator's output, and P_2 was a measure of how patchy the output was (similar to Kitchen & Rosenfeld's (1981) edge coherence measure). They found that the Hueckel operator was clearly the worst on both measures. On the P_2 measure, the DOB and DOOG operators were quite similar, while, on P_1, the DOOG operator gave the best performance.

In commenting on both Rosenfeld's and Macleod's operators, Argyle (1971, 1972) argued that a central discontinuity is essential in an operator designed to detect rectangular step edges, because the optimal filter is the matched filter: the maximum SNR will occur when the filter kernel is perfectly correlated with the feature to be detected. Argyle (1972) argued that, if an edge is defined as a discontinuity, then the operator should be discontinuous, but to detect a blurred edge, a smooth operator is optimal. The next section explores the question of whether the optimal detector for rectangular step edges is smooth or discontinuous, while section 1.4.4 describes operators designed to detect blurred edges.

1.4.2 Operators for detecting rectangular step edges

Canny (1986) attempted a formal analysis of detection of rectangular step edges. He proposed three performance criteria: an edge detector should have a high SNR, should locate the edges accurately, and should give only one response to a particular edge. The presence of an edge was to be indicated by a peak in the filter output. He defined two performance measures, Σ and Λ, which reflected, respectively, a filter's SNR and its

localization accuracy in response to a step edge. He showed that, if the filter width increased by a factor w, Σ increased by $\sqrt{w}$, and Λ decreased by the same factor, so that the product, $\Sigma\Lambda$, remained constant. He therefore aimed to find the operator that maximized $\Sigma\Lambda$. The optimal operator, by this measure, was the DOB used by Rosenfeld, supporting Argyle's argument that a central discontinuity is optimal for detection of rectangular step edges. However, this analysis ignores the third criterion, that there should be only one response to an edge. The DOB has a very high bandwidth, causing it to be very sensitive to noise, so a noisy edge would give rise to many peaks in the vicinity of the correct edge location. Canny derived an expression for x_{max}, the mean distance between successive maxima in the filter's response to the noise. He then derived a set of operators, each one chosen to maximize $\Sigma\Lambda$ for a particular value of x_{max}. For small values of x_{max}, the optimal operator was quite similar to a DOB but, for larger values, the optimal operator was similar to a Gaussian 1st derivative. Thus, the requirement that there should not be spurious noise peaks close to the true edge location forces the operator to be smooth, even when it is optimized for detecting rectangular step edges.

Unfortunately, Canny's analysis has turned own to be flawed. His intention was that the localization accuracy measure, Λ, should be inversely proportional to the variance of the localization error, but Tagare & deFigueiredo (1990) found an error in his derivation of the formula for calculating Λ. Their point was initially misunderstood by Sarkar & Boyer (1991), who went on to repeat Canny's mistake, but Boyer & Sarkar (1994) later acknowledged the error. Koplowitz & Greco (1994) attempted to derive a correct expression for Λ. They derived the probability density function for the localization error, but found that its variance was infinite, so that Canny's intuitively reasonable-sounding localization accuracy measure could not, in fact, exist. Tagare & deFigueiredo (1994) also criticized Canny's multiple response criterion on the grounds that it was derived from the filter's response to the noise only, and may not be applicable when there is a signal as well as noise.

Despite these criticisms of Canny's attempt to derive an optimal filter, the main point still stands that a smooth filter will produce fewer spurious edges due to noise than a filter containing a discontinuity. In fact, using various other criteria, other researchers have found that smooth Gaussian-like filters are optimal. After demolishing Canny's attempt to

derive an optimal filter, Tagare & deFigueiredo (1990) attempted a derivation of their own which, like Canny's, assumed a rectangular step edge profile. They argued that a filter should have a high density of maxima around the true edge location, and that the density should drop off quickly for locations away from the edge. They proposed a measure, which they claimed reflected this property, and showed that a Gaussian 1st derivative filter was the optimal filter, according to this criterion. However, Tagare & deFigueiredo's analysis has not been accepted uncritically (Boyer & Sarkar, 1994).

It must be said that not all filters derived to be "optimal" are smooth. Lee *et al.* (1993) fixed localization accuracy at a particular value and derived a filter that optimized the SNR. The derived smoothing filter had a bell-shaped appearance, but its 1st derivative had a discontinuity in the centre. They actually implemented the 2nd derivative of the smoothing filter, which had a value of $-\infty$ in the centre, due to the discontinuity in the 1st derivative. They claimed that this did not present a problem in their discrete implementation, but there is no indication of how the filter performed with respect to Canny's third criterion, that there should only be one response to an edge. It was this criterion that forced Canny's filter to be smooth.

Canny showed that the DOB operator, which matches the rectangular step edge profile, maximizes the SNR, as is the case with all matched filters. Ben-Arie & Rao (1993a, 1993b) argued against using the conventional SNR measure to assess the performance of edge detection operators. The problem is that the conventional SNR is determined by the ratio between the squared response to the signal *at the feature point* and the variance of the response to the noise. This measure thus ignores the response in the vicinity of the feature point. Conventional matched filters, as well as Gaussian derivative operators, give rise to wide response profiles, which are difficult to localize. Ben-Arie & Rao (1993a, 1993b) argued that the ideal filter output would be a delta function (i.e. a single marked point) at the location of the feature to be detected, and no response anywhere else. Matched filters are not penalized for their wide response distributions because of the nature of the SNR. Ben-Arie & Rao proposed a new measure, called the *discriminative signal-to-noise ratio* (DSNR). This is determined by the ratio of the squared filter response at the feature location to the sum of the squared filter responses everywhere else. Thus, *any* response away from the feature point is considered noise. Unlike the conventional SNR, this measure encapsulates all of Canny's three criteria. They developed a method

that, given any template, generates the operator that minimizes the DSNR. The image is then filtered with this operator, rather than with the template itself. Ben-Arie & Rao showed that this method was able to give sharp, isolated peaks at the feature locations, and was far less disturbed by partial occlusion or superimposition of another image than matched filtering techniques. Rao & Ben-Arie (1993, 1994) used the expansion matching technique for detection of rectangular step edges. The filter that optimized the DSNR contained a discontinuity. The output in response to a noisy edge was a very narrow, high-amplitude peak, whose location was not distorted (unlike the Canny operator with which they compared their operator). The response at other regions contained many peaks due to noise, as would be expected from an operator that contained a discontinuity, but these were low-amplitude, and Rao & Ben-Arie argued that they could be removed with a threshold.

In summary, the answer to the question of what filter is optimal for detecting step edges depends on the definition of optimality. Smooth filters produce few peaks dues to noise, but produce broad response profiles that are difficult to localize. In addition, the blurring caused by smooth filters can cause systematic distortions of edge location (see section 1.7). Rao & Ben-Arie's (1993, 1994) filter, designed to maximize their DSNR measure, produces sharp, accurately located peaks in response to features, surrounded by many much smaller peaks due to noise.

1.4.3 Regularization operators

Poggio and co-workers have argued in favour of Gaussian-like filters (Poggio & Torre, 1984; Torre & Poggio, 1986; Poggio *et al.*, 1988). Their arguments were again related to noise reduction, but they approached the problem from the standpoint of regularization theory. They saw "*edge detection* as the process of computing derivatives in the two-dimensional intensity image" (Torre & Poggio, 1986, p. 148). Differentiation is an ill-posed problem, according to Hadamard (1923), who defined a well-posed problem as one that has a unique solution that depends continuously on the data. Roughly speaking, this last condition requires that small changes in the data lead to small changes in the solution. Differentiation fails to meet this condition because adding very low amplitude noise to an image can lead to a huge change in its derivative if the noise has a sufficiently high spatial

frequency content: as the amplitude of the noise tends to zero, its gradient can be made arbitrarily high by increasing its spatial frequency, as shown by equation (1.14).

A branch of mathematics known as regularization theory has been developed to find solutions to ill-posed problems (e.g., see Tikhonov & Arsenin, 1977). In cases where there are too many solutions, the approach is to find the solution that minimizes a certain functional, chosen to be appropriate for the task at hand. In the case of differentiation, where there is a unique solution, but it does not depend continuously on the data, the approach is to find an approximate solution that does depend continuously on the data. This can be achieved by smoothing the image, using a smooth filter kernel, thus reducing the high-frequency content.

Suppose the image is I. Ideally we want to find I', but this problem is ill-posed. Therefore, we instead find g', where g is a function similar to I, but smoothed, so that finding its derivative is a well-posed problem. g is found by minimizing the following expression:

$$\sum_i \left(I_i - g(x_i)\right)^2 + \lambda \int \left(g''(x)\right) dx \,.$$

The first term ensures that g is similar to I, and the second term (the simplest of Tikhonov's stabilizing functionals) ensures that g is smooth (i.e. that its average 2nd derivative is low). The trade-off between these two requirements is controlled by the *regularization parameter*, λ.

Poggio, *et al.* (1988)[6] showed that the function, g, that minimizes the above expression can be found by convolving I with a cubic spline function that is very similar to a Gaussian. Because differentiation commutes with convolution, we can find g' by convolving with the 1st derivative of this kernel. The regularization parameter, λ, controls the width of the kernel: a low value prioritizes the similarity between I and g, and gives rise to a narrow kernel that only slightly smoothes the image, whereas a high value emphasises the smoothness requirement, leading to a large smoothing kernel.

---

[6] A preprint of this paper is also available as widely-cited MIT AI Memo (Poggio *et al.*, 1985).

In summary, the regularization of differentiation (i.e. making it into a well-posed problem) leads to the derivation of a filter kernel close to the Gaussian derivative. Unlike Canny's derivation, no assumptions were made about the type of feature that is to be detected. Of course, the derivation of this particular filter kernel depended on the choice of the particular stabilizing functional. Poggio *et al.* (1988) did not give any justification for this choice, other than that it is the simplest of Tikhonov's stabilizing functionals. Since the stabilizing functional was chosen in a fairly arbitrary way, it must be concluded that the derived "optimal" filter kernel is arbitrary to some extent. However, it is likely that any derived kernel would be smooth, rather than discontinuous, since a discontinuous kernel has a high bandwidth and will not carry out the essential smoothing function required by this approach. Lacroix (1990) argued that, if the regularizing operator is to be circular (i.e. rotation-invariant) and separable (i.e. can be expressed as the product of two 1D functions in orthogonal directions: $f(x,y) = f_1(x) f_2(y)$), then the only acceptable operator is the Gaussian.

1.4.4 Operators for detecting blurred edges

1.4.4.1 The Laplacian of Gaussian operator

The first operator explicitly designed for detecting blurred edges was introduced by Marr & Hildreth (1980). Until then, most researchers had considered edges to be abrupt changes in grey level (e.g., Macleod, 1970; Rosenfeld, 1970; Rosenfeld & Thurston, 1971; Rosenfeld *et al.*, 1972; Hueckel, 1971, 1973, 1974; Robinson, 1977; Frei & Chen, 1977). Hueckel (1971) suggested that "exceptions may occur in a scene containing patches of fog or in a picture containing objects out of focus" (p. 113, footnote 1), but Marr (1976) argued that it is scenes that contain only abrupt changes that are the exceptions, and that most scenes contain blurred edges caused by shadows, curved surfaces and focal blur. Images contain a range of intensity changes with different *scales*[7], or blurs, i.e. the spatial distance over which the change occurs. Marr & Hildreth (1980)

---

[7] As pointed out by Shah *et al.* (1986), the term *scale* has been used to refer both to the width of an edge and to the size of the smoothing operator applied to the image. It is important to keep in mind the distinction between the two usages of this term.

argued that no single filter would be optimal for all edge scales, so an edge detection system should have filters with a range of scales, each optimized for a particular edge scale. In order to be sensitive both to an edge's position and its scale, the filter should be narrow in both the spatial and spatial frequency domains. Unfortunately, these two requirements conflict due to the uncertainty principle, so a filter that is narrow in one domain must be wide in the other. The filter that minimizes the product of the bandwidths in the spatial and frequency domains is the Gaussian (Leipnik, 1960).

Marr & Hildreth proposed filtering the image with a Gaussian operator and then taking the 2nd derivative, using the ZCs to locate the gradient peaks. The 2nd derivative should be taken in a direction perpendicular to the local line of ZCs. If the luminance along and parallel to a line of ZCs is constant or varies linearly (Marr & Hildreth's *condition of linear variation*), then this direction is also the direction of maximum slope across the ZC. Marr & Hildreth showed that, assuming the condition of linear variation, the set of ZCs returned by the 2nd directional derivative is the same as the set returned by the Laplacian operator, $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$, i.e. the sum of the directional 2nd derivatives in two orthogonal directions. Further, because differentiation commutes with convolution, the Gaussian blurring and Laplacian operation can be performed in one stage, by convolving with the Laplacian of a Gaussian, $\nabla^2 G$, or LoG. This operator is isotropic, (i.e. circular), and is bandpass in both the spatial and frequency domains. It has a central circular inhibitory region surrounded by an excitatory region.

1.4.4.2 Blurred template operators

Nalwa & Binford (1986) defined an edge as a discontinuity in intensity, but acknowledged that any optical system is imperfect and would give rise to a certain amount of blurring, which they assumed could be modelled by convolution with a Gaussian kernel. This implies that a rectangular step edge in a stimulus will give rise to an intensity profile in the image that is a Gaussian integral (henceforth, the term *Gaussian edge* will refer to an edge with a Gaussian integral profile). Rather than using Gaussian integrals, they constructed a set of edge templates using the tanh function, presumably because it was quicker to compute, but they showed that it was very close to a Gaussian integral. Each template

had 4 parameters: orientation, mean luminance, step height, and a sub-pixel offset. Only one template scale was used. At each point in the image, they found the best-fitting template and the best-fitting quadratic surface (which also had 4 parameters – orientation, and the 3 parameters that described its parabolic cross-section). An edge was marked at any point at which the best-fitting template fitted better than the best-fitting quadratic surface. The idea behind this was that, for an edge to be marked, the local image region should fit better to a function that has been specifically constrained to be edge-like than to a relatively unconstrained function with the same number of parameters. False positives could be reduced by eliminating any edges with a step size below a threshold value. The performance of this algorithm was quite comparable to that of Canny's algorithm (Heath *et al.*, 1997, 1998).

Rao & Ben-Arie's (1993, 1994) expansion matching scheme for step edge detection could be extended to blurred edges, since expansion matching is a completely general template matching process. Wang *et al.* (1996) used this method to detect blurred edges formed from two exponential functions.

Boie *et al.* (1986) argued that, since the matched filter gives the highest signal-to-noise ratio at the edge location, this filter should be used for edge detection. Assuming white noise, the matched filter has the same profile as the feature to be detected, so matched filtering is essentially a template matching operation. Although the matched filter is optimal for *detection* of edges, its output is not ideal for localization, because the feature being detected usually gives rise to a smooth peak, whose position could easily be disrupted by a small amount of noise. For localization, Boie *et al.* required a filter response that crossed through a threshold level (e.g. zero) at the edge point. They argued that, for accurate localization, filter should maximize the ratio of the filtered output *slope* to noise at the edge position. This is because the position of a zero-crossing with a shallow gradient can easily be disrupted by low-amplitude noise, whereas the position of a zero-crossing with a high gradient is less sensitive to noise. Boie *et al.* showed that the filter that maximizes the slope-to-noise ratio is the derivative of the optimal detection filter.

Boie *et al.* then addressed the question of the choice of signal in which to search for edges. Most template matching schemes match templates to the luminance/grey level signal. Boie *et al.* argued against this approach because, if the filter is matched to the

luminance profile of the edge, then an assumption must be made about how far the regions either side of the edge should extend. In most models, the "ideal edge" is a monotonic step function or sigmoid function, in which the central intensity change is flanked by two regions that extend away from the intensity change for an infinite distance. As pointed out by Boie *et al.*, the use of an infinitely wide filter of this form does not fit well with Marr & Hildreth's (1980) observation that edges are spatially localized, so that "the contributions to each point in the filtered image should arise from a smooth average of nearby points, rather than any kind of average of widely scattered points." (Marr & Hildreth, 1980, pp. 189-191): if the filter integrates over a large area, then different edges will interfere with each other. One solution is to allow the filter kernel to end abruptly at some distance from the edge, as with Rosenfeld's DOB kernels (Rosenfeld, 1970; Rosenfeld & Thurston, 1971; Rosenfeld *et al.*, 1972). The problem with this is that the choice of where the kernel should end is arbitrary. The filter will only be optimal if the width of the flanking regions in the kernel happen to match those in the image. Thus, the suitability of the kernel for detecting one particular edge is determined by the positions of other edges in the image: decisions must be made about parts of the image that form no part of the intensity change under consideration. Argyle (1971) was probably the first to raise this point in the context of edge detection (although Rosenfeld's (1971) rather heated reply to Argyle makes clear that Rosenfeld's edge detection methods were not matched-filtering algorithms). Argyle's solution was to use a "split Gaussian" filter formed by taking a standard Gaussian function and multiplying the values on the left hand side by -1. This allowed the filter to return to zero gradually, avoiding the decision about how far the flanking regions should extend. Although Argyle was correct in arguing that the split Gaussian filter would be less sensitive to noise than the DOB filter, it does not solve the problem that he had identified, because the filter is still only optimal for an edge that has a split Gaussian profile: Argyle's approach still makes an assumption about the profile of the flanking regions.

Boie *et al.*'s solution was to apply a matched filter in the gradient domain, rather than the luminance domain. If an ideal edge is differentiated, the result is a line or bar, surrounded by values equal to or close to zero. A filter kernel matched to the first derivative of the edge therefore integrates over a small area, without assumptions being made about the positions of flanking edges. Because of the commutativity of convolution and differentiation, the differentiation and matched filtering can be accomplished in one step,

by filtering with the derivative of the filter that matches the edge's first derivative. The matched filter in the gradient domain is therefore the second derivative of the edge profile. Boie *et al.* modelled edges as Gaussian integrals so, assuming white noise, the optimal detection filter for matching in the gradient domain is the first derivative of a Gaussian (i.e. a Canny filter), and the optimal localization filter is the second derivative of a Gaussian (i.e. a Marr-Hildreth filter). Boie *et al.* also derived optimal detection and localization filters under the assumption of a more realistic (non-white) model of television camera noise. The idea of template matching in the gradient domain is central to the Gaussian derivative template model of edge perception in humans, presented in Chapter 2.

Boie & Cox (1987) extended Boie *et al.*'s algorithm to two dimensions. Along one direction, their 2D filters had the profiles of the 1D filters derived by Boie *et al.*; along the orthogonal direction, the 2D filters had the profile of a Wiener filter. A Wiener filter is derived from the signal and noise power spectra, and its output is a "best guess" of the profile of the signal before the noise was added, in the sense that the filter minimizes the expected mean square error between the filtered output and the original signal.

Lacroix (1990) pointed out that regularized differentiation using his recommended Gaussian 1st derivative operator is equivalent to template matching with a Gaussian 1st derivative template. He argued that the Gaussian 1st derivative template could be applied at each orientation at each point in the image; the orientation with the strongest response would indicate the edge orientation, and the response magnitude would indicate the edge strength. It is not clear whether Lacroix considered his suggested operation to be template matching in the gradient domain, or just template matching in the image domain, with a vaguely edge-like template that would give the maximum response at the correct position and orientation.

1.5 The need for multiple spatial channels

Fram & Deutsch (1975) did Rosenfeld a disservice when they compared his operator with those of Macleod and Hueckel, because they failed to implement the most interesting

feature of his algorithms, which was their multiscale property. Rosenfeld (Rosenfeld, 1970; Rosenfeld & Thurston, 1971; Rosenfeld *et al.*, 1972) proposed a number of algorithms that used a range of operators of different size (these are described in more detail in Section 1.8). This idea was enthusiastically extended by Marr (Marr, 1976, 1982; Marr & Hildreth, 1980). Indeed, Marr & Hildreth's LoG operator was designed specifically to respond to a narrow range of spatial frequencies, so that the information at different spatial scales could be processed by different channels.

The problem with using a single large filter to remove high-frequency noise is that some high-frequency components of the image may arise from small, fine-detailed objects, or texture. A Hildreth (1983) put it, "it is common to find high frequency changes in intensity, due to the surface texture, superimposed on a more coarsely varying background intensity" (p. 2). De Micheli *et al.* (1989) distinguished between two types of noise in an image. The first type is caused by the image acquisition system (camera, eye, etc.), and should be removed if possible. They described the second type as *semantic noise*. This is high-frequency information, due to texture, etc., which exists in the signal, and is not an artifact of the image acquisition system. Semantic noise could be considered to be noise if the task is to find the boundary of a large object, but it may also carry important information. Thus, rather than being eliminated, the high-frequency information should merely be separated from the low-frequency information (Witkin, 1983). This can be achieved using multiple scales.

Elder & Zucker (1998) demonstrated what happens if only a single operator scale is used. They processed an image twice, once with a large Canny operator, and once with a small one. They showed that the small operator resolved fine details well, but produced a mass of broken line segments on the blurred edge of a shadow; the large operator handled the shadow well, but failed to capture the details in some parts of the image, and distorted the locations of some edges.

Most modern edge detection schemes use a range of operator scales, but there is no consensus on how best to combine information across scales. In the 1970s and early 1980s, researchers were hampered by the lack of a formal understanding of the effects of blurring an image. This deficit has been corrected by the flourishing field of scale-space theory.

1.5.1 Scale-space theory

One of the central ideas of scale-space theory is that, when representing data, it is important to use the correct scale. The "correct" scale depends on what structures in the data we want to know about, and the size of those structures. A good example of this is a map. The purpose of a street map is to tell us about the layout of streets. It needs to be detailed enough to show all the individual streets but, if it contained more detail than this, such as the layout of individual properties, the extra details would be an unnecessary burden, as they would not help us to understand the layout of the streets. On the other hand, a plan of an individual property must include precisely these details.

Images contain structures at a range of scales. The largest (the *outer scale*) is determined by the size of the image, and the smallest (the *inner scale*) is determined by the image resolution (Koenderink, 1984). The visual system cannot know *a priori* which scales should be used, so the only reasonable solution is to represent the image at all scales. Representations of the image at different scales can be found by blurring by different amounts.

One of the problems with multi-scale representations is how to link the representations at different scales. Witkin (1983) argued that this problem could be solved by making the filter scale, σ, a continuous parameter, and generating a family of blurred images. If $I(x,y)$ is a point in the image, then $I(x,y,\sigma)$ is a point in the *scale-space image*, a function of space and scale. The function $I(\cdot,\cdot,\sigma)$ is the image blurred with scale σ. Thus, $I(\cdot,\cdot,0)$ is the original image. The space that forms the domain of the scale-space image is called the *scale space*. Witkin's original presentation dealt only with 1D signals. This gave rise to a 2D scale space, with a spatial dimension, x, and a scale dimension, σ. Witkin was particularly interested in extrema or ZCs in the image or its derivatives. Each extremum (or ZC) in the image (or its derivative) corresponds to a contour in the scale-space image, which shows how its spatial position varies as scale increases (reflecting the fact that blurring an image distorts the locations of edges). As the scale increases, contours from extrema of opposite sign join together, forming n-shaped curves. Witkin

showed that these curves could be used to construct a hierarchical description of structures that exist at different scales in the image.

If two contours join together at a certain scale, to form an n-shaped curve, then they will not be present at larger scales in the scale-space image. This reflects the fact that blurring can remove features from the image. Witkin argued that it was important that blurring should always simplify the image – it should not create structures that did not exist in the original image. Thus, every contour in the scale-space image should originate in the unblurred image, where $\sigma = 0$. Individual contours can then be characterized by their location in the unblurred image, and the scale at which they disappear.

The non-creation of spurious features due to blurring has been referred to as *nice scaling behaviour* (Yuille & Poggio, 1986)[8], or *causality* (Koenderink, 1984), reflecting the requirement that each feature in the blurred image must be traceable back to a "cause" in the original image. The causality criterion was one of Marr's motivations for using a Gaussian filter:

> The reason why one chooses the Gaussian for this purpose, rather than blurring with a cylindrical pillbox function (for instance), is that the Gaussian distribution has the desirable characteristic of being smooth... And the reason, in turn, why this should be a desirable property of our blurring function is that if the blurring is as smooth as possible, both spatially and in the frequency domain, it is least likely to introduce any changes that were not present in the original image. (Marr, 1982, p. 56).

Marr's intuition turned out to be highly prescient: firstly, increasing the scale of a Gaussian filter *never* gives rise to new features; secondly, the Gaussian and its derivatives form the *only* family of filters that has this property. Witkin (1983) referred to an unpublished proof of this fact for 1D signals but, by the time it had been published (Babaud *et al.*, 1986), Koenderink (1984) had already provided the proof for 2D images. Yuille & Poggio (1986) gave another proof, and showed that, although any linear Gaussian derivative operator satisfies the causality criterion, a nonlinear directional operator operating in the gradient direction at each point[9], will not show nice scaling behaviour, whatever the filter shape. It

---

[8] A preprint of this paper is also available as a widely-cited MIT AI Memo (Yuille & Poggio, 1983).

[9] The computation of the directional derivative may make use of two or more oriented linear operators, but the way that they are combined makes the overall operation nonlinear.

should be noted that these proofs all apply to continuous signals. When dealing with discrete signals in a digital computer implementation, it is not strictly correct to use sampled Gaussian operators. Lindeberg (1990, 1993a) showed that modified Bessel functions should be used instead.

Scale-space theory has now developed into a formal, axiomatic system of mathematics (e.g., Lindeberg, 1994, 1997), which allows rigorous proofs of the behaviour of extrema and ZCs as the image and its derivatives are blurred. One useful result is that, as the scale increases, the intensity gradient of phantom edges created by 2nd-derivative filters increases, whereas the gradient of authentic edges decreases (Clark, 1988). This gives a means of distinguishing between phantom and authentic edges. Clark introduced the idea of a *reduced scale space* in which the phantom edge contours are removed.

Although Witkin's (1983) paper is almost universally credited as being the first published work on scale-space theory, it is interesting to note that axiomatic Gaussian scale-space theory was first developed in Japan in 1959 by Iijima, initially for the purpose of optical character recognition (see Weickert *et al.*, 1999). Unfortunately, this research was completely ignored by European and American researchers. By the time Witkin's paper was published, Japanese scale-space researchers had already made many discoveries that were subsequently rediscovered in Europe and America.

Most modern edge detection algorithms use the ideas of scale-space theory to some extent, and a selection of the most important ones is described in Section 1.8. Before describing these algorithms in detail, it is appropriate to describe similar developments concerning multi-scale representations that have occurred in the areas of physiology and psychophysics.

1.6 The link with physiology and psychophysics

So far, the analysis has come entirely from a computational standpoint. Many researchers, using different optimality criteria, have concluded that linear Gaussian

derivative filters (or similar filters) are optimal for detecting edges in noise. The fields of physiology and psychophysics have produced overwhelming evidence that similar linear filters exist in biological visual systems.

Kuffler (1953) showed that ganglion cells in the cat's retina have circular receptive field profiles that have a central excitatory or inhibitory region, and a surround that has the opposite sign to the centre (these cells are termed *on-centre* and *off-centre*). This finding was confirmed by Hubel (1960) and Wiesel (1960), and similar cells were found in the lateral geniculate nucleus (LGN) of the cat (Hubel, 1960). Similar results apply to monkeys (Hubel & Wiesel, 1960). These receptive fields can be modelled as the difference of two Gaussians of different scale, centred on the same point, known as DOG operators (Rodieck, 1965; Enroth-Cugell & Robson, 1966). Marr & Ullman (1981) showed that responses of LGN X cells to edges (Dreher & Sanderson, 1973) and retinal ganglion X cells to bars (Rodieck & Stone, 1965) are qualitatively similar to the responses of the LoG operator to these stimuli. Marr & Hildreth (1980) were so impressed with this similarity that they proposed that LGN X cells are the biological implementation of the LoG operator. The plausibility of this claim is discussed in Section 1.8.1.

Hubel and Wiesel also recorded from cells in the visual cortex of the cat (Hubel & Wiesel, 1959, 1962) and monkey (Hubel & Wiesel, 1968), and found oriented receptive fields. These can be modelled as Gaussian directional derivative operators, or DOOG operators (Young, 1985, 1986; Young *et al.*, 2001; Young & Lesperance, 2001).

Despite the qualitative similarity between receptive field and operator shapes, it was important to establish whether or not these cells could be considered to be carrying out the *linear* filtering operation required by the computational algorithms described earlier. Enroth-Cugell & Robson (1966) reported that retinal ganglion cells could be divided into two main classes, which they called X and Y. The X cells showed linear spatial summation, while the Y cells were distinctly non-linear. Hochstein & Shapley (1976a, 1976b) confirmed and extended Enroth-Cugell & Robson's work, with more comprehensive tests of linearity. LGN cells can be divided into linear (X) and nonlinear (Y) cells on the same basis (Shapley & Hochstein, 1975; Derrington & Fuchs, 1979). A similar dichotomy was established between *simple* and *complex* cells in the visual cortex (Hubel & Wiesel, 1962). Hubel & Wiesel classified cells as "simple" if they fitted the

following description: their receptive fields were subdivided into excitatory and inhibitory regions, there appeared to be summation within a region and antagonism between regions, and the optimal stimulus could be predicted from the spatial arrangement of inhibitory and excitatory regions. Complex cells had receptive fields that did not fit this description. Hubel & Wiesel's definition of simple cells implies that they are linear. They did not perform any quantitative tests of linearity, but, building on the work of Enroth-Cugell & Robson and Hochstein & Shapley, several researchers demonstrated that simple cells show approximately linear spatial summation, whereas complex cells are highly non-linear (Movshon *et al.*, 1978a, 1978c; Andrews & Pollen, 1979; Glezer *et al.*, 1980; Kulikowski & Bishop, 1981).

It is important to understand that the "linearity" shown by cells refers to their underlying membrane potential, rather than their overt response. Enroth-Cugell & Robson (1966) made clear that the cell responses were not linear functions of the stimulus amplitude. A "linear" cell can be modelled as an underlying linear operator followed by an output nonlinearity. Enroth-Cugell & Robson introduced a clever way of by-passing the effect of the output nonlinearity, by measuring the cell's *contrast sensitivity*, rather than its response amplitude. Suppose the stimulus is a sine wave, with spatial frequency, f, and contrast, C. Then the response, L, of a linear operator will be a sine wave with the same frequency. L, is given by

$$L = Ch(f),$$

where h is the spatial frequency tuning function of the underlying linear operator. If a cell is modelled as a linear operator, followed by a nonlinear transducer, T, then the response, R, of the cell, will be given by:

$$R = T(L) = T(Ch(f)).$$

If T is injective (i.e., no two elements in the domain map onto the same element in the co-domain, such as is the case with a strictly monotonic function), we can define the inverse function, T^{-1}. This gives:

$$Ch(f) = T^{-1}(R).$$

The contrast sensitivity, $S(f)$, at each frequency is the reciprocal of the contrast required to keep the response, R, at a constant level, R_0. Thus,

$$S(f) = \frac{1}{C} = \frac{h(f)}{T^{-1}(R_0)}.$$

Since R_0 is constant, $T^{-1}(R_0)$ will be constant, whatever the nature of the output nonlinearity, and will act as a constant scaling factor, the only effect of which will be to translate the contrast sensitivity curve vertically on a log-log plot. Thus, the nonlinear transducer has no effect on the shape of the contrast sensitivity function.

Although, with retinal ganglion cells, it is important to find the contrast sensitivity, rather than the response amplitude for a fixed contrast, both methods give very similar results with simple cells, as long as the contrast is not too high when using the response amplitude method (Movshon *et al.*, 1978b, Figure 1; Kulikowski & Bishop, 1981, Figure 4). Movshon *et al.* (1978c) argued that the principal nonlinearity in the output of simple cells is a half-wave rectifier with a threshold. The nonlinearities in simple cells are discussed in detail in section 7.1.3.

The peak spatial frequencies of LGN X cells and simple cells in visual cortical area V1 vary over a range of about 3 octaves, suggesting a similar range of receptive field sizes (Maffei & Fiorentini, 1973; Movshon *et al.*, 1978b; Andrews & Pollen, 1979; Derrington & Fuchs, 1979). These results suggest that biological visual systems use multiple scales of analysis. Campbell & Robson (1968) introduced the idea that the human visual system contains a set of independent channels, each tuned to a particular spatial frequency. This was supported by Pantle & Sekuler (1968) and Blakemore & Campbell (1969), who found that adaptation to sine wave gratings selectively suppressed sensitivity to gratings of a similar spatial frequency. Wilson & Bergen (1979) presented a model of early vision which used DOG filters of different spatial frequency. Marr & Hildreth (1980) showed that these DOG operators were almost indistinguishable from a LoG operator.

In summary, the physiological and psychophysical evidence strongly suggests that biological visual systems make use of linear operators that can be modelled as various forms of Gaussian derivative (or similar functions), and that a range of operators of different spatial scale are used. It has been less clear how the information from different spatial frequency channels is used. Section 1.8 describes the various ways that researchers have suggested that multiple scales could be used in edge detection. Before this, section 1.7 describes some of the edge location errors that occur with Gaussian derivative operators.

1.7 Edge location performance of Gaussian derivative operators

We have seen that many researchers, using different criteria, have deduced that the optimal operator for detecting edges is a smooth function that either is or approximates a Gaussian derivative, and that biological visual systems may contain operators of this kind. There is considerable variety within this general class of functions. Operators may be 1st or 2nd derivative; 1st derivative operators must be directional, but 2nd derivative operators may be either isotropic, like the Laplacian, or directional. All of these operators can correctly locate an isolated symmetrical edge of infinite extent but, in real images, the edges have finite length, and nearby edges can interfere, due to the blurring process. A number of researchers have studied the edge location errors made when applying different filters to different image features. This section describes the main results. In each case, the 1st derivative filters referred to take the 1st derivative in the direction of steepest gradient.

1.7.1 Edges flanked by ramps

Ulupinar & Medioni (1988, 1990) showed that, if a luminance discontinuity is flanked by ramps, the ZC in the output of a LoG operator deviates from the correct edge location. They presented several methods of correcting this error, described in Section 1.8.3.2.

1.7.2 Ramp edges

A ramp edge is a linear intensity change between two points in the image. Huertas & Medioni (1986) and Shah *et al.* (1986) have both analysed the performance of a LoG operator in response to this type of stimulus. With its theoretically infinite support, a LoG operator will locate a ramp, with a single ZC in the centre of the ramp. A practical implementation of this operator will have finite support and, if the ramp edge is wider than the support of the implemented LoG operator, then it will be undetected, as it will give rise to a plateau, rather than a ZC.

A 1st derivative filter often produces a mass of broken edges on ramps. This is because, when the intensity gradient is high over a large area, the gradient signal will be unaffected by any gradient threshold. After thresholding, edges are often taken to be points where the gradient has a peak in the gradient direction, even if the fluctuations in gradient are very small. Thus, slight errors in the gradient estimation on ramps can give rise to many spurious edges. Fleck (1992b) showed that these spurious edges can be eliminated by requiring not only that the gradient at the edge point is higher than the surrounding gradients, but also that the difference between the edge-point gradient and the gradients in adjacent pixels exceeds a threshold.

1.7.3 Bars and staircases

Two edges with opposite polarity in close proximity to each other form a *bar* (also known as a *pulse* or *line*). If the edges have the same polarity, they form a *staircase*. The two edges in these features interact with each other in characteristic ways.

Huertas & Medioni (1986) analysed the response of a LoG filter to a bright bar with a rectangular intensity profile of width, d. Let the diameter of the central excitatory region of the LoG filter be w. They claimed that, "for any d smaller than w the response is the same and the zero-crossings are displaced by $(w-d)/2$" (p. 652). They also claimed that, when d is larger than w, "the zero-crossings represent accurately the position of the edges" (p. 653). Neither of these statements was supported by mathematical proofs, and

they are both incorrect. Shah *et al.* (1986) derived an expression for the output, E, of a LoG filter in response to a rectangular bar formed by adding a rectangular step edge with amplitude 1, at position $x = 0$, to another one with amplitude P at position $x = d$:

$$E = -\frac{x}{\sigma^2}\exp\left(\frac{x^2}{2\sigma^2}\right) + P\frac{x-d}{\sigma^2}\exp\left(\frac{(x-d)^2}{2\sigma^2}\right). \qquad (1.16)$$

Equation (1.16) is their equation (11) with some changes in nomenclature. Huertas & Medioni's bar was symmetrical, so $P = 1$. As pointed out by Shah *et al.*, at positions $x = 0$ and $x = d$, E is non-zero, ruling out Huertas & Medioni's second statement. Now, the width, w, of the central region of the LoG operator is 2σ. If Huertas & Medioni were correct in saying that, for $d < w$, "the zero-crossings are displaced by $(w-d)/2$", then, when $d < 2\sigma$, substituting $-(2\sigma - d)/2$ for x in (1.16) should give $E = 0$. This can be disproved by counter-example.

Shah *et al.* (1986) studied the behaviour of bars and staircases in scale space using a LoG filter. Williams & Shah (1990) made a similar study using a Gaussian 1st derivative filter, deriving equations that described the paths through scale space that the gradient maxima followed. Both studies assumed that the bars and staircases had rectangular profiles. Zhang & Bergholm (1997) generalized these results to bars and staircases made from Gaussian edges.

The LoG operator causes the ZCs of a bar to move away from each other, by an amount that increases with the scale of the operator. If the sides of the bar are of unequal height, the low-amplitude ZC moves more than the high-amplitude ZC. The same applies to gradient peaks derived from a Gaussian 1st-derivative operator.

With staircases, the ZCs or gradient extrema travel towards each other when filtered with a LoG or Gaussian 1st derivative filter. If the two edges have the same amplitude then, as the filter scale increases, the ZCs/extrema eventually meet and become one; otherwise, the ZC/extremum corresponding to the weaker edge eventually disappears, and the stronger one asymptotes towards a position determined by the distance between the two edges in the image, and their relative amplitudes. With the LoG operator, as well as ZCs

corresponding to the two edges, there is a false ZC, of opposite polarity to the other two, that corresponds to the gradient minimum between the two edges (Shah *et al.*, 1986; Richter & Ullman, 1986). When the operator scale is substantially smaller than the distance between the edges, this false ZC may be missed, as it is barely distinguishable from a plateau. As the operator scale increases (or the distance between the edges decreases), the false ZC appears. With further increases in operator scale (or decreases in edge separation), the false ZC disappears again. If the two edges have equal amplitude, the false ZC disappears at the scale that the two other ZCs meet; otherwise, the false ZC eventually meets the weaker of the true ZCs, and they annihilate each other.

1.7.4 Corners

In the vicinity of a corner, the blurred image strongly violates Marr & Hildreth's (1980) condition of linear variation, i.e. the assumption that the intensity variation is linear along the edge. Consider two straight rectangular step edges extending away from a corner for an infinite distance. With the LoG filter, the ZC contour passes through the corner point, but does so in a curved trajectory, so that it bulges out around the corner (Berzins, 1984; De Micheli *et al.*, 1989). A 1st derivative filter misses the corner point, the contour following a smooth trajectory on the inside of the corner (Bergholm, 1987). In both cases, the error increases with increasing filter scale.

1.7.5 Finite-sized objects

As mentioned above, all Gaussian derivative filters can correctly locate an edge of infinite extent. In the case of a finite-sized object, such as a rectangle, the finite length of the edges introduces an additional source of error. Berzins (1984) showed that, when the sides of the rectangle are much longer than the scale of the filter, the ZCs from a LoG filter follow a trajectory around the corners similar to the infinite case, although the contour lies on the outside of the corners, no longer passing through the corner points: there is a bulge at each corner, with the minimum displacement occurring at the midpoint of each edge. As the rectangle gets smaller, the maximum deviation increases. For the

case of a square, when the width is less than 3.23σ, the ZC contour changes qualitatively, so that it becomes roughly circular, with the *maximum* displacement at the midpoint of each edge.

1.7.6 Curved contours

A straight-edged corner has infinite curvature. Berzins (1984) looked at the deviation of LoG ZCs from the correct locations as a function of curvature, using a homogeneous disc as the stimulus. He found that the deviation was a function of the ratio of the disc radius to the filter scale, with deviation decreasing with increasing disc radius (i.e., increasing with increasing curvature).

1.7.7 Trihedral vertices

De Micheli *et al.* (1989) examined the responses to trihedral vertices, which are points where three straight edge contours terminate, such as the vertex of a cube viewed so that the three surfaces in contact with the vertex are visible. A LoG filter gives rise to two ZC contours. One passes through the vertex, and approximately corresponds to two of the three edges; the remaining contour approaches the vertex approximately following the remaining edge, but then turns away from the vertex, to produce a spurious contour. Unlike the LoG filter, a 1st derivative filter produces contours that do not pass through the vertex, and there is no spurious contour.

1.7.8 Contour geometry

Torre & Poggio (1986) showed that ZCs contours from a LoG filter "are closed curves or curves that terminate at the boundary of the image" (p. 157). This gives rise to characteristic edge contours that look like an arrangement of rubber bands. Directional derivative filters do not necessarily give rise to closed contours, which means that they can produce short edge segments that do not contact the image boundary.

1.7.9 False edges

Section 1.1.1 explained how 2nd derivative filters give rise to spurious ZCs that correspond to gradient minima. That section also described strategies for eliminating these false edges. Marr & Hildreth's algorithm did not include any mechanism for rejecting false edges, which Clark (1989) considered a weakness, since about 15% of edges detected by 2nd derivative filters are false.

Fleck (1992b) showed that directional 1st and second derivative operators produce spurious rings around junctions of 4 or more regions (such as the corners of the squares in a checkerboard). She accounted for this as follows:

> The gradient must be zero at the vertex, near zero in the middle of each region, but significantly different from zero along the diagonal line segment joining these two points... Since the gradient direction for points on the diagonal lies along the diagonal, there must be a gradient peak somewhere along it. (p. 343)

The ZC contours resulting from a checkerboard filtered with a LoG do not deviate from the correct edge locations.

1.7.10 Edge Coherence

The purpose of using smooth Gaussian-like filters is to remove noise. Lunscher & Beddoes (1986) used Kitchen & Rosenfeld's (1981) edge coherence measure to assess the quality of the edges produced by a LoG filter. They found that the LoG ZC contours were more coherent than contours produced by the Sobel and Prewitt differential operators, and all of the template operators given in Table 1.1.

1.8 Edge detection algorithms

Section 1.4 discussed the design of edge detection operators, section 1.5 explained why a range of operators of different size are needed, and section 1.6 presented some

physiological and psychophysical evidence that multiple spatial channels exist in biological visual systems. Section 1.7 described some of the distortions in edge location that occur when using large operators. This section describes some edge detection algorithms in detail. Most of these algorithms are multi-scale, and the emphasis is on the methods used to combine information across scales. The algorithms of Marr & Hildreth (1980) and Canny (1986) are described first, since most subsequent work on edge detection has been strongly influenced by these two papers.

In order to achieve some kind of structure, I have attempted to classify the remaining algorithms into five categories. Inevitably, some algorithms have characteristics from more than one category, and these are classified according to their most prominent feature. Most of these algorithms are aimed at finding a solution to the fundamental problem, studied by Canny (1986), that small operators are sensitive to noise, whereas large operators distort the locations of edges. One type of algorithm uses large operators to reduce noise, and then attempts to correct the edge location errors. Another type of algorithm exploits the edge migration across scale space (as well as other scale space properties) and uses this information to help to characterize the edges. These two approaches are not mutually exclusive, as shown by Zhang & Bergholm (1997). Local scale selection algorithms vary the operator size across the image, choosing the "best" scale at each point. Many of these algorithms work by varying the operator scale inversely with the SNR. Filter combination algorithms apply the full range of scales to each image point, but then combine filter outputs across scale or orientation. Finally, a completely different approach is described, which uses phase congruency across scale to find edges. Section 1.9 focuses strategies for characterizing an edge in terms of its blur and contrast, and section 1.10 discusses the evidence relating to how edges are located by humans.

1.8.1 Marr & Hildreth (1980)

Marr & Hildreth's (1980) theory of edge detection was part of an ambitious theory of vision, outlined in Marr's (1982) book, *Vision*. The key idea behind Marr's approach was that, before starting to develop an algorithm, whether it is a computer program or a model of a brain process, it is essential to state precisely what the algorithm is supposed to be computing. This level of understanding of the problem is termed the *computational level*.

Marr defined two other levels: the *algorithmic level* is an abstract description of the algorithm that will carry out the computation, and the *implementational level* is a description of how the algorithm is to be implemented, either in computer hardware or biological cells. The aim was to replace the *ad hoc* heuristic approaches to vision that had existed previously with a rigorous approach, in which each vision problem was precisely defined, so that algorithms could be developed that could be proved to be "optimal in some sense or ... guaranteed to function correctly" (Marr, 1982, p. 19).

Marr & Hildreth's theory of edge detection can be viewed in terms of these three levels. The computational level requires the detection of intensity changes in the image. The algorithmic level states that these can be detected by locating ZCs in the output of a LoG filter. The implementational level states that the LoG filter is implemented by the X cells in the LGN. Previous derivative-based approaches to edge detection had used 1st derivative filters, which correctly carried out the task required by the computational level: they identified points at which the intensity changed sharply across the image. In contrast, the LoG operator introduced by Marr fails to carry out this task correctly, because it produces false ZCs, corresponding to points of minimal intensity change. It is ironic that, having placed so much emphasis on rigorous, provably correct algorithms, Marr then introduced a new operator which had a fundamental flaw, often producing the opposite of what was required by the computational level.

At the implementational level, Marr & Hildreth (1980) proposed that the on-centre X cells would carry the positive parts of the LoG signal and the off-centre cells would carry the negative parts. If two adjacent cells, one on-centre and one off-centre, were both firing, then a ZC must occur between them. Marr & Hildreth proposed that simple cells in the visual cortex could detect short, oriented lines of ZCs by acting as AND gates. The inputs to a simple cell would be two parallel lines of LGN X cells – one line of on-centre cells and one line of off-centre cells. The simple cell would fire when all the input cells were firing.

This proposed implementation has been questioned. First, it has been argued that the ratio of centre to surround diameters of X cells is too large to form an acceptable approximation of the LoG operator (Robson, 1983). However, Robson's argument was based on the characteristics of retinal ganglion cells, whereas Marr & Hildreth's specific proposal was that LGN X cells implement the LoG operator. The claim that a cell

implements a 2nd derivative operation predicts that, on a log-log plot, the spatial frequency tuning curve will have a slope of 2 on the low-frequency side, where the high-frequency attenuation does not have much effect. This is because the response is proportional to the square of the frequency (see equation (1.14)). Enroth-Cugell & Robson's (1966) data rule out retinal ganglion cells, as their slopes are usually around 1 or less. However, the tuning curves of LGN X cells can show slopes of around 2 (Maffei & Fiorentini, 1973; Troy, 1983), so the proposal that LGN X cells carry out a 2nd derivative operation is not entirely implausible from the physiological data. Stronger evidence against Marr & Hildreth's proposal was provided by Georgeson (1992), who studied the perception of plaid patterns formed from the sum of two or more sine wave gratings. He showed that the perceived edge structure of plaid patterns could be distorted by adaptation to or masking by oriented gratings. Since LGN cells are isotropic, they should show no orientation-specific effects – adaptation to an oriented grating should lead to, if anything, an overall decrease in response, but the zero-crossing positions should not change. Georgeson showed that, if a plaid was forward-masked by a grating aligned with one of the plaid components, the perceived edge structure was altered in a way that was perceptually equivalent to reducing the contrast of the plaid component aligned with the mask, suggesting that the perceived edge structure is determined by cells with oriented receptive fields, rather than LGN cells.

Marr & Hildreth's proposal that simple cells act as AND gates has also been criticised. Schumer & Movshon (1984) set out to test the predictions of the AND model. They argued that, with the AND model, there should be a criterion edge length below which a simple cell did not fire, since not all its inputs would be firing. They found, instead, that the simple cell response increased gradually with length over a wide range. In addition, they found that contrast could be traded for length in just the way that would be expected from a cell showing linear summation. Marr & Ullman (1981) admitted that simple cells would probably show a weaker nonlinearity than a pure AND function, and that, consequently, "it is hard to make quantitative predictions about the response of such units to arbitrary stimuli" (p. 171). Nevertheless, the quasi-linear spatial summation shown by simple cells makes it unlikely that they are implementing an AND function (Tolhurst & Dean, 1987). If a cell were acting as anything approximating an AND gate, it should not be possible to trade off stimulus contrast against spatial extent in the linear fashion shown in Schumer & Movshon's data.

One possibility, mentioned by Marr & Ullman, is that only a small minority of simple cells act as AND gates, the other cells having other functions. Richter & Ullman (1986) recording from simple cells while presenting staircase stimuli consisting of two edges. They reported finding some cells that may have been detecting ZCs. As described in section 1.7.3, when the edge separation in a staircase is substantially larger than the operator width, there are two true ZCs. Then, as the edge separation decreases, a false ZC with opposite polarity appears. With further decreases in edge separation, the two true ZCs merge into one, and the false ZC disappears. Richter & Ullman found simple cells that were selective for edges of a particular polarity moving in a particular direction. If the staircase was arranged so that the true ZCs had the cell's preferred polarity then, as the stimulus moved through the cell's receptive field, the cell gave one or two pulses, depending on the edge separation. When the polarity of the staircase was reversed, and it was moved in the same direction, the cell gave a single pulse, even though the stimulus contained no edges of that cell's preferred polarity. Richter & Ullman argued that the cell might be responding to the false ZC, which would have the cell's preferred polarity when the staircase was reversed in polarity. In support of this, they found that the cell responded to the polarity-reversed staircase only for intermediate edge separations (when the false ZC would be expected to occur). Further, they found that the smallest edge separation for which the polarity-reversed staircase elicited a response was the edge separation at which the response to the non-reversed staircase changed from one pulse to two. This edge separation could correspond to the separation at which the two true ZCs meet, and the false ZC disappears.

One of the most unsatisfactory aspects of Marr & Hildreth's theory of edge detection is their method of combining information across spatial frequency channels. First, the ZCs are grouped into short, straight, oriented ZC segments. If a ZC segment is present with the same location and orientation in a contiguous range of channels, they are combined, and an edge segment is marked at that position. Their justification for this algorithm is that "physical edges will produce roughly coincident zero-crossings in channels of nearby sizes. The spatial coincidence assumption asserts that the converse of this is true, that is the coincidence of zero-crossings is sufficient evidence for the existence of a real physical edge." (Marr & Hildreth, 1980, pp. 211-212). But Witkin (1983) showed that, in most complex signals, the ZC locations tend to move as the channel scale increases. This

means that looking for ZCs that are coincident across scales is likely to fail. Marr & Hildreth were aware that ZCs can move as the scale changes, and they devised a complex system of rules for combining ZCs across scale. These rules were never published in full, but a sketch of some of them was given by Marr & Hildreth (1980) and Marr (1982). Marr (1982) wrote that "the actual details of the rules are quite complicated because a number of special cases have to be taken into account" (p. 70). These special cases seem to diminish the mathematical elegance of the theory, and suggest that the end-product is no less *ad hoc* than the previous approaches to edge detection, of which Marr (1982) had been quite critical.

One special case is that of narrow bars, which, as described in Section 1.7.3, show movement in ZCs as the operator scale increases. Marr & Hildreth's (1980) solution was that "in these circumstances, the larger channels must be ignored, and the description formed solely from small channels of which the zero-crossing segments do superimpose." (p. 204). This completely undermines their multi-scale approach: information from larger scales is ignored unless it agrees precisely with information from smaller scales, in which case it is redundant. There would therefore be no point in having more than one channel, unless there were zero-crossings at large scales which did not correspond to zero-crossings at smaller scales, but this situation cannot arise (Koenderink, 1984; Yuille & Poggio, 1986; Babaud *et al.*, 1986).

Another special case is the disappearance of ZCs with increasing scale. Marr & Hildreth stated that "if a discernable zero-crossing is present in a channel centred on wavelength λ_0, there should be a corresponding zero-crossing at the same spatial location in channels for wavelengths $\lambda > \lambda_0$" (p. 202). Research into scale-space theory has proved this to be incorrect. If a ZC is present at wavelength λ_0, a corresponding ZC will be present at all *smaller* channels (not necessarily at the same location), but ZCs tend to disappear as the scale increases (Koenderink, 1984; Yuille & Poggio, 1986; Babaud *et al.*, 1986). Marr & Hildreth were aware that ZCs could disappear as the scale increases, when "two or more local intensity changes are being averaged together in the larger channel" (p. 202), but they seemed to regard this as an exception, rather than a general occurrence.

In summary, Marr & Hildreth's theory of edge detection fails to carry out the basic task set at the computational level, and requires so many *ad hoc* rules to deal with special cases, that the full set of rules was never published. Their method of combining information across scale is deeply flawed in that it makes false assumptions about the behaviour of ZCs across scale, and ignores large-scale information when it conflicts with information at smaller scales, rendering the large-scale channels redundant.

1.8.2 Canny (1986)

Section 1.4.2 described Canny's (1986) flawed derivation of "optimal" operators that could be approximated by Gaussian 1st derivatives. The derivation was performed in 1D. In 2D, each operator is a partial 1st derivative of a Gaussian in a particular direction. At each point, the direction, $\mathbf{n}$, of steepest gradient was found, and a 1st derivative operator, $G_{\mathbf{n}}$, with that direction was applied. "An edge point is defined to be a local maximum (in the direction $\mathbf{n}$) of the operator $G_{\mathbf{n}}$ applied to the image I." (p. 691)

Canny used the term *feature synthesis* to refer to his method of combining information across scales. The algorithm starts by marking the edges at peaks in the smallest scale filter outputs, since these will be localized with the most accuracy. Then, assuming these are the only edges, the larger-scale filter outputs are predicted. Peaks are marked in the next-largest filter output only where it is "significantly" (not defined) greater than the predicted output. This allows the algorithm to handle blurred edges but, unlike Marr & Hildreth's algorithm, does not rely on the hopeless task of trying to match up ZCs from different scales.

Canny introduced several novel ideas. One was his *hysteresis thresholding* scheme. Like many other algorithms, he applied a threshold to the gradient signal, to remove noise. A single threshold may cause streaking, even with low noise levels, because low-gradient contours may fluctuate above and below threshold along their length. The hysteresis technique uses two thresholds, a low one and a high one. Any edge points above the high threshold are marked, and the whole of any edge segment containing a marked point is then marked if all the points in the contour are above the low threshold. This works on

the assumption that an isolated weak edge point is likely to be noise, whereas a connected line of them that includes a strong edge point is likely to be a genuine boundary. Streaking is reduced because it only occurs if the edge gradient fluctuates above the high threshold and below the low one.

Another important idea that Canny introduced was stimulus-dependent combination of operators. He showed that operators had better signal-to-noise ratio and localization accuracy if they were elongated along the length of a straight edge contour. But, if elongated operators were hard-wired into the system, the result was "false responses to highly curved edges and extension of edges beyond corners". (p. 695). The solution is to create elongated operators by adding the outputs of a line of operators (say 5) along the edge in those parts of the image where elongated operators are appropriate. The variance of the outputs indicates whether this filter combination is appropriate: if it is low, then there is a good fit to the edge, and the outputs can be summed to produce an elongated operator; if it is high, then the non-elongated operators described above are used. This filter combination procedure is highly relevant to the question of how edges are detected in biological visual systems, because there is psychophysical evidence that filter combination occurs in the human visual system, both across orientation and spatial frequency (Watt & Morgan, 1984; Olzak & Thomas, 1991, 1992, 1999; Georgeson, 1992, 1998; Meese & Freeman, 1995; Meese & Georgeson, 1996a; Georgeson & Meese, 1997, 1999). This evidence is described in Section 1.8.6.

A further idea introduced by Canny was the suggestion that the operator scale at each point in the image could be chosen according to an estimate of the local noise energy. At points where the noise is high, a small operator would be too unreliable, so a large operator should be used to smooth out the noise; at points with low noise, a small operator should be used, to ensure minimal distortion due to blurring. The idea is to use, at each point, the smallest operator with an acceptable SNR. Canny gave few details of how this scheme could be implemented, but the basic idea of choosing the smallest reliable operator at each point has since been taken up by several other researchers, whose algorithms are described in section 1.8.5 (Fleck, 1992a; Morrone *et al.*, 1995; Elder & Zucker, 1996, 1998; Marimont & Rubner, 1998).

1.8.3 Algorithms that correct spatial distortions

The algorithms described in this section all use large operators to reduce noise, and then take steps to correct the distortions in edge location that these operators are expected to introduce.

1.8.3.1 Bergholm (1987): edge focusing

Bergholm (1987) used large Canny operators to detect the edges, removing noise, but then reversed the blurring process by tracking the edges through scale-space from coarse to fine scales. This process is termed *edge focusing*.

First, the whole image is blurred with a Gaussian filter, and peaks in the first directional derivative are marked in an edge map. Then a slightly smaller filter is applied only in the neighbourhood of each edge point. The old edge points in these neighbourhoods are thrown away, and replaced by the edge points derived from the smaller filter. This procedure is repeated until the smallest filter scale has been reached. The size of the step in scale is chosen to be sufficiently small, so that an edge is unlikely to move more than 1 pixel in each focusing step. At some points, an edge will spit, so the final edge map has more detail than the original.

This procedure is based on the assumption that the significant edges are those that survive at high blurring scales. Elder & Zucker (1998) argued that survival at high blurring scales is not a reliable indication of edge significance: "generally, edges will survive at a higher scale if they are high contrast, in focus, and isolated." (p. 701). They argued that many significant edges fail to meet these requirements. To be fair, though, Bergholm discussed the possibility of extending the scheme so that edges that disappear at high blurring scales could also be identified and focused: "a worthwhile extension of the focusing scheme would be to add edges which did not appear at the coarsest scale, but which gave a strong response at some intermediate scale." (p. 737).

Bergholm noted that blurred edge contours, when focussed, tend to break up into a cloud of pixels. This might be seen as a disadvantage but, undaunted, Sjöberg & Bergholm (1988) exploited this finding and used it to classify edges as blurred or sharp. Their algorithm classified an edge as diffuse if a sufficient number of new edge points appeared at smaller scales. This allowed them to distinguish shadows from boundaries of objects. This algorithm gives a very crude binary-valued estimate of blur. A disadvantage, pointed out by Sjöberg & Bergholm, is that areas with a lot of intricate detail may be incorrectly classified as shadows.

1.8.3.2 Ulupinar & Medioni (1988, 1990)

Ulupinar & Medioni (1988, 1990) modelled an edge as a luminance discontinuity with ramps on either side. When filtered with a LoG operator, such an edge profile will give rise to a ZC shifted from the true edge location. They reported several methods to correct this bias. One method involved using LoG filters with two different scales. They showed that ZC location is a function of LoG scale with two unknowns, the edge location, and the ratio of slope difference to step height. Using two scales gives simultaneous equations that can be solved for both of these unknowns. A problem with this method is that it requires corresponding ZCs at different scales to be matched up and, as shown by Marr & Hildreth (1980), this is a difficult problem.

Ulupinar & Medioni also reported a single-scale method that made use of the positions of the local maximum and minimum either side of the ZC, as well as the ZC position itself. They showed that the true edge location could be found simply by adding the positions of the maximum and minimum and subtracting the ZC position. In addition, the step height and slope difference could be found analytically. They then generalized this method and showed that the true edge location could be found from the ZC location and the filter outputs at any two locations away from the ZC. Although their method did not explicitly deal with interactions between edges, it was able to correct for these interactions to some extent. For example, a bar could be modelled approximately as a discontinuity with steep slopes of opposite polarity on each side. The usefulness of this method depends on whether edge features in general can be modelled in this way with sufficient accuracy for the edge location correction to give satisfactory results.

1.8.3.3 Chen & Medioni (1989)

Chen & Medioni (1989) used a LoG filter to remove noise and used an iterative method to restore the correct edge locations. In the initial presentation of their algorithm, they assumed that the image could be constructed by adding rectangular step edges together. Each edge, *j*, has position, z_j, and step height, s_j. ZCs in the LoG filter output give the initial estimates of each z_j. On each iteration, the following sequence occurs. For each edge, *j*, the step height, s_j, is estimated analytically. The image can then be reconstructed by adding together a set of rectangular step edges with the estimated positions and heights given by each z_j and s_j, and the reconstructed image can be filtered with the LoG filter. This reconstruction of the filtered image is performed analytically, and the result is used to find the next estimate of the edge positions, z_j. The iteration continues until the estimated edge locations have reached stable states. The iteration does not diverge, although oscillations can occur with low-contrast edges, so these are removed by thresholding.

Chen & Medioni showed that this algorithm was successful at restoring the positions of edges that had been distorted by the filtering process. Their algorithm was designed for a 1D signal, but they applied it to the individual raster lines of a 2D image and found that they could reconstruct the image quite well by adding step edges with the estimated positions and step sizes.

Since the original algorithm assumed that all edges were rectangular step edges, it could not be expected to give good results with images containing blurred edges. They extended the algorithm by assuming that a blurred edge is a Gaussian integral. They showed that applying the rectangular step edge model caused the step size to be underestimated:

$$\frac{\text{reconstructed step size}}{\text{true step size}} = \frac{1}{\sqrt{1+\left(\frac{\sigma_b}{\sigma_m}\right)^2}},$$

where σ_b is the edge blur and σ_m is the scale of the filter. This equation has two unknowns: σ_b and the true step size. Running the algorithm at two different filter scales produces simultaneous equations that can be solved for both unknowns.

1.8.4 Algorithms that characterize edges in scale space

1.8.4.1 Korn (1988)

Korn used Gaussian 1st derivative operators, but normalized them by multiplying the response by $\sigma\sqrt{2\pi}$, where σ is the filter scale. This means that the integral on each side of the ZC of the operator is equal to 1 for every scale. Clearly, this causes the response to rectangular step edges to be constant across scale.

Korn detected gradient maxima at the finest scale, and then tracked edges on paths through scale space towards coarse scales, i.e. in the opposite direction to Bergholm (1987). At each scale along the path, the filter response was found. The function mapping filter scale onto filter output along a path was termed the *Gaussian difference signature* by Zhang & Bergholm (1997). Korn determined the scale of the edge from first peak in the Gaussian difference signature, moving from fine to coarse scales. He applied this scheme to a signal containing many edges of different blur and showed that the selected scale for each edge did seem to correspond to the edge blur (Korn, 1988, his Figure 16). However, there is a theoretical flaw in this method. It is easily shown that the Gaussian difference signature in response to an isolated Gaussian edge increases monotonically[10] (Zhang & Bergholm, 1997). A peak only occurs in the Gaussian difference signature when the edge is in the vicinity of other edges. In this case, the scale at which the peak occurs depends on the proximity of the other edges, as well as the scale of the edge being examined, and so the scale of the peak is not a good measure of edge scale. A further problem with Korn's method is that, by tracking edges from fine to coarse scales, the algorithm is very sensitive to noise. Any noise-reduction advantage of using large scale operators is lost, because the edges are detected using the smallest operators.

[10] In the limiting case of a rectangular step edge, the Gaussian difference signature is flat.

1.8.4.2 Williams & Shah (1993)

Building on their earlier work, described in section 1.7 (Shah *et al.*, 1986; Williams & Shah, 1990), Williams & Shah (1993) examined the behaviour of edges when filtered using Korn's (1988) normalized Gaussian 1st derivatives. A ramp gives a response that increases linearly with scale for small operators but, when the tails of the operator start to fall substantially outside the extent of the ramp, the response starts to approach an asymptote, which equals the edge amplitude. The operator scale at which the increase in response starts to deviate substantially from linearity indicates the ramp width, and the response from the largest filter indicates the edge height. The ramp slope can be found, if required, by dividing the height by the width. Williams & Shah's work was extended to Gaussian-blurred edges by Zhang & Bergholm (1997), whose algorithm is described next.

1.8.4.3 Zhang & Bergholm (1997)

Zhang & Bergholm (1997) used Bergholm's (1987) edge focusing method with Korn's (1988) normalized Gaussian 1st derivatives. Each gradient maximum in the large-scale filtered image gave rise to a *scale path* through scale-space. Their method therefore began in a very similar way to Korn's algorithm, except that the scale paths originated in the coarsest scale, rather than the finest scale. For each scale path, they found the Gaussian difference signature (defined in section 1.8.4.1), which they used to classify and characterize each edge.

Zhang & Bergholm classified edges into 2 broad types: *step edge* and *double edge*. A step edge is a monotonic intensity change, and a double edge is two intensity changes in close proximity. As stated in section 1.8.4.1, Zhang & Bergholm showed that a step edge has a Gaussian difference signature that is a monotonically increasing function of operator scale, whereas a double edge has a Gaussian difference signature that peaks at a particular scale. This fact was used to classify each edge as a step edge or double edge.

The double edge classification can be further subdivided into *bar edge* and *line edge*. A bar edge has approximately equal intensities on each side, whereas the intensities differ on either side of a line edge. Edges can be classified by means of the observation that, with a

bar, the two edges move away from each other as scale increases whereas, with a line edge, the weaker edge eventually disappears. For each candidate double edge, the algorithm searches for another nearby edge at the smallest scale and the largest scale. The candidate edge is classified as a line edge if (1) no matching edge can be found at the coarsest scale, or (2) the candidate edge is closer to the large-scale matching edge than the small-scale matching edge. Otherwise, the candidate edge is classified as a bar edge.

After the edges have been classified into edge type, their attributes are estimated. Zhang & Bergholm modelled step edges as Gaussian integrals, and double edges as the sum of two Gaussian integrals. A step edge and line edge have attributes of height and diffuseness (blur). Each combination of attributes in these ideal edges gives rise to a Gaussian difference signature template. The attributes of each edge are found by taking its Gaussian difference signature and finding the best-fitting Gaussian difference signature template. A bar edge has parameters of diffuseness and width (i.e. the distance between the edges); the height of the bar is determined by these two parameters. The Gaussian difference signature has no spatial dimension, so it is not appropriate for measuring the width parameter. Instead, the attributes of a bar edge are estimated using the *apparent width signature*, which gives the distance between two edges as a function of operator scale. As with the Gaussian difference signature, each combination of parameters gives rise to a different apparent width signature template, and the edge parameters are estimated by finding the best-fitting template.

1.8.5 Local scale selection algorithms

The algorithms described in this section address the problem of combining information across scales by selecting (usually) one "best" scale at each point.

1.8.5.1 Rosenfeld & Thurston (1971) and Rosenfeld *et al.* (1972)

Rosenfeld & Thurston (1971) presented the first edge detection algorithm to perform local scale selection. Rosenfeld *et al.*'s (1972) algorithm is similar, but is simpler and performs better, so it is the latter that is described here. The algorithm uses difference-of-boxes (DOB) operators with a range of sizes, k, and four different orientations (horizontal, vertical, and the two diagonals). The method runs as follows. Convolve the image with each operator, giving a set of convolution images. At each point, pick the best orientation, i.e. the orientation of the operator that gives the highest response magnitude at that point. Then, at each point, pick the best scale at the chosen orientation, i.e. the largest scale, k, for which the next smallest scale does not give a substantially higher response. Then erase a point if there is a higher point within a distance $k/2$ in a direction perpendicular to the point. Rosenfeld *et al.* noted that there are some image configurations for which their method will put edges in the wrong place, but they claimed that overall it does a reasonable job.

1.8.5.2 Marr (1976)

Rosenfeld *et al.*'s (1972) algorithm was designed to detect large-scale edges in noise. Because of this, their algorithm was biased towards large operators – a smaller operator was only used if its response was substantially higher than that of a larger operator. Marr (1976) presented an algorithm, using DOB operators, that was very similar to Rosenfeld's algorithm, except that it was biased towards smaller operators: "mask size s is selected at point P in the image whenever (a) masks slightly smaller than s give an appreciably smaller peak at P, and (b) slightly larger masks give a peak that is not appreciably larger" (p. 488). He plotted peak response as a function of filter size and found that, while the step edge gives the same response for all filters, a ramp edge gives a response that increases with filter scale, asymptoting towards the peak response to the step edge. Using Rosenfeld's scale selection scheme would always choose the highest scale for a ramp edge, whereas Marr's scheme chooses the scale that corresponds roughly to the scale at which any further reduction in scale would cause a relatively large decrease in peak response.

In both cases, if there is a clear single peak in response over scale, then the scale corresponding to that peak will be selected. But the schemes behave differently with a near-plateau. Rosenfeld's scheme will choose a scale on the large-scale side of the plateau, whereas Marr's scheme will choose a scale on the small side. Additionally, Marr allowed more than one scale to be selected at each point, whereas Rosenfeld selected the largest scale that fitted the selection criteria.

Marr's graphs of filter response as a function of operator size were a precursor of Zhang & Bergholm's (1997) Gaussian difference signatures, described in sections 1.8.4.1 and 1.8.4.3.

1.8.5.3 Perona & Malik (1990)

If an image is blurred with a Gaussian kernel, the pattern of intensity is modified in the same way as a pattern of heat after diffusing over a uniform conducting surface (Kovaznay & Joseph, 1955). Thus, the evolution of the image over scale in scale space is an exact model of the evolution of a pattern of heat over time. In fact, the *diffusion equation*, which describes the diffusion of heat, plays a crucial role in the proof that only Gaussian smoothing shows nice scaling behaviour. Essentially, the proof amounts to demonstrating that, for the scale-space image to show nice scaling behaviour, it must satisfy the diffusion equation. Assuming that the conducting surface has uniform conductivity, convolution with the Gaussian kernel uniquely describes the solution to the diffusion equation, so this kernel is the only kernel with nice scaling behaviour. Perona & Malik (1990) pointed out that the assumption of uniform conductivity is unnecessary and causes many of the problems associated with blurring, since it leads to a space-invariant blurring kernel. Although blurring has the beneficial effect of removing noise within an image region, it also distorts the edges between regions, because the different regions diffuse into each other. Perona & Malik argued that the benefits of blurring could be maintained, without the disadvantages, if the blurring kernel was smaller at the edges. This is equivalent to making the conduction coefficient lower at the edges, allowing rapid diffusion within a region, but not between regions. This process is called *anisotropic diffusion*. Since the diffusion equation is still obeyed, anisotropic diffusion satisfies the causality criterion.

In order to make the conduction coefficient low at the edges, it is necessary to know where they are. This is problematic, because the position of the edges is what the algorithm is trying to find. The solution is to use the intensity gradient as a current best guess of the edge locations. The conduction coefficient is defined as a monotonically decreasing function of the gradient magnitude. A number of different functions will suffice, one such candidate being the Gaussian. Perona & Malik's anisotropic diffusion process proceeds in an iterative manner: at each iteration, the conduction coefficient is set at each point, using a decreasing function of the gradient magnitude, and the diffusion process is simulated using some simple nearest-neighbour operations. The result is a series of images that get progressively simpler, yet the locations of the edges do not move.

The problem with this method is that the initial estimates of the edge positions are estimated before the filtering takes place, and so the assignment of conduction coefficients to locations can be disrupted by noise. Perona & Malik admitted that "in images in which the brightness gradient generated by the noise is greater than that of the edges, and the level of the noise varies significantly across the image the scheme that we have described proves insufficient to obtain a correct multiscale segmentation" (p. 637).

1.8.5.4 Jeong & Kim (1992)

Jeong & Kim (1992) produced versions of the Canny and Marr-Hildreth algorithms, modified so that a single "best" scale, σ, was chosen at each point, (x, y). The aim was to find a scale function $\sigma(x, y)$ that had three properties:

1. $\sigma(x, y)$ is large in uniform intensity areas, to smooth out noise;
2. $\sigma(x, y)$ is small at significant intensity changes, to recover edge location accurately;
3. $\sigma(x, y)$ does not change rapidly, preventing broken edges.

They defined an energy functional, E, which took the function, σ, as its argument:

$$E(\sigma) = \iint (I - G * I)^2 + \lambda \left| \nabla \frac{1}{\sigma(x,y)} \right|^2 dxdy ,$$

where I is the image and G is a Gaussian filter with scale σ, which varies over space; λ is a stabilizing parameter, chosen by hand. The function, σ, which minimized this energy function, was then found. They wrote that "it is easy to see that the first and second properties correspond to the first term ..., whereas the third property corresponds to the second term of this equation." (p. 581). However, it is doubtful that the first term would ensure that the first of the three conditions listed above is met, because the effect of this term is to force the filtered image to be as close to the real image as possible, and that means making σ as small as possible at all points, even those from uniform intensity regions. In fact, with $\lambda = 0$, the energy function would be minimized by having $\sigma(x,y) = 0$ for all (x,y).

Another problem with this scheme is that the minimization of E over the image is complicated, and the outcome is somewhat dependent on the initial values of the filter scales. With some synthetic noisy images, Jeong & Kim's algorithm produced results which they admitted were worse than the single-scale algorithms. With a natural image, they claimed that the adaptive scale algorithm was able to "detect edges very remarkably" (p. 584), but no formal evaluation procedure was used.

1.8.5.5 Morrone, Navangione & Burr (1995)

Morrone *et al.* (1995) varied the operator scale with the aim of keeping the SNR in the filter output constant across the image. They derived an equation that described how the operator scale should vary with distance from the edge. Since this distance was not known *a priori*, they estimated it using a recursive strategy. The filter was applied to the first pixel at a preset scale. Then the distance between that pixel and the nearest edge was found using the filter output at that pixel and the maximum filter output in the vicinity. This estimated distance was then used to calculate the scale for the filter applied to the next pixel, and so on. They applied their scheme to noisy rectangular step edge signals

and showed that a low operator scale was selected in the vicinity of the edge, and a large operator scale was selected elsewhere.

1.8.5.6 Fleck (1992a)

Fleck (1992a) used very simple finite-difference operators to compute directional 1st and 3rd derivatives, and a Laplacian 2nd derivative. Operator scale was varied by varying the distance between the sampling points, rather than varying the *number* of sampling points. With blurred edges, the large-scale operators tended to respond more strongly than small-scale ones, because the difference between the intensity values increased with the distance between the sampling points. At each point in the image, the algorithm selects the smallest-scale operator whose response is statistically reliable. A reliable response is defined as one that is three times the standard deviation of the operator's response to noise (which, of course, depends on the camera being used in the implementation). After selecting the operators, a *detection map* is generated, marking out all regions where the 1st and 3rd derivatives have opposite signs, and a *location map* is generated from the responses of the Laplacian operators. Edges are marked at ZCs in the location map if the detector map indicates that the 1st and 3rd derivatives have opposite signs.

1.8.5.7 Elder & Zucker (1996, 1998)

Elder & Zucker (1996, 1998) presented an algorithm based on very similar principles to that of Fleck (1992a), but they used directional Gaussian derivative filters. Like Fleck, they selected, at each point, the smallest operator whose response was statistically reliable. Each scale has a criterion response level, which is the lowest response that can be reliably distinguished from zero. For Gaussian 1st derivative operators, the criterion response is inversely proportional to the square of the operator scale; for 2nd derivative operators, it is inversely proportional to the cube of the operator scale. The *minimum reliable scale* at each point is the scale of the smallest operator that is responding above its criterion.

At each point, the algorithm finds the gradient direction using the smallest reliable 1st derivative operator. Elder & Zucker argued that it would be wrong to use the peaks in the

1st derivative to locate edges. The use of the minimum reliable scale ensures that the 1st derivative is reliably non-zero, but the *peaks* within this region may not be reliable. A peak in the 1st derivative corresponds to a ZC in the 2nd derivative. Thus, a reliable peak in the 1st derivative corresponds to a point at which the 2nd derivative goes from being reliably positive to reliably negative. If ZCs are found in the 2nd derivative at the minimum reliable scale, then these will identify points at which the 2nd derivative reliably changes sign, indicating positions of reliable peaks in the 1st derivative, i.e. edges.

In summary, the smallest reliable 1st derivative operator is used at each point to find the gradient direction at that point. Then the smallest reliable 2nd derivative operator is applied at each point *in the gradient direction*. Edges are marked at ZCs in the output of the 2nd derivative operator. The blur of each edge is found from the distance between the extrema of the 2nd derivative either side of the ZC, after correcting for the blur introduced by the operator itself.

Elder & Zucker's algorithm produces impressive results on real images. It produces smooth, unbroken contours in response to blurred edges, such as shadows, because these edges have a low gradient, so only the large operators respond above their criterion. Fine details are not lost because they give rise to high gradients, so even the smallest operators respond above their criterion. A possible problem with this approach is that only one scale is selected at each position. If a blurred shadow fell across a region of fine detail, it might be beneficial to represent that region at two scales.

1.8.5.8 Marimont & Rubner (1998)

Marimont & Rubner (1998) presented an algorithm that was strongly influenced by Elder & Zucker's work. Their algorithm takes an edge to be a ZC in the 2nd derivative such that the product of the 1st and 3rd derivatives is negative. They used Lindeberg's (1990, 1993a) discrete operators with a range of scales to estimate 1st, 2nd and 3rd derivatives at each point. The algorithm uses the 2nd and 3rd derivatives to calculate the probability of an edge at each point, for each operator scale. The algorithm also computes a *confidence probability* at each point, which is a measure of the reliability of the 2nd and 3rd derivative responses that are used to compute the edge probability. At each point, the

minimum reliable scale is chosen on the basis of the confidence probability. An edge probability map is then generated, using the edge probabilities from the minimum reliable scale at each point.

1.8.5.9 Lindeberg (1996, 1998)

Lindeberg (1996, 1998) took a very familiar edge detection method and developed it in terms of scale-space theory. Let $L_v(x,y;\sigma)$, $L_w(x,y;\sigma)$, and $L_{ww}(x,y;\sigma)$ be the 1st, 2nd and 3rd derivatives in the gradient direction at point (x,y) and scale σ. Lindeberg proposed marking edges as peaks in the gradient, L_v, by finding points (x,y) such that $L_w(x,y;\sigma)=0$ and $L_{ww}(x,y;\sigma)<0$, as first suggested by Haralick (1984). The novel feature of the algorithm was the scale selection method. Large operators work better on blurred edges, whereas small operators work better on sharp edges. Lindeberg's solution was to select the operator scale to match the blur of the image at each point. He defined an *edge strength* measure, σL_v^2. At the steepest point of a Gaussian edge with scale σ_e, this edge strength measure would peak at operator scale $\sigma=\sigma_e$; selecting the scale that corresponded to a peak in edge strength would therefore give an operator scale equal to the edge scale. Edge elements were found as follows. First, peaks in gradient along the gradient direction were found at each scale. A contour in the image would give rise to a contour at a range of scales. Across scale space, the contours joined up to form an *edge surface in scale space*. Edge elements were marked at points on the edge surface that were also peaks in edge strength along the scale dimension. This method has the advantage that, as well as finding edges at an appropriate scale at each point, the operator scale selected for each edge gives an estimate of the edge's blur.

Since the square root function is monotonic, Lindeberg could have chosen $\sigma^{1/2}L_v$ as his edge strength measure – the position of the peak would have been the same. This is a normalized derivative very similar to that of Korn (1988), except that the scaling factor is $\sigma^{1/2}$, rather than $\sigma\sqrt{2\pi}$ [11].

---

[11] Further work on normalized derivatives and scale selection was described by Lindeberg (1993b).

Lineberg's algorithm has much in common with those of Korn (1988) and Zhang & Bergholm (1997). The two main differences are that he used a different normalized derivative, and the *detection* of edges occurs at the scale given by the peak in the normalized derivative. In contrast, the edges in the other two schemes are detected at a fixed scale: either the smallest scale (Korn, 1988) or the largest (Zhang & Bergholm, 1997). The ability to select the scale at which an edge is detected, depending on the local blur of the image, is a key advantage of Lindeberg's algorithm. This idea plays a fundamental role in the Gaussian derivative template model, presented in Chapter 2.

1.8.6 Filter combination algorithms

The opposite of filter selection is probably filter combination. A number of researchers have proposed algorithms for combining the outputs of different operators.

1.8.6.1 Rosenfeld (1970)

Rosenfeld (1970) was probably the first to present an edge detection algorithm using multiple scales. He noticed that the output of small DOB filters is noisy, whereas the output of large DOB filters gives a wide hump around the edge, rather than a thin line. He found that, if the outputs of different-sized filters were multiplied together at each point, the result was a fairly noise-free thin line along the edge. He explained this informally by reasoning that, for the product of responses to be high, all the filters would have to give a strong response, and that this would only happen at a "major" edge.

1.8.6.2 Watt & Morgan: MIRAGE

After Marr & Hildreth (1980), the next model of edge detection in humans was Watt & Morgan's MIRAGE algorithm. The final version of this model was described in Watt & Morgan (1985), an unusual paper which, with subheadings of "Exposition", "Development", "Recapitulation", and "Coda", is probably the only scientific paper to have been structured explicitly in the form of a classical sonata. The key idea in the

MIRAGE algorithm is that information about spatial scale is lost before edges have been detected. Processing occurs as follows. The retinal image is filtered with a set of Gaussian 2nd-derivative filters with scales ranging from 0.35' to 2.8'. Each filtered signal is separated into positive and negative responses. All the positive responses are added together across scale, as are all the negative responses. This leads to two signals: the sum of the positive responses and the sum of the negative responses. The algorithm then identifies zero-bounded response distributions (RESPs) in the two signals and parses them according to a simple set of rules. A RESP with another RESP on one side only is interpreted as an edge. A RESP with another RESP on both sides or neither side is interpreted as a bar. The polarities of the edges and bars depend on the signs of the RESPs. The perceived position of an edge is determined by the position of the RESPs (or possibly just the positive one – Watt & Morgan, 1983, p. 1474). The blur of the edge is determined from the distance between the RESPs. There are several ways of defining the "location" of a RESP: one could use the centroid, the peak, or the ZC between adjacent RESPs. Using a computer simulation, Watt & Morgan (1985) showed that the centroid was the most reliable measure.

The separation of positive and negative responses has substantial physiological support. The dendrites of on- and off-centre retinal ganglion cells lie in separate layers of the inner plexiform layer of the retina (Famiglietti & Kolb, 1976; Nelson *et al.*, 1978) and the on and off channels remain functionally segregated at least until the LGN (Schiller, 1984).

Watt & Morgan presented most of their psychophysical evidence for MIRAGE in three papers (Watt & Morgan, 1983, 1984, 1985). Watt & Morgan (1983) concentrated on the representation of edge blur and location. In their first experiment, they measured blur discrimination thresholds as a function of base blur, for three different types of edge profile. The results showed dipper functions, similar to those seen in studies of contrast discrimination (e.g., Legge & Foley, 1980): as base blur increased from 0 (i.e. a step edge), the blur difference threshold first decreased, and then began a monotonic increase with further increases in base blur. When blur difference threshold was expressed as a function of the distance between the peak and trough in the 2nd derivative, the three different types of edge gave almost identical functions, supporting Watt & Morgan's view that perceived blur is determined by the distance between the edge's two RESPs. The

monotonically increasing portion of these functions described a power law with exponent 1.5.

Watt & Morgan (1983) also measured location discrimination as a function of edge scale and contrast. They found that the location discrimination threshold was proportional to $\sqrt{W/C}$, where C is the contrast, and W is the width of the positive or negative half of the second derivative at half peak amplitude (which is also proportional to the scale of a Gaussian edge). Watt & Morgan (1985) presented the results of simulations that showed that MIRAGE predicted this relationship between contrast and location discrimination threshold, as long as the signal-to-noise ratio was not too low.

A further experiment reported by Watt & Morgan (1983) measured edge separation discrimination as a function of spatial separation, and found a Weber's law relationship, so that separation difference threshold was proportional to edge separation.

The four main findings of Watt & Morgan's (1983) paper can be summarized as follows:

1. blur discrimination threshold is proportional to edge scale raised to the power of 1.5
2. location discrimination threshold is inversely proportional to the square root of edge contrast;
3. location discrimination threshold is proportional to the square root of edge scale;
4. separation discrimination threshold is proportional to edge separation.

Watt & Morgan (1983) argued that the first finding can be explained by findings 3 and 4. They argued that blur discrimination involves locating the RESPs and then determining their spatial separation, which will be proportional to the edge scale. Therefore, there are two sources of error, both functions of the edge scale: the location discrimination error, which follows a power law with exponent 0.5, and spatial separation discrimination error, which follows a power law with exponent 1 (Weber's law). Watt & Morgan (1983) noted that "adding the exponents (0.5 and 1.0) together gives an expected power law with an exponent of 1.5, thereby modelling the results of Experiment 1 rather well" (pp. 1474-1475). But it is incorrect to simply add the exponents of the two sources of error. Each

threshold in Watt & Morgan's experiments corresponded to the 83% correct point of the psychometric function, and can be assumed to be the standard deviation of the perceived value of each stimulus. When two sources of error combine linearly, the variances add, so the standard deviations (and hence thresholds) should add in Pythagorean sum. From Watt & Morgan's (1983) data, if σ_e is the edge scale, then location threshold is given by $a\sigma_e^{0.5}$, and separation threshold is given by $b\sigma$, where *a* and *b* are constants. Thus, the blur threshold should be given by $\sqrt{a^2\sigma + b^2\sigma^2}$, which approximates a power law with exponent 0.5 for small blurs and 1 for high blurs. In fact, Hess *et al.* (1989) and Wuerger *et al.* (2001) found that blur discrimination did indeed follow Weber's law for high blurs. The discrepancy between these later findings and Watt & Morgan's (1983) data may have arisen because Watt & Morgan used very short edges, which were only 12' long. Georgeson & Freeman (1994) showed that shortening an edge made it look sharper. The sharpening effect is determined by the edge length as a proportion of edge scale. A fixed edge length will therefore sharpen large-scale edges more than small-scale ones. As edge scale increases from zero, the internal blur code will initially rise at the normal rate, but will start to rise more slowly than normal if the edge scale continues to increase, and the sharpening effect takes hold. Thus, with quite blurred edges, a larger-than-normal increase in physical scale would be required to give rise to a just-noticeable difference in the internal blur code, while sharper edges would be unaffected. This would increase the effective exponent in the blur discrimination function. Watt & Morgan (1983) performed a pilot experiment to check that the location threshold was unaffected by the short edge length, but they did not carry out a similar experiment to check that blur threshold was also unaffected. In summary, the apparent consistency between Watt & Morgan's data and theoretical expectations seems to have arisen from a fortuitous coincidence of a theoretical error and a methodological flaw. Correcting the theoretical error shows that MIRAGE correctly predicts the data of Hess *et al.* (1989) and Wuerger *et al.* (2001), collected in better conditions. It is hard to understand why Watt & Morgan thought that the exponents of the two sources of error should add. In their next paper, they gave the correct analysis, similar to that given above, which predicts Weber's law for blur discrimination at high blur levels (Watt & Morgan, 1984, p. 1392, equations 2 and 3).

Watt & Morgan's (1984) paper concentrated on the combination of information across spatial scale. The key piece of evidence for the MIRAGE filter combination process was

an experiment measuring the Vernier acuity for relative location of two Gaussian edges, positioned one above the other. The upper edge had a high-frequency cosine wave mask added to it. In one set of conditions, the mask had a Gaussian window; in the other set, the mask window was the complement of the Gaussian, so the grating only appeared to the left and right, but not in the middle. Watt & Morgan argued that a low-frequency filter would be optimal to perform this task, as it would not be sensitive to the mask. They showed that the response profile of the filter would be the same shape with each type of window, predicting that, if the filter outputs were not combined, the type of mask would not make any difference. In fact, they found quite different results with the two types of mask. Gaussian-windowed mask had its greatest effect when its scale was equal to the edge scale, whereas the mask with a Gaussian complement window showed the lowest masking around this value. They presented some modelling data showing that MIRAGE could predict that the Gaussian-windowed mask would reduce Vernier acuity, while the complement mask increased it, but they did not address the issue of whether this explanation could account for the whole shape of the masking functions across mask size: the effect of the Gaussian-windowed mask was quite sharply tuned, so that a substantial masking effect only occurred when the Gaussian mask window had a similar scale to the edge.

MIRAGE explains a number of visual illusions. In the Chevreul illusion, the luminance increases in discrete steps, with uniform luminance between each edge, but the region of uniform luminance appears non-uniform. MIRAGE explains this by marking bars of opposite polarity at each end of each uniform region. MIRAGE also explains the existence of Mach bands, which are illusory bars seen where a ramp meets a plateau. MIRAGE produces an isolated RESP at the location of the Mach band, which is interpreted as a bar. There seems to be little computational justification for interpreting an isolated RESP as a bar, since a real bar gives rise to three adjacent RESPs, but this rule allows the model to predict the appearance Mach bands. With both of these illusions, MIRAGE fails to capture all the perceived structure. For example, Fiorentini *et al.* (1990) pointed out that, although MIRAGE detects the Mach bands at each side of a ramp, it fails to mark an edge on the ramp itself. They showed that this deficiency could be remedied by applying MIRAGE's parsing rules to each scale separately. Similarly, depending on the spatial scale, MIRAGE can detect either the bars seen in the Chevreul illusion, or the edges, but not both (Kingdom & Moulden, 1992). Kingdom & Moulden presented a

model, called MIDAAS, which had parsing rules similar to those of MIRAGE, but applied them to each spatial scale separately. Kingdom & Moulden showed that, if the outputs of each channel were averaged together *after* symbolic processing, rather than before, the model correctly predicted the perceived luminance profile. Thus, it seems that the most distinctive feature of MIRAGE, the early inflexible summing of information across channels after rectifying, is its greatest weakness.

Although the summing across scale may not occur in quite the way that Watt & Morgan had envisaged, they appear to be correct in their proposal that visual perception does not have direct access to the low-level visual filters. Olzak & Thomas (1991, 1992, 1999) have presented a substantial amount of evidence that information is combined across channels. Specifically, they argued that information about orientation is carried by mechanisms that sum across spatial frequency within an orientation channel; information about contrast and spatial frequency is carried by mechanisms that sum across orientation within a spatial frequency channel.

1.8.6.3 Georgeson & Meese: filter combination

Another substantial body of evidence for filter combination was presented by Georgeson & Meese. Georgeson (1992) presented naive subjects with plaid pattern formed from the sum of two or more sine wave gratings, and asked them to judge the edge structure. The perceived edge structure matched the pattern of ZCs predicted by Marr & Hildreth's (1980) theory of edge detection. As described in section 1.8.1, Georgeson then showed that the perceived edge structure could be distorted by adaptation to or masking by an oriented grating, the effect being perceptually equivalent to reducing the contrast of the plaid component aligned with the adapting/masking grating. This behaviour would not be expected from isotropic filters. Since the unadapted/unmasked edge structure matched that predicted by ZCs in a LoG filter, Georgeson proposed that a LoG filter was constructed by adding the outputs of a set of oriented operators whose orientations spanned 360°. This scheme would be more flexible than a hard-wired LoG operator because, if the stimulus conditions demanded, the combination across orientation could be broken, so that filters were constructed dynamically to suit the characteristics of the image,

an idea first suggested by Canny (1986), who allowed filters to combine across space, where appropriate.

Georgeson & Meese then set out to discover the mechanisms that controlled filter combination. They performed psychophysical experiments, using plaid patterns, in which subjects had to judge whether the edges were oriented along the ZCs in the LoG filter output (filter combination) or along the individual plaid components (filter segregation). Filter combination was found to be encouraged by increasing component contrast (Meese & Freeman, 1995; Meese & Georgeson, 1996a) and by decreasing the angle between the components (Georgeson & Meese, 1997). Combination is discouraged by adapting to a grating that bisects the angle between the components of the test plaid (Meese & Georgeson, 1996b), by adapting to a plaid with components that bisect the angles between the components of the test plaid (Meese & Georgeson, 1996a), or by adding a third harmonic in square-wave phase with one of the plaid components (the *switching effect* - Georgeson & Meese, 1997, 1999). Most of these results are readily explained by proposing that filters of different orientation will combine if and only if filters of intermediate orientation are sufficiently active (the *bridge channel hypothesis*). To use this explanation to account for the switching effect it is necessary to postulate inhibitory connections that allow filters of one orientation to inhibit a "bridge" filter tuned to a third of the inhibiting filter's spatial frequency, and oriented at 45° to the inhibiting filter. In support of this, Meese & Georgeson (1996a) found that filter combination was encouraged by adapting to a plaid with components aligned with the test plaid, but three times the spatial frequency. All of these findings were incorporated into a model of edge detection in human vision, presented by Georgeson (1998).

1.8.7 Phase congruency algorithms

1.8.7.1 Energy models

Morrone & Owens (1987) proposed an approach to feature detection that was very different to everything that had gone before. They noted that, at a step edge, all the Fourier components of the image have a phase of 0 while, at a line, or bar, all the Fourier components have a phase of $\pm\pi/2$. They used the term *phase congruency* to refer to

the similarity of phase across components, and proposed that features are perceived at peaks of phase congruency. The central tendency of the phase values at the peak indicates the type of feature: a phase of 0 indicates an edge, a phase of $\pm\pi/2$ indicates a line, and any other value indicates a combination of edge and line.

The algorithm for finding peaks of phase congruency runs as follows. Let F be the image, and H its Hilbert transform (found by phase shifting all the Fourier components by $\pi/2$). The energy, E, is found by adding the squares of F and H at each point:

$$E(x) = F^2(x) + H^2(x).$$

Morrone & Owens proved that peaks of the energy function coincide with peaks of phase congruency. In most subsequent presentations of the energy model (except Venkatesh & Owens, 1990), energy has been defined as the square root of the term used by Morrone & Owens:

$$E(x) = \sqrt{F^2(x) + H^2(x)}.$$

This variation in definitions is not important, since the square root is a monotonic function, so both versions of the energy function will peak in the same place. Using the second definition of energy, if $F(x)$ and $H(x)$ are plotted on the horizontal and vertical axes, respectively, of a Cartesian coordinate system, then the length of the vector $(F(x), H(x))^T$ is the energy, $E(x)$, and its direction, $\text{atan2}(F(x), H(x))$, gives the central tendency of the phase. As explained above, the phase indicates the type of feature(s) present. See Venkatesh & Owens (1990) for more details of feature classification.

In practice, it is computationally expensive to find the Hilbert transform of the image, so Morrone & Owens proposed that F and H should be the results of filtering the image with, respectively, even and odd filters that have the same amplitude spectrum, i.e. *quadrature pairs* of filters. Then H will still be the Hilbert transform of F, and, if the filters have a

sufficient bandwidth, F will be similar to the image. This approach was inspired by Adelson & Bergen's (1985) motion energy model.

Morrone & Owens' algorithm was intended for machine vision, but the idea had arisen from psychophysical work on the perception of Mach bands by humans (Morrone *et al.*, 1986). Mach bands are seen on trapezoidal waves, where the ramp meets the plateau. Morrone *et al.* (1986) noted that the Fourier components of the trapezoidal waves had maximal phase congruency at the points where the Mach bands were seen. These peaks of phase congruency depended critically on some high-frequency, phase-reversed, low-contrast components. If those components were removed, the peaks of phase congruency would shift from the Mach band locations to the mid-points of the ramps. They found that, as predicted by the energy model, when the stimulus contrast was reduced, the Mach bands disappeared at a contrast that corresponded to the detection threshold of the phase-reversed harmonics. This finding was confirmed by Ross *et al.* (1989), who also found the same effect when the stimuli were masked with a sine-wave grating. Ross *et al.* also confirmed Mach's finding that Mach bands are not seen more than 3° into the periphery, where phase sensitivity is low (see Ratliff, 1965).

The results of Morrone *et al.* (1986) and *Ross et al.* (1989) suggested a role for phase congruency in feature detection in human vision. There is also evidence supporting the use of odd and even filters in quadrature phase in biological visual systems. Pollen & Ronner (1981) studied simple cell recordings in which action potentials from two distinct cells could be identified. They found that adjacent cells tended to differ in phase by 90°, which would simplify the implementation of the energy computation. Burr *et al.* (1989) studied the discrimination of phase-congruent patterns and found that, when the pattern was in sine or cosine phase, a pedestal[12] with phase orthogonal to the main pattern did not facilitate discrimination, but facilitation did occur with an orthogonal pedestal when the pattern had a phase intermediate between sine and cosine. This suggests that the pattern discrimination was performed by filters in sine or cosine phase: the orthogonal pedestal could only have an effect when the pattern was activating filters responding to both the pattern and the pedestal.

---

[12] A pedestal is another stimulus, added to the main stimulus.

Morrone & Burr (1988) presented a version of the model intended as a model of feature detection in humans. This model was very similar to that of Morrone & Owens (1987), except that the single pair of broadband filters was replaced by a set of psychophysically plausible band-pass filters. Morrone & Burr (1988) did not commit themselves on how the information was combined across spatial frequency channels.

The models presented by Morrone & Owens (1987) and Morrone & Burr (1988) were limited to one spatial dimension. Morrone & Burr (1993) extended the model to two dimensions. Images were filtered with quadrature pairs of filters selective for orientation and spatial scale, and a feature map was found separately at each spatial scale. Morrone & Burr (1993) described the energy peak detection in 2D as follows: "features are marked by maxima in the local energy functions along the orientation where energy has maximum gradient" (pp. 49-50). It is not clear what this means, since the energy gradient will always be zero at the maxima. Having found the peaks at each scale, Morrone & Burr again did not commit themselves as to whether the feature maps should remain separate or be combined. See Burr & Morrone (1990, 1992) for further discussion of the energy models described so far.

Kisworo *et al.* (1994) described a model which made initial estimates of edge location using the standard energy model, but then made a more detailed estimate of the edge profile and position to sub-pixel accuracy by fitting a template which was sufficiently flexible to serve as a model for step edges, ramps, and roof edges. Unlike conventional matched-filter template models, the fitting was achieved by performing the energy computation on the fitted edge model and adjusting the parameters until the result of the energy computation best matched the energy computation performed on the image.

1.8.7.2 Kovesi (2000)

Although features are defined in terms of phase congruency across Fourier components, Morrone's various energy models do not require a Fourier analysis to be performed: using a single pair of broadband filters, the energy peaks coincide with the peaks of phase congruency. Kovesi (2000) presented a phase congruency model in which a Fourier analysis was performed explicitly, using log Gabor filters. He argued that Morrone's local

energy function does not properly measure phase congruency because it is affected by changes in image luminance, contrast, and size: phase congruency should be a dimensionless value between 0 and 1, that is invariant with respect to these changes. His measure of phase congruency can be understood by considering the fact that, if $(F(x),H(x))^T$ is plotted as a vector, then its length gives the energy, $E(x)$, and its direction, $\text{atan2}(F(x),H(x))$, gives the phase. This vector will be the vector sum of the corresponding vectors of the individual Fourier components. If the Fourier components all have the same phase, then their vectors will all have the same direction, and the sum of their lengths will equal the energy, $E(x)$. If the Fourier components differ in phase, then the vectors representing the Fourier components will have different directions, and in order for their vector sum to equal $(F(x),H(x))^T$, the sum of their lengths must exceed the energy. Kovesi proposed that phase congruency should be measured by dividing $E(x)$ by the sum of the lengths of the Fourier components, giving a maximum value of 1 when the Fourier components all have the same phase. This method amplifies noise considerably, since even low amplitude noise can have high phase congruency, so a threshold based on an estimate of noise has to be applied to the local energy function.

1.8.7.3 Peli (2002)

Peli (2002) presented a phase congruency model that was very different from the energy models described above. The image was first filtered with a set of physiologically plausible band-pass filters that differed in spatial frequency. These filters could be considered to be in cosine phase, since they only affected the amplitude of the signal (rather than its phase). Let F_i be the image filtered by filter i. An edge-map was created, in which pixels were black, white or grey. A pixel (x,y) in the edge map was white if, for all i, $F_i(x,y) \geq 0$, and black if, for all i, $F_i(x,y) < 0$; otherwise it was grey. Thus, the algorithm marked pixels as black or white if the image exhibited some kind of phase congruency across scales at that point. A line in the image gave rise to a black or white line in the edge map (the colour of the l[illegible]th [illegible]map indicating the polarity of the line in the image). An edge in the im[illegible]djacent black and white lines in the edge map (the

direction of intensity increase in the image indicted by direction of intensity increase in the edge map).

Peli found that the basic algorithm suffered from noise, so he introduced a threshold after the filtering stage. Each filter, i, had a threshold, T_i, based on psychophysical measurements of contrast sensitivity for that filter's peak frequency. A pixel (x, y) in the edge map was white if, for all i, $F_i(x, y) \geq T_i$, and black if, for all i, $F_i(x, y) \leq -T_i$; otherwise it was grey. This led to a substantial improvement in the edge map.

As well as having filters and thresholds based on psychophysical and physiological data, the algorithm captures the separation of responses into on-centre and off-centre responses, found in physiological studies of the early visual system (Famiglietti & Kolb, 1976; Nelson *et al.*, 1978; Schiller, 1984).

Although Peli's model shows reasonable performance with natural images, it has serious problems with the most common of psychophysical stimuli, the sine wave grating. Since the filters all have cosine phase, the response of each filter to a sine wave grating would be identical in every way apart from the amplitude: each filter would respond with a sine wave grating with the same phase and spatial frequency as the stimulus. Thus, every pixel would have the same sign across all filters, and the model would mark an edge at every point. As shown in section 1.10, the other phase congruency algorithms also have problems with sine wave gratings but, rather than marking too many edges, they mark too few.

1.9 Edge characterization

Section 1.8, and indeed most research on edge detection, has focussed on the problem of detecting and locating edges. Many algorithms do no more than this. However, edges have a number of other parameters: orientation, contrast, and blur/width. Orientation is useful in contour integration (e.g. Robinson, 1977; Nevatia & Babu, 1980). The contrast of an edge contains important information about the nature of the physical source of the

edge feature in the image. The usefulness of blur is not quite so obvious. Ignoring gradual changes in surface pigmentation, blurred edges arise mainly from three situations:

1. shadows;
2. gently curved surfaces, where the surface shading changes gradually across the surface;
3. focal blur, when the eye/camera is focusing on a depth plane different from that of the edge.

Thus, if an edge is sharp, it probably corresponds to an abrupt change in the geometry of a real surface (as opposed to a shadow), that lies in the depth plane that the eye is currently focusing on; if the edge is blurred, it probably corresponds to a shadow, a gently curved surface, or an abrupt change in a surface that lies in a depth plane different to the one currently in focus. Thus, the ability to classify an edge as blurred or sharp can help to narrow down the range of possible interpretations. In the case of a blurred edge, the local edge blur alone is not sufficient to tell which of the three situations has arisen, since all three can give rise to locally indistinguishable edge profiles (Elder & Zucker, 1998). However, if the physical source of the blur could be established using the edge blur in conjunction with other information, then the blur level can help to fine-tune the interpretation. If the source of the blur is a shadow, the blur level indicates the distance between the object casting the shadow and the surface onto which the shadow is cast. If the source of the blur is a curved surface, the blur level indicates the radius of curvature, and could be used to determine 3D shape from shading. If the source of the blur is optical defocus, then the blur level blur indicates the difference in depth between the edge and the plane currently in focus. In addition, the blur of the boundary between two textured surfaces that differ in blur can indicate which surface is nearer. There is evidence that these depth cues are used by humans (Marshall *et al.*, 1996; Mather, 1996, 1997; Mather & Smith, 2002).

This section summarizes the ways that have been found to characterize edges in terms of orientation, contrast and blur. Orientation is straightforward, and is usually measured by finding the direction of steepest gradient. The compass gradient operators, described in section 1.2.2, found the orientation by applying eight templates of different orientation and finding which one fitted the best; however, since template matching is equivalent to

differentiation for these operators, these algorithms were effectively estimating the direction of steepest gradient, to the nearest 45°. Contrast is less straightforward, because several different definitions of contrast are used, the most common ones being Michelson contrast (Michelson, 1927) and step height. The different definitions of contrast are all linearly related to each other, though, so the multiplicity of definitions is not too troublesome. The most difficult parameter to define is edge blur, because it cannot be defined without making an assumption about the form that the edge takes. For any edge model that has definite end-points (such as a linear ramp, or a half-period of a sine wave), the natural definition of physical blur is the width over which the intensity change extends. But some widely-used edge models, such as the Gaussian integral, do not have any end-points. The Gaussian edge is a particularly attractive edge model because it has mathematically convenient properties, and the blur in imaging systems can often be modelled as convolution with a Gaussian kernel, in which case a rectangular step edge will be transformed into a Gaussian integral. For the Gaussian edge, the natural definition of its blur is the scale (or standard deviation) of the Gaussian of which it is an integral.

Marr & Hildreth (1980) defined the amplitude of a ZC segment as the slope of the filter response as it crosses zero. Hildreth (1983) noted that the slope was a function of edge contrast, blur and filter scale. The filter scale is known so, if there are at least two scale channels, then there are two simultaneous equations that can be solved for the two unknowns, edge contrast and blur. Chen & Medioni's (1989) algorithm (described in section 1.8.3.3) was similar to this, in that it took measurements at two different scales and solved two simultaneous equations to give step height and edge blur. In this case, the measurement taken was, not slope, but reconstructed step height, a measurement which was invariant with filter scale for sharp edges, but varied with filter scale for blurred edges.

Sjöberg & Bergholm's (1988) measure of edge blur was based on the observation that, with Bergholm's (1987) edge focusing algorithm, a very blurred edge would break up into a cloud of edge elements. This allowed them to perform a very crude classification of edges as blurred or sharp.

One approach to blur coding uses normalized Gaussian derivative filters, and measures an edge's parameters using the edge's signature as it evolves over filter scale. Korn (1988)

and Lindeberg (1996, 1998) used different types of normalized derivatives and measured the edge blur from the scale of the filter with the strongest response; in both cases they used the Gaussian integral as their edge model. As described in section 1.8.4.1, there are problems with Korn's approach because, using his normalized derivatives, an isolated Gaussian edge will not give rise to a peak in operator response over scale. Zhang & Bergholm (1997) used the whole of the function mapping operator scale to response (termed the Gaussian difference signature), rather than just the peak. Edge blur and contrast were found from the Gaussian difference signature template that fitted best to the obtained signature. Williams & Shah (1993) showed that, using Korn's normalized derivative, the width of a linear ramp could be found from the scale at which the filter response stopped rising linearly with operator scale. With increasing filter scale, the filter response approached an asymptote that was the edge height.

So far, the discussion has focussed on how blur could be coded in any unspecified visual system, but not how it is coded in the human visual system. Watt & Morgan's MIRAGE algorithm was the first serious model of edge blur coding in humans. Their measure of blur could not be based on scale space properties, because MIRAGE throws away information about spatial scale at an early stage. Watt & Morgan (1983) proposed that the edge blur was determined from the distance between the peak and trough either side of a ZC in the 2nd derivative of the stimulus. In support of this, they measured blur discrimination thresholds as a function of base blur for three different types of edge profiles, and found that, when blur was expressed as the peak-trough distance, the blur discrimination functions for the three edge types were superimposed. Watt & Morgan (1985) revised their model slightly, and proposed that blur should be determined from the distance between the *centroids* of the two RESPs corresponding to an edge, rather than the peaks, because the centroid-based measure was more reliable. They also proposed that the edge contrast could be found from the mass of the RESPs, after taking edge blur into account. Elder & Zucker (1996, 1998) proposed an almost identical measure of blur to Watt & Morgan, proposing that the blur should be found by measuring the peak-trough distance either side of a ZC, and then correcting for the blur introduced by the filter itself. Because of this correction, Elder & Zucker's algorithm provides a veridical estimate of blur for Gaussian edges, unlike Watt & Morgan's (1985) algorithm.

Georgeson (1994) proposed a derivative-based model of edge coding in which edges were seen at peaks in the 1st derivative and bars at peaks in the 2nd derivative. He showed that the scale of a Gaussian edge is equal to the square root of the ratio of 1st to 3rd derivative magnitudes at the edge location. This model very accurately predicted the results of an experiment in which subjects had to compare the perceived blurs of Gaussian and sine wave edges. The distance between peak and trough in the 2nd derivative, proposed by Watt & Morgan (1983) and Elder & Zucker (1996, 1998), underestimated the perceived blur of sine wave edges. Georgeson argued that contrast could be obtained by multiplying the edge scale by the ratio of 1st derivative to luminance at the edge location. He suggested that the division by the luminance could be carried out by divisive retinal gain controls.

1.10 Edge detection and localization by humans

One might suppose that a critical test that all models of human feature detection must pass is that they should mark features at the positions that humans see them, and nowhere else. However, Watt argued that MIRAGE should not have to sit this particular exam:

> we never supposed that our account was intended to mark the location of a feature in a direct and completely linear geometrical mapping of the image. We supposed that our model was providing the locations in some sample version of the image that could be used to create spatial relations between different features — that's very different. We weren't concerned with having a transformation of the image that produced a discrete peak bang on where subjects saw an edge. We were quite happy to go for something that found response either side and then inferred, subsequently, that the edge lay in between. It seems to me that one could create a very strong distinction between two classes of model: those that are firmly rooted in mappings that are strictly within the image and produce edge maps, and those that are in some sense far more abstract (which is where I would put ours). (Discussion of Burr & Morrone, 1994, p. 143).

But Watt's argument is invalid, because any model that created even an abstract description of the "spatial relations between different features" should be able to perform a feature marking task perfectly well. In the feature marking tasks reported by Morrone &

Burr (1988) and Georgeson & Freeman (1997), the subject sees the stimulus with a cursor against it, and has to adjust the position of either the cursor or the stimulus so that the cursor is aligned with a feature. All that is required is that the subject can tell which side of the feature the cursor lies, i.e. whether the cursor lies between the feature and the left side of the screen, or between the feature and the right side of the screen. It is not necessary to have an internal representation that is a "completely linear geometric mapping of the image". If the representations of the spatial relations between features are so abstract that the model cannot even offer an opinion about whether one feature is on the left or right of another, then it is doubtful that such a representation could be of much value. Any useable representation of the spatial relations between features should make these relations explicit enough to allow the cursor to be adjusted so that it was aligned with the feature. Therefore, even MIRAGE must correctly predict the results of feature marking experiments in order to be considered a valid model of human vision. Unfortunately, there appears to have been no research into the accuracy with which MIRAGE can locate features.

Morrone & Burr seem to consider feature marking to be a valid test, since one of the key experiments cited in support of their model uses this paradigm (Morrone & Burr, 1988, experiment 1). A stimulus which raises serious questions about their model is the sine wave grating. The energy model finds no features at all in a sine wave grating, because the energy is constant everywhere. The odd filter generates a sine wave with the same phase as the stimulus, and the even filter generates a cosine wave with the same phase. Thus, the energy at each point is $a\sin^2 x + a\cos^2 x = a$. This would seem to disqualify the energy model from being a serious model of feature detection in humans, because the most obvious description of a sine wave grating is that it consists of a series of alternating dark and light bars with blurred edges in between. Morrone & Owens (1987) were well aware of this problem, but dismissed it with the remarkable claim that, in a sine wave grating, "no feature is perceived, only a smooth graduation of luminance" (legend of their Figure 1, p. 304). In particular, they claimed that "no edge is perceived in a sine-wave grating at the point where the luminance crosses its mean value, that is, the point at which the gradient is maximum." (p. 303). No empirical data were provided to support this claim. Subsequently, Georgeson (1998) conducted a feature marking experiment that showed that naive subjects do in fact mark edges at the gradient peaks in a sine wave

grating. One might argue that Georgeson's subjects may have been biased by his instructions in some way but, even if we ignore Georgeson's feature marking data, there is a further difficulty for the energy model. Georgeson (1994) noted that, if a half-cycle sine wave edge is placed between two regions of uniform luminance, the energy model does mark an edge at the gradient peak. Within the half-cycle, this edge has exactly the same luminance profile as the edges in a patch of sine wave grating with the same spatial frequency. It seems highly unsatisfactory that the half-cycle sine wave edge should be detected when isolated, but not when flanked by other similar features. Georgeson (1994) showed that humans are able to make blur judgments about both types of edge, perceiving them to be equally blurred.

The failure of a model to detect an edge that is perceived by humans in one particular kind of stimulus should not lead to the immediate rejection of the model. After all, it might be that a minor modification would remedy the problem. However, the energy model's inability to detect features in sine wave gratings is an intrinsic characteristic of the energy computation: indeed, Adelson & Bergen (1985) introduced the energy concept to vision in order to account for the unmodulated outputs of complex cells in response to drifting sine wave gratings. If the energy model could be modified in some way to detect sine wave gratings, then it would no longer be an energy model. One could argue that, if the retina applied a nonlinear transducer to the image, then the sine wave grating would be distorted from a true sine wave, and features might become visible. But, as pointed out by Meese (1999), the energy model would then predict that, if the inverse of the transducer function were applied to the sine wave grating, the perceived edges and bars would disappear. The nonlinear functions that have been proposed as models of retinal transduction are all monotonic functions (e.g., Naka & Rushton, 1966), and so they are all invertible, the inverse in each case being another monotonic function. It seems highly unlikely that applying any monotonic function to a sine wave grating would yield a stimulus with no perceived bars or edges.

Georgeson (1994) presented evidence that the energy model's problem with the sine wave grating is actually a general problem with spectrally simple stimuli. He constructed stimuli from two sine wave gratings with spatial frequencies in the ratio 3:1, and found that subjects marked edges and bars at extrema in the 1st and 2nd derivatives, respectively, with a slight distortion that could be explained by an early nonlinear transducer. When the

grating components were in square-wave phase, the subjects saw six bars and six edges per cycle, as predicted by the derivative model, whereas the energy model only predicted two features per cycle. This, and other failures of the energy model with low-frequency stimuli, led Burr & Morrone (1994) to suggest that low frequencies are processed in a different way from high frequencies. However, the failure of the energy model to account for Georgeson's (1994) data are due to the spectral simplicity of the stimuli, not their frequency level. It would be interesting to repeat these experiments with high-frequency versions of Georgeson's (1994) stimuli.

Georgeson (1994) described as "tentative" his suggestion that bars were marked at extrema in the 2nd derivative. The more straightforward model, of marking bars at extrema in the luminance function, is ruled out by the appearance of Mach bands, which appear where a ramp meets a plateau. The 2nd derivative has extrema at these features, but it also has extrema either side of a Gaussian edge, and informal observations suggest that Mach bands are not perceived at these locations. Georgeson & Freeman (1997) investigated the nature of the "edge" and "bar" filters more closely. Edges have odd symmetry and bars have even symmetry; for a peak to occur in the filter output, the filter must have the same symmetry as the feature being detected. Georgeson & Freeman proposed a model in which edges were detected by an odd filter and bars were detected by an even filter, and the slope of the spatial frequency spectrum of each filter (on a log-log plot) was a free parameter. The proposal that edges and bars are detected by peaks in the 1st and 2nd derivatives, respectively, predicts that the slope of the frequency spectra will be 1 and 2, respectively (see equation (1.14)). Georgeson & Freeman performed feature-marking experiments with stimuli similar to those used by Georgeson (1994), and found that the data were best fitted with filter slopes of 0.5 for both filters. This was not incompatible with the idea that the filters had the spectrum of a Gaussian 1st derivative filter because, although a perfect 1st derivative filter has a slope of 1, with a Gaussian derivative filter, the slope will obviously decrease around the peak of the tuning curve. Georgeson & Freeman calculated that, if the peak of the filter's tuning curve had matched the higher of the two spatial frequency components in the stimuli, the slope of the curve would effectively have been about 0.5 in these experiments. They proposed that edges and bars are detected by odd and even filters with the same amplitude spectrum, i.e. in quadrature phase. The linear filtering stage in this model is therefore the same as that in the energy models, proposed by Morrone and colleagues but, instead of squaring and

summing the two sets of responses, the model simply finds extrema in each set. The data from Pollen & Ronner (1981) and Burr *et al.* (1989), supporting the use of pairs of filters in quadrature phase, are perfectly consistent with Georgeson & Freeman's model, and so do not offer exclusive support for the energy model.

Morrone & Burr (1988) presented some evidence that appeared to support the energy model over the derivative-based models. They constructed stimuli that had the amplitude spectrum of a low-pass filtered square wave, but with different phase values at the point of phase congruency (always the origin). These stimuli were much like the (f + 3f) gratings used by Georgeson (1994). It was noticed that, with prolonged viewing, some features showed perceptual instability, varying between an edge and a bar. This was particularly noticeable for phases close to $\pi/4$, indicating an edge/bar combination. They asked subjects to wait until a feature was clearly an edge or bar, and then to report its identity and position. They found that, as predicted by the energy model, all features were marked at the point of phase congruency, whether an edge or a bar. They interpreted the perceptual instability as rivalry between the two interpretations (edge or bar), analogous to the phenomenon of *monocular rivalry*, reported by Atkinson & Campbell (1974). Since the derivative-based models predicted a shift in perceived position, Morrone & Burr claimed that this evidence supported the energy model in preference to others. However, Georgeson (1984) had already shown that monocular rivalry could be explained by a combination of eye-movements and afterimages, and Georgeson & Freeman (1997) presented a similar account of Morrone & Burr's results: after an eye movement, the afterimage would add to the stimulus, changing the relative phases of the components of the effective image. Morrone & Burr's subjects waited until the percept was a clear bar or edge; these events probably occurred when the effective phases were 0 and $\pi/2$, which would give rise to unambiguous edges and bars, respectively, at the origin.

Morrone & Burr (1988) created a family of 1D stimuli that had the amplitude spectrum of a square wave grating, but the phase value at the point of phase congruency (the origin) could be varied. When all the components where in sine phase, the percept at the origin was an edge; when the components were in cosine phase, the percept was a bar. This was predicted by the energy model, but most feature detection models would give similar results, since the points of phase congruency in these two stimuli corresponded to peaks

in gradient and luminance for the edge and bar stimuli, respectively. One might suggest that, logically, edges should be marked either side of a bar. In feature marking experiments, subjects must be given some instructions about what to look for, and reasonable instructions would be to mark edges at boundaries between regions of different brightness, and to mark bars at peaks of brightness. Given these definitions, a bar must logically be flanked by two edges. Morrone & Burr's energy model failed to mark edges either side of the bar. However, this is not a serious problem because, if a model fails to find flanking edges for every bar, it could easily be extended with a rule that marks edges either side of a bar if they have not been detected initially. Curiously, several proponents of energy models have argued that marking edges either side of a bar is not only unnecessary, but is actually undesirable. Owens *et al.* (1989) stated that "a desirable property of any edge detector is that it be a projection in the mathematical sense, that is, when it is applied to its own output it produces no further change" (p. 233). They argued that a line-drawing is similar to the feature map produced by the visual system, and that "when we look at line-drawings the visual system does not appear to produce a second feature-map, that is, a primal sketch consisting of features of the features" (p. 233). They showed that the energy model is indeed a projection, while gradient-based models, such as Marr & Hildreth's model produce two edges for every line, so that the edges proliferate every time the model is applied to its own output. Apart from their (psychophysically untested) assertion that lines do not appear to have edges, they provided no explanation of why it should be either necessary or useful for a feature detector to produce no further change when applied to its own output. A similar view was taken by Kovesi (2000, 2002), who created a phase-shifted square-wave stimulus in which the phase was varied along the vertical axis. This stimulus has an edge at the top, where the components line up in sine phase and, as the vertical position moves towards the bottom of the stimulus, the edge gradually mutates into a line. Kovesi showed that a Canny edge detector found a single edge at the top of the stimulus, and a double edge either side of the line at the bottom. Like Owens *et al.*, Kovesi considered this to be a problem for the Canny detector, and argued that "this example shows that points of maximum gradient do not necessarily correspond to the locations where we perceived features" (Kovesi, 2002, p. 823). But, as argued above, using intuitively reasonable perceptual definitions of edges and lines, flanking edges are a logical requirement, so the output of the Canny edge detector is exactly what one should expect. Neither Owens *et al.*, nor Kovesi, provided psychophysical data to support their assertions that lines are not perceived to have edges.

Barbieri-Hesse & Georgeson (2003) used Morrone & Burr's stimulus in a feature marking experiment and showed that subjects mark an edge when the components line up in sine phase, and mark a bar flanked by two edges when the components line up in cosine phase, exactly as predicted by gradient-based models.

In conclusion, derivative-based feature detection algorithms can explain the edge marking data, but energy models do not find enough features, especially in spectrally simple stimuli. It is hard to disagree with Georgeson & Freeman's (1997) conclusion that "this discrepancy between energy peaks and observed features is so large, and so general, that we must conclude that human vision does not use energy peaks to select feature locations" (p. 137).

1.11 The usefulness of edges

Very many papers on edge detection in both human and machine vision begin with some kind of assertion to the effect that edge detection is an obviously useful, or even necessary, first stage in image analysis. Rigorous arguments in support of such assertions have been much less plentiful. At the opposite extreme, Belhumeur & Kriegman (1998) have stated that "edges do not contain all of the information useful for recognition" (p. 1). They pointed out the success of recent object recognition schemes that do not use edge-based representations. There are two issues here. One is whether or not edge maps contain enough information; the other is whether edge maps present the information in a way that is useful to higher-level processing systems.

1.11.1 The information content of edge maps

Marr proposed that representations of edges in an image are contained in a *raw primal sketch* that is used by all higher processing stages (Marr, 1976, 1982; Marr & Hildreth, 1980). He proposed that the raw primal sketch contains primitives that represent the edges, bars and blobs in an image, and that these primitives should have attributes of

position, contrast, length, width and orientation. Marr conjectured that this primal sketch contained sufficient information to allow the image to be reconstructed completely (Marr *et al.*, 1979; Marr & Hildreth, 1980; Marr, 1982). Although reconstruction of the retinal image is not the purpose of edge detection, a demonstration that a complete reconstruction is possible would confirm that edge maps contain enough information for any further processing stage. Marr's conjecture that the raw primal sketch contained all the information in the image was based on Logan's (1977) theorem, which states that a 1D signal with a bandwidth of one octave or less is completely specified (up to a multiplicative constant) by the positions of its zero-crossings, as long as the function has no zero-crossings in common with its Hilbert transform. An image filtered with a LoG operator does not meet these requirements, since it is 2D and has a bandwidth larger than one octave. However, Marr conjectured that the additional information contained in the raw primal sketch, such as edge contrast, width and orientation, might be sufficient to allow reconstruction of the image. Several researchers have presented algorithms for reconstructing an image from the zero-crossing locations, along with other information (Hummel & Moniot, 1989; Mallat & Zhong, 1992; Elder, 1999). Elder (1999) presented an algorithm that reconstructed images from edge maps generated using Elder & Zucker's (1996, 1998) edge detector. The reconstruction algorithm uses the intensity value on each side of the edge, the edge blur value, and a very coarse estimate of orientation. First, the edge intensity values are diffused across the image, using an algorithm that minimizes the Laplacian of the image at all points except the edge points. Diffusion is not allowed to occur across edges; because of this, all the edges at this stage in the reconstruction are sharp. The algorithm then takes the blur values of the edges and diffuses blur across the image in the same way as intensity. The image is then convolved with a blurring kernel which has a scale that varies across the image, according to the obtained blur value at each image point. The reconstructed images are very close to the originals. Particularly impressive are the algorithm's results with amorphous images, such as clouds, which one might think could not be adequately described by an edge-based representation.

1.11.2 The appropriateness of edge maps

It appears that edge maps contain enough information to perform a good reconstruction of the image, and therefore should contain enough information for any higher-level

processes. A further question is whether the information contained in edge maps is represented in a way that is useful to higher-level mechanisms. One of the central ideas behind Marr's primal sketch was that it should make explicit the information required by other mechanisms.

One way of addressing this issue has been to compare the relative efficacy of edge maps and photographs in tests of object recognition, using human subjects. The rationale is that, if object recognition uses edge maps, then photographs should not provide any advantage over edge maps in object recognition tests. Biederman & Ju (1988) presented evidence that line drawings were no harder to recognize than photographs. However, Sanocki *et al.* (1998) argued that Biederman & Ju's line drawings were not true edge maps. Instead, these drawings contained abstractions and simplifications not present in the images. Structural elements not detectable as edges in the images were included in the corresponding line drawings, and details in the photographs that did not aid recognition (and may have hindered it) were missed out in the line drawings. Sanocki *et al.* carried out object recognition tests with photographs and edge maps generated using Canny's algorithm. They found that edge maps were significantly less easy to recognize. One criticism of this approach is that the result could simply reflect inadequacies in Canny's algorithm. Another criticism is that Sanocki *et al.*'s edge maps used lines to represent edges. This loses all information about the polarity, blur and amplitude of the edges; this information is crucial for the reconstruction algorithms described in the previous section. In addition, most models of edge detection will mark two edges for each line, the only exceptions being the energy models (Owens *et al.*, 1989). Thus, the perceptual processing of the edge maps used by Sanocki *et al.* may have generated two edges for each edge in the photograph, and this could have been disruptive to the object recognition processes.

In summary, it appears that edge maps contain sufficient information for any perceptual processes, but it is not certain that this information is represented in a way that is ideal for further processing. It is not clear whether human vision bases all higher-level processing on a primal sketch of the kind envisaged by Marr, or even whether such a primal sketch exists at all.

It might be argued that it is premature to conduct research into how humans detect edges when it has not been established that they do. A counter-argument would be that, if the results of a large number of psychophysical experiments can be correctly predicted using an edge-based model of perception, then we have the evidence that we need. The next chapter introduces a model of edge detection in humans that has already had considerable success in predicting the results of experiments on the perception of edge blur.

Chapter 2
The Gaussian derivative template model

Although Georgeson (1994) had had some success with the derivative ratio model of blur coding, the model failed to predict correctly the perceived blur of edges formed by adding two Gaussian edges of different scale, or by sharpening a Gaussian edge with a compressive transducer (Georgeson, 2001). The peak-trough distance in the 2nd derivative also failed to predict these data, as did a model based on the width of the 1st derivative signal. Georgeson (2001) began to experiment with Gaussian edge templates, to see if the perceived scale was equal to the scale of the best-fitting template. A straightforward luminance template was unsuccessful. The next option was to differentiate the image and apply a 1st-derivative template, and this model fitted the data better. The data best supported a model that applied a 2nd-derivative template to the 2nd derivative of the image. The edge's 2nd derivative was termed its *signature*.

The Gaussian derivative template (GDT) model consists of a set of channels, selective for orientation and blur. Figure 2.1 shows the processing stages that would occur in one channel, whose orientation is aligned with a particular edge in the image. For simplicity, only one spatial dimension has been included. The upper row of Figure 2.1 shows the processing of an edge with a gradient that is positive relative to the orientation of the channel; the lower row shows the processing of an edge with a negative gradient. The edge is differentiated twice, by convolving with a gradient operator, and then the 2nd derivative signature is filtered with the template, which has an orientation and scale unique to that channel[13]. As explained in section 1.2.2, filtering the signature with the template

[13] The template is the 2nd derivative of a Gaussian edge, and a Gaussian edge is a Gaussian integral, so the template is the 1st derivative of a Gaussian. The gradient operators are also Gaussian 1st derivatives, so the two derivative operations, followed by filtering with the template, can be considered to be a 3rd derivative operation. The only difference is that the two derivative operations involve *convolution*, whereas the

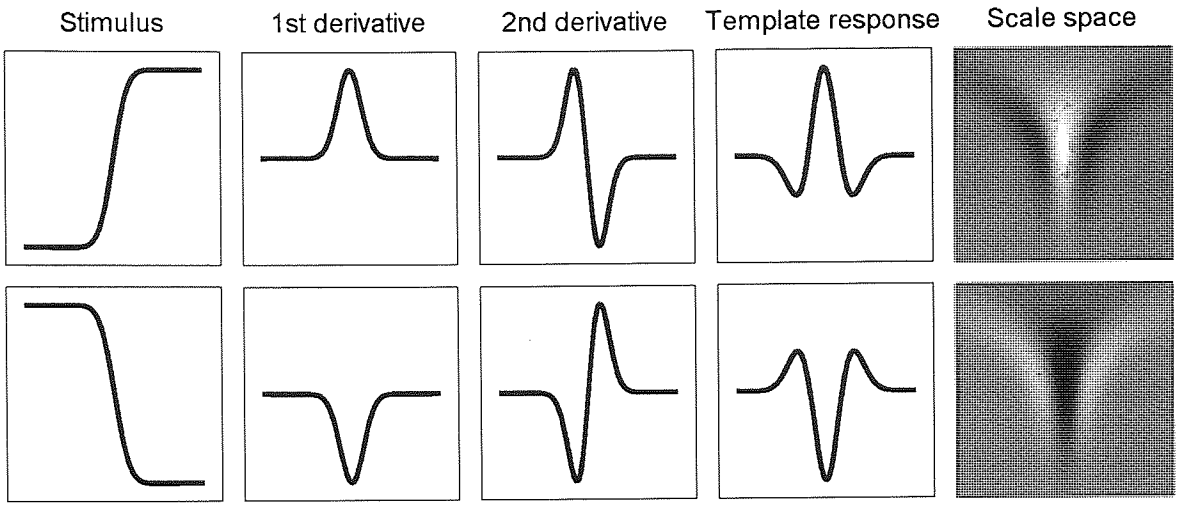


Figure 2.1. Processing stages in an early version of the template model. The top row shows the processing of an edge with a positive gradient; the bottom row shows the processing of an edge with a negative gradient. The four columns on the left show the processing in the blur channel that matches the scale of the edge; in these graphs, the vertical axis represents response level. The right hand column shows scale-space representations that give the template response for all blur channels; here, the vertical axis represents scale, and the response level is represented by the grey level at each point. With the positive edge, a single peak occurs in the scale-space representation, indicating the position and scale of the edge. With the negative edge, two peaks occur either side of the main edge. These can be removed by placing a half-wave rectifier after the 1st derivative operation. The negative edge is then removed from the channel. The rectifier causes each channel to process only edges that have a gradient that is positive relative to the channel's orientation.

gives the correlation between the signature and the template at each spatial position. The responses from the different scales are combined together into a scale-space representation, shown in the right hand column of Figure 2.1. The horizontal axis represents spatial position and the vertical axis represents scale. The grey level at each point represents the correlation between the signature and a template with that particular spatial position and scale. Peaks in scale space represent local maxima in the correlation between signature and template. This suggests that, to detect edges, the model just has to locate the peaks in scale space: each peak could represent an edge element, and the position of each peak in scale space would indicate the spatial position and scale of the best-fitting template. Although this algorithm works with the positive-gradient edge, in the upper half of Figure 2.1, the lower half of this figure indicates a problem. Here, the edge

template matching stage involves *filtering*. Convolution is the same as filtering with the kernel rotated by 180°. In the case of an odd-symmetric filter (like the Gaussian 1st derivative), rotation by 180° is equivalent to multiplying the filter kernel values by -1. Thus, filtering with the template is equivalent to convolving with it and multiplying the result by -1. The result of these operations is therefore equivalent to a 3rd derivative operation multiplied by -1.

has a negative gradient, and the polarity of the signature is reversed. In this case, there are two local maxima of correlation between the signature and template, flanking the main edge. According to the algorithm just described, these would be interpreted as edges. Although some edge profiles do give rise to Mach bands flanking the central edge (e.g. a ramp edge), informal observations suggest that a Gaussian edge does not, so these extra edges would be undesirable.

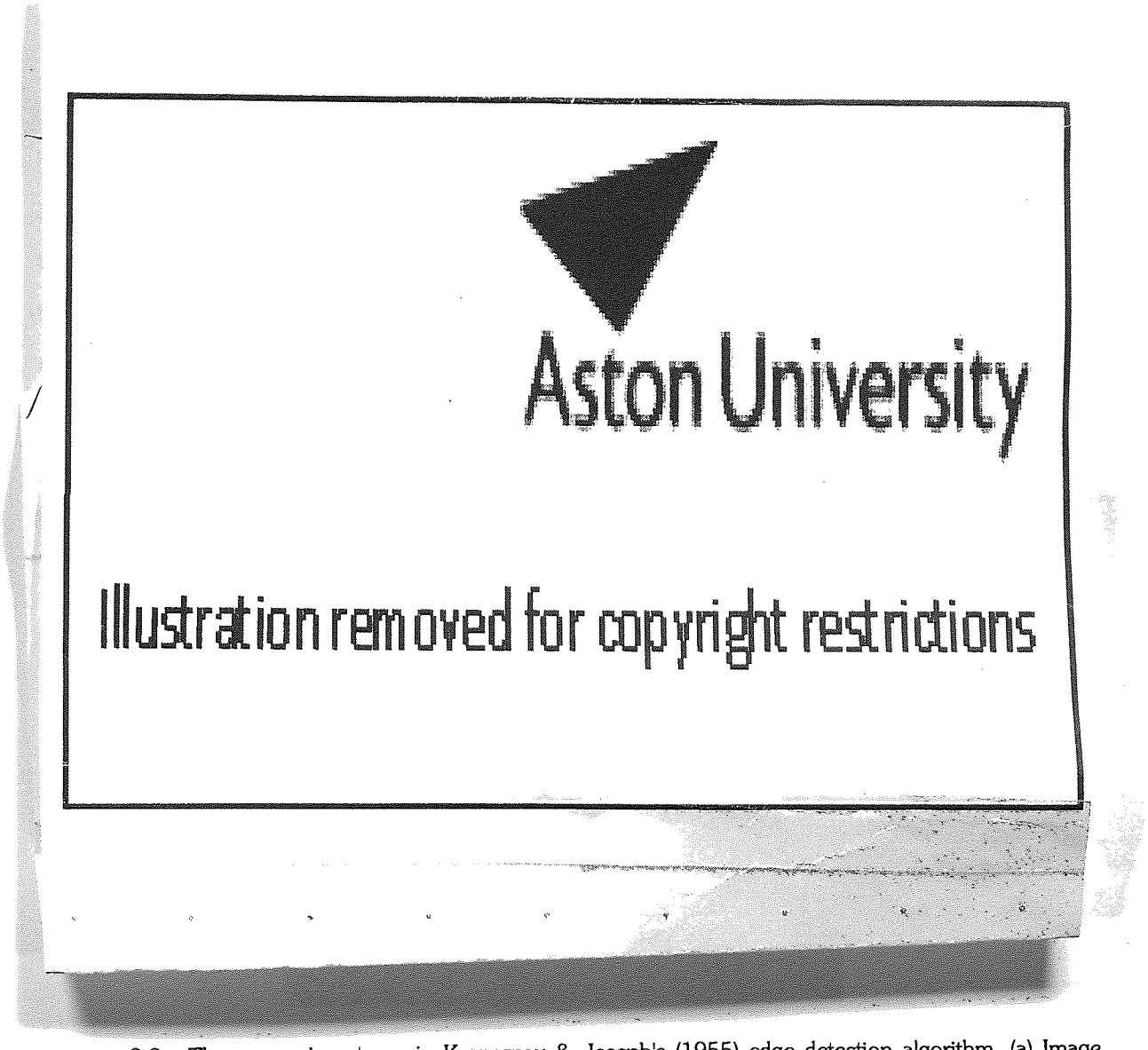


Figure 2.2. The processing stages in Kovasznay & Joseph's (1955) edge detection algorithm. (a) Image luminance; (b) 1st derivative; (c) half-wave rectification with bias; (d) 2nd derivative of parts of signal that pass through the rectifier; (e) trigger driven by the 2nd derivative; (f) pulse formed from the trigger.

The solution can be seen by examining what is probably the earliest published edge detection algorithm. The processing stages in Kovasznay & Joseph's (1955) edge detector are shown in Figure 2.2. This algorithm was implemented using analogue electronics to perform derivative operations. "Orientation channels" were implemented by scanning in different directions. Figure 2.2 shows the processing in two orientation channels of opposite polarity: the solid lines represent the derivatives taken from left to right, and the dashed lines represent derivatives taken in the opposite direction. There were two derivative operations, followed by an operation that marked points at valid ZCs in the 2nd derivative. As discussed in section 1.1.1, the valid ZCs are those with a slope of opposite sign to the 1st derivative. The removal of false ZCs was achieved by placing a half-wave rectifier between the 1st and 2nd derivative operators. This caused edges with opposite polarity to be processed by different channels. The ZCs were found by a circuit that marked one side of each positive section of the 2nd derivative response. The side that was marked depended on the polarity of the channel. In this way, only ZCs with slopes of opposite sign to the 1st derivative were marked. Figure 2.3 shows that the algorithm detected all the major high-contrast edges.

Figure 2.3. The output from Kovasznay & Joseph's (1955) edge detection algorithm. The top row shows the processing of a geometric figure. The bottom row shows that, even back in 1955, the test stimulus of choice was still a woman in an alluring pose. Depending on how the apparatus was set up, the edge map could be black on white (centre) or white on black (right).

The GDT model can remove spurious edges in the same way as Kovasznay & Joseph's model. A rectifier is placed between the 1st and 2nd derivative operations. This has no effect on the positive edge, but completely removes the negative edge from the channel. This edge would instead be processed by channels in which the edge gradient was positive, relative to the channel orientation. This completes the model. In each channel, the edge is differentiated, half-wave rectified, and differentiated again to produce the signature. The signature is filtered with a 2nd-derivative edge template with orientation and scale unique to that channel. The template outputs of all the channels are combined together into a scale-space representation, and peaks in this representation identify the position and scale of any edge elements in the image.

As a historical note, the rectifier was first introduced to improve the fit of the model to some sets of blur-matching data, and it was only later that it was realized that it would remove spurious edges, allowing the model to *detect* edges, as well as judge their blur. Barbieri-Hesse & Georgeson (2002) reported the results of blur-matching experiments that showed that edges formed from one half-period of a sine wave grating were perceived to have the same blur as the edges in a full period of a grating that had the same spatial frequency. The GDT model without the rectifier predicts that the edges in the full period of the grating will look more blurred; with the rectifier, the parts of the signal with negative gradient are removed, so the signatures of the two types of stimulus are identical. Barbieri-Hesse & Georgeson (2002) also reported several other experiments in which the rectifier improved the fit to the data.

This section has presented an informal description of the GDT model. The remainder of this chapter presents a detailed description, including formal derivations of all the equations, and proofs that the model operates correctly. First, the model is developed in 1D. It is then shown that the 2D model can be derived from the 1D model with almost no extra work.

2.1 Stimulus definitions

In 2D, visual stimuli have the following general form:

$$L(x,y,t) = L_0\left(1 + I(x,y,t)\right). \tag{2.1}$$

L is the luminance, which is a function of the two spatial dimensions, x and y, and time, t; I is the image function, which can vary between -1 and 1; L_0 is the mean luminance, i.e. the value of L when $I = 0$. The GDT model takes the function, I, as its input. Thus, it assumes that divisive retinal gain control processes have normalized the image with respect to mean luminance.

For a vertical dark-light Gaussian edge, centred on $x = 0$, the function, I, is defined as follows:

$$I(x,y,t) = E(x;\sigma_e) = C\,\mathrm{erf}\left(\frac{x}{\sigma_e\sqrt{2}}\right), \tag{2.2}$$

where C is the Michelson contrast and σ_e is the edge's spatial scale or blur. E is a Gaussian integral, with scale, σ_e, that varies from $-C$ on the left to C on the right. Thus, its first derivative, E', is a Gaussian function that has an integral (between limits of $\pm\infty$) of $2C$. This gives: $E'(x;\sigma_e) = 2CG(x;\sigma_e)$, where G is the unit-area Gaussian described by the following equation:

$$G(x;\sigma_e) = \frac{1}{\sigma_e\sqrt{2\pi}}\exp\left(\frac{-x^2}{2\sigma_e{}^2}\right).$$

So the full formula for E' is

$$E'(x;\sigma_e) = \frac{2C}{\sigma_e\sqrt{2\pi}}\exp\left(\frac{-x^2}{2\sigma_e{}^2}\right). \tag{2.3}$$

The maximum gradient, E'_{max}, will occur when $x = 0$. Thus,

$$E'_{max} = \frac{2C}{\sigma_e \sqrt{2\pi}}. \tag{2.4}$$

The third derivative, E''' is given by

$$E'''(x;\sigma) = \frac{2C}{\sigma_e{}^3 \sqrt{2\pi}} \left(\frac{x^2}{\sigma_e{}^2} - 1 \right) \exp\left(\frac{-x^2}{2\sigma_e{}^2} \right). \tag{2.5}$$

2.2 The GDT model in 1D

2.2.1 Orientation/blur channels

In the 2D model, each channel is selective for edge orientation and spatial scale. Orientation is a characteristic that does not exist in a 1D signal. Because of this, the channels in the 1D model cannot be orientation-selective: instead, they are selective for edge polarity, i.e., positive (changing from dark to light towards the right – henceforth, dark-light) versus negative (light-dark). Polarity can be thought of as a 1D version of orientation because, when an edge is rotated by 180°, its polarity changes between positive and negative.

At each polarity, there are N channels. Each channel, n, has a template, which is the 2nd derivative of a Gaussian edge with scale, $\sigma_t(n)$. The model parameters that specify the channel scales are:

1. N, the number of channels;
2. $\sigma_t(1)$, the scale of the smallest template;
3. $\sigma_t(N)/\sigma_t(1)$, the ratio of the largest to smallest template scales.

Stepping through the channels, the scale increases by a constant proportion each time. Thus,

$$\sigma_t(n) = \sigma_t(1)\left(\frac{\sigma_t(N)}{\sigma_t(1)}\right)^{\frac{n-1}{N-1}}. \qquad (2.6)$$

When $n=1$, we have $\sigma_t(n)=\sigma_t(1)$ and, when $n=N$, we have $\sigma_t(n)=\sigma_t(N)$. Each channel's template scale differs from the previous one by a factor

$$\left(\frac{\sigma_t(N)}{\sigma_t(1)}\right)^{\frac{1}{N-1}}.$$

2.2.2 Operators

As explained in footnote 13 on page 30, all the operators have the form of Gaussian 1st derivatives, i.e. the 1st partial derivative of a unit area Gaussian. If g is the operator, then

$$g(x;\sigma) = \frac{-x}{\sigma^3\sqrt{2\pi}}\exp\left(\frac{-x^2}{2\sigma^2}\right) \qquad (2.7)$$

σ is the operator scale. In channel n, the scale for the template is $\sigma_t(n)$, given by (2.6). The scale, $\sigma_g(n)$, of the gradient operators in channel n is a constant proportion, m, of the template scale:

$$\sigma_g(n) = m\sigma_t(n). \qquad (2.8)$$

The value of m used for all the simulations described here is 0.25. This value best accounted for Georgeson & Freeman's (1994) data on the effect of edge length on perceived blur (Georgeson, personal communication).

Equation (2.7) describes the operators in a channel with positive polarity. The negative channels take derivatives in the opposite direction, and have templates with opposite polarity. This is achieved by multiplying the operator in (2.7) by -1.

The operator described by (2.7) is a continuous function with an infinite domain. For a practical implementation, the filter kernel must be a discrete function with finite domain $Z = \{i \in \mathbb{Z} \mid i_{\min} \le i \le i_{\max}\}$. The discrete version of the model plays an essential part in the theoretical development, so it is worthwhile to express the operator in its discrete form:

$$g(i) = \begin{cases} \dfrac{-i\delta x}{\sigma^3\sqrt{2\pi}} \exp\left(\dfrac{-(i\delta x)^2}{2\sigma^2}\right) & i \in Z \\ 0 & i \notin Z \end{cases} . \qquad (2.9)$$

δx is the visual angle subtended horizontally by a single pixel. For simplicity, it is assumed that the central element of the kernel has an index $i = 0$, and that negative indices are allowed. As mentioned in section 1.5.1, Lindeberg (1990, 1993a) showed that, unlike in a continuous scale space, the operators that satisfy the causality criterion in a discrete scale space are not Gaussian, so this sampled Gaussian operator is not ideal. However, the Gaussian is easier to handle algebraically and, with a fine sampling resolution, the use of a sampled Gaussian should not be problematic.

2.2.3 The sequence of operations

The image is differentiated, half-wave rectified, differentiated again, and then filtered with the template. The terms I_x and I_{xx} will refer to the 1st and 2nd derivatives found by convolving with Gaussian gradient operators of non-zero scale (which introduces some extra blur); the term I'' will refer to the pure 2nd derivative, with no additional blurring. I_{xx} is the edge's *signature*. In the informal presentation of the model given above, the signature was defined as the edge's 2nd derivative, without making a distinction between I_{xx} and I''. As will be explained in section 2.2.4, this distinction is very important. H will represent the output of the half-wave rectifier; T will be the result of filtering the signature

with the template. For each channel, n, the gradient operator will be represented by the function $g(\cdot;n)$, and the template by the function $t(\cdot;n)$; these operators are generated using equation (2.7) in the continuous case, and equation (2.9) in the discrete case. If we consider the signals to be continuous functions, the derivative and template matching stages can be expressed using the standard formulae for convolution and filtering. The four stages are:

$$I_x(x;n) = \int_{\xi\in\mathbb{R}} g(x-\xi;n)I(\xi)d\xi\,, \tag{2.10}$$

$$H(x;n) = \lfloor I_x(x;n) \rfloor, \tag{2.11}$$

$$I_{xx}(x;n) = \int_{\xi\in\mathbb{R}} g(x-\xi;n)H(\xi;n)d\xi\,, \tag{2.12}$$

$$T(x;n) = \int_{\xi\in\mathbb{R}} t(\xi-x;n)I_{xx}(\xi;n)d\xi\,. \tag{2.13}$$

Note that, in (2.11) $\lfloor a \rfloor = \max(a,0)$. Note also the slight difference between the convolution equations, (2.12), and the filtering equation, (2.13). Equation (2.13) is essentially a continuous form of (1.4); in (2.10) and (2.12), the index of the kernel is inverted, to bring about the necessary reflection of the kernel.

The operations are expressed most straightforwardly in terms of continuous functions. However, the correlation that is performed is easier to analyze using discrete forms. In addition, the implementation must be discrete. Therefore, equations (2.10) to (2.13) are expressed as discrete functions, which would use the discrete form of the operator, given by (2.9):

$$I_x(i;n) = \sum_{\alpha\in A} g(i-\alpha;n)I(\alpha)\delta x\,, \tag{2.14}$$

$$H(i;n) = \lfloor I_x(i;n) \rfloor, \tag{2.15}$$

$$I_{xx}(i;n) = \sum_{\alpha \in A} g(i-\alpha;n) H(\alpha;n) \delta x, \tag{2.16}$$

$$T(i;n) = \sum_{\alpha \in A} t(\alpha - i;n) I_{xx}(\alpha;n) \delta x. \tag{2.17}$$

Equations (2.14) to (2.17) are closely related to equations (2.10) to (2.13)[14]. The index, i, corresponds to x; the index, α, corresponds to ξ; A is the set of values that α can take; δx corresponds to $d\xi$. The standard formula for discrete filtering (equation (1.4)) does not include the δx term; this term is essential if the filtering process is to implement an integral.

2.2.4 The virtual template

In the terminology of section 1.2.2, equations (2.13) and (2.17) give the unnormalized correlation coefficient between the signature and template. The computational-level requirement of the GDT model is that we match a template to the 2nd derivative of the edge. Because the first two derivative operations introduce blurring, the signature is not a pure 2nd derivative. However, the fact that convolution commutes with differentiation allows us to access the pure 2nd derivative. The scales of all the operators add in Pythagorean sum. This means that convolving with the two derivative operators (with

---

[14] Although the equations of the discrete version of the model look very similar to the continuous versions, there are some hidden difficulties because the indices generated by these equations can fall outside the set of indices over which the image or kernel is defined. Equation (2.9) takes care of this for the kernel because it gives a value of zero for any index not included in this set. Similarly, the image can be assumed to have a value of zero for any index that falls outside of the range of indices for which it is defined. An additional difficulty is that it is assumed that the central index of the filter kernel has an index of zero, and that negative indices are allowed. In MATLAB, arrays are indexed with positive indices, starting at 1, so the central pixel cannot have a zero index. In such a case, the index of $g(\cdot;n)$ in (2.14) and (2.16) would have to be adjusted so that it still addresses the central pixel when $i = \alpha$. Suppose $g(\cdot;n)$ is a vector of length M (which is an odd number). Then the central element of $g(\cdot;n)$ will have an index $(M+1)/2$. Thus, $g(i-\alpha;n)$ in (2.14) and (2.16) should be replaced with $g((M+1)/2+i-\alpha;n)$. A similar modification should be made to (2.17). Additionally, the definition of the gradient operator, given in equation (2.9), would have to be modified in a similar way.

scale σ_g), followed by filtering with the template (with scale σ_t) is equivalent to filtering the pure 2nd derivative of the edge with a *virtual template* with scale $\sqrt{2\sigma_g{}^2 + \sigma_t{}^2}$. Using (2.8), we can see that, for each channel, n, the virtual template scale, $\sigma_v(n)$, is given by

$$\sigma_v(n) = \sigma_t(n)\sqrt{2m^2 + 1} \ . \tag{2.18}$$

The virtual template operator, v, would be generated in the same way as the others, using (2.7) or (2.9) with scale $\sigma_v(n)$. It must be emphasized that the virtual template is not explicitly created or used: equations (2.13) and (2.17) show that the template response, T, is generated by filtering with the template, t, not the virtual template, v. However, these operations are exactly equivalent to the filtering operations shown in (2.19) and (2.20), which do use the virtual template:

$$T(x;n) = \int_{\xi \in \mathbb{R}} v(\xi - x;n) I''(\xi) d\xi \ , \tag{2.19}$$

$$T(i;n) = \sum_{\alpha \in A} v(\alpha - i;n) I''(\alpha) \delta x \ . \tag{2.20}$$

I'' represents the pure 2nd derivative of the image. Instead of regarding equation (2.13) or (2.17) as matching the *template* to the *signature*, the process should be thought of as matching the *virtual template* to the *edge's 2nd derivative*. The template and virtual template are only identical if the two gradient operations are performed using infinitesimal operators, giving the pure 2nd derivative; in this case, m is effectively zero.

2.2.5 Normalizing the correlation

Equations (2.19) and (2.20) give an unnormalized correlation coefficient, so it is necessary to normalize the result. This is of crucial importance: it turns out that the normalization factor varies across scale so, without this factor, the peak over scale would not give the scale for which the normalized correlation was highest. A correlation between two

functions is most easily understood by considering the discrete case. Equation (2.20) can be re-written

$$T(i;n) = \sum_{\alpha \in A} u(\alpha;n) I''(\alpha) \delta x \,, \tag{2.21}$$

where $u(\alpha;n) = v(\alpha - i;n)$. $u(\cdot;n)$ is the virtual template kernel, $v(\cdot;n)$, shifted to the right by i pixels, so that it is centred on pixel i in the image. This shift to the right will not affect the sum of squared values, i.e.:

$$\sum_{\alpha \in A} u^2(\alpha;n) = \sum_{\alpha \in A} v^2(\alpha;n) \,. \tag{2.22}$$

Let r be the normalized correlation between $u(\cdot;n)$ and I'', i.e. the cosine of the angle between their vector forms. Then, from (1.12),

$$r = \frac{\sum_{\alpha \in A} u(\alpha;n) I''(\alpha)}{\sqrt{\sum_{\alpha \in A} u^2(\alpha;n) \sum_{\alpha \in A} I''^2(\alpha)}} \,. \tag{2.23}$$

So, from (2.21) to (2.23), it can be deduced that the correlation, $r(i;n)$, between I'' and the virtual template in channel n, centred on pixel i, is given by:

$$r(i;n) = \frac{T(i;n)}{\delta x \sqrt{\sum_{\alpha \in A} v^2(\alpha;n) \sum_{\alpha \in A} I''^2(\alpha)}} \,. \tag{2.24}$$

Local peaks in the 2D scale-space representation, $r(i;n)$, are local peaks in the correlation.

The denominator in (2.24) ensures that the correlation coefficient is scaled to lie between -1 and 1, with 1 representing a perfect correlation. This removes all information about contrast since, for a Gaussian edge, the correlation between the 2nd derivative and the best-fitting virtual template will always be 1, whatever the contrast. Information about

contrast can be restored by dropping the $\sum_{\alpha \in A} I''^2(\alpha)$ term, giving a new scale-space representation, $s(i;n)$:

$$s(i;n) = \frac{T(i;n)}{\delta x \sqrt{\sum_{\alpha \in A} v^2(\alpha;n)}}. \tag{2.25}$$

$\sum_{\alpha \in S} I''^2(\alpha)$ is a constant across all blur channels, because it is the sum of squares of all the elements in the 2nd derivative of the *stimulus*. If it is removed, the positions of the peaks across scale will be unchanged, but now the height of a peak will be proportional to contrast within each blur channel (although the constant of proportionality differs across blur channels - see section 2.2.8). It is very important to keep the $\sum_{\alpha \in S} v^2(\alpha;n)$ term, as this decreases with increasing channel scale – dropping this term would shift the peaks in the scale-space representation towards smaller blur channels, so that their positions no longer indicated the positions of local peaks in the correlation between I'' and the virtual template. Note that the denominator in (2.25) contains the sum of squares of the *virtual template*, *v*, rather than the *template*, *t*. This is because we are interested in the correlation between the virtual template and the pure 2nd derivative, not between the template and the signature. Even though the algorithm finds T by filtering the signature with the template (described by equations (2.13) and (2.17)), the normalization proceeds as if T had been found by the equivalent procedure of filtering the pure 2nd derivative with the virtual template (described by (2.19) and (2.20)).

Any peak in $s(i;n)$ is taken to represent an edge element. Finding the position (*i*) and blur channel (*n*) where each peak occurs allows the model to determine both where the edge element is and what its scale is. The edge position is simply the spatial position of the point occupied by pixel i in the image; the edge scale is the scale of the virtual template in channel *n*.

Apart from the δx term in (2.25), the model presented so far is identical to that developed by Georgeson (Georgeson *et al.*, 2002; Barbieri-Hesse & Georgeson, 2002).

The following modification, and the material presented in section 2.2.5.1, and sections 2.2.6 to 2.2.8, are original work undertaken since then.

By dropping the $\sum_{\alpha \in A} I''^2(\alpha)$ term, we are effectively multiplying $r(i;n)$ by $\sqrt{\sum_{\alpha \in A} I''^2(\alpha)}$. But $\sum_{\alpha \in A} I''^2(\alpha)$ is proportional to the image resolution, $1/\delta x$. Thus, by dropping the $\sum_{\alpha \in A} I''^2(\alpha)$ term, we are making $s(i;n)$ proportional to $1/\sqrt{\delta x}$. Ideally, we require that the model's output is not affected by superficialities such as the resolution of the images used. This can be achieved by multiplying $s(i;n)$ by $\sqrt{\delta x}$ (since δx is a constant, this does not affect the positions of the peaks). So the final scale-space processing stage of the model, $S(i;n)$, is given by

$$S(i;n) = \frac{T(i;n)}{\sqrt{\sum_{\alpha \in A} v^2(\alpha;n)\delta x}}. \tag{2.26}$$

S will be referred to as *edge strength,* after an analogous measure introduced by Lindeberg (1996, 1998).

Equation (2.26) has two advantages over (2.25): firstly, it is independent of image resolution; secondly, both the numerator and denominator contain a sum multiplied by δx, and the existence of the δx term makes it easy to convert these sums into continuous integrals, allowing algebraic proofs of some of the model's properties.

$$\text{As } \delta x \to 0, \ \sum_{\alpha \in A} v^2(\alpha;n)\delta x \to \int_{x \in \mathbb{R}} v^2(x;n)\,dx.$$

Therefore, we can generate a continuous version of (2.26):

$$S(x;n) = \frac{T(x;n)}{\sqrt{\int_{x \in \mathbb{R}} v^2(x;n)\,dx}}. \tag{2.27}$$

The reciprocal of the denominator of (2.27) will be referred to as the *normalization factor*.

2.2.5.1 Deriving an expression for the normalization factor

The key part of the denominator of (2.27) is $\int_{x\in\mathbb{R}} v^2(x;n)\,dx$.

From (2.7),

$$\int_{-\infty}^{\infty} v^2(x;n)\,dx = \frac{1}{2\pi\sigma_v^{\,6}(n)} \int_{-\infty}^{\infty} x^2 \exp\left(\frac{-x^2}{\sigma_v^{\,2}(n)}\right) dx .$$

Since v^2 has even symmetry, we can write:

$$\int_{-\infty}^{\infty} v^2(x;n)\,dx = \frac{1}{\pi\sigma_v^{\,6}(n)} \int_{0}^{\infty} x^2 \exp\left(\frac{-x^2}{\sigma_v^{\,2}(n)}\right) dx . \qquad (2.28)$$

A standard result is:

$$\int_{0}^{\infty} x^2 \exp\left(-ax^2\right) = \frac{\sqrt{\pi}}{4a^{3/2}} \qquad (2.29)$$

(e.g., see Spiegel, 1968). From (2.28) and (2.29) we get

$$\int_{-\infty}^{\infty} v^2(x;n)\,dx = \frac{1}{4\sigma_v^{\,3}(n)\sqrt{\pi}} . \qquad (2.30)$$

Substituting (2.30) into (2.27) gives an algebraic expression for the edge strength, in terms of the spatial position, x, assuming infinitely high image resolution:

$$S(x;n) = 2\pi^{1/4}\sigma_v^{3/2}(n)T(x;n). \qquad (2.31)$$

The normalization factor for channel n is $2\pi^{1/4}\sigma_v^{3/2}(n)$.

2.2.6 Deriving an expression for the edge strength in response to Gaussian edges

In (2.31), $T(\cdot;n)$ is the result of convolving the image with the three sets of filters in the model – two derivative operators, with scale $m\sigma_t(n)$, and the template, with scale $\sigma_t(n)$. The first two operators have a half-wave rectifier between them but, with a monotonic function, such as an edge, the gradient within a channel will be either entirely negative (in which case there is no signal in any subsequent processing stages), or entirely positive (in which case the rectifier has no effect). As explained in footnote 13 on page 30, this sequence of operations can be considered to be a Gaussian 3rd derivative operation multiplied by -1. Suppose the image is a Gaussian edge with scale σ_e (without loss of generality, we can consider the edge to be located at position $x = 0$). Then $T(\cdot;n)$ will be -1 multiplied by the 3rd derivative of a Gaussian edge with scale $\sqrt{\sigma_e^2 + \sigma_t^2(n)(2m^2+1)}$, which, from (2.18), equals $\sqrt{\sigma_e^2 + \sigma_v^2(n)}$. Therefore, from (2.5), we have

$$T(x;n) = \frac{2C}{\left(\sigma_e^2 + \sigma_v^2(n)\right)^{3/2}\sqrt{2\pi}}\left(1 - \frac{x^2}{\sigma_e^2 + \sigma_v^2(n)}\right)\exp\left(\frac{-x^2}{2\left(\sigma_e^2 + \sigma_v^2(n)\right)}\right), \qquad (2.32)$$

where $\sigma_v(n)$ is calculated using (2.6) and (2.18). Inserting this expression for $T(x;n)$ into (2.31) gives an algebraic expression for the edge strength across scale space in response to Gaussian edges:

$$S(x;n) = C\pi^{-1/4}\left(\frac{2\sigma_v(n)}{\sigma_e^2 + \sigma_v^2(n)}\right)^{3/2}\left(1 - \frac{x^2}{\sigma_e^2 + \sigma_v^2(n)}\right)\exp\left(\frac{-x^2}{2\left(\sigma_e^2 + \sigma_v^2(n)\right)}\right). \quad (2.33)$$

Equation (2.33) describes the kind of scale space response shown in the right hand column of Figure 2.1.

2.2.7 Proof that the template model correctly estimates the position and scale of a Gaussian edge

The template model is designed to select the virtual template that fits best to the edge's 2nd derivative. The estimated edge position and scale are taken to be the position and scale of the best-fitting virtual template. The 2nd derivative of an unwindowed Gaussian edge will exactly match the best-fitting virtual template, and so the position and scale of the best-fitting virtual template will be equal to the position and scale of the edge, and the model should recover these edge properties veridically.

Equation (2.33) gives an opportunity to prove algebraically that the peak in edge strength in response to a Gaussian edge correctly indicates the edge's position and scale, allowing a confirmation of the correctness of the algebra presented so far. For any given virtual template scale, $\sigma_v(n)$, (2.33) has the form of an inverted Gaussian 2nd derivative, which has a single positive peak at $x = 0$. Since the derivation of (2.33) made the assumption that the edge was centred on $x = 0$, we can see that the position of the peak correctly indicates the edge's spatial position. To show that it also correctly indicates the edge's scale, let us set x to 0 and, for convenience, redefine S as a function of $\sigma_v(n)$, rather than n:

$$S(0;\sigma_v) = C\pi^{-1/4}\left(\frac{2\sigma_v}{\sigma_e^2 + \sigma_v^2}\right)^{3/2}.$$

The partial derivative of S with respect to σ_v is

$$\frac{\partial S}{\partial \sigma_v} = 3C\pi^{-1/4}\left(\frac{2\sigma_v}{\sigma_e^{\,2}+\sigma_v^{\,2}}\right)^{1/2}\left(\frac{\sigma_e^{\,2}-\sigma_v^{\,2}}{\left(\sigma_e^{\,2}+\sigma_v^{\,2}\right)^2}\right).$$

The peak along the blur dimension occurs when $\partial S / \partial \sigma_v = 0$. Setting $\partial S / \partial \sigma_v = 0$ and squaring both side gives, after simplification,

$$\frac{\sigma_v\left(\sigma_e^{\,2}-\sigma_v^{\,2}\right)^2}{\left(\sigma_e^{\,2}+\sigma_v^{\,2}\right)^5} = 0.$$

So either $\sigma_v = 0$ or $\left(\sigma_e^{\,2}-\sigma_v^{\,2}\right) = 0$. Setting $\sigma_v = 0$ in (2.33) gives $S = 0$ for any value of x (unless σ_e is also 0, in which case S is undefined) so, since the peak is known to be positive, there is not a peak at $\sigma_v = 0$. This leaves us with $\left(\sigma_e^{\,2}-\sigma_v^{\,2}\right) = 0$. Thus, $\sigma_v^{\,2} = \sigma_e^{\,2}$ and, since σ_v and σ_e are both positive, we have $\sigma_v = \sigma_e$, so the virtual template scale in the channel where the peak occurs is equal to the edge scale. Thus, the position of the peak in the edge strength correctly indicates both the position and scale of a Gaussian edge.

2.2.8 Recovering edge contrast from peak height

The previous section proved that, with a Gaussian edge with scale, σ_e, at position $x = 0$, the peak in the edge strength occurs when $x = 0$ and $\sigma_v = \sigma_e$. From (2.33), the height of the peak at $(0;n)$, where n is the number of the channel with $\sigma_v = \sigma_e$, is given by

$$S(0;n) = C\pi^{-1/4}\sigma_v^{\,-3/2}(n). \qquad (2.34)$$

A formula for estimating the contrast, C, of the edge from the height of the peak can be derived by rearranging (2.34):

$$C = \pi^{1/4}\sigma_v^{3/2}(n)S(0;n). \tag{2.35}$$

Note that it is important not to multiply the whole of S by $\pi^{1/4}\sigma_v^{3/2}(n)$ *before* finding the peaks, as this would shift the peaks towards larger-scale channels, so they would no longer correspond to the positions of the peaks of correlation between the virtual template and the image's 2nd derivative. The locations of the peaks must be found first, and only then should the contrast be found, according to (2.35).

2.3 The GDT model in 2D

2.3.1 Orientation/blur channels

Although images are 2D, an edge element is still a 1D entity in that it varies along only one dimension. For this reason, the 2D implementation is very similar to the 1D version. Instead of just two polarity channels, there is a set of orientation channels, spanning the complete range of orientations. Each channel is selective for orientation and scale. The gradient operators take the partial derivative along the direction perpendicular to the channel's orientation.

2.3.2 Operators

As in the 1D model, the operators all have the same form. They are defined by taking the partial derivative of a unit-volume Gaussian surface, G:

$$G(x,y;\sigma) = \frac{1}{2\pi\sigma^2}\exp\left(\frac{-(x^2+y^2)}{2\sigma^2}\right).$$

Let the Gaussian derivative operator in blur channel n and orientation channel θ be $g(\cdot,\cdot;n,\theta)$. The vertical operator, $g(\cdot,\cdot;n,0)$, oriented at 0° from vertical, is given by

$$g(x,y;\sigma,0)=\frac{-x}{2\pi\sigma^4}\exp\left(\frac{-(x^2+y^2)}{2\sigma^2}\right). \tag{2.36}$$

The equation for the gradient operator with orientation, θ, is found by substituting $x\cos(\theta)+y\sin(\theta)$ for x and $y\cos(\theta)-x\sin(\theta)$ for y in Equation (2.36), and then simplifying:

$$g(x,y;\sigma,\theta)=\frac{-(x\cos(\theta)+y\sin(\theta))}{2\pi\sigma^4}\exp\left(\frac{-(x^2+y^2)}{2\sigma^2}\right). \tag{2.37}$$

A discrete version can be constructed by sampling from this function in a similar way to the 1D case, described in section 2.2.2. The gradient operators, template and virtual template are all forms of this operator.

2.3.3 Sequence of operations

The sequence of operations in the 2D model is identical to that in the 1D model, except that 2D operators are used, generated according to (2.37):

$$I_x(i,j;n,\theta)=\sum_{\beta\in B}\sum_{\alpha\in A} g(i-\alpha,j-\beta;n,\theta)I(\alpha,\beta)\delta x\delta y\,,$$

$$H(i,j;n,\theta)=\lfloor I_x(i,j;n,\theta)\rfloor,$$

$$I_{xx}(i,j;n,\theta)=\sum_{\beta\in B}\sum_{\alpha\in A} g(i-\alpha,j-\beta;n,\theta)H(\alpha,\beta)\delta x\delta y\,,$$

$$T(i,j;n,\theta)=\sum_{\beta\in B}\sum_{\alpha\in A} t(\alpha-i,\beta-j;n,\theta)I_{xx}(\alpha,\beta;\theta)\delta x\delta y \tag{2.38}$$

$$=\sum_{\beta\in B}\sum_{\alpha\in A} v(\alpha-i,\beta-j;n,\theta)I''(\alpha,\beta;\theta)\delta x\delta y\,. \tag{2.39}$$

$I''(\cdot,\cdot;\theta)$ is the 2nd partial derivative of the image along the direction θ from horizontal. A and B are the sets of indices, α and β. The operators g, t, and v, are the 2D versions of the corresponding operators in the 1D model. Equation (2.38) describes the actual filtering process that the model performs; equation (2.39) describes the equivalent operation that corresponds to filtering the pure 2nd derivative with the virtual template.

2.3.4 Correlation and peak-finding

Having found the template response, T, the next step is to find the virtual template with a scale that matches the edge. This can be achieved very easily by using the results from the 1D model. Given an ideal Gaussian edge, the sequence of operations in the channel aligned with the edge orientation in the 2D model will yield the same result as the corresponding sequence of operations in the 1D model: both sets of operations will give rise to a template response, T, which is the 3rd derivative of an edge with a scale that is the Pythagorean sum of the stimulus edge scale and the scales of all the operators in that channel. In the 1D model, T will be a 1D function; in the 2D model, T will be a function that varies only along 1 dimension, and which has exactly the same cross-sectional profile as the corresponding function in the 1D model. Thus, the height of T at the edge location will be identical in corresponding channels of both 1D and 2D models. Therefore, the same normalization factor should be used to convert the template response, T, into edge strength, S:

$$S(x,y;n,\theta) = 2\pi^{1/4}\sigma_v^{3/2}(n)T(x,y;n,\theta). \tag{2.40}$$

The partial derivative of an edge profile is largest in the direction perpendicular to the edge. This means that, since T is a partial derivative in a direction perpendicular to the channel orientation, the peak in edge strength at the edge location will occur in the channel aligned to the edge, so the peak in edge strength indicates the edge orientation as well as the scale. In the 1D model, the edge strength had a 2D domain, with dimensions of location and scale, and the location and scale of the each edge is determined from the position of the corresponding peak. In the 2D model, the edge strength has a 4D domain, with 2 spatial dimensions, a scale dimension, and an orientation dimension. The

straightforward extension from the 1D model would be to find peaks in edge strength over its 4D domain. This would not work, though, because then only a few points along an edge contour would be selected, i.e. those with local peaks of edge strength. In common with all non-maximum suppression algorithms (e.g. Rosenfeld & Thurston, 1971; Rosenfeld *et al.*, 1972), the peaks must be found along a spatial direction perpendicular to the local contour orientation.

The proposed solution is strongly influenced by Lindeberg's (1996, 1998) algorithm, described in section 1.8.5.9. The process runs as follows. At each spatial position within each scale, the local orientation is determined from the peak of edge strength across orientation at that position and scale. Provisional edges at each scale should then be marked at points that are local maxima in edge strength along a direction perpendicular to the local orientation, at that scale. Contours in the image will therefore give rise to contours of provisional edge elements in a contiguous range of scales, which join up to form an edge surface in scale space. Points on this surface are then marked as edges if they are peaks in edge strength across scale.

2.4 Relations to other literature

In formal terms, the GDT model is most closely related to Lindeberg's (1996, 1998) algorithm. Lindeberg expressed his model in terms of Gaussian derivatives, rather than templates. In his model, provisional edges are marked at peaks in the 1st derivative along the gradient direction, creating an edge surface in scale space. Edges on this surface are then marked at peaks in edge strength over scale. The edge strength operator is a Gaussian 1st derivative, normalized so that edge strength peaks at a scale equal to the edge scale. Thus, Lindeberg's normalized 1st derivative operation is equivalent to the GDT model's template matching operation, except that it corresponds to a 1st-derivative template, rather than a 2nd derivative (see section 1.4.4.2 for an explanation of how a Gaussian 1st derivative operation is equivalent to template matching in the gradient domain). The psychophysical data presented by Georgeson (2001) show that, for modelling human perception, a 2nd derivative template is preferable.

As explained in footnote 13 on page 30, the GDT model has an equivalent interpretation in terms of normalized Gaussian derivatives. The sequence of operations is effectively a Gaussian 3rd derivative operation, except that there has to be a rectifying nonlinearity between 1st and 2nd derivative operations, to remove the spurious peaks that would otherwise arise from the extra derivative operations. In the vicinity of an edge, though, the rectifier does not affect the signal in a channel with the same orientation as the edge. Having found the 3rd derivative, the model then normalizes it, to produce an edge strength measure that is exactly analogous to Lindeberg's edge strength measure derived from normalized Gaussian 1st derivatives.

A difference between the two models is that, in Lindeberg's model, the local orientation, and peaks in the spatial domain are found from the unnomalized 1st derivatives, and it is only the peaks over scale that are found using the normalized edge strength measure. In the GDT model, the peaks across local orientation, space, and scale are *all* found using peaks in edge strength.

Korn (1988) proposed a similar method of scale selection, described in section 1.8.4.1. He claimed that peaks over scale would indicate the local blur of the image. He used Gaussian 1st derivatives, like Lindeberg, but used a different normalization factor. Zhang & Bergholm (1997) showed that Korn was using his normalized derivatives in an inappropriate way, because these derivatives would not give rise to a peak across scale in response to an isolated Gaussian edge.

2.4.1 Summary of previous template models

Several previous edge detection algorithms have been expressed explicitly in the form of template algorithms, although very few used template matching to find an edge's blur. The earliest template matching algorithms were the compass gradient algorithms, described in section 1.2.2. These used template matching to find the orientation at each point, and then the gradient was found from the response of the template with the best

orientation, an operation that exploited the equivalence of template matching and differentiation when using these operators.

The compass gradient operators were all very small – most were 3×3 pixels in size – and this made them very sensitive to high frequency noise. Nevatia & Babu (1980) presented a matched filtering algorithm with 5×5 operators, but Lyvers & Mitchell (1988) found its performance to be very poor. Hueckel (1971, 1973, 1974) developed the first edge detector to perform template matching with large operators, as described in section 1.4.1. Despite an ingenious analytical method for finding the best-fitting template, this algorithm performed considerably worse than its main contemporaries (see Fram & Deutsch, 1975).

Nalwa & Binford (1986) presented a template matching algorithm, described in section 1.4.4.2, that modelled edges as tanh functions. This algorithm took the slightly unusual step of marking an edge at any point at which the best-fitting tanh template fitted better than the best-fitting quadratic surface. The performance of this algorithm was quite comparable to that of Canny's algorithm (Heath *et al.*, 1997, 1998).

Boie and colleagues (Boie *et al.*, 1986; Boie & Cox, 1987) introduced the idea of matched filtering in the gradient domain, described in section 1.4.4.2. This approach allowed them to derive an edge detection filter that was spatially localized, without making any assumptions about the positions of edges flanking the edge under consideration. They also argued that the optimal edge *localization* filter was the 1st derivative of the optimal edge *detection* filter.

Lacroix (1990) also suggested using a Gaussian 1st derivative template, as described in section 1.4.4.2. It seems that Lacroix was only thinking about matching over orientation, rather than scale. He noted that matching over space with a 1st derivative template was equivalent to finding the peak of gradient.

A new method of template matching, called expansion matching, was presented by Ben-Arie & Rao (1993a, 1993b), described in section 1.4.2. Rather than filtering with the template itself, as in matched filtering algorithms, expansion matching takes a template and finds the linear filter that optimizes the discriminative SNR; this measure treats as

noise any response away from the location of the target feature. They presented versions of this algorithm for detecting rectangular step edges (Rao & Ben-Arie, 1993, 1994) and blurred edges (Wang *et al.*, 1996). The response profiles showed high, narrow peaks at the feature locations, and a multitude of small peaks, due to noise, at other locations. The small peaks due to noise could be removed using a threshold. Wang *et al.* did not explore the possibility of using templates of different scale to judge the scale of the edges.

All of the template matching algorithms described so far in this section were presented as single scale models. Nalwa & Binford suggested that, with images containing many different amounts of blur, it would be desirable to use several templates with different scales, and integrate the information across scale, but they did not indicate how this could be done. Kisworo, *et al.* (1994) presented an algorithm, described in section 1.8.7.1, in which edges were *detected* by locating peaks in the energy function, but were then characterized in terms of width, etc., by fitting a flexible spatial template. An energy computation was performed on both the edge and the template, and the algorithm selected the template which had an energy profile that best matched that of the edge. Zhang & Bergholm (1997) presented a multi-scale edge detection algorithm, described in section 1.8.4.3, that used templates to characterize edges in scale space. The edges were *detected* using Korn's (1988) normalized derivatives in combination with Bergholm's (1987) edge focusing method. Edges were then classified as step edges, bar edges, or line edges according to the shape of the Gaussian difference signature, which was the function mapping normalized derivative scale to response magnitude. Having classified the edges, the contrast and scale of a step edge or line edge was found by matching the edge's Gaussian difference signature to a template. The scale and width of a bar edge was found by matching the edge's apparent width signature to a template. The apparent width signature gave the distance between the two edges within the bar, as a function of operator scale. This approach differs quite substantially from the GDT model, because the templates in the GDT model are purely spatial, and the scale is found by finding the best-fitting 2D template across scale. In Zhang & Bergholm's algorithm, the templates are 1D functions of scale.

2.4.2 Conclusions

The GDT model appears to be unique in that fact that it uses template matching to detect and locate edges and to characterize them in terms of orientation and scale. The closest previous algorithm was that of Lindeberg (1996, 1998). Lindeberg presented his algorithm in terms of normalized Gaussian derivatives, rather than templates, but a formally equivalent template-based algorithm could be described. Even so, the template-like behaviour of the model was only used to select the local scale. The position and orientation of edges was found using conventional derivatives, which were not normalized in a way that would make them equivalent to template operators. The GDT model is the first model of edge detection in humans to make use of template matching.

2.5 Appendix: a deeper look at the edge strength measure in the 2D model

Section 2.3.4 presented a very simple argument that the normalization factor for converting the template response, T, into edge strength, S, should be the same in both 1D and 2D models. The argument was so simple that one might suspect that a sleight of hand had taken place. This section derives the normalization factor in the 2D model from first principles, in an analogous way to the derivation in the 1D model.

With the 1D model, the aim was to find the virtual template which fitted best to the edge's 2nd derivative. For a Gaussian edge stimulus, the best-fitting template would fit exactly to the 2nd derivative of the stimulus, since the virtual template is itself the 2nd derivative of a Gaussian edge. In the 2D model, the templates are Gaussian along the length of the edge, whereas the ideal edge stimulus is flat along its length. Therefore, although the edge's 2nd derivative and the virtual template may be perfectly correlated along any line perpendicular to the edge, the correlation will be poor when sampled over a 2D area. The model therefore needs to find the correlation between the edge profile and the virtual

template that would have occured if the template did not vary along the length of the edge. This can be achieved by expressing the virtual template as the product of 1D functions that vary in the horizontal and vertical directions.[15] Substituting $\sigma_v(n)$ for σ in (2.36),

$$v(x,y;n,0) = v_H(x;n)v_V(y;n),$$

where

$$v_H(x;n) = \frac{-x}{2\pi\sigma_v^4(n)}\exp\left(\frac{-x^2}{2\sigma_v^2(n)}\right) \tag{2.41}$$

and

$$v_V(y;n) = \exp\left(\frac{-y^2}{2\sigma_v^2(n)}\right). \tag{2.42}$$

We therefore want to find the correlation between the edge's 2nd derivative along the horizontal direction and v_H, the horizontal profile of the virtual template. This can be found by normalizing T, as explained in the remainder of this section.

As with the 1D case, we rewrite (2.39) as

$$T(i,j;n,0) = \sum_{\beta\in B}\sum_{\alpha\in A} u(\alpha,\beta;n,0)I''(\alpha)\delta x\delta y, \tag{2.43}$$

where $u(\alpha,\beta;n,0)$ is $v(\alpha-i,\beta-j;n,0)$, the discrete version of the vertical virtual template centred on (i,j). Since we are considering a vertical edge, I'' can be indexed with just the horizontal coordinate. u, can be written as the product of 1D functions, u_H and u_V, that vary in the horizontal and vertical directions:

$$u_H(\alpha;n) = v_H(\alpha-i;n)$$

---

[15] The analysis presented here considers the simplest case of a vertical edge and vertical virtual template, but could be generalized to parallel edges and operators of any orientation, simply by using the same analysis in a different coordinate system. It has already been argued, in section 2.3.4, that the peak will occur in a channel aligned with the edge, so it is unnecessary to consider the case of edges and operators that do not have the same orientation.

and

$$u_V(\beta;n) = v_V(\beta - j;n).$$

Replacing u with the product of u_H and u_V in (2.43) gives

$$T(i,j;n,0) = \sum_{\beta\in B}\sum_{\alpha\in A} u_H(\alpha;n)u_V(\beta;n)I''(\alpha)\delta x\delta y.$$

Applying the distributive rule twice gives

$$T(i,j;n,0) = \left(\sum_{\beta\in B} u_V(\beta;n)\delta y\right)\left(\sum_{\alpha\in A} u_H(\alpha;n)I''(\alpha)\delta x\right). \tag{2.44}$$

From (1.12), the correlation between I'' and the horizontal profile of the virtual template centred on (i,j) is given by

$$r(i,j;n,0) = \frac{\sum_{a\in A} u_H(\alpha;n)I''(\alpha)}{\sqrt{\sum_{a\in A} {u_H}^2(\alpha;n)\sum_{a\in A} I''^2(\alpha)}}. \tag{2.45}$$

From (2.44) and (2.45), we have

$$r(i,j;n,0) = \frac{T(i,j;n,0)}{\sum_{\beta\in B} u_V(\beta;n)\delta y\delta x\sqrt{\sum_{a\in A} {u_H}^2(\alpha;n)\sum_{a\in A} I''^2(\alpha)}}.$$

As with the 1D model, we should drop the $\sum_{a\in A} I''^2(\alpha)$ term, so that information about contrast is not lost. To maintain independence of image resolution, we again need to multiply by $\sqrt{\delta x}$, because $\sum_{a\in A} I''^2(\alpha)$ is proportional to $1/\sqrt{\delta x}$. This gives an edge strength, $S(\cdot,\cdot;n,0)$ in vertical channel n given by

$$S(i,j;n,0) = \frac{T(i,j;n,0)}{\sum_{\beta\in B} u_V(\beta;n)\delta y \sqrt{\sum_{\alpha\in A} {u_H}^2(\alpha;n)\delta x}}. \tag{2.46}$$

This is equivalent to the edge strength in the 1D model: the peaks in this function indicate the positions and blurs of the edges, and the height can be used to recover the edge contrast.

Assuming infinite image resolution, we can derive an analytical formula for the denominator of (2.46).

As $\delta x, \delta y \to 0$, $\sum_{\beta\in B} u_V(\beta;n)\delta y \sqrt{\sum_{\alpha\in A} {u_H}^2(\alpha;n)\delta x} \to \int_{y\in\mathbb{R}} u_V(y;n)dy \sqrt{\int_{x\in\mathbb{R}} {u_H}^2(x;n)dx}$.

This gives us the continuous form of (2.46):

$$S(x,y;n,0) = \frac{T(x,y;n,0)}{\int_{y\in\mathbb{R}} u_V(y;n)dy \sqrt{\int_{x\in\mathbb{R}} {u_H}^2(x;n)dx}}. \tag{2.47}$$

The denominator of (2.47) is evaluated as follows. First, consider the vertical profile of the virtual template:

$$\begin{aligned}\int_{y\in\mathbb{R}} u_V(y;n)dy &= \int_{y\in\mathbb{R}} v_V(y;n)dy \\ &= \int_{y\in\mathbb{R}} \exp\left(\frac{-y^2}{2{\sigma_v}^2(n)}\right)dy.\end{aligned}$$

This is a unit-area Gaussian multiplied by $\sigma_v(n)\sqrt{2\pi}$. Thus,

$$\int_{y\in\mathbb{R}} u_V(y;n)dy = \sigma_v(n)\sqrt{2\pi}. \tag{2.48}$$

Next, consider the horizontal profile of the virtual template:

$$\int_{x\in\mathbb{R}} {u_H}^2(x;n)\,dx = \int_{x\in\mathbb{R}} {v_H}^2(x;n)\,dx$$
$$= \int_{x\in\mathbb{R}} \frac{x^2}{4\pi^2{\sigma_v}^8(n)} \exp\left(\frac{-x^2}{{\sigma_v}^2(n)}\right) dx.$$

Making use of (2.29),

$$\int_{x\in\mathbb{R}} {u_H}^2(x;n)\,dx = \frac{1}{8\pi^{3/2}{\sigma_v}^5(n)}. \tag{2.49}$$

From (2.48) and (2.49) we have

$$\int_{y\in\mathbb{R}} u_V(y;n)\,dy \sqrt{\int_{x\in\mathbb{R}} {u_H}^2(x;n)\,dx} = \frac{1}{2\pi^{1/4}{\sigma_v}^{3/2}(n)}. \tag{2.50}$$

Finally, (2.47) and (2.50) yield an analytical formula for the edge strength in the 2D model:

$$S(x,y;n,0) = 2\pi^{1/4}{\sigma_v}^{3/2}(n)T(x,y;n,0). \tag{2.51}$$

Equation (2.51) is identical to (2.40), except for the extra spatial and orientation parameters, confirming that the normalization factor is the same in both 1D and 2D models.

Chapter 3
Implementing the GDT model

3.1 Introduction

The 1D template model was implemented using MATLAB version 6.0, release 12. The stimuli used in the experiments generally had a circular spatial envelope. Since the 1D model has only one spatial dimension, the stimulus for the model was taken to be the horizontal line of pixels halfway up the stimulus, which corresponded to the widest part of the spatial envelope in all experimental stimuli that did not have rectangular windows. Most of the experiments required subjects to make a judgement about the blur or contrast of the central edge. The model searched for the peak in the central spatial position of the edge strength representation and determined the edge's perceived blur from the blur channel in which the peak occurred, using equations (2.6) and (2.18). Perceived contrast was calculated from the height of the peak, using equation (2.35).

3.2 Basic parameters

The 1D model described in Chapter 2 has a small number of parameters. These are listed in Table 3.1, along with the values used for the simulations. With the m parameter set to 0.25, the smallest operator in the first derivative stage had a scale of 1 pixel, which gave an acceptable approximation to a Gaussian 1st derivative shape (see Figure 3.1). With an operator this size, there is a negligible error in gradient estimation due to sampling (see Figure 3.2). The image resolution of 240 pixels per deg visual angle gave rise to a minimum template scale of 1 arc min (4 pixels). There were 256 blur channels, with the ratio of largest to smallest scales being 128. This meant that each channel's scale differed

from the next one by a factor of $128^{1/(256-1)} \simeq 1.02$, or 2%. (see section 2.2.1). Thus, the scale of the edge differed from that of the best-fitting virtual template by at most 1%.

Table 3.1. The basic parameters used in the implementation of the GDT model.

Parameter	Value
Image resolution	240 pixels per deg visual angle.
N, the number of channels	256
$\sigma_t(1)$, the scale of the smallest template	4 pixels
$\sigma_t(N)/\sigma_t(1)$, the ratio of the largest channel scale to the smallest channel scale	128
m, the ratio of gradient operator scale to template scale in each channel, defined in equation (2.8)	0.25

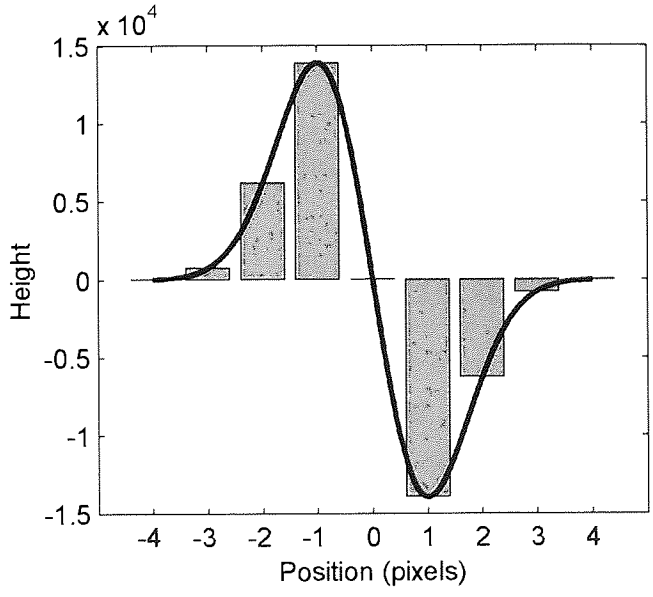


Figure 3.1. The bars show the central 9 pixels of a Gaussian 1st-derivative operator with scale equal to 1 pixel, the smallest operator scale used in the simulations. The height of the operator is scaled to give the gradient in units/deg when the image resolution is 240 pixels/deg. The curve shows the corresponding continuous function.

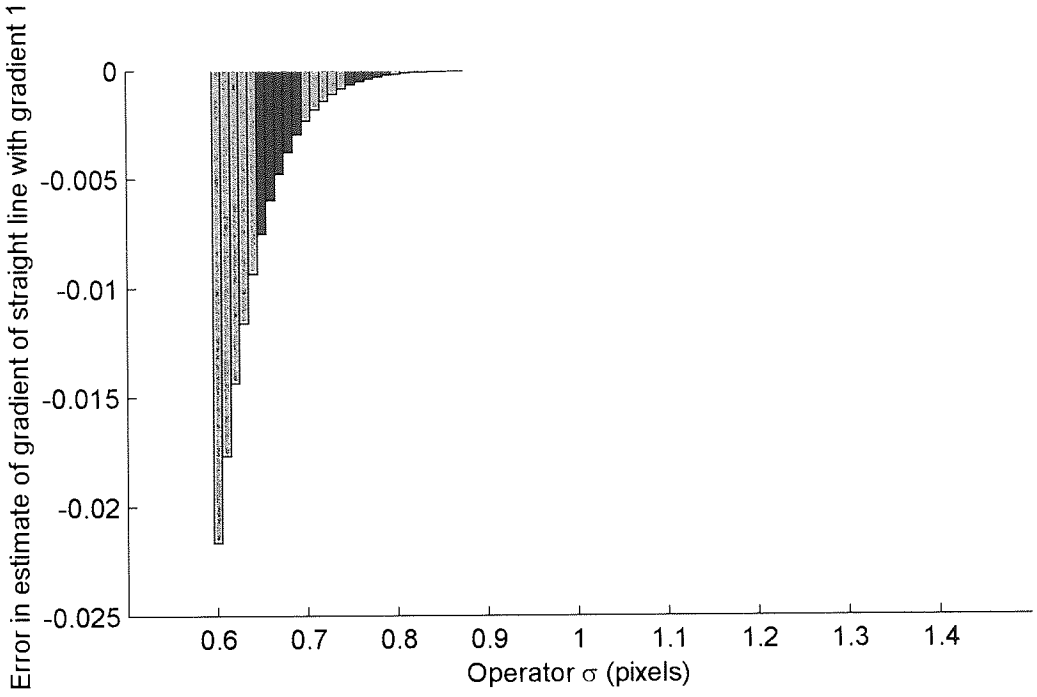


Figure 3.2. Gaussian 1st-derivative operators were constructed with scale (σ) ranging from 0.6 to 1.4 pixels in steps of 0.01 pixels. Each operator was 31 pixels wide. These operators were used to estimate the gradient of a straight line with gradient 1. The graph shows the estimated gradient minus 1.

3.3 Operators

3.3.1 Combining operators

In Chapter 2, the model was presented as having 3 stages of convolutions: 1st derivative, 2nd derivative, and template. In each stage, the output of the previous stage is convolved with a Gaussian 1st-derivative operator (to use convolution, rather than filtering in the template stage, the operator must be inverted). These stages are kept separate in the description of the model to emphasise the idea that the template is being compared with the 2nd derivative signature; in practise, the final two operators can be combined into a single Gaussian 2nd-derivative operator. If the template has scale σ_t, then the gradient operator that precedes it has scale $m\sigma_t$, and the scale of the combined operator will be

the Pythagorean sum of the individual scales, $\sigma_t\sqrt{1+m^2}$. The combined operator thus has the following form:

$$f(x)=\frac{-1}{\sigma^3\sqrt{2\pi}}\left(\left(\frac{x}{\sigma}\right)^2-1\right)\exp\left(\frac{-x^2}{2\sigma^2}\right),$$

where $\sigma=\sigma_t\sqrt{1+m^2}$. This operator is a Gaussian 2nd-derivative operator multiplied by -1, the inversion being necessary because the template operator is inverted. The output of the rectification stage can be either filtered or convolved with this combined operator (the combined operator has even symmetry, so the reversal of the elements that distinguishes convolution from filtering makes no difference). Combining the two operators into one caused a significant increase in the speed of the model. It was not possible to combine all three operators into a single operator, because the first two operators are separated by the rectifying nonlinearity.

3.3.2 Spatial extent of operators

Initially, it seemed appropriate to make each operator's size a constant multiple of its own scale, but this proved to be problematic. The number of pixels in an operator could only be an approximate multiple of the scale, because it had to be rounded to the nearest integer. Since neighbouring channel scales differed by a factor of 2%, the smallest channels differed by a small fraction of a pixel. Thus, as the scale increased, the number of pixels in the operators would stay the same for several channels. This meant that the width that was being sampled from the operator was a progressively smaller multiple of its scale, and so the magnitude of the gradient would be underestimated by a progressively larger amount. Eventually, the channel scale would reach a value that caused the operator width to step up by one pixel on each side, and there would be a sudden jump in the ratio of operator width to scale, causing a sudden jump in the estimated gradient (see Figure 3.3). These jumps were very small, but were still large enough to give rise to detectable spurious peaks in the edge strength.

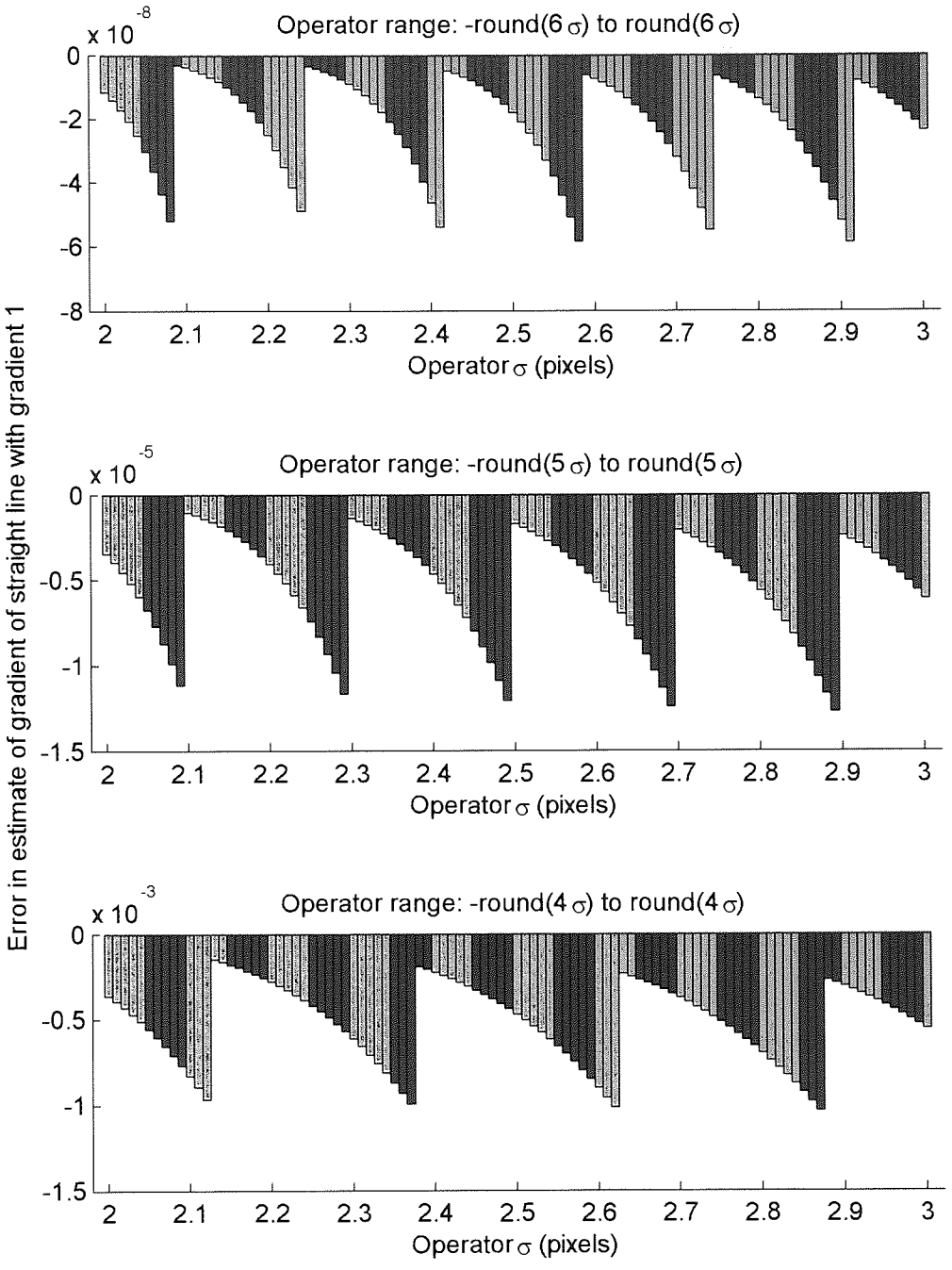


Figure 3.3. Gaussian 1st-derivative operators were constructed with scale (σ) ranging from 2 to 3 pixels in steps of 0.01 pixels. Taking the central pixel position to be 0, each operator extended spatially from -round($n\sigma$) pixels to round($n\sigma$) pixels, where n = 4, 5, or 6, and round(x) is x, rounded to the nearest integer. These operators were used to estimate the gradient of a straight line with gradient 1. The graphs show the error, i.e. the estimated gradient minus 1. As σ increases, the gradient is underestimated by a progressively larger amount, until the operator size steps up by 1 pixel on each side. At this point, there is a jump in the estimated gradient. This jump occurs whenever $n\sigma = m + 0.5$, for $m \in \mathbb{Z}$, because these values of $n\sigma$ mark the boundary between two different values of round($n\sigma$).

The solution to this problem was to keep the operator width the same across channels. At each convolution stage of the model, the scale, σ_{max}, of the operator in the largest channel, was determined. Taking the central pixel position to be 0, the operator in each channel extended spatially from $-\text{round}(4\sigma_{max})$ pixels to $\text{round}(4\sigma_{max})$ pixels, where $\text{round}(x)$ is x, rounded to the nearest integer. Whenever an operator extended over the edge of the stimulus, the area outside of the stimulus was taken to have a zero value for the purposes of the convolutions.

3.4 Finding peaks in the edge strength

The problem of finding peaks in the 2D scale-space edge strength representation was nontrivial. At first sight, it might seem that all that was required was to take each pixel and see if its value was larger than that of its eight immediate neighbours. The problem with this approach is that a symmetrical peak centred exactly on the border between two pixels would give rise to an adjacent pair of identically valued pixels. This occurred very frequently because the stimuli always had an even number of pixels, so the centre of the stimulus, on which the edges were generally centred, lay exactly on the border between the two central pixels. Because of this, a peak was defined as any cluster of equally-valued pixels that extended not more than 2 pixels horizontally, vertically, or diagonally, and which had a value greater than that of any of the pixels adjacent to the cluster. This gave rise to the ten different configurations shown in Figure 3.4. The peak-finding algorithm simply searched for any of these configurations. MATLAB proved to be too slow to implement this algorithm, so it was written in C, and compiled as a mex file, which could be called from MATLAB.

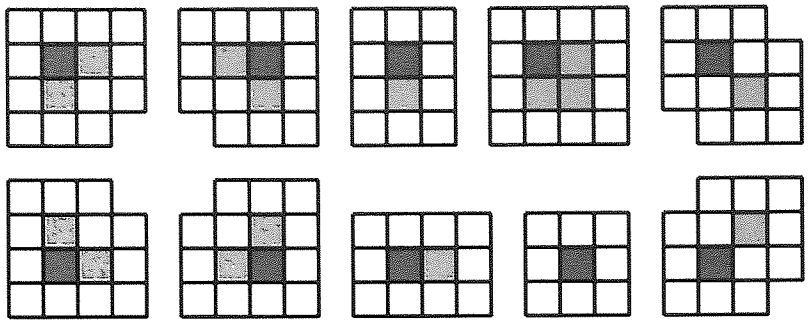


Figure 3.4. The ten configurations of pixels that were counted as peaks. In each configuration, all the grey pixels have the same value, which is greater than that of any of the white pixels; the dark grey pixel indicates the pixel that was chosen as the location of the peak.

Before searching for the configurations shown in Figure 3.4, the edge strength representation was surrounded by a border, 1 pixel wide, which had a value which was a negative number with a very large magnitude, lower than any plausible edge strength. This meant that a peak was marked for any configuration of pixels at the border of the image which might have been a peak, if the image had extended further. This liberal criterion was necessary because, if the model was given an edge whose blur was below that of the smallest channel, the height of the scale-space representation would increase up to the lowest channel, but would not decrease before the border of the scale-space representation had been reached. The very low border ensured that any edge sharper than the smallest channel would be marked in the smallest channel, rather than not at all. In such a case, the estimated blur would obviously be higher than the veridical blur; in addition, the contrast would be over-estimated because the peak height would be multiplied by an inappropriately large scaling factor, as described by equation (2.35).

3.5 Spurious peaks

Figure 3.5 shows the edge strength representation from the positive channel in response to a Gaussian edge with scale 16 arc min and contrast 1. There is one obvious peak in the centre, which corresponds to the correct blur and position of the edge. This peak has a height of 5.45, which, when scaled according to equation (2.35), gives an estimated

contrast of 1.01, which is acceptably close to the correct value (the slight overestimation in contrast occurs because the blur is slightly overestimated, at 16.1 arc min, an error of less than 1%). But, to the left and right, there is a multitude of tiny peaks, which cannot be seen by looking at the image (Figure 3.5(c)), but which are genuine peaks nonetheless. There are two pairs of isolated peaks along the bottom of the image (in the smallest channel). Within each pair, the heights are 2.50×10^{-5} and 4.43×10^{-6}; the maximum height among the remaining peaks is 5.58×10^{-14}. The two isolated pairs of peaks in the smallest blur channel most probably arise because of the liberal criterion for marking peaks discussed in section 0: where there is doubt, because a configuration is on the border of the template output, a peak is marked, and this can lead to the inappropriate marking of peaks. The very low height of these peaks (equivalent to an edge contrast of the order of 10^{-8}) could allow them to be rejected. None of the remaining peaks are on the border, so another explanation must be sought.

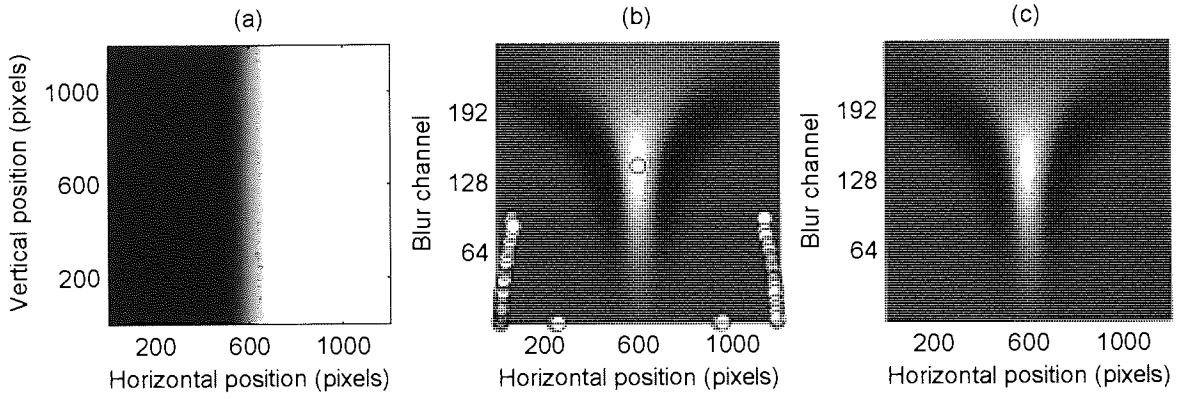


Figure 3.5. (a) The stimulus, a vertical Gaussian edge with scale 16′, and contrast 1. The image spanned 1200 pixels left-to-right, which corresponded to 5 deg of visual angle. (b) The edge strength representation in the positive channel (the channel that detects edges that are dark on the left); the white disks mark all the edges found in this channel. (c) The same edge strength representation without the edges marked. The model was run with the parameter values given in Table 3.1.

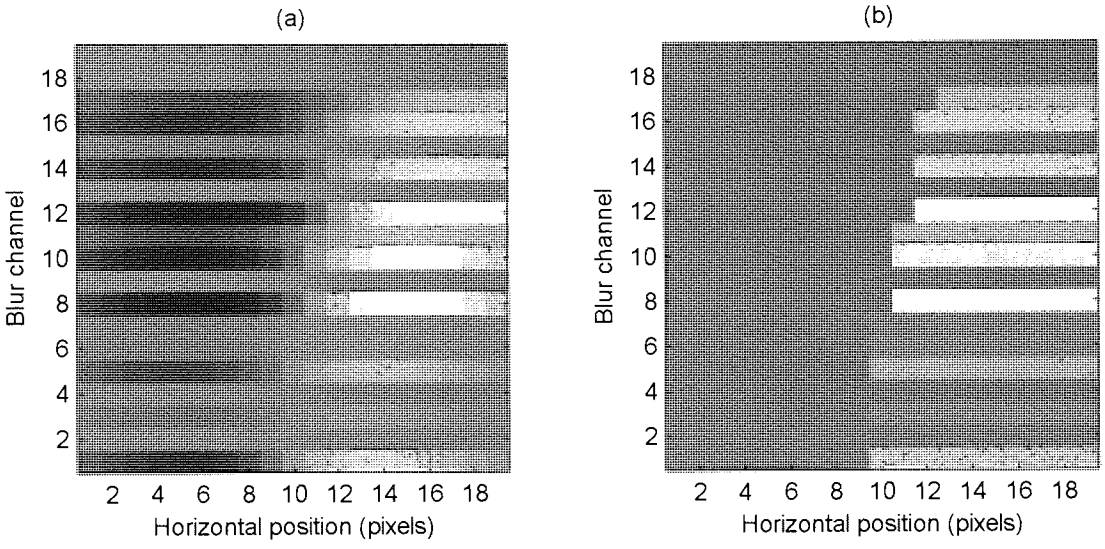


Figure 3.6. (a) An enlargement of the bottom-left corner of Figure 3.5(c), with brightness re-scaled so that black represents -4.64×10^{-14} and white represents 4.64×10^{-14}. (b) The corresponding region of the rectified 1st derivative, with white representing 1.01×10^{-14}. In both images, blur channels 2, 4, 9, 13, 18, and 19 have a value of exactly zero at each spatial position.

As stated in footnote 14, on page 119, the areas outside of the stimulus are taken to have a zero value for the purpose of the convolutions. This means that the left and right borders of the stimulus effectively contain negative step edges. When a 1st derivative operator is close to the border of the stimulus, it should be affected much more by the negative step edge than by the positive central edge, and should have a negative value that is set to zero by the rectifier. Figure 3.6(a) shows an enlargement of the bottom-left corner of Figure 3.5(c), with the brightness rescaled to reflect the range of values in this region; Figure 3.6(b) shows the corresponding region of the rectified 1st derivative. In some blur channels, the rectified 1st derivative has a zero value, as expected, but, in other channels, there is a very small positive value. These non-zero channels in the rectified 1st derivative match up perfectly with the non-zero channels in the template output, in which the spurious peaks occur.

The fact that some gradient operators were giving small positive responses in the vicinity of a negative edge suggested that the spurious peaks might be resulting from rounding errors in MATLAB's arithmetic functions, which overcame the negative response that should have occurred: in some channels the response would stay negative, and would be

set to zero by the rectifier, while in other channels, the response would become positive, leading to the strips of non-zero response levels in selected channels in Figure 3.6. This view was supported by the fact that the channels with non-zero responses seemed to be distributed randomly. This possibility was investigated by running the MATLAB function SumOp, whose listing is shown in Box 3.1. This function creates the central 9 pixels of the gradient operator shown in Figure 3.1, and sums its elements in two different orders. Since the operator has odd symmetry, its elements should sum to zero. The first 2 lines set up the scale of the operator (which is 0.25 times the minimum template SD), and line 3 calculates dx, the width of a pixel in deg visual angle. Line 4 creates a vector of 9 values, which step upwards in steps of dx, with the central value (x(5)) equal to 0. Line 5 creates the gradient operator, g, which is the first derivative of a unit-area Gaussian curve. Elements of g are indexed using brackets. Line 6 creates a variable, sum1, which is the sum of the elements of g, made by taking pairs of corresponding elements equidistant from the central element, adding each pair together, and then summing the results. Line 7 creates a variable, sum2, by adding the same elements together from right to left. Each sum is multiplied by dx, as in the convolution described in equation (2.14).

Box 3.1 Listing of the MATLAB function SumOp

```
function SumOp

m = 0.25;
SD = m*1/60;
dx = 1/240;

x = (-4:4)*dx
g = Gx(x,SD)
sum1 = ((g(1)+g(9)) + (g(2)+g(8)) + (g(3)+g(7)) + (g(4)+g(6)) + g(5))*dx
sum2 = (((((((g(9) + g(8)) + g(7)) + g(6)) + g(5)) + g(4)) + g(3)) + g(2)) + g(1))*dx

function y = Gx(x,s)
% 1st derivative of a unit area gaussian
y = (-x/((s*s*s)*sqrt(2*pi))).*exp(-(x.*x)/(2*s*s));
```

After running SumOp, the value of sum1 is exactly 0, whereas the value of sum2 is -5.12×10^{-15}, which has a similar order of magnitude to the incorrect positive values in Figure 3.6(b). This failure of the rules of commutativity and associativity of addition came as a surprise initially but, with hindsight, it is perfectly understandable. Real numbers are represented in computers in a form such that a number and its additive inverse always sum to exactly 0; the numbers in each bracketed pair that form sum1 are additive inverses of each other, so sum1 is exactly 0. But adding numbers that are not related in this way can lead to small rounding errors, so that adding the numbers in a different order can give a different result.

The magnitude of the rounding error is exactly the level that would be expected to give rise to the spatial position of the incorrect positive responses shown in Figure 3.6(b). The positive response in channel 1 begins when the centre of the operator is 10 pixels from the border. The height of this operator 10 pixels from its centre is $\pm4.43\times10^{-17}$, so an error of the order of 10^{-15} would be expected to overcome the small negative value expected theoretically, causing the unexpected perturbations in the template output. The height of this operator 9 pixels from its centre is $\pm5.33\times10^{-13}$. When the operator is 9 pixels from the border, its genuine negative response is large enough not to be wiped out by the rounding errors.

In conclusion, the evidence is very strong that the spurious peaks in Figure 3.5 are caused by rounding errors in MATLAB's arithmetic functions: the rounding errors have been shown to exist, and are of the same order of magnitude as the incorrect positive 1st-derivative responses that occur in the vicinity of a negative edge. The channels in which these positive responses occur are the channels in which the spurious peaks occur, and the positive responses occur in spatial positions where the expected negative response has an order of magnitude which is lower than the order of magnitude of the rounding errors. Two strategies were used to remove these peaks:

1. The spurious peaks in Figure 3.5 were all less than 1×10^{-13} so, in all the simulations, peaks with a height smaller than this were ignored.
2. In all the experiments, the stimuli had several edges in different locations, and the subject was asked to judge the central edge. In simulating these experiments, the

only spatial positions that were examined were the two central pixels. As Figure 3.5 indicates, the spurious peaks occurred away from the centre: this is because the template response surface was sufficiently steep in the central spatial positions to overcome any slight perturbation caused by rounding errors. Thus, spurious peaks did not occur in the central positions, and could be ignored by restricting the search to the two central pixels.

Chapter 4
General psychophysical methods

4.1 Subjects

Three subjects were used: KAM, MAG and PAA. All had corrected-to-normal vision. KAM performed all the experiments; MAG performed all the main experiments; PAA performed all the main experiments, except experiments 6 and 7.

4.2 The matching paradigm

All the experiments except experiment 3 were aimed at finding the perceived blur or perceived contrast of an edge. The edge whose blur or contrast was being assessed was called the *test edge*. The perception of the test edge was assessed by taking a *comparison edge* (always a Gaussian edge) and adjusting its scale (or contrast) to find the scale (or contrast) at which it matched the test edge in the required perceptual dimension.

4.2.1 Responses

The matching experiments used a 2-interval forced-choice method. On each trial there were two intervals separated in time: one interval contained a test edge, and the other contained that test edge's comparison edge (the order was randomized). Subjects were asked to indicate which stimulus appeared more blurred (in a blur matching experiment) or which had the higher contrast (in a contrast matching experiment). Responses were collected using a mouse: clicking the left button indicated that the edge with the higher

blur/contrast had appeared in the 1st interval; clicking the right button indicated that it had appeared in the 2nd interval. Responses made while the stimulus was being displayed had no effect. The next trial started a short time after the response.

4.2.2 Staircases

The scale/contrast of each comparison edge was controlled by a staircase. The staircase reduced the comparison edge scale/contrast after trials in which the comparison edge blur/contrast was judged higher than that of the test edge, and increased the comparison edge scale/contrast after trials in which it was judged to be lower. This scheme converges on a comparison edge scale/contrast at which the comparison edge is equally likely to be judged higher or lower than the test edge (Wetherill & Levitt, 1965; Levitt, 1971), and can be considered to match the test edge in the perceptual dimension being judged.

The step sizes were measured in dB, where an increase of 1 dB is multiplication by a factor of $10^{1/20}$, and a decrease of 1 dB is division by this factor. The step size started off at 8 dB, and halved with each reversal of direction, until it had reached 1 dB. The staircase terminated immediately after the 12th reversal of direction.

A staircase like this usually lasted for about 25 trials, which took about 1 minute to complete. Several staircases were interleaved within a session. On each trial, the staircase was selected according to the following algorithm, suggested by Meese (personal communication). For each staircase, the algorithm finds the number of reversals remaining, and then assigns each staircase a probability of being selected, which is equal to the number of reversals remaining on that staircase, as a proportion of the total. So, if $P(i)$ is the probability that staircase i is selected, R_i is the number of reversals remaining on staircase i, and N is the number of staircases, then

$$P(i) = \frac{R_i}{\sum_{r=1}^{N} R_r}.$$

Thus, the closer a staircase is to terminating, the lower its probability of being selected. When it has no reversals left (i.e. it has terminated), its probability of being selected is zero. This method causes all the staircases finish at approximately the same time. An alternative would have been simply to choose, on each trial, the staircase with the lowest number of reversals remaining, but this can sometimes lead to one staircase being selected for several consecutive trials, until the next reversal occurs. The random element in the chosen method makes this less likely.

4.2.3 Data analysis

The purpose of the experiments was to find the comparison edge scale/contrast for which the comparison edge appeared to have the same blur/contrast as the test edge. This corresponds to the scale/contrast at which the comparison edge has a 0.5 probability of being selected as having a higher blur/contrast than the test edge. One common way to find this value is simply to average the stimulus values for the peaks and valleys that occur over the sequence of trials (Wetherill & Levitt, 1965). The method chosen here is to find, for each comparison scale/contrast level, the proportion of trials for which the comparison stimulus was judged higher than the test edge, and to fit a psychometric function to these data. Thus, the data analysis was much like the analysis of a method-of-constant-stimuli paradigm, with the staircase employed merely to find a suitable range of stimulus values around the 50% point of the psychometric function.

The psychometric function fitted to the data was a Gaussian integral. The horizontal axis of the psychometric function represented the comparison edge scale/contrast level (in 1 dB steps), and the vertical axis represented the proportion of times a comparison edge at each scale/contrast level was selected as being higher than the test edge. Because the horizontal axis was a logarithmic function of the physical stimulus level, this analysis assumed that the internal perceptual values are an approximately logarithmic function of the corresponding physical values. This assumption is supported by the finding that, at above-threshold levels, blur discrimination thresholds obey Weber's law (Hess *et al.*, 1989; Wuerger *et al.*, 2001), suggesting that a constant *difference* in internal values corresponds to a constant *ratio* of physical values. Contrast discrimination thresholds also increase

with stimulus contrast, but the exponent is less than 1 (Legge, 1981) – a near-miss to Weber's law.

The psychometric function was fitted to the data using psychofit, version 2.5.1 (recently renamed to psignifit; see http://bootstrap-software.org/psignifit/), a software package which implements the maximum-likelihood method described by Wichmann & Hill (2001a). This software generates a psychometric function, ψ, which maps the comparison stimulus scale/contrast, x (in dB), onto the probability that a comparison edge with scale/contrast x will be selected as being higher than the test edge. ψ has 4 parameters, $\alpha, \beta, \gamma, \lambda$, and is described in general as follows:

$$\psi(x;\alpha,\beta,\gamma,\lambda) = \gamma + (1-\lambda-\gamma)F(x;\alpha,\beta). \tag{4.1}$$

F is a 2-parameter sigmoid function ranging from 0 to 1 (in this case, a Gaussian integral); α and β are the parameters of F, fitted to the data using the maximum-likelihood method. γ and λ are fitted parameters that determine the range of values of ψ: ψ varies from γ to $(1-\lambda)$. In the case of the matching paradigm, ψ would be expected to vary from close to 0 (for values of x for which the comparison edge is rarely judged to be higher than the test edge) to close to 1 (for values of x for which the comparison edge is usually judged to be higher than the test edge). Because of this, the fitted values of γ and λ are constrained to fall between 0 and 0.05. Wichmann & Hill (2001a) note that these parameters play an important role in avoiding biases caused by response errors (i.e responses that differ from the intended response, or responses that are essentially random because the subject failed to see the stimulus due to a lapse of concentration). Without the γ and λ parameters, ψ would vary from 0 to 1, which incorrectly assumes that response errors are highly unlikely. If a response error occurs when the blur/contrast of the comparison edge is clearly different from that of the test edge, then the fitted psychometric function may have to be drastically distorted to accommodate this event (since the probability of this event has been greatly underestimated). Wichmann & Hill (2001a) reported the results of a simulation in which a psychometric function was fitted to data from 6 stimulus levels, with 50 trials at each level. They showed that, if a response

error occurred on just one trial at the highest stimulus level, the slope of the fitted psychometric function was approximately halved.

In summary, F can be considered to be the underlying psychometric function, and ψ is the actual psychometric function, after response errors have been taken into account. Since we are interested in the underlying psychometric function, rather than the response errors, the matching comparison scale/contrast level is determined from the 50% point of F, rather than ψ. Specifically, the scale/contrast level at which the comparison edge is considered to match the test edge is taken to be $F^{-1}(0.5)$, where F^{-1} is the inverse of the function F.

Having found the matching comparison edge level, 5% and 95% confidence limits were found using the percentile bootstrap method implemented by psychofit (see Wichmann & Hill, 2001b); each confidence limit was based on 10000 simulations. All error bars shown on the graphs in the following chapters indicate bootstrap confidence limits generated with this method. One important point made by that Wichmann & Hill (2001b), regarding the interpretation of bootstrap confidence limits, is that, unlike standard errors, bootstrap confidence limits do not give an estimate of the variability in the subject's performance. Indeed, the crucial *bridging assumption* underlying the bootstrap method is that the psychometric function underlying the subject's performance is the same on every trial. For each simulation, the bootstrap algorithm takes the psychometric function that was fitted to the psychophysical data and uses this to generate simulated data, using the same sampling scheme as used in the real experiment (i.e. the same stimulus levels and the same number of trials at each level). Then, a new psychometric function is fitted to the simulated data, and $F^{-1}(0.5)$ is found. The confidence limits are based on the distribution of values of $F^{-1}(0.5)$ fitted to the simulated data sets. Thus, the confidence interval indicates the expected variability in estimated match points assuming that, on each trial, the subject's underlying psychometric function is the same. Because they do not take into account the variability in the subject's performance, the bootstrap confidence limits do not delineate the boundary within which the "true" value is likely to fall; rather, they indicate how accurately the match point can be estimated from the given quantity and distribution of data across the stimulus levels.

Despite the caveat about the interpretation of the bootstrap confidence limits, a very large confidence interval might still indicate unreliable performance. This is because very noisy data (that would be expected from unreliable performance) will give rise to a psychometric function that has a shallow slope, and a psychometric function with a shallow slope will give rise to a wide distribution of $F^{-1}(0.5)$ values across the simulated data sets. Thus, a large bootstrap confidence interval might be an indirect indicator of unreliable performance, but it might also indicate completely consistent performance based on an underlying psychometric function that is very shallow.

The psychometric functions were usually fitted to data collected from 8 staircases, spread over two or more sessions. In general, each subject's data were analysed twice: in one analysis, the data were separated according to the temporal order of test and comparison edges in a trial (giving about 100 trials per psychometric function); in the other, the data were collapsed across stimulus order before fitting the psychometric functions (giving about 200 trials).

4.3 Apparatus

The experiments were run on a Pentium III PC computer with a VSG 2/3 graphics card (made by Cambridge Research Systems). The VSG software installation was version 6.085. The VSG card has an 8-bit frame store that stores the images. These 8-bit values are mapped onto 15-bit values using look-up tables (LUTs), and converted to analogue outputs using two 8-bit digital-to-analogue converters (DACs). The analogue output from one DAC is multiplied by 128, using analogue electronics, and is electronically added to the values from the other DAC, giving pseudo-15-bit grey-level resolution. Without the mapping to pseudo-15-bit grey-level values, the contrast of the image could not be reduced without reducing the image resolution. The mapping to a pseudo 15-bit resolution allows contrast to be reduced while keeping the image resolution at 8 bits so, although this system is inferior to a 15-bit frame store, it is better than an 8-bit frame store with only a single 8-bit DAC. Stimuli were displayed on an Eizo FlexScan 6600-M monitor.

Experiments were run using MATLAB with custom-written software. This software generated stimuli that were transferred to the VSG graphics card using custom-written C mex files that could be called from MATLAB. Luminance calibration was carried out with an OptiCal luminance meter (Cambridge Research systems), using the standard calibration procedure included in the VSG software library. Half maximum luminance (L_0 in equation (2.1)) was 45 cd/m^2. The graphics card was configured to generate a screen display that measured 800 pixels horizontally by 600 pixels vertically. The stimulus matrix occupied a square within this display, whose size varied across experiments (dimensions in pixels, cm and deg visual angle are given in the methods sections of the individual experiments). The area outside the stimulus matrix was set to a luminance of L_0. The visual stimulus did not always fill the stimulus matrix: it was often windowed to a smaller size by setting pixels within the stimulus matrix to L_0.

To generate the stimuli, the system first generated a matrix of floating-point values that corresponded to the function, I, in equation (2.1). This matrix was then linearly scaled and rounded to 8-bit frame store values. The matrix was scaled to fill the largest possible range of 8-bit frame store values such that the minimum value of I mapped onto a frame store value of 0, and the value $I = 0$ mapped onto an integer frame store value. The requirement that 0 maps onto an integer frame store value was necessary because, since the frame store holds integer values, only values of I that map onto integer frame store values can be reproduced accurately. A value of $I = 0$ corresponds to L_0, the background luminance. If this value cannot be reproduced accurately, then there will be a slight mismatch between the background and the areas within the image matrix that are supposed to have a luminance of L_0.

The contrast of the stimulus was set by generating LUTs that converted the 8-bit frame store values to values ranging from 0 to 1, which represented the luminance as a proportion of the maximum luminance. The VSG software library then converted these values to the appropriate 15-bit DAC values to achieve the desired luminance for each grey level. All stimuli were invariant over time in every aspect except their contrast. The contrast had a temporal profile that was flat for the central 250 ms, and flanked by raised cosine edges of half-period 25 ms, so the entire stimulus lasted for 300 ms. These stimuli could be generated with a single stimulus matrix, and the contrast was varied over time by

generating a sequence of LUTs. In the following chapters, when the stimuli are described, the temporal aspect is ignored and, whenever the contrast of a stimulus is given, the contrast referred to is the contrast during the flat central period of the stimulus.

Chapter 5
Effect of contrast on perceived blur

5.1 Introduction

In section 2.2.7 it was proved that the blur of a Gaussian edge is estimated veridically by the GDT model: the blur value that the model assigns to a Gaussian edge is equal to that edge's scale, and is therefore independent of the edge contrast. This prediction is contradicted by Georgeson's (1994) finding that edges look sharper when their contrast is reduced. This was a potentially serious problem for the model and, since there had been no other reports of the effect of contrast on perceived blur, experiment 1 attempted to replicate this effect.

5.2 Experiment 1: effect of contrast on perceived blur

5.2.1 Introduction

Experiment 1 examined the effect of contrast on perceived blur for test edges with different scales, using the blur-matching procedure described in section 4.2. The experiment was divided into four separate experiments, with experiment 1.1 being the main one.

5.2.1.1 Experiment 1.1: main experiment

Experiment 1.1 had two independent variables: test edge contrast and scale. Contrast had 4 levels: 0.05, 0.1, 0.2, and 0.4; scale had 6 levels: 6, 8, 12, 16, 24, 32 arc min. The experiment thus had a 4×6 factorial design. The dependent variable was the scale of a 0.4 contrast comparison edge that matched the test edge in perceived blur.

5.2.1.2 Experiment 1.2: effect of noise surround

Experiment 1.2 was the same in every way to experiment 1.1, except that the stimulus was surrounded by binary noise with contrast 0.2 (see Figure 5.1). There were several reasons for running this additional version:

1. It kept the overall contrast of the stimuli fairly constant, despite the variation of the edge contrast, thereby limiting the possible action of contrast gain control mechanisms, which are not taken into account by the model.
2. It allowed comparison with the results of experiment 3, which used a noise surround in an attempt to make the edge of the spatial window less detectable than the central edge.
3. It was thought that the noise surround might de-emphasize the edges of the spatial window, making it easier to concentrate on the central edge.

Only subject KAM performed this experiment.

5.2.1.3 Experiment 1.3: effect of inter-stimulus interval duration

Experiment 1.3 was designed to examine the effect of inter-stimulus interval (ISI) duration, in case the afterimage from the first stimulus in a trial affected the perception of the second stimulus. Georgeson & Turner (1985) argued that afterimages are formed by a temporal integrator that has a time-span of at least 4 seconds, so it was possible that

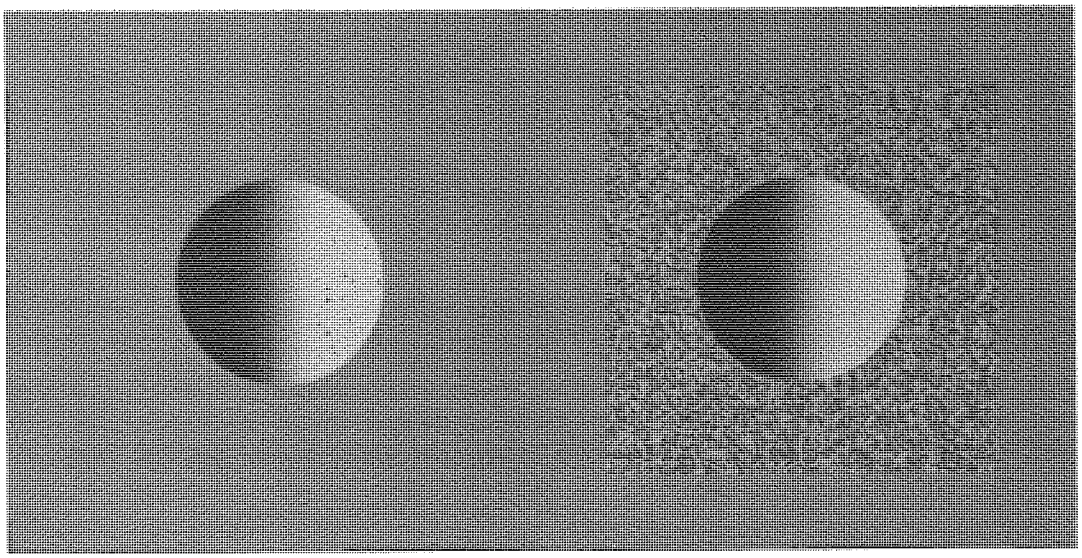
afterimages could have some effect. They found that afterimage strength was an increasing (but decelerating) power function of contrast. Georgeson & Turner used a contrast matching method to assess the perceived contrast of a stimulus presented immediately after the afterimage-inducing stimulus, and it was found that the afterimage adds nearly linearly to the subsequently presented stimulus. Thus, the largest effect of afterimages in experiment 1 should occur when the comparison edge (always at 0.4 contrast) preceded a test edge with a contrast of 0.05, and the smallest effect should occur when the temporal order of these same stimuli is reversed. To assess the impact of afterimages, the 0.05 contrast conditions from experiment 1.2, with test edge scales of 6′, 8′, 24′ and 32′, were repeated at a range of ISI durations. Only KAM performed this experiment.

5.2.1.4 Experiment 1.4: larger window diameter and longer inter-stimulus interval duration

Experiment 1.4 repeated the 12′ and 16′ test edge conditions of experiment 1.1 with a larger spatial window diameter and a longer ISI duration. The stimuli did not have a noise surround.

5.2.2 Stimuli

In all parts of experiment 1, each trial consisted of a vertical test edge and a vertical comparison edge, one following the other in a random order. The ISI between the test and comparison edges was a blank field (at mean luminance) that lasted for 500 ms in experiments 1.1 and 1.2, 1200 ms in experiment 1.4, and was varied as an independent variable in experiment 1.3.



(a) (b)

Figure 5.1. (a) An example of the stimuli used in experiment 1.1. (b) An example of the stimuli used in experiment 1.2, in which the stimuli were surrounded by binary noise. These edges have an orientation of 0°.

5.2.2.1 Test edges

Examples of the test edges are shown in Figure 5.1. The test edges were Gaussian edges, generated according to equation (5.1):

$$E(x,y) = Cw(x,y)\,\mathrm{erf}\left(\frac{x\cos(\theta) + y\sin(\theta)}{\sigma_e\sqrt{2}}\right), \tag{5.1}$$

where (x,y) are the spatial coordinates, θ is the edge orientation from vertical, σ_e is the edge scale, C is the contrast, and w is a circular spatial envelope. The orientation parameter can be used to control the edge polarity: an edge with an orientation of 0° is dark on the left; one with an orientation of 180° is dark on the right. The edge scales and corresponding spatial envelope diameters used in experiments 1.1 and 1.2 are given in Table 5.1. The perimeter of the spatial envelope was a sharp edge that was anti-aliased to avoid jaggedness. Anti-aliasing of the perimeter was achieved by making its radial profile a raised cosine with half-period equal to the diagonal distance across a pixel. Thus, the spatial envelope function, w, had the following definition:

$$w(x,y)=\begin{cases} 1 & r \le R_i \\ 0 & r \ge R_o \\ \frac{1}{2}\left(\cos\left(\frac{\pi}{p}(r-R_i)\right)+1\right) & R_i < r < R_o \end{cases} \quad (5.2)$$

where $r=\sqrt{x^2+y^2}$ (the radial distance of point (x,y) from the centre of the window), and R_i and R_o are, respectively, the inner and outer radii of the window, defined as follows:

$$R_i=(D-d)/2$$
$$R_o=(D+d)/2$$

where D is the window diameter and d is the diagonal distance across one pixel.

Table 5.1. Scales, spatial window diameters and viewing distances of test stimuli in experiments 1.1 and 1.2.

scale (arc min)	Window diameter (arc min)	Viewing distance (cm)
6	80	260
8	80	260
12	160	130
16	160	130
24	320	65
32	320	65

Stimulus matrices were arrays of 256×256 pixels, which measured 11.4×11.4 cm on the screen. As Table 5.1 indicates, the conditions were spread over three viewing distances: 260, 130 and 65 cm; at these viewing distances, the stimulus matrix subtended visual angles of, respectively, 2.5, 5, and 10 deg. The 6′, 12′ and 24′ test stimuli were physically identical on the screen, and differed in scale only because of the different viewing distances; the same applied to the 8′, 16′ and 32′ test stimuli.

5.2.2.2 Comparison edges

Comparison edges were generated according to equation (5.1). The contrast of the comparison edge was fixed at 0.4, and its scale was controlled by a staircase. Within a trial, the comparison edge always had the same orientation (i.e. polarity) as the test edge.

5.2.2.3 Noise surround in experiments 1.2 and 1.3

In experiments 1.2 and 1.3, the circular window was surrounded by binary noise with contrast 0.2, as illustrated in Figure 5.1(b). The noise was generated by dividing the 256×256 pixel stimulus matrix into 2×2 pixel squares and assigning each a random value of either 0.2 or −0.2. The binary noise function was windowed with a photographic negative of the edge window, so the full equation of the stimulus surrounded by noise is

$$E(x,y) = Cw(x,y)\,\mathrm{erf}\left(\frac{x\cos(\theta) + y\sin(\theta)}{\sigma_e\sqrt{2}}\right) + \left(1 - w(x,y)\right)N(x,y),$$

where N is the binary noise, taking values of 0.2 or −0.2.

5.2.2.4 Spatial envelope diameter and ISI duration in experiment 1.4

Experiment 1.4 repeated the 12′ and 16′ test edge conditions of experiment 1.1, with a larger spatial window diameter (300′) and a longer ISI duration (1200 ms). The stimuli did not have a noise surround. The viewing distance was 130 cm, as for these test edge scales in experiments 1.1 and 1.2.

5.2.3 Methods

5.2.3.1 Experiments 1.1 and 1.2

In experiments 1.1 and 1.2, each subject completed six sessions, two at each viewing distance. Within a session, four variables were manipulated, as described in Table 5.2. Each of the 32 combinations of levels of these variables was assigned to a different staircase. The sessions were run in the order ABCBCA, where a different viewing distance was assigned (randomly for each subject) to each of the letters A, B, and C. This ordering attempted to balance out any practice effects and, unlike the standard ABCCBA ordering, it avoided two consecutive sessions at the same viewing distance.

Table 5.2. Variables manipulated within a session of experiments 1.1 and 1.2

Variable	Number of levels per session	Possible values
Test edge scale	2	6′ or 8′ at viewing distance of 260 cm 12′ or 16′ at viewing distance of 130 cm 24′ or 32′ at viewing distance of 65 cm
Test edge contrast	4	0.05, 0.1, 0.2, or 0.4
Orientation/ polarity	2	0° (dark on left) or 180° (dark on right)
Initial comparison scale for staircase	2	random value between 10 and 14 dB above test scale or random value between 10 and 14 dB below test scale

5.2.3.2 Experiment 1.3

In experiment 1.3, the independent variables were the same as in Table 5.2, except that the test edge scales of 12′ and 16′ were not used, and the only test edge contrast level used was 0.05. In addition, there was an extra independent variable of ISI duration, which took values of 0.5, 1, and 2 seconds. Thus, within a session, there were 2 test edge scales, 1 test edge contrast, 2 orientations, 2 staircase starting levels, and 3 ISI durations, giving 24 combinations of levels. Each combination was assigned to a different staircase.

5.2.3.3 Experiment 1.4

In experiment 1.4, the edge polarity was randomized, rather than being assigned to different staircases (although it was still the case that the test and comparison edges within a trial had the same polarity). The other variables in Table 5.2 were manipulated in the same way within a session, so there were 16 different combinations of levels in a session. Each combination was assigned to a different staircase within a session. In this experiment, only the sessions at a viewing distance of 130 cm were run, giving test edge scales of 12′ and 16′. Each subject performed 4 identical sessions at this viewing distance.

5.2.4 Results

5.2.4.1 Experiment 1.1

The data were analysed as described in section 4.2.3. Before fitting the psychometric functions, the data from different sessions, stimulus polarities and staircase start positions were collapsed together. Each subject's data were analysed twice: in one analysis, the data were separated according to the temporal order of test and comparison edges in a trial; in the other, the data were collapsed across stimulus order before fitting the psychometric functions. The results of both types of analysis are shown in Figure 5.2. The vertical axis of each panel represents the ratio (henceforth termed the σ-ratio) of the comparison and test edge scales when they perceptually matched in blur. When the test and comparison edge contrasts were equal, they could only differ in scale and, as would be expected, they appeared to match in blur when they were physically identical, giving a σ-ratio of 1. As the contrast of the test edge was reduced, the scale of the matching comparison edge decreased, as reported by Georgeson (1994). Assuming that the perceived blur of the comparison edge is an increasing function of scale, these data show that edges look sharper when their contrast is reduced. There was only a small effect of stimulus order, suggesting that afterimages were not affecting the perceived blur too much.

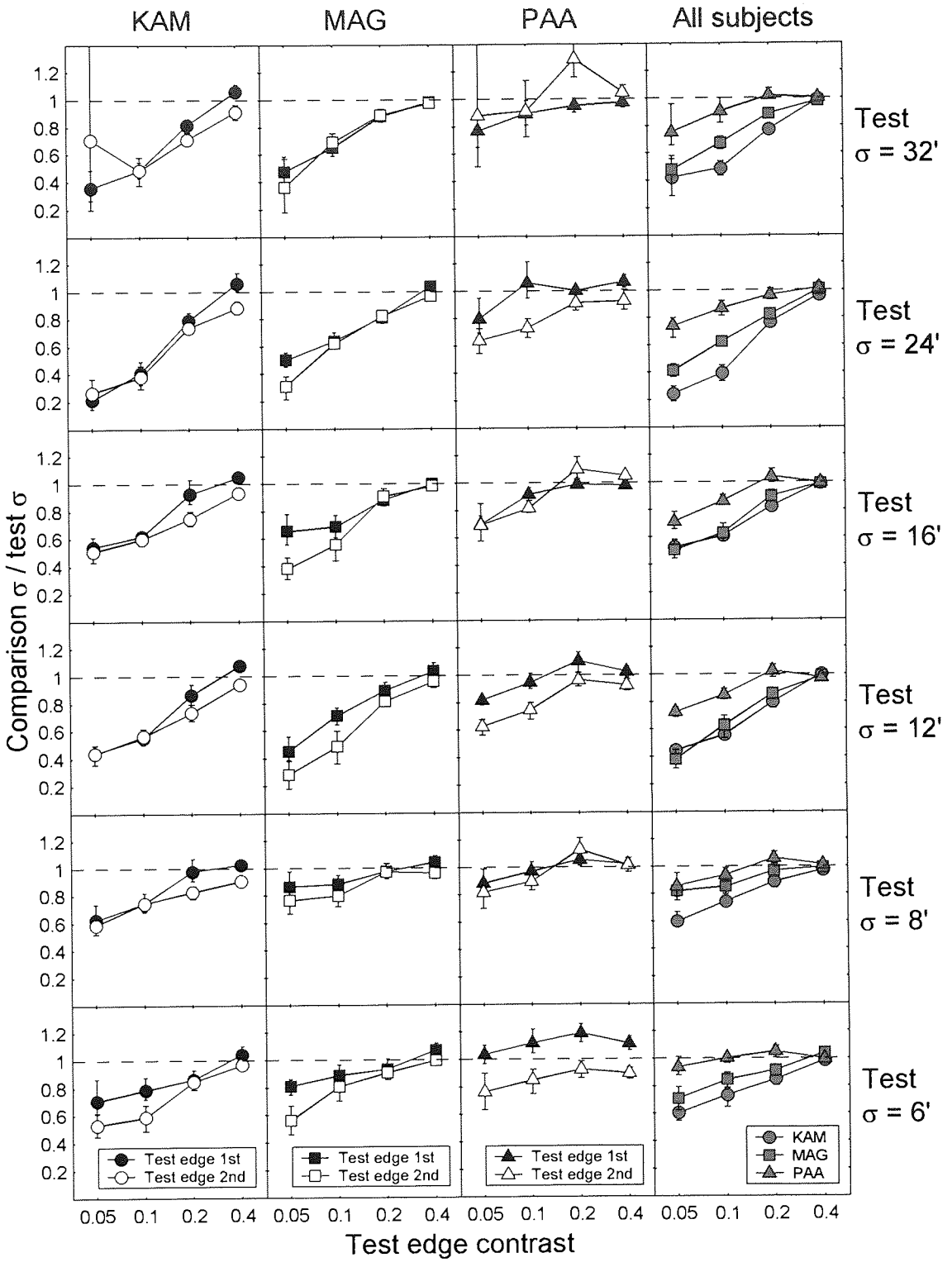


Figure 5.2. Results of experiment 1.1. Each row of panels shows data for a particular test edge scale. Each of the three leftmost columns of panels shows data for one subject; filled symbols show data from trials in which the test edge preceded the comparison edge, and open symbols represent trials with the opposite order. The right hand column shows data for all three subjects, collapsed across stimulus order. Within each panel, the horizontal axis represents the test edge contrast, and the vertical axis represents the ratio of the comparison and test edge scales when they perceptually matched in blur. The dashed line represents the null hypothesis that the edges would match in blur when their scales were equal.

The effect of contrast on σ-ratio was slightly greater for the large-scale test edges (24′ and 32′) than for the small-scale ones (6′ and 8′). The data from the right hand column of Figure 5.2 are re-plotted in Figure 5.3, to show this interaction more clearly. As mentioned earlier, the 6′, 12′ and 24′ test stimuli were physically identical on the screen, as were the 8′. 16′ and 32′ test stimuli: the scale differed only because of the different viewing distances. The finding of an increase in the effect with increasing scale (i.e. decreasing viewing distance) shows that the blur perception process is not completely scale invariant, i.e. the perceived blur does not simply scale up in proportion to the magnification of the image.

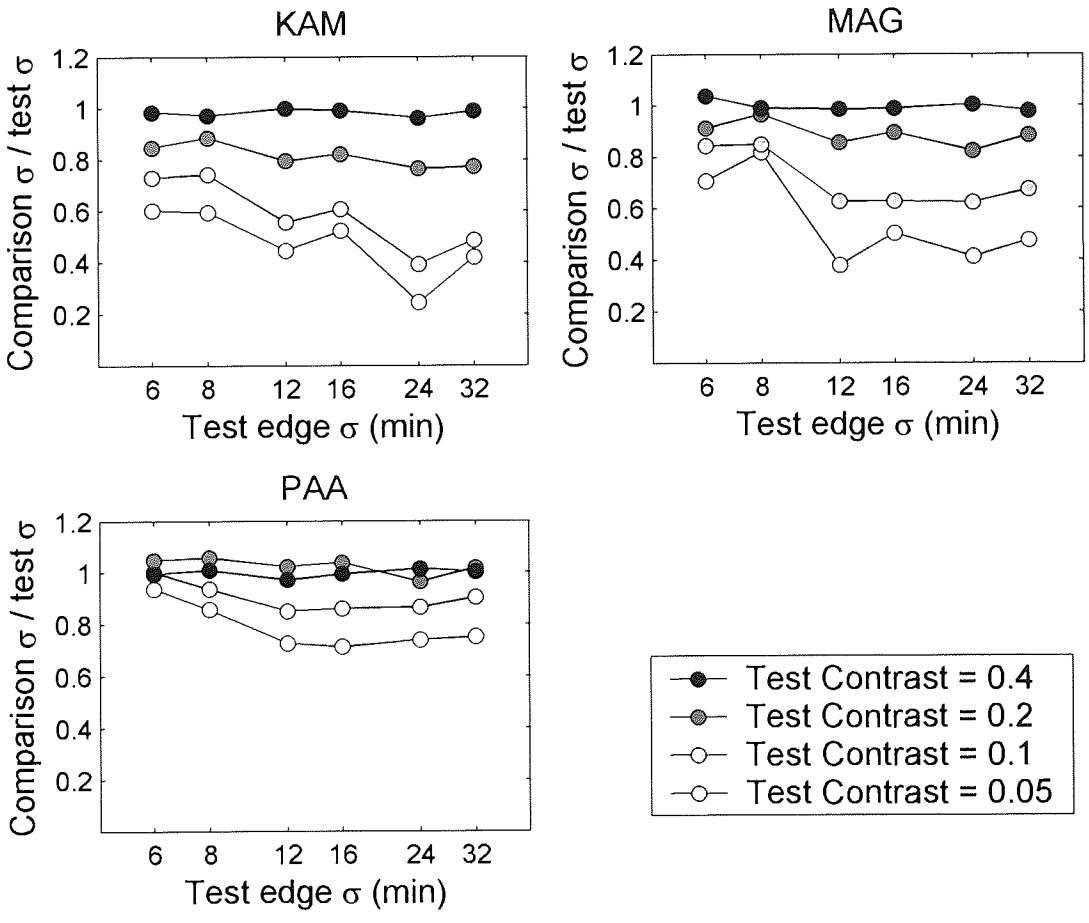


Figure 5.3. Data from the right hand column of Figure 5.2 (i.e. collapsed across stimulus order), re-plotted to show how test edge scale interacts with contrast in determining the σ-ratio.

The test edge scales can be grouped into three pairs: (6′, 8′), (12′, 16′), and (24′, 32′). As shown in Table 5.1, these pairs had spatial window sizes of, respectively, 80′, 160′ and 320′. For each pair, the window diameter was 10 times larger than the scale of the larger-scale edge, and 13⅓ times larger than the scale of the smaller-scale edge. Although there was a general tendency for the effect of contrast on σ-ratio to be larger with larger-scale edges, within each pair, there was a small but very reliable tendency for the effect to be larger for the smaller-scale edge: for KAM and MAG, at each contrast level (except 0.4, where the σ-ratio is expected to be 1), the smaller-scale edge of each pair had a lower σ-ratio than the larger-scale edge, except KAM's (6′, 8′) pair at a contrast of 0.05. PAA also showed the same tendency, with 6 of the 9 pairs at contrasts 0.2, 0.1, and 0.05 showing a lower σ-ratio for the smaller-scale edge.

5.2.4.2 Experiment 1.2

Experiment 1.2 was identical in design to experiment 1.1, except for the noise surround. Figure 5.4 compares the results with and without noise. In general, the noise made very little difference to the results. When the test edge preceded the comparison edge, there was very close agreement between the two noise conditions for all test edge scales and contrasts. There was a substantial effect of noise only with the largest-scale test edge at the lowest contrast, when the test edge came second. The data point corresponding to this combination of conditions with the noise surround deviates strongly from the pattern observed in all the other data, in that the σ-ratio is substantially greater than 1. This particular combination of conditions (highest blur, lowest contrast) gives rise to the lowest test edge gradient, which may have made the test edge maximally difficult to see. In addition, the comparison edge would have given rise to a negative afterimage, which, if it had persisted long enough, would have reduced the effective test edge gradient still further. It may have been that, on some trials, the edge was not detected, and the apparently blank stimulus was mistaken for a very blurred edge, while, on other trials, the edge was detected as normal. This would have caused the threshold and slope of the psychometric function to increase, giving rise to a high σ-ratio and an enlarged confidence interval, as seen here.

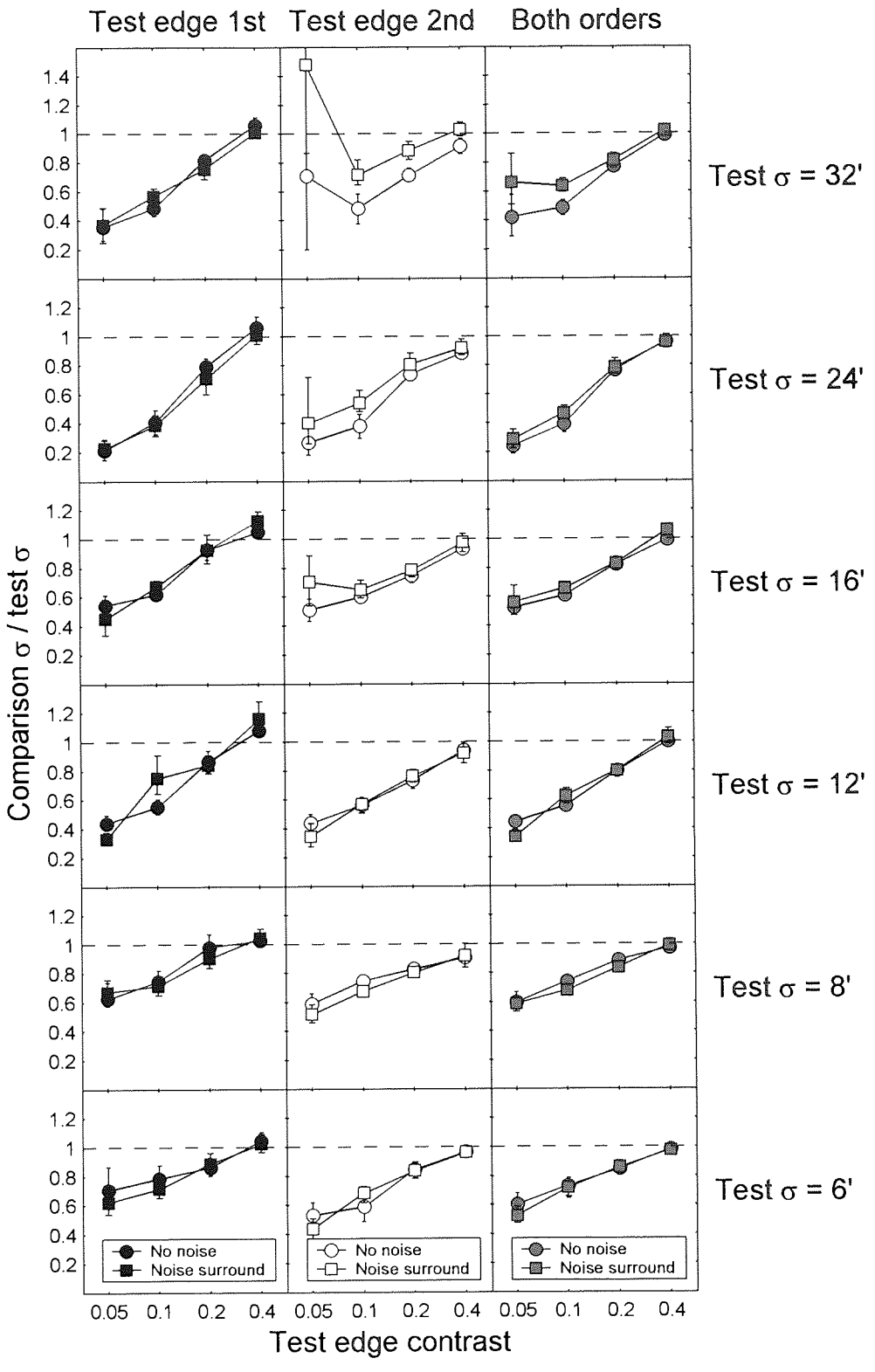


Figure 5.4. Comparison between KAM's results from experiments 1.1 and 1.2. Circles represent data from experiment 1.1, with no noise; squares represent data from experiment 1.2, in which stimuli were surrounded by a border of binary noise. The symbol colours correspond to the colours used in Figure 5.2.

5.2.4.3 Experiment 1.3

The results of experiment 1.3 are shown in Figure 5.5. The 0.5 sec conditions in this experiment were replications of the 0.05 contrast condition in experiment 1.2 (data shown in Figure 5.4). The data from this replication are very similar except that, in the current experiment, the condition in which a 32 min test edge came after the comparison edge did not give rise to the wayward data point seen in Figure 5.4, strengthening the argument (given in Section 5.2.4.2) that this was an aberration, caused by the subject's failure to detect the stimulus on some trials.

As Figure 5.5 shows, there is no clear effect of ISI duration up to 2 sec. If afterimages were affecting perception substantially, one would expect there to be a systematic trend in σ-ratio with ISI duration, as the afterimage would have faded more with a longer ISI. There is also no convincing effect of stimulus order. As argued in Section 5.2.1, the test stimuli in this experiment all had the lowest contrast used in experiments 1.1 and 1.2, and should give rise to the largest discrepancy between stimulus orders due to afterimages. This experiment therefore confirms that afterimages were probably not having a large effect on the results of the main experiment.

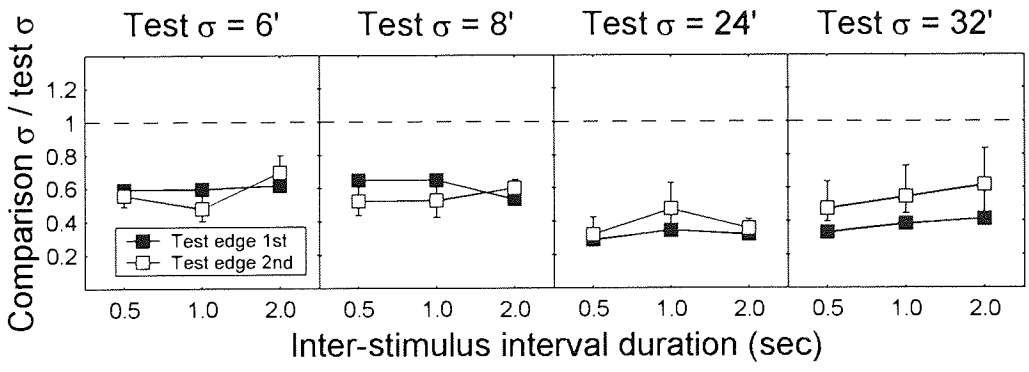


Figure 5.5. Results of 1.3. Subject was KAM. Each panel shows data for a different test edge scale. Filled symbols show data from trials in which the test edge preceded the comparison edge, and open symbols represent trials with the opposite order. Test edge contrast was 0.05 in all conditions.

5.2.4.4 Experiment 1.4

Comparison of the results of experiments 1.1 and 1.4 allows an examination of the effect of spatial window size. Ideally, the ISI should be the same in both experiments but, given that the ISI duration apparently has no effect between 0.5 and 2 seconds, this should not be a problem. Figure 5.6 shows that the effect of contrast on σ-ratio is very similar in the two experiments. Figure 5.7 re-plots the data to show the effect of contrast on σ-ratio interacts with test edge scale. As with experiment 1.1 (Figure 5.3), the effect was slightly stronger for the 12′ edge than the 16′ edge.

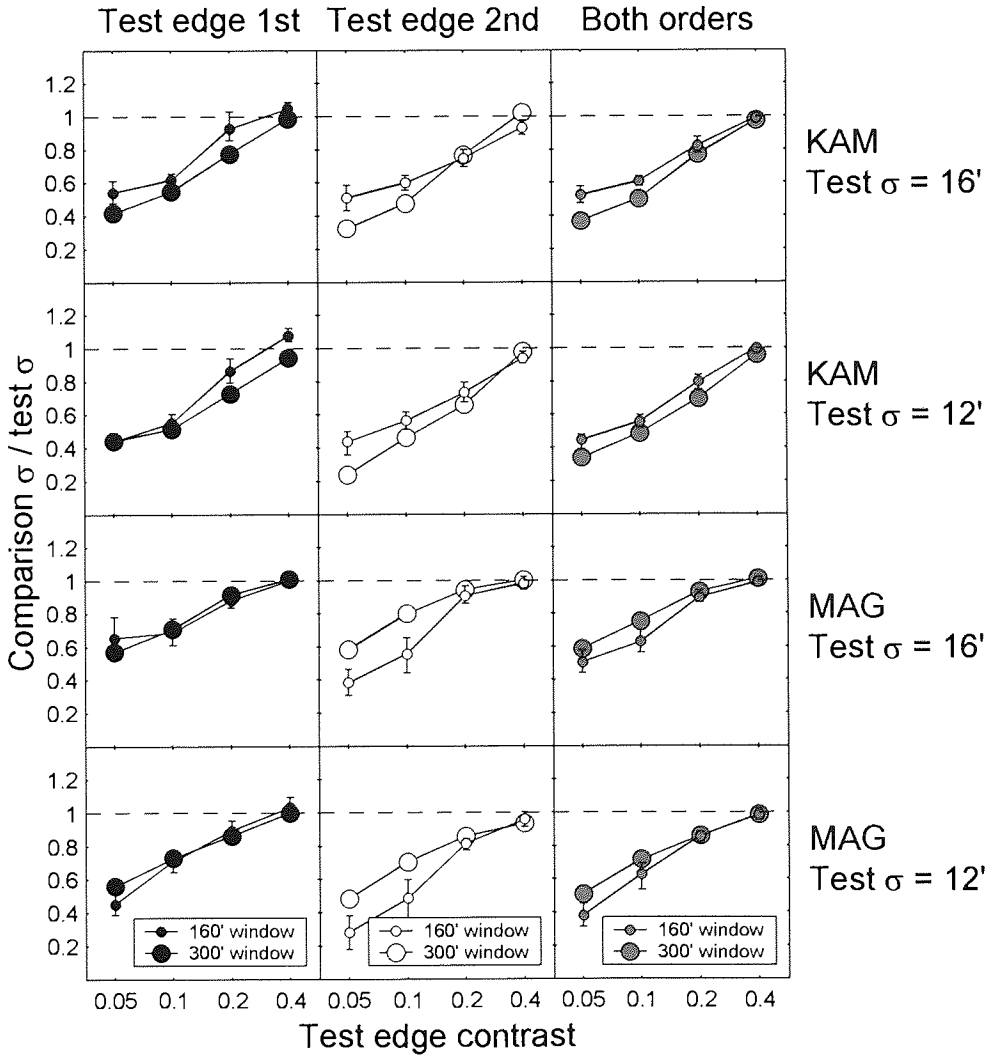


Figure 5.6. Comparison of results of experiment 1.4 with the corresponding conditions of experiment 1.1. Large symbols represent the data from experiment 1.4, which used a 300′ window diameter, and small symbols represent data from experiment 1.1, which used a 160′ window diameter with the test edge scales used in experiment 1.4. The symbol colours correspond to the colours used in Figure 5.2 and Figure 5.4. The top two rows of panels show data from subject KAM; the bottom two rows show data from MAG.

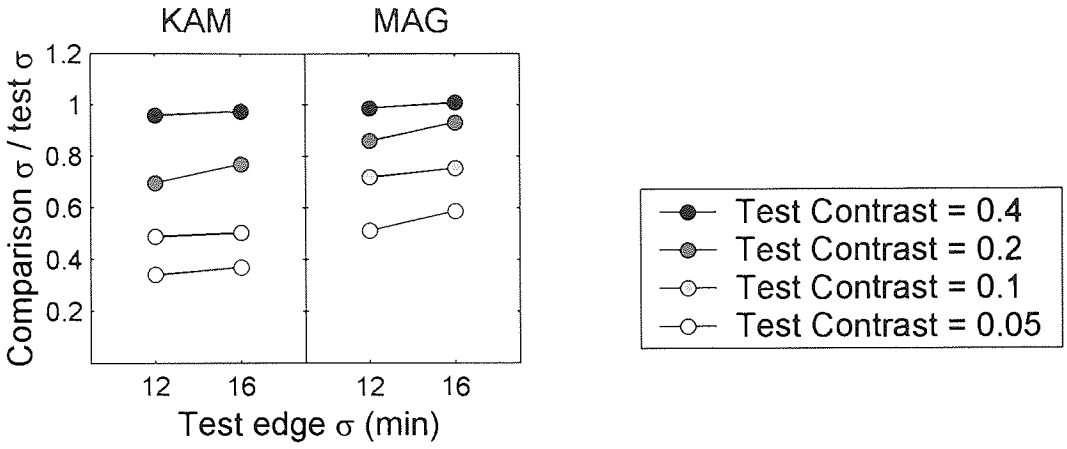


Figure 5.7. Data from the right hand column of Figure 5.6 (i.e. collapsed across stimulus order), re-plotted to show how test edge scale interacts with contrast in determining the σ-ratio.

5.2.5 Discussion of experiment 1

The main result of experiment 1 is that reducing the contrast of an edge makes it look sharper. With subjects KAM and MAG, as the test edge contrast was reduced from 0.4 to 0.05, the scale of the blur-matched comparison edge usually fell by around 40-60%. PAA showed a qualitatively similar effect, but with a smaller magnitude. It may have been that PAA's visual system is genuinely different from that of KAM or MAG. Alternatively, it may be that PAA was reluctant to judge the low-contrast edges as sharp, because they were quite difficult to see and often appeared "grainy". On some trials, the edges looked "patchy", so that the edge was not seen along its entire length. The written instructions given to the subjects stressed that graininess is not the same as blur, and asked subjects to judge the blur of the region that they could see, if they could only see part of the edge. But despite these instructions, it may have been difficult to distinguish between an edge that is blurred and one that is difficult to see.

The results of experiments 1.1 and 1.2 suggest that the effect of contrast on the σ-ratio increases slightly with increasing test edge scale, across the range of test edge scales (6′ to 32′). However, this effect was complicated by the effect of the spatial window size. The test edge scales can be grouped into three pairs: (6′, 8′), (12′, 16′), and (24′, 32′). Within

each pair, the window diameter was the same, so comparisons within each pair allow the effect of test edge scale to be examined while keeping the spatial window diameter constant. These comparisons showed that the effect of contrast on σ-ratio actually *decreased* slightly with increasing test edge scale. It might be argued that the effect generally decreases with increasing test edge scale, but increases with increasing spatial window diameter, and it was the increase in spatial window diameter that caused the larger effect with larger-scale edges. However, experiment 1.4 rules out this hypothesis. In this experiment, the window diameter for the 12′ and 16′ edges was increased to 300′, which was very close to the 320′ window used with the 24′ and 32′ edges in experiments 1.1 and 1.2, yet there was no detectable increase in effect of contrast on σ-ratio: in fact, MAG showed a very slightly smaller effect with the larger window.

At this point, it is instructive to examine the data reported by Georgeson (1994) because, in his experiment, the circular spatial window was kept constant for a large range of test edge scales (the window had a flat circular region with diameter 200′, and the radial profile of the edge of the window was a raised sine with half-period 50′, giving an outer diameter of 300′). Georgeson presented the geometric mean of the blur matching data from two subjects, MAG and TCAF. The mean data, as well as the data from the individual subjects, are re-plotted in Figure 5.8, using the σ-ratio form used with the present experiments. For both subjects, the effect of contrast on perceived blur increased with increasing test edge scale up to 5.65′. Further increases in test edge scale produced no further change in the effect size for MAG; for TCAF, the effect size became smaller.

It is not clear why TCAF's data are so different from MAG's, and in particular why the σ-ratio goes substantially higher than 1 with the 22.6′ test edge. It might be argued that the edges with the lowest contrasts and highest scales would be difficult to see, but TCAF's most anomalous data points occurred at contrasts of 0.08 and 0.16, which should be easily visible.

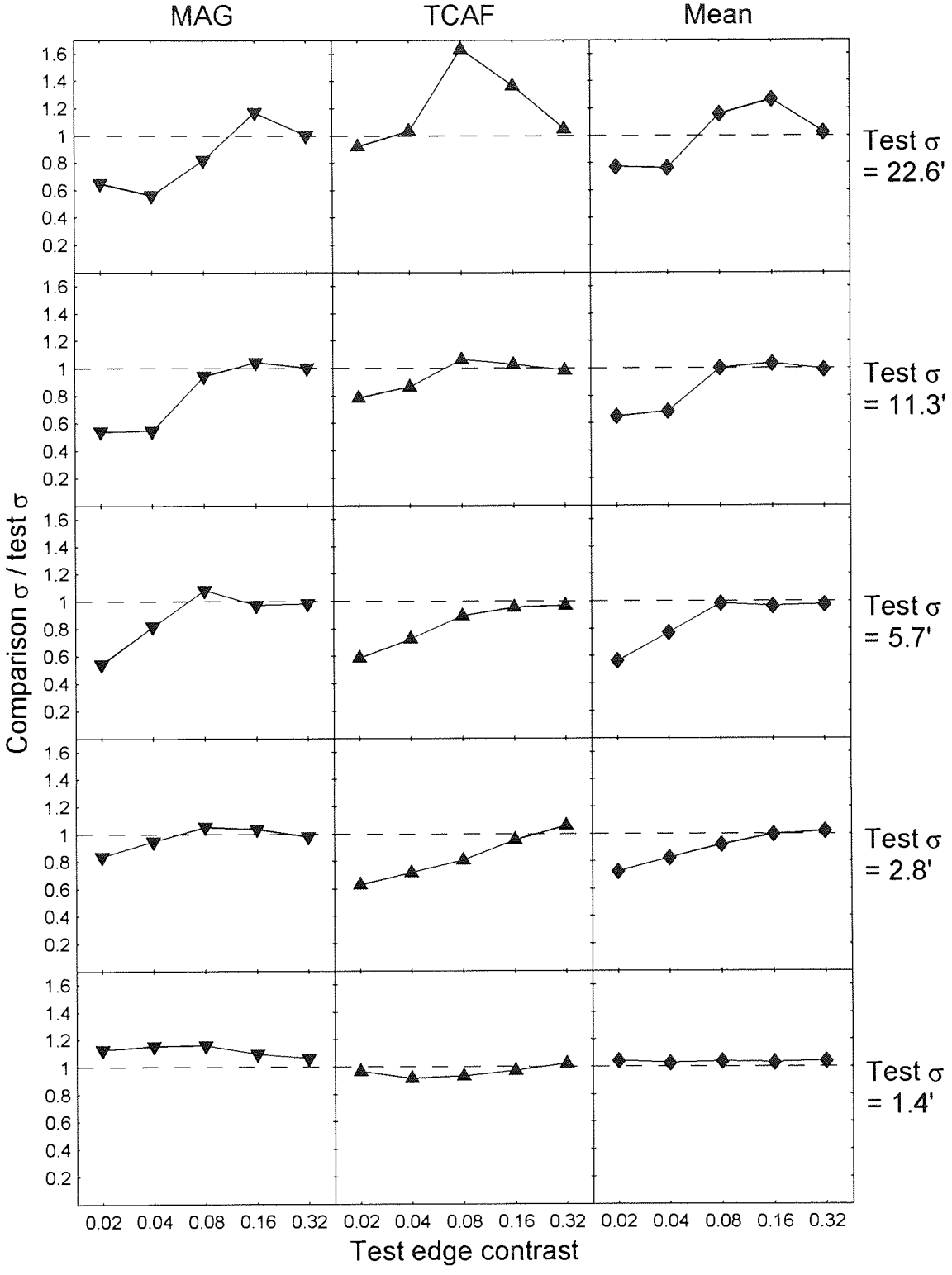


Figure 5.8. Georgeson's (1994) data, re-plotted in the form used in this thesis. Each row of panels shows data for a different test edge scale, and the columns show data from subjects MAG, TCAF, and the geometric mean across the two subjects.

MAG's data show a similar pattern to the results of experiments 1.1 and 1.2, suggesting that the effect of contrast on σ-ratio really does increase with increasing test edge scale, and that the increase in the effect found in experiments 1.1 and 1.2 is not due entirely to the increasing window diameter. Nevertheless, TCAF's data are far more equivocal. It was important to establish how the effect of contrast on σ-ratio varies with scale, as this might have implications for any candidate model of the effect, so a new experiment was run to confirm that the effect of contrast on σ-ratio increases with increasing test edge scale when the spatial window size is held constant.

5.3 Experiment 2: effect of contrast on perceived blur with fixed window size

5.3.1 Introduction

The aim of experiment 2 was to find how the effect of contrast on perceived blur changed across edge scale, using a fixed window size. To give a more complete view of the change in the effect size across scale, experiment 2 used a wider range of edge scales than experiment 1. Experiment 2 again used the blur-matching procedure described in section 4.2. The main experiment was performed by subjects KAM, MAG and PAA. KAM also performed two slight variations on the main experiment, described in the next section.

5.3.2 Stimuli

The stimuli in the main experiment were all vertical Gaussian edges, generated according to equation (5.1), using a different window function, w, which was a sharp-edged square with width and height 320′. Stimulus matrices were arrays of 320×320 pixels, which measured 14.2×14.2 cm on the screen. The viewing distance was set to 152.3 cm, which made each pixel 1′ wide. The ISI duration between the test and comparison edges

was 500 ms, as with experiments 1.1 and 1.2. After collecting the data, a strong order effect was found with the sharper edges. This might have been caused by a reduction of effective test edge contrast due to an afterimage from the comparison edge, when the comparison edge preceded the test edge, so KAM repeated the experiment with an ISI of 1200 ms. This manipulation had very little effect, so KAM repeated the experiment again, this time with an ISI of 500 ms, but with a horizontal comparison edge. The idea behind this manipulation was that the addition of an afterimage of an orthogonal edge would not change the effective contrast of the stimulus. In the versions with only vertical stimuli, the polarity of the test and comparison edges was always the same within a trial, and was chosen randomly for each trial. In the version with orthogonal stimuli, the polarity of each stimulus was chosen randomly.

5.3.3 Methods

The design of experiment 2 was very similar to that of experiments 1.1 and 1.2. As before, there were two independent variables, test edge contrast and scale. Contrast had 4 levels: 0.05, 0.1, 0.2, and 0.4; scale had 5 levels: 2, 4, 8, 16, 32 min of visual angle. The experiment thus had a 4×5 factorial design. The dependent variable was the scale of a 0.4 contrast comparison edge that matched the test edge in perceived blur. Each of the 20 combinations of levels of the independent variables was assigned to a different staircase, which controlled the comparison edge scale, and all 20 staircases were tested within one session. The subject completed 8 similar sessions. For each odd-numbered session, half of the staircases were randomly chosen to start above the test edge scale, and the other half started below the test edge scale; the magnitude of the difference between the scales of the test edge and initial comparison edge was randomly chosen to be an integer between 10 and 14 dB. For the succeeding even-numbered session, the starting point of each staircase was chosen to be above the test edge scale if the corresponding staircase in the previous session had been below the test edge scale, otherwise the starting point was below the test edge scale. Again, the magnitude of the difference was randomly chosen to lie between 10 and 14 dB.

5.3.4 Results and discussion

The data were analysed as described in section 4.2.3. The results of the main experiment are shown in Figure 5.9. As in experiment 1, the results for KAM and MAG showed a reduction in σ-ratio with decreasing contrast. For the intermediate scales, 4′ to 16′, the overall size of the effect was constant across scale; for the largest (32′) and smallest (2′) scales, the effect was smaller. These results are therefore highly consistent with those of Georgeson (1994). PAA showed a complete reversal of the effect that he had shown in experiment 1.1: the only substantial effect of contrast on blur was an *increase* in σ-ratio with decreasing contrast for the larger-scale test edges, when the test edge came 2nd. A similar effect had occurred in experiment 1.2 with subject KAM and, as argued in section 5.2.4.2, this probably occurred because the low-contrast, large-scale stimuli were sometimes invisible, and the apparently blank stimuli may have been mistaken for very blurred edges. It is not clear why PAA showed this reversal in experiment 2 and not experiment 1.1: the stimuli in the 32′ conditions for both experiments had the same window width, the only difference being that the window was circular in experiment 1.1 and square in experiment 2.

Although KAM showed effects in the same direction as those shown in experiment 1, he showed a very large order effect, particularly with the smallest-scale edges. Figure 5.10 and Figure 5.11 show that this order effect was almost as strong with a longer ISI duration and with orthogonal stimuli, ruling out the hypothesis that the order effect was caused by a change in the effective stimulus contrast, caused by afterimages. In general, Figure 5.11 shows a striking agreement among the three versions of experiment 2. Figure 5.11 also shows data from any conditions of experiment 1.1 which had the same edge scale and contrast as a condition of experiment 2. Again, the data for corresponding conditions in the different experiments are very similar.

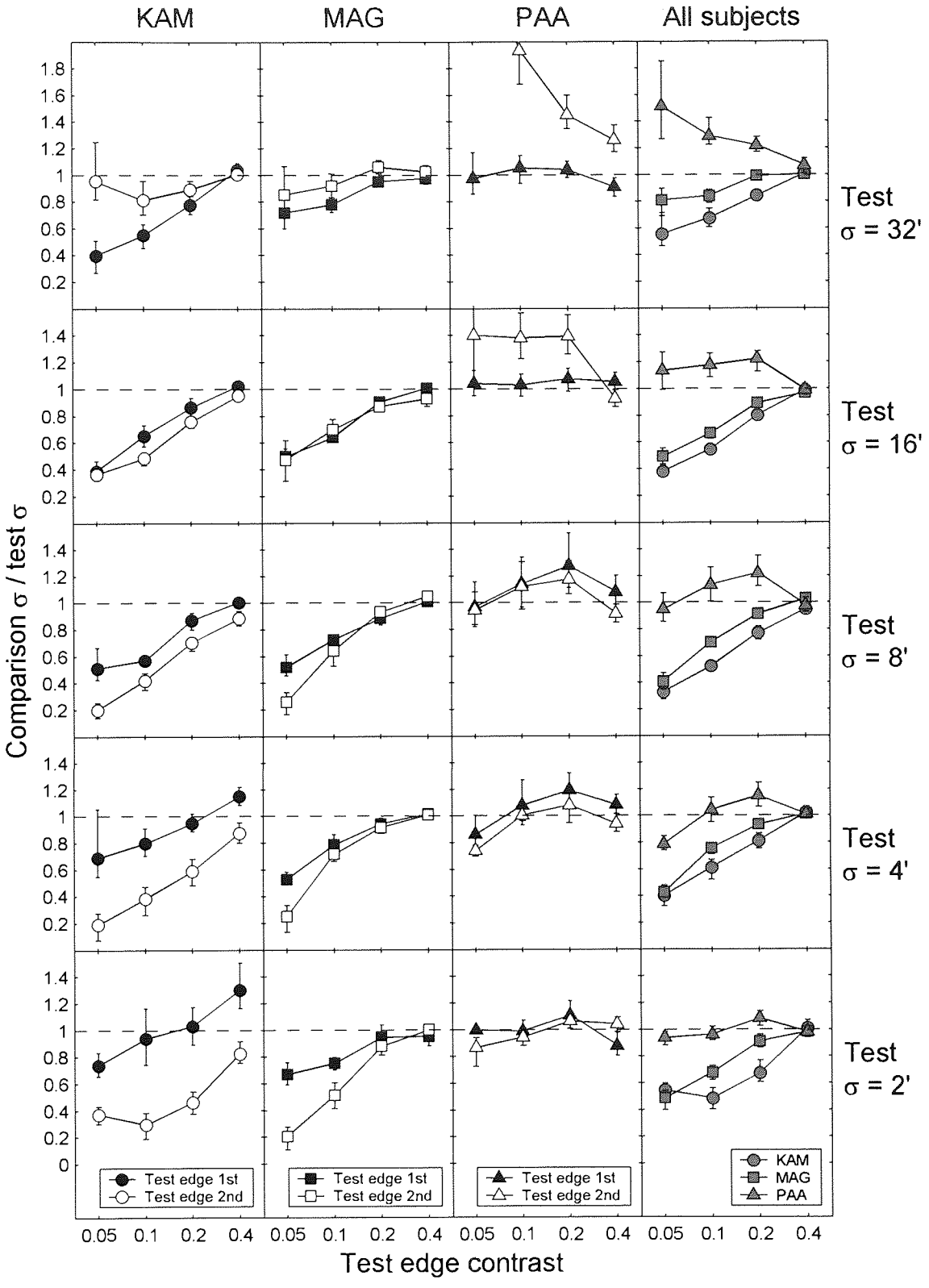


Figure 5.9. Results of the main part of experiment 2. One data point could not be plotted on these axes, because it was too wayward. This was PAA's data for the test edge with scale 32′ and contrast 0.05, when the test edge came 2nd. The σ-ratio for this point was 2.96.

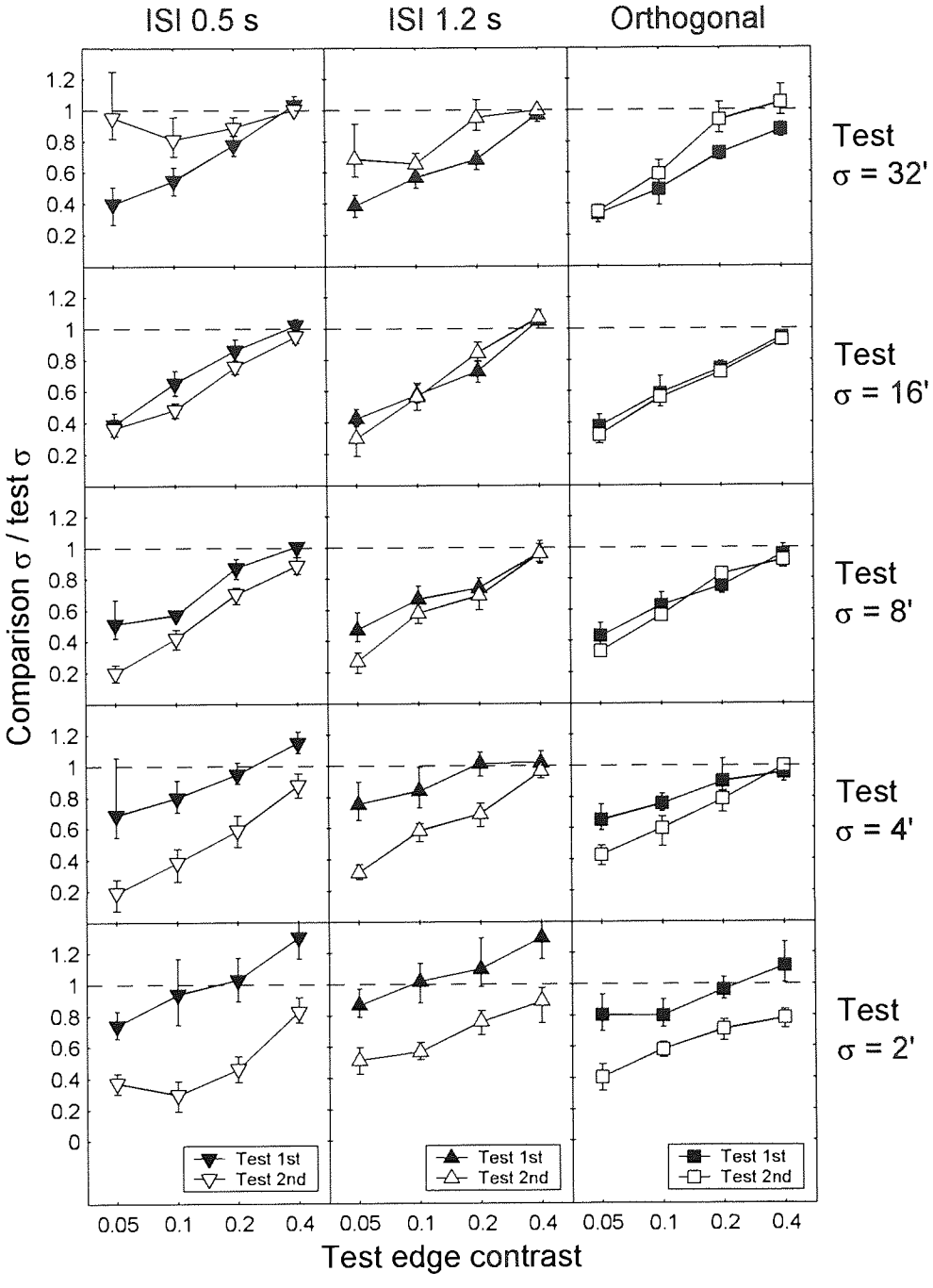


Figure 5.10. Results of the three versions of experiment 2 that KAM performed. The left column shows the results of the main experiment (taken from the left column of Figure 5.9). The middle column shows the results with a longer ISI duration (1200ms) and the right column shows the results with orthogonal test and comparison edges (ISI duration 500ms, as in the main experiment).

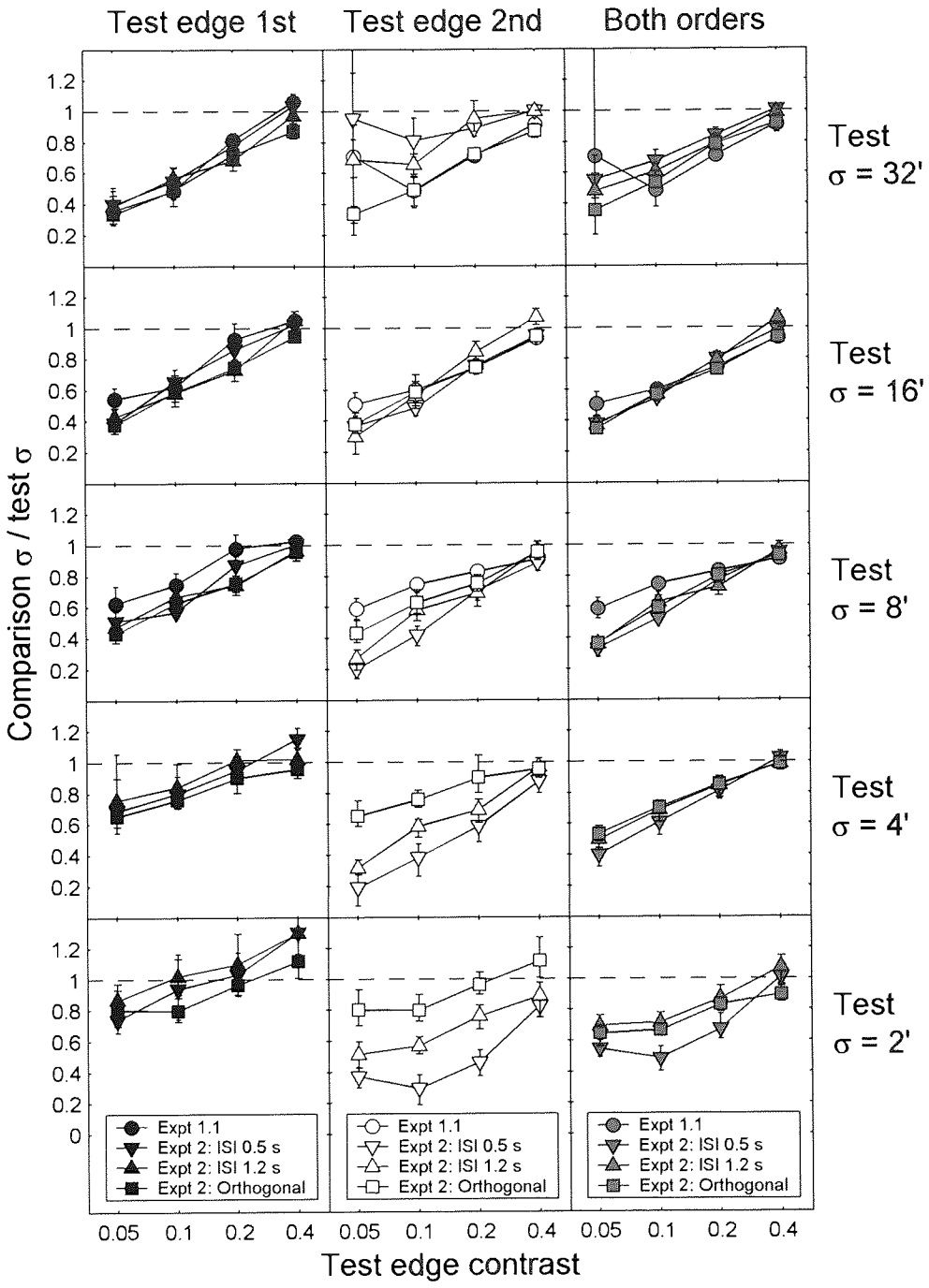


Figure 5.11. Results KAM's 3 versions of experiment 2, plotted so that the corresponding conditions in the three versions can be compared directly. All the data from Figure 5.10 appear in the two leftmost columns of this figure, represented by the same symbols. The right column shows the results of fitting the psychometric functions after collapsing across stimulus order. Also shown are KAM's data from the 8', 16' and 32' conditions of experiment 1.1, which were identical to the corresponding conditions of experiment 2, apart from the window shape and/or width. The data from experiment 1.1 are represented by circles, which are the same symbols used to represent these data in Figure 5.2.

5.4 General discussion: explaining the effect of contrast of perceived blur

The key result from this chapter is that reducing the contrast of a blurred edge makes it look sharper. The size of this effect is fairly constant across edge scale, but is slightly larger with edges of intermediate scale (4′ to 24′). There are indications that the slight reduction in the effect size for 32′ edges may be due to occasional invisibility of the central edge in the stimulus. Subject KAM, who performed a number of variations of the basic experiments, showed considerable consistency across changes of ISI duration (experiment 1.3, Figure 5.5; experiment 1.1 vs. 1.4, Figure 5.6; experiment 2, Figure 5.11), window size (experiment 1.1 vs. 1.4, Figure 5.6; experiment 1.1 vs. 2, Figure 5.11), surrounding noise (experiment 1.1 vs. 1.2, Figure 5.4), and relative orientation of test and comparison stimuli (experiment 2, Figure 5.11).

The model described in Chapter 2 predicts that blur perception will be contrast-invariant, so it requires modification in order to account for the results. Since this model had been very successful at predicting the results of many previous blur-matching experiments performed at moderate contrasts, it was important to find a modification that would not drastically alter its performance with moderate-contrast stimuli. One possible solution can be found by returning to Kovasznay & Joseph's (1955) algorithm. As previously noted, the first three stages of Kovasznay & Joseph's algorithm are essentially the same as the template model (except that their algorithm did not have blur channels). One slight difference is that, in their model, the half-wave rectifier has a non-zero threshold. Thus, the rectifier has the form: $H = \lfloor I_x - T \rfloor$, where T is the threshold. This process is referred to as *rectification with bias* (Kovasznay & Joseph, 1955) or *over-rectification* (Heeger, 1992a, 1992b). Figure 5.12 explains how over-rectification could make low-contrast edges appear sharper. The threshold truncates the 1st-derivative signal, so that only the central (above-threshold) region survives. This causes the 2nd derivative signature to be truncated, too. With low-contrast edges, more of the 1st derivative falls below threshold, so the truncation is more severe, and the 2nd derivative fits a narrower

template than the 2nd derivative from a higher contrast edge. Thus, the low-contrast edge is perceived to be sharper than the high-contrast edge.

Threshold models have a long history in machine vision, physiology and psychophysics. In machine vision, thresholds are used to remove noise. Indeed, that was the motivation for the threshold in Kovasznay & Joseph's model, as well as the enhancement/ thresholding edge detectors described in section 1.2. In physiology, neurons have been described using threshold models, because each neuron has a threshold contrast level below which it does not fire (Ikeda & Wright, 1974; Dean, 1981; Tolhurst *et al.*, 1981; Albrecht & Hamilton, 1982). In psychophysics, studies of contrast perception have shown that the visual system behaves as if it has access to the physical stimulus contrast minus the detection threshold (Kulikowski, 1976; Cannon, 1979; Georgeson, 1991b). These issues are examined in detail in Chapter 7.

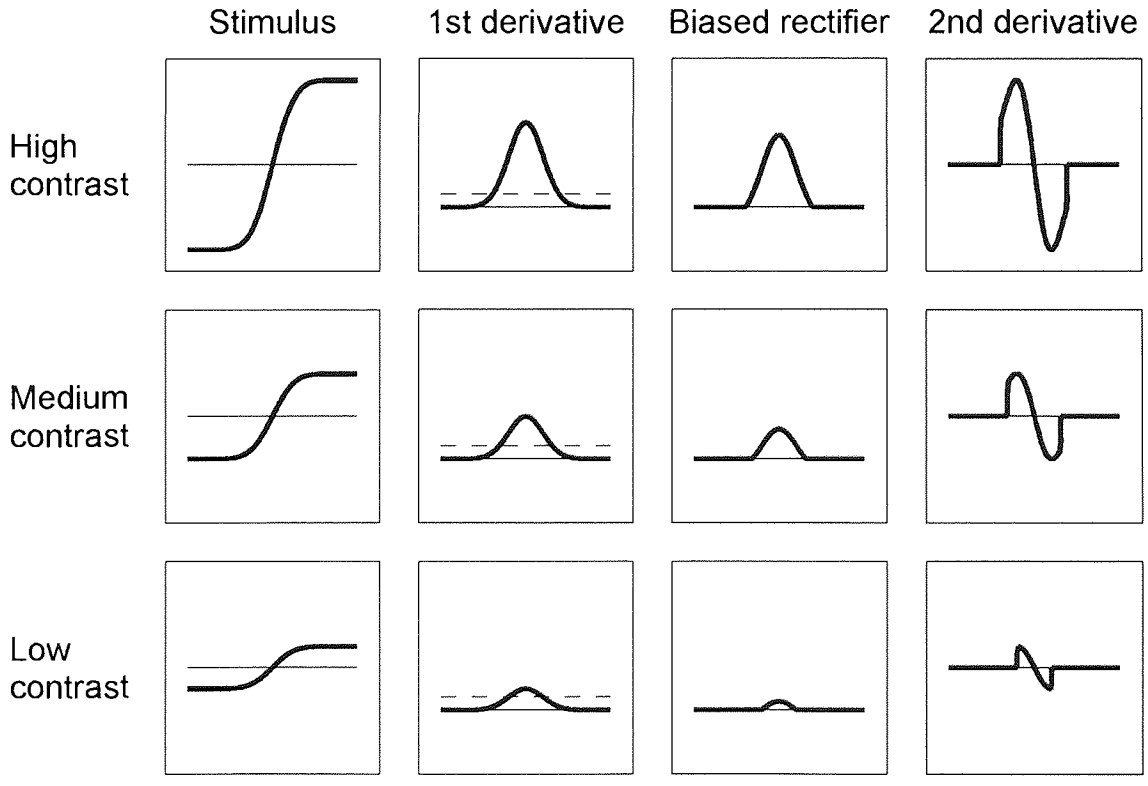


Figure 5.12. An illustration of why half-wave rectification with a threshold makes low-contrast edges appear sharper. Each row shows the processing of an edge of different contrast. Each column shows a different processing stage. The threshold level is indicated by the dashed line in the 1st derivative stage. The threshold causes the 2nd derivative to be truncated. The truncation is more severe with the low-contrast edge, so the 2nd derivative is narrower, and fits a narrower template, and therefore looks sharper than the high-contrast edge.

Chapter 6
Contrast threshold for edge detection

6.1 Introduction

In the previous chapter, it was hypothesised that the edge detection mechanism applies a threshold to the 1st-derivative signal. If this were the case, the system would be unable to detect any edge with a 1st-derivative signal falling below the gradient threshold in all channels. In the absence of any other threshold, and with a noise level substantially below the gradient threshold, the contrast threshold for edge detection would be determined by the gradient threshold. The highest gradient response is always in the smallest channels, since they introduce the least blurring, and hence reduce the edge gradient by the smallest amount. Thus, if the threshold were constant across all channels, then, as the contrast was reduced, the smallest channel would be the last to fall below threshold, so the contrast threshold would be determined by the smallest channel: an edge would be at threshold when its gradient (after blurring by the smallest channel) was equal to the gradient threshold. If we assume that the blurring introduced by the smallest channel is negligible, then the contrast threshold should be proportional to edge scale, since edge contrast is proportional to edge scale when the peak gradient is held constant.

Of course, this is a very naive model: the idea that there is a fixed sensory threshold that separates perceived things from unperceived things has been largely discredited (e.g., Swets, 1961). Even so, the existence of an explicit threshold in the system could possibly have an effect on the detection threshold, and contrast thresholds for edge detection might indicate an upper bound on the range of plausible values for the gradient threshold.

6.2 Experiment 3: contrast threshold for edge detection

6.2.1 Introduction

One aim of this experiment was to find the contrast of just-detectable edges with different scales. The other aim was to see if the contrast threshold for judging the polarity of the edges matched that for the detection, i.e. to see if the polarity of a stimulus was correctly identified whenever it was detected. The experiment used the 2×2-alternative forced choice procedure used previously by several authors (e.g., Nachmias & Weber, 1975; Tolhurst & Dealy, 1975; Watson & Robson, 1981), in which, on each trial, the subject is asked to indicate, not just which interval contained the target, but also which, of two alternatives, the target was. In Watson & Robson's (1981) terminology, edges of opposite polarity that can be discriminated at detection threshold are *perfectly discriminated*, and must be detected by different *labelled detectors* whose responses can be distinguished from each other. In some edge detection models, particularly the energy models, edges are detected by mechanisms that are insensitive to the polarity of the edge, so edges of opposite polarity would be detected by the same mechanism, and might not be discriminable at the detection threshold. Identification of polarity is performed by a later mechanism, which might have a higher contrast threshold[16]. In contrast, in the GDT model, the same peak of activity serves both to detect the presence of an edge, and to identify its parameters, so this model predicts that, if an edge can be detected, it can be identified. A more rigorous proof of this prediction follows.

Assume that, in a detection task, the chosen interval is that which gave rise to the highest peak in the edge strength representation. If the highest peak occurred in a channel with orientation between 0° and 180° anticlockwise from vertical, the peak is consistent with the target being dark on the left; if the highest peak occurs in a channel with orientation between 0° and −180°, the peak is consistent with the target being dark on the right. It is

---

[16] Of course, energy models would also be consistent with the finding that edges of opposite polarity are perfectly discriminated, because the later polarity identification stage might not have a higher threshold than the detection stage.

reasonable to assume that the polarity judgement is determined by which of these two sets of orientation channels contains the highest peak. We can define the propositional variables[17] D and I as follows:

D: the highest peak in the edge strength representation occurred in the target interval (so the subject made a correct Detection judgement)

I: the highest peak in the edge strength representation occurred in the set of orientation channels consistent with the target polarity (so the subject made a correct Identification judgement).

These propositional variables can be combined to form the four compound propositions listed in Table 6.1, which correspond to the four possible (mutually exclusive) response outcomes.

Table 6.1. Compound propositions corresponding to the four possible response outcomes in experiment 3.

Compound proposition	Interpretation
$(D \wedge I)$	Correct detection and correct identification
$(D \wedge \neg I)$	Correct detection and incorrect identification
$(\neg D \wedge I)$	Incorrect detection and correct identification
$(\neg D \wedge \neg I)$	Incorrect detection and incorrect identification

If $P(X)$ is the probability of outcome X, then:

$$P(D) = P(D \wedge I) + P(D \wedge \neg I) \tag{6.1}$$

and

$$P(I) = P(D \wedge I) + P(\neg D \wedge I). \tag{6.2}$$

$(\neg D \wedge I)$ is true when the highest peak occurs in the non-target interval in one of the orientation channels consistent with the target polarity, and $(\neg D \wedge \neg I)$ is true when the highest peak occurs in the non-target interval in one of the orientation channels

[17] Propositional variables are variables that take the truth values TRUE or FALSE.

inconsistent with the target polarity. Assuming no afterimage from the preceding stimulus (reasonable, since the staircased stimuli in this experiment quickly reach a very low contrast), the peak due to a non-target is equally likely to occur in either set of orientation channels, so:

$$P(\neg D \wedge I) = P(\neg D \wedge \neg I). \tag{6.3}$$

$(D \wedge \neg I)$ is true when the highest peak occurs in the target interval in one of the orientation channels inconsistent with the target polarity. In the target interval, the rectifier ensures that there will be no signal from the target edge in these orientation channels. These orientation channels should therefore react in the same way in the target interval as they would in the non-target interval. Thus,

$$P(D \wedge \neg I) = P(\neg D \wedge \neg I). \tag{6.4}$$

So, using (6.1) to (6.4), we have:

$$\begin{aligned} P(I) &= P(D \wedge I) + P(\neg D \wedge I) \\ &= P(D \wedge I) + P(\neg D \wedge \neg I) \\ &= P(D \wedge I) + P(D \wedge \neg I) \\ &= P(D). \end{aligned}$$

This proves that the GDT model predicts that the detection rate will always equal the identification rate, so the contrast thresholds for detection and identification should be the same. Tolhurst & Dealy (1975) found that, for step edges, identification performance was only slightly worse than detection performance at each contrast level. Experiment 3 aimed to see if this finding could be replicated for blurred edges with a range of scales and window sizes, including those used in experiment 1.

6.2.2 Stimuli

One problem with assessing the detection/identification threshold for an edge is that any edge stimulus that is smaller than the entire visual field must contain more than one edge. Take, for example, the stimulus in Figure 5.1(a). As well as the central vertical edge, there must be edges around the window of the stimulus that allow the luminance to return to the mean level. Since any of these edges could contribute to the subject's performance, if a stimulus like that in Figure 5.1(a) were used, it would be difficult to know which part of the stimulus was determining the threshold. The attempted solution was to surround the central stimulus with binary noise, as shown in Figure 5.1(b). In this case, for the surrounding edges to contribute to performance, the visual system would have to detect a difference between the central stimulus luminance and the *mean* luminance of the noisy border, rather than simply detecting a luminance discontinuity, and it was hoped that the former task would require a higher contrast, so that the contrast threshold was determined by the central edge.

To check that the edges were more difficult to detect/identify on the basis of the difference between the central stimulus luminance and the mean noise luminance, a condition was included in which the central edge was occluded by a vertical opaque strip of binary noise, as shown in Figure 6.1. The width of the strip was 0.3 times the diameter of the circular region of the stimulus. The semi-circular regions either side of the strip had uniform luminance: for a Michelson contrast of C, these values would be $L_0(1 \pm C)$, where L_0 is the mean luminance. In this stimulus, there are two kinds of luminance difference that could allow detection of the target. One is the luminance difference between the two semi-circular regions, and the other is the difference between each semi-circular region's luminance and the mean noise luminance. If the contrast threshold for detection/identification of this stimulus were higher than for the standard edge stimuli, this would support the claim that it was the central edge of the latter stimuli that determined the threshold.

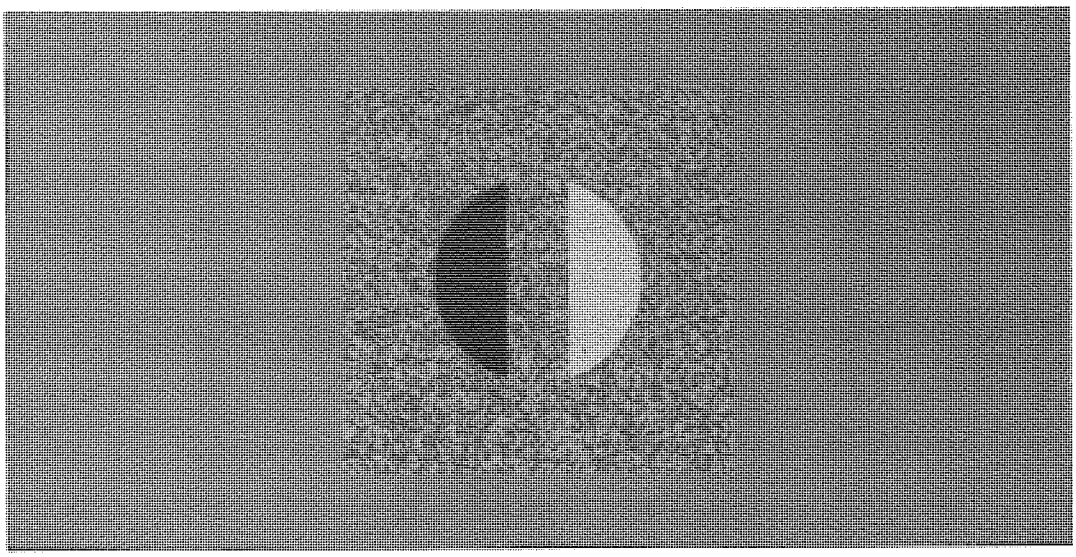


Figure 6.1. An example of the control stimuli used in experiment 3. The central edge is completely occluded by an opaque strip of binary noise. The semicircular regions have uniform luminance $L_0(1 \pm C)$, , where C is the Michelson contrast.

The contrast of the binary noise was always 0.2. The circular perimeter of the stimulus was anti-aliased according to the method used in experiment 1 (described in section 5.2.2.3). Since the width of the noise strip in the control stimulus was not an integer multiple of the pixel width, the edges of this were anti-aliased in a similar way to the circular perimeter, except that the raised cosine edge had a half-period equal to the width of a pixel.

Table 6.2 shows the stimulus parameters for the different conditions. Stimulus matrices were arrays of 256×256 pixels, which measured 11.4×11.4 cm on the screen. At each viewing distance, the two most blurred target edges had the same scale and window size as the test stimuli at that viewing distance in experiments 1.1 and 1.2.

Table 6.2. Stimulus parameters for all conditions in experiment 3.

Viewing distance (cm)	Target stimulus					Circular window diameter (arc min)	Noise rectangle width (deg)	Noise pixel width (screen pixels)
	Edge stimulus with given sigma (min)				Control stimulus			
65	8	16	24	32	as described	320	10	2
130	6	8	12	16	as described	160	5	2
260	2	4	6	8	as described	80	2.5	2

6.2.3 Method

Each trial consisted of a vertical target stimulus and a non-target stimulus, one following the other in a random order. The target was either an edge stimulus, like that in Figure 5.1(b), or a control stimulus, like that in Figure 6.1. The non-target had the same shape of noise surround as the target in that trial (i.e. edge or control stimulus), but had an edge contrast of zero. A different random noise pattern was generated for each stimulus on every trial. The ISI between the stimuli within a trial was a blank field (at mean luminance) that lasted for 500 ms. The polarity of the target on each trial was randomized.

On each trial, the subjects had to make two judgements. For the *Detection* judgment, they had to judge which interval contained the target; for the *Identification* judgment, they had to judge the polarity of the edge. The two possible responses for each judgment combined to give four possible response options, and the subjects indicated the combination of judgments by pressing one of four keys. The response keys were arranged in a 2×2 square in which the row indicated the chosen target interval (upper row = 1st interval; lower row = 2nd interval) and the column indicated the polarity (left column = target dark on left; right column = target dark on right). This arrangement of keys allowed each judgment to be associated with one of the two spatial dimensions of the keypad, and none of the subjects reported any difficultly with the task. With a different key for each combination of responses, both judgments could be indicated simultaneously, and this had two advantages over the alternative procedure of indicating each judgment with a separate response. Firstly, requiring subjects to make separate responses might have encouraged them to make the judgments in a particular order, which might not have corresponded to their natural behaviour; secondly, if two separate key presses had been made on each trial, the performance on the second response could have been impaired by the act of making the first response and by the longer duration between stimulus and response. After each response, a sound (fanfare or buzz) was played to indicate the correctness of the *Detection* judgement, followed by another sound (same options as before) to indicate the correctness of the *Identification* judgment.

The contrast of the stimuli was controlled by staircases. Each condition within a session had two staircases – one that converged to a performance level of 71% correct, and one

that converged to a performance level of 79% correct (Wetherill & Levitt, 1965; Levitt, 1971). The responses for the *Detection* judgment were used to drive the staircases. For the staircase that converged to 71% correct, the contrast was reduced after two consecutive correct responses and increased after one incorrect response; for the staircase that converged to 79% correct, the contrast was reduced after three consecutive correct responses and increased after one incorrect response. Each staircase had a starting point that was chosen randomly, with a uniform probability density between 10 dB and 20 dB, and rounded to the nearest integer value in dB. The contrast in dB, C_{dB}, is defined in terms of C_M, the Michelson contrast as follows:

$$C_{dB} = 20\log_{10}(100C_M).$$

The initial step size was 8 dB, and halved after each reversal until it reached 1dB, at which it stayed until the staircase terminated. Staircases terminated after six reversals. Selection of the staircase on any given trial was made according to the algorithm described in section 4.2.2.

Each subject completed six sessions, two at each viewing distance, using the ABCBCA ordering, described in section 5.2.3.1. Within a session, there were five different conditions: the four target edge scales, and the control condition. The conditions are summarized in Table 6.2. Each condition in a session was assigned to two staircases, converging to *Detection* performance levels of 71% and 79%, as described earlier.

6.2.4 Results and discussion

6.2.4.1 Data analysis

For each condition, the proportion of correct responses for each judgement (*Detection* and *Identification*) at each contrast level were used to derive psychometric functions, as described in section 4.2.3. Experiment 3 differed from the other experiments in this thesis in that the task was objective, so that each response could be judged correct or incorrect, and so, unlike in the other experiments, the psychometric functions from experiment 3

would be expected to vary from 0.5 (chance level) to $(1-\lambda)$ for $\lambda \ll 1$. Therefore, γ in equation (4.1) was set to 0.5 and, as in the other experiments, λ was constrained to fall between 0 and 0.05. As in the other experiments, the threshold was taken to be $F^{-1}(0.5)$, where F^{-1} is the inverse of the function F in equation (4.1). This corresponded to a performance level (proportion correct) of $(0.75-\lambda/2)$.

6.2.4.2 Are edges of opposite polarity perfectly discriminated?

If edges of opposite polarity are perfectly discriminated, the polarity of an edge can be identified whenever it is detected, and so the thresholds for detection and polarity identification should be the same. The results of experiment 3 are shown in Figure 6.2. One thing that stands out is that the thresholds for detection and identification are quite similar: in each graph, the two lines follow each other very closely. Bizarrely, the thresholds tend to be slightly lower for identification than detection. This is a small effect, but very reliable: 40 out of the 45 pairs of data points in Figure 6.2 show a lower threshold for the identification task. This means that, in the trials in which the wrong interval was chosen as the target, the target polarity was judged correct more than half the time. It is difficult to imagine an edge detection model which can judge the polarity of edges that it cannot detect. One briefly considered possibility was that, in trials in which the target came first, even if the target had not been seen, its afterimage might have been seen during the presentation of the subsequent zero-contrast non-target. However, the afterimage would have been of opposite polarity to the target, making the subject less, rather than more likely to make a correct polarity judgement. In any case, an analysis of the data after separating the trials according to the order of target and non-target within a trial (shown in Figure 6.3 and Figure 6.4) shows the same effect in both data sets, so it is unlikely that the effect could be caused by an interaction between the stimuli within a trial. A possible explanation is that the detection judgment may have had more response errors (i.e. responses that differed from the intended response) than the polarity judgment. This is a promising possibility because of the way that response options were assigned to key positions: the assignment of key position to target interval was arbitrary (up/down), whereas the assignment of key position to polarity was non-arbitrary (left/right). A subject

on "autopilot" might therefore be less likely to make response errors for the polarity judgement. In addition, the assignment of key position to target interval differed from the usual left/right mouse button assignment that all the subjects had used previously to indicate the target interval in standard 2-interval forced choice experiments, which could have been a further cause of difficulty.

It might be argued that, in fitting the psychometric function to data, the λ parameter in equation (4.1) would have been adjusted to compress the psychometric function vertically, so that the threshold was not affected by response errors. Indeed, that was Wichmann & Hill's (2001a) motivation for including this parameter. However, this adjustment would only happen if response errors occurred at high contrasts, when the stimuli were easily visible. It is more likely that the response errors occurred in the low contrast range, because the staircases only occupied high contrasts at the start of a session (when the subjects would have been less fatigued). The staircases quickly converged to lower contrasts, and so response errors due to fatigue near the end of a session would have been in the low contrast range. This would have tended to push the threshold towards higher contrasts.

Whatever the explanation of the lower contrast thresholds for identification than detection, the results demonstrate quite convincingly that, if an edge can be detected, its polarity can be identified, i.e. edges of opposite polarity are "perfectly discriminated", as predicted by the GDT model. Thus, the GDT model passed a test that it might have failed.

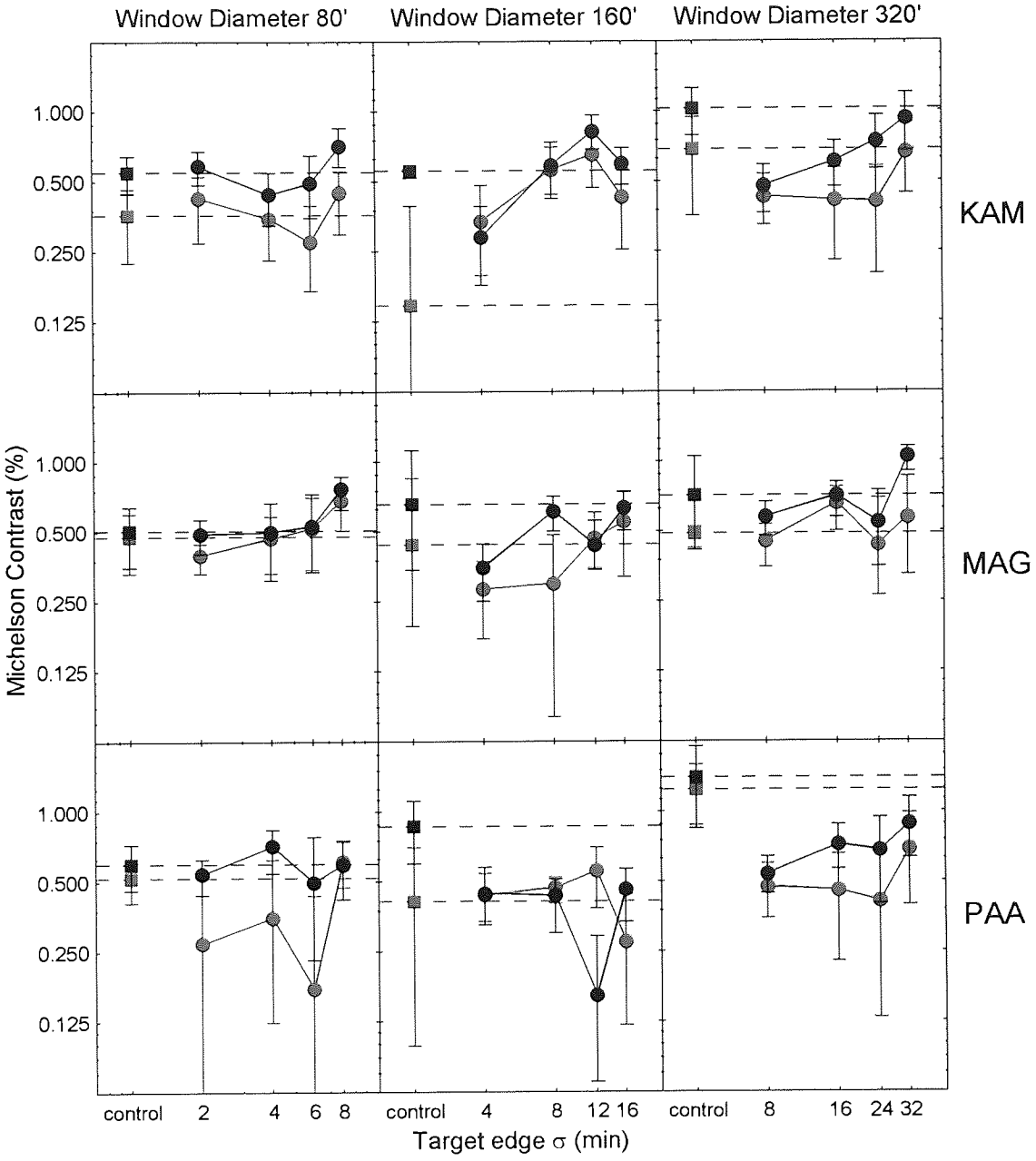


Figure 6.2. Results of experiment 3. Black symbols represent detection thresholds, and red symbols represent identification thresholds. Circles represent Gaussian edges, and squares represent the control stimuli. The dashed lines pass through the square symbols, to facilitate comparisons between the control thresholds and thresholds for each of the Gaussian edges.

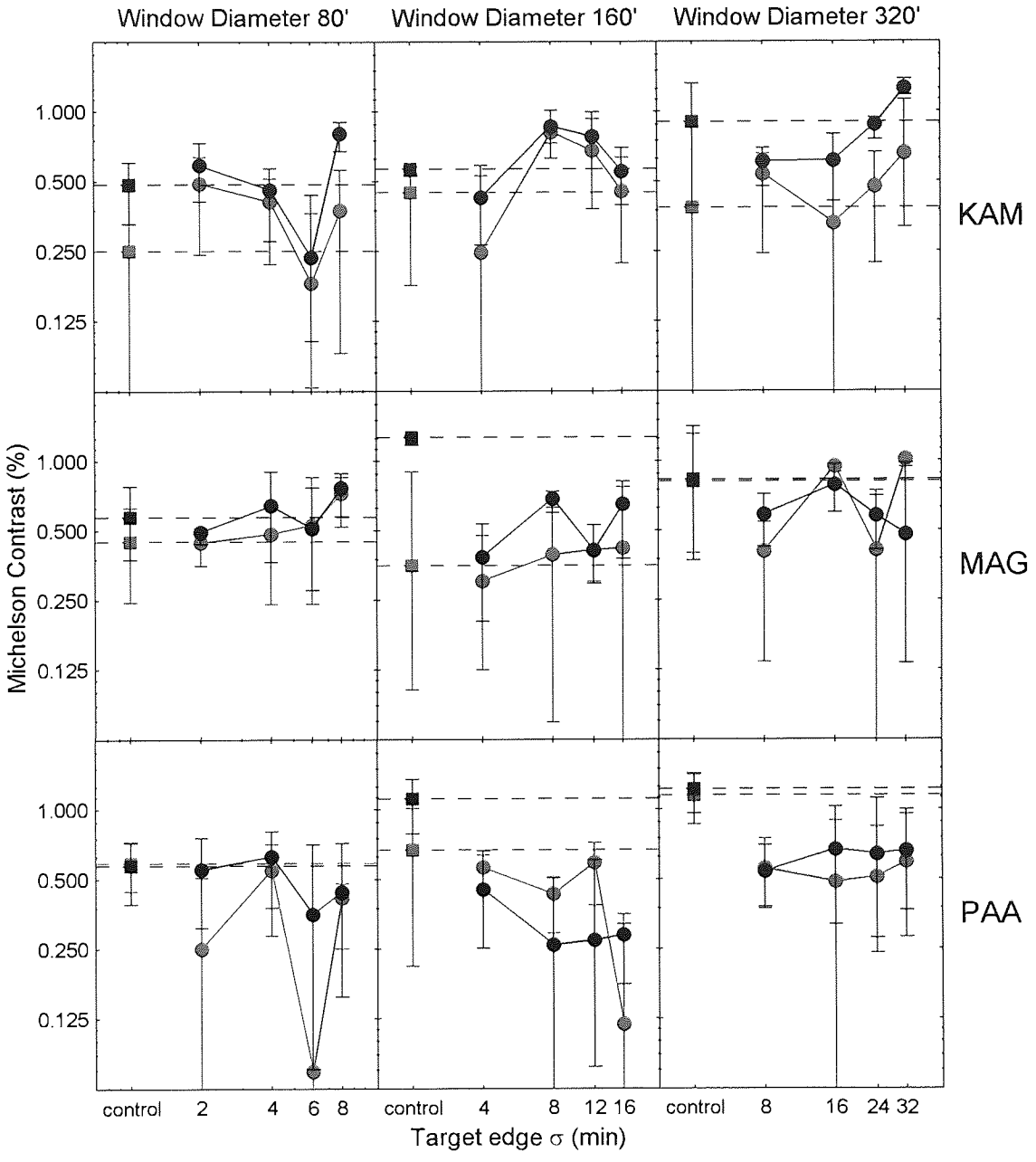


Figure 6.3. Data from trials in experiment 3 in which the target preceded the non-target. Symbol colour and shape as in Figure 6.2.

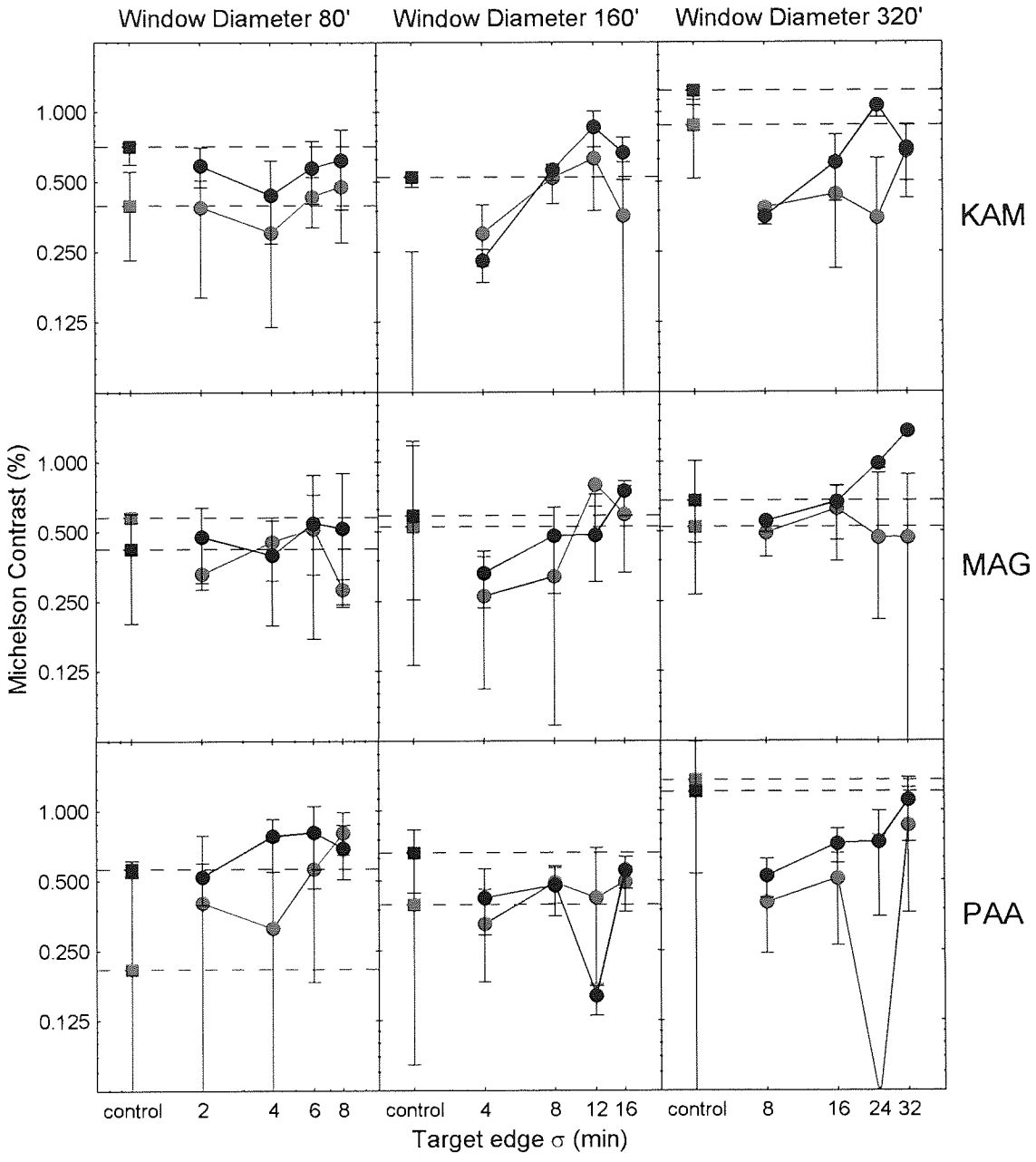


Figure 6.4. Data from trials in experiment 3 in which the non-target preceded the target. Symbol colour and shape as in Figure 6.2.

6.2.4.3 Variation of detection threshold with edge scale

With the smallest circular window diameter, Figure 6.2 shows that the detection threshold was approximately constant across all scales, but this result is difficult to interpret because the control stimulus had a very similar detection threshold. Thus, with the 80′ window, the detection threshold for the central edge was no lower than the threshold for detection of other luminance differences in the stimulus. Performance could therefore be based on these other luminance differences, and one would expect no variation in threshold with scale of the central edge. With the other window diameters, the detection threshold for the central edge was generally lower than for the control stimulus. In these conditions, there was a general tendency for the threshold to increase with increasing edge scale, although the trend was not strong. The geometric mean contrast threshold across subjects was 0.355% for the 4′ edge, and 0.948% for the 32′ edge. According to equation (2.4) (repeated as equation (6.5), below), these contrasts corresponded to peak gradients of 0.0425 and 0.0142 units/deg, respectively.

If contrast threshold were proportional to edge scale, as hypothesised in section 6.1, then the data in Figure 6.2 should fall on a straight line of gradient 1, since they are plotted on log-log axes. The slope is clearly less than one. The prediction that contrast threshold would be proportional to edge scale was based on the assumption that the gradient signal in the smallest channel was equal to the gradient of the edge. If the smallest channel introduced significant blurring (or if there were other sources of blurring, such as optical blurring (Williams *et al.*, 1994) or other intrinsic blur (Levi & Klein, 1990; Wuerger *et al.*, 2001)), then the sharpest edges would be the most strongly affected, and would behave as if they were more blurred, which would make the slope of the data flatten out. This possibility was investigated by finding the combination of blur scale and gradient threshold which best fitted the data.

The calculation was performed as follows. Section 2.1 derived the peak gradient of a Gaussian edge:

$$E'_{max} = \frac{2C}{\sigma_e \sqrt{2\pi}}. \qquad (6.5)$$

If the edge is subjected to Gaussian blurring of scale, σ_b, then the effective peak gradient will be given by

$$E'_{\max} = \frac{2C}{\sqrt{(\sigma_e^2 + \sigma_b^2)2\pi}}.$$

If the contrast threshold for edge detection corresponds to a constant gradient threshold value, T, applied to the edge after Gaussian blurring with scale σ_b, then the predicted contrast threshold will be $(T/2)\sqrt{(\sigma_e^2 + \sigma_b^2)2\pi}$. The predicted threshold values were found for the 160′ and 320′ window conditions for a range of σ_b values (0.05′ to 30′ in steps of 0.05′) and T values (0.001 to 0.05 units/deg in steps of 0.001). Only the 160′ and 320′ window diameter conditions were modelled because, in the 80′ condition, the threshold was being determined by luminance differences other than the central edge. Using the least-squares method, the best-fitting pair of parameters was found for each subject's data, and for the data averaged across the three subjects. The contrast threshold values were expressed in dB, both for averaging across subjects and for finding the sum of squared differences between predicted and actual values. Thus, the mean contrast threshold values were the geometric mean of the Michelson contrast values. The best-fitting parameters are shown in Table 6.3, and Figure 6.5 plots the predicted detection thresholds against the data. Although the predictions fit the data reasonably well, the amount of blurring required (around 16′) is implausibly high. Intrinsic blur has previously been estimated to be about 1 arc min (Levi & Klein, 1990; Wuerger *et al.*, 2001). There is some evidence that intrinsic blur increases at lower contrasts, but the highest value reported was 1.5′ at 3% contrast (Wuerger *et al.*, 2001). The best-fitting gradient threshold was found when σ_b was fixed at 1′ or 2′ but, as Figure 6.5 shows, the predicted slopes are considerably steeper than those shown in the data. In conclusion, it would seem that the detection thresholds do not arise from a constant gradient threshold after blurring.

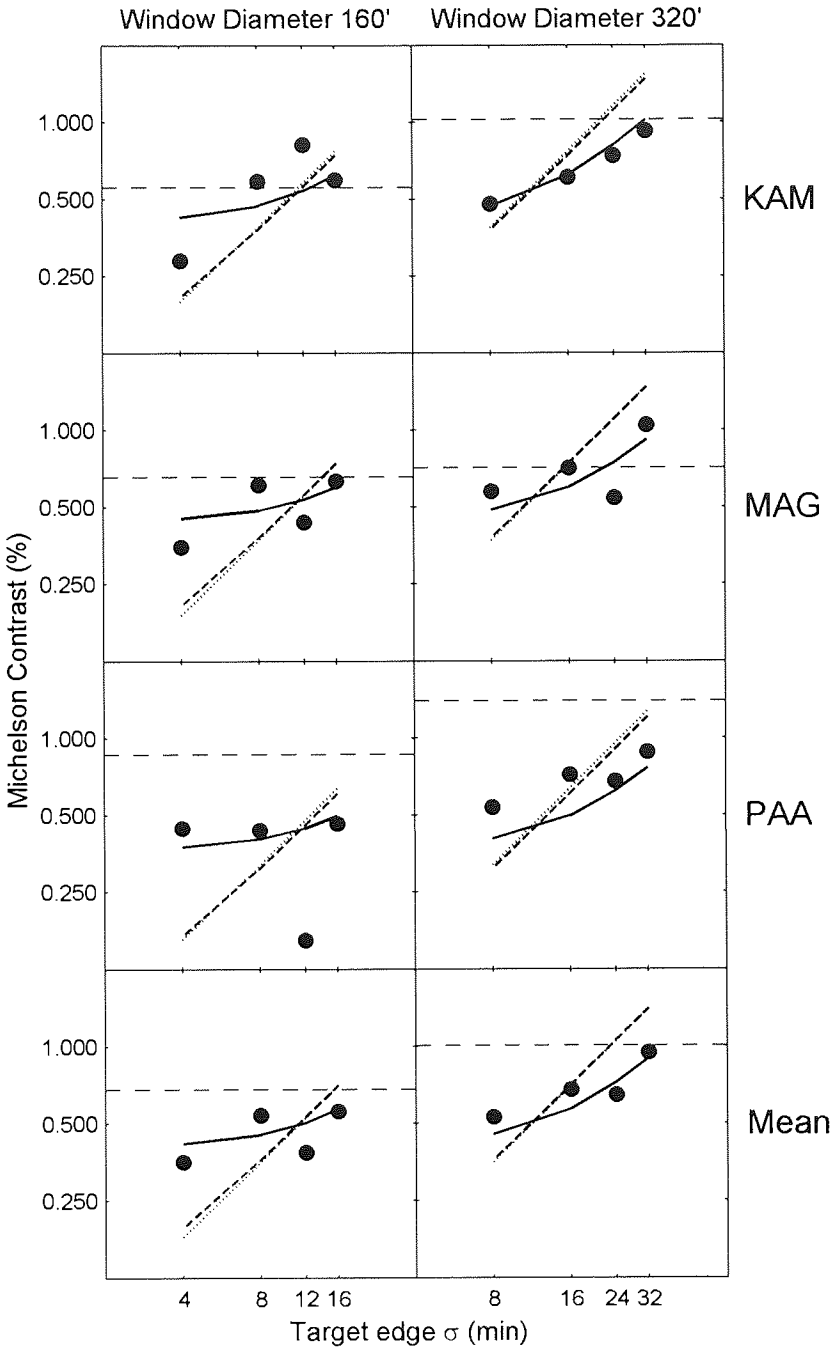


Figure 6.5. Circles represent detection thresholds for Gaussian edges from Figure 6.2, and the horizontal dashed lines represent the detection thresholds for the control stimulus. The solid lines represent the thresholds predicted from the best-fitting parameters in Table 6.3. Sloping dotted lines and dashed lines represent the best-fitting predictions when σ_b is fixed at 1′ and 2′, respectively.

Table 6.3. Best-fitting parameters for predicting the contrast threshold in experiment 3, assuming that the contrast threshold corresponds to a gradient threshold of T after blurring with a Gaussian kernel with scale σ_b. The two columns on the far right give the best-fitting T value when σ_b is fixed at 1′ or 2′. The parameters in the "Mean" row are fitted to the geometric mean across the 3 subjects.

Subject	Best-fitting T and σ_b parameters		Best T when σ_b = 1 arc min	Best T when σ_b = 2 arc min
	T (units/deg)	σ_b (arc min)		
KAM	0.014	13.90	0.023	0.022
MAG	0.012	17.60	0.022	0.022
PAA	0.010	17.55	0.019	0.018
Mean	0.012	16.15	0.021	0.021

Chapter 7
Modifying the rectifier

In Chapter 5 it was suggested that the data might be explained by modifying the rectifier so that it had a threshold. Thus, the rectifier becomes

$$H(I_x) = \lfloor I_x - T \rfloor . \tag{7.1}$$

This kind of function is called a *threshold-linear* function. Section 7.1 reviews the literature on threshold models, and related models, in computer science, psychophysics and physiology. Section 7.2 introduces two further modifications to the rectifier.

7.1 Threshold models in computer science, psychophysics and physiology

7.1.1 Computational threshold models

Most edge detection algorithms in machine vision have thresholds to remove noise. Usually, unimportant edges are assumed to have a low gradient, and a threshold is applied to the gradient signal, as in equation (7.1). This approach was used in the earliest edge detection algorithms, such as Kovasznay & Joseph's (1955) algorithm, and the enhancement/thresholding algorithms described in section 1.2. It has been used in many subsequent algorithms. Canny (1986) introduced a more sophisticated thresholding scheme, which he called hysteresis thresholding (described in section 1.8.2), but this was

still a threshold on the gradient. Gradient thresholds are not always successful in removing noise, because high frequency noise can have a high gradient, and many other thresholding schemes have been proposed (reviewed by Rosin, 1997).

A threshold on the gradient signal is a critical element of many computational models of lightness perception (reviewed by Hurlbert, 1986). The key task is to separate the surface reflectance function from the illumination function. It is assumed that any changes in illumination will be gradual, giving rise to low gradients. The image is spatially differentiated, and then the low gradients generated by variations in illumination are removed with a threshold. The thresholded gradient signal is then integrated to recover the reflectance function.

7.1.2 Psychophysical threshold models of contrast coding

Georgeson & Sullivan (1975) found that above-threshold gratings of different spatial frequency appear to match in contrast when their physical contrasts are almost equal, despite vast differences in contrast threshold. They called this phenomenon *contrast constancy*. They attributed it to a gain control mechanism in each spatial frequency channel, which compensates for the attenuation of that channel's frequencies by early visual mechanisms, including the optics of the eye.

Kulikowski (1976) showed that contrast constancy could be explained by assuming that the internal contrast code had the following form:

$$C' = k\lfloor C - C_0 \rfloor^n, \tag{7.2}$$

where C' is the perceived contrast, C is the physical stimulus contrast, and C_0 is the contrast threshold. If two stimuli are both substantially above threshold, the C_0 terms become insignificant, and the stimuli match in contrast when their physical contrasts are about the same. Kulikowski estimated the exponent, n, to be 1, so that the internal contrast code was a threshold-linear function of physical contrast. Cannon (1979) used

subjective ratings of perceived contrast, and found that the data were well-fitted by equation (7.2), again with $n = 1$. Subsequently, there has been some variation in the estimated exponent (see Georgeson, 1991b, for a review), but it seems clear that the threshold term plays an important role.

Georgeson (1991b) argued that Kulikowski's explanation of contrast constancy cannot be the whole story because, although it fits the data very well, the visual system does not have access to the physical contrast: the image is known to be attenuated by multiplicative factors, so that the signal that is delivered to each spatial frequency channel will have contrast of the form $g\lfloor C - C_0 \rfloor^n$, where g the channel's attenuation, the value of which varies across channels. In order for contrast constancy to occur, there must be a multiplicative gain control to compensate for this attenuation. Georgeson (1991b) presented a model that contained both subtraction of a threshold and multiplicative gain control.

7.1.3 Models of simple cells

7.1.3.1 Threshold-linear models

In section 1.6, a distinction was drawn between the underlying membrane potential of a simple cell, and its overt response (i.e. firing rate). It has been suggested that the underlying membrane potential shows linear spatial summation, and that the overt response is an essentially linear function of membrane potential, except for half-wave rectification (Movshon *et al.*, 1978c; Schumer & Movshon, 1984). Many apparent non-linearities can be accounted for simply by assuming that the rectifier has a threshold.

Researchers have tested simple cells for linearity in a number of ways. Enroth-Cugell & Robson's (1966) original *null test* for linearity in retinal ganglion cells examined whether there was a spatial phase at which a sine wave grating elicited no response from the cell. Their reasoning was that, when the input to a linear system is a sine wave grating, the output is a sine wave with the same frequency. The output therefore has two zero, or null, points per cycle. By varying the input spatial phase, it should be possible to vary the output phase so that the stimulus elicits no response. Ganglion cells that showed a null

phase were termed X-cells, and cells that did not were termed Y-cells. Hochstein & Shapley (1976b) pointed out that the null test is quite a weak test of linearity, because "it might be possible to hypothesize a fortuitous cancellation of nonlinearities yielding the existence of null responses" (p. 245). So, although linearity implies null phases, the converse of this proposition is not necessarily true: having null phases does not imply that the cell is linear. In fact, Movshon *et al.* (1978c) showed that null phases are a necessary property of a linear system only if the excitatory and inhibitory regions differ in spatial or temporal phase by an integer multiple of π. In general, as the spatial phase of the stimulus is varied, the amplitude and temporal phase of the response of a linear cell should trace out an ellipse, centred on the origin, when plotted in polar coordinates. The response amplitude never falls to zero, unless the excitatory and inhibitory regions differ in spatial or temporal phase by a multiple of π, in which case the ellipse collapses into a line segment that passes through the origin. The excitatory and inhibitory regions of ganglion cells have the same spatial phase, so these cells would be expected to have null phases; in general, though, simple cells would not be expected to have this property. Movshon *et al.* (1978c) plotted the response amplitude against temporal phase for one simple cell and found what they described as a "wasp-waisted ellipse", a finding replicated by several other researchers (Reid *et al.*, 1987, 1991; Tolhurst & Dean, 1991). Movshon *et al.* showed that this result could be predicted by assuming that the overt response arose from over-rectification of the membrane potential: adding a threshold of 8 spikes/s to each amplitude allowed the data to fall almost perfectly onto an ellipse.

Another test of linearity is to compare a cell's impulse response function (i.e. its receptive field profile) with its spatial frequency tuning curve. For a linear system, the tuning curve should be the Fourier transform of the impulse response function. The majority of researchers have found that the measured tuning curve is narrower than predicted by the impulse response function, or, equivalently, that the impulse response function has fewer excitatory or inhibitory regions than predicted from the tuning curve (Maffei *et al.*, 1979; Andrews & Pollen, 1979; Glezer *et al.*, 1980; Kulikowski & Bishop, 1981; Field & Tolhurst, 1986). A threshold provides an obvious qualitative explanation of this non-linearity, because low-amplitude ripples far from the receptive field centre could fall below threshold. Tadmor & Tolhurst (1989) performed a quantitative analysis of the predictions of the threshold model, and found that it could indeed account for this discrepancy. Not all researchers have reported a discrepancy between impulse response function and tuning

curve, however. Movshon *et al.* (1978c) were the first to examine this issue, and reported an excellent agreement between observed and predicted line-weighting functions; in particular, they reported that excitatory or inhibitory regions "were never absent from our line-weighting measurements when their appearance was predicted." (Movshon *et al.*, 1978c, p. 72). Dean & Tolhurst (1983) also found a good agreement between observed and predicted line-weighting functions, as long as the cell showed no overlap between adjacent excitatory and inhibitory regions, as required by Hubel & Wiesel's (1962) original definition of simple cells. In addition, Jones & Palmer (1987) found a very good agreement between measurements made in the spatial and frequency domains. Palmer suggested, in personal communication to Heeger (1992a), that the discrepancies in other studies may have arisen because spatial frequency tuning curves are typically measured using larger stimuli than those for measuring the impulse response function.

There have been several other results that have an obvious interpretation in terms of a threshold-linear model. Robson *et al.* (1988) measured the response of a simple cell to a sine wave grating, as a function of spatial phase of the grating. Instead of the sine wave response predicted from linear systems theory, the response function appeared to be a sine wave after over-rectifying the positive and negative portions separately.

Tolhurst & Dean (1987) tested simple cells to see if they possessed the additivity property of linear systems, i.e. the property that the response to the sum of two stimuli should be the sum of the responses to the individual stimuli. The stimuli were bars whose luminance was modulated sinusoidally over time about the background luminance level. They presented one bar in the excitatory region of a simple cell's receptive field, and another, in antiphase, in the inhibitory region. They found that the sum of the responses to the two stimuli presented separately was less than the response to the stimuli presented simultaneously. This can be explained easily by the threshold-linear model, because the sum of responses to the individual stimuli would involve subtraction of two thresholds (one for each response), whereas only one threshold would be subtracted from the response to the summed stimulus. Tolhurst & Dean also measured the response to a bar presented in the excitatory region as a function of contrast. In one condition, the bar was presented alone, and in the other conditions it was accompanied by a fixed-contrast conditioning bar presented in the inhibitory region. The conditioning bar was either in phase or in antiphase with the bar presented in the excitatory region, so it either inhibited or facilitated

the response. With no conditioning bar, the contrast-response function had the appearance of a biased rectifier: there was no response to below a certain threshold contrast, and, above threshold, the response was approximately a straight-line function of contrast. With a facilitatory conditioning bar, the response was raised above threshold for all contrasts, and was a straight-line function of contrast; with an inhibitory conditioning bar, the response function was well predicted by the threshold-linear model.

7.1.3.2 Power-law/normalization models

Despite its many attractions, the threshold-linear model has three major problems. One is that simple cells can be inhibited by stimuli of almost any kind, including stimuli which do not evoke a response in the cell when presented alone. Although it is possible that the threshold-linear model could be modified to include nonspecific suppression, it does not do so naturally. A second problem is that the threshold-linear model incorrectly predicts that orientation and spatial-frequency tuning will become broader with increasing contrast, because the tails of the tuning curves will rise above threshold; it has been shown that orientation and spatial frequency tuning of cells is contrast-invariant (Sclar & Freeman, 1982; Skottun *et al.*, 1987). The third main problem is that the response of cells saturates with increasing contrast, deviating considerably from a straight line. Albrecht & Hamilton (1982) evaluated four different functions to see which best described the contrast-response function. The candidate functions were a straight line, a logarithmic function, a power law, and the Naka-Rushton equation (Naka & Rushton, 1966):

$$R(C) = \frac{R_{max} C^n}{S^n + C^n} .$$

R is the response of the cell, C is the stimulus contrast, and S is the semi-saturation constant, equal to the contrast at which the cell responds at half its maximum rate. The Naka-Rushton equation was the clear winner, providing the best description of the contrast response function for 70% of the cells, and still giving an acceptable fit for the other cells. The best-fitting exponent, n, had a mean value of 2.9. Sclar *et al.* (1990) also used the Naka-Rushton equation and obtained a median exponent of 2.4. Carandini *et al.*

(1997) used a very similar equation and found median values of n between 2.37 and 2.61 in different experiments.

Heeger (1991, 1992a, 1992b, 1993) used the Naka-Rushton equation as the basis of a formal model of cell responses, henceforth called the *normalization model*. Cells were modelled as half-squared linear operators, i.e. linear operators whose response is half-wave rectified and then squared:

$$H = \lfloor L \rfloor^2 .$$

L is the output of the linear operator, and H is the output of the half-squaring process. The response, R_i of simple cell i is then determined by a form of the Naka-Rushton equation:

$$R_i = \frac{kH_i}{S^2 + \sum_j H_j},$$

where k and S are constants for the cell. The inclusion of H_i in the summation term in the denominator ensures that the cell saturates in a realistic way; the other summation terms generate nonspecific suppression. The nonspecific suppression was initially thought to arise entirely from inhibition from other cortical cells. It is likely that some suppression phenomena, such as surround inhibition, do have a cortical origin (Carandini *et al.*, 2002; Freeman *et al.*, 2002). However, Freeman *et al.* (2002) have recently presented evidence suggesting that mechanisms in the LGN are responsible for suppression arising from masks presented within a cell's receptive field. Firstly, cortical cells can be suppressed by rapidly drifting stimuli that stimulate LGN cells but do not elicit much of a response from cortical cells. Secondly, adaptation to the mask orientation does not reduce the mask's suppressive effect. Since adaptation substantially reduces the responses of V1 cells (Maffei *et al.*, 1973; Albrecht *et al.*, 1984; Ohzawa *et al.*, 1985), but hardly affects responses of LGN cells (Shou *et al.*, 1996; Sanchez-Vives *et al.*, 2000), this second result is difficult to explain with an intracortical suppression model, but has a straightforward explanation in terms of suppression from the LGN. Carandini *et al.* (2002) presented a model which

showed that many suppression phenomena can be explained by synaptic depression at thalamocortical synapses.

Heeger preferred to use an exponent of 2 because the sum of four half-squared linear operators, differing in phase by 90°, is an energy mechanism. Adelson & Bergen (1985) had proposed a similar energy mechanism as a model of complex cells. A particular advantage of an energy mechanism is that it gives an unmodulated response to drifting sine wave gratings, as seen with some complex cells.

Response saturation and nonspecific suppression are properties of cells that are not accounted for by the threshold-linear model, but emerge naturally from the normalization model. Another improvement over the threshold-linear model is that, in the normalization model, the ratio of responses to any two stimuli is independent of contrast, so the model shows contrast-invariant orientation and spatial frequency tuning. In addition, Heeger showed that his model could account for all of the nonlinearities that had previously been attributed to thresholds. Tolhurst & Heeger (1997) conducted a head-to-head comparison of the threshold-linear and normalization models. They found that, in most cases, the two models made very similar predictions but, when they differed appreciably, the normalization model was usually superior. For example, one cell showed a wasp-waisted ellipse that required a high threshold to fit the data, but the contrast-response function suggested that this cell had a very low threshold; the normalization model was able to give a reasonable fit to both sets of data.

7.1.3.3 Reconciling threshold-linear and power-law models

More recently, techniques have been developed that allow researchers to measure the membrane potential directly, as well as recording the firing rate. Jagadeesh *et al.* (1993) presented some direct evidence for Movshon *et al.*'s (1978c) proposal that the underlying membrane potential shows linear spatial summation, and that the firing rate is a non-linear function of membrane potential. They measured the directional index (a measure of directional selectivity) of several cells in response to moving gratings. For a linear cell, it should be possible to estimate the directional index from the cell's response to stationary

contrast-modulated gratings. When using firing rate as the response measure, the directional index obtained using stationary contrast-modulated gratings is typically underestimated by a factor of 2 or 3 (Reid *et al.*, 1987, 1991; Albrecht & Geisler, 1991; Tolhurst & Dean, 1991). Jagadeesh *et al.* found that, using the membrane potential as the response measure, the responses to both moving and modulated gratings gave very similar estimates of directional index. Consistent with the previous results, they found that, when firing rate was used as the response measure, the directional index obtained from moving gratings was about 3 times higher than that obtained from stationary contrast-modulated gratings.

An important issue is the nature of the non-linearity that intervenes between membrane potential and firing rate. Jagadeesh *et al.* attributed their findings to a threshold: a cell which is only slightly directionally selective on the basis of its membrane potential may appear more directionally selective if its responses to the nonpreferred direction fall below threshold. However, this nonlinearity can also be accounted for using the squaring nonlinearity of the normalization model (Tolhurst & Heeger, 1997). Initially, it might seem that the normalization model is preferable, because, unlike the threshold model, it predicts that orientation tuning is contrast invariant. Miller & Troyer (2002) proved that a power-law transducer is both necessary and sufficient for contrast-invariant tuning. However, Carandini & Ferster (2000) measured instantaneous spike rate as a function of instantaneous membrane potential, and found that the threshold-linear model was well-supported. This apparent paradox can be resolved by proposing that, although the *instantaneous* spike rate is a threshold-linear function of *instantaneous* membrane potential, the spike rate of threshold-linear units averaged over time, or over a population of cells, can approximate a power function if noise is added to the membrane potential. Suarez & Koch (1989) proved that adding a uniform distribution of noise to a threshold-linear function results in a quadratic power-law function, as required by Heeger's half-squaring model. Miller & Troyer (2002) showed numerically that adding Gaussian noise to the membrane potential of a threshold-linear unit results in a good approximation to a power law, as long as the membrane potential does not rise much more than 1.5 standard deviations of noise above threshold. Anderson *et al.* (2000) showed that a threshold-linear unit with added noise showed contrast-invariant orientation tuning, as would be expected if the noise gave rise to a power law function. One reason that Miller & Troyer preferred their scheme to that of Suarez & Koch is that Suarez & Koch's method can only

yield an exponent of exactly 2, whereas, in Miller & Troyer's scheme, the best-fitting exponent is a decreasing function of the noise standard deviation. Section 7.2.1 shows how Suarez & Koch's scheme can be extended to yield a power-law function with any exponent and any threshold, for inputs within any specified limits. The exponent is manipulated by varying the shape of the noise distribution. The resulting function is called the *power-function rectifier*.

7.2 Modifications to the half-wave rectifier

This section describes a number of possible modifications to the half-wave rectifier. Section 7.2.1 describes the power-function rectifier. With different parameters, this can take the form of a pure threshold-linear function, or a pure power function, and these forms of the power-function rectifier are described in sections 7.2.2 and 7.2.3. In section 7.2.4, a function based on Legge & Foley's (1980) transducer is described, as an alternative modification to the half-wave rectifier. Sections 7.2.5 and 7.2.6 report the results of simulations 1 and 2, which explore the ability of the modified models to mark edges at the correct locations in a sine wave grating.

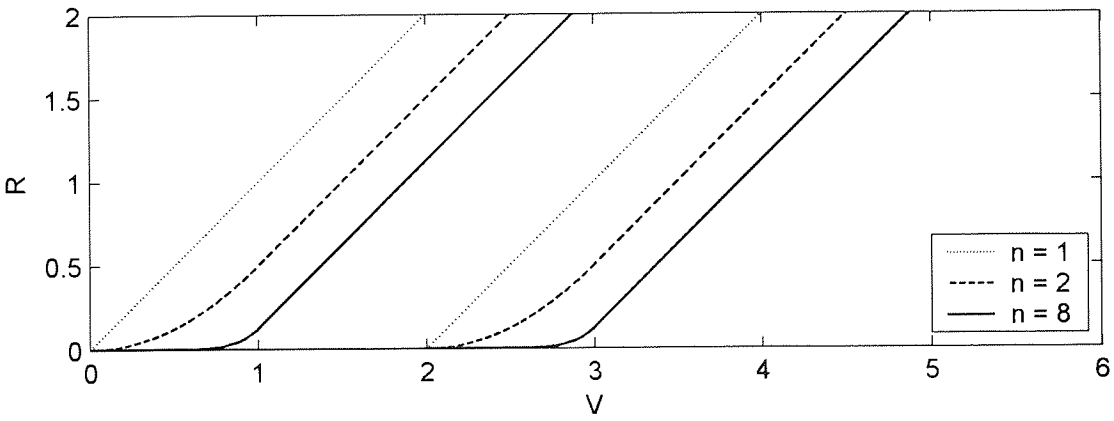


Figure 7.1. The power-function rectifier. The graph shows two sets of three functions. Within each set, the T_1 and T_2 parameters are the same, and the exponent, n, varies, as indicated by the line style. The set on the left has $T_1 = 0$ and $T_2 = 1$; the set on the right has $T_1 = 2$ and $T_2 = 3$. With exponent, $n = 1$, the function is a half-wave rectifier with threshold T_1. With $n > 1$, the function is a power function of order n between T_1 and T_2, and a straight line for values $V > T_2$. As n increases, the function approaches a half-wave rectifier with threshold T_2, but it always remains smooth.

7.2.1 The power-function rectifier

7.2.1.1 Defining the function

The power-function rectifier is illustrated in Figure 7.1. It can serve as a pure half-wave rectifier (with or without a threshold), a pure power function (with or without a threshold), or a smoothed rectifier-like function. For values $V \leq T_2$, the response, R, has the form

$$R = k_1 \lfloor V - T_1 \rfloor^n \tag{7.3}$$

and, for values $V > T_2$,

$$R = V - k_2. \tag{7.4}$$

Thus, R is zero for $V < T_1$, a power function for $T_1 \leq V \leq T_2$, and a straight line with gradient 1 for $V > T_2$. k_1 and k_2 are chosen so that, at $V = T_2$, where the function changes from a power function to a straight line with gradient 1, the two sections of the function have the same value (so the function is continuous) and the same gradient (so the function is smooth at all points). k_1 and k_2 are determined from n, T_1, and T_2 as follows.

For $T_1 \leq V \leq T_2$,

$$dR/dV = nk_1 (V - T_1)^{n-1}.$$

The requirement that $dR/dV = 1$ when $V = T_2$ gives

$$k_1 = \frac{1}{n(T_2 - T_1)^{n-1}}.$$

Substituting this value into (7.3) reveals that, when $V = T_2$, $R = (T_2 - T_1)/n$. After substituting these values into (7.4), we can solve for k_2, to give the following complete definition of the power-function rectifier:

$$R = \begin{cases} \dfrac{\lfloor V - T_1 \rfloor^n}{n(T_2 - T_1)^{n-1}} & V \le T_2 \\ V - \dfrac{(n-1)T_2 + T_1}{n} & V > T_2 \end{cases}. \qquad (7.5)$$

If $n = 1$, then $R = \lfloor V - T_1 \rfloor$, i.e. it is a half-wave rectifier with threshold T_1. If $n \neq 1$ and T_2 is higher than all possible values of V, then R is effectively a pure power function, with optional threshold T_1, which can be set to 0 if no threshold is required. As $n \to \infty$, R approaches a half-wave rectifier with threshold, T_2. This is proved as follows. For $V < T_1$, $R = 0$, as required. For $T_1 \le V \le T_2$, we have

$$R = \left(\frac{V - T_1}{T_2 - T_1}\right)^{n-1} \left(\frac{V - T_1}{n}\right). \qquad (7.6)$$

Since $(V - T_1)/(T_2 - T_1) < 1$, both multiplicands in (7.6) are decreasing functions of n, so as $n \to \infty$, $R \to 0$. For $V > T_2$, $R \to V - T_2$. Thus, for $V \in \mathbb{R}$, as $n \to \infty$, $R \to \lfloor V - T_2 \rfloor$. For finite values of n, with values of T_2 that fall within the range of possible values of V, R is a smoothed rectifier.

7.2.1.2 Constructing the power-function rectifier from threshold-linear units

The power function rectifier can be constructed by adding noise to a threshold-linear function or, equivalently, by adding the outputs of a set of threshold-linear functions with different thresholds. Take the standard threshold-linear function, with fixed threshold, T:

$$R = \lfloor V - T \rfloor.$$

Suppose the input has added noise, v, with mean value $\bar{v}$:

$$R = \lfloor V + v - T \rfloor .$$

The mean response, $\bar{R}$, will be the same as the mean response to a set of noiseless linear threshold functions, $\lfloor V - t \rfloor$, with mean threshold, $\bar{t} = T - \bar{v}$, and with the distribution of thresholds about the mean, $\bar{t}$, equal to the distribution of noise about the mean, $\bar{v}$. Thus, we have two formally equivalent models: one in which noise is added to a single threshold-linear function, and the other in which the outputs of a set of noiseless threshold-linear functions with different thresholds are added together. In the first model, we can manipulate the distribution of the noise; in the second model we can manipulate the distribution of frequency of occurrence of different threshold values.

Let us consider the second version. There now follows a proof that, in order to generate the power-function rectifier defined in (7.5), the thresholds, t, should range between T_1 and T_2, with a probability density function given by $p(t) = m(t - T_1)^{n-2}$, where m is a scaling factor that ensures that the integral of p is equal to 1. First, an expression for m is derived. We require

$$\int_{T_1}^{T_2} p(t)dt = \int_{T_1}^{T_2} m(t - T_1)^{n-2}\, dt = 1 .$$

The integral can be performed using the substitution, $u = t - T_1$.

$$\begin{aligned} \int_{T_1}^{T_2} m(t - T_1)^{n-2}\, dt &= \int_{0}^{T_2 - T_1} mu^{n-2} du \\ &= \frac{m(T_2 - T_1)^{n-1}}{n-1} . \end{aligned} \qquad (7.7)$$

Setting this integral to 1, as required, gives

$$p(t) = \frac{(n-1)(t - T_1)^{n-2}}{(T_2 - T_1)^{n-1}} .$$

The probability density function of the noise is therefore a polynomial of order $(n-2)$. The mean response, $\bar{R}$, of all the threshold-linear functions is given by

$$\bar{R} = \int_{T_1}^{T_2} p(t)\lfloor V - T \rfloor dt = \frac{n-1}{(T_2 - T_1)^{n-1}} \int_{T_1}^{T_2} (t - T_1)^{n-2} \lfloor V - t \rfloor dt .$$

There are 3 cases, described in sections 7.2.1.2.1 to 7.2.1.2.3, below.

7.2.1.2.1 Case 1: $V < T_1$

In this case, $\lfloor V - t \rfloor = 0$, for all thresholds, so $\bar{R} = 0$.

7.2.1.2.2 Case 2: $V > T_2$

In this case, $\lfloor V - t \rfloor = V - t$ for all thresholds, so

$$\bar{R} = \frac{n-1}{(T_2 - T_1)^{n-1}} \left(\int_{T_1}^{T_2} V (t - T_1)^{n-2} \, dt - \int_{T_1}^{T_2} t (t - T_1)^{n-2} \, dt \right). \tag{7.8}$$

The first integral can be evaluated using (7.7), and the second integral can be performed using the substitution, $u = t - T_1$:

$$\begin{aligned} \int_{T_1}^{T_2} t (t - T_1)^{n-2} \, dt &= \int_0^{T_2 - T_1} (u + T_1) u^{n-2} du \\ &= \int_0^{T_2 - T_1} u^{n-1} du + \int_0^{T_2 - T_1} T_1 u^{n-2} du \\ &= \frac{(T_2 - T_1)^n}{n} + \frac{T_1 (T_2 - T_1)^{n-1}}{n-1}. \end{aligned}$$

Substituting these results into (7.8) and simplifying gives

$$\bar{R} = V - \frac{(n-1)T_2 + T_1}{n},$$

as required by the definition of $\bar{R}$ (equation (7.5)).

7.2.1.2.3 Case 3: $T_1 \le V \le T_2$

In this case, $\lfloor V - t \rfloor = V - t$ for all thresholds between T_1 and V, and $\lfloor V - t \rfloor = 0$ for all thresholds between V and T_2. Thus, the equation for the mean response, $\bar{R}$, is the same as (7.8), except with an upper limit of V on the integrals, instead of T_2:

$$\bar{R} = \frac{n-1}{(T_2 - T_1)^{n-1}} \left(\int_{T_1}^{V} V (t - T_1)^{n-2} \, dt - \int_{T_1}^{V} t (t - T_1)^{n-2} \, dt \right).$$

Using the results from the previous sections, we get

$$\bar{R} = \frac{n-1}{(T_2 - T_1)^{n-1}} \left(\frac{V (V - T_1)^{n-1}}{n-1} - \frac{(V - T_1)^n}{n} - \frac{T_1 (V - T_1)^{n-1}}{n-1} \right),$$

which simplifies to

$$\bar{R} = \frac{(V - T_1)^n}{n (T_2 - T_1)^{n-1}}.$$

Combining cases 1 and 3 gives

$$\bar{R} = \frac{\lfloor V - T_1 \rfloor^n}{n (T_2 - T_1)^{n-1}},$$

as required by (7.5).

7.2.1.3 Concluding remarks about the power-function rectifier

In summary, the power-function rectifier can take the form of a power function with any exponent (with or without a threshold), a half-wave rectifier (with or without a threshold), or a smoothed rectifier-like function. This function has three parameters: T_1, T_2 and n. The function can be constructed either by adding noise to a threshold-linear function, or by adding the outputs of a set of threshold-linear functions with different threshold values. In the latter case, each threshold-linear function is given by

$$R = \lfloor V - t \rfloor,$$

where t is the threshold. t should have values ranging from T_1 to T_2, with a relative frequency of occurrence, $p(t)$, given by

$$p(t) = \frac{(n-1)(t-T_1)^{n-2}}{(T_2-T_1)^{n-1}}.$$

In the noise model, the noise, v, is added to the membrane potential, V:

$$R = \lfloor V + v - T \rfloor.$$

T is a fixed threshold. This model is equivalent to the multi-threshold model, if we let $t = T - v$. Therefore, to create a power-function rectifier with parameters, T_1, T_2, and n, v should vary between $T - T_2$ and $T - T_1$, with a distribution, $p(v)$, given by

$$p(v) = \frac{(n-1)(T-v-T_1)^{n-2}}{(T_2-T_1)^{n-1}}.$$

The value of the threshold, T, is not critical, because the limits of the distribution of v can be adjusted to give rise to any required values of T_1 and T_2.

To the best of my knowledge, the work presented here is the first proof that any arbitrary power function could arise, with complete mathematical precision, from a noisy threshold-linear mechanism (or the sum of a set of noiseless threshold-linear mechanisms). Suarez & Koch (1989) derived the simple case of a quadratic power function. In this case, $n = 2$, and the distribution of noise or threshold levels is flat. Miller & Troyer (2002) criticized their approach, because it could only yield an exponent of exactly 2. The work presented here addresses this criticism. Suarez & Koch mentioned in passing that a power function of any exponent, *n*, could be generated from a flat noise distribution if the function, *R*, was a polynomial of order $(n-1)$. Here, it has been shown that a power function of any exponent, *n*, would be generated if *R* is a threshold-linear function with a noise probability density function that is a power function with an exponent $(n-2)$. Miller & Troyer also criticised Suarez & Koch's approach because it required the upper limit of the noise to be exactly equal to the unit's threshold in order for a perfect power function to be generated (i.e. $R = \lfloor V \rfloor^n$, rather than $R = \lfloor V - T_1 \rfloor^n$), and so was not robust. This is true, but Miller & Troyer's own function was only an approximation of a power law, and failed to approximate it closely for values of *V* greater than 1.5 standard deviations of the noise. The approach described here has the potential to generate a power function that is mathematically precise between any arbitrarily specified limits (determined by the parameters T_1 and T_2). Any deviation from a perfect power law, due to an undesirable upper limit of the noise, can be quantified analytically, unlike in Miller & Troyer's numerical approach.

7.2.2 A simple threshold-linear model

By setting the exponent in the power-function rectifier to 1, the function becomes the pure threshold-linear function, described in equation (7.1). The simplest possible threshold-linear model has a threshold that is constant across all blur channels. Each channel introduces some blurring, which increases with channel scale, so the threshold in each channel is a threshold on the gradient of the edge after the blurring introduced by that channel's 1st derivative operator. If the threshold is set too high, an edge with a low gradient (low contrast or high scale) will fall completely below threshold in all channels and

will be undetected by the model. Section 7.2.2.1 derives a formula for finding the highest threshold that would allow detection of any particular Gaussian edge.

7.2.2.1 Maximum threshold for detectability of a Gaussian edge

In a channel with a template scale of σ_t, the 1st derivative operator has a scale of $m\sigma_t$. When a Gaussian derivative operator is applied to a Gaussian edge, the scales of the edge and operator add in Pythagorean sum, so the response, I_x, of the 1st derivative operator to a Gaussian edge with scale σ_e will be the 1st derivative of a Gaussian edge with scale $\sqrt{\sigma_e^{\,2}+(m\sigma_t)^2}$. Substituting this value for the scale into equation (2.3) gives the following equation for I_x:

$$I_x = \frac{2C}{\sqrt{2\pi\left(\sigma_e^{\,2}+(m\sigma_t)^2\right)}}\exp\left(\frac{-x^2}{2\left(\sigma_e^{\,2}+(m\sigma_t)^2\right)}\right).$$

The maximum value of I_x occurs when $x = 0$:

$$\max(I_x) = \frac{C}{\sqrt{\frac{\pi}{2}\left(\sigma_e^{\,2}+(m\sigma_t)^2\right)}}.$$

Clearly, the highest peak will occur in the smallest channel (with scale $\sigma_{t_{min}}$), and the threshold must be less than this peak value for the edge to be detected. Thus, an edge with contrast, C, and scale, σ_e, can only be detected if the threshold, T, satisfies the following condition:

$$T < \frac{C}{\sqrt{\frac{\pi}{2}\left(\sigma_e^{\,2}+(m\sigma_{t_{min}})^2\right)}}. \qquad (7.9)$$

This inequality can be rearranged to give the conditions that C and σ_e must satisfy:

$$C > T\sqrt{\frac{\pi}{2}\left(\sigma_e{}^2 + \left(m\sigma_{t_{\min}}\right)^2\right)}$$

or, alternatively,

$$\sigma_e < \sqrt{\frac{2C^2}{\pi T^2} - \left(m\sigma_{t_{\min}}\right)^2}\,.$$

7.2.3 A pure power-law transducer

The parameters of the power-function rectifier can be set up so that the function acts as a half-wave rectifier followed a pure power function, i.e. $H = \lfloor V \rfloor^n$. This is achieved by setting T_1 to zero, and setting T_2 above the gradients of all the stimuli used in the experiment. The approximate similarity between this function and a threshold-linear function has been noted by Carandini *et al.* (1997). Therefore, one might ask whether a power-law transducer could have a similar sharpening effect at low contrast as a threshold-linear function, but without rendering low-gradient edges invisible (since no positive signal is set to zero). But, in fact, having a power function transducer after the half-wave rectifier will produce contrast-invariant blur estimates. The proof is straightforward. Suppose I_1 is a particular image displayed with a contrast of 1, and I_C is the same image at contrast C. Let G_1 and G_C be the outputs of the gradient operator, let H_1 and H_C be the outputs of the transducer, and let S_C and S_1 be the edge strength measure. Then:

$$I_C = CI_1\,,$$

and, from the homogeneity of linear systems,

$$G_C = CG_1\,.$$

Raising both sides to the power of n gives

$$H_C = C^n H_1 .$$

Since the 2nd-derivative and template stage are both linear operations, by homogeneity, we have

$$S_C = C^n S_1 . \tag{7.10}$$

Thus, adjusting the contrast of the image has the effect of multiplying the whole edge strength representation by a constant factor. The position of the peak (in both spatial and scale domains) will be unaffected, so the perceived blur will not vary with contrast.

The contrast-invariant blur estimates that arise from a power-law function can be considered an example of the contrast-invariant tuning shown by power-law transducers (Heeger, 1992b; Miller & Troyer, 2002). The simplest way to prove that tuning along a particular dimension is contrast-invariant is to show that the contrast does not affect the relative responses to stimuli that differ along that dimension (e.g., Heeger, 1992b, p. 186). From (7.10), it is obvious that the ratio between the responses at any two points in the edge strength representation, S, will be contrast-invariant.

7.2.4 A Legge-Foley transducer

An alternative to the power-function rectifier is the transducer proposed by Legge & Foley (1980). In order to explain their psychophysical data on contrast masking, Legge & Foley proposed a model in which the signal was filtered with a set of linear filters tuned to different spatial frequencies, and the output of each filter was passed through a nonlinear transducer that had the following general form:

$$R = \frac{F^p}{S^q + F^q}, \tag{7.11}$$

where R is the response of the transducer, F is the output of the preceding filter, and S, p, and q are constants. In their model, $p = 2.4$ and $q = 2$. Legge & Foley's transducer is a generalization of the Naka-Rushton equation, used by Albrecht & Hamilton (1982), Sclar *et al.* (1990), and Heeger (1991, 1992a, 1992b, 1993), and others. In the Naka-Rushton equation, the exponents are all equal (for example, in Heeger's model, $p = q = 2$). When the exponents are equal, the response saturates, with an asymptote of $R = 1$. When $0 < p - q < 1$, as in Legge & Foley's model, the function never saturates, but it becomes compressive for high inputs. If $p - q = 1$, then the function asymptotes towards the line $R = F$, i.e. the line that describes the standard half-wave rectifier.

If we rectify F, in equation (7.11), then the first two stages of Legge & Foley's model are not unlike the first two stages of the GDT model, because the first stage of the GDT model is a linear filter, and values of S, p, and q can be chosen so that Legge & Foley's transducer function is very similar to the power-function rectifier. The Legge-Foley transducer has a possible advantage over the power-function rectifier in that it has the ability to saturate, and nonspecific suppression can easily be incorporated into it by adding terms to the denominator, as in Heeger's normalization model.

Before examining the success of any of the transducers at predicting the blur-matching data, it was important to check that they did not mark spurious edges. This was examined in simulations 1 and 2.

7.2.5 Simulation 1: feature marking with a Legge-Foley transducer

7.2.5.1 Methods and stimuli

A 1 c/deg sine wave grating stimulus with contrast 1 was created in an image 5 deg wide. The model was run with a Legge-Foley transducer with parameters given in Table 7.1. For each value of S, stimuli were constructed with Legge & Foley's values for p and q (2.4 and 2.0), and two other pairs of values, subject to the constrain that $p - q = 1$: one

of these pairs had the same p value as the standard Legge-Foley transducer (2.4), and the other had the same q value (2.0). All other parameters were as described in Table 3.1.

Table 7.1. Parameters of the Legge-Foley transducer for the six conditions of simulation 1.

S	p	q
0.10	2.4	2.0
0.10	2.4	1.4
0.10	3.0	2.0
0.48	2.4	2.0
0.48	2.4	1.4
0.48	3.0	2.0

7.2.5.2 Results

The results are shown in Figure 7.2. With the higher value of S, the standard Legge-Foley transducer marks edges at the gradient peaks (except near the edge of the stimulus, where the operators would overlap the border of the stimulus considerably). With the lower value, many spurious edges are marked. These spurious edges had estimated Michelson contrasts of around 0.01 or 0.02, which were above detection threshold, as shown by experiment 3. The lower value of S causes the Legge-Foley transducer to become compressive at lower values. The results therefore suggested that compressive transducers are problematic for the model. This idea was explored further in simulation 2.

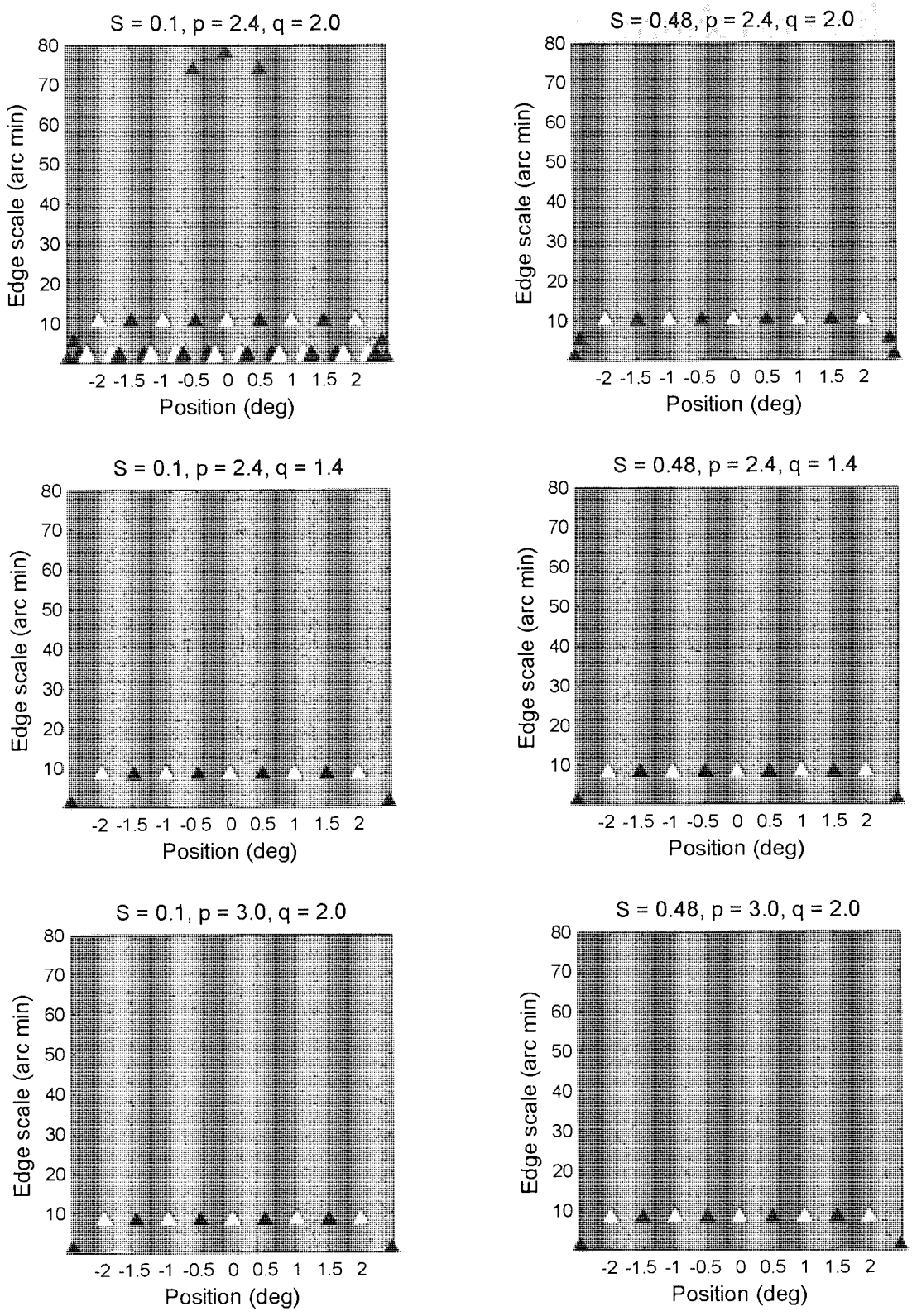


Figure 7.2. Results of simulation 1: feature marking with different Legge-Foley transducer parameters. The vertical axis represents the scale of the marked edges, and the horizontal axis represents their positions. The background of each axis shows the sine wave grating stimulus, scaled to fit the horizontal axis. White symbols represent positive (dark-light) edges, and dark symbols represent negative (light-dark) edges.

7.2.6 Simulation 2: feature marking with the power-function rectifier

Simulation 1 indicated that spurious edges may be marked if the rectifier is replaced with a transducer that is compressive for small values. This was investigated further by examining the features marked when using a power-function rectifier with an exponent less than 1.

7.2.6.1 Methods and stimuli

A 1 c/deg sine wave grating stimulus of contrast 1 was created in an image 5 deg wide. The model was run with a power-function rectifier with the parameters given in Table 7.2. Note that the infinite values of T_2 give a pure power function that does not straighten out. All other parameters were as described in Table 3.1.

Table 7.2. Parameters of the power-function rectifier for the four conditions in simulation 2.

T_1	T_2	n
0	∞	0.5
0.05	∞	0.5
0	∞	2
0.05	∞	2
0	0.3	2
0.05	0.3	2

7.2.6.2 Results

The results are shown in Figure 7.3. The compressive power functions (those with exponent 0.5) marked many spurious edges. The power functions with exponent 2 did not mark spurious edges. This strengthens the conclusion of simulation 1, that spurious features are marked when the rectifier is replaced with a compressive transducer.

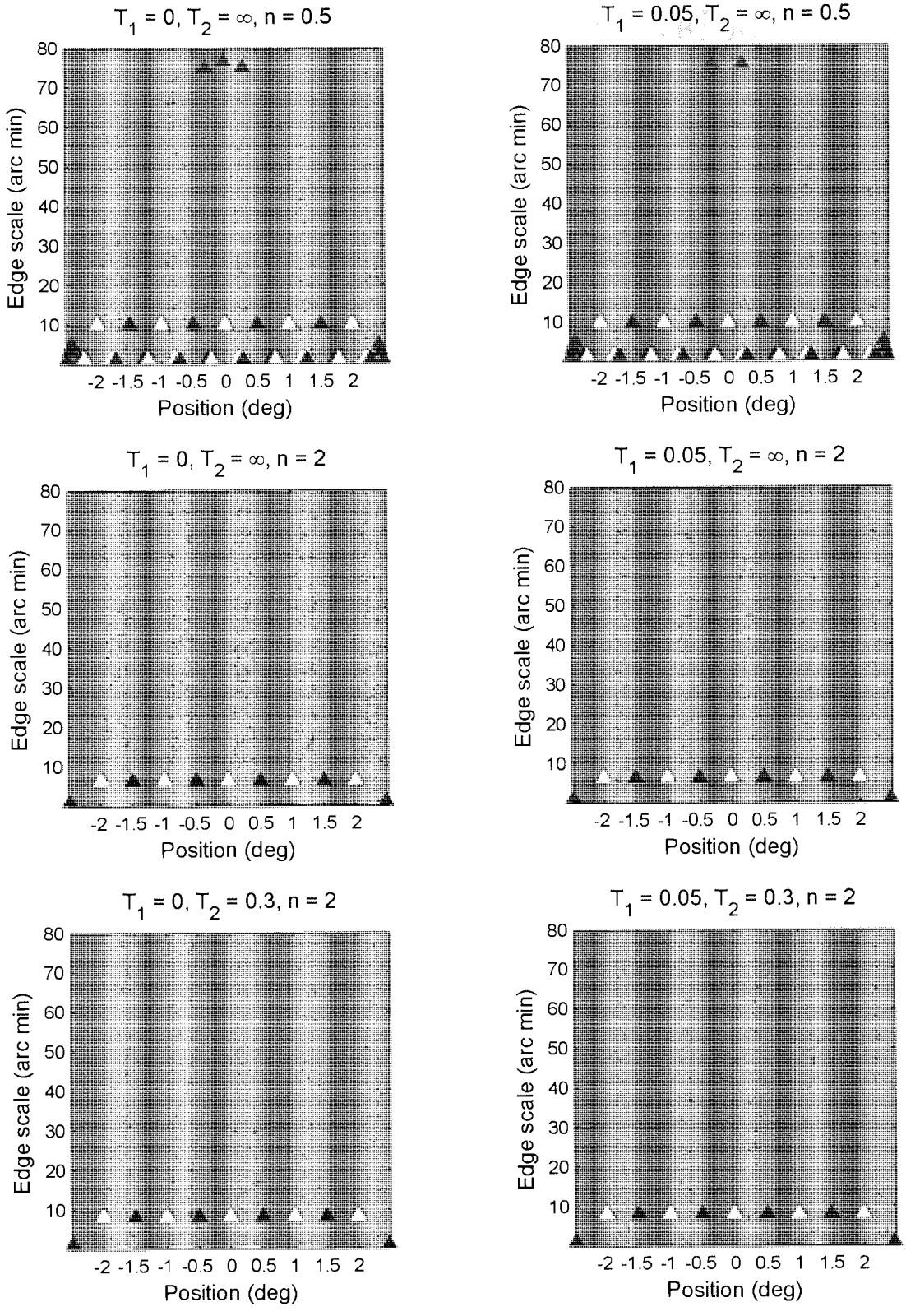


Figure 7.3. Results of simulation 2: feature marking with different power-function rectifier parameters. The top row shows results using a compressive square-root function. The middle row shows a squaring function. The bottom row shows the power-function rectifier which is a straight line for inputs greater than 0.3. The left column shows models with the T_1 parameter set to 0. The right column shows models with the T_1 parameter set to 0.05. For an explanation of the figure, see the legend of Figure 7.2.

7.2.7 Summary of feature marking simulations

Simulation 1 suggested that the Legge-Foley transducer with Legge & Foley's parameters would mark spurious edges. Simulation 2 suggested that this was because of the compressive shape of this function: spurious edges were marked by the power-function rectifier model only when the exponent was less than 1. Simulation 1 showed that the Legge-Foley transducer with exponents that differed by exactly 1 did not mark spurious edges. This function is a smooth rectifier-like function that is quite similar to the power-function rectifier with exponent greater than 1: it is a smooth accelerating function for low inputs, and asymptotes towards the half-wave rectifier function for larger inputs.

In order to avoid spurious edges, the models used for simulating the experiments were the power-function rectifier with exponent not less than 1, and the restricted version of the Legge-Foley transducer with exponents that differed by exactly 1. The Legge-foley transducer used in the simulations is described by equation (7.12):

$$H = \frac{\lfloor I_x \rfloor^p}{S^{p-1} + \lfloor I_x \rfloor^{p-1}}. \qquad (7.12)$$

This transducer has been used previously by Zenger & Sagi (1996) and Meese & Holmes (2002).

Before describing the simulations, Chapter 8 introduces the general modelling methods.

Chapter 8
General modelling methods

8.1 Introduction

The modifications to the half-wave rectifier introduced in Chapter 7 made the simulations of the experiments more complex and time-consuming than they would have been if the pure form of the model, described in Chapter 2, had been used. It was proved in section 2.2.7 that the unmodified model would estimate the scale of a Gaussian edge correctly, and section 2.2.8 derived a formula for correctly estimating the contrast of a Gaussian edge. The veridical perception of Gaussian edges by the model meant that, in any experiment with Gaussian comparison stimuli, the perceived scale or contrast of the test edge would be equal to the physical scale or contrast of the matched comparison edge: simply running the test stimuli through the model would give the predicted results of a blur or contrast matching experiment. But, with the modified half-wave rectifier, the comparison edge is no longer perceived veridically. Therefore, having found the perceived scale or contrast of a test edge, it was necessary to search for the comparison edge that matched the test edge. A binary search method was employed, which is described in the next sections.

8.2 Blur matching with a binary search method

Blur matching experiments find the scale of the comparison edge that matches each test edge in perceived blur. These experiments were simulated by running the model to find its blur estimate for each test edge, and then using a binary search method to find the comparison edge that was perceived by the model to be equally blurred. First, the

perceived blur of the test edge was found. This perceived blur became the target of the search. An initial comparison scale was chosen, and the perceived blur of the comparison edge with this scale was found. If the perceived comparison edge blur was higher than the perceived test edge blur, then the comparison scale was reduced by an initial step size; otherwise, the comparison scale was increased. This process was then repeated with the new comparison scale. The step size stayed the same until the first reversal in step direction. Starting with the first step in the opposite direction to the initial step, each step was half the size of the previous step. The process continued until the perceived comparison edge blur exactly matched the perceived target blur. The requirement of an exact blur match was reasonable: the blur is determined by which of the 256 blur channels contains the peak in the template response, and so the perceived blur of a stimulus could take one of only 256 different values. For each test stimulus, this search procedure usually found the scale of the blur-matched comparison edge within at most 7 steps.

8.3 Modified binary search termination criterion for contrast matching

With the simulations of the blur matching experiments, the perceived blur of a stimulus could take one of only 256 discrete values, since that was the number of channels in the model. Thus, the search could continue until an exact match had been found. With contrast perception, the contrast is a pseudo-continuous quantity calculated from the height of the peak in the edge strength representation, so the number of different possible contrast values was vastly greater, and an exact match in perceived contrast was neither practical nor necessary. Clearly, the search needed to stop when the match was sufficiently good, if not exact.

There are two ways of judging how good the match is. The most obvious is to look at the perceived contrast of the comparison edge after each step, and stop the search if it is sufficiently close to the perceived contrast of the test edge. But this ignores the fact that the purpose of the search is to find the *physical* contrast, C, of the comparison edge that

matches the test edge in perceived contrast. Thus the relevant issue at each step of the search is, not how close the *perceived* contrasts of the test and comparison are, but how close the *physical* contrast of the current comparison edge is to C.

To take an extreme example, suppose there were a stimulus for which a large increase in physical contrast had only a small effect on the perceived contrast. The psychometric functions from the psychophysical experiment would be very shallow but, with enough data, it would be possible to find C with a reasonable degree of accuracy. When modelling this experiment, the *perceived* contrasts of two different comparison edges could be close even if their *physical* contrasts were quite different. Thus, if the search termination criterion was the closeness of the *perceived* contrasts of the test and comparison edges, the search might stop when the physical contrast of the comparison edge was quite different from C, giving an unreliable result.

With the binary search method, it is possible to estimate the physical contrast, C, to any specified degree of accuracy. After the first reversal in the binary search, there is a known interval (a,b) within which C must lie. The size of this interval halves after each step. After each comparison, the current value of the comparison contrast will be at one end of the interval (either a or b). The next step size is $\Delta x = (b-a)/2$, which will take the comparison contrast to a new value $x = (a+b)/2$. Thus,

$$a = x - \Delta x$$
$$b = x + \Delta x .$$

Since C lies within this interval, we have

$$x - \Delta x < C < x + \Delta x , \tag{8.1}$$

and so

$$|x - C| < \Delta x . \tag{8.2}$$

If x is the estimate of C at each step, then Δx is the largest possible error in estimation of C. If the search ends when $\Delta x \leq \Delta x_{min}$, then we can be sure that the error in estimation of C is no higher than Δx_{min}. Rather than setting an upper limit on the *absolute* size of the error, it is preferable to set an upper limit on the error as a *proportion*, p, of C. Thus we require

$$|x - C| < pC .$$

This can be guaranteed if the search is terminated when the following is true:

$$\Delta x \leq p(x - \Delta x). \tag{8.3}$$

This is proved as follows. From (8.1) and (8.3), it follows that

$$\Delta x \leq pC , \tag{8.4}$$

and, from (8.2) and (8.4), we get $|x - C| < pC$, as required. The simulations of contrast matching experiments used $p = 0.01$, and terminated when condition (8.3) was true. Thus, the estimated contrast of the contrast-matched comparison stimulus was always within 1% of the correct value.

8.4 Fitting the model to the data

For each experiment, the model was fitted to the data by simulating the experiment many times, with different sets of parameters, and the best-fitting model was found using a least-squares measure. For a given set of model parameters, the binary search method was used to find the predicted matching comparison scale or contrast for each test edge in the experiment. Scales and contrasts were expressed in dB, defined as follows:

$$x_{dB} = 20 \log_{10}(x) .$$

x is the scale (in arc min) or contrast (Michelson contrast times 100), and x_{dB} is the scale or contrast in dB. For each test edge, the predicted scale or contrast was subtracted from the value found in the experiment[18]. The sum of squared differences was then found across all test edges in the experiment. The best-fitting set of parameters was that which gave the lowest sum of squared differences. The model was fitted to each subject's data individually, because it was anticipated that each subject's visual system might have had different parameters.

This method of finding the best-fitting parameters was used for all the simulations described in this thesis. In addition, the best-fitting parameters across all the experiments for each subject was found by finding set of parameters that minimized the sum of squared differences across all conditions of all experiments. For details of this over-all best fitting procedure, see section 13.3. For most of the simulations, the predictions of the model are shown for the parameter set that fits best to that experiment, and for the parameter set that fits best to all the experiments for each subject.

8.5 1D and 2D simulations

The binary search method was quite time-consuming and had to be repeated many times with different transducer parameters. Although it was quite feasible to use this method with the 1D model, the 2D model ran far too slowly to make this a practical proposition for all the simulations. Because of this, the 1D model was used to simulate the experiments with the binary search method, and the 2D model was run on the test stimuli from each experiment, with a couple of different sets of parameters, to check that it gave the same results as the 1D model for those stimuli.

---

[18] A constant difference in log units, such as dB, corresponds to a constant ratio of linear units. This meant that the choices of units of scale (arc min or deg) or contrast (Michelson or Michelson times 100) are irrelevant, because a linear change of units maintains the ratio between values, so the difference in log units is unchanged. The use of log units is appropriate because blur discrimination follows Weber's law for blur values of greater than about 1′ (Hess *et al.*, 1989; Wuerger *et al.*, 2001), and contrast discrimination functions are reasonably close to Weber's law (Legge, 1981). If the data had been expressed in linear units, then the fit would have been dominated by errors made with large scales and contrasts.

Chapter 9
Simulating the effect of contrast on perceived blur

9.1 Simulation 3: simulating experiment 1.1 with a single threshold model

This simulation used the simplest possible model: a threshold-linear transducer with a single fixed threshold across all blur channels, as described in section 7.2.2. Experiment 1.1 was simulated with a range of different threshold values. The test edge in experiment 1.1 with the lowest gradient had a contrast of 0.05 and a scale of 32'. From equation (7.9), the maximum threshold that would allow this edge to be detected is 0.0748. Informal modelling suggested that this threshold value would be insufficient to explain the effect of contrast on the sharpest test edges, so the range of thresholds used in simulation 3 actually extended beyond this value, despite the knowledge that the lowest-gradient edges would not be detected.

9.1.1 Methods and Stimuli

Each test stimulus from experiment 1.1 was generated using the software that was used to create these stimuli in the experiment, although at a higher resolution. Table 9.1 shows the widths, in both arc min and pixels, of the three different spatial envelopes used in the simulation. The main simulation used the 1D model; the image matrix given to the model was the horizontal line of pixels halfway up the stimulus. The width of the image matrix was the same size as the spatial envelope. This was justified by simulation 3.1, described

below. Simulation 3.2 confirmed that the 1D model gave the same blur estimates as the 2D model for the test edges. The comparison stimulus always had the same image matrix size as the corresponding test stimulus. This was justified because the blur-matched comparison stimuli would never have a larger scale than the corresponding test stimuli, and so an image matrix that was large enough for the test stimuli would be large enough for the comparison stimuli.

For each test stimulus, the binary search method described in section 8.2 was used to find the comparison stimulus that matched in estimated blur. The model was run with threshold values of 0, 0.05, 0.10, 0.15, 0.20 and 0.25. Other parameters are given in Table 3.1.

Table 9.1. Width of the spatial envelope used in simulation 3, in arc min and pixels.

Edge scale (arc min)	Envelope width (arc min)	Envelope width (pixels)
6, 8	80	320
12, 16	160	640
24, 32	320	1280

9.1.1.1 Simulation 3.1: effect of size of image matrix

One issue was whether the border of the spatial envelope of a stimulus should coincide with the border of the image matrix, or whether there should be a margin of zero values surrounding the envelope. Allowing the border of the envelope to coincide with the border of the image matrix would make no difference to the first convolution stage of the model, as the convolution function assumes that the area outside the image has a zero value. But, in the subsequent stages, this area would effectively continue to have a value of zero, rather than the correct non-zero value that should occur because of the blurring introduced by the first stage. To address this issue, the test stimuli were modelled twice with the 1D model: once with an image matrix the same width as the envelope, and once with the image matrix twice as wide as the envelope. The model was run with no threshold, and with the parameters given in Table 3.1. Rather than run each test edge scale at each contrast, the contrast of all the edges was set to 1. The choice of contrast

was arbitrary because, with no threshold, the contrast makes no difference at all to the perceived blur of the edge, and the perceived contrast is a constant multiple of the stimulus contrast. Using a contrast of 1 is preferable because it makes explicit the proportional error in estimated contrast.

For all the test stimuli, the zero-value margin around the stimulus made no difference to the blur channel selected by the model, and the largest difference in perceived contrast was less than 1×10^{-8}, so the main simulation was run with an image matrix the same width as the spatial window. It was felt to be unnecessary to repeat this preliminary simulation with an above-zero threshold, because the threshold would have the effect of setting to zero most of the values within the margin, making the model more likely to give the same result for the two different sizes of image matrix. Thus, the zero-threshold condition was the most likely to reveal any effect of image matrix size.

Since the model was being run with no threshold, it was a faithful implementation of the model described in Chapter 2, and should therefore estimate the edge blur and contrast veridically. The estimated contrasts and blurs of the stimuli are given in Table 9.2. The estimated contrasts were all very close to the actual contrast of 1 used in the simulation. For all edges except the 12′ edge, the selected blur channel was that with scale closest to the edge scale. With the 12′ edge, the two closest channels were almost exactly the same distance from the correct value, and the channel that was slightly further away was selected.

Table 9.2. Estimated blur and contrast of the 0.2 contrast test edges in simulation 3.1. All values rounded to 3 significant figures. The border of zero values made no difference to the perceived blur, and made a difference of less than 1×10^{-8} to the perceived contrast of each stimulus.

Edge scale (arc min)	Perceived contrast using 1D model	Perceived blur using 1D model (arc min)	Channel virtual template scale closest to edge scale.
6	0.998	5.99	5.99
8	0.994	7.97	7.97
12	1.014	12.1	11.9
16	1.011	16.1	16.1
24	1.002	24.0	24.0
32	0.999	32.0	32.0

9.1.1.2 Simulation 3.2: comparison of 1D and 2D models

The 1D and 2D models described in Chapter 2 were set up so that they gave the same results as each other for ideal Gaussian edges. The stimuli in experiment 1 were not ideal because they were windowed. The 2D stimulus had a circular window; to model this in 1D, the profile of the stimulus at its widest point was used. It was possible that the effect of the 2D circular window was different from the corresponding 1D window, so the test stimuli of experiment 1.1 were run through the 2D model with threshold parameters of 0 and 0.05, to check that it gave the same results as the 1D model. The model and stimulus parameters are given in Table 9.3. Note that, although the visual angle subtended by the spatial envelope varied with edge scale (as detailed in Table 5.1), the image resolution was adjusted so that the envelope diameter corresponded to a constant number of pixels for each stimulus.

Table 9.3. The model and stimulus parameters used in simulation 3.2. See Table 3.1 for an explanation of the rightmost two columns.

Edge scale (arc min)	Image resolution (pixels/deg)	Envelope diameter (pixels)	N, the number of channels	$\sigma_t(1)$ (pixels)	$\sigma_t(N)/\sigma_t(1)$
6, 8	120	160	86	4	5
12, 16	60	160	86	4	5
24, 32	30	160	86	4	5

The results are shown in Table 9.4 and Table 9.5. These tables compare the results of simulation 3.2 with the test edge data from the 1D model used in the main part of simulation 3. It can be seen that the 1D model gave almost identical results to the 2D model, justifying the use of the 1D model in the full simulation of experiment 1.1.

Table 9.4. Comparison between 1D and 2D models with zero threshold. The shaded areas give the models' estimated blur for each of the test edges.

			Test edge contrast			
			0.05	0.1	0.2	0.4
Test edge scale	32′	1D	32.0	32.0	32.0	32.0
		2D	31.9	31.9	31.9	31.9
	24′	1D	24.0	24.0	24.0	24.0
		2D	24.0	24.0	24.0	24.0
	16′	1D	16.1	16.1	16.1	16.1
		2D	16.0	16.0	16.0	16.0
	12′	1D	12.1	12.1	12.1	12.1
		2D	12.0	12.0	12.0	12.0
	8′	1D	7.97	7.97	7.97	7.97
		2D	7.98	7.98	7.98	7.98
	6′	1D	5.99	5.99	5.99	5.99
		2D	6.01	6.01	6.01	6.01

Table 9.5. Comparison between 1D and 2D models with threshold of 0.05. The shaded areas give the models' estimated blur for each of the test edges.

			Test edge contrast			
			0.05	0.1	0.2	0.4
Test edge scale	32′	1D	15.8	24.5	28.5	30.8
		2D	15.9	24.5	28.5	30.7
	24′	1D	15.2	19.9	22.3	23.1
		2D	15.3	19.9	22.3	23.1
	16′	1D	12.1	14.4	15.2	15.8
		2D	12.2	14.3	15.4	15.7
	12′	1D	10.0	11.0	11.7	11.9
		2D	9.95	11.1	11.6	11.8
	8′	1D	7.11	7.67	7.82	7.97
		2D	7.13	7.69	7.83	7.98
	6′	1D	5.55	5.88	5.99	5.99
		2D	5.57	5.79	5.90	6.01

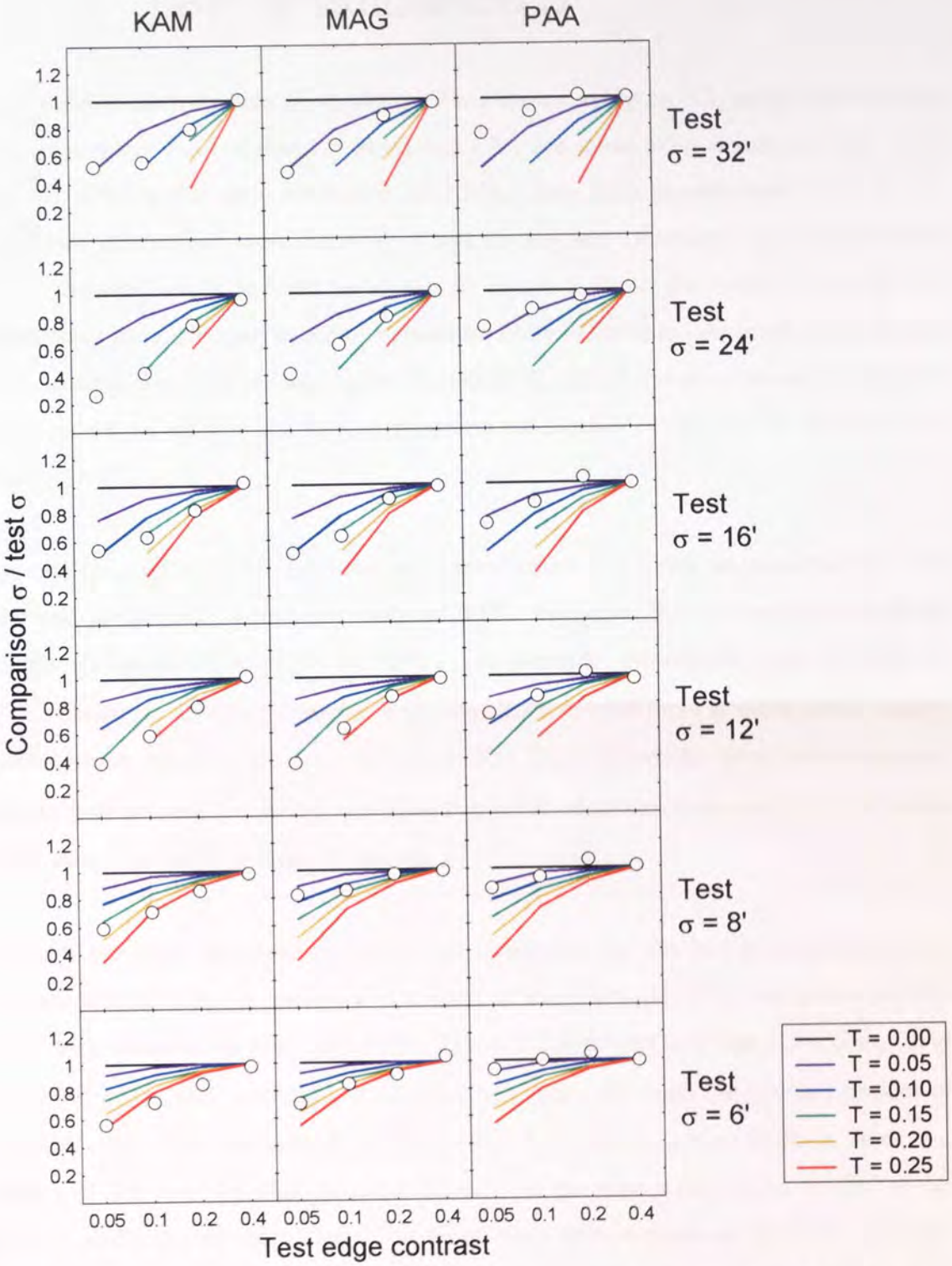


Figure 9.1. Results of the main part of simulation 3, which simulated experiment 1 with a fixed-threshold rectifier. Each row of panels shows data for a particular test edge scale. Each column of panels shows data for a particular subject. The circles show the subject's data, collapsed across stimulus order. The data for MAG and PAA are the same as those shown in the right hand column in Figure 5.2. The data for KAM are the geometric mean of KAM's data from experiments 1.1 and 1.2 (both shown in the right hand column of Figure 5.4). Each subject's data are plotted against the same set of coloured lines, which show the model's predictions with different threshold values, *T*.

9.1.2 Results and discussion of the main part of simulation 3

The results of the main part of simulation 3 are shown in Figure 9.1, along with the data from experiment 1. The data for MAG and PAA are those from experiment 1.1. The data for KAM is the geometric mean of KAM's data from experiments 1.1 and 1.2. These two experiments were identical, except for the noise surround, and KAM showed almost identical results in both (as shown in Figure 5.4), so the results from the two experiments were averaged in order to minimize noise in the data. As predicted in section 7.2.2.1, when the threshold was higher than 0.0748, not all the stimuli could be detected by the model, so some of the lines representing the model's predictions do not extend to all the stimuli.

The data from subject PAA, who showed a small effect of contrast on perceived blur, are quite well predicted by a threshold value of 0.05. However, there is no single threshold that fits all the data from KAM or MAG. For example, a threshold value of 0.05 fits MAG's data quite well for the largest scale edge (32′), but the three smallest scales require a much higher threshold of about 0.15 or 0.20. These threshold values are implausibly high, as they prevent the model from detecting a 32′ edge with a contrast 0.1, a stimulus that is easily detected by human observers.

Although the single threshold model is able to account for the fact that perceived blur increases with contrast, it predicts that the size of the effect of contrast on perceived blur should vary considerably with edge scale. Figure 9.2 illustrates why this is the case, using 6′ and 32′ edges, with a contrast of 0.05, as examples. Although the gradient profiles of Gaussian edges have theoretically infinite width, the gradient is very close to zero at a distance of 2σ from the steepest point, where σ is the edge scale, so the "width" of the gradient profile can be considered to be about 4σ. With a threshold of 0.05, the width of the 32′ edge gradient profile is substantially reduced, causing a substantial amount of perceptual sharpening, whereas the width of the 6′ profile is hardly changed. At a contrast of 0.4, the gradient profiles would be 8 times higher, and neither profile would be substantially truncated by the threshold. Thus, the perceived blur varies with contrast

much more with the 32′ edge than the 6′ edge. The 6′ edge profile is substantially reduced by the higher threshold of 0.25, but the peak gradient of the 32′ edge falls well below this threshold, and the edge is not detected by the model.

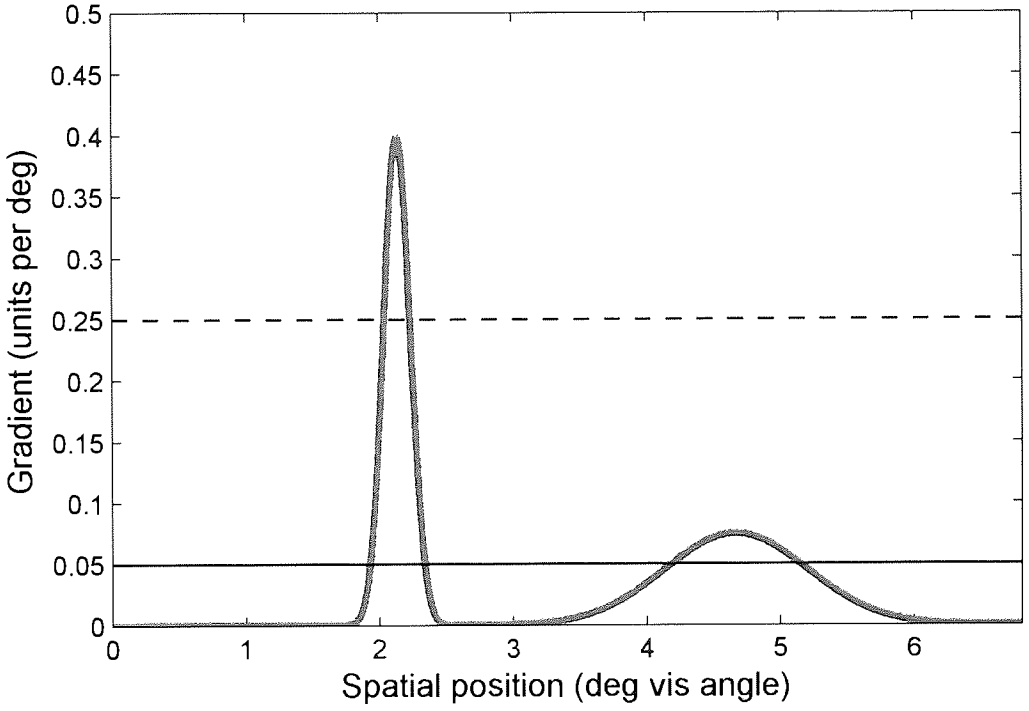


Figure 9.2. The thick grey line shows the gradient profiles of a 6′ edge on the left and 32′ edge on the right. Both edges have a contrast of 0.05. A threshold of 0.05 (solid black line) substantially reduces the width of the 32′ edge's gradient profile, but only makes a small difference to the width of the 6′ profile. A threshold of 0.25 (dashed black line) substantially reduces the width of the 6′ profile, but the gradient of the 32′ edge falls well below this threshold.

9.2 Simulation 4: varying the threshold with channel scale

As explained in section 9.1.2, the single threshold model needs a low threshold to explain the effect of contrast on large-scale edges, and a high threshold to explain the effect of contrast on small-scale edges. This suggested that one promising approach might be to make the threshold decrease with increasing channel scale. Two options were considered.

1. Let the threshold be inversely proportional to the channel scale: $T(n) = k/\sigma_v(n)$, where k is a constant, and $T(n)$ and $\sigma_v(n)$ are, respectively, the threshold and virtual template scale in channel n.
2. Let the threshold be inversely proportional to the square of the channel scale: $T(n) = k/\sigma_v^2(n)$.

The second approach was based on the idea that the purpose of the threshold is to reduce noise. Elder & Zucker (1998) derived a formula for the minimum reliable response of a Gaussian gradient operator, i.e. the lowest response that could be statistically distinguished from zero, given a known sensor noise level. They showed that the minimum reliable response was inversely proportional to the square of the operator scale. Thus, setting a threshold that is inversely proportional to the square of the gradient operator scale will remove all responses that are statistically unreliable (at a particular level of statistical significance), while maintaining reliable responses. Since, in the GDT model, the virtual template scale in each channel is proportional to the gradient operator scale, if the threshold is proportional to the square of the virtual template scale, it is also proportional to the square of the gradient operator scale.

Rather than simulating experiment 1, the effect of contrast on perceived blur was investigated for a single edge, with the thresholds varied according to the two schemes described above.

9.2.1 Methods and stimuli

The stimuli were unwindowed vertical Gaussian edges with scale 16′, and contrasts of 0.05, 0.1, 0.2 and 0.4. The image matrix was 5×5 deg, or 1200×1200 pixels. The 1D model was run with the parameters given in Table 3.1. Thresholds were varied according to the two schemes described in the previous section. In both cases, the constant of proportionality, k, took values of 0, 0.02, 0.04, and 0.08.

9.2.2 Results

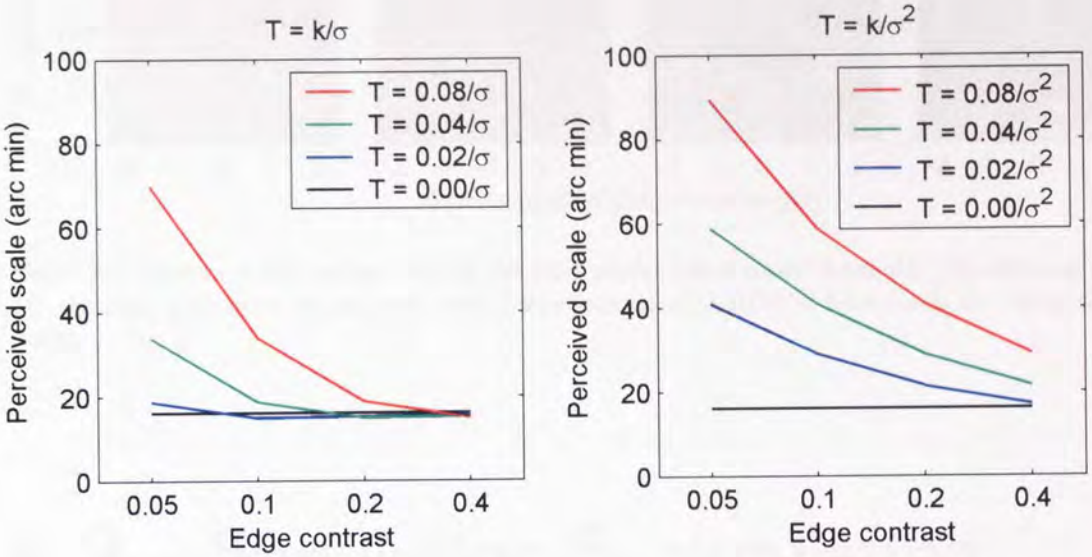


Figure 9.3. Results of simulation 4. The graph on the left shows the effect of contrast on perceived blur of a 16′ edge when the threshold, *T*, is inversely proportional to the virtual template scale, σ. The graph on the right shows the effect when the threshold is inversely proportional to σ^2.

The results are shown in Figure 9.3. Unlike the fixed threshold model, the scaled threshold model perceives blur to *increase* with decreasing contrast. This phenomenon can be understood by looking at Figure 9.4, which shows the output of the rectifier stage for one set of model parameters in this simulation. With decreasing contrast, the central peak does get narrower within each blur channel, but the thresholds are so much higher in the smaller blur channels that, when the contrast is decreased, the signal falls completely below threshold in the smaller blur channels, and is only above threshold in the large channels. Thus, although the signal is quite narrow, the peak in the edge strength representation occurs in the large blur channels, even though the signature does not fit these large templates very well.

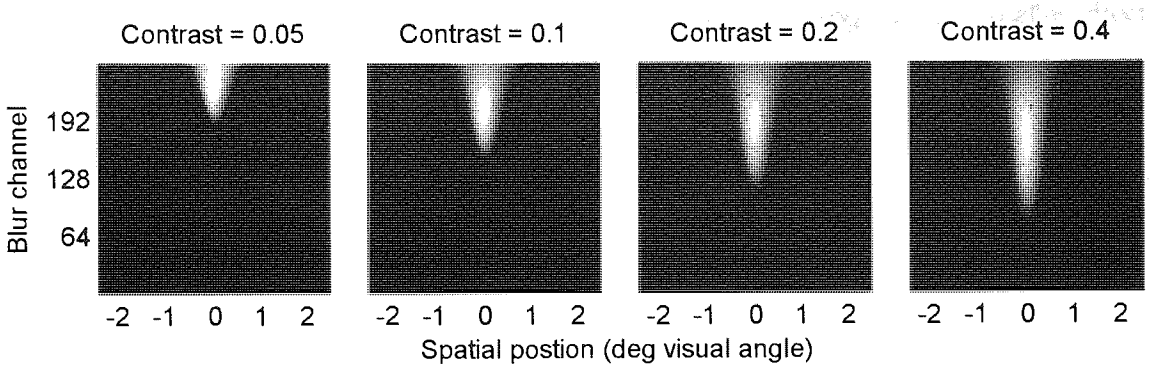


Figure 9.4. Output of the rectifier stage of the GDT model with a scaled threshold. The stimulus was the 16′ Gaussian edge from simulation 4, with a threshold equal to $0.08/\sigma$, where σ is the virtual template scale.

9.3 Simulation 5: simulating experiment 1 with the power-function rectifier

In order to explain the effect of contrast on the 6′ edge, the single threshold model required a threshold so high that the 32′ edge was undetectable. The power-function rectifier seemed promising because it can have a shape which is very similar to a high-threshold rectifier, without falling to zero anywhere.

9.3.1 Methods and stimuli

The methods and stimuli were identical to those of simulation 3, except that, instead of a threshold-linear function, the model used the power-function rectifier. The model was run with 384 different combinations of parameters. T_1 took 4 different values, ranging from 0 to 0.06 in steps of 0.02; T_2 took 8 different values, ranging from 0.1 to 0.8 in steps of 0.1; n took 12 different values, which all had the form 2^p, where p varied from 0.25 to 3 in steps of 0.25. The reason for the geometric progression in the sequence of exponent

values was that, as the exponent increases, a given step has a progressively smaller effect on the shape of the power-function rectifier.

The model was fitted to the data from experiment 1.1 for MAG and PAA, and to the geometric mean of the σ-ratios from experiments 1.1 and 1.2 for KAM, using the fitting procedure described in section 8.4.

9.3.2 Results

The best-fitting parameters are shown in section 13.3.1. Figure 9.5 shows the data from experiment 1, along with the predicted results using the best-fitting power-function rectifier for each subject. It can be seen that the model accounts for the data very well, both with the parameters chosen to fit best to these data, and also with the parameters that fit best to all the experiments.

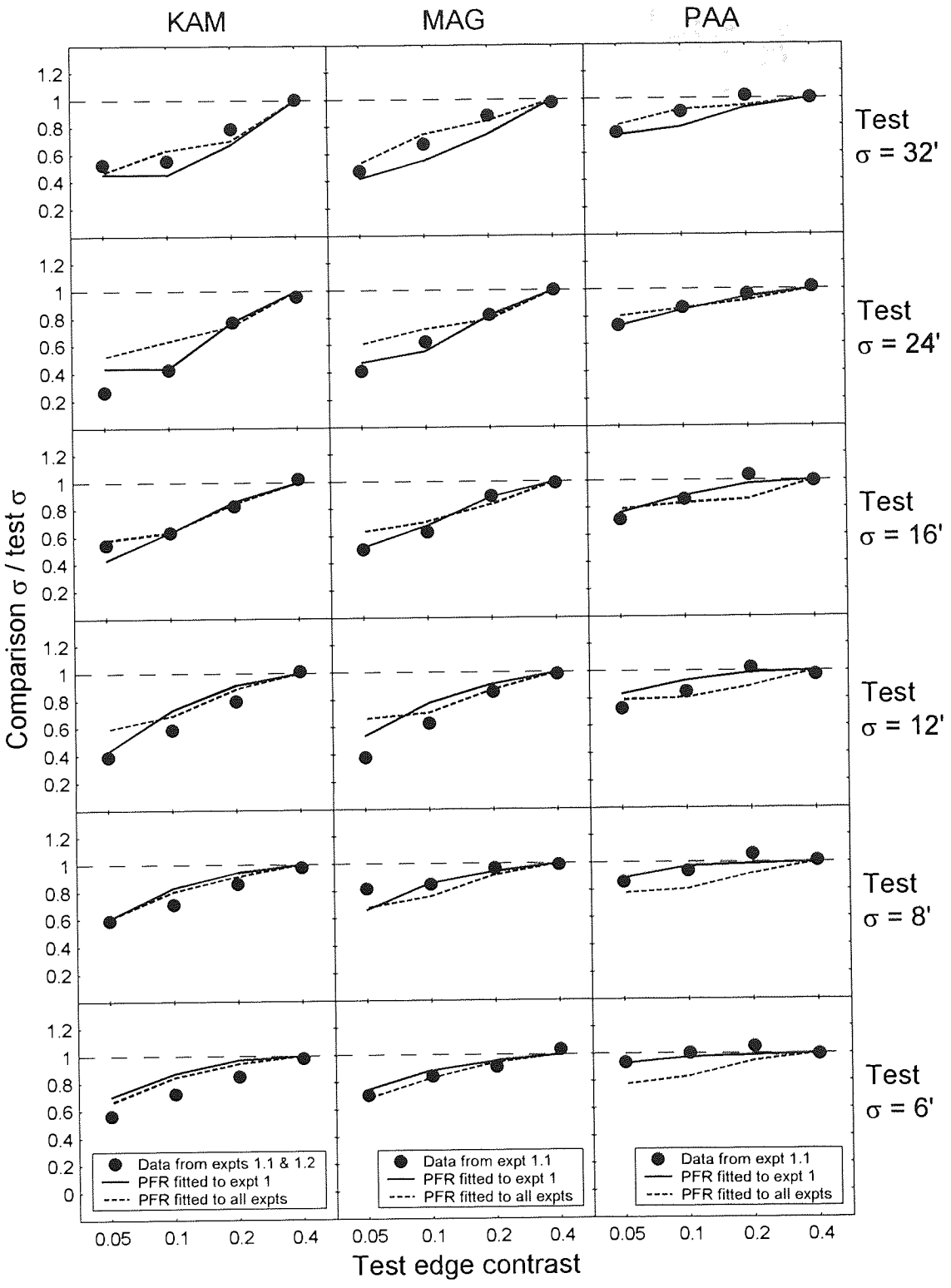


Figure 9.5. Results of simulation 5, which simulated experiment 1 with a power-function rectifier (PFR). The solid line in each graph shows the results of fitting the PFR to the data from experiment 1: for each subject, one set of parameters was fitted to the whole data set. The dashed line in each graph shows the results of fitting the PFR to each subject's data from all the experiments.

9.4 Simulation 6: simulating experiment 1 with a Legge-Foley transducer

9.4.1 Methods and stimuli

Experiment 1 was simulated following exactly the same procedure as simulation 5, except that, instead of the power-function rectifier, the Legge-Foley transducer, described by equation (7.12), was used. The S parameter could take any value from the set {0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10, 0.12, 0.14, 0.16, 0.18, 0.2, 0.24, 0.28, 0.32, 0.36, 0.40, 0.48, 0.56, 0.64, 0.72, 0.80, 0.96}, and the p parameter could take any value from the set {1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.2, 2.4, 2.6, 2.8, 3.0, 3.2, 3.4, 3.6, 3.8, 4.0, 4.4, 4.8, 5.2, 5.6, 6.0}. Thus, there were 650 different pairs of parameters. The experiment was simulated with each pair. In each simulation, $q = p - 1$.

9.4.2 Results

The best-fitting parameters are shown in section 13.3.2. Figure 9.6 shows the data from experiment 1, along with the predicted results using the best-fitting Legge-Foley transducer for each subject; also shown are the predictions of the best-fitting power-function rectifier from simulation 5 (Figure 9.5). The predictions from the two types of model are very similar, both with the parameters fitted to experiment 1, and the parameters fitted to all the experiments. Both sets of parameters for both models fit the data well.

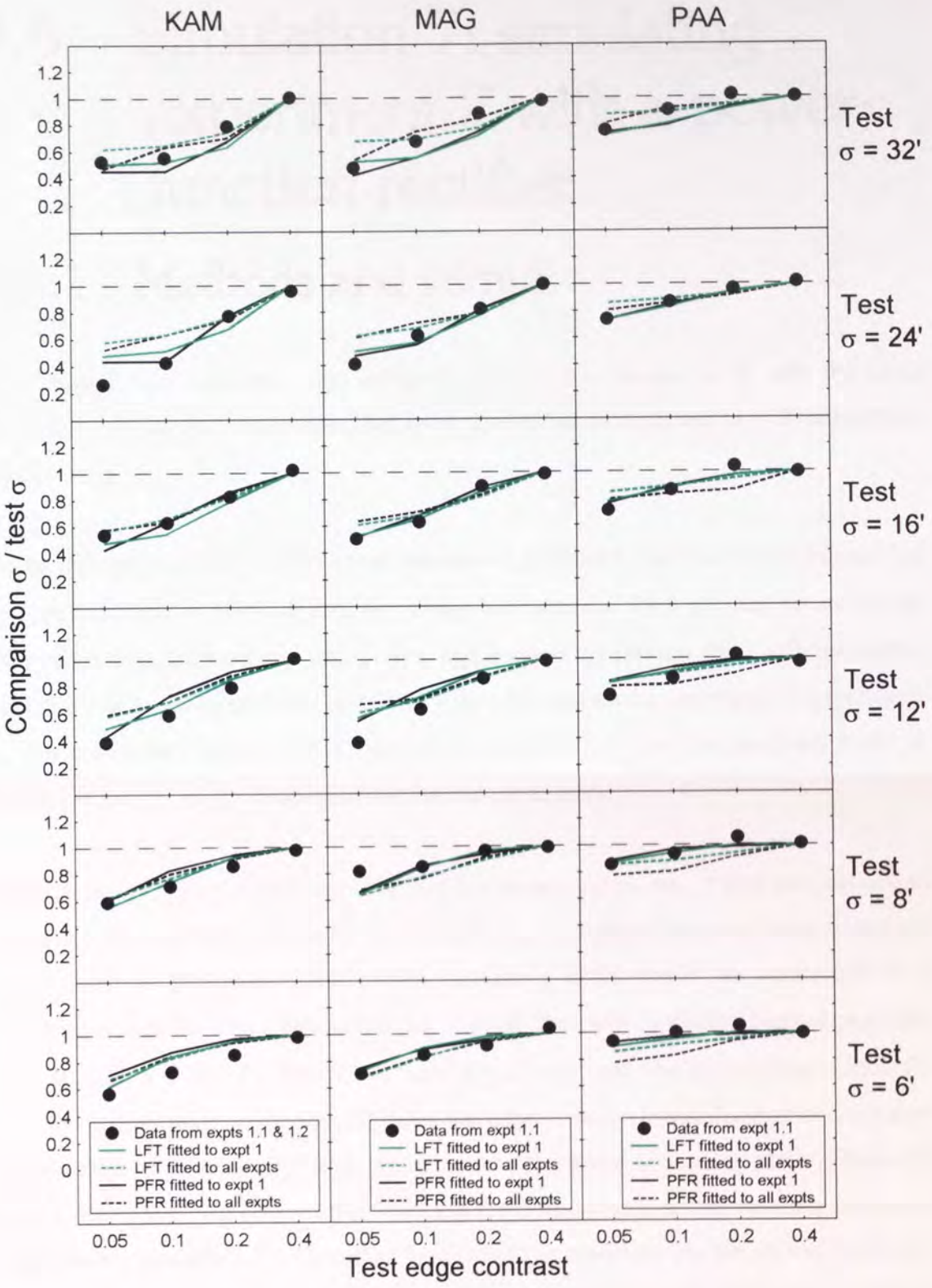


Figure 9.6. Results of simulation 6, which simulated experiment 1 with a Legge-Foley transducer (LFT). The predictions of the best-fitting Legge-Foley transducer are shown in green. These are compared with the predictions of the power-function rectifier (PFR), which are shown in black, as in Figure 9.5. Solid lines show the best-fitting parameters for experiment 1, and dashed lines show the best-fitting parameters for all the experiments.

9.5 Simulation 7: simulating experiment 2 with a power-function rectifier

9.5.1 Methods and stimuli

Experiment 2 was simulated using an identical method to simulation 5, with the same power-function rectifier parameters, but using the stimuli from experiment 2, rather than experiment 1.

The model was not fitted to PAA's data because all parameter combinations predicted that the σ-ratio would fall below 1 with decreasing contrast, and PAA showed no substantial effect of contrast for most test edge scales, and showed the reverse effect with the largest scale; this result was inconsistent with PAA's data for very similar conditions in experiment 1. The model was fitted to MAG's data from experiment 2, and the geometric mean of KAM's σ-ratios for all three versions that he had performed.

Although the stimuli in experiment 2 all had the same window size, it was unnecessary to keep the image matrix the same for all stimuli in the simulation, because the smaller-scale edges would be processed using smaller operators, which would be unaffected by a reduction in stimulus size. Simulation 3.1 showed that, with Gaussian edge stimuli with scales 8′, 16′, and 32′, the model estimated the contrast and blur of the edges veridically when the image matrix width was 10 times the edge scale, so it was decided to use image matrix widths of 10 times the edge scale, in order to reduce processing time. These are given in Table 9.6. To confirm that reducing the image matrix width made no difference to the results, simulation 7.1 (described below) ran the model on the test stimuli with both the full and reduced image sizes. The comparison stimulus always had the same image matrix size as the corresponding test stimulus.

Table 9.6. Widths of image matrices used in simulation 7, in arc min and pixels. All stimuli the experiment 2 had window widths of 320', but smaller stimuli were used in the simulation to reduce processing time.

Edge scale (arc min)	Image matrix width (arc min)	Image matrix width (pixels)
2	20	80
4	40	160
8	80	320
16	160	640
32	320	1280

9.5.1.1 Simulation 7.1: effect of size of image matrix

The stimuli from experiment 2 were constructed using the 320′ window used in the experiment, and also with an image matrix set to 10 times the edge scale. The model was run through the model with an unbiased half-wave rectifier. As argued in Section 9.1.1.1, this setting would be the most likely to reveal any effect of stimulus size. All other parameters were as given in Table 3.1. The contrast of the edges was set to 1, as in simulation 3.1.

For each stimulus, the estimated blur was identical, whether the window size was 320′ or 10 times the edge scale, and the estimated contrasts differed by less than 1×10^{-6}, confirming that the reduced stimuli could be used in the main simulation. The estimated values are shown in Table 9.7.

Table 9.7. Results of simulation 7.1.

Edge scale (min visual angle)	Estimated contrast using 1D model	Estimated blur using 1D model (arc min)	Channel virtual template scale closest to edge scale.
2	0.988	1.99	1.99
4	1.006	4.02	4.02
8	0.994	7.97	7.97
16	1.011	16.1	16.1
32	0.999	32.0	32.0

9.5.2 Results of the main part of simulation 7

The best-fitting parameters are shown in section 13.3.1. Figure 9.7 shows the data from experiment 2, along with the predicted results using the best-fitting power-function rectifier for each subject. Unlike in the simulations of experiment 1, the parameters that fitted best to the data from experiment 2 gave rise to predictions that fitted the data substantially better than the parameters fitted to all the experiments. Both sets of parameters predicted the data well for the more blurred edges, but with the sharper edges, the model required a value of T_2 that was substantially higher than the value that fitted best to all the experiments. Even with this high value, the model was unable to predict an effect quite as large as that shown in the data with the 2′ edge. The subjects behaved as if the 2′ edge had a higher blur. This suggested that the data might be explained by introducing some intrinsic blur into the model. This was investigated in simulation 8.

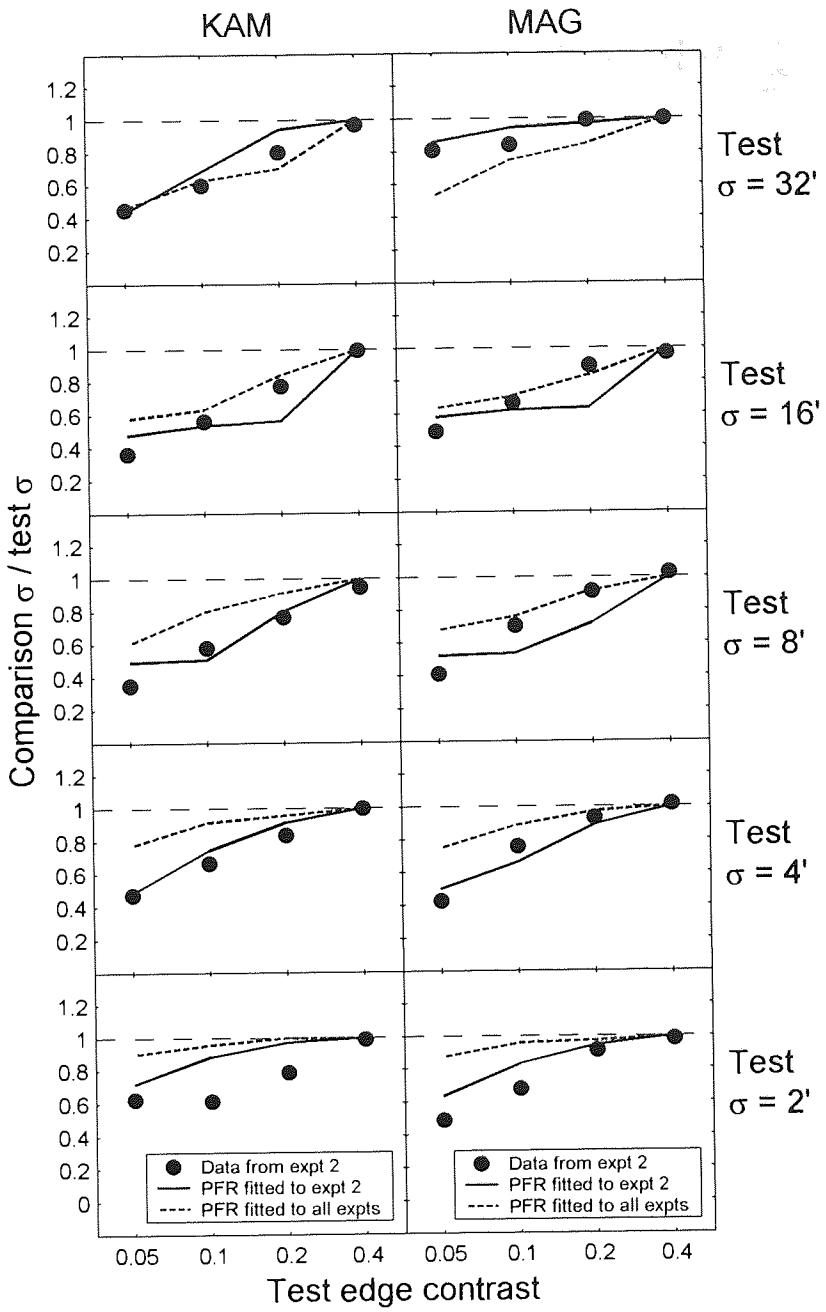


Figure 9.7. Results of simulation 7, which simulated experiment 2 with a power-function rectifier (PFR). The solid line shows the predictions from the parameters that fit best to the data from experiment 2, and the dashed line shows the predictions from the parameters fitted to all the experiments. The circles indicate the experimental data. In the case of KAM, these data are the geometric mean of all three versions of this experiment that KAM performed.

9.6 Simulation 8: simulating experiment 2 with intrinsic blur

It has been estimated that the human visual system contains intrinsic blur that is equivalent to blurring with a Gaussian kernel with a scale of about 1′ (Levi & Klein, 1990; Wuerger *et al.*, 2001). Simulation 8 repeated simulation 7 with intrinsic blur in an attempt to improve the fit of the model to the data for the 2′ edge in experiment 2.

9.6.1 Methods and stimuli

The methods were identical to simulation 7, in which experiment 2 was simulated with the power-function rectifier. The stimuli were identical except that they were blurred by convolving with a Gaussian kernel with a scale of 1′. Because the scale of each edge was made larger by the intrinsic blur, simulation 8 used slightly larger window sizes than simulation 7, and these are shown in Table 9.8. These values were 10 times the scale of the Gaussian edges after blurring, rounded to the nearest even integer. Simulation 8.1 confirmed that these image matrix sizes were sufficient. Again, the comparison stimulus always had the same image matrix size as the corresponding test stimulus.

Table 9.8. Width of stimuli used in simulation 8, in arc min and pixels.

Edge scale (arc min)	Stimulus width (arc min)	Stimulus width (pixels)
2	20	90
4	40	164
8	80	322
16	160	642
32	320	1280

9.6.1.1 Simulation 8.1: effect of size of image matrix

The method was identical simulation 7.1, except that the blurred stimuli were used, and the simulation was run with image matrix sizes given in Table 9.8, as well as the image matrix sizes used in simulation 7.1. The image matrix sizes given in Table 9.8 gave rise to the same blur estimates as a window size of 1280 pixels for all stimuli, and these are given in Table 9.9. The image matrix sizes used in simulation 7 (shown in Table 9.6) gave rise to the same blur estimates as the 1280 pixel image matrix, except for the 2′ edge, which gave rise to a blur estimate of 2.27′, slightly higher than the value of 2.23 estimated using the larger image matrices. This confirmed that it was necessary to use slightly larger image matrices in simulation 8 than had been used in simulation 7. The estimated contrast for the image matrix sizes given in Table 9.8 differed from the estimated contrast using the 1280 pixel image matrix by less than 1×10^{-3}.

Table 9.9. Results of simulation 8.1, using image matrices with size 1280 pixels, or the sizes given in Table 9.8 (the results were the same to 3 significant figures).

Edge scale (min visual angle)	Estimated contrast using 1D model	Estimated blur using 1D model (arc min)	Effective edge scale after blurring
2	0.992	2.23	2.24
4	0.989	4.10	4.12
8	1.01	8.12	8.06
16	1.01	16.1	16.0
32	0.998	32.0	32.0

9.6.2 Results of the main part of simulation 8

The best-fitting parameters are shown in section 13.3.1. Figure 9.8 shows the data from experiment 2, along with the predicted results using the best-fitting power-function rectifier for each subject. For comparison, Figure 9.8 also shows the results of simulation 7, from Figure 9.7. The intrinsic blur marginally slightly improved the fit of the model to the data (see Table 13.1 and Table 13.2), but did not make a substantial difference; in particular,

the model could not fit KAM's data with the 2′ edge. It is not clear whether this is a failure of the model, or just noisy data.

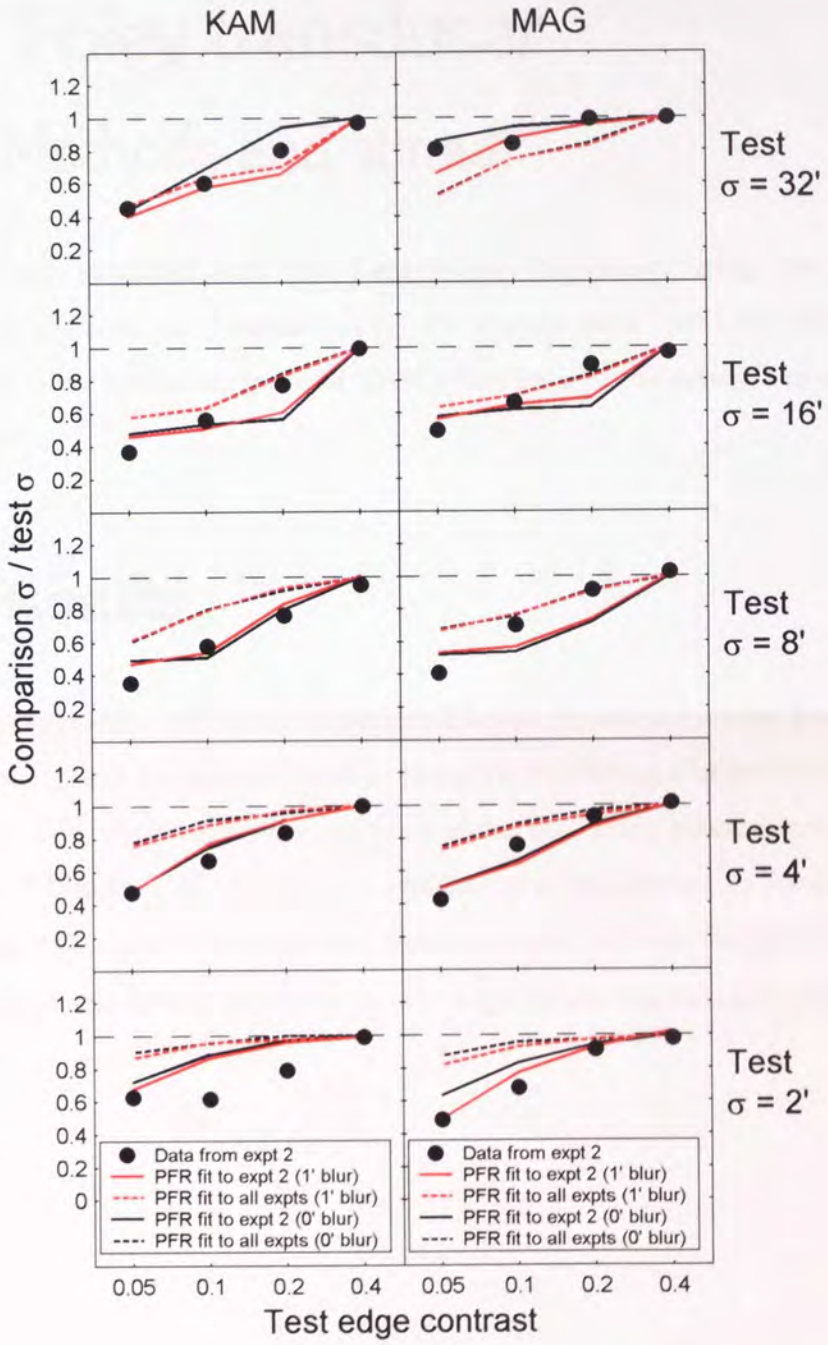


Figure 9.8. Results of simulation 8, which simulated experiment 2 using a power-function rectifier and intrinsic blur of 1′. The results of simulation 8 are shown in red, and are compared with the results of simulation 7, which are plotted in black, as in Figure 9.7. Circles show the same data as Figure 9.7.

9.7 Simulation 9: simulating experiment 2 with a Legge-Foley transducer

9.7.1 Methods and stimuli

Experiment 2 was simulated with the Legge-Foley transducer, using the stimuli of simulation 7 and the method of simulation 6. Parameters were found that fitted best to MAG's data and to the geometric mean of KAM's data for all three versions of experiment 2.

9.7.2 Results

The best-fitting parameters are shown in section 13.3.2. Figure 9.9 shows the data from experiment 2, along with the predicted results using the best-fitting Legge-Foley transducer for each subject. Also shown are the predictions of the best-fitting power-function rectifier from simulation 7 (Figure 9.7). As with the simulations of experiment 1, the Legge-Foley transducer gave very similar results to the power-function rectifier for most conditions. The only exception was KAM's data with the 32′ edge, where the power-function rectifier was superior.

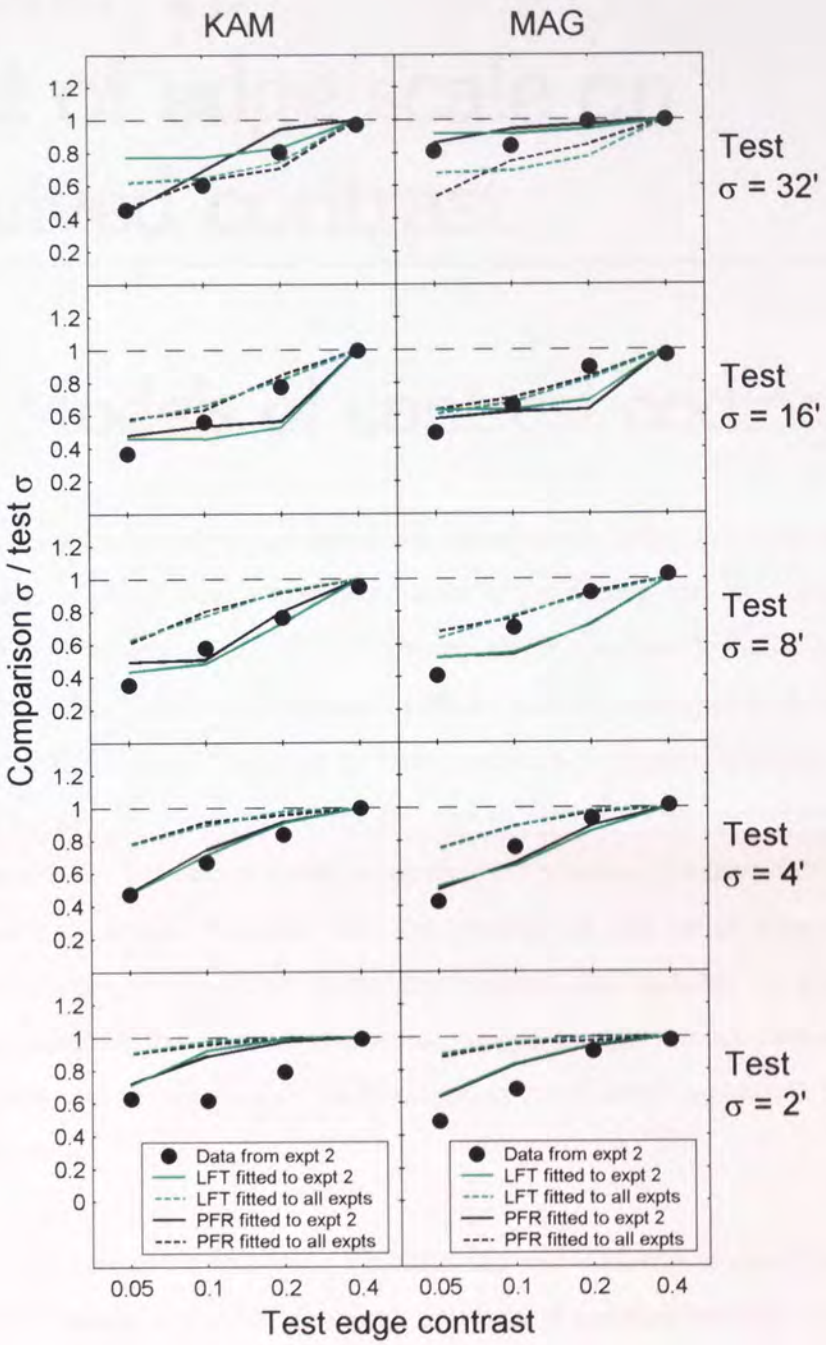


Figure 9.9. Results of simulation 9, which simulated experiment 2 with a Legge-Foley transducer (LFT). The predictions of the best-fitting Legge-Foley transducer are shown in green; predictions of the power-function rectifier (PFR) are shown in black, as in Figure 9.7. Solid lines show the best-fitting parameters to experiment 2, and dashed lines show the best-fitting parameters to all the experiments. Circles show the same data as Figure 9.7.

Chapter 10
Effect of edge scale on perceived contrast

10.1 Models of contrast coding

Most models of contrast coding are based on experiments with sine wave gratings or plaids, and take, as input, the stimulus contrast (e.g., Georgeson & Sullivan, 1975; Kulikowski, 1976; Hamerly *et al.*, 1977; Cannon, 1979; Cannon & Fullenkamp, 1988, 1991a, 1991b, 1993, 1996; Georgeson, 1991a, 1991b; Georgeson & Shackleton, 1994; Langley, 2002). Since the input to these models is a physical stimulus parameter, rather than an image, these models cannot be applied directly to the perception of other, more complex stimuli. It is inappropriate to apply the Michelson contrast definition to the whole of a complex image, because then the contrast of the whole image would be defined by the luminance of just two pixels: the brightest and darkest. In any case, it is unclear what is meant by the contrast of a whole image. Images contain features of many different contrasts, and condensing all this information into a single number is bound to be an oversimplification.

A measure of *local* contrast is required. The difficulty with defining contrast locally is that contrast is an intrinsically 2nd-order property: any kind of contrast involves a comparison between at least two things. Watson *et al.* (1983) and Peli (1990) introduced definitions of contrast that compare the luminance at each point with the mean luminance. Watson *et al.* used the function, I, in equation (2.1) as a measure of the "contrast at each point". In this definition, the contrast at each point is the difference between the luminance at that point and the mean luminance of the stimulus, expressed as a proportion of the mean luminance. Peli (1990) introduced a *local band-limited contrast* measure. This gave a contrast measure at each point in the image, for each spatial frequency band. The image

was filtered with a set of band-pass filters. For each band-pass filtered image, the contrast at each point was defined as the amplitude of that filter response, divided by the amplitude of a low-pass filtered version of the image, which contained all the energy below that band. This low-pass filtered image was intended as a measure of the local mean luminance. The problem with Watson *et al.*'s definition is that a very uniform region that differed substantially from the overall mean luminance would be assigned a high contrast. In Peli's definition, local mean luminance is used, so any region that is very uniform will be assigned a low contrast, even if it differs substantially from the overall mean luminance.

An alternative approach has been to compare the luminance values at different points within a feature. For example, Badcock (1984) defined local contrast measures on the basis of peak and trough luminance values in local features of gratings consisting of a sine wave grating with an added 3rd harmonic. Similar definitions were used by Hess & Pointer (1987) and Akutsu & Legge (1995). This kind of contrast definition does not assign a contrast value to each point; instead, it assigns contrast values to pre-defined features. This approach is limited in that it can only be applied to classes of stimuli for which features have been defined.

Badcock's feature-based contrast definition would be more useful if it could be applied to images which did not have a restricted form. The GDT model is designed to find features in any image. The model generates an estimate of the contrast of each edge element it detects, and therefore provides a promising approach to studying the perceived contrast of complex images. Having found the position, x, and blur channel, n, of each peak in the edge strength, S, the model uses equation (2.35), repeated below as equation (10.1), to calculate the contrast.

$$C = \pi^{1/4}\sigma_v^{3/2}(n)S(x;n). \qquad (10.1)$$

$\sigma_v(n)$ is the estimated scale of the edge.

The modifications to the half-wave rectifier, described in Chapter 7, have the effect of reducing the estimated contrast. This is because the estimated edge scale, $\sigma_v(n)$, is reduced by the modified rectifier; equation (10.1) shows that, when $\sigma_v(n)$ is reduced, the

estimated contrast, C, is also reduced. The modifications to the half-wave rectifier have a greater effect on low-gradient stimuli, since the modified functions are quite similar to a standard half-wave rectifier for large inputs. Thus, we might expect the perceived contrast to be disproportionately low for low-gradient stimuli: perceived contrast should decrease with increasing edge scale, particularly at low contrasts. Experiment 4 was designed to test this prediction and, in particular, to see if the GDT model could account for the results quantitatively.

10.2 Experiment 4: effect of edge scale on perceived contrast

10.2.1 Methods and stimuli

The methods and stimuli were very similar to those of the main part of experiment 2 (the version performed by both KAM and MAG), except that the task was the contrast matching procedure, described in section 4.2, rather than blur matching. There were four test edge contrasts (0.05, 0.1, 0.2, and 0.4), and four test edge scales (2′, 4′, 8′, and 16′). Each test stimulus was identical to the test stimulus with that contrast and scale in experiment 2, i.e. it was a vertical Gaussian edge with a square window with width and height 320′. The perimeter of the window was a sharp edge. Subjects were asked to judge the perceived contrast of the central edge in each stimulus, and to ignore the rest of the stimulus. The comparison edge in each trial had a scale fixed at 2′, and its contrast was adjusted with a staircase. The test and comparison edges were both vertical, and were separated by an ISI of 500 ms. The polarity of test and comparison edges was the same within a trial, and was generated randomly on each trial. The viewing distance and screen resolution were the same as for experiment 2, described in section 5.3.2. Unlike in the previous experiments, each stimulus was accompanied by a 440 Hz tone. This was because the comparison contrast was being adjusted by a staircase and it was possible that it could fall below the detection threshold, so the tone was provided to indicate when each stimulus was being presented. The amplitude of the tone had three phases: attack, sustain, and decay. In the attack phase, the amplitude increased linearly with time over

the first 25 ms; in the sustain phase, the amplitude was held constant for 250 ms; in the decay phase, the amplitude decreased linearly with time over the final 25 ms. Thus, the attack, sustain and decay phases were synchronized exactly with the temporal contrast envelope of the stimulus, described in section 4.3.

Each of the 16 combinations of levels of the independent variables (4 contrasts and 4 scales) was assigned to a different staircase, and all 16 staircases were tested within one session. Each subject completed 8 similar sessions. Starting points of the staircases were determined using the same procedure as in experiment 2, described in section 5.3.3.

10.2.2 Results

The data were analysed as described in section 4.2.3. For each test edge, the physical contrast of the comparison edge with the same perceived contrast was expressed as a proportion of the test edge contrast, giving a contrast ratio that was exactly analogous to the σ-ratio used in the blur matching experiments. These contrast ratios are plotted in Figure 10.1 as a function of test edge scale for each test contrast. As predicted, the perceived contrast decreased with increasing test edge scale, particularly when the test edge contrast was low. In order to be consistent with the graphs in Chapter 5 and Chapter 9, the contrast ratios in Figure 10.1 are re-plotted in Figure 10.2 as a function of test edge contrast for each test scale. The only data point that substantially violated the trend of decreasing contrast ratio with increasing test edge scale was PAA's data point for 16′ edge with a contrast of 0.2. This was a strange result because, when the data were separated according to the order of stimuli, this condition was well behaved. But combining the data from the two stimulus orders before fitting the psychometric function gave rise to a contrast ratio that was substantially different from that derived from the two individual data sets. Experiment 4 was simulated in simulations 10 and 11.

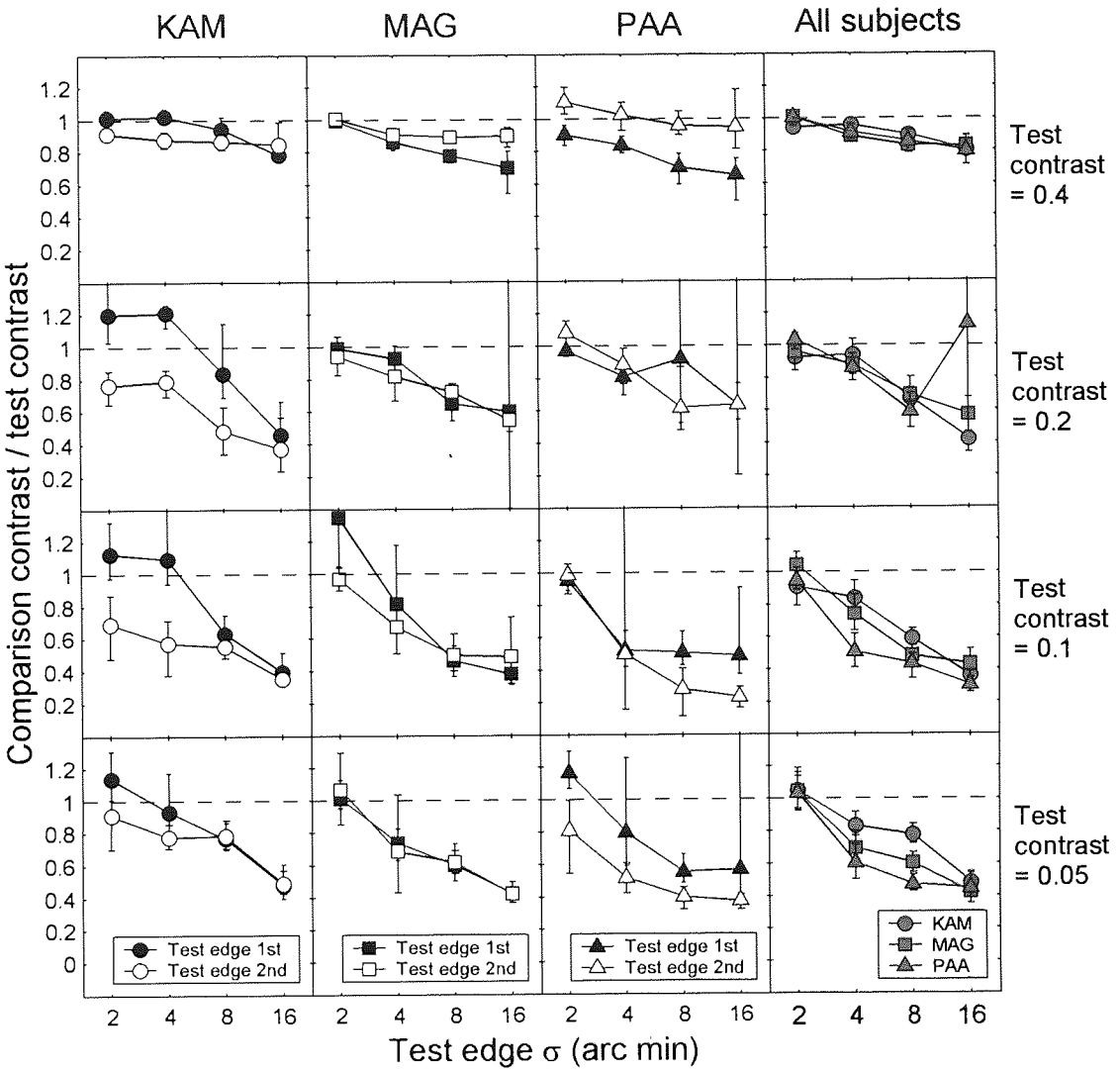


Figure 10.1. Results of experiment 4. Each row of panels shows data for a particular test edge contrast. The three leftmost columns of panels show data for the individual subjects; filled symbols show data from trials in which the test edge preceded the comparison edge, and open symbols represent trials with the opposite order. The right hand column shows data for all three subjects, collapsed across stimulus order. Within each panel, the horizontal axis represents the test edge scale, and the vertical axis represents the ratio of the comparison and test edge contrast when they perceptually matched in contrast. The dashed line represents the null hypothesis that the edges would match in contrast when their contrasts were equal.

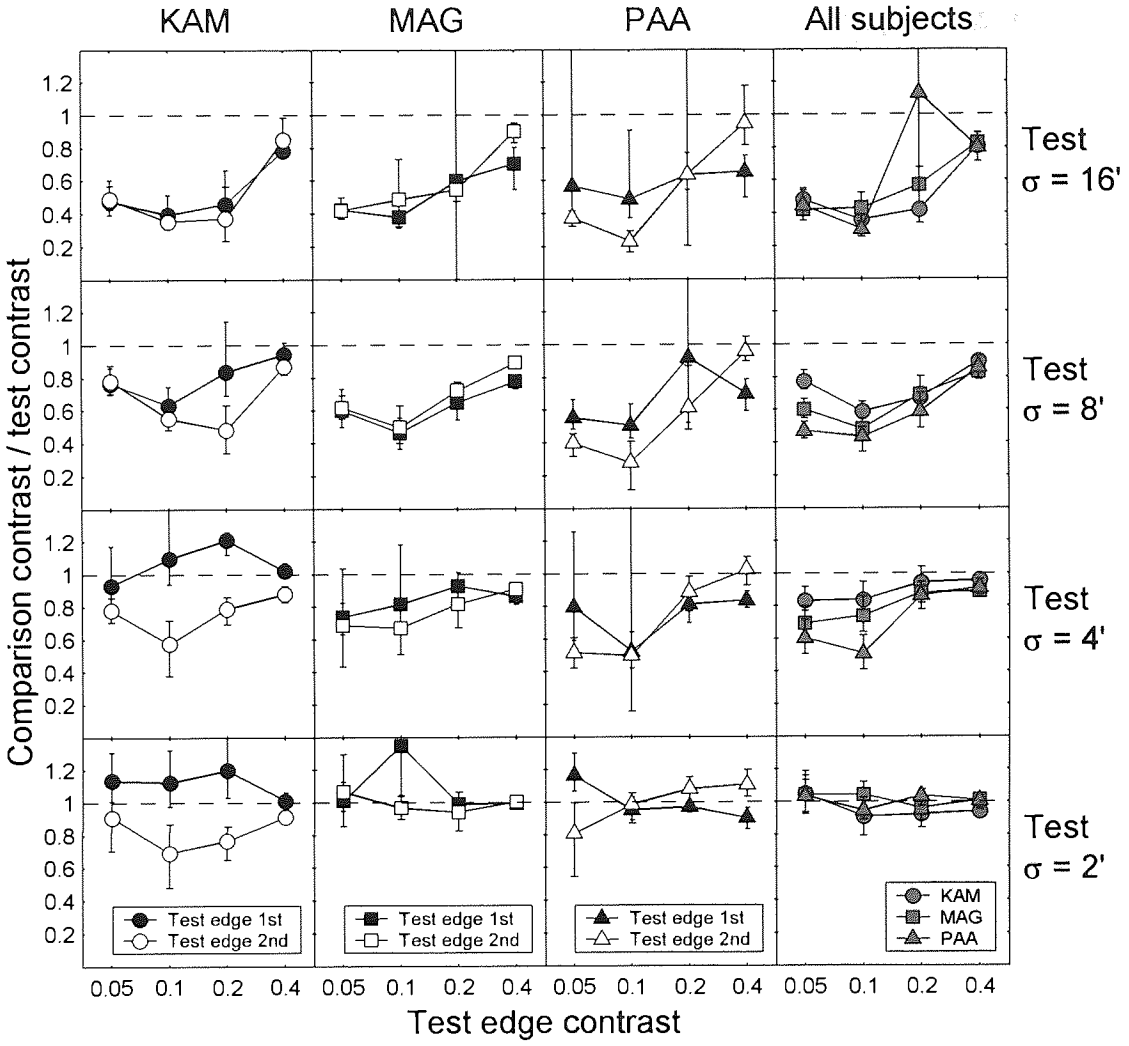


Figure 10.2. Results of experiment 4. The data are identical to those in Figure 10.1, but they are plotted differently, to be consistent with the graphs in Chapter 5 and Chapter 9. Each row of panels shows data for a particular test edge scale. Within each panel, the horizontal axis represents the test edge contrast, and the vertical axis represents the ratio of the comparison and test edge contrasts when they perceptually matched in contrast.

10.3 Simulation 10: simulating experiment 4 with a power-function rectifier

10.3.1 Methods and stimuli

Since the test stimuli in experiment 4 were a subset of the test stimuli used in experiment 2, the test stimuli used in simulation 10 were the corresponding subset of the test stimuli used in simulation 7, which stimulated experiment 2. The sizes of the image matrices are given in Table 9.6. The comparison stimulus in experiment 4 always had a scale of 2′. Simulation 7.1 had shown that, for an edge with this scale, an image matrix of 20′ (80 pixels) was sufficient, so the image matrix of the comparison stimulus was this size for every condition. Experiment 4 was simulated using the contrast-matching procedure described in section 8.3. The model was run using the power-function rectifier, with the set of parameters used in simulation 5. All other parameters were as described in Table 3.1.

10.3.2 Results

The best-fitting parameters are shown in section 13.3.1. Figure 10.3 shows the data from experiment 4, along with the predicted results using the best-fitting power-function rectifier for each subject. The model accounts for the data very well, both with the parameters chosen to fit best to these data, and also with the parameters that fit best to all the experiments.

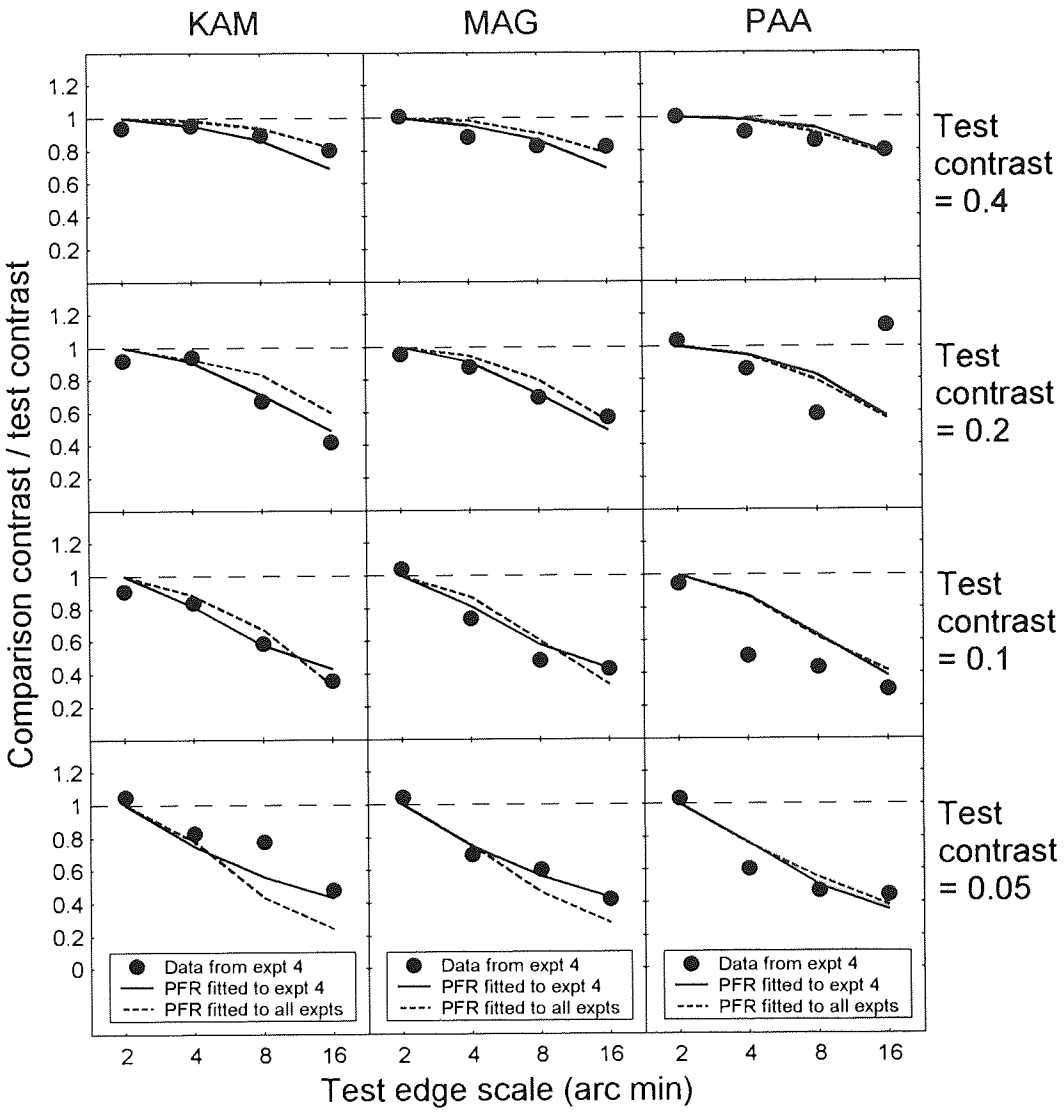


Figure 10.3. Results of simulation 10, which simulated experiment 4 with a power-function rectifier.

10.4 Simulation 11: simulating experiment 4 with a Legge-Foley transducer

10.4.1 Methods and stimuli

The methods and stimuli were identical to those of simulation 10, except that the Legge-Foley transducer was used, with the parameters used in simulation 6.

10.4.2 Results

The best-fitting parameters are shown in section 13.3.2. The predicted results are shown in Figure 10.4. The best-fitting Legge-Foley transducer made very similar predictions to the best-fitting power-function rectifier.

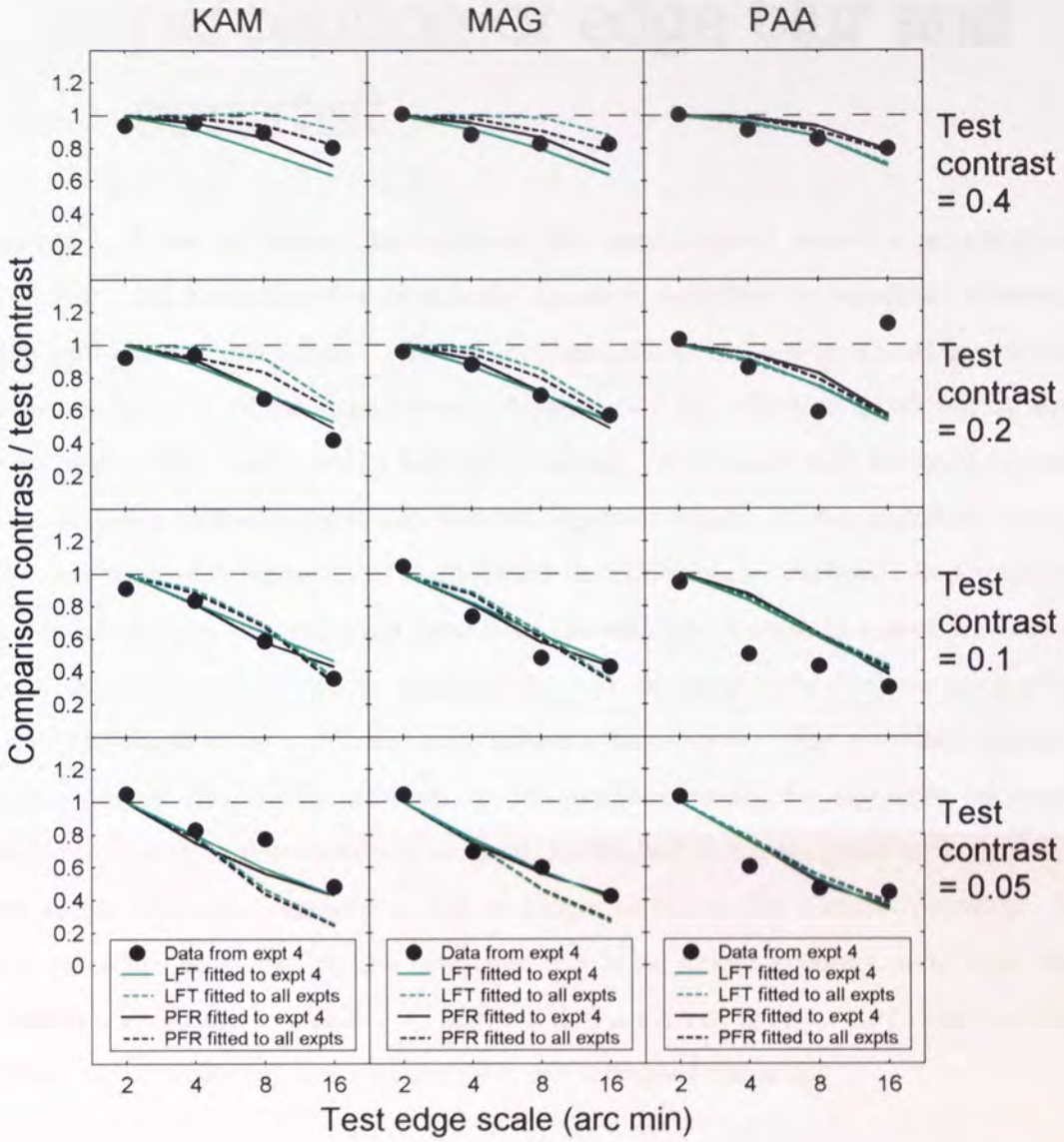


Figure 10.4. Results of simulation 11, which simulated experiment 4 with a Legge-Foley transducer. The black lines show the results predicted by the power-function rectifier in simulation 10 (Figure 10.3).

10.5 General discussion of the perception of edge blur and contrast

Experiments 1, 2 and 4 showed that reducing the edge contrast causes a reduction in perceived blur, and increasing the edge scale causes a reduction in perceived contrast, especially with low contrast edges. The sizes of these effects are well predicted by the set of parameters fitted to all the experiments. Neither of these effects is predicted by the pure form of the GDT model, with a half-wave rectifier. This model was designed to give a veridical estimate of the contrast and scale of Gaussian edges. In the modified model, the half-wave rectifier is replaced by a smoothed threshold-linear function. As shown in section 7.2.1.2, such a function could arise from the addition of noise to a pure half-wave rectifier or threshold-linear function; alternatively, it could result from the averaging of a number of threshold-linear functions with different thresholds. The modified rectifier causes errors in the model's calculations. In the modified model, the signature becomes increasingly truncated as the gradient is reduced, so the best-fitting template no longer has the scale of the edge, and equation (2.35) no longer estimates the contrast correctly. A question that arises is, given that the estimated blur is no longer accurate, why does the visual system stick rigidly to equation (2.35)? Why does it not apply a correction so that the contrast, or possibly both contrast and blur, are estimated correctly?

The ability of the model to correct for the errors depends on the nature of the function mapping physical edge contrast and scale, (C,σ), onto estimated edge properties, $(\hat{C},\hat{\sigma})$. If this function is *injective* (i.e. no two different physical Gaussian edges map onto the same estimated pair of parameters), then an inverse function can be generated, and a veridical estimate of the physical properties, (C,σ), can be obtained from the incorrectly estimated edge properties, $(\hat{C},\hat{\sigma})$. If the function is not injective, the system cannot know which pair of physical properties, (C,σ), gave rise to any given pair of estimated properties, $(\hat{C},\hat{\sigma})$.

The test edge data from simulations 5 and 7 were used to examine whether the function mapping physical parameters to estimated parameters was injective. Figure 10.5 shows isoscale contours, which give the estimated contrast and scale for an edge of a given physical scale, as it varies in contrast. Figure 10.6 shows isocontrast contours, which give the estimated contrast and scale of an edge of a given physical contrast as it varies in scale. If any contours overlap, then two physically different Gaussian edges give rise to the same pair of estimated properties, $\left(\hat{C},\hat{\sigma}\right)$, and no inverse function can be generated; if no contours overlap, then each pair of estimated properties, $\left(\hat{C},\hat{\sigma}\right)$, arises from a unique pair of physical properties, $\left(C,\sigma\right)$, and an inverse function can be generated which corrects for the errors. In both figures, the contours show no sign of overlapping. Isoscale contours from similar scales remain well separated. This suggests that the mapping from physical to estimated properties is injective. If any overlap were to occur, it would only occur for edges that were physically very similar, so an inverse function could be generated that would recover the correct physical properties of the edge with a high degree of accuracy.

It is clear that it is perfectly possible, in principle, to correct the errors made by the modified model. Each blur channel could have its own function mapping estimated contrast, $\hat{C}$, onto $\left(C,\sigma\right)$ for that channel's scale, $\hat{\sigma}$. Why the visual system does not perform such a correction is a mystery. One possibility is that the required mapping is difficult or impossible to build from the input-output functions generated by neurons. If most important edges are sufficiently high-contrast, then the errors will be small, and a correction would be an unnecessary expense.

This issue validates Marr's (1982) view that it is important to have a computational theory of what the brain is trying to achieve. Without a computational theory, we might still be able to model the effect of edge scale on perceived contrast, using conventional psychophysical models, but this would give no insight into *why* the visual system is set up to show a systematic decrease in perceived contrast with scale. The computational theory suggests a reason why these errors are made: they arise from the application of an equation (equation (2.35)) which should give the correct answer but, because of noise in an earlier stage of the model, no longer does so.

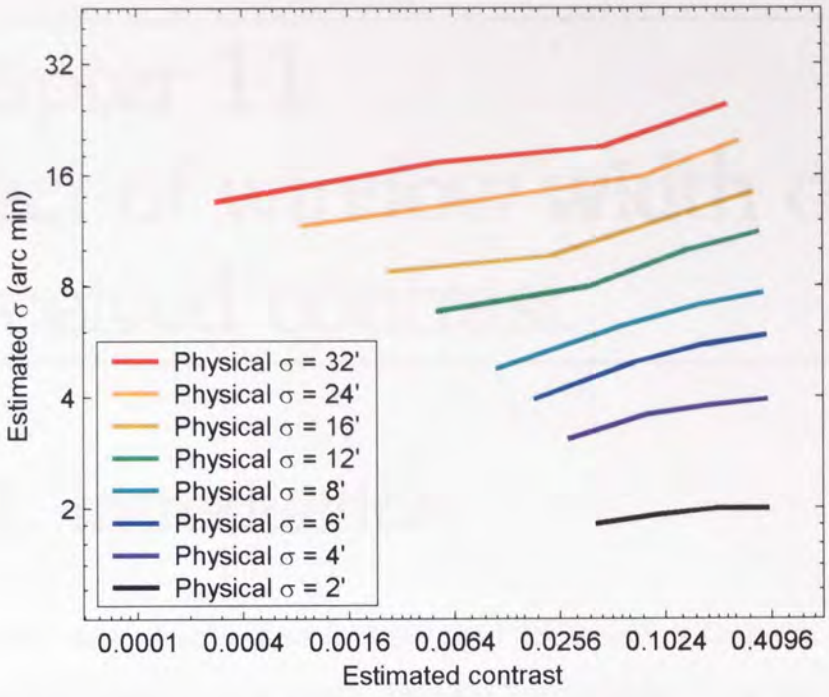


Figure 10.5. Isoscale contours. Each line gives the estimated contrast and scale for an edge of the given physical scale, as its contrast is varied from 0.05 to 0.4. The model was run with KAM's overall best-fitting parameters, $T_1 = 0.04$, $T_2 = 0.3$, $n = 2.38$. The data for the 2′ and 4′ test edges came from simulation 7, and the remaining data came from simulation 5.

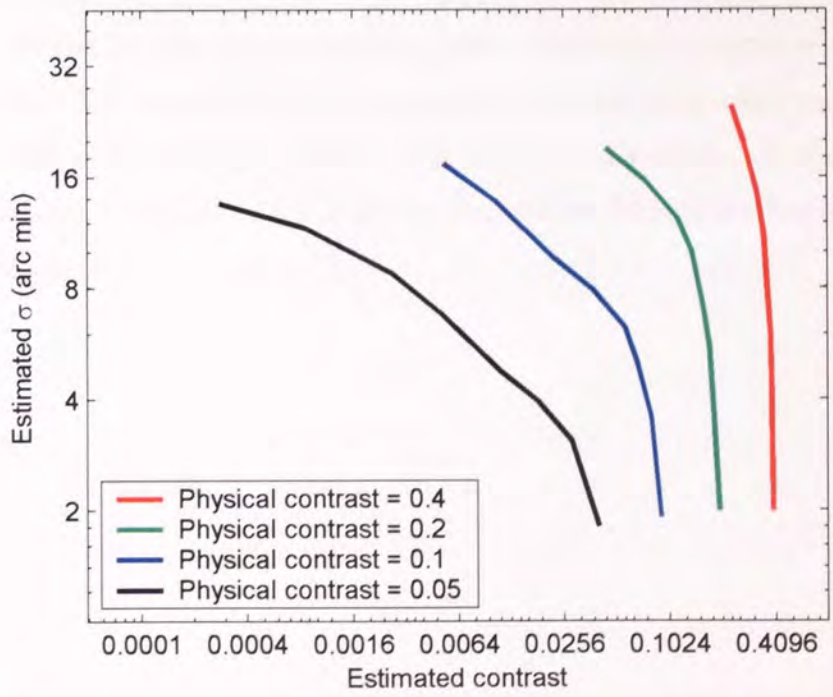


Figure 10.6. Isocontrast contours. This figure is plotted using the same data as Figure 10.5, but each line shows the estimated contrast and scale for an edge of the given physical contrast as its scale is varied from 2′ to 32′.

Chapter 11
Effect of window width on perceived contrast

11.1 Introduction

All the experiments described so far have used edges with a Gaussian integral profile. These are ideal edges for the model, because the templates have this profile. It was important to test the robustness of the model by using edges that were non-Gaussian. One way of generating a non-Gaussian edge is to take a Gaussian edge with a sharp window, and reduce the width of the window so that the edge profile is truncated. For a fixed window size, as the scale of the edge increases, the Michelson contrast will decrease, because the edge profile will reach the edge of the window before the maximum and minimum luminance has been reached. Such a stimulus presents a considerable to challenge the GDT model, because the equation for estimating edge contrast ((2.35) or (10.1)) was derived assuming a Gaussian edge profile, and estimated contrast therefore has no direct relation to the Michelson contrast. Experiment 5 found the perceived contrast of this kind of stimulus.

11.2 Experiment 5: effect of edge scale and window width on perceived contrast

11.2.1 Methods and stimuli

The test edges were vertical Gaussian edges with 4 different scales (2′, 4′, 8′, 16′), and rectangular windows with 5 different widths (20′, 40′, 80′, 160′, 320′). All the edges had a window height of 320′. The windows had sharp edges. The profiles of the test edges are shown in Figure 11.1. The viewing distance and screen resolution were the same as for experiment 2, described in section 5.3.2. Subjects were asked to judge the perceived contrast of the central edge in each stimulus, and to ignore the rest of the stimulus. The contrast of the test edge was always 0.2 before windowing, but the window caused the Michelson contrast to decrease with decreasing window width and increasing edge scale. The comparison stimulus was a vertical Gaussian edge with a scale of 2′ and a window with width 20′ and height 320′. The test and comparison stimuli were separated by an ISI of 500 ms. The polarity of test and comparison edges was the same within a trial, and was generated randomly on each trial. For each test edge, the contrast-matching procedure described in section 4.2 was used to find the contrast of the comparison edge that perceptually matched the test edge in perceived contrast.

Each of the 20 combinations of levels of the independent variables (4 scales and 5 window widths) was assigned to a different staircase, and all 20 staircases were tested within one session. Each subject completed 8 similar sessions. Starting points of the staircases were determined using the same procedure as in experiment 2, described in section 5.3.3.

11.2.2 Results

The results are shown in Figure 11.1. The trace at the bottom of each panel shows the test edge profile. These profiles show that the Michelson contrast decreases with increasing test edge scale, and that the size of this effect decreases with increasing window width. The perceived contrast shows the same pattern. However, the results are complicated by the fact that, as the window width increases, there is an overall decrease in the perceived contrast. This effect does not follow the Michelson contrast, which always increases with increasing window width.

11.3 Simulation 12: simulating experiment 5 with a power-function rectifier

11.3.1 Methods and stimuli

Each stimulus from experiment 5 was generated using an image matrix that had the same width as the stimulus window. This was justified by simulation 12.1, described below. Simulation 12.2 confirmed that the 1D model gave the same results as the 2D model. The experiment was simulated using the contrast-matching procedure described in section 8.3. The comparison stimulus always had a scale of 2′, as in the experiment. The model was run using the power-function rectifier, with the set of parameters used in simulation 5. All other parameters were as described in Table 3.1.

It was known that the model would be unable to account for the overall decrease in perceived contrast with large window widths, because simulation 7.1 had demonstrated that increasing the window width from 10 times the edge scale to 320′ had only a negligible effect on estimated contrast. This effect must be caused by mechanisms that are not included in the current version of the model. Because of this, best-fitting parameters were found by fitting the model to the data for the 20′ and 40′ width conditions only.

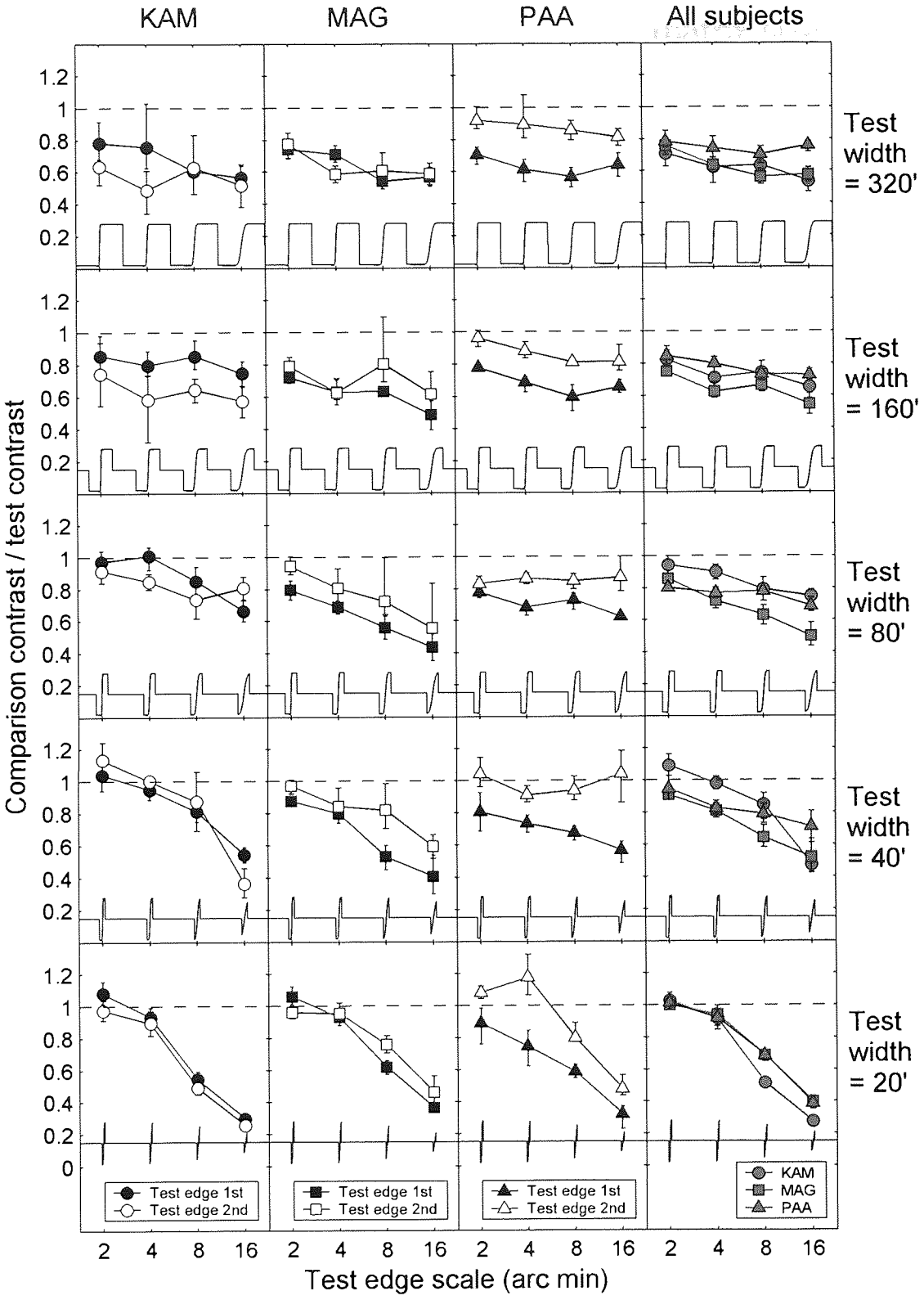


Figure 11.1. Results of experiment 5. Each row of panels shows data for a particular test stimulus width. Otherwise, the data presentation follows the same pattern as Figure 10.1. The trace along the bottom of each panel shows the profile of the test stimulus in each condition.

11.3.1.1 Simulation 12.1: effect of image matrix size

The stimuli from experiment 5 were constructed using an image matrix that was either 320′ wide or was set to the width of the stimulus window. Using the image resolution in the main simulations (given in Table 3.1), the numbers of pixels in the 20′, 40′, 80′, 160′ and 320′ windows were 80, 160, 320, 640, and 1280, respectively. These stimuli were run through the model with an unbiased half-wave rectifier. As argued in section 9.1.1.1, this setting would be the most likely to reveal any effect of image matrix size. All other parameters were as given in Table 3.1. Using the larger image matrix size had no effect on the perceived blur of the stimuli, and made a difference of less than 1×10^{-16} to the perceived contrast.

11.3.1.2 Simulation 12.2: comparison of 1D and 2D models

The test stimuli of experiment 5 were run through the 2D model, to check that it gave the same results as the 1D model. The model was run with an unbiased half-wave rectifier, and with the power-function rectifier with the parameters that fitted best to KAM's data for all the experiments ($T_1 = 0.04$, $T_2 = 0.3$, $n = 2.38$). The data from the 1D model were obtained from the main part of simulation 12. Since it was expected that the 2D model would produce similar, if not identical, results to the 1D model, the 2D stimulus and model parameters were adjusted to be compatible with the results from the 1D model. For each stimulus, the image resolution was set so that 1 pixel was 1/16 of the scale estimated from the 1D model. The scale of the smallest template was set to 4 times this scale (so the gradient operators had a scale of 1 pixel). The ratio of largest to smallest template scales, $\sigma_t(N)/\sigma_t(1)$, was set to 8. This meant that the highest template scale was twice the scale estimated with the 1D model, so if the 2D model's blur estimate was within about 1 octave of the 1D model's estimate, it would fall within the range of templates used. The height and width of the image matrix used with the 2D model was 10 times the edge scale, whatever the stimulus width. The model was run with 256 channels.

Table 11.1 shows the estimated contrasts, and Table 11.2 shows the estimated blurs. For each test edge, the 1D and 2D models give almost identical results, justifying the use of the 1D model in the main part of simulation 12.

Table 11.1. Comparison between 1D and 2D models with two different parameter settings. The shaded areas give the models' estimated contrast for each of the test edges.

			$T_1 = 0$, $T_2 = 0$, $n = 1$				$T_1 = 0.04$, $T_2 = 0.3$, $n = 2.38$			
			Test edge scale				Test edge scale			
			2′	4′	8′	16′	2′	4′	8′	16′
Test edge scale	320′	1D	0.198	0.241	0.182	0.104	0.197	0.203	0.149	0.068
		2D	0.201	0.241	0.183	0.104	0.196	0.202	0.149	0.069
	160′	1D	0.198	0.201	0.238	0.185	0.197	0.182	0.165	0.113
		2D	0.201	0.200	0.241	0.185	0.196	0.185	0.165	0.116
	80′	1D	0.198	0.201	0.199	0.235	0.197	0.182	0.159	0.114
		2D	0.201	0.200	0.200	0.237	0.196	0.185	0.162	0.115
	40′	1D	0.198	0.201	0.199	0.202	0.197	0.182	0.159	0.114
		2D	0.201	0.200	0.200	0.200	0.196	0.185	0.162	0.115
	20′	1D	0.198	0.201	0.199	0.202	0.197	0.182	0.159	0.114
		2D	0.201	0.200	0.200	0.200	0.196	0.185	0.162	0.115

Table 11.2. Comparison between 1D and 2D models with two different parameter settings. The shaded areas give the models' estimated blur for each of the test edges.

			$T_1 = 0$, $T_2 = 0$, $n = 1$				$T_1 = 0.04$, $T_2 = 0.3$, $n = 2.38$			
			Test edge scale				Test edge scale			
			2′	4′	8′	16′	2′	4′	8′	16′
Test edge scale	320′	1D	1.99	4.34	5.66	6.11	1.99	4.02	5.55	5.88
		2D	2.01	4.34	5.67	6.12	1.97	3.99	5.52	5.94
	160′	1D	1.99	4.02	8.60	11.4	1.99	3.80	7.25	10.6
		2D	2.01	3.99	8.69	11.5	1.97	3.83	7.26	10.8
	80′	1D	1.99	4.02	7.97	17.1	1.99	3.80	7.11	12.3
		2D	2.01	3.99	7.99	17.2	1.97	3.83	7.18	12.4
	40′	1D	1.99	4.02	7.97	16.1	1.99	3.80	7.11	12.3
		2D	2.01	3.99	7.99	16.0	1.97	3.83	7.18	12.4
	20′	1D	1.99	4.02	7.97	16.1	1.99	3.80	7.11	12.3
		2D	2.01	3.99	7.99	16.0	1.97	3.83	7.18	12.4

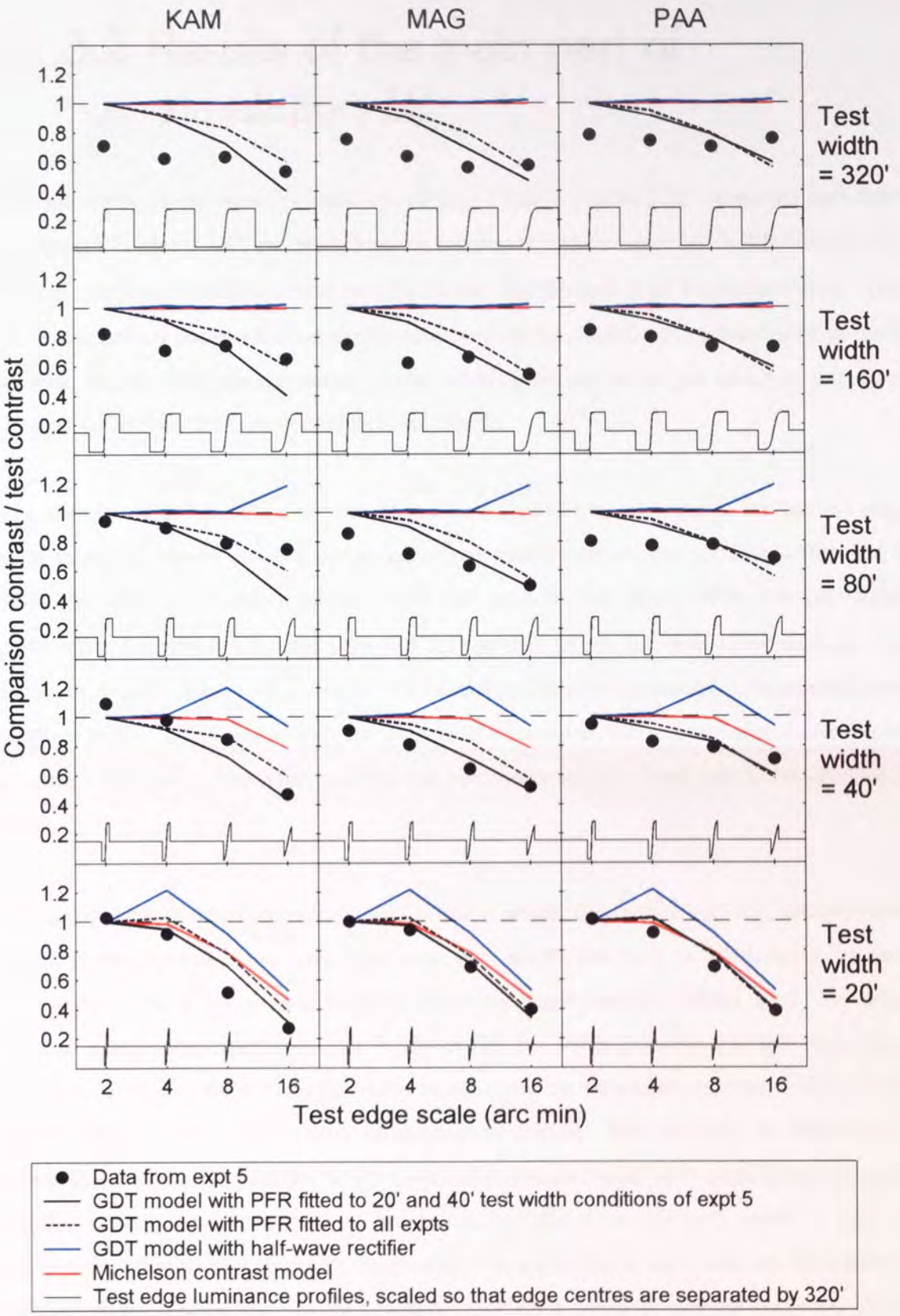


Figure 11.2. Results of simulation 12, which simulated experiment 5 with a power-function rectifier. The black lines show the best-fitting models. The blue line shows the predictions of the pure form of the GDT model, described in Chapter 2. The red line shows the predicted results if the edges match in perceived contrast when they match in Michelson contrast.

11.3.2 Results of the main part of simulation 12

The best-fitting parameters are shown in section 13.3.1. Figure 11.2 shows the data from experiment 5, along with the predictions of the power-function rectifier that fits best to the 20′ and 40′ width conditions, and also the model that fits best to all the experiments. This figure also shows the predictions of the pure form of the model, with a standard half-wave rectifier, and the Michelson contrast model, which predicts that edges match in perceived contrast when their Michelson contrasts are equal.

One clear result is that the pure form of the model performs quite badly. As the test edge scale increased, there was a peak in perceived contrast when the window width was 5 times the edge scale, but this effect was not seen in the data. With the best-fitting parameters, the model accounts very well for the data in the 20′ and 40′ conditions. As expected, it was unable to account for the overall decrease in perceived contrast with large window widths. The parameter set that fitted best to all the experiments gave rise to predictions that were almost as good as the parameter set that fitted best to experiment 5 alone.

One striking result was that, with the 20′ window width, the best-fitting GDT model made almost identical predictions to the Michelson contrast model (both of which fitted the data very well). This is quite remarkable because, as mentioned in section 11.1, the edge contrast is estimated using equation (2.35) (or (10.1)), which assumes that the edge has a Gaussian profile, so there is no particular reason why the estimated contrast should match the Michelson contrast for a non-Gaussian edge profile. The blue line in Figure 11.2 shows that, in fact, the pure form of the algorithm (implemented with a standard half-wave rectifier) produces estimates that do not match the Michelson contrast model so well. It appears that the error that would occur using this algorithm is cancelled out by the error that occurs because of the modified rectifier, yielding a veridical contrast estimate. With the 40′ window, the decrease in Michelson contrast with increasing test edge scale is much smaller. Here, the data show a stronger effect than the Michelson contrast model. The

best-fitting GDT model accounts for these data in the same way that it accounts for the results of experiment 4.

Equation (2.35) (or (10.1)) gives an insight into how the estimated contrast decreases with decreasing stimulus width with the GDT model. The truncation of the stimulus gives rise to a truncated 2nd derivative signature, so the scale, σ_v, of the best-fitting virtual template is reduced. Equation (2.35) shows this will lead to a reduced contrast estimate.

11.4 Simulation 13: simulating experiment 5 with a Legge-Foley transducer

11.4.1 Methods and stimuli

The methods and stimuli were identical to those of simulation 12, except that the Legge-Foley transducer was used, with the parameters used in simulation 6.

11.4.2 Results

The best-fitting parameters are shown in section 13.3.2. The predicted results are shown in Figure 11.3. The best-fitting Legge-Foley transducer made very similar predictions to the best-fitting power-function rectifier.

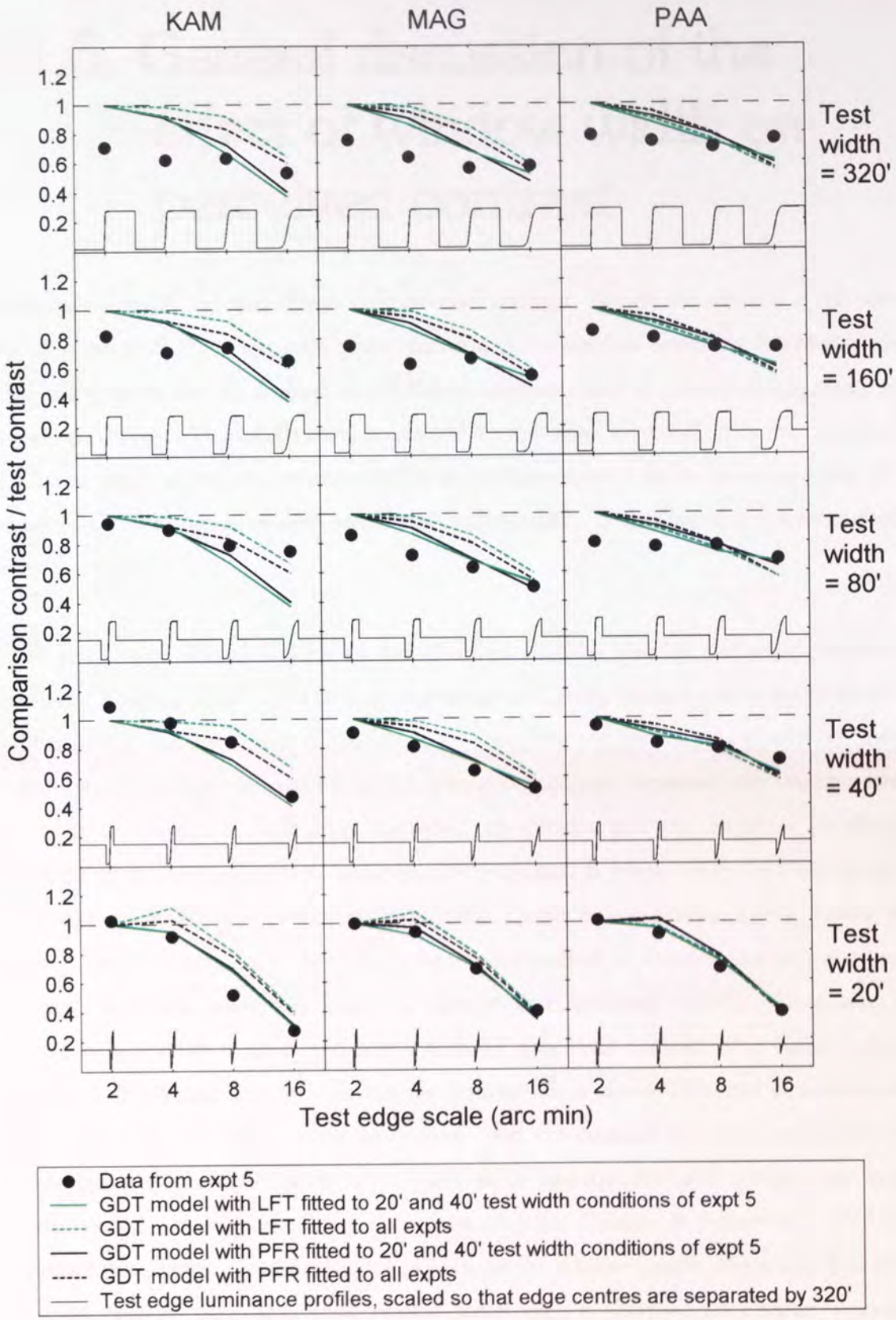


Figure 11.3. Results of simulation 13, which simulated experiment 5 with a Legge-Foley transducer. The black lines show the results from simulation 12 (Figure 11.2).

11.5 General discussion of the effect of window width on perceived contrast

The window width had two effects on perceived contrast. When the window width was less than about 5 times the edge scale, decreasing the window width (or increasing the edge scale) gave rise to a drop in Michelson contrast, and a corresponding drop in perceived contrast. The GDT model accounted for this effect very well. The second effect of window width is that, for window widths larger than about 5 times the edge scale, the perceived contrast decreases with *increasing* window width. This effect was not accounted for by the model.

There have been several reports of the effect of stimulus size on perceived contrast. Takahashi & Ejima (1984) used sine wave gratings with sharp-edged rectangular windows, and varied the width or height of the window. Using the contrast matching method, they found that, for grating contrasts below 0.1, perceived contrast *increased* with stimulus size up to about 14 cycles, replicating the effect of stimulus size on detection threshold (Hoekstra *et al.*, 1974; Savoy & McCann, 1975; Howell & Hess, 1978; McCann *et al.*, 1978; Robson & Graham, 1981; Wright, 1982; Quinn & Lehmkuhle, 1983; Tootle & Berkley, 1983). At contrasts above 0.1, there was no effect of stimulus size on perceived contrast. A similar effect was found by Cannon & Fullenkamp (1988). They used a magnitude estimation method and found that the perceived contrast of a Gabor patch increased with stimulus size for stimulus contrasts less than about 0.06, and was invariant with stimulus size for higher contrasts. They used the contrast matching procedure to confirm that the perceived contrast of a Gabor patch was invariant with stimulus size at a contrast of 0.3. Using stimuli with a contrast of 0.25, Cannon & Fullenkamp (1991b) confirmed the size-invariance of perceived contrast for a Gabor patch, but found that the perceived contrast of a sine wave grating patch with a sharp-edged circular window *decreased* with increasing size. These studies suggest that there is a spatial summation mechanism that works only at low contrasts, giving rise to an increase in perceived contrast with stimulus size for stimuli of lower contrast than those used in experiment 5,

and an inhibitory mechanism that can cause the perceived contrast to decrease with increasing size at the contrast levels used in experiment 5. It is not clear why the inhibitory effect was seen with sharp-edged windows and not Gabor patches.

The experiments on the effect of stimulus size on grating stimuli are not directly applicable to the edge stimuli used in experiment 5. With grating stimuli, the number of features in the signal increases with stimulus size. On the other hand, with the edge stimuli, the number of features remains the same, but they are moved further apart as the stimulus size increases.

There are several possible classes of explanation of the effect of stimulus size. The effect could be a facilitatory effect due to small stimuli, or an inhibitory effect due to large stimuli. In addition, the facilitation/inhibition could occur across spatial frequency or across space. Firstly, consider interactions across spatial frequency: as the stimulus increases in size, its low-frequency content increases, so an inhibitory mechanism that obtained most of its input from the low-frequency band would give rise to a decrease in contrast with increasing stimulus size. Secondly, consider interactions across space: when the window is small, the edges of the window could enhance the response to the central edge; alternatively, when the window is large, the edges of the window could inhibit the response to the central edge. Of course, any combination of these explanations could also apply.

The explanation involving inhibition across frequency from low spatial frequencies is ruled out by several studies. Chubb *et al.* (1989) showed that a noise surround suppressed the perceived contrast of a central patch of noise. In one experiment, the central patch was band-pass filtered to contain frequencies between 5.8 and 11.6 c/deg. They found that a band-pass filtered surround had the greatest inhibitory effect when its spatial frequency matched that of the central patch. Similarly, Cannon & Fullenkamp (1991b) found that the inhibitory effect of a vertical surround grating was tuned to the spatial frequency of a vertical central grating of frequency 8 c/deg. They found that when the central frequency was 2 c/deg, the surround inhibition was not tuned for spatial frequency. Using very similar stimuli to Cannon & Fullenkamp's (1991b) 8 c/deg stimuli, Yu *et al.* (2001) found very different results. They found that a surround with the same spatial frequency as the centre caused inhibition, but that increasing the surround frequency had very little effect,

and decreasing the surround frequency led to an *increase* in perceived contrast of the central patch. They also repeated this experiment with the surround orientated orthogonally to the centre, and found facilitation for surrounds with a lower frequency than the centre, and no effect for surrounds with a higher frequency. In addition, Xing & Heeger (2000) found evidence for facilitation when the surround frequency was 2 octaves lower than the centre. It seems that, depending on the frequency of the centre, and individual differences between subjects, reducing the frequency of the surround, so its frequency no longer matches the centre, can have no effect on the strength of inhibition, or can weaken the inhibition, or give rise to facilitation. The only pattern of results that has not been observed is a stronger inhibitory effect from a low-frequency surround, which is required for the cross-frequency inhibition explanation of the results of experiment 5.

The possible spatial-domain interactions are facilitation when the edges of the window are close to the central edge, or inhibition when these edges are far from the central edge. The inhibitory explanation seems unlikely. Cannon & Fullenkamp (1991b) examined the effect of a narrow ring of grating surrounding a central patch. The distance between the inner and outer diameter of the ring was held constant, and the inner diameter was varied. These stimuli were more comparable to the stimuli in experiment 5, because there was a region of uniform luminance between the centre and surround. Cannon & Fullenkamp showed that the strongest inhibitory effect occurred when the surrounding ring was immediately adjacent to the central patch, and the effect diminished as the ring expanded away from the central patch, so that the perceived contrast of the central patch increased with increasing ring diameter. It therefore seems unlikely that the decrease in perceived contrast with increasing window size shown in experiment 5 could be explained by increasingly strong inhibition as the edge of the window moves away from the centre.

The other possibility is that, when the edges of the window are close to the central edge, they enhance the perceived contrast. Cannon & Fullenkamp (1993) argued for the existence of a short-range facilitatory mechanism. They found large individual differences between subjects, and classified the subjects into three groups: type 1 subjects (suppressors) showed a reduction in perceived contrast with increasing surround size; type 3 subjects (enhancers) showed an increase in perceived contrast with increasing surround size; type 2 subjects showed an initial increase in perceived contrast as the surround increased, followed by a decrease. They argued that the performance of suppressors was

dominated by an inhibitory mechanism, performance of enhancers was dominated by a facilitatory mechanism, and the performance of type 2 subjects arose from an interaction between these two mechanisms. The fact that enhancement occurred only when the surround size was small suggests that the inhibitory system "appears to pool from a larger area than the enhancement system" (p. 1694). This might explain the effect of stimulus size on perceived contrast seen in experiment 5: when the stimulus is narrow, the edges of the window enhance the perceived contrast of the central edge, due to a facilitatory mechanism that pools over a small area. When the window width increases, the edges of the window fall outside of the receptive field of the facilitatory mechanism, so the perceived contrast reduces. Unfortunately, Cannon & Fullenkamp's data show that the facilitatory mechanism was effective only at low contrasts (0.125 and below). At a contrast of 0.25 (the nearest to the contrast of the stimuli in experiment 5), all the type 3 subjects showed almost no facilitation, so it seems unlikely that such a mechanism could account for the fairly strong effect seen in experiment 5. Ejima & Takahashi (1985) also found evidence for facilitatory and inhibitory mechanisms that work in opposition. They used stimuli that had a central rectangular strip of vertical grating surrounded by two vertical gratings with the same spatial frequency as the centre, also with rectangular windows. In one set of conditions, the central strip was horizontal, with the surrounding gratings vertically adjacent to it; in the other set, the central strip was vertical, with the surrounding gratings horizontally adjacent. When the surround was in phase with the centre, they found the same pattern whether the surround was horizontally or vertically adjacent: for both subjects, when the surround was higher in contrast than the centre, it suppressed the perceived contrast of the centre; when the surround was lower in contrast than the centre, it enhanced the perceived contrast of the centre[19]. The latter effect decreased with increasing centre contrast, unlike the former effect. They argued that, at low contrasts, the operators processing the central stimulus show spatial summation, explaining the enhancing effect of the surround at low contrasts (this mechanism would also explain Takahashi & Ejima's (1984) finding that perceived contrast of a grating

---

[19] It should be noted that Cannon & Fullenkamp failed to replicate Ejima & Takahashi's (1985) pattern of results. Cannon & Fullenkamp (1991b) found inhibition whether the surround contrast was above or below the centre contrast. Cannon & Fullenkamp (1993) found one facilitatory effect that depended on the relative contrast of centre and surround, but this interaction was in the opposite direction of that of Ejima & Takahashi: Cannon & Fullenkamp's (1993) type 2 subjects showed an initial increase in perceived contrast as the surround increased, followed by a decrease, but this facilitation only occurred when the surround contrast was *higher* than the centre, and did not occur at all when the surround contrast was equal to or lower than the centre contrast.

increases with window size at low contrasts). They also argued that the operators processing the surround could inhibit those processing the centre. If the centre had a higher contrast than the surround, then the excitatory mechanism would dominate, whereas, if the surround operators had the higher contrast, then the inhibitory mechanism would dominate, explaining their pattern of results. This proposal still does not explain the results of experiment 5 because, like Cannon & Fullenkamp's (1993) subjects, Takahashi & Ejima's subjects only showed facilitation at low contrasts: the facilitatory effect was absent at a contrast of 0.2.

Many of the mechanisms that have been proposed to explain the effects described above have been phase-invariant second-order contrast-sensitive processes. It is likely that the first-order process of brightness induction also plays a role in contrast perception. Ejima & Takahashi (1985) were the first to suggest this. When their surround gratings were out of phase with the centre, they found differences between the two spatial configurations: when the surround was horizontally adjacent to the centre, the perceived contrast of the centre decreased with increasing surround contrast; when the surround was vertically adjacent to the centre, the perceived contrast of the centre increased with increasing surround contrast. They attributed the perceived increase in contrast to a brightness induction effect: when the out-of-phase surrounding gratings were vertically adjacent to the (vertical) central grating, the dark bars of the surround lined up with the bright bars of the centre, and *vice-versa*, and brightness induction would tend to enhance the contrast of the gratings. This effect would not occur with the horizontally adjacent surround, because the gratings in this configuration were still vertical, so the bars of the centre and surround would not line up. Yu *et al.* (2001) directly tested the brightness induction hypothesis by introducing a small gap between centre and surround, which should reduce the brightness induction effect. They found that this gap abolished the contrast enhancement caused by out-of-phase surrounds, suggesting that this enhancement was caused by brightness induction.

A brightness induction mechanism is an attractive possibility for explaining the effect of window width on perceived contrast in experiment 5. It could be that the induction effect acts quite locally. When the window width is small, the mean-luminance background is close to the central edge. The dark side of this edge is made darker by the light side, and also by the nearby background; the light side of this edge is made lighter by the dark side,

and also by the nearby background. When the window width is large, the background is far from the central edge, so one source of contrast enhancement has been removed: the dark side of the central edge would have only darkness on one side, and the light edge would have only lightness on one side. This provides a possible explanation of the results, but it is clear that a great deal more work would have to be done in order to progress from the speculative suggestions offered here.

Chapter 12
Effect of linear ramp on edge perception

12.1 Introduction

Very many edge detection algorithms operate by finding ZCs in the 2nd derivative. In particular, some of the most well-known models of edge detection in humans have used Gaussian 2nd derivative filters as their first stage (Marr & Hildreth, 1980; Watt & Morgan, 1985; Kingdom & Moulden, 1992). Most researchers have proposed that the edge properties, such as contrast and blur, could be determined from local cues in the vicinity of the edge (e.g., Marr & Hildreth, 1980; Watt & Morgan, 1985; Georgeson, 1994; Elder & Zucker, 1998).

If edge properties are determined by local cues in the 2nd derivative, then adding a linear ramp to an edge stimulus should make no difference to the appearance of the edge. This can be understood by examining Figure 12.1. The top half of this figure shows the effect of adding a ramp with positive gradient to a positive edge; the bottom half shows the effect of adding a ramp with a negative gradient to the same edge. Although the luminance profiles of the resulting stimuli are quite different, their gradient profiles are identical, except for a vertical shift. Therefore, the 2nd derivative of each stimulus will be the same in the vicinity of the edge. If the edge properties are determined by local cues in the 2nd derivative, then these properties will not be changed by the addition of a ramp[20].

[20] Although the local 2nd derivative will be unchanged by the addition of a ramp, the 2nd derivative signals at the points where the where the ramps end will differ according to the ramp gradient. Thus, models that take the 2nd derivative and then estimate edge characteristics using more global cues could predict an effect of ramp on the perceived properties of the edge.

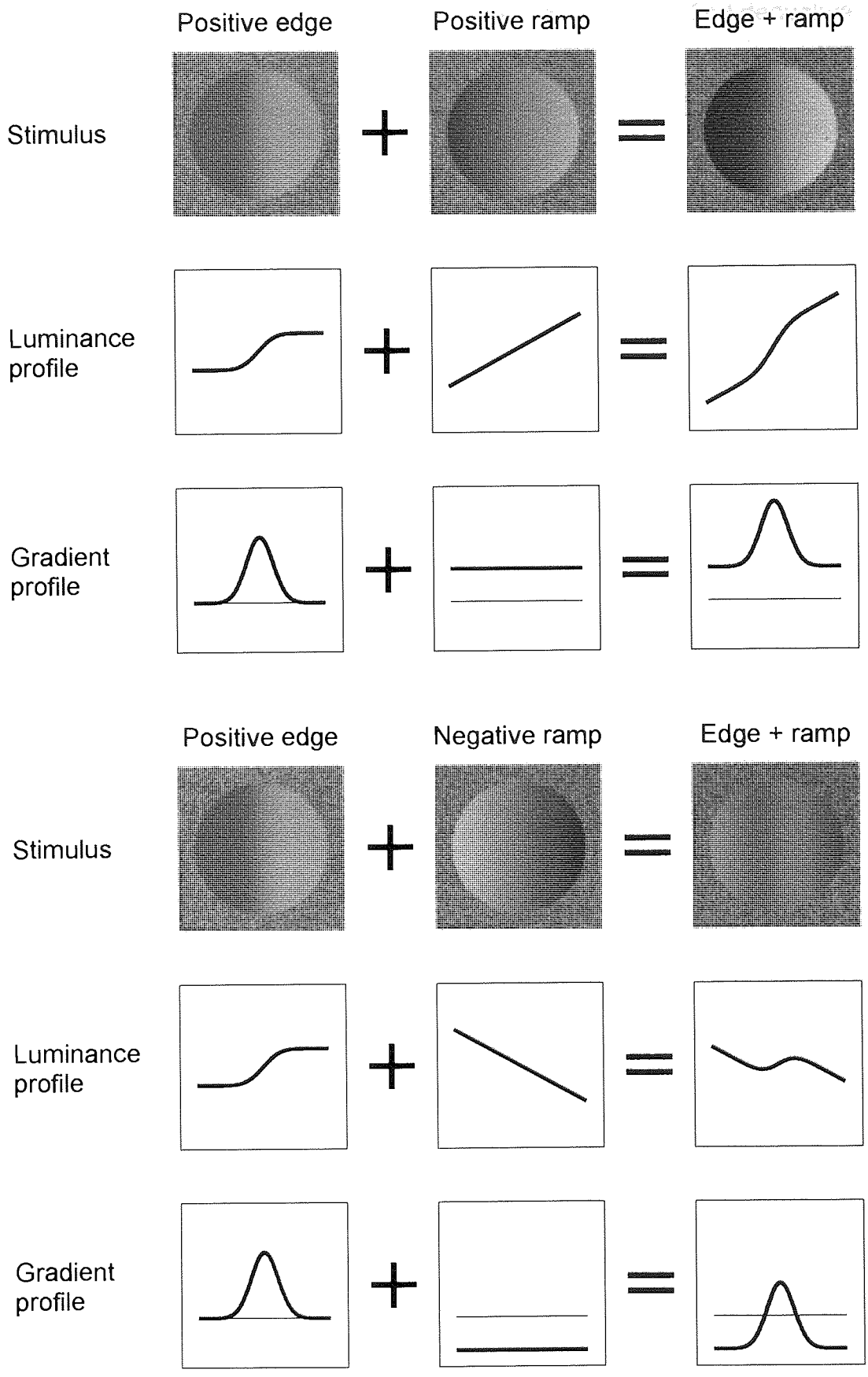


Figure 12.1. Effect of adding a linear ramp to an edge: luminance and gradient profiles.

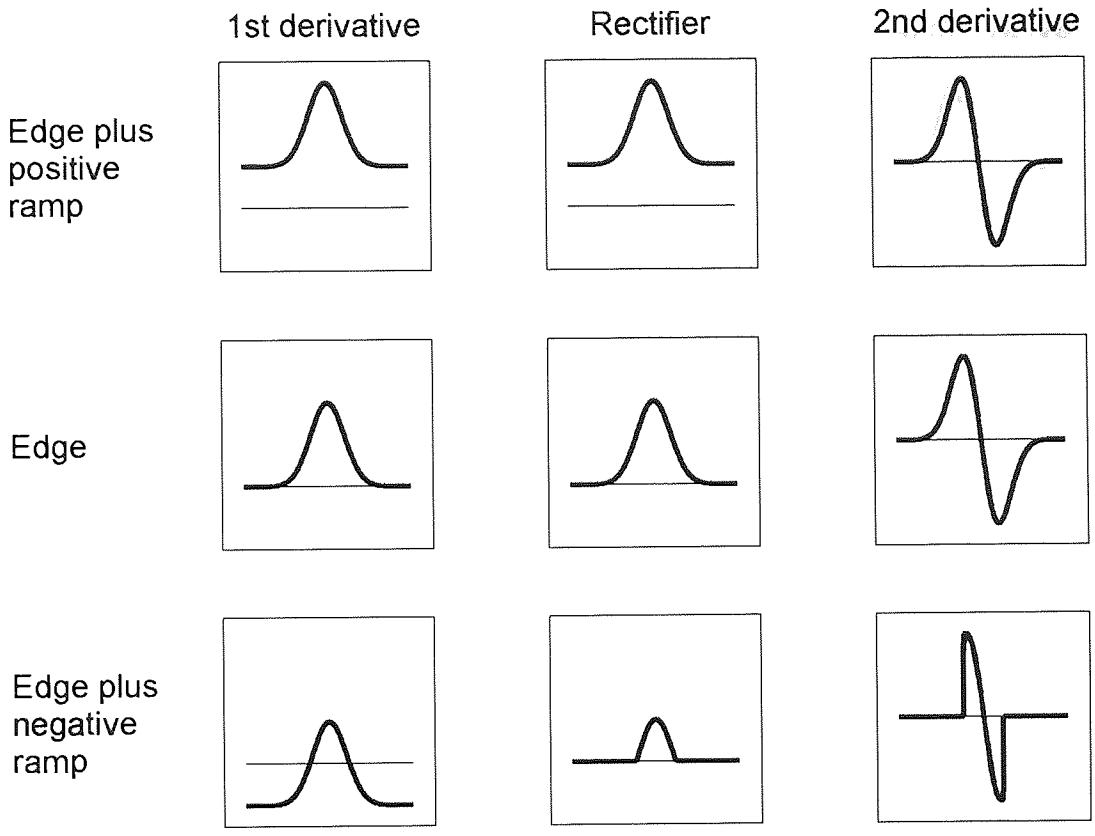


Figure 12.2. Predicted effects of adding a positive or negative ramp to a positive edge, using the GDT model with an unbiased half-wave rectifier. The positive ramp makes no difference to the 2nd derivative signature, so the edge should have the same perceived blur and contrast as the standard edge. The negative ramp causes the gradient profile to be truncated by the rectifier, and this leads to a narrower 2nd derivative signature, which fits a narrower template, making the edge look sharper.

Figure 12.2 shows how adding a positive or negative ramp would affect the processing in the GDT model. When a positive ramp is added to a positive edge, the gradient profile is shifted upwards, so the rectifier has no effect, and the 2nd derivative is unchanged, as with a pure 2nd derivative model. Thus, like the pure 2nd derivative models, the GDT model predicts that adding a ramp with the same polarity as the edge should not change the edge's appearance. However, when a negative ramp is added, the extremities of the gradient profile fall below zero, and the gradient signal is truncated, leading to a truncated 2nd derivative profile. The truncated 2nd derivative signature will fit a narrower template than the signatures for the other two stimuli, and so adding a ramp of opposite polarity to the edge should make it look sharper.

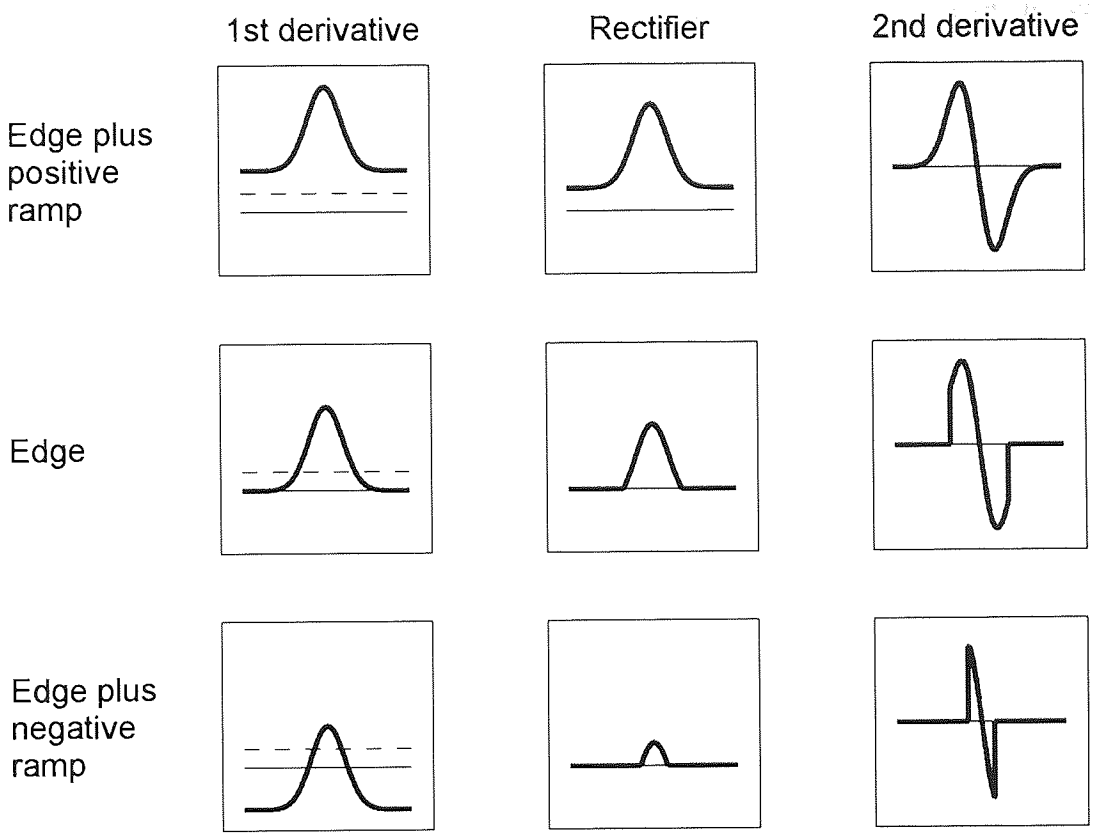


Figure 12.3. Predicted effects of adding a positive or negative ramp to a positive edge, using the GDT model with a biased half-wave rectifier. Unlike in Figure 12.2, the 2nd derivative signature of the standard edge is truncated slightly because the rectifier has a threshold, making the edge look sharper. The positive ramp lifts the gradient signal above the threshold, so it is no longer subject to the sharpening effect, and so it looks more blurred.

The predictions shown in Figure 12.2 assume an unbiased half-wave rectifier. Figure 12.3 shows the predicted effect of adding a ramp when the rectifier has a threshold. Here, the standard edge is truncated slightly by the rectifier. When the positive ramp is added, it lifts the whole gradient signal above the threshold, so the 2nd derivative signature is no longer subject to the sharpening effect of the threshold. In this case, adding the positive ramp should make the edge look more blurred.

Experiment 6 examined how adding a linear ramp affected the perceived blur of an edge, in order to test the predictions outlined in this section. Experiment 7 examined the effect of a linear ramp on perceived contrast. The bottom half of Figure 12.1 shows that the negative ramp reduces the Michelson contrast based on the local maximum and minimum

of luminance. It might therefore be expected that the perceived contrast of the edge would be reduced by the addition of a negative ramp.

12.2 Experiment 6: effect of linear ramp on perceived blur of Gaussian edges

12.2.1 Methods and stimuli

Experiment 6 took the 0.2 contrast test edges from experiment 1 and added a vertical linear ramp before windowing. Viewing distances and diameters of the test and comparison edge windows were the same as in experiment 1, given in Table 5.1. The ramp had a value of zero in the centre of the stimulus (the edge location). The ramp gradient was always expressed as a proportion of the peak edge gradient. The proportions used were -0.75, -0.50, -0.25, 0, 0.25, and 0.5. The window widths and the range of gradients and was limited by the fact that stimuli with a high-gradient ramp descend quickly to zero luminance either side of the edge. This experiment was performed by subjects KAM and MAG. For both subjects, the stimuli were surrounded by 0.2 contrast binary noise, as in experiment 1.2, for the reasons given in section 5.2.1.2. KAM also ran a version of the experiment with no noise surround. The viewing distances, screen resolutions, comparison edges and blur-matching procedure were identical to those in experiment 1.

The blur-matching task was quite difficult. It was felt that one of the difficulties was the qualitative difference between the different types of stimulus, as shown in Figure 12.1. It was felt that the task would be easier if the stimuli were blocked according to the ramp gradient level. The variables manipulated within a block are listed in Table 12.1. Each of the 8 combinations of levels of variables was assigned to a difference staircase. Each session consisted of 6 blocks, one for each ramp gradient. The whole session was conducted at one viewing distance. As in experiment 1, there were 3 viewing distances, with 2 sessions at each viewing distance, and the sessions were conducted in the order

ABCBCA, with A, B and C assigned to viewing distances randomly for each subject. The order of the blocks within the first session at a particular viewing distance was random, and the order of the second session at each viewing distance was the opposite of the order in the first session at that viewing distance.

Table 12.1. Variables manipulated within a block of experiment 6.

Variable	Number of levels per block	Possible values
Test edge scale	2	6′ or 8′ at viewing distance of 260 cm 12′ or 16′ at viewing distance of 130 cm 24′ or 32′ at viewing distance of 65 cm
Orientation/ polarity	2	0° (dark-light) or 180° (light-dark)
Initial comparison scale for staircase	2	random value between 10 and 14 dB above test scale or random value between 10 and 14 dB below test scale

12.2.2 Results

The data for each combination of test edge scale and ramp gradient were analysed as described in section 4.2.3. The results, using stimuli with a noise surround, are shown in Figure 12.4. Figure 12.5 compares KAM's results with and without the surrounding binary noise. As in experiment 1, the noise made a negligible difference. As predicted by the GDT model, the negative ramp made the test edges look sharper. The positive ramps caused a slight increase in the perceived blur of the test edges. This effect was stronger for KAM than MAG.

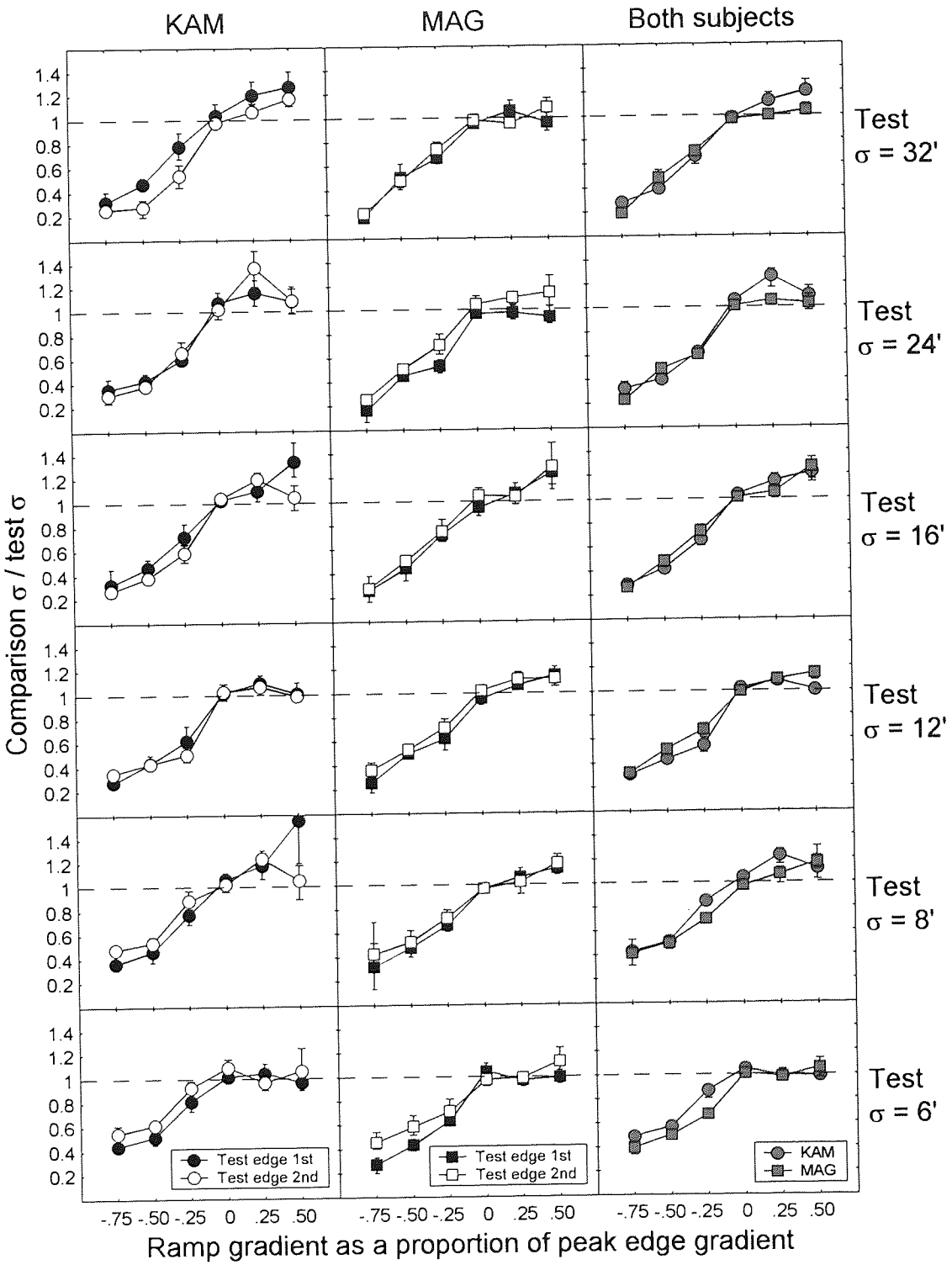


Figure 12.4. Results of experiment 6, using stimuli with a binary noise surround.

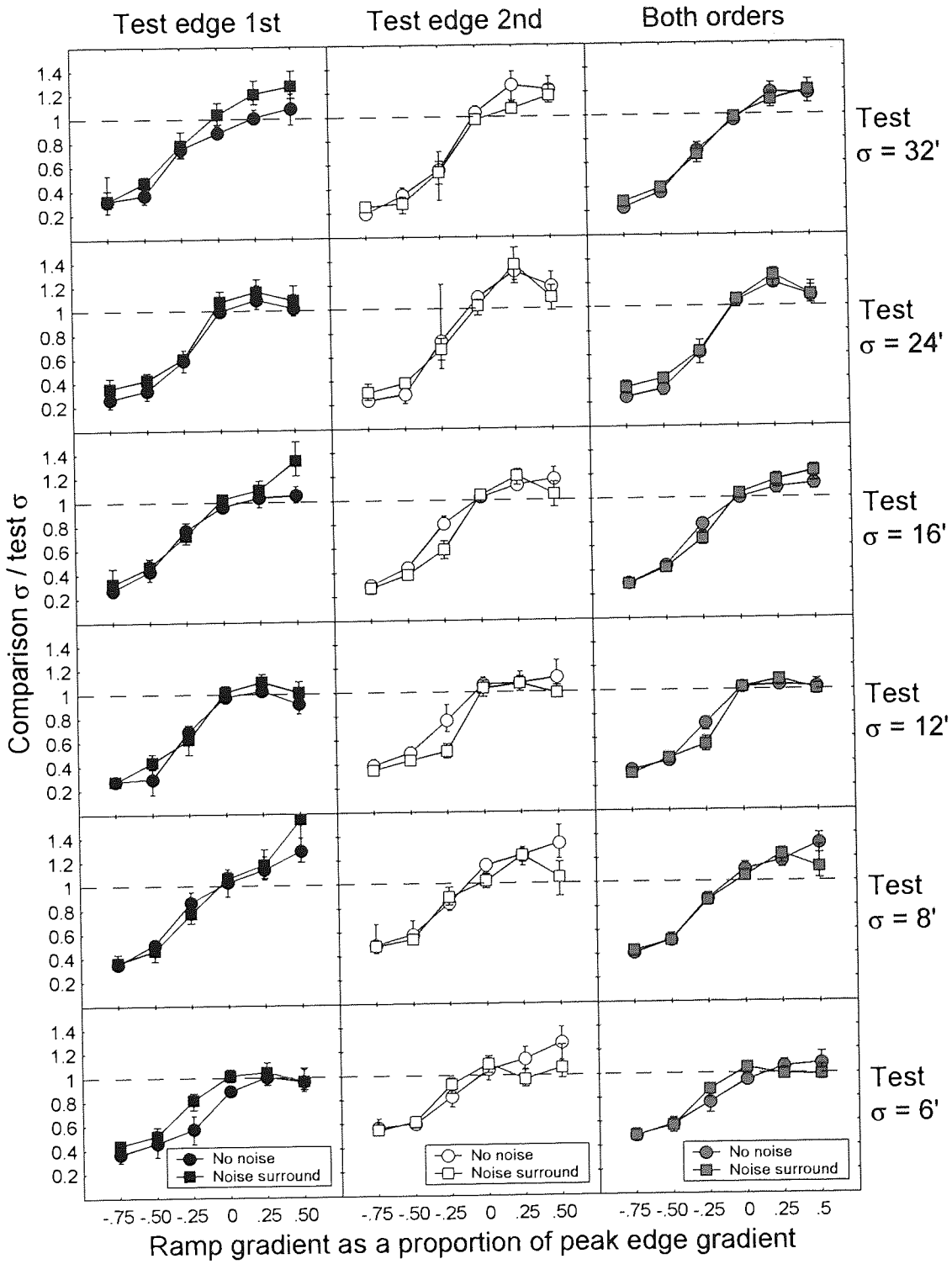


Figure 12.5. Comparison of KAM's results with and without surrounding binary noise in experiment 6.

12.3 Simulation 14: simulating experiment 6 with a power-function rectifier

Each stimulus from experiment 6 was generated. The image matrix sizes for each test edge scale were as used in simulations 3 and 5, given in Table 9.1. The model parameters and methods were the same as in simulation 5. Simulation 14.1 confirmed that the 1D model would give the same results as the 2D model. One complication with the simulations of the experiments with positive ramps is that the steepest positive ramp stimuli often gave rise to two edges superimposed at the same location: one had a perceived blur of around the same order of magnitude as the test edge, and the other was very much more blurred, and corresponded to the entire ramp, rather than the central edge. In these cases, the model used the sharper of the two edges.

12.3.1 Simulation 14.1: comparison of 1D and 2D models

The test stimuli of experiment 6 were run through the 2D model, to check that it gave the same results as the 1D model. The model was run with an unbiased half-wave rectifier, and with the power-function rectifier with the parameters that fitted best to KAM's data for all the experiments ($T_1 = 0.04$, $T_2 = 0.3$, $n = 2.38$). The data from the 1D model were obtained from the main part of simulation 14.

The model and stimulus parameters are given in Table 12.2. Note that, although the visual angle subtended by the spatial envelope varied with edge scale (as detailed in Table 5.1), the image resolution was adjusted so that the envelope diameter corresponded to a constant number of pixels for each stimulus.

The estimated blurs are shown in Table 12.3 and Table 12.4. For each test edge, the 1D and 2D models give almost identical results, justifying the use of the 1D model in the main part of simulation 14.

Table 12.2. The model and stimulus parameters used in simulation 14.1.

Edge scale (min visual angle)	Image resolution (pixels/deg)	Envelope diameter (pixels)	N, the number of channels	$\sigma_t(1)$ (pixels)	$\sigma_t(N)/\sigma_t(1)$
6, 8	192	256	103	4	7
12, 16	96	256	103	4	7
24, 32	48	256	103	4	7

Table 12.3. Comparison between 1D and 2D models with unbiased half-wave rectifier. The shaded areas give the models' estimated blur for each of the test edges in simulation 14.

			Ramp gradient as a proportion of peak edge gradient					
			−0.75	−0.50	−0.25	0	0.25	0.50
Test edge scale	32′	1D	13.6	20.3	26.4	32.0	32.0	32.0
		2D	13.5	20.2	26.3	31.9	31.9	32.5
	24′	1D	10.2	15.2	19.9	24.0	24.0	24.0
		2D	10.1	15.1	19.8	23.9	23.9	23.9
	16′	1D	6.85	10.2	13.3	16.1	16.1	16.1
		2D	6.75	10.1	13.2	15.9	15.9	16.2
	12′	1D	5.15	7.67	10.0	12.1	12.1	12.1
		2D	5.07	7.57	9.89	12.0	12.0	12.0
	8′	1D	3.39	5.05	6.59	7.97	7.97	7.97
		2D	3.38	5.04	6.58	7.97	7.97	8.12
	6′	1D	2.55	3.80	4.95	5.99	5.99	5.99
		2D	2.54	3.79	4.95	5.98	5.98	5.98

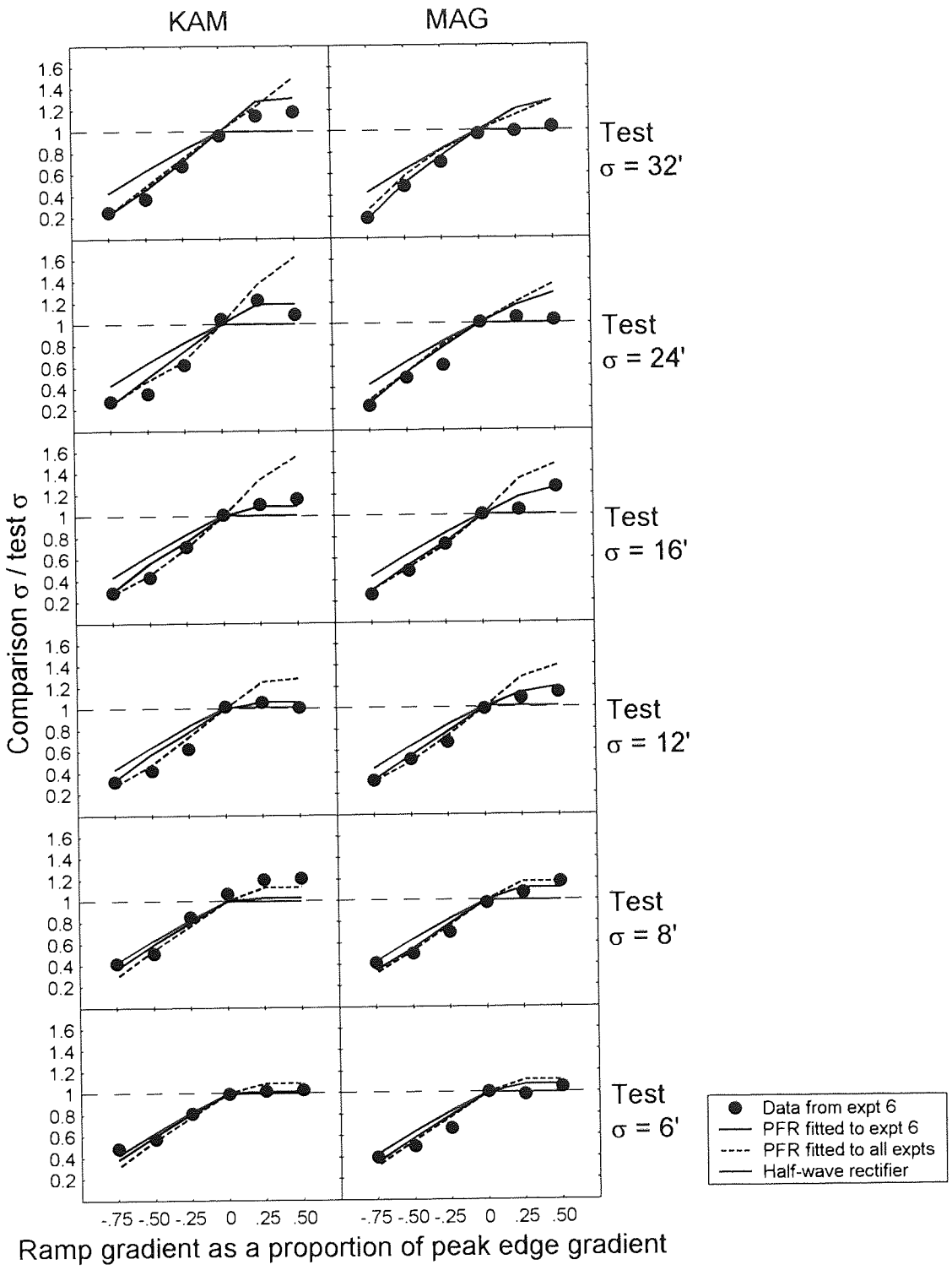


Figure 12.6. Results of simulation 14, which simulated experiment 6 with a power-function rectifier.

In summary, for both subjects, the negative ramps caused a perceived sharpening that was very well predicted by the power-function rectifier model fitted to all the experiments. The positive ramps caused a smaller perceived blurring than predicted by this model. For MAG, even the model that was fitted to MAG's data from experiment 6 failed to predict a sufficiently large effect of positive ramps.

12.4 Simulation 15: simulating experiment 6 with a Legge-Foley transducer

12.4.1 Methods and stimuli

The methods and stimuli were identical to those of simulation 14, except that the Legge-Foley transducer was used, with the parameters used in simulation 6.

12.4.2 Results

The best-fitting parameters are shown in section 13.3.2. The predicted results are shown in Figure 11.3. For most conditions, the best-fitting Legge-Foley transducer made very similar predictions to the best-fitting power-function rectifier. The Legge-Foley transducer that fitted best to MAG's data from experiment 6 appeared more successful than the power-function rectifier at predicting MAG's lack of a blurring effect with positive ramps. However, the Legge-Foley transducer that fitted best to all the experiments made almost identical predictions to the power-function rectifier that fitted best to all the experiments.

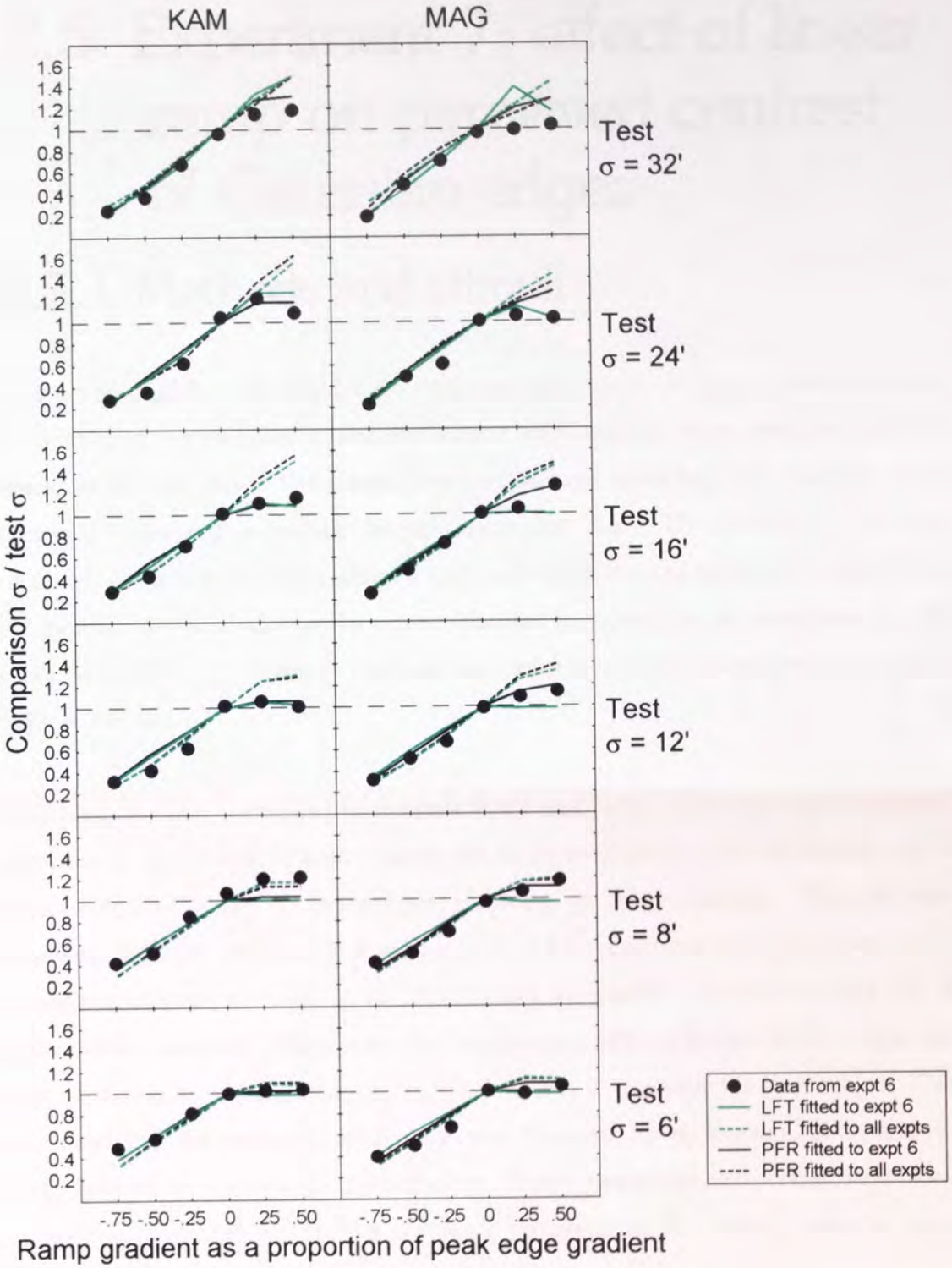


Figure 12.7. Results of simulation 15, which simulated experiment 6 with a legge-foley transducer. The black lines show the predicted results using the power-function rectifier in simulation 14 (Figure 12.6).

12.5 Experiment 7: effect of linear ramp on perceived contrast of Gaussian edges

12.5.1 Methods and stimuli

Experiment 7 used the test stimuli from experiment 6 without the binary noise surround. The comparison stimulus was a vertical Gaussian edge with the same scale and window diameter as the test edge. The comparison contrast was controlled by a staircase, using the contrast matching procedure describe in section 4.2. The polarity of test and comparison edges was the same within a trial, and was generated randomly on each trial. Each stimulus was accompanied by a tone identical to those used in experiment 4. ISI duration was 500 ms. Viewing distances and screen resolution were the same as in experiments 1 and 6.

This experiment was performed by subjects KAM and MAG, after they had completed experiment 6. By this time, both subjects felt more comfortable with the stimuli, and it was no longer necessary to block them according to ramp gradient. The variables manipulated within a session are shown in Table 12.5. Each combination of levels was assigned to a different staircase, giving 24 staircases per session. 12 sessions were run, 4 at each viewing distance. They were run in the order ABC ACB CBA BCA, where the viewing distances were assigned to the letters A, B and C randomly for each subject. This order meant that the mean ordinal position was the same for A, B and C (6.5), and no viewing distance was used twice in succession. It also meant that each consecutive block of 3 sessions contained each of the 3 viewing distances, so the viewing distances were distributed evenly throughout the sequence of sessions.

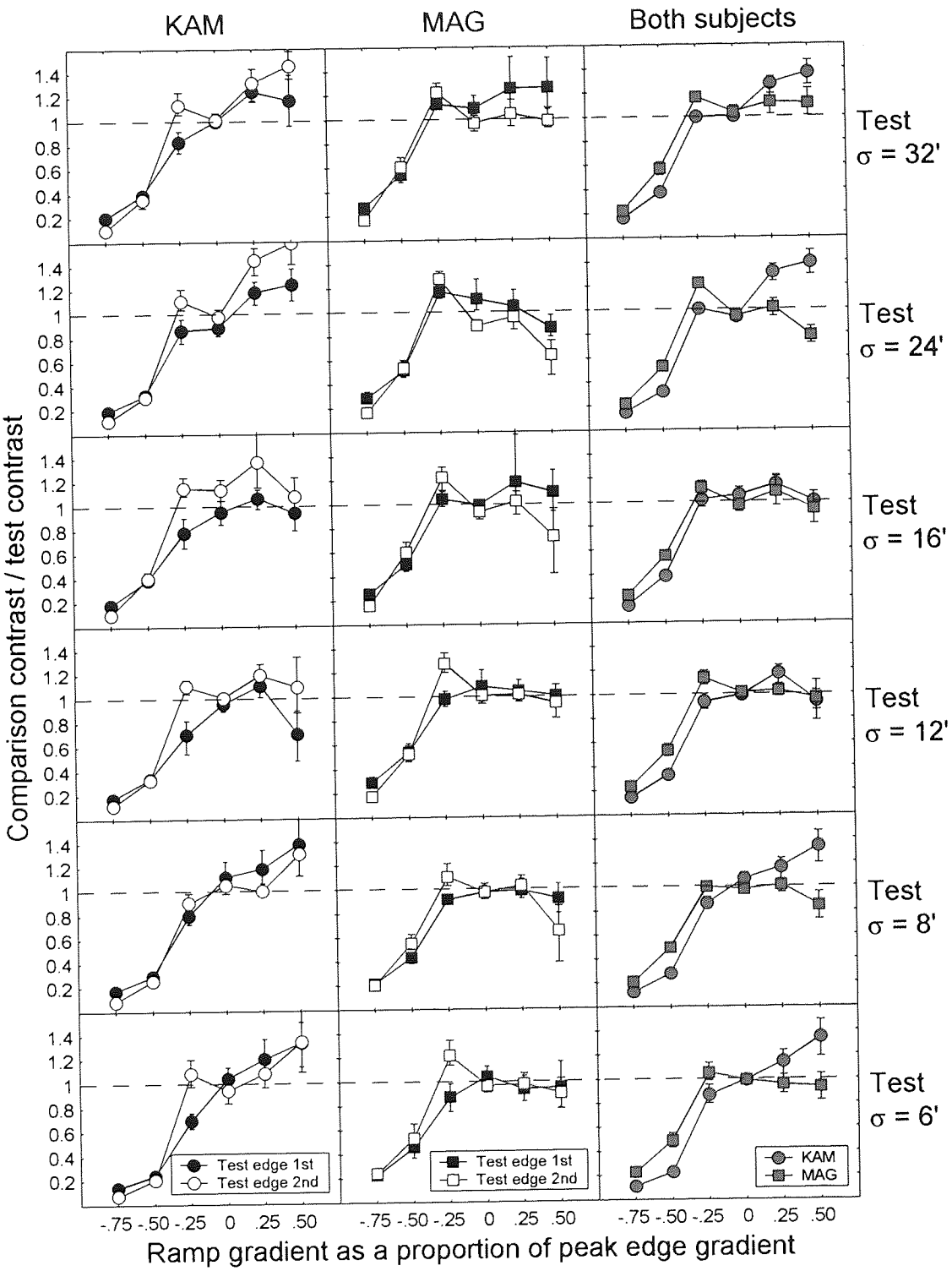


Figure 12.8. Results of experiment 7.

12.5.2 Results

The results are shown in Figure 12.8. As expected, the negative ramp caused a reduction in perceived contrast. The positive ramp had virtually no effect for MAG, but caused an increase in perceived contrast for KAM. These results were similar to those of experiment 6, where the positive ramp had a smaller effect on MAG's results.

Table 12.5. Variables manipulated within a session of experiment 7.

Variable	Number of levels per session	Possible values
Test edge scale	2	6′ or 8′ at viewing distance of 260 cm 12′ or 16′ at viewing distance of 130 cm 24′ or 32′ at viewing distance of 65 cm
Ramp gradient as a proportion of peak test edge gradient	6	−0.75, −0.50, −0.25, 0, 0.25, and 0.5
Initial comparison scale for staircase	2	random value between 10 and 14 dB above test scale or random value between 10 and 14 dB below test scale

12.6 Simulation 16: simulating experiment 7 with a power-function rectifier

12.6.1 Methods and stimuli

The methods and stimuli were the same as for simulation 14, except that the contrast matching procedure described in section 8.3 was used. Simulation 16.1 confirmed that the 1D model would give the same results as the 2D model.

12.6.1.1 Simulation 16.1: comparing 1D and 2D models.

The test edge contrast data from the 2D model used in simulation 14.1 were compared with the test edge contrasts from the 1D model used in the main part of simulation 16. These data are shown in Table 12.6 and Table 12.7. The two models give almost identical contrast estimates for each test edge, justifying the use of the 1D model in the main part of simulation 16.

Table 12.6. Comparison between 1D and 2D models with unbiased half-wave rectifier. The shaded areas give the models' estimated contrast for each of the test edges in simulation 16.

			Ramp gradient as a proportion of peak edge gradient					
			-0.75	-0.50	-0.25	0	0.25	0.50
Test edge scale	32′	1D	0.0246	0.0720	0.136	0.200	0.200	0.200
		2D	0.0244	0.0715	0.136	0.199	0.199	0.205
	24′	1D	0.0247	0.0722	0.137	0.200	0.200	0.200
		2D	0.0245	0.0717	0.136	0.199	0.199	0.199
	16′	1D	0.0249	0.0728	0.138	0.202	0.202	0.202
		2D	0.0244	0.0715	0.136	0.199	0.199	0.205
	12′	1D	0.0250	0.0731	0.138	0.203	0.203	0.203
		2D	0.0245	0.0717	0.136	0.199	0.199	0.199
	8′	1D	0.0245	0.0716	0.136	0.199	0.199	0.199
		2D	0.0244	0.0715	0.136	0.199	0.199	0.205
	6′	1D	0.0246	0.0719	0.136	0.200	0.200	0.200
		2D	0.0245	0.0717	0.136	0.199	0.199	0.199

Table 12.7. Comparison between 1D and 2D models with a power-function rectifier with parameters $T_1 = 0.04$, $T_2 = 0.3$, $n = 2.38$. The shaded areas give the models' estimated contrast for each of the test edges in simulation 16.

			Ramp gradient as a proportion of peak edge gradient					
			-0.75	-0.50	-0.25	0	0.25	0.50
Test edge scale	32′	1D	0.0001	0.0038	0.0168	0.0445	0.0903	0.143
		2D	0.0001	0.0038	0.0171	0.0442	0.0897	0.142
	24′	1D	0.0004	0.0073	0.0296	0.0733	0.132	0.180
		2D	0.0004	0.0072	0.0293	0.0727	0.131	0.184
	16′	1D	0.0013	0.0161	0.0566	0.114	0.175	0.202
		2D	0.0013	0.0162	0.0555	0.115	0.178	0.199
	12′	1D	0.0025	0.0259	0.0749	0.138	0.196	0.203
		2D	0.0025	0.0261	0.0734	0.139	0.193	0.199
	8′	1D	0.0056	0.0399	0.0937	0.159	0.199	0.199
		2D	0.0056	0.0398	0.0936	0.159	0.199	0.205
	6′	1D	0.0090	0.0474	0.104	0.173	0.200	0.200
		2D	0.0092	0.0472	0.104	0.173	0.199	0.199

12.6.2 Results of the main part of simulation 16

The best-fitting parameters are shown in section 13.3.1. Figure 12.9 shows the data from experiment 7, along with the predictions of the model that fits best to experiment 7, and also the model that fits best to all the experiments. This figure also shows the predictions of the pure form of the model, with an unbiased half-wave rectifier. All the models successfully predicted the results of the two steepest negative ramp conditions. However, the data points from the −0.25 ramp gradient conditions consistently deviated from the predictions. The models predicted a substantial decrease in perceived contrast, but no such effect was seen in the data. Figure 12.8 shows that this failure of the data to fit the model more often applied to the data from the trials in which the test edge came second. For most conditions, the data are very similar for both stimulus orders but, in the −0.25 ramp gradient conditions, the contrast ratio is consistently higher when the test edge came second. In particular, KAM's data for the "test edge 1st" condition fitted the model considerably better. It is not clear why this discrepancy between the two test orders should occur for this one condition.

The best-fitting model was quite successful at predicting the blurring effect of the positive ramps for KAM's data. MAG showed no effect of positive ramp, contradicting the predictions of the best-fitting model.

Figure 12.9 also shows the predictions of the Michelson contrast model. With the negative ramp stimuli, there is a local maximum and minimum of luminance either side of the edge (see Figure 12.1). With these stimuli, the comparison edge was considered to match the test edge in Michelson contrast when the local peak-trough luminance difference in the test edge matched the luminance difference across the comparison edge. Figure 12.9 shows that, for the negative ramp stimuli, the Michelson contrast model gives very similar predictions to the other models, and fits the data well. As with the simulations of experiment 5 (in Chapter 11), this is quite remarkable, because there is no particular reason why the model's estimated contrast of a non-Gaussian edge should match the edge's local Michelson contrast. The reason that the negative ramp causes a reduction in estimated contrast is similar to the reason for the reduction in estimated contrast due to

edge truncation, shown in Chapter 11: the negative ramp leads to a reduction in estimated blur (explained in Figure 12.2 and Figure 12.3), and according to equation (2.35) (or (10.1)), this leads to a reduced contrast estimate.

With the positive ramp stimuli, where is no local peak and trough, so it is more difficult to apply the Michelson contrast model. In this case, the only consistent option was to use the Michelson contrast of the entire test stimulus. As Figure 12.9 shows, this leads to a massive over-estimation of the perceived contrast.

12.7 Simulation 17: simulating experiment 7 with a Legge-Foley transducer

12.7.1 Methods and stimuli

The methods and stimuli were identical to those of simulation 16, except that the Legge-Foley transducer was used, with the parameters used in simulation 6.

12.7.2 Results

The best-fitting parameters are shown in section 13.3.2. The predicted results are shown in Figure 12.10. For most conditions, the best-fitting Legge-Foley transducer made very similar predictions to the best-fitting power-function rectifier. The pattern of results was very similar to those of simulation 15: the Legge-Foley transducer that fitted best to MAG's data from experiment 7 was more successful than the power-function rectifier at predicting MAG's lack of an effect with positive ramps, but the Legge-Foley transducer that fitted best to all the experiments made almost identical predictions to the power-function rectifier that fitted best to all the experiments.

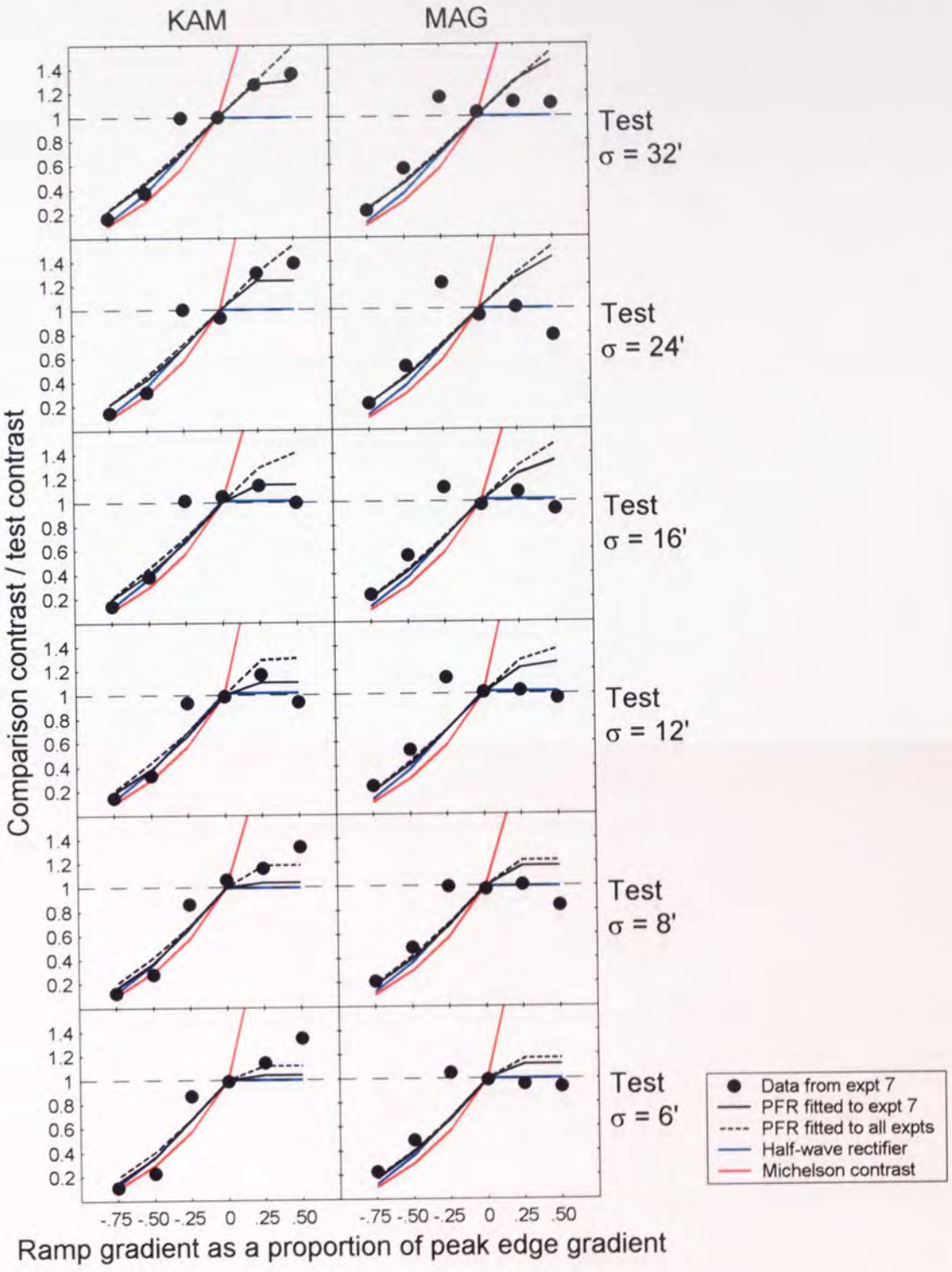


Figure 12.9. Results of simulation 16, which simulated experiment 7 with a power-function rectifier. The blue lines show the predictions using an unbiased half-wave rectifier. The contrast ratios predicted by the Michelson contrast model were too high to plot on these axes. The ratios for the 0.25 and 0.50 ramp gradient conditions were 0.399 and 0.599 for the 6′, 12′, and 24′ stimuli, and 0.466 and 0.732 for the 8′, 16′ and 32′ stimuli.

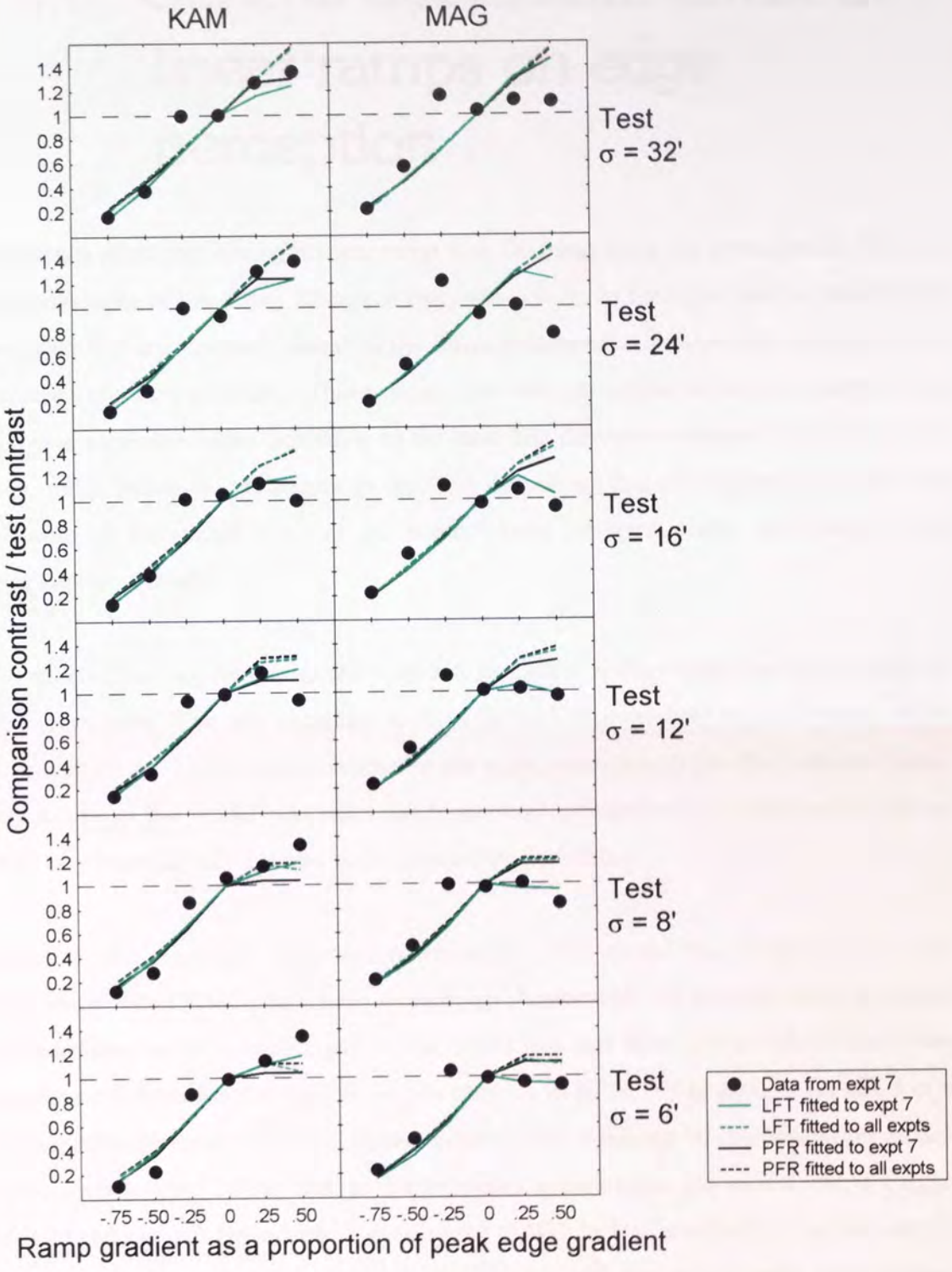


Figure 12.10. Results of simulation 17, which simulated experiment 7 with a Legge-Foley transducer.

12.8 General discussion: effect of linear ramps on edge perception

The results show that adding a linear ramp to a Gaussian edge has a substantial effect on the appearance of the edge. Ramps of opposite polarity to the edge cause a reduction in perceived blur and contrast; ramps of the same polarity can cause a slight increase in the perceived blur and contrast. These results rule out any model of edge perception that estimates edge parameters according to the local 2nd derivative, because the addition of a linear ramp makes no difference to the 2nd derivative: the only difference in the 2nd derivative of the stimuli occur at the points where the ramp stops, and that is some distance from the edge.

The effect of the negative ramp was very well predicted by the model that fitted best to all the experiments. The only exception to this was the lowest-gradient negative ramp, which had no effect on the perceived contrast of the edge, even though an effect was predicted. This failure of the model was restricted to contrast perception: the same ramp had an effect on perceived blur that was well predicted by the model.

The effect of the positive ramp was more erratic. The model that fitted best to all the experiments fitted KAM's data from experiment 7 quite well, but overestimated the effect of the positive ramp in experiment 6; the model that was fitted just to KAM's data from experiment 6 fitted the data quite well. In contrast to KAM, MAG showed no effect of a positive ramp in some conditions where an effect was predicted by the best-fitting model. Of the models fitted to the data of the individual experiments, the model with a Legge-Foley transducer was better able to account for MAG's lack of an effect of positive ramps. However, the Legge-Foley model that fitted best to all the experiments made almost identical predictions to the corresponding power-function rectifier model. It is possible that the reduced effect of the positive ramp is caused by a gain control mechanism acting on the half-wave rectifier output. The positive ramp would lead to a high gradient signal over a large area, whereas the negative gradient signal from the negative ramp would be

removed by the rectifier. Thus a gain control mechanism would selectively reduce the effect of the positive ramp. All attempts to introduce a gain control mechanism to the model gave rise to perturbations in the edge strength measure across scale space, and these resulted in spurious peaks, which were interpreted as extra edges.

Chapter 13
Summary and conclusions

13.1 The GDT model

Chapter 2 introduced a new model of edge detection, the Gaussian derivative template (GDT) model. The model consists of a set of channels, selective for orientation and blur. In each channel, the model spatially differentiates the luminance profile, half-wave rectifies the first derivative, and then differentiates again. At each point, the correlation is found between the 2nd derivative and a 2nd-derivative edge template with orientation and blur specific to the channel. Each stage of processing, including the correlation, consists of filtering with a Gaussian 1st derivative operator, a physiologically plausible mechanism. Peaks in the correlation across scale space indicate the position and scale of each edge element in the image. The height of each peak indicates the edge contrast.

Various earlier versions of this model have been presented at conferences (Georgeson, 2001; Georgeson *et al.*, 2002; Barbieri-Hesse & Georgeson, 2002; Georgeson *et al.*, 2003; May & Georgeson, 2003a, 2003b), but this is the first time that the model has been presented in full, with all the equations describing its behaviour. Two new developments are the derivation of an algebraic expression for the normalization factor (section 2.2.5.1) and the derivation of an equation to calculate edge contrast from peak height (equation (2.35), section 2.2.8).

Although the pure form of the GDT model presented in Chapter 2 has been successful in explaining many previous results (Georgeson, 2001; Georgeson *et al.*, 2002; Barbieri-Hesse & Georgeson, 2002), it failed to predict most of the results presented here. However, it has been shown that a simple, physiologically plausible, modification to the half-wave rectifier can account for most of the results: the half-wave rectifier is replaced with a smoothed threshold-linear function.

There are a number of reasons why the visual system might replace the ideal half-wave rectifier with a smoothed threshold-linear function. One possibility is that a threshold is applied to the gradient signal in order to remove noise. A threshold-linear function of the gradient has often been used for this purpose in edge detection algorithms intended for machine vision. The signal at each point in scale space should be carried by more than one neuron: section 7.2.1 shows that, if the neurons have a range of different thresholds, then the summed response will be a smoothed threshold-linear function with exactly the form used in the model. Another possibility is that the smoothed threshold-linear function is *caused by* the noise. Section 7.2.1 shows that adding noise to a half-wave rectifier or threshold-linear function can result in the smoothed threshold-linear function used in the model. It is also plausible that both of these possibilities could apply: the underlying function could be a threshold-linear function, which is smoothed both by having noise and a range of thresholds.

There has recently been some interest in the idea that noise could play a crucial role in determining the characteristics of simple cell responses. Carandini & Ferster (2000) showed that the function relating instantaneous membrane potential to instantaneous spike rate was well described by a threshold-linear function, but this model is unable to account for the contrast invariant tuning reported by Sclar & Freeman (1982) and Skottun *et al.* (1987). Andersen *et al.* (2000) showed that adding noise to a threshold-linear function results in contrast invariant tuning. Miller & Troyer (2002) provided a theoretical understanding of this phenomenon by showing that only a power-function will give contrast-invariant tuning, and that adding Gaussian noise converts a threshold-linear function into an approximate power function for small input values. Previously, Suarez & Koch (1989) had proved that adding a uniform distribution of noise to a threshold-linear function results in a quadratic power-law function. Miller & Troyer argued that this mechanism was not robust because it could only produce an exponent of exactly 2. The work presented in section 7.2.1 shows that Suarez & Koch's mechanism can be extended so that, by manipulating the shape of the noise distribution, any threshold-linear function can be converted into a power function with any exponent, with complete mathematical precision between any arbitrarily specified limits.

13.2 Experimental findings

Experiment 3 showed that Gaussian edges of opposite polarity are perfectly discriminated, i.e. the polarity of an edge can be identified whenever it is detected. The GDT model predicts that this should be the case. Tolhurst & Dealy (1975) had previously found a similar result for rectangular step edges, but this issue had not been examined with blurred edges. The contrast thresholds for edge detection increased slightly with increasing edge scale, but not enough to correspond to a fixed gradient threshold.

Experiments 1, 2 and 4 examined the perception of edge scale and contrast, using standard Gaussian edges with different physical scales and contrasts. It was found that reducing the contrast of an edge reduced the perceived scale, an effect that was roughly constant across scale; increasing the scale of an edge reduced the perceived contrast, especially for low-contrast edges.

Experiment 5 explored the effect of window width on the perceived contrast of a Gaussian edge, using a sharp, rectangular window. The window that varied in width, but its height, along the length of the edge, remained constant. Two effects of window width were found. When the window was so narrow that it encroached upon the signature of the edge, the perceived contrast was reduced. This effect was quite well predicted by the change in Michelson contrast resulting from the truncation of the edge. When the window width was substantially larger than the edge signature, *increasing* the window width led to a reduction in perceived contrast.

Experiments 6 and 7 investigated effect of adding a linear ramp to a Gaussian edge. It was found that adding a ramp of opposite polarity to the edge led to a substantial reduction in the perceived scale and contrast; adding a positive ramp slightly increased the perceived blur and contrast for one subject, KAM, and had almost no effect for the other subject, MAG. The perceived contrast of the edge with a negative ramp was well predicted by the local Michelson contrast. These are important results because they rule out any model of edge perception that estimates edge parameters according to the local 2nd derivative. This includes the two most well-known derivative-based models of edge detection in humans (Marr & Hildreth, 1980; Watt & Morgan, 1985).

The simulations reported in the previous chapters fitted the model to the data for individual experiments. The simulation results were also used to find the best-fitting model across all the experiments. This procedure is described in the next section, which also gives the best-fitting parameters for all the individual simulations.

13.3 Overall best-fitting parameters

After completing all the simulations, the best-fitting model parameters were found across all experiments for each subject. KAM and MAG had completed all the experiments, and the best-fitting parameters were found across all the blur and contrast matching experiments (i.e. all experiments except experiment 3). Of the blur and contrast matching experiments, PAA only completed experiments 1, 2, 4 and 5. PAA's data from experiment 2 were wildly different from those of KAM and MAG, and from PAA's own data from very similar conditions in experiment 1. The fact that PAA's data were very different for the two test orders suggested that these data were contaminated by invisibility of the stimulus on some trials. Because of these problems, the models were fitted only to experiments 1, 4 and 5 for PAA. In section 11.3.1 it was explained that, because the model would be unable to account for the overall decrease in perceived contrast with large window widths, best-fitting parameters were found by fitting the model to the data for the 20′ and 40′ width conditions only. Therefore, only these window widths were included when finding the overall best-fitting parameters.

The best-fitting parameters across all experiments were found by finding the sum of squared differences between model and data across all included conditions of all included experiments. Thus, experiments with more conditions carried more weight. The model and predicted data were expressed in dB before subtraction, as in the individual simulations.

The best-fitting parameters for the two types of model are given in sections 13.3.1 and 13.3.2. Each table has a column headed *SS/N*, which is the total sum of squared differences between model and data, divided by the number of conditions. Taking the square root of the *SS/N* value give the root mean squared (RMS) error. It can be seen

that the RMS error for the overall best-fitting models is between 1 and 2 dB, which is quite acceptable.

For the overall best-fitting models, the best-fitting power-function rectifiers and Legge-Foley transducers give very similar errors. Across all the individual simulations, there was no significant difference between the two types of model. A matched-pairs t-test was performed to compare the two simulations for each experiment and each subject. No significant difference was found.

It is interesting to note that the best-fitting exponent for the power-function rectifier across all experiments was around 2 for all three subjects, agreeing with Heeger's (1991, 1992a, 1992b, 1993) theoretical preference, and also the physiological data of Albrecht & Hamilton (1982), Sclar *et al.* (1990), and Carandini *et al.* (1997).

13.3.1 Best-fitting parameters for the power-function rectifier

Table 13.1. Power-function rectifier parameters for subject KAM.

Simulation	Experiment	T_1	T_2	n	SS/N
5	1	0.00	0.2	5.66	1.92
7	2	0.04	0.6	4.00	2.17
8	2	0.04	0.5	4.00	1.76
10	4	0.00	0.7	1.68	1.10
12	5	0.06	0.4	4.00	1.40
14	6	0.02	0.1	3.36	1.37
16	7	0.00	0.7	1.19	2.85
Overall best fit		0.04	0.3	2.38	3.10

Table 13.2. Power-function rectifier parameters for subject MAG.

Simulation	Experiment	T_1	T_2	n	SS/N
5	1	0.04	0.2	2.83	1.09
7	2	0.02	0.8	4.00	1.92
8	2	0.04	0.8	3.36	1.35
10	4	0.00	0.7	1.68	0.57
12	5	0.00	0.8	1.68	0.60
14	6	0.06	0.4	1.41	0.87
16	7	0.04	0.2	2.38	5.68
Overall best fit		0.04	0.4	2.00	2.38

Table 13.3. Power-function rectifier parameters for subject PAA.

Simulation	Experiment	T_1	T_2	n	SS/N
5	1	0.00	0.1	2.00	0.25
7	2	-	-	-	-
8	2	-	-	-	-
10	4	0.00	0.4	2.00	5.65
12	5	0.00	0.8	1.41	0.63
14	6	-	-	-	-
16	7	-	-	-	-
Overall best fit		0.02	0.5	1.68	1.94

13.3.2 Best-fitting parameters for the Legge-Foley transducer

Table 13.4. Legge-Foley transducer parameters for subject KAM.

Simulation	Experiment	S	p	SS/N
6	1	0.32	4.4	1.94
9	2	0.72	5.2	4.32
11	4	0.72	1.9	1.51
13	5	0.64	3.2	1.54
15	6	0.14	3.8	1.24
17	7	0.80	1.4	2.74
Overall best fit		0.32	3.4	3.32

Table 13.5. Legge-Foley transducer parameters for subject MAG.

Simulation	Experiment	S	p	SS/N
6	1	0.24	4.0	1.25
9	2	0.96	4.0	1.98
11	4	0.64	2.0	0.81
13	5	0.96	1.9	0.32
15	6	0.10	6.0	1.24
17	7	0.16	4.4	5.49
Overall best fit		0.36	3.2	2.43

Table 13.6. Legge-Foley transducer parameters for subject PAA.

Simulation	Experiment	S	p	SS/N
6	1	0.10	2.6	0.23
9	2	-	-	-
11	4	0.48	2.4	5.76
13	5	0.96	1.6	0.30
15	6	-	-	-
17	7	-	-	-
Overall best fit		0.48	2.2	1.91

13.4 The model's performance

The best-fitting models were generally very successful at predicting the data. For each subject, one set of parameters predicted most of the data from all the blur and contrast matching experiments.

There were some exceptions to this success. In experiment 2, the size of the effect of contrast on perceived blur was largely constant across scale. The models were able to account for this to some extent, but were unable to predict such a large effect with the sharpest edge (2′). This edge scale approaches the resolution limit of the human visual system (Levi & Klein, 1990; Wuerger *et al.*, 2001), and it is perhaps not surprising that the theoretical predictions failed with this edge scale.

Experiment 5 showed two effects of window width on perceived contrast. A very narrow window reduced the perceived contrast, due to the truncation of the edge; increasing the width of an already wide window reduced the perceived contrast. The GDT model accounted for the first effect very well (better than the Michelson contrast model), but was unable to account for the second effect.

The GDT model accurately predicted the effects of a negative ramp on perceived blur and contrast, and was fairly successful at predicting the effects of a positive ramp on KAM's performance, although the model that fitted best to all the experiments overestimated the effect of a positive ramp on perceived blur; the effect on perceived contrast was quite well predicted. In contrast to KAM, MAG showed a very small effect of positive ramp on both blur and contrast perception. Even the model that was fitted to MAG's data for the individual experiments (6 and 7) was unable to account for the lack of an effect of a positive ramp for the largest edge scales.

One striking finding was that the contrast reduction, caused both by truncation with a narrow window and addition of a negative ramp, was well predicted by both the Michelson contrast model and the GDT model. This is remarkable because there is no particular reason why the GDT model should give contrast estimates that correspond to the local

Michelson contrast for non-Gaussian edges. In both of these cases, the physical contrast reduction was caused by a manipulation that caused the signal to be truncated: in experiment 5, the stimulus was literally truncated and, in experiment 7, the negative ramp caused the signal to be truncated by the rectifier. This truncation reduced the width of the 2nd derivative signature, so that it fitted a narrower template, and looked sharper. The reduction in perceived edge scale caused the reduction in estimated contrast because, in equation (2.35), which gives the estimated contrast, contrast is an increasing function of edge scale.

13.5 Future extensions to the GDT model

Although the model generally performed very well, it would need to be extended in order to account for all the results. One extension should be an inhibitory or facilitatory mechanism to explain the decrease in perceived contrast with increasing window width, discussed in section 11.5. The problem with such an extension is that, at present, the effects of surround on perceived contrast are very poorly understood, with large differences in results between different subjects and also between slightly different stimulus configurations. It may be the effect of window width on perceived contrast is not an intrinsic part of the edge processing mechanism, but occurs at an earlier stage of processing. Thus, the model itself might not have to be modified – it could just be embedded within a larger model of early visual processing.

One previous result that the model cannot currently explain is Thomas's (1989) finding that a 6 c/deg grating looks lower in contrast when added to a bright low-frequency bar than when added to a dark low-frequency bar. The model does not account for this because the local mean luminance is discarded by the derivative operators. This result could easily be accommodated by introducing a compressive transducer, such as the Naka-Rushton function, at the retinal stage: if the response saturated at high luminance levels, then the peak-to-trough difference in response would decrease with increasing mean luminance. Naka & Rushton (1966) first used this function to describe the responses of retinal cells. Several researchers have reported data on edge locatization that can be

explained by applying the Naka-Rushton function to the luminance image before processing it with an edge coding model. Morgan *et al.* (1984) used this approach to explain the perceived shift in edge location with a change of blur; Georgeson & Freeman (1997) used it to improve the fit of their model to the data from feature marking experiments. It would be interesting to see if this transducer could also improve the model's performance with positive ramps. The positive ramp stimuli had the largest extremes of luminance of any of the stimuli, and a compressive function would have the effect of decreasing the gradient, especially towards the bright end of the ramp. This might reduce the effect of the positive ramp, bringing the simulated results more into line with the empirical data.

Another possible extension might be a gain control process to limit the effect of a positive ramp, as described in section 12.8. Unlike the mechanisms described in the two previous paragraphs, this mechanism would have to be an integral part of the edge processing mechanism, as it must come immediately after the rectifier. It may be that, for studying the processing of the kind of images that occur naturally, such a process may not be necessary, since natural images are unlikely to have steep ramps that extend over a large area – any ramps in natural images are likely to be either very shallow (in the case of illumination gradients) or of limited spatial extent. The fact that the existing model accounts for the positive ramp data reasonably well for one subject (KAM) suggests that this is not a particularly serious issue.

A more substantial extension would be to add another feature type, such as bars. It is possible that bars are not primitive features, and are simply a combination of two edges. However, many models do include bars as primitive features (e.g., Watt & Morgan, 1985; Morrone & Owens, 1987; Morrone & Burr, 1988; Kingdom & Moulden, 1992; Georgeson, 1994; Georgeson & Freeman, 1997), and the GDT model may benefit from having bar templates as well as edge templates. Early explorations with bar templates have suggested that this would be a promising line of enquiry.

In the current model, any pixel that has a higher edge strength than the eight surrounding pixels is marked as a peak, and is interpreted as an edge element. This method is perfectly satisfactory for processing the highly artificial, noiseless stimuli used in the experiments, but it would be completely inadequate for processing natural images using an

implementation that included neural noise: the slightest noise would give rise to perturbations in the edge strength representation, which would be interpreted as a multitude of edges. A more robust solution would be to use some kind of population coding scheme that finds the peaks based on the distribution of activity. It turns out that many common population coding schemes are equivalent to template matching methods (Deneve *et al.*, 1999; Pouget *et al.*, 2000). It might seem that nothing has been achieved if the result of the template matching operation can only be decoded reliably by performing yet another template matching operation. However, the edge strength representation is much more tightly constrained than the original image. If a peak occurs in a particular channel, then the distribution should have a particular width in the spatial and scale domains, determined by the channel bandwidths. The template that would be used to find the peak in each channel would therefore have a fixed width. This would make the peak-finding operation much more robust than the process of finding peaks in an unconstrained image. For example, any high-frequency ripples that occurred in large-scale channels could be attributed to noise, because a genuine peak in that channel would not have a high spatial frequency content. Deneve *et al.* (1999) showed that a biologically plausible network of units can find the smooth peak that fits best to a noisy peak, if the expected width of the peak is known. The model takes the noisy peak as its input and, over time, converges towards the best-fitting smooth peak. This convergence is controlled by lateral connections between units, which have the effect of pooling the activity of nearby units. The width of the smooth template is determined by the spatial extent of these pooling connections. Thus, if, for a given channel, the expected width of the peak in edge strength is known, then the noise can be removed using lateral connections, and peaks in the smoothed output could be marked.

Section 1.8.6.3 discussed Georgeson & Meese's evidence that outputs from spatial filters can be summed across orientation before the edges are found. This summation across orientation channels is a stimulus-dependent process that seems to occur whenever intermediate channels are sufficiently active. This effect was by studied using plaid stimuli with orthogonal components. When the intermediate channel activation was reduced, using several different experimental manipulations, very large changes in the perceived edge structure were reported. These results challenge the current GDT model because, in this model, all the channels are independent. This issue was not addressed by the current

experiments, since all the stimuli were 1D. It remains to be seen whether Georgeson & Meese's results can be accommodated by the current or future versions of the GDT model.

13.6 Conclusions

A purely theoretical model of edge processing was presented, which characterized edges using Gaussian derivative templates. Despite the success of the model in predicting the results of previous experiments, the model failed to accurately predict most of the results presented here. In particular, it could not account for the decrease in perceived blur with decreasing contrast, or the decrease in perceived contrast with increasing blur. The model was modified slightly by replacing a half-wave rectifier with a smoothed threshold-linear function that was described by three parameters. This simple modification allowed the model to predict most of the results using one set of parameters for each subject. The robustness of the model is demonstrated by its ability to accurately predict the perceived contrast and blur of edge profiles that differ markedly from the ideal Gaussian edges on which the templates were based. The model is very simple and does not include gain control processes, interactions across space or spatial frequency, or non-linear luminance transduction. Such complications are known to exist in the visual system, and it is argued that introducing these additional mechanisms might improve the model's fit to the data in the few cases where it cannot explain the results. In the case of non-linear luminance transduction, and interactions across space or spatial frequency, it is argued that these mechanisms could be accommodated by embedding the model within a more comprehensive model of early visual processing, rather than modifying the model itself.

References

Abdou, I. E. & Pratt, W. K. (1979). Quantitative design and evaluation of enhancement/thresholding edge detectors. *Proceedings of the IEEE*, **67**, 753-763.

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

Akutsu, H. & Legge, G. E. (1995). Discrimination of compound gratings: spatial-frequency channels or local features? *Vision Research*, 35, 2685-2695.

Albrecht, D. G., Farrar, S. B., & Hamilton, D. B. (1984). Spatial contrast adaptation characteristics of neurones recorded in the cat's visual cortex. *Journal of Physiology*, **347**, 713-739.

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

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

Anderson, J. S., Lampl, I., Gillespie, D. C., & Ferster, D. (2000). The contribution of noise to contrast invariance of orientation tuning in cat visual cortex. *Science*, **290**, 1968-1972.

Andrews, B. W. & Pollen, D. A. (1979). Relationship between spatial frequency selectivity and receptive field profile of simple cells. *Journal of Physiology*, **287**, 163-176.

Argyle, E. (1971). Techniques for edge detection. *Proceedings of the IEEE (Letters)*, 59, 285-286.

Argyle, E. (1972). Comments on "Techniques for edge detection" - reply. *Proceedings of the IEEE (Letters)*, 60, 344.

Atkinson, J. & Campbell, F. W. (1974). The effect of phase on the perception of compound gratings. *Vision Research*, **14**, 159-162.

Babaud, J., Witkin, A. P., Baudin, M., & Duda, R. O. (1986). Uniqueness of the Gaussian kernel for scale-space filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8**, 26-33.

Badcock, D. R. (1984). Spatial phase or luminance profile discrimination? *Vision Research*, **24**, 613-623.

Barbieri-Hesse, G. S. A. & Georgeson, M. A. (2002). Template model for blur coding: the role of early nonlinearity in edge segmentation. *Perception*, **31, Supplement**, 54.

Barbieri-Hesse, G. S. A. & Georgeson, M. A. (2003). Edges and bars: where do people see features in 1-D images? *submitted.*

Belhumeur, P. N. & Kriegman, D. J. (1998). What is the set of images of an object under all possible illumination conditions? *International Journal of Computer Vision*, **28**, 245-260.

Ben-Arie, J. & Rao, K. R. (1993a). A novel approach for template matching by nonorthogonal image expansion. *IEEE Transactions on Circuits and Systems for Video Technology*, **3**, 71-84.

Ben-Arie, J. and Rao, K. R. (1993b). Image expansion by non-orthogonal basis functions extended for optimal multiple template matching. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 145-148). Minneapolis, MN.

Bergholm, F. (1987). Edge focusing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **9**, 726-741.

Berzins, V. (1984). Accuracy of Laplacian edge detectors. *Computer Vision, Graphics, and Image Processing*, **27**, 195-210.

Biederman, I. & Ju, G. (1988). Surface versus edge-based determinations of visual recognition. *Cognitive Psychology*, **20**, 38-64.

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

Boie, R. A. and Cox, I. J. (1987). Two dimensional optimum edge recognition using matched and Wiener filters for machine vision. In *Proceedings of the International Conference on Computer Vision* (pp. 450-456).

Boie, R. A., Cox, I. J., and Rehak, P. (1986). On optimum edge recognition using matched filters. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 100-108).

Boyer, K. L. & Sarkar, S. (1994). On the localization performance measure and optimal edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16, 106-110.

Burr, D. C. & Morrone, M. C. (1990). Feature detection in biological and artificial visual systems. In Blakemore, C. (Ed), *Vision: coding and efficiency*, (pp. 185-194). Cambridge University Press: Cambridge, U.K.

Burr, D. C. & Morrone, M. C. (1992). A nonlinear model of feature detection. In Pinter, R. B. & Nabet, B. (Eds), *Non-linear vision*, (pp. 309-327). CRC Press.

Burr, D. C. & Morrone, M. C. (1994). The role of features in structuring visual images. *Higher-Order Processing in the Visual System*, **184**, 129-141.

Burr, D. C., Morrone, M. C., & Spinelli, D. (1989). Evidence for edge and bar detectors in human vision. *Vision Research*, **29**, 419-431.

Campbell, F. W. & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, **197**, 551-566.

Cannon, M. W. (1979). Contrast sensation: a linear function of stimulus contrast. *Vision Research*, **19**, 1045-1052.

Cannon, M. W. & Fullenkamp, S. C. (1988). Perceived contrast and stimulus size: experiment and simulation. *Vision Research*, **28**, 695-709.

Cannon, M. W. & Fullenkamp, S. C. (1991a). A transducer model for contrast perception. *Vision Research*, 31, 983-998.

Cannon, M. W. & Fullenkamp, S. C. (1991b). Spatial interactions in apparent contrast: inhibitory effects among grating patterns of different spatial frequencies, spatial positions and orientations. *Vision Research*, 31, 1985-1998.

Cannon, M. W. & Fullenkamp, S. C. (1993). Spatial interactions in apparent contrast: individual differences in enhancement and suppression effects. *Vision Research*, 33, 1685-1695.

Cannon, M. W. & Fullenkamp, S. C. (1996). A model for inhibitory lateral interaction effects in perceived contrast. *Vision Research*, **36**, 1115-1125.

Canny, J. F. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8**, 679-698.

Carandini, M. & Ferster, D. (2000). Membrane potential and firing rate in cat primary visual cortex. *Journal of Neuroscience*, **20**, 470-484.

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

Carandini, M., Heeger, D. J., & Senn, W. (2002). A synaptic explanation of suppression in visual cortex. *Journal of Neuroscience*, **15**, 10053-10065.

Carlson, A. B., Crilly, P. B., & Rutledge, J. C. (2002). *Communication systems: an introduction to signals and noise in electrical communication.*, fourth edition ed. McGraw-Hill, New York.

Chen, J. S. & Medioni, G. (1989). Detection, localization, and estimation of edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 191-198.

Chubb, C., Sperling, G., & Solomon, J. A. (1989). Texture interactions determine perceived contrast. *Proceedings of the National Academy of Sciences of the United States of America*.

Clark, J. J. (1988). Singularity theory and phantom edges in scale space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **10**, 720-727.

Clark, J. J. (1989). Authenticating edges produced by zero-crossing algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 43-57.

Davies, E. R. (1984). Circularity - a new principle underlying the design of accurate edge orientation operators. *Image and Vision Computing*, **2**, 134-142.

De Micheli, E., Caprile, B., Ottonello, P., & Torre, V. (1989). Localization and noise in edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**, 1106-1117.

Dean, A. F. (1981). The relationship between response amplitude and contrast for cat striate cortical neurones. *Journal of Physiology*, **318**, 413-427.

Dean, A. F. & Tolhurst, D. J. (1983). On the distinctness of simple and complex cells in the visual cortex of the cat. *Journal of Physiology*, **344**, 305-325.

Deneve, S., Latham, P. E., & Pouget, A. (1999). Reading population codes: a neural implementation of ideal observers. *Nature Neuroscience*, 2, 740-745.

Derrington, A. M. & Fuchs, A. F. (1979). Spatial and temporal properties of X and Y cells in the cat's lateral geniculate nucleus. *Journal of Physiology*, **293**, 347-364.

Dreher, B. & Sanderson, K. J. (1973). Receptive field analysis: responses to moving visual contours by single lateral geniculate nucleus neurons in the cat. *Journal of Physiology*, **234**, 95-118.

Duda, R. O. & Hart, P. E. (1973). *Pattern classification and scene analysis.* Wiley, New York.

Ejima, Y. & Takahashi, S. (1985). Apparent contrast of a sinusoidal grating in the simultaneous presence of peripheral gratings. *Vision Research*, **25**, 1223-1232.

Elder, J. H. (1999). Are edges incomplete? *International Journal of Computer Vision*, **34**, 97-122.

Elder, J. H. and Zucker, S. W. (1996). Scale space localization, blur and contour-based image coding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 27-34).

Elder, J. H. & Zucker, S. W. (1998). Local scale control for edge detection and blur estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**, 699-716.

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

Famiglietti, E. V. & Kolb, H. (1976). Structural basis for on- and off-center responses in retinal ganglion cells. *Science*, **194**, 193-195.

Field, D. J. & Tolhurst, D. J. (1986). The structure and symmetry of simple-cell receptive-field profiles in the cat's visual cortex. *Proceedings of the Royal Society of London B*, **228**, 379-400.

Fiorentini, A., Baumgartner, G., Magnusson, S., Schiller, P. H., & Thomas, J. P. (1990). The perception of brightness and darkness. In Spillman, L. & Werner, J. S. (Eds), *Visual perception: the neurophysiological foundations*, Academic Press: San Diego, California.

Fleck, M. M. (1992a). Multiple widths yield reliable finite differences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**, 412-429.

Fleck, M. M. (1992b). Some defects in finite-difference edge finders. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**, 337-345.

Fram, J. R. & Deutsch, E. S. (1975). On the quantitative evaluation of edge detection schemes and their comparison with human performance. *IEEE Transactions on Computers*, **24**, 616-628.

Freeman, T. C. B., Durand, S., Kiper, D. C., & Carandini, M. (2002). Suppression without inhibition in visual cortex. *Neuron*, **35**, 759-771.

Frei, W. & Chen, C. C. (1977). Fast boundary detection: a generalization and a new algorithm. *IEEE Transactions on Computers*, **26**, 988-998.

Georgeson, M. A. (1984). Eye movements, afterimages and monocular rivalry. *Vision Research*, **24**, 1311-1319.

Georgeson, M. A. (1991a). Contrast overconstancy. *Journal of the Optical Society of America, A*, 8, 579-586.

Georgeson, M. A. (1991b). Over the limit: encoding contrast above threshold in human vision. In Kulikowski, J. J., Walsh, V., & Murray, I. J. (Eds), *Vision and visual dysfunction, Vol. 5: Limits of visual perception*, (pp. 106-119).

Georgeson, M. A. (1992). Human vision combines oriented filters to compute edges. *Proceedings of the Royal Society of London B*, **249**, 235-245.

Georgeson, M. A. (1994). From filters to features: location, orientation, contrast and blur. *Higher-Order Processing in the Visual System*, **184**, 147-165.

Georgeson, M. A. (1998). Edge-finding in human vision: a multi-stage model based on the perceived structure of plaids. *Image and Vision Computing*, **16**, 389-405.

Georgeson, M. A. (2001). Seeing edge blur: receptive fields as multi-scale neural templates. *Journal of Vision*, **1**, 438a.

Georgeson, M. A., Barbieri-Hesse, G. S. A., & Freeman, T. C. A. (2002). The primal sketch revisited: locating and representing edges in human vision via Gaussian-derivative filtering. *Perception*, **31, Supplement**, 1.

Georgeson, M. A. & Freeman, T. C. A. (1994). A short sharp shock: perceived edge blur in two-dimensional images. *Perception*, **24, Supplement**, 76.

Georgeson, M. A. & Freeman, T. C. A. (1997). Perceived location of bars and edges in one-dimensional images: computational models and human vision. *Vision Research*, **37**, 127-142.

Georgeson, M. A., May, K. A., & Barbieri-Hesse, G. S. A. (2003). Perceiving edge blur: the Gaussian-derivative template model. *Journal of Vision*, **3**, 360a.

Georgeson, M. A. & Meese, T. S. (1997). Perception of stationary plaids: the role of spatial filters in edge analysis. *Vision Research*, **37**, 3255-3271.

Georgeson, M. A. & Meese, T. S. (1999). Adaptive filtering in spatial vision: evidence from feature marking in plaids. *Perception*, **28**, 687-702.

Georgeson, M. A. & Shackleton, T. M. (1994). Perceived contrast of gratings and plaids: non-linear summation across oriented filters. *Vision Research*, **34**, 1061-1075.

Georgeson, M. A. & Sullivan, G. D. (1975). Contrast constancy: deblurring in human vision by spatial frequency channels. *Journal of Physiology*, **252**, 627-656.

Georgeson, M. A. & Turner, R. S. E. (1985). Afterimages of sinusoidal, square-wave and compound gratings. *Vision Research*, **25**, 1709-1720.

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

Hadamard, J. (1923). *Lectures on the Cauchy problem in linear partial differential equations*. Yale University Press, New Haven.

Hamerly, J. R., Quick, R. F., & Reichert, T. A. (1977). A study of grating contrast judgement. *Vision Research*, 17, 201-207.

Haralick, R. M. (1980). Edge and region analysis for digital image data. *Computer Graphics and Image Processing*, 12, 60-73.

Haralick, R. M. (1984). Digital step edges from zero crossings of 2nd directional derivatives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 58-68.

Haralick, R. M. & Watson, L. (1981). A facet model for image data. *Computer Graphics and Image Processing*, 15, 113-129.

Heath, M., Sarkar, S., Sanocki, T., & Bowyer, K. (1998). Comparison of edge detectors: A methodology and initial study. *Computer Vision and Image Understanding*, 69, 38-54.

Heath, M. D., Sarkar, S., Sanocki, T., & Bowyer, K. W. (1997). A robust visual method for assessing the relative performance of edge-detection algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19, 1338-1359.

Heeger, D. J. (1991). Nonlinear model of neural responses in cat visual cortex. In Landy, M. & Movshon, J. A. (Eds), *Computational models of visual processing*, (pp. 119-133). MIT Press: Cambridge, MA.

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

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

Heeger, D. J. (1993). Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. *Journal of Neurophysiology*, 70, 1885-1898.

Hess, R. F. & Pointer, J. S. (1987). Evidence for spatially local computations underlying discrimination of periodic patterns in fovea and periphery. *Vision Research*, 27, 1343-1360.

Hess, R. F., Pointer, J. S., & Watt, R. J. (1989). How are spatial filters used in fovea and parafovea? *Journal of the Optical Society of America, A*, 6, 329-339.

Hildreth, E. (1983). The detection of intensity changes by computer and biological vision systems. *Computer Vision, Graphics, and Image Processing*, **22**, 1-27.

Hochstein, S. & Shapley, R. M. (1976a). Linear and nonlinear spatial subunits in Y cat retinal ganglion cells. *Journal of Physiology*, **262**, 265-284.

Hochstein, S. & Shapley, R. M. (1976b). Quantitative analysis of retinal ganglion cell classifications. *Journal of Physiology*, **262**, 237-264.

Hoekstra, J., van der Goot, D. P. J., van den Brink, G., & Bilsen, F. A. (1974). The influence of number of cycles upon the visual contrast threshold for spatial sine patterns. *Vision Research*, **14**, 365-368.

Howell, E. R. & Hess, R. F. (1978). The functional area for summation to threshold for sinusoidal gratings. *Vision Research*, **18**, 369-374.

Hubel, D. H. (1960). Single unit activity in lateral geniculate body and optic tract of unrestrained cats. *Journal of Physiology*, **150**, 91-104.

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

Hubel, D. H. & Wiesel, T. N. (1960). Receptive fields of optic nerve fibres in the spider monkey. *Journal of Physiology*, **154**, 572-580.

Hubel, D. H. & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, **160**, 106-154.

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

Hueckel, M. H. (1971). An operator which locates edges in digitized pictures. *Journal of the Association for Computing Machinery*, **18**, 113-125.

Hueckel, M. H. (1973). A local visual operator which recognizes edges and lines. *Journal of the Association for Computing Machinery*, **20**, 634-647.

Hueckel, M. H. (1974). Erratum for "A local visual operator which recognizes edges and lines". *Journal of the Association for Computing Machinery*, **21**, 350.

Huertas, A. & Medioni, G. (1986). Detection of intensity changes with subpixel accuracy using Laplacian-Gaussian masks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8**, 651-664.

Hummel, R. & Moniot, R. (1989). Reconstructions from zero crossings in scale space. *IEEE Transactions on Acoustics Speech and Signal Processing*, **37**, 2111-2130.

Hurlbert, A. (1986). Formal connections between lightness algorithms. *Journal of the Optical Society of America, A*, **3**, 1684-1693.

Ikeda, H. & Wright, M. J. (1974). Sensitivity of neurones in visual cortex (area 17) under different levels of anaesthesia. *Experimental Brain Research*, **20**, 471-484.

Jagadeesh, B., Wheat, H. S., & Ferster, D. (1993). Linearity of summation of synaptic potentials underlying direction selectivity in simple cells of the cat visual cortex. *Science*, **262**, 1901-1904.

Jeong, H. & Kim, C. I. (1992). Adaptive determination of filter scales for edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**, 579-585.

Jones, J. P. & Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, **58**, 1233-1258.

Kingdom, F. & Moulden, B. (1992). A Multi-channel approach to brightness coding. *Vision Research*, **32**, 1565-1582.

Kirsch, R. A. (1971). Computer determination of the constituent structure of biological images. *Computers and Biomedical Research*, **4**, 315-328.

Kirsch, R. A., Cahn, L., Ray, C., and Urban, G. H. (1957). Experiments in processing pictorial information with a digital computer. In *Proceedings of the Eastern Joint Computer Conference* (pp. 221-229). Washington, D.C.

Kisworo, M., Venkatesh, S., & West, G. (1994). Modeling edges at subpixel accuracy using the local energy approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**, 405-410.

Kitchen, L. & Rosenfeld, A. (1981). Edge evaluation using local edge coherence. *IEEE Transactions on Systems Man and Cybernetics*, **11**, 597-605.

Kittler, J. (1983). On the accuracy of the Sobel edge detector. *Image and Vision Computing*, 1, 37-42.

Kittler, J., Illingworth, J., & Paler, K. (1983). The magnitude accuracy of the template edge detector. *Pattern Recognition*, 16, 607-613.

Koenderink, J. J. (1984). The structure of images. *Biological Cybernetics*, 50, 363-370.

Koplowitz, J. & Greco, V. (1994). On the edge location error for local maximum and zero-crossing edge detectors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16, 1207-1212.

Korn, A. F. (1988). Toward a symbolic representation of intensity changes in images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10, 610-625.

Kovasznay, L. S. G. & Joseph, H. M. (1953). Processing of two-dimensional patterns by scanning techniques. *Science*, 118, 475-477.

Kovasznay, L. S. G. & Joseph, H. M. (1955). Image processing. *Proceedings of the IRE*, 43, 560-570.

Kovesi, P. (2000). Phase congruency: a low-level image invariant. *Psychological Research*, 64, 136-148.

Kovesi, P. (2002). Edges are not just steps. In *Proceedings of the fifth Asian conference on computer vision, Melbourne, January 22-25, 2002* (pp. 822-827).

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

Kulikowski, J. J. (1976). Effective contrast constancy and linearity of contrast sensation. *Vision Research*, 16, 1419-1431.

Kulikowski, J. J. & Bishop, P. O. (1981). Linear analysis of the responses of simple cells in the cat visual cortex. *Experimental Brain Research*, 44, 386-400.

Lacroix, V. (1990). Edge detection: what about rotation invariance? *Pattern Recognition Letters*, 11, 797-802.

Langley, K. (2002). A parametric account of contrast adaptation on contrast perception. *Spatial Vision*, **16**, 77-93.

Lee, D., Wasilkowski, G. W., & Mehrotra, R. (1993). A new zero-crossing-based discontinuity detector. *IEEE Transactions on Image Processing*, **2**, 265-268.

Legge, G. E. (1981). A power law for contrast discrimination. *Vision Research*, **21**, 457-467.

Legge, G. E. & Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America*, **70**, 1458-1471.

Leipnik, R. (1960). The extended entropy uncertainty principle. *Information and control*, **3**, 18-25.

Levi, D. M. & Klein, S. A. (1990). Equivalent intrinsic blur in spatial vision. *Vision Research*, **30**, 1971-1993.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the acoustical society of America*, **49**, 467-477.

Lindeberg, T. (1990). Scale-space for discrete signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**, 234-254.

Lindeberg, T. (1993a). Discrete derivative approximations with scale-space properties: A basis for low-level feature detection. *Journal of Mathematical Imaging and Vision*, **3**, 349-376.

Lindeberg, T. (1993b). On scale selection for differential operators. In *Proceedings of the 8th Scandinavian Conference on Image Analysis* (pp. 857-866). Tromsø, Norway.

Lindeberg, T. (1994). *Scale-space theory in computer vision.* Kluwer Academic Publishers, Dordrecht, Netherlands.

Lindeberg, T. (1996). Edge detection and ridge detection with automatic scale selection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 465-470). San Francisco.

Lindeberg, T. (1997). On the axiomatic foundations of linear scale-space: combining semi-group structure with causality vs. scale invariance. In Sporring, J., Nielsen, M.,

Florack, L., & Johansen, P. (Eds), *Gaussian Scale-Space Theory*, (pp. 75-98). Kluwer Academic Publishers: Dordrecht, Netherlands.

Lindeberg, T. (1998). Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision*, **30**, 117-154.

Logan, B. F. (1977). Information in the zero-crossings of bandpass signals. *Bell Systems Technical Journal*, **56**, 487-510.

Lunscher, W. H. H. J. & Beddoes, M. P. (1986). Optimal edge detector evaluation. *IEEE Transactions on Systems Man and Cybernetics*, **16**, 304-312.

Lyvers, E. P. & Mitchell, O. R. (1988). Precision edge contrast and orientation estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **10**, 927-937.

Macleod, I. D. G. (1970). On finding structure in pictures. In Kaneff, S. (Ed), *Picture Language Machines*, (pp. 231-256). New York: Academic.

Macleod, I. D. G. (1972). Comments on "Techniques for edge detection". *Proceedings of the IEEE (Letters)*, **60**, 344.

Maffei, L. & Fiorentini, A. (1973). Visual cortex as a spatial frequency analyser. *Vision Research*, **13**, 1255-1268.

Maffei, L., Fiorentini, A., & Bisti, S. (1973). Neural correlate of perceptual adaptation to gratings. *Science*, **182**, 1036-1038.

Maffei, L., Morrone, C., Pirchio, M., & Sandini, G. (1979). Responses of visual cortical cells to periodic and nonperiodic stimuli. *Journal of Physiology*, **296**, 27-47.

Mallat, S. & Zhong, S. (1992). Characterization of signals from multiscale edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**, 710-732.

Marimont, D. H. and Rubner, Y. (1998). A probabilistic framework for edge detection and scale selection. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 207-214). Bombay, India.

Marr, D. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society of London B*, **275**, 483-519.

Marr, D. (1982). *Vision.* WH Freeman, New York.

Marr, D. & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London B*, **207**, 187-217.

Marr, D., Poggio, T., & Ullman, S. (1979). Bandpass channels, zero-crossings, and early visual information processing. *Journal of the Optical Society of America*, **69**, 914-916.

Marr, D. & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London B*, **211**, 151-180.

Marshall, J. A., Burbeck, C. A., Ariely, D., Rolland, J. P., & Martin, K. E. (1996). Occlusion edge blur: a cue to relative visual depth. *Journal of the Optical Society of America, A*, **13**, 681-688.

Mather, G. (1996). Image blur as a pictorial depth cue. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **263**, 169-172.

Mather, G. (1997). The use of image blur as a depth cue. *Perception*, **26**, 1147-1158.

Mather, G. & Smith, D. R. R. (2002). Blur discrimination and its relation to blur-mediated depth perception. *Perception*, **31**, 1211-1219.

May, K. A. & Georgeson, M. A. (2003b). Perceiving edge blur: linear filtering and a rectifying nonlinearity. *Perception*, **32**, 388.

May, K. A. & Georgeson, M. A. (2003a). Perceiving edge blur: linear filtering and a rectifying nonlinearity. *Perception*, **32**, Supplement, 46.

McCann, J. J., Savoy, R. L., & Hall, J. A. (1978). Visibility of low-frequency sine-wave targets: dependence on number of cycles and surround parameters. *Vision Research*, **18**, 891-894.

Meese, T. S. (1999). A model of human pattern perception: association fields for adaptive spatial filters. *Spatial Vision*, **12**, 363-394.

Meese, T. S. & Freeman, T. C. A. (1995). Edge computation in human vision: anisotropy in the combining of oriented filters. *Perception*, **24**, 603-622.

Meese, T. S. & Georgeson, M. A. (1996a). Spatial filter combination in human pattern vision: channel interactions revealed by adaptation. *Perception*, **25**, 255-277.

Meese, T. S. & Georgeson, M. A. (1996b). The tilt aftereffect in plaids and gratings: channel codes, local signs and "patchwise" transforms. *Vision Research*, **36**, 1421-1437.

Meese, T. S. & Holmes, D. J. (2002). Adaptation and gain pool summation: alternative models and masking data. *Vision Research*, **42**, 1113-1125.

Michelson, A. A. (1927). *Studies in optics.* University of Chicago Press, Chicago, Illanois.

Miller, K. D. & Troyer, T. W. (2002). Neural noise can explain expansive, power-law nonlinearities in neural response functions. *Journal of Neurophysiology*, **87**, 653-659.

Morgan, M. J., Mather, G., Moulden, B., & Watt, R. J. (1984). Intensity-response nonlinearities and the theory of edge localization. *Vision Research*, **24**, 713-719.

Morrone, M. C. & Burr, D. C. (1988). Feature detection in human vision: a phase-dependent energy model. *Proceedings of the Royal Society of London B*, **235**, 221-245.

Morrone, M. C. & Burr, D. C. (1993). A model of human feature detection based on matched filters. In Dario, P., Sandini, G., & Aebischer, P. (Eds), *Robots and biological systems: towards a new bionics?*, Springer-Verlag: Berlin.

Morrone, M. C., Navangione, A., & Burr, D. (1995). An adaptive approach to scale selection for line and edge detection. *Pattern Recognition Letters*, **16**, 667-677.

Morrone, M. C. & Owens, R. A. (1987). Feature detection from local energy. *Pattern Recognition Letters*, 6, 303-313.

Morrone, M. C., Ross, J., Burr, D. C., & Owens, R. (1986). Mach bands are phase dependent. *Nature*, **324**, 250-253.

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

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978b). Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. *Journal of Physiology*, **283**, 101-120.

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

Nachmias, J. & Weber, A. (1975). Discrimination of simple and complex gratings. *Vision Research*, **15**, 217-223.

Naka, K. I. & Rushton, W. A. H. (1966). S-potentials from colour units in the retina of fish (cyprinidae). *Journal of Physiology*, **185**, 536-555.

Nalwa, V. S. & Binford, T. O. (1986). On detecting edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8, 699-714.

Nelson, R., Famiglietti, E. V., & Kolb, H. (1978). Intracellular staining reveals different levels of stratification for on- and off-center ganglion cells in cat retina. *Journal of Neurophysiology*, **41**, 472-483.

Nevatia, R. & Babu, K. R. (1980). Linear feature extraction and description. *Computer Graphics and Image Processing*, **13**, 257-269.

Ohzawa, I., Sclar, G., & Freeman, R. D. (1985). Contrast gain control in the cat's visual system. *Journal of Neurophysiology*, **54**, 651-667.

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

Olzak, L. A. & Thomas, J. P. (1992). Configural effects constrain Fourier models of pattern discrimination. *Vision Research*, **32**, 1885-1898.

Olzak, L. A. & Thomas, J. P. (1999). Neural recoding in human pattern vision: model and mechanisms. *Vision Research*, **39**, 231-256.

Owens, R., Venkatesh, S., & Ross, J. (1989). Edge detection is a projection. *Pattern Recognition Letters*, 9, 233-244.

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

Peli, E. (1990). Contrast in complex images. *Journal of the Optical Society of America, A*, 7, 2032-2040.

Peli, E. (2002). Feature detection algorithm based on a visual system model. *Proceedings of the IEEE*, **90**, 78-93.

Perona, P. & Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**, 629-639.

Poggio, T. A. and Torre, V. (1984). Ill-posed problems and regularization analysis in early vision. MIT AI Memo 773.

Poggio, T. A., Voorhees, H., and Yuille, A. (1985). A regularized solution to edge detection. MIT AI Memo 833.

Poggio, T. A., Voorhees, H., & Yuille, A. (1988). A regularized solution to edge detection. *Journal of Complexity*, 4, 123.

Pollen, D. A. & Ronner, S. F. (1981). Phase relationships between adjacent simple cells in the visual cortex. *Science*, **212**, 1409-1411.

Pouget, A., Dayan, P., & Zemel, R. (2000). Information processing with population codes. *Nature Reviews Neuroscience*, **1**, 125-132.

Prewitt, J. (1970). Object enhancement and extraction. In Lipkin, B. & Rosenfeld, A. (Eds), *Picture Processing and Psychopictorics*, (pp. 75-149). New York: Academic.

Quinn, P. C. & Lehmkuhle, S. (1983). An oblique effect of spatial summation. *Vision Research*, **23**, 655-658.

Rao, K. R. and Ben-Arie, J. (1993). Edge detection and feature extraction by non-orthogonal image expansion for optimal discriminative SNR. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 791-798). New York.

Rao, K. R. & Ben-Arie, J. (1994). Optimal edge detection using expansion matching and restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**, 1169-1182.

Ratliff, F. (1965). *Mach bands*. Holden-Day, San Fransisco.

Reid, R. C., Soodak, R. E., & Shapley, R. M. (1987). Linear mechanisms of directional selectivity in simple cells of cat striate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, **84**, 8740-8744.

Reid, R. C., Soodak, R. E., & Shapley, R. M. (1991). Directional selectivity and spatiotemporal structure of receptive fields of simple cells in cat striate cortex. *Journal of Neurophysiology*, **66**, 505-529.

Richter, J. & Ullman, S. (1986). Non-linearities in cortical simple cells and the possible detection of zero-crossings. *Biological Cybernetics*, **53**, 195-202.

Roberts, L. G. (1965). Machine perception of three-dimensional solids. In Tippett, J. T., Berkowitz, D. A., Clapp, L. C., Koester, C. J., & Vanderburgh, A. (Eds), *Optical and electro-optical information processing*, (pp. 159-197). Cambridge, MA: MIT Press.

Robinson, G. S. (1977). Edge detection by compass gradient masks. *Computer Graphics and Image Processing*, **6**, 492-501.

Robson, J. G. (1983). Frequency domain visual processing. In Braddick, O. J. & Sleigh, A. C. (Eds), *Physical and biological processing of images*, (pp. 73-87). Springer-Verlag: Berlin.

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

Robson, J. G., Tolhurst, D. J., Freeman, R. D., & Ohzawa, I. (1988). Simple cells in the visual cortex of the cat can be narrowly tuned for spatial frequency. *Visual Neuroscience*, 1, 415-419.

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

Rodieck, R. W. & Stone, J. (1965). Analysis of receptive fields of cat retinal ganglion cells. *Journal of Neurophysiology*, **28**, 833-849.

Rosenfeld, A. (1970). A nonlinear edge detection technique. *Proceedings of the IEEE (Letters)*, **58**, 814-816.

Rosenfeld, A. (1971). Techniques for edge detection - reply. *Proceedings of the IEEE (Letters)*, 59, 286-287.

Rosenfeld, A. & Thurston, M. (1971). Edge and curve detection for visual scene analysis. *IEEE Transactions on Computers*, **20**, 562-569.

Rosenfeld, A., Thurston, M., & Lee, Y. H. (1972). Edge and curve detection: further experiments. *IEEE Transactions on Computers*, **21**, 677-715.

Rosin, P. L. (1997). Edges: saliency measures and automatic thresholding. *Machine Vision and Applications*, **9**, 139-159.

Ross, J., Morrone, M. C., & Burr, D. C. (1989). The conditions under which Mach bands are visible. *Vision Research*, **29**, 699-715.

Sanchez-Vives, M. V., Nowak, L. G., & McCormick, D. A. (2000). Membrane mechanisms underlying contrast adaptation in cat area 17 *in vivo*. *Journal of Neuroscience*, **20**, 4267-4285.

Sanocki, T., Bowyer, K. W., Heath, M. D., & Sarkar, S. (1998). Are edges sufficient for object recognition? *Journal of Experimental Psychology-Human Perception and Performance*, **24**, 340-349.

Sarkar, S. & Boyer, K. L. (1991). On optimal infinite impulse response edge detection filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13**, 1154-1171.

Savoy, R. L. & McCann, J. J. (1975). Visibility of low-spatial-frequency sine-wave targets: dependence on number of cycles. *Journal of the Optical Society of America, A*, **65**, 343-350.

Schachter, B. J. & Rosenfeld, A. (1978). Some new methods of detecting step edges in digital pictures. *Communications of the ACM*, **21**, 172-176.

Schiller, P. H. (1984). The connections of the retinal on and off pathways to the lateral geniculate nucleus of the monkey. *Vision Research*, **24**, 923-932.

Schumer, R. A. & Movshon, J. A. (1984). Length summation in simple cells of cat striate cortex. *Vision Research*, **24**, 565-571.

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

Sclar, G., Maunsell, J. H. R., & Lennie, P. (1990). Coding of image contrast in central visual pathways of the macaque monkey. *Vision Research*, **30**, 1-10.

Shah, M., Sood, A., & Jain, R. (1986). Pulse and staircase edge models. *Computer Vision, Graphics, and Image Processing*, **34**, 321-343.

Shapley, R. M. & Hochstein, S. (1975). Visual spatial summation in two classes of geniculate cells. *Nature*, **256**, 411-413.

Shou, T., Li, X., Zhou, Y., & Hu, B. (1996). Adaptation of visually evoked responses of relay cells in the dorsal lateral geniculate nucleus of the cat following prolonged exposure to drifting gratings. *Visual Neuroscience*, **13**, 605-613.

Sjöberg, F. & Bergholm, F. (1988). Extraction of diffuse edges by edge focusing. *Pattern Recognition Letters*, **7**, 181-190.

Skottun, B. C., Bradley, A., Sclar, G., Ohzawa, I., & Freeman, R. D. (1987). The effects of contrast on visual orientation and spatial frequency discrimination: a comparison of single cells and behaviour. *Journal of Neurophysiology*, **57**, 773-786.

Spiegel, M. R. (1968). *Mathematical handbook of formulas and tables.* McGraw-Hill, New York.

Suarez, H. & Koch, C. (1989). Linking linear threshold units with quadratic models of motion perception. *Neural Computation*, **1**, 318-320.

Swets, J. A. (1961). Is there a sensory threshold? *Science*, **134**, 168-177.

Tadmor, Y. & Tolhurst, D. J. (1989). The effect of threshold on the relationship between the receptive-field profile and the spatial-frequency tuning curve in simple cells of the cat's striate cortex. *Visual Neuroscience*, **3**, 445-454.

Tagare, H. D. & deFigueiredo, R. J. P. (1990). On the localization performance measure and optimal edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**, 1186-1190.

Tagare, H. D. & deFigueiredo, R. J. P. (1994). On the localization performance measure and optimal edge detection - reply. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**, 108-110.

Takahashi, S. & Ejima, Y. (1984). Dependence of apparent contrast for a sinusoidal grating on stimulus size. *Journal of the Optical Society of America, A*, **1**, 1197-1201.

Thomas, J. P. (1989). Independent processing of suprathreshold spatial gratings as a function of their separation in spatial frequency. *Journal of the Optical Society of America, A*, **6**, 1102-1111.

Tikhonov, A. N. & Arsenin, V. Y. (1977). *Solution of Ill-posed problems.* Winston & Sons, Washington, D.C.

Tolhurst, D. J. & Dealy, R. S. (1975). The detection and identification of lines and edges. *Vision Research*, **15**, 1367-1372.

Tolhurst, D. J. & Dean, A. F. (1987). Spatial summation by simple cells in the striate cortex of the cat. *Experimental Brain Research*, **66**, 607-620.

Tolhurst, D. J. & Dean, A. F. (1991). Evaluation of a linear model of directional selectivity in simple cells of the cat's striate cortex. *Visual Neuroscience*, **6**, 421-428.

Tolhurst, D. J. & Heeger, D. J. (1997). Comparison of contrast-normalization and threshold models of the responses of simple cells in cat striate cortex. *Visual Neuroscience*, **14**, 293-309.

Tolhurst, D. J., Movshon, J. A., & Thompson, I. D. (1981). The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Experimental Brain Research*, **41**, 414-419.

Tootle, J. S. & Berkley, M. A. (1983). Contrast sensitivity for vertically and obliquely oriented gratings as a function of grating area. *Vision Research*, **23**, 907-910.

Torre, V. & Poggio, T. A. (1986). On edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8**, 147-163.

Troy, J. B. (1983). Spatio-temporal interactions in neurones of the cat's dorsal lateral geniculate nucleus. *Journal of Physiology*, **344**, 419-432.

Ulupinar, F. and Medioni, G. (1988). Refining edges detected by a LoG operator. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 202-207).

Ulupinar, F. & Medioni, G. (1990). Refining edges detected by a LoG operator. *Computer Vision, Graphics, and Image Processing*, **51**, 275-298.

Venkatesh, S. & Owens, R. (1990). On the classification of image features. *Pattern Recognition Letters*, **11**, 339-349.

Wang, Z. Q., Rao, K. R., & Ben-Arie, J. (1996). Optimal ramp edge detection using expansion matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**, 1092-1097.

Watson, A. B., Barlow, H. B., & Robson, J. G. (1983). What does the eye see best? *Nature*, **302**, 419-422.

Watson, A. B. & Robson, J. G. (1981). Discrimination at threshold: labelled detectors in human vision. *Vision Research*, **21**, 1115-1122.

Watt, R. J. & Morgan, M. J. (1983). The recognition and representation of edge blur: evidence for spatial primitives in human vision. *Vision Research*, **23**, 1465-1477.

Watt, R. J. & Morgan, M. J. (1984). Spatial filters and the localization of luminance changes in human vision. *Vision Research*, **24**, 1387-1397.

Watt, R. J. & Morgan, M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Research*, **25**, 1661-1674.

Weickert, J., Ishikawa, S., & Imiya, A. (1999). Linear scale-space has first been proposed in Japan. *Journal of Mathematical Imaging and Vision*, **10**, 237-252.

Wetherill, G. B. & Levitt, H. (1965). Sequential estimation of points on a psychometric function. *British Journal of Mathematical and Statistical Psychology*, **18**, 1-10.

Wichmann, F. A. & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, **63**, 1293-1313.

Wichmann, F. A. & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, **63**, 1314-1329.

Wiesel, T. N. (1960). Receptive fields of ganglion cells in the cat's retina. *Journal of Physiology*, **153**, 583-594.

Williams, D. J. & Shah, M. (1990). Edge contours using multiple scales. *Computer Vision, Graphics, and Image Processing*, **51**, 256-274.

Williams, D. J. & Shah, M. (1993). Edge characterization using normalized edge detector. *CVGIP: Graphical Models and Image Processing*, **55**, 311-318.

Williams, D. R., Brainard, D. H., McMahon, M. J., & Navarro, R. (1994). Double-pass and interferometric measures of the optical quality of the eye. *Journal of the Optical Society of America, A*, **11**, 3123-3135.

Wilson, H. R. & Bergen, J. R. (1979). A four mechanism model for spatial vision. *Vision Research*, **19**, 19-32.

Witkin, A. P. (1983). Scale-space filtering. In *Proceedings of the International Joint Conference on Artificial Intelligence* (pp. 1019-1022). Karlsruhe, Germany.

Wright, M. J. (1982). Contrast sensitivity and adaptation as a function of grating length. *Vision Research*, **22**, 139-149.

Wuerger, S. M., Owens, H., & Westland, S. (2001). Blur tolerance for luminance and chromatic stimuli. *Journal of the Optical Society of America, A*, **18**, 1231-1239.

Xing, J. & Heeger, D. J. (2000). Center-surround interactions in foveal and peripheral vision. *Vision Research*, **40**, 3065-3072.

Young, R. A. (1985). The Gaussian derivative theory of spatial vision: analysis of cortical cell receptive field line-weighting profiles. Technical Report GMR-4920, General Motors Research Laboratories, Computer Science Department, Warren, MI 48090-9055.

Young, R. A. (1986). The Gaussian derivative model for machine vision: visual cortex simulation. Technical Report GMR-5323, General Motors Research Laboratories, Computer Science Department, Warren, MI 48090-9055.

Young, R. A. & Lesperance, R. M. (2001). The Gaussian derivative model for spatial-temporal vision: II. Cortical data. *Spatial Vision*, **14**, 321-389.

Young, R. A., Lesperance, R. M., & Meyer, W. W. (2001). The Gaussian derivative model for spatial-temporal vision: I. Cortical model. *Spatial Vision*, **14**, 261-319.

Yu, C., Klein, S. A., & Levi, D. M. (2001). Surround modulation of perceived contrast and the role of brightness induction. *Journal of Vision*, 1, 18-31.

Yuille, A. L. and Poggio, T. A. (1983). Scaling theorems for zero-crossings. MIT AI Memo 722.

Yuille, A. L. & Poggio, T. A. (1986). Scaling theorems for zero-crossings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8, 15-25.

Zenger, B. & Sagi, D. (1996). Isolating excitatory and inhibitory nonlinear spatial interactions involved in contrast detection. *Vision Research*, 36, 2497-2513.

Zhang, W. & Bergholm, F. (1997). Multi-scale blur estimation and edge type classification for scene analysis. *International Journal of Computer Vision*, 24, 219-250.